UNIVERSIDADE ABERTA

HOW CAN DATA MINING SUPPORT THE DECISION-MAKING PROCESS IN THE BEVERAGE INDUSTRY

Tania Regina Fernandes

Master´s Dissertation in Management

2019

UNIVERSIDADE ABERTA

HOW CAN DATA MINING SUPPORT DECISION-MAKING PROCESS IN THE BEVERAGE INDUSTRY

Tania Regina Fernandes

Master´s Dissertation in Management

Dissertation supervised by Prof. Maria do Rosário Matos Bernardo

2019

Resumo

Os dados têm vindo a tornar-se essenciais em todos os aspetos da nossa vida e vieram alterar a forma como vivenciamos as pessoas, os negócios e o ambiente que nos rodeia. Atualmente, os dados são usados para: desenvolver carros autónomos, construir cidades inteligentes, desenvolver aplicações móveis entre outros, melhorando as nossas interações do dia-a-dia. O IDC (2017) estima que até 2025 os dados no mundo irão crescer para 163 zeta bytes, representando dez vezes mais os 16.1 ZB de dados criados em 2016 (Reinsel, Gantz, & Rydning, 2017). Contudo, o universo digital, isto é, a soma de todos os dados criados, colecionados ou replicados, tem crescido exponencialmente, resultando num aumento de complexidade no que toca ao processamento, armazenamento, gestão e segurança dos dados e da informação. O desafio para os gestores de negócios está em focar-se nos dados mais importantes e úteis através da identificação de subconjuntos de dados únicos que os possam apoiar a criar impato na experiência do consumidores e ajudar a resolver problemas complexos. A presente investigação é um caso de estudo onde o objetivo é perceber como o uso da prospeção de dados (Data Mining) pode ser usado no setor das bebidas para gerar novos insights e apoiar o processo de tomada de decisão. A metodologia usada no projeto é uma investigação exploratória que recorre a métodos mistos para obter uma visão holística do problema. O estudo consiste num questionário realizado numa empresa internacional no setor das bebidas, acompanhado por uma revisão bibliográfica sobre o tema bem como dados de relatórios específicos do setor. Os resultados não podem ser representativos da empresa e da indústria devido à baixa percentagem de respostas, mas sugerem que o setor lida com enormes volumes de dados e está a construir capacidades a nível de Big Data. Também se demonstrou que algumas empresas enfrentam dificuldades na implementação do conhecimento adquirido dentro do processo de decisão indicando necessidade de futura investigação nesta área. Este estudo sugere que através do uso de Data Mining e Big Data as empresas podem compreender melhor o ambiente que as rodeia, aumentar a velocidade de decisão e melhorar as decisões táticas e estratégicas. Para tal, as empresas devem combinar recursos humanos, técnicas analíticas, cultura organizacional e liderança para extrair conhecimento útil dos dados e melhor avaliar oportunidades, o que poderia resultar em vantagem competitiva.

Palavras-chave: Data Mining, Sistemas de Suporte à Decisão (DSS), Análise de Big Data, descoberta de conhecimento accionável (AKD), Indústria das bebidas

Abstract

Data is becoming essential in all aspect of life and it has changed how we experience people, business and our surrounding environment. Data is now being used to, among other things to develop autonomous cars, create humanoid robots, build smart cities and smart homes, develop apps, improving our day-to-day living interactions. IDC (2017) estimates that by 2025 the world data will grow to 163 zettabytes, representing ten times the 16.1ZB data generated in 2016 (Reinsel et al., 2017). However, the digital universe, defined as the sum of all data created, captured and replicated, is expanding exponentially and it has resulted in increased complexity in to process, store, manage and secure data and information. The challenge for business leaders and managers is now to focus on the most important and useful data by identifying unique subsets of data which can help them impact the consumer experience and solve complex business problems. This research project is a case study where the goal is to understand how data mining could be used in the beverage industry to generate new insights and to support the decision-making process. The research methodology used is an exploratory research which will use mixed research methods to obtain a holistic view of the problem. The study performs a representative survey in an international company within the beverage industry supplemented by an academic bibliographic review on the subject and complementing it with industry specific reports on the beverage industry from the world leading consulting firms. The results could not be representative of the company in study or the whole industry due to the low response rate, but they suggest the beverage industry is already collecting, storing and managing vast amounts of data and building big data analytics capabilities. It also shows some companies struggle to implement the insights from analytics into the tactical or strategic business decisions indicating further research should be conducted to address that. This study suggests that by using data mining and big data companies could better understand their environment, increase the speed of decisions and improve tactical and strategic decision-making. To achieve this, companies in the beverage industry should weave talent, analytical tools, organizational culture and leadership to further derive useful insights from data and to be able to better seize opportunities, which could result in competitive advantage.

Keywords: Data Mining, Decision Support Systems, Decision-Making Process, Big Data Analytics, Actionable Knowledge Discovery, Beverage Industry

"Consumer data will be the biggest differentiator in the next two to three years. Whoever unlocks the realms of data and uses it strategically will win."

Angela Ahrendts, Apple

"In God we trust, all others bring data."

W. Edwards Deming

Table of Contents

Index of Tables

Index of Figures

List of Abbreviations and Acronyms

AKD	“Actionable knowledge discovery”
BDA	“Big Data Analytics”
BDA-DMF	“Big Data Analytics and Decision-Making Framework”
CIO	“Chief Information Officer”
DDD	“Data-driven decision-making”
DDDM or D^3M	“Domain-Driven Data Mining”
DM	“Data Mining”
DSS	“Decision Support Systems”
ETL	“Extraction, transformation, and loading “
IDC	“International Data Corporation”
IoT	“Internet of Things”
IS	“Information Systems “
IT	“Information technology”
KBV	“Knowledge Based View”
MNEs	“Multinational Enterprises”
OOS	“Out-of-stocks”
PCI	“Perceived Characteristics of Innovation”
POS	“Point of Sale”
ROI	“Return on Investment”
RQ	“Research Question”
SD	“Standard Deviation”
SKU	“Stock Keeping Unit”
UAb	“Univesidade Aberta”

Introduction

The current broad availability of data throughout organizations and the advancements in data storage technologies has led to the increasing interest in methods for extracting useful information and knowledge from data (Provost & Fawcett, 2013). Shim et al also agree that the current rapidly expanding volume of real-time data from the internet and e-commerce, has contributed to the demand for data mining tools (Shim et al., 2002). Data mining (DM), i.e. the extraction of hidden predictive information from databases, is a powerful new technology with potential to help companies focus on the most relevant information in customer data, transactions and in warehouses (Lodhi, 2012).

This information can be used to make better decisions and to create new knowledge the managers can act upon. Provost and Fawcett (2013) assert that “there is convincing evidence that data-driven decision-making and big data technologies improve business performance” (Provost & Fawcett, 2013:17), but this requires a “close interaction between the data scientists and the business people responsible for the decision-making” (Provost & Fawcett, 2013:13). Ichija & Ikujiro add that the success of a company will be determined by the extent to which its leaders can develop intellectual capital through knowledge creation and knowledge-sharing (Ichijo & Nonaka, 2006). Moreover, considering that knowledge as a resource can become obsolete in the future, new knowledge must be created on a continued basis. New capture, search, discovery, and analysis tools, like data mining, can help organizations gain insights from their unstructured data, and can generate metadata (data about data) automatically. According to Gantz & Reinsel, metadata “ is growing twice as fast as the digital universe as a whole” (Gantz & Reinsel, 2011:2). For Hormozi & Giles data mining is proving to be a valuable tool, because it identifies potentially useful information from the large amounts of data collected and enables an organization to gain an advantage over its competitors (Hormozi & Giles, 2004).

Researchers need to study and document use cases that explain how specific, novel data, so-called Big Data, can be used to support decision-making (Power, 2013). Piatetsky-Shapiro (2000) gives a good example to characterize current data mining research area: “we see many papers proposing incremental refinements in association rules algorithms, but very few papers describing how the discovered association rules are used” (Pechenizkiy, Puuronen, & Tsymbal, 2008:246). But even as data analytics comes of age, many companies aren’t seeing the return on investment (ROI) they expected as “they struggle to move from employing analytics in a few successful use cases to scaling it across the enterprise, embedding it in organizational culture and everyday decision-making" according to McKinsey, a global consulting company (McKinseyAnalytics, 2018:3). Erik Brynjolfsson et al and Penn's Wharton School conducted a study on how data-driven decision-making (DDD) affects firm performance (2011) by developing a measure that rates companies by how strong they use data for decision making and they show that statistically, the more data driven the firm is, the more productive it is – even controlling for a wide range of possible confounding factors: one standard deviation higher on DDD scale is associated with a 4% -6% increase in productivity (Provost & Fawcett, 2013).

In this research project, the focus will be on how to leverage data mining to support the decision-making process using the beverage sector as a case study. The beverage sector is part of the consumer-packaged-goods (CPG), and these companies have long had access to vast amounts of transaction data, since every day companies capture information about every stock keeping unit (SKU) sold to every customer at every store, every machine or through e-commerce (Breuer & Moulton, 2013). A type of descriptive analysis will be undertaken to understand the practice of business analytics in the industry and empirical techniques like surveys and interviews will be used for exploration. This is an exploratory research, where the data used by the company will be used as representative sample of the industry.

The methodology used includes mixed research methods (quantitative and qualitative) these allow: (1) Completeness, by providing better overall framework of the phenomenon; (2) Compensation as the mixed methods compensate the weakness of the use of only one of the methods quantitative/qualitative; (3) Diversity, using mixed methods to gain different insights on the same phenomenon (Venkatesh et al, 2013). This research methodology uses sequential mixed methods for the data collection, analysis and presentation. Firstly, a qualitative analysis of survey data is performed and presented by using column and bar charts and pie charts and, where applicable, a quantitative analysis of the results is done applying descriptive statistics on the survey results. To complement the mixed methods, NVivo software - a leading tool for qualitative analysis results that helps collect and analyze text-based research data- was used to analyze the survey results and identify sentiment and main themes.

The structure and organization of the present research project reflect the main goal of this study which focuses on understanding how data mining can support the decision making in the beverage industry. The paper is divided in three chapters, where Chapter I provides the theoretical framework and clarification of the main concepts that relate and help to explain the research object, mainly the definition of data mining and presentation of data mining models and tools, an overview of the decision making process and definition of Decision Support Systems (DSS), Actionable Knowledge Discovery (AKD) and finally, a definition of Big Data and an overview of the Information Value Chain.

Chapter II presents the case study, starting with the presentation of the industry and the case-study company. Then, a brief description of the research methodology is done, including data collection techniques and data analysis, followed by the presentation and interpretation of the result based on the theoretical framework presented in Chapter I. It concludes by referring some existing used cases in the beverage industry found during bibliometric research. In Chapter III, final considerations are done addressing some insights on how to create value from data mining, discussing current challenges and limitations, and possible areas for research.

CHAPTER I - Theoretical Framework

1.1. Data Mining and Knowledge Discovery in Databases

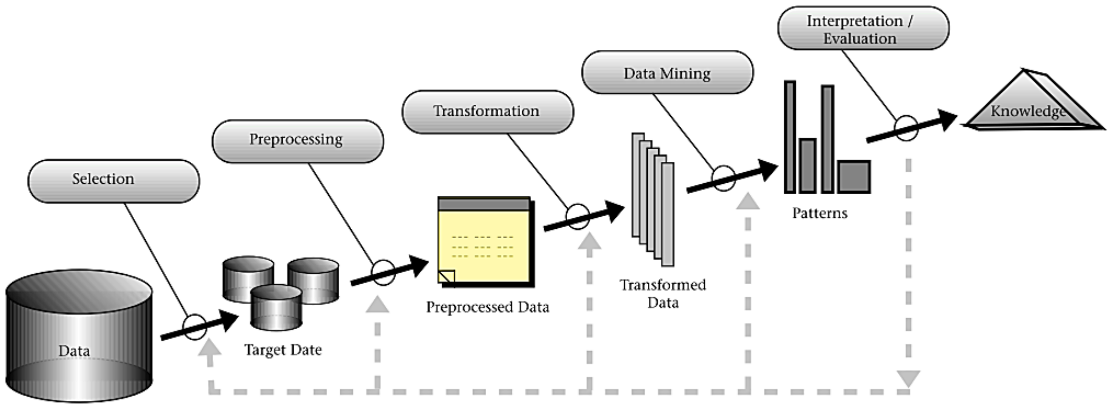
Organizations in general use data to get competitive advantage, to increase efficiency, and to provide more valuable services to customers. The data captured about the surrounding environment are the basic evidence used to build theories and models of the world and because computers have enabled humans to gather more data than they can process, they turn to computational techniques to help discover meaningful patterns and structures from massive volumes of data (Fayyad et al., 1996). Knowledge Discover in Databases (KDD) is an attempt to answer a challenge that was brought by the digital information: data overload.

Data mining can be “defined as the identification of interesting structure in data” (Fayyad & Uthurusamy, 2002:28). The goal of data mining is to produce new knowledge that the user can act upon and it does this by building a model of the real world based on data collected from a variety of sources, which may include corporate transactions, customer histories and demographic information, process control data and relevant external databases such as credit bureau information or weather data, etc. (Two Crows Corporation, 2005). Data mining is mainly concerned with making it easy, convenient, and practical to “explore very large databases for organizations and users with lots of data but without years of training as data analysts” (Fayyad & Uthurusamy, 2002:30).

The Gartner group defines data mining as the process of discovering meaningful correlations, patterns and trends by shifting large amount of data stored in repositories, using pattern recognition technologies as well as statistical and mathematical techniques (Gartner, 2018). For other scholars data mining is “a step in the KDD process that consists of applying data analysis and discovery algorithms that, under acceptable computational efficiency limitations, produce a particular enumeration of patterns (or models) over the data” (Fayyad, Piatetsky-Shapiro, & Smyth, 1996).

As shown in Figure 1.1, the KDD process involves using the database for the selection, preprocessing, subsampling, and transformation of data; then applying data mining methods (algorithms) to enumerate patterns from it; and finally, evaluating the products of data mining to identify the subset of the enumerated patterns deemed knowledge (Fayyad et al., 1996:41). Data mining is a major step in this process and involves the application of particular algorithms to identify relevant structures in data, where the structure designates patterns, statistical or predictive models from the data, and the relationships among parts of the data (Fayyad & Uthurusamy, 2002).

Figure 1.1 Overview of the Steps that compose the KDD Process



Source: Fayyad et al. (1996), p.41

In the context of data mining, the terms patterns and models have a concrete definition. For Fayyad & Uthurusamy a pattern "is a parsimonious summary of a subset of the data (e.g. people who own minivans have children), a model of data is a model of the entire data set and it can be predictive" (e.g. predict customer behavior by looking at historical data of interaction with the company) (Fayyad & Uthurusamy, 2002:28) and the data mining task is to identify intesting struture and useful patterns amoung multiple possibilities, using algortihms, and perfom this quicky over very large databases (Fayyad & Uthurusamy, 2002). The main goals addressed by data mining fall into specific categories presented in the Table 1.1 below:

Table 1.1 Different categories of goals of data mining adapted from Fayyad & Uthurusamy

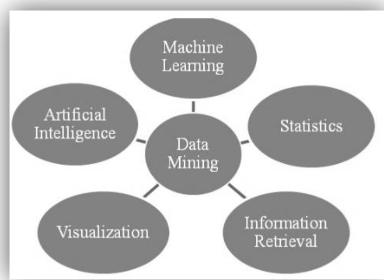
DM Goal	Description
Scaling analysis to large databases	Can abstract data-access primitives embedded in databases systems provide mining algorithms with information to drive a search for patterns?
	How might we avoid having to scan an entire very large database while reliably searching for patterns?
DM Goal	**Description**

Scaling to high-dimensional data and models	A model derived from automated discovery and search process can be used to find lower-dimensional subspaces where people find it easier to understand aspects of the problem that are interesting
Automating search	Besides replying on human analyst to list and create hypothesis, the algorithm performs much of this tedious and data-intensive work automatically
Finding patterns and models understandable and interesting	Methodologies for scoring models can be done based on notions of accuracy (how well a model predicts data) and utility (how to measure the benefit of a derived pattern, e.g. money saved)-
	Data mining focus on additional measures as well, such as the understandability of a model, the novelty of a pattern or how to simplify a model for interpretability. More importantly, the algorithm should help users gain insight by focusing on the extraction of pattern that are easily understood and turned into meaningful reports and summaries, by trading off complexity for understandability

Source: Fayyad & Uthurusamy 2002, p.30

DM and KDD bridge many technical areas, Figure 1.2 presents data mining as an interdisciplinary area which involves databases, machine learning, pattern recognition, statistics, and visualization among others (Newton & Singh, 2013; Philip Chen & Zhang, 2014).

Figure 1.2 Association of data mining with other fields



Source: Kautkar (2014), p.187

The data mining component often involves repeated iterative application of DM methods (Fayyad et al., 1996) and fitting models to, or determining patterns from, observed data. Fayyad et al state that the fitted models play the role of inferred knowledge: Whether the models reflect useful/interesting knowledge is part of the overall KDD process where subjective human judgment is also required (Fayyad et al., 1996).

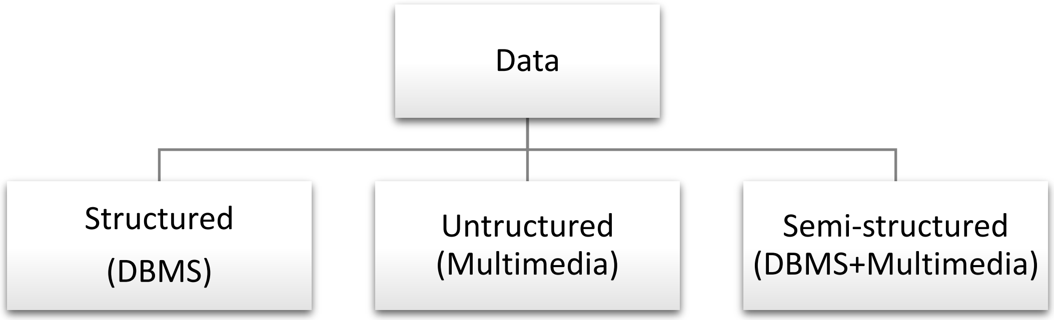
Even though the definitions of data mining are somewhat different, they all have the same idea: to extract important information from existing data and enable better decision making throughout an organization. Data mining can improve decision making by searching for relationships and patterns from the extensive data collected by organizations and by reducing the information overload (Hormozi & Giles, 2004), it enables an organization to focus on the most important information in the database, which allows managers to make more knowledgeable decisions by predicting future trends and behaviors (Hormozi & Giles, 2004). For example, customer acquisition can be targeted appropriately using data mining to find patterns in a customer database or data mining can be used for customer retention, by identifying customers who are likely to leave allowing the company to offer those customers incentives to stay. If a company knows the customers who are most profitable or more likely to respond positively to an offer, there is less chance of losing those customers to competitors (Hormozi & Giles, 2004).

Types of data

Databases are increasing its size in two ways: the number N of records or objects in the database and the number of fields or attributes to an object. Databases containing on the order of N = 109 objects are becoming increasingly common (Fayyad et al., 1996). As we have no capacity to digest millions of records, each having tens or hundreds of fields, the analysis work needs to be automated, at least partially, as mentioned by Fayyad et al (1996).

Besides the data volume, there is another challenge in the data related the different data types now accessed by organizations. Figure 1.3 depicts the types of data like structured, semi-structured and unstructured. Structured data is usually data stored in Database Management System (DBMS), and it can be easily organized and retrieved. Sources of structured data are: (1) Machine Generated Data, such as Sensory Data (GPS data, manufacturing sensors, etc.), Point-of-Sale data (location of sale, product information), Web Server Logs (page requests, other server activity) and (2) Human generated data, which is any data inputted into a computer (dates, customer names, addresses, product names and numbers). Unstructured data refers to Multimedia and it contains many different data formats like text files, email, Social Media, mobile data, audio and videos files, images etc. and it not easily retrievable. The semi-structured data is a combination of the previous two.

Figure 1.3 Types of data



Source: Kautkar, (2014), p.186

To gratify the demand of analyzing the available large volume data sets, more efficiently the data mining employs diverse fields like Machine learning, Information retrieval, Statistics, and Visualization as referred by Kautkar (2014). Moreover, from the data available, it is important to differentiate between Big Data and Small Data. To this intention, Poleto et al (2015) proposed a table with the main differentiators, presented in Table 1.2 below.

Table 1.2 Difference between Big Data and Small Data

Aspects	Big Data	Small data
Goals	Projected from a predetermined goal and have a greater level of flexibility, considering the context of the problem. (e.g. market scenario analysis to identify forms of accelerate sales)	Designed to answer a specific question and control in a particular context. (e.g. inventory control, by getting only information concerning the entry and exit of goods, acting based on current market)

Aspects	Big Data	Small data
Data location	Aggregates data spread across different media, i.e. in Internet servers. The architecture consists of a distributed computing where multiple servers work together to store and process information. It has high power scalability, low cost of implementation	In general, data come from internal organization and data files
Data structure	Able to absorb unstructured data, i.e. Big Data, and work with many variables simultaneously as well as rendering images in minimal time and efficiently.	Usually contain structured data- data represented by uniform records in orderly spreadsheet. (e.g. enterprise resource planning (ERP) with a predefined architecture and a structured way to work with data
Data preparation	Data collected from different sources and prepared by several users. People who use the data are rarely those who prepare it and therefore different organizational roles contribute to disseminate information	In general, data users prepare their own data for their own purposes and according their specific context
Analysis	Is done in incremental steps. First data is extracted, revised, normalized, processed, visualized, interpreted and then analyzed with different methods. (e.g. use of complex techniques of data analysis combining data mining and artificial intelligence)	In principle, all data for the project can be analyzed all at once. The structure is pre-defined and based on the specific context. (e.g. use of Structured Query Language (SQL) combined with programming languages to create procedures to mining and analyze data)

Source: Poleto, de Carvalho & Costa (2015), p.13

1.1.1. Data Mining Methods and Algorithms

This section presents a description of the data mining methods and the algorithms that incorporate these methods. Fayyad (1996) note that, although different algorithms and applications might look different on the surface, it is usual to find out many share common components. Understanding data mining and model induction at this component level clarifies the behavior of any data-mining algorithm, but each technique typically suits some problems better than others and there is no universal data-mining method according to Fayyad (Fayyad et al., 1996). Moreover, a large portion of the application effort can go into properly formulating the problem rather than into optimizing the algorithmic details of a particular data-mining method (Langley & Simon, 1995).

For Fayyad et al (1996) the data mining methods can be seen as extensions of a few basic techniques and principles since the actual underlying model representation being used by a particular method typically comes from a composition of a small number of well-known options: polynomials, splines, kernel and basic functions, threshold-Boolean functions, and so on (Fayyad et al., 1996). The mathematical formalisms used in model fitting are statistical, which allow both nondeterministic effects in the model and logical (deterministic). The most widely used basis for data mining applications is the statistical approach due to the presence of uncertainty in real-world data-generating processes.

There are two high-level primary goals of data mining which are prediction and description:

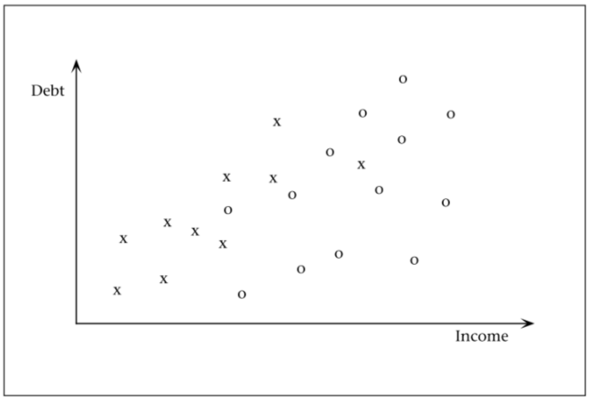
- **Prediction** involves using some variables or fields in the database to predict unknown or future values of other variables of interest and to infer patterns, and
- **Description** focuses on finding human-interpretable patterns describing essential characteristics or general properties of the data (Fayyad et al., 1996; Kannimuthu, Premalatha, & Usha, 2015).

Predictive mining consists of tasks like **Classification, Regression** and **Deviation detection**, which are presented in detail below. These models are often judged by the empirical prediction accuracy on some test set.

To illustrate the main models, a bank dataset has been classified into two classes: the x’s representing persons who have defaulted on their loans and the o’s represent persons whose loans are in good status with the bank, as shown in Figure 1.4. This simple artificial data could contain useful knowledge from the point of view of a bank making loans.

The examples presented are merely representative since “in actual KDD applications, there are typically many more dimensions (as many as several hundreds) and many more data points (many thousands or even millions)” (Fayyad et al., 1996:43).

Figure 1.4 Simple bank data set with two classes for illustrative purposes

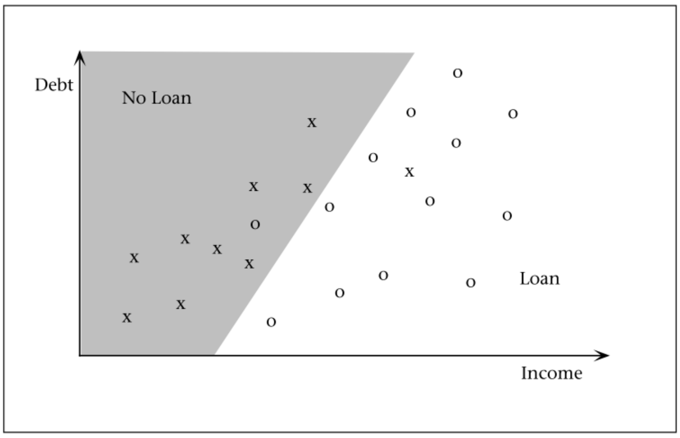


Source: Fayyad et al. (1996), p.43

Classification

Refers to learning a function that classifies or maps a data item into one of various predefined classes (Weiss and Kulikowski 1991; Hand 1981 apud Fayyad et al., 1996) Examples of classification methods used as part of KDD applications include classifying trends in financial markets (Apte and Hong 1996 apud Fayyad et al., 1996). Figure 1.5 shows a simple Linear Classification Boundary for the example of loan data set.

Figure 1.5 Simple Linear Classification Boundary for Loan Data Set. (The shaped region shows class no loan)

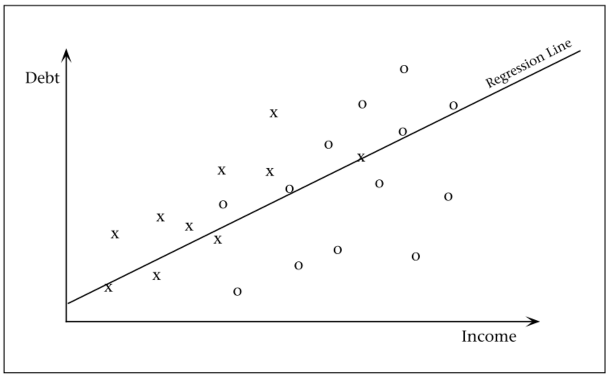


Source: Fayyad et al. (1996), p.44

Regression

Refers to a learning a function that maps a data item to a real-valued prediction variable and it allows to model a relationship between a single continuous variable (outcome) and one or more explanatory variables (predictors). It does this by constructing a line or a surface that minimizes the distance between outcome and predictors. Some examples of regression applications are predicting consumer demand for a new product as a function of advertising expenditure or predicting time series where the input variables can be time-lagged versions of the prediction variable (Fayyad et al., 1996). A graphical representation is presented in Figure 1.6.

Figure 1.6 Simple Linear Regression for the Loan Data Set



Source: Fayyad et al. (1996), p.44

Change and deviation detection

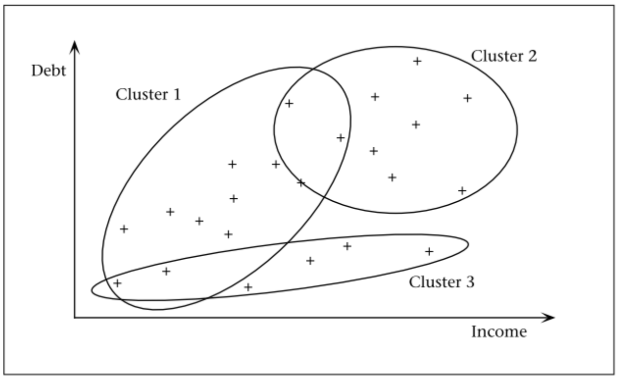
Is concerned with discovering the most significant changes in the data from previously measured or normative values (Berndt and Clifford 1996; Guyon, Matic, and Vapnik 1996; Kloesgen 1996; Matheus, Piatetsky-Shapiro, and McNeill 1996 apud Fayyad et al., 1996).

On the other hand, descriptive mining includes tasks like **Clustering, Association** and **Sequence analysis**. These models can be evaluated along the dimensions of predictive accuracy, novelty, utility, and understandability of the fitted model.

Clustering

The goal is to identify a finite set of categories or clusters to describe the data. The categories can be mutually exclusive and exhaustive or consist of a richer representation, such as hierarchical or overlapping categories (Jain and Dubes 1988; Titterington, Smith, and Makov 1985 apud (Fayyad et al., 1996)). Examples in KDD applications include discovering homogeneous subpopulations for consumers in marketing databases. Figure 1.7 shows a possible clustering of the loan data set into three clusters; note that the clusters overlap, allowing data points to belong to more than one cluster.

Figure 1.7 Simple Clustering of Loan Data Set, p.44 (Fayyad et al., 1996)



Source: Fayyad et al. (1996), p.44

Association rules

These models basically give the relation between different attributes and their values. Association rule utilizes the concept of support and confidence to determine the usefulness of rule. Patterns are practicable for probing the buying habits of customer for business analysis (Kautkar, 2014).

Sequence analysis

In sequential pattern mining, the model will, given a set of sequences, find the complete set of frequent subsequences (for example, satisfying the min_sup thereshold) (Han, 2018). Examples of sequence data include DNA, protein, customer purchase history, web surfing history, and more (G. Dong, 2009).

One important consideration after the model choice, is the relevance of attributes, i.e. choose data attributes that are relevant to the discovery task; there is no amount of data that will allow prediction if the attributes they are based on do not capture the information needed. Furthermore, considerations need to be done with regards to maintaining low noise levels (few data errors) since high amounts of noise make it hard to identify patterns (Fayyad et al., 1996).

Another important component is prior knowledge, which is useful to know something about the domain (e.g. What are the important fields; What are the likely relationships; What is the user utility function; What patterns are already known? etc.).

Once the model is selected, the next step is to construct specific algorithms to implement the general methods outlined. Fayyad et al identify three primary components in any data-mining algorithm: model representation, model evaluation, and search (Fayyad et al., 1996).

Table 1.3 Components of data mining algorithms adapted from Fayyad et al

Components of data mining algorithms	Description
Model representation	It is the language used to describe discoverable patterns. The analysts must fully understand the representational assumptions inherent in a method and an algorithm designer should clearly state which assumptions are being made by the algorithm. Fayyad et al note that increased representational power for models can also increase the danger of overfitting the training data – leading to reduced prediction accuracy
Model evaluation	The criteria are quantitative statements (or fit functions) of how well a pattern (a model and its parameters) meets the goals of the KDD process
Search	Consists of two components: parameter search and model search. In parameter search, the algorithm must search for the parameters that optimize the model-evaluation criteria given observed data and a fixed model representation; In model

	search, the algorithms must find models from the selected family that optimizes the evaluation criteria.

Source: Fayyad et al. (1996), p.45

Finally, it should be clear that there are no established criteria for deciding which data mining methods to use in which circumstances, and

> "many of the approaches are based on crude heuristic approximations to avoid the expensive search required to find optimal, or even good, solutions. Hence, the reader should be careful when confronted with overstated claims about the great ability of a system to mine useful information from large (or even small) databases"
>
> (Fayyad et al., 1996:49).

1.1.2. Data Mining Tools and Technologies

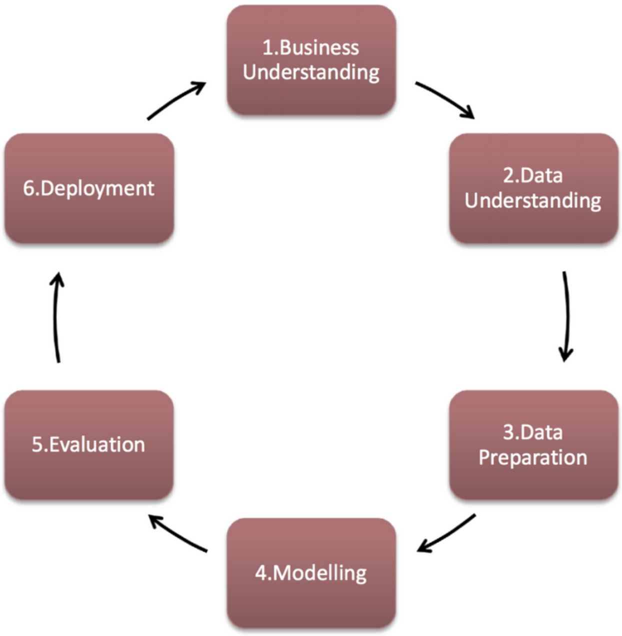
Data mining tools can be classified into one of three categories (Newton & Singh, 2013):

- **Traditional Data Mining Tools.** Traditional data mining programs help companies establish data patterns and trends by using several complex algorithms and techniques;
- **Dashboards**. Installed in computers to monitor information in a database, dashboards reflect data changes and updates onscreen — often in the form of a chart or table — enabling to see overview of the company's performance.
- **Text-mining Tools**. Ability to mine data from different kinds of text — from Microsoft Word and Acrobat PDF documents to simple text files, for example

There are major challenges remaining for data warehouse content management, because the extraction, transformation, and loading (ETL) functions in a data warehouse are considered the most time-consuming and expensive portion of the development lifecycle (Srivastava and Chen 1999 apud Newton & Singh, 2013). But the developments in database management and increase of data volumes being accumulated also brought interest in DM. Even though there are numerous DM algorithms being developed and others being improved, there is relatively little on-going research directed to development of DM Frameworks (Pechenizkiy, Puuronen, & Alexey, 2005:2). According to Pechenizkiy et al, the existing DM Frameworks are:

- Theory-oriented framework – mainly based on statistical paradigms, or on the data compression paradigm (“compress the dataset by finding some structure or knowledge for it”), the machine learning paradigm (“let the data suggest a model” that) and the database paradigm (“there is no such thing as discovery, it is all in the power of the query language”);
- Process-oriented framework – DM is seen as a sequence of iterative processes that include data cleaning, feature transformation, algorithm and parameter selection, evaluation, interpretation and validation; For example, cross-industry process for data mining (CRISP-DM) is an industry standard for DM Projects (Figure 1.8);
- Foundations-oriented frameworks – try to search for mathematical bricks for DM.

Figure 1.8 Cross-industry standard process for data mining, CRISP-DM



Source: Smart Vision (2018), p.1

This diversity in theoretical foundations and research methods is a good thing according to Pechenizkiy et al (2005) and the suggest it would be more reasonable to search for an umbrella-framework that would cover the existing theory (Pechenizkiy et al., 2005:2).

1.1.3. Actionable Data Mining

Data mining can be described as the process of turning data into information, then information into action, and finally action into value (Ling, Chen, Yang, & Chen, 2002; Yang, Yin, Lin, & Chen, 2003). The data mining process includes a series of steps or stages from data cleaning, selection and transformation to pattern evaluation and visualization, but one of the main challenges in "DM is how to turn the mined patters and knowledge into actionable insights" (He, Xu, & Deng, 2005:1).

He et al (2005) state that there are two major terms in data mining research and applications which are patterns and interest. Patterns can be discovered through the use of classification, association or clustering; while interest refers to patterns that are useful or meaningful in business applications, for example, an "important measure of interestingness is whether it can be used in the decision making process of a business to increase its profit" (He et al., 2005:1).

In general, the goal of data mining is taking actions to make profit for an organization, but "current post-processing techniques are limited to producing visualization results, not directly suggesting actions" (He et al., 2005:2). On the other hand, there are often many patterns mined but "business people are not interested in them or do not know what follow- up actions to take to support their business decisions" (He et al., 2005:1), this fact that might have hindered the widespread employment of advanced data mining techniques in promoting operational quality and productivity in organizations according to Cao et al (Cao et al., 2010).

The development of effective approaches for pattern discovery should be of "technical significance", but they should also "satisfy business expectations" and "indicate the possible actions" that can be taken (Cao et al., 2010:1299). One of the great challenges mentioned by scholars in the development of the next generation of KDD methodologies and system is actionable knowledge discovery (AKD) (Ankerst, 2002).

Actionable knowledge discovery (AKD) is defined as "a closed optimization problem- solving process from problem definition, framework/model design to actionable pattern discovery, and it is designed to deliver operable business rules" that can be easily associated or integrated with business processes and systems (Cao et al., 2010:1299). "The term "actionability" measures the ability of a pattern to suggest a user to take some concrete actions to his/her advantage in the real world and it mainly measures the ability to suggest business decision-making actions" (Cao et al., 2010:1300)

The nature of the existing work on actionable interestingness development is mostly focused on technical-significance-oriented and little focus has been taken in enhancing KDD system infrastructure to address organizational and social complexities in real-world applications (Cao et al., 2010).

Furthermore, surveys of data mining for business applications have also demonstrated that business people cannot effectively take over and interpret the identified patterns for business use (Cao, Yu, & Zhang, 2008) which might result from several challenges such as, (1) the existence of many patterns mined which are not informative or clear to business people, (2) many of the identified patterns may be commonsense or of no specific interest for the business, (3) not a clear direction from managers on what actions can be taken on the discovered patterns in order to support business decision-making (Cao et al., 2010).

This denotes that there is a gap between academic deliverables and business expectations and between data miners and business analysts (Cao & Zhang, 2007a, 2007b), and it is therefore critical to develop general methodologies and techniques for actionable knowledge discovery (Cao et al., 2010). AKD deliverables should be business-friendly enough for business people to interpret, validate, action and be able to be "seamlessly embedded" into existing business processes and systems (Cao et al., 2010:1300).

If this happens, Cao et al (2010) state that data mining would have great potential to lead to productivity gain, smarter operation and decision making in business. And support in shifting the KDD paradigm from traditional technical aspects towards a more business-use-oriented and domain-driven actionable knowledge discovery (Cao, Yu, & Zhang, 2009).

Cao et al (2010) present some AKD frameworks (Table 1.4) from the system viewpoint and that follow the methodology of Domain-Driven Data Mining (DDDM or D^3M). Some of the these frameworks have shown to be effective and flexible for extracting actionable knowledge in complex real-world situation as well as supporting data mining practitioners with their business requirements and decision-making actions on findings and deliverables in business context (Cao et al., 2010:1300).

Table 1.4 Actionable Knowledge Discovery (AKD) Frameworks adapted from Cao et al 2010

Notations	Explanations	Description
PA-AKD	Post analysis-based AKD	It is a 2-step process, 1st, general patterns are mined based on technical significance; 2nd the learned patterns are filtered and summarized in terms of business expectations and converted into operationalizable business rules for business people's use
UI-AKD	Unified interestingness based AKD	Develops unified interestingness that aggregates and balances technical significance and business expectation. Mined patterns are then converted into deliverables based on domain knowledge and semantics
CM-AKD	Combined mining-based AKD	It is a multistep pattern mining on the data set according to a certain combination strategy. Mined patterns are fed into another mining procedure to guide its feature construction and its pattern mining. Finally, individual patterns identified from each step are then merged into final deliverables based on combination strategy, on domain knowledge and on business needs
MSCM-AKD	Multi-source + combined mining-based AKD	Deals with AKD in multiple data sources or in large quantities of data. One of the data sets is selected for initial pattern mining. Then, some learned patterns are selected to guide feature construction and pattern mining on the next data set(s). This interactive mining stops when all data sets are mined, and corresponding patterns are then merged/summarized into actionable deliverables

Source: Cao et al. (2010), p.1300

Concluding, data mining applications face significant issues with workability of the deployed algorithms, tools and resulting deliverables but these AKD frameworks support “closed-optimization-based problem solving from a business problem definition, assist in “actionable pattern discovery and in operable business rule conversion”, (Cao et al., 2010:1310) . Cao et el (2010) add that the deliverables extracted in this way are both of technical significance and capable of a smooth integration into business processes.

1.1.4. Uses of data mining

Data mining can have many uses, depending on the business problem and data availability, some of use cases are:

- **Market segmentation** – which analyze customer databases to identify different customer groups and forecast their behavior (Fayyad et al., 1996);
- **Customer churn** – aims to predict which customers are likely to leave some x-company and go to a competitor;
- **Direct marketing** - identify which prospects should be included in a mailing list to obtain the highest response rate;
- **Market basket analysis** – aims at understanding what products or services are commonly purchased together; which find patterns such as, “If customer bought X, he/she is also likely to buy Y and Z.” Such patterns are valuable to retailers (Fayyad et al., 1996).
- **Trend analysis** - uncover the difference between a typical customer present and future (Newton & Singh, 2013).

Some of the challenges faced by data mining currently are connected to:

(1) Larger databases: Databases with hundreds of fields and tables and millions of records and of a multigigabyte size are commonplace, and terabyte (1012 bytes) databases are beginning to appear;

(2) **High dimensionality**: Databases have large number of fields (attributes, variables); so, the dimensionality of the problem is high. Prior knowledge could help identify irrelevant variables.

(3) **Overfitting**: When the algorithm searches for the best parameters for one particular model, it can model not only the general patterns in the data but also any noise specific to the data set, resulting in poor performance of the model on test data. This could be addressed using cross-validation, regularization, and other sophisticated statistical strategies.

(4) **Assessing of statistical significance**: if a system tests models at the 0.001 significance level, then on average, with purely random data, N/1000 of these models will be accepted as significant. This point is frequently missed by many initial attempts at KDD. Fayyad et al refer that this could be overcome by using methods that adjust the test statistic as a function of the search, for example, Bonferroni adjustments for independent tests or randomization testing (Fayyad et al., 1996).

(5) **Changing data and knowledge**: Rapidly changing (nonstationary) data can make previously discovered patterns invalid. In addition, the variables measured in a given application database can be modified, deleted, or augmented with new measurements over time.

(6) **Missing and noisy data**: Important attributes can be missing if the database was not designed with discovery in mind (Fayyad et al., 1996)).

(7) **Understandability of patterns:** make the discoveries more understandable by humans by using graphic representations (Buntine 1996; Heckerman 1996), rule structuring, natural language generation, and techniques for visualization of data and knowledge.

(8) **User interaction and prior knowledge**: The use of domain knowledge is important in all the steps of the KDD process. Bayesian approaches, for example, Cheeseman (1990) use prior probabilities over data and distributions as one form of encoding prior knowledge (Fayyad et al., 1996)

(9) **Integration with other systems**: Typical integration issues include integration with a database management system (for example, through a query interface),

integration with spreadsheets and visualization tools, and accommodating of real-time sensor readings.

Coppock (2003) analyzed the failure factors of DM-involved projects and his main conclusion is that leadership, communications skills and an understanding of the culture of the organization are not less important than the traditionally emphasized technological job of turning data into insights (Pechenizkiy et al., 2005).

A good way to apply advanced data mining techniques is to have a flexible and interactive data mining tool that is fully integrated with a database or data warehouse. According to Newton & Singh (2013) when a data mining tool is integrated with the data warehouse, it simplifies the application of mining results and as the warehouse grows with new decisions and results, the organization can mine best practices continually and apply them to future decisions (Newton & Singh, 2013).

1.2. Decision Making Process

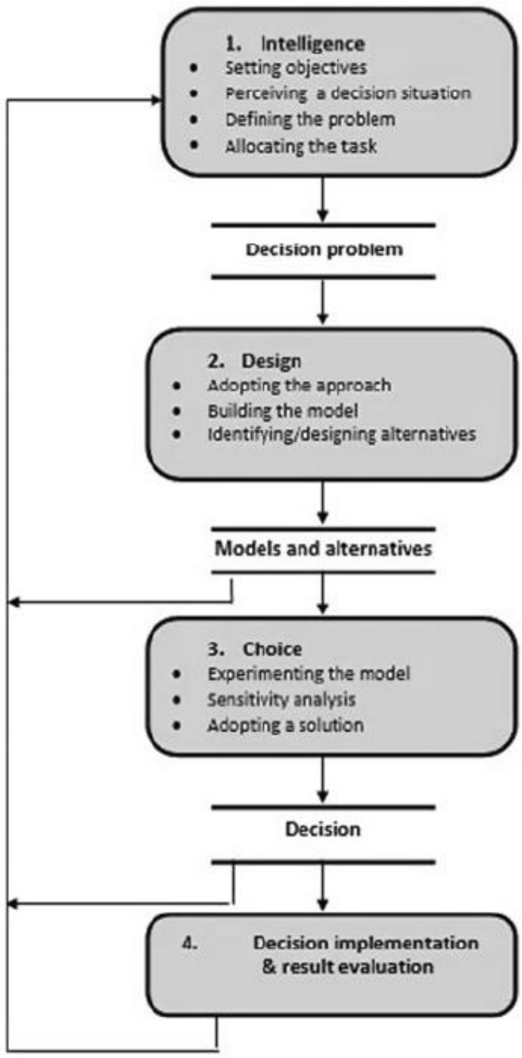
Information and knowledge are valuable assets for the decision-making process in an organization and they require an instrument to process data into valuable and useful information for use in organizational processes (Poleto, Heuer de Carvalho, & Cabral Seixas Costa, 2015). This instrument could be represented by Information Systems (IS), for example, Decision Support Systems (DSS) which is IS focused on the decision-making process. Over the years, various other tools and techniques have been used by businesses to support decision-making.

Herbert Simon (1960) studies aimed at understanding the processes that participate in human decision making and although he did not directly work on the concept of DSS much of his research can be regarded as theoretical foundation of DSS (Pomerol & Adam, 2004). Simon proposed a Process Model of Decision-Making (Figure 1.9), which was made up of four stages: Intelligence, Design, Choice and Implementation.

In the first stage of the decision process is Intelligence, where the problem and objectives are defined, the second is Design and it deals with building the model and identifying alternatives. In the third stage – Choice - a sensitivity analysis and model experimenting can be done, and a solution is chosen. In the final step comes the decision Implementation and result evaluation.

For Simon, the problems that trigger decision are not factual data, but constructs, and the decision "is a matter of compromise" since decision makers have several more or less contradictory objectives in mind, which highlights the multi-criteria aspect of decision-making (Pomerol & Adam, 2004). He also noted that each of the stages can be considered recursively as a decision in itself (Simon, 1977:43).

Figure 1.9 Process Model of Decision-Making proposed by Herbert Simon



Source: Filip, Zamfirescu, & Ciurea (2017), p.33

As Simon's studies gradually turned toward the computer, he introduced another aspect of decision: the difference between programmed decision and non-programmed decision (Simon, 1977) stating that,

> "decisions are programmed to the extent that they are repetitive and routine, to the extent that a definite procedure has been worked out for handling them so that they don't have to be treated from scratch each time they occur (…) decisions are non-programmed to the extent that they are novel, unstructured and unusually consequential"
>
> (Simon, 1977:46)

Each phase of Simon's decision model is susceptible to the use of methods and tools from organizational and tecnological perpectives according to Poleto et al (2015). In an organization, the information technology (IT) function is tasked with managing and integrating data as an "enabler" of data-driven business processes and decision making (Abbasi, Sarker, & Chiang, 2016).

In figure 1.10 Saggi & Jain propose various types of decision making data analytics , where **data mining** extracts the "knowledgeable data from organizational data by employing a broad variety of statistical approaches, data visualization, and the pattern recognition approaches" (Saggi & Jain, 2018:779).

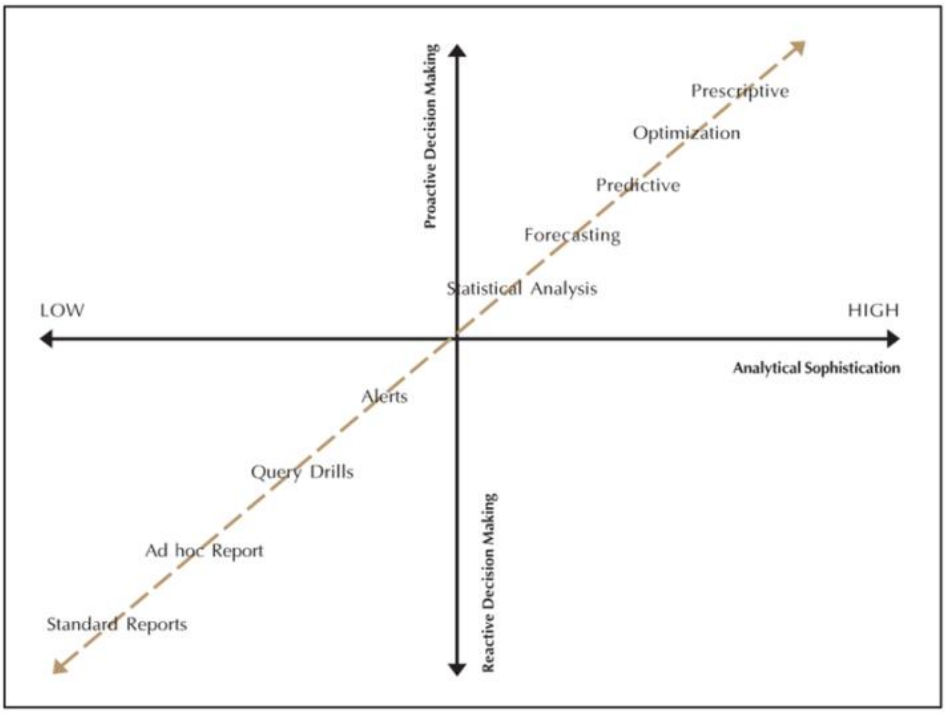
Figure 1.10 Types of decision-making data analytics, p 779 (Saggi & Jain, 2018)

Source: Saggi & Jain (2018), p.779

Predictive Modeling is used to identify organization threats and opportunities by developing patterns discovered in historical and transactional data; Optimization Modeling can help to define the best decision by offering various optimization models; Prescriptive Methods are used to create better solutions by exploiting the application of predictive models together with optimization technology; Business Intelligence is used to "increase and sustain a competitive edge by utilizing the modern techniques in data mining and analysis" (Saggi & Jain, 2018:779).

In the beginning, computer-assisted decision-making was based on finding out answers to questions such as what happened, how and where it happened (e.g. ad hoc reports, query, etc.), but with the advent of advanced analytics, businesses can now initiate proactive decision-making by using predictive and prescriptive analytics to answer questions on why it happened, what may happen next, and how it can be solved (Banerjee, Bandyopadhyay, & Acharya, 2013). Figure 1.11 demonstrates the various analytical tools that can be used for decision making, from reactive decision making to proactive decision making and from low to high analytical sophistication as presented by Banerjee et al.

Figure 1.11 The use of data analytics in Decision-making



Source: Banerjee et al. (2013), p.6

Banerjee et al assert that “the core activities of analytics in many organizations today remain primarily in the lower SW quadrant, while the aspiration would be to reach the NE quadrant many companies “are restricted due to the lack appropriate data in the organizations” (Banerjee et al., 2013:6). Recently, technologies have been developed to analyze such data (i.e., big data analytics) and these are now being used to inform decision-making.

– **Descriptive analytics** provides insights into what has happened in their internal and external environment (Chen et al., 2012; Chui et al., 2011);
– **Predictive analytics** predicts what will happen by using machine learning and algorithms to find patterns and capture relationships in multiple (un)structured data sources (Gandomi and Haider, 2015 apud M. van Rijmenam et al., 2018).
– **Prescriptive analytics** offer recommendations and uses a variety of algorithms and data modelling techniques to understand the environment and improve business performance (M. van Rijmenam et al., 2018).

Due to the decreasing costs of information technology, decision-making is becoming decentralized. Decentralized organizations are better positioned to benefit from big data analytics (Galbraith, 2014a apud M. van Rijmenam et al., 2018), as real-time insights enable anyone, not only executives, to make decisions rapidly, resulting in more agile organizations (M. van Rijmenam et al., 2018).

1.3. Decision Support Systems (DSS)

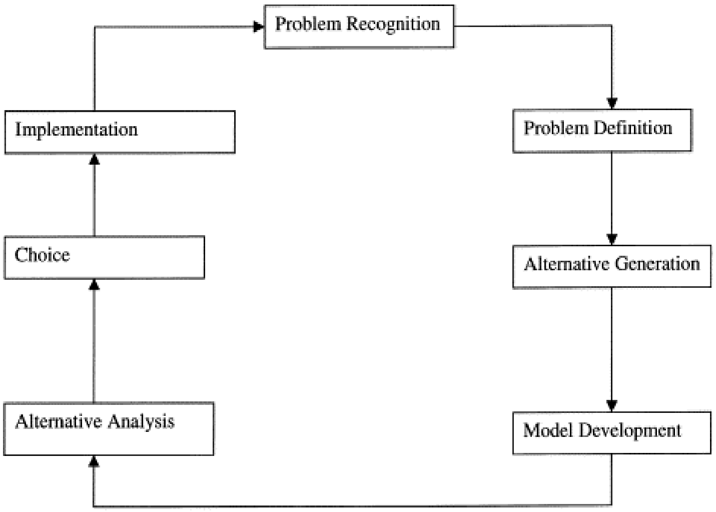
The term DSS has its origin in two streams: the original studies of Simon’s research team in the late 1950/60s and the technical works on interactive computer systems by Gerrity’s research team in the 1960s (Liu, Duffy, Whitfield, & Boyle, 2010). DSS are interactive, computer-based IS that help decision-makers utilize data, models, solvers, visualizations, and the user interface to solve semi-structured or unstructured problems. DSS are built using a DSS Generator (DSSG) as an assembling component (C.-S. J. Dong & Srinivasan, 2013).

A decision support process draws on selected enterprise transactional data from trusted source systems over regular intervals of reporting periods, such as quarterly over the fiscal year, for the purposes of data analysis and decision making (Moss 2003; Gandhi 2005; Kulkarni 2006; Turban 2006; Howson 2007; Watson 2008 apud (Kesner, 2010)).

The DSS has a strict link with intelligence – design - choice model proposed by Simon, but with more emphasis in the choice phase (Poleto et al., 2015). Their main objective is to support a decision by determining the most appropriate alternatives to solve the problem. Although the choice is made by a human agent (a manager, treated as a decision-maker within this process), the DSS role is to provide a friendly interface where the agents can build scenarios and simulate and obtain reports and visualizations to support their decisions (Daas, Overbeek, Bouwman, & Hurkmans, 2012; Newton & Singh, 2013). A good DSS must allow users to intuitively, quickly, and flexibly manipulate data using familiar terms, in order to provide analytical insight (Newton & Singh, 2013).

Simon based his reflection on the idea that organizations are systems designed for “complex information processing” (Simon, 1977:15), comparable to computers. He then proposed a distinction between programmed decisions and nonprogrammed decisions, by stating that "decisions are programmed to the extent that they are repetitive and routine, to the extent that a definite procedure has been worked out for handling them so that they don’t have to be treated from scratch each time they occur" (Simon, 1977:46), i.e. which obey computer programs or other programs that are computerizable and that can be modeled in DSS; Nonprogrammed decisions are so "to the extent that they are novel, unstructured and unusually consequential" (Simon, 1977:46), they come under the heading "problem solving" , and problems that cannot be modeled should be treated outside the realm of DSS (Simon, 1977:64).

Figure 1.12 DSS Decision making process Shim et al 2002



Source: Shim et al. (2002), p.113

According to Poleto et al (2015) each phase of Simon's decision model (Fig. 1.9) is susceptible to the use of methods and tools from organizational and tecnological perpectives. In Figure 1.12 Shim et al (2002) describe what probably came to be a more commonly used model for decision-making process in a DSS environment. The emphasis is put on the model development and on the problem analysis; once the problem is recognized, it is defined in terms that facilitate the creation of models (Shim et al., 2002). After that, alternative solutions are created, and models are then developed to analyze the various alternatives, and finally, the choice is made and implemented consistent with Simon's description (Shim et al., 2002). Since no decision process is clear-cut in an ill-structured situation, many times the phases can overlap and blend together, with frequent looping back to earlier stages as more is learned about the problem, as solutions fail, and so forth (Shim et al., 2002).

Decision support processes "assist managers in their tactical and strategic planning and management" (Kesner, 2010:5) since DSS serve many roles in organizations as they can be used to "forecast the consumption of raw material, factory capacity, labor hours, financial assets and so forth" (Kesner, 2010:4); DSS is commonly used to optimize operations within a fixed business setting (warehouse, distribution center, delivery area, or factory) and to "identify statistically significant patterns within large bodies of data to help discover possible causal relationships among the various moving parts of a business process, or a set of business indicators or metrics" (Kesner, 2010:4). It does this by using standard measures and benchmarks, to measure performance across business units, product lines, groups of employees and even individual contributors quickly and objectively.

According to Bâra & Lungu (2012) to build an efficient DSS there techniques and methods must be combined several to improve performance and accuracy of the analysis from two major perspectives: historical data and forecasts. This can be obtained by combining data warehousing, OLAP, data mining and business intelligence tools for analyzing and reporting into a flexible architecture that must contain:

- a data model's level, where an Extract-Transformation-Load (ETL) process must be applied to clean and load data into a data warehouse or data marts;
- an application level, with analytical models where multidimensional reporting like OLAP and data mining can be combined for historical and forecast analysis;
- an interface level, where dashboards and reports can be built with business intelligence tools (Newton & Singh, 2013).

The integration of data mining with decision support system, which involves joint data preprocessing, standards for model exchange and meta-learning can provide decision support when choosing the best data mining tools for a specific problem (Newton & Singh, 2013:2). The following tools can be used in integrating data mining and decision support:

- A pre-processing tool allowing access to various data sources;
- A common representation language supporting the exchange of data mining and decision support models for different application and visualization tools;
- Meta-learning tools for classifier selection and ROC methodology for model selection;
- The RAMSYS methodology for solving data mining and decision support problems, requiring remote collaboration of project partners (Newton & Singh, 2013).

However, according to Charest et al (2008) there is a lack of DM-DSS integration which can be exemplified by the "myriad of commercial data mining tools that provide limited support in the form of rudimentary "wizard-like" interfaces", these interfaces "include numerous, abstruse parameters and settings that mainly appeal to expert data miners, but that make hard assumptions about the level of background knowledge required by decision makers" (Charest et al., 2008:212).

Erik Brynjolfsson et al and Penn's Wharton School conducted a study on how data-driven decision-making affects firm performance (2011) and they show that statistically, the more data driven the firm is, the more productive it is – even controlling for a wide range of possible confounding factors: one standard deviation higher on DDD scale is associated with a 4% -6% increase in productivity (Provost & Fawcett, 2013).

1.4. Big data analytics (BDA) and Information Value Chain

Davenport (2013) stated that in the foreseeable future, analytics represents a primary arena for innovation and competition (Davenport, 2013) and others authors suggest that big data analytics have become a prerequisite to understanding the business environment and to remaining competitive (Bean, 2016; Siemens and Long, 2011 apud (M. van Rijmenam et al., 2018).

Big Data analytics can be seen as a tool that can provide support to the decision-making process by using technology to rapidly analyze large amounts of data from different types and from a variety of sources, to produce a stream of actionable knowledge (Berman, 2013). Big Data is characterized by a combination called the 4Vs: Volume (large amounts of data), Velocity (huge, steady stream of data), Variety (different types of data), and Value (hidden in the data) (Poleto et al., 2015). One of the most widely accepted definition of big data was presented by Gartner:

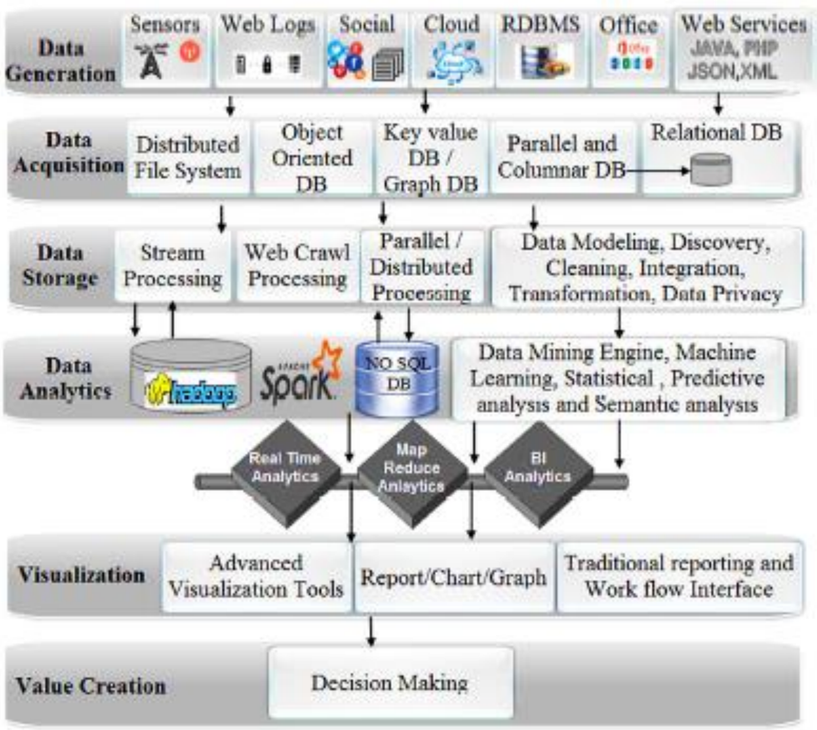
> "Big data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation"
>
> ("Gartner IT Glossary," 2018)

Chen et al (2012) defined big data analytics as "analytical techniques in applications that are so large (from terabytes to exabytes) and complex (from sensor to social media data) that they require advanced and unique data storage, management, analysis, and visualization technologies (Chen, Chiang, & Storey, 2012:1166). Big Data has the potential to aid in identifying opportunities related to decision in the intelligence phase of Simon's model (see Fig.1.9), where the term "intelligence" refers to knowledge discovery with mining algorithms (Poleto et al., 2015) and it could be aligned with the application of Business Intelligence (BI) tools to provide an intelligent aid for organizational processes.

It is important to mention that big data is not only about the original content stored or being consumed, but it is also about the information around its consumption. For example, Smartphones produce additional data sources that are being captured and that include geographic location, text messages, browsing history, and even motion or direction (M. van Rijmenam et al., 2018). The facilitation of better decision- making is identified as one of the greatest benefits of big data. In a survey by Tata Consulting Services, 80% of businesses found that implementing big data initiatives had improved their decision making (Tata Consultancy Services, 2013).

Big Data Analytics (BDA) tools, techniques and technologies have received much attention due to its wide application as multi-purpose tools, borrowing techniques from Natural Language Processing (NLP), Data Mining (DM), Machine Learning (ML), Deep Learning (DL) etc., benchmarking software technologies such as Hadoop/Map-Reduce based processing frameworks, NoSQL databases, graph databases and analytical frameworks (Saggi & Jain, 2018).

Figure 1.13 Architecture of big data analytics proposed by Saggi & Jain 2018



Source: Saggi & Jain (2018), p.768

BDA technologies allow firms to improve existing applications by offering business-centric practices and methodologies that provide a competitive advantage(H. Chen et al., 2012; Davenport, 2013). Saggi and Jain refer that Big Analytics is about turning information into knowledge by using a combination of existing and new approaches, however big data and data science are only useful for insight if they can be turned into an action and if actions are carefully defined and evaluated (Saggi & Jain, 2018) and they have proposed an architecture for big data analytics shown in Figure 1.13.

The value creation phase of Big Data is driving the creation of new tools and systems to facilitate intelligence in consumer behavior, economic forecasting, and capital markets and market domination may be driven by which companies can absorb and use the best data the fastest (Poleto et al., 2015).

Big data presents two challenges for organization: business leaders must implement new technologies and then prepare for a potential revolution in the collection and measurement of information, and the organization, as a whole, must adapt to this new philosophy about how decisions are made by understanding the real value of Big Data (Poleto et al., 2015). Some of the key benefits of BDA are understanding customers better; improving products and services; improving the management of existing data; creation of new revenue streams; better management of governance, risk and compliance and improving the detection and prevention of fraud (Saggi & Jain, 2018).

The ultimate value of a big data implementation will be judged based on one or more of three criteria:

- Does it provide more useful information?
- Does it improve the fidelity of the information?
- Does it improve the timeliness of the response? (Gantz & Reinsel, 2011).

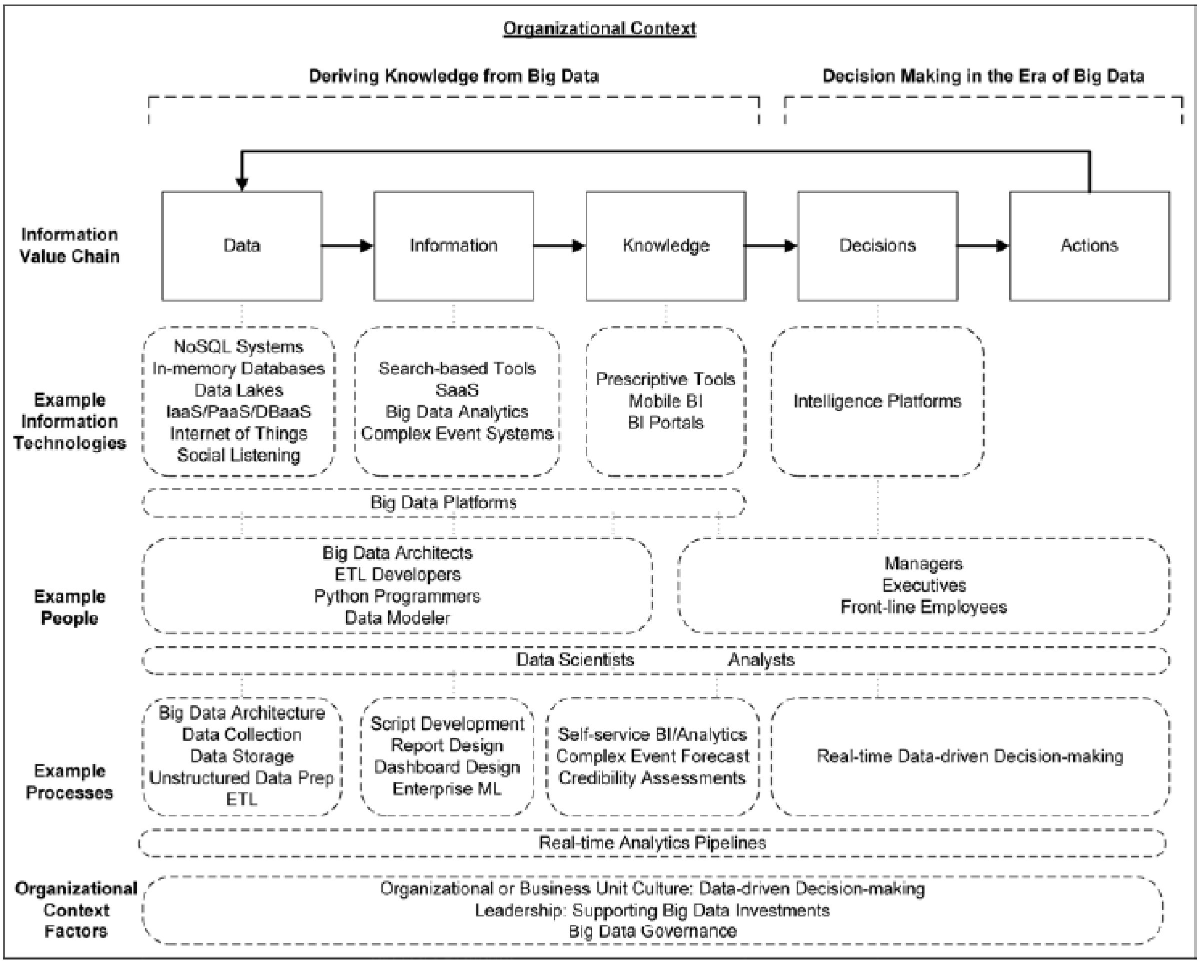
1.4.1 The information value chain

Information value chain refers to the cyclical set of activities necessary to convert data into information and to transform information into knowledge (Fayyad et al., 1996), which individuals use to make decisions and take action. Figure 1.13 illustrates the information value chain in the era of big data as presented by Abbasi et al (2016) showing the interplay between people, processes, and technologies across the stages of the big data information value chain (Abbasi et al., 2016). Certain processes and technologies could be placed differently depending on how one interprets data, information and knowledge and how decisions and actions are delineated (Abbasi et al., 2016).

BDA is a source of knowledge management, allowing firms to "add value primarily at the beginning of the information value chain and helping knowledge to flow to achieve business excellence "(Chau & Xu, 2012; Popovič et al., 2012 apud Côrte-Real, Oliveira, & Ruivo, 2017:381).

Abbassi et al (2016) suggest that companies need to reflect how decisions are made in organizations using big data and big data analytics, for example, given that people and culture can potentially act as impediments to the adoption of big data analytics, organizations should ask what theories and models are appropriate for avoiding implementation failures due to human and cultural issues.

Figure 1.14 The Big Data Information Value Chain and Examples of related People, Processes and Technologies



Source: Abbasi et al. (2016), p.vi

Additionally, the integration and use of rich data sources have great potential for driving competitive differentiation and strategic value; however, there are substantial obstacles to manage, as “data sources are often unstructured, noisy, and difficult to integrate into a focused and cohesive view of the customer” (Kitchens, Dobolyi, Li, & Abbasi, 2018:542). A main goal for deriving strategic value from big data should be the creation an integrated big data infrastructure that aids agile development of advanced customer analytics “without overinvestment in worthless data or underinvestment in data that could add significant value” (Kitchens, Dobolyi, Li, & Abbasi, 2018:542).

CHAPTER II - The case study "OrBev"

2.1. Beverage Industry

The beverage industry is classified as a sub segment of the fast-moving consumer goods (FMCG) – goods that have a useful life shorter than one year such as food and beverages, personal care, household, tobacco and which are bought frequently. The FMCG industry has a long history of generating reliable growth, mostly due to mass brands, but the industry is now facing new challenges due to shifting consumer behaviors and changes is the channel landscape. By 2010, "the FMCG industry had grown the total return to shareholders (TRS) by almost 15 percent a year for 45 years" - placing second, after the materials industry (Kelly, Kopka, Küpper, & Moulton, 2018:1).

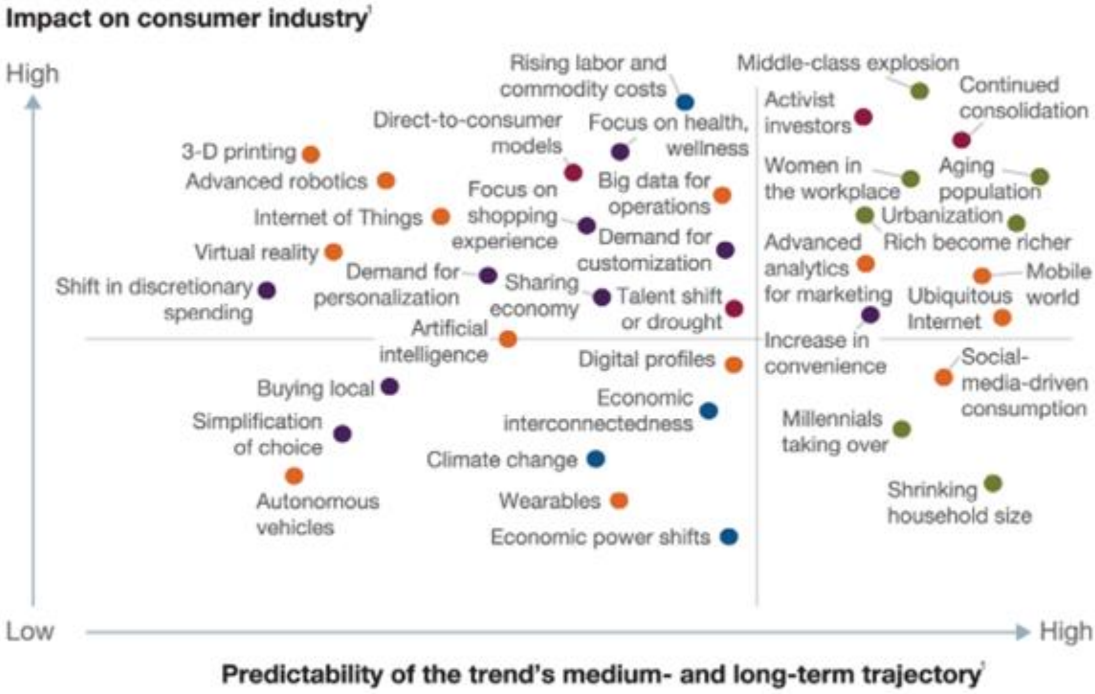
According to Nielsen, a global information, data and measurement company, and mentioned by Bhatia (2017), with the current challenging economy and intense market competition, an "increasing number of mainstream consumer packaged goods (CPG) companies and retailers have dug deeper into their petabytes of data to develop a clear understanding of consumers in their categories and sub-categories" (Bhatia, 2017). Moreover, a recent KPMG survey noted that 30% of CPG companies are using cutting-edge data analytics tools and that the use of predictive analytics has doubled from 24% to 49%. (Bhatia, 2017:2). McKinsey states (2018) that the between 2012 and 2015, the FMCG industry grew organic revenue at 2.5 percent, a figure lower than global GDP over the same period, but "companies with net revenue of more than $8 billion grew at only 1.5 percent, while companies under $2 billion grew at twice the large company rate" (Kelly et al., 2018:2). This shows that large companies are facing a growth penalty, and this is relevant because of the importance of organic growth in the consumer-goods industry. There are also other challenges in the environment which are affecting the industry, such as:

- Consumer behavior: millennials are almost four times more likely than baby boomers to avoid buying products from the big food companies (Kelly et al., 2018);
- Digital economy: Is changing the way how consumers and companies learn about and engage with each other. For example, digital-device penetration, the IoT, and digital profiles are increasing the volume of data collected, increasing companies' capabilities, but also consumer expectations.;

- Competition: The explosion of small brands capitalizing on millennial preferences and digital marketing to grow fast. They can be hard to spot because they are mostly sold online or in channels;
- New Trends: Consumers are redefining what healthy means, and demanding more products that are natural, green, organic and/or free from sugar, gluten, pesticides, and other additives and packaged-food players are racing to meet their needs (Kelly et al., 2018) and with healthy living climbing up the agenda, consumers are also expecting their consumable of choice to taste good and look cool too (Medeiros, Lauster, Veldhoen, & Soundararajan, 2018);
- Mass-retailer relationships: Disruptions are happening due to E-commerce giants (Alibaba Group, Amazon), discounters (ALDI and LIDL) which typically grow to secure market share of 20 percent or more in each grocery market they enter; and the rise of local competitors disrupting the developing-market creation (Kelly et al., 2018).

McKinsey present an overview with the biggest trends that could affect the consumer companies like beverage companies, and presented the conceptual matrix shown in Figure 2.1. It is to note that this matrix will differ for each company depending on their product categories, geography and business context.

Figure 2.1 Trends that could most affect consumer companies, McKinsey, 2015



Source: Benson-Armer, Noble, & Thiel, McKinsey Co (2015), p.5

From this matrix, we can identify high-predictability, high-impact trends in the upper right quadrant and a base case could be developed for the beverage industry in 10 years' time, with majority of consuming population being urban, average consumer will be older, middle-class in emerging markets will explode and there is high demand for business analytics. Many FMCGs have started to embrace digital, but they still have a long way to go, especially in adopting truly data-driven marketing and sales practices refers McKinsey (2018) (Kelly et al., 2018).

Developing markets still have tremendous growth potential and they are to generate new consumer sales of $11 trillion by 2025, but as new competitors offer locally relevant products and win local talent, "FMCG companies will need to respond - which will challenge the centralized decision-making models that most of them use" (Kelly et al., 2018:12). FMCGs will also need to drive costs down by zero-based budgeting, by using supply-chain and sales-and-operations planning with a high degree of automation; and advanced analytics and digital technologies to improve manufacturing performance (predictive maintenance, use of augmented reality to enable remote troubleshooting, and advanced analytics for real-time optimization of process to increase throughput yield of good-quality product) (Kelly et al., 2018).

With regards to western markets, growth potential is being dragged down by ageing societies, middle class erosion and the reduction of physical shelf space with smaller households looking for proximity and convenience. Emerging markets, meanwhile, offer growth, but challenge investors with multiple structural issues, such as infrastructure bottlenecks (Africa), security (Nigeria), exchange rate volatility (Russia, Turkey), and often several of these obstacles are combined (Schmidt, 2018). In fact, McKinsey analysis indicates that close to 20% of companies in the consumer sector are already in financial distress today, being more vulnerable those companies that are primarily active in mature and low-growth markets (Benson-Armer, Noble, & Thiel, 2015).

On the other side, Mckinsey (2015) states that spikes in input costs could increase cost of goods sold (COGS) and depress gross margins, and greater product complexity and higher labor costs "could push up operating expenses by 3 to 5 percentage points" (Benson-Armer et al., 2015:6). Investments in automation and digitization could also "increase depreciation on capital expenditures by 2-3 percentage points" even though they can enable efficiency gains in the long term (Benson-Armer et al., 2015:7). Figure 2.2 shows the median growth, the revenue CAGR 2013-2017 and top performers in the FMCG sector. An interesting fact is that within the Top 7 performers in terms of organic growth in 2017, four of them are beverage companies.

Figure 2.2 KPMG Report: Organic Growth barometer 2018

Median growth

2013-2017	3.2% pa
2017	2.5% pa

- Median CAGR growth reduced to 3.2% (from 3.4% in 2016), based on 49 companies reporting 5 year organic growth.
- Median 2017 growth (YoY) has reduced to 2.5% in comparison to 2016 at 3.0%.

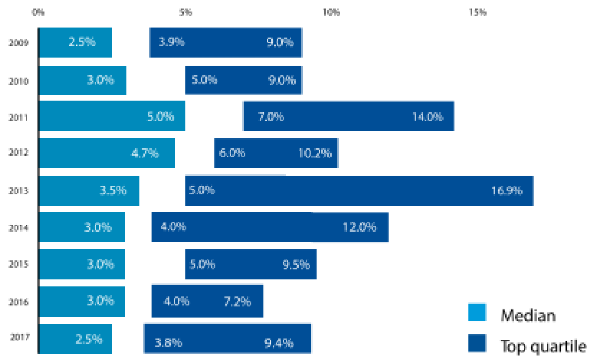
Revenue CAGR 2013-2017 and 2017 data

% pa growth	2013-2017	2016	2017
Top quartile threshold	**4.1**	**4.0**	**3.8**
Median	3.2	3.0	2.5
Lower quartile threshold	2.2	1.4	0.5
Highest growth company	**7.0**	**7.2**	**9.4**
Lowest growth company	**-1.3**	**-3.0**	**-5.0**

New entrant top performers

Short Form Name	2017 organic growth
1 Philip Morris International	**9.4%**
2 Constellation Brands	**9.0%**
3 Coca-Cola HBC	**5.9%**
4 Beiersdorf AG	**5.7%**
5 Anheuser Busch InBev	**5.1%**
6 Diageo plc	**4.3%**
7 Nomad Foods Limited	**3.9%**

Nine year upper quartile performance



Source: Medeiros et al (2018), p.5

Banks (2012) defines “success” in the consumer-packaged goods (CPG) industry in one word: “SPEED - the speed at which the best companies take advantage of the latest consumer trends and defend their market share against the competition” (Banks, 2012:1). In order for these companies to battle for retail shelf space and increase customer sales they should leverage on Big Data and the capacity to analyze vast amounts of data in real-time, for example, accurate sales forecasts and actual point-of-sale (POS) data are essential to predict production volume and fine-tune pricing and promotions (Banks, 2012).

Competing effectively means collecting large amounts of data and, while developing capabilities for storing, processing, and translating the data into actionable business insights. McKinsey´s report on analytics (2018) emphasizes that companies that will derive the most value from data are those that progress from traditional siloed customer analytics to “advanced customer analytics”, integrating a variety of relationship-oriented data that enable a deep understanding of customers, driving actionable insights and outcomes for acquisition, retention (McKinseyAnalytics, 2018).

In the beverage sector, most companies are already generating petabytes of data from various sources, such as multi-channel retail data, customer profile data from loyalty programs, social media data, supply chain data, competitor data, sales and shipment data as well as transaction and merchandising data (V. Rijmenam, 2018). But understanding the social context of individuals’ and organizations’ actions means that a company can track not only what their customers do, but also get closer to learning why they do what they do (Poleto et al., 2015).

2.1.1. The Case-study company: “OrBev”

The beverage sector is dominated by a few mass brands, where we see companies like Coca-Cola, Pepsi, Constallation Brands or Anheuser Busch InBev in top positions, but local brands continued to gain share in 2017, taking 64.6% of all brand spend, versus global brands’ 35.4% share with every 0.1% gained now worth $500 million according to the 2018 Kantar Brand Footprint report (KantarWorldpanel, 2018). According to this report, which measures which brands are being bought by the most consumers the most often, Coca-Cola is the world’s most chosen brand, picked from the shelves 5.8 billion times in a year.

The case-study company for this research project is one of the largest bottlers of drinks of The Coca-Cola Company, for confidentiality reasons the name of the company will not be disclosed, and it will be designated going forward by “OrBev”. The Coca-Cola Company System is one of the most extensive soft drinks distribution systems in the world with over 300 bottling partners. OrBev buys concentrates, drinks bases and syrups from The Coca-Cola Company, which they use to manufacture, package, merchandise and distribute finished branded products to their partners and consumers across over a dozen countries.

The demand creation is under the responsibility of The Coca Cola Company through consumer marketing and brand development and as a bottling partner, OrBev is responsible for meeting this demand though manufacturing, packaging and merchandising the final product and it is responsible for customer marketing and outlet execution.

Their business is selling more than 1 billion-unit cases every year offering the consumers a big range of non-alcoholic ready-to-drink beverages (NARTD). The company´s strategy for occasion, brand, package, price and channel (OBPPC) establishes clear category and brand priorities to offer consumers the right product, in the right pack, at the right price to suit the occasion. OrBev has a diverse footprint and is exposed both in established and emerging markets, with greater incidence in the latter.

The company promotes and supports sustainable plants and warehouses by focusing on processes and solutions for their products and processes, such as reducing the carbon emissions per liter of soft drink produced, investing in renewable energies and recyclable packaging, investing in advanced technology to optimize infrastructure and supporting their communities.

2.2. Research methodology

This exploratory work aims to explain how data mining can support the decision-making process in the beverage industry and how it can be implemented in business operations and management. To address the main research question, some secondary research questions have been proposed, namely:

RQ1: What is the type of data stored in the database or accessed by the company and which has influence in the decision making?

RQ2: Capture the insights from the Data & Analytics Team and IT as well as Business Functions on the use and application of data mining methods and techniques (association, visualization, regression, etc.);

RQ3: What is the existing decision-making process? How can Data Mining support the tactical and strategic decision-making?

RQ4: What is the adoption rate of data mining models and tools in the company and in the beverage industry?

RQ5: Which DM models could be additionally explored to solve business problems?

To answer these research questions and objectives, the present study performs a representative survey in an international company within the beverage industry. This is an exploratory research, where a survey instrument is be used to analyze the use of data mining in the beverage industry and to help understand the existing types of data and how data can be used to support the decision-making. The survey instrument was built partially based on existing data survey on Big Data from leading firms like New Vantage Partners (NewVantagePartners, 2018) and based on the European Commission survey on European data-Driven Economy(CNECT, 2015) and it was developed and distributed using the free online software tool SurveyMonkey.

On a first step, the company data is segmented into 4 verticals, by data generators, by type, by data type and by application – this information has been collected during interviews with a Data Analytics Managers at OrBev and it is presented in table format in the next chapter. Secondly, a survey instrument has been prepared and its validity has been confirmed by one established academic IS Researcher from Univesidade Aberta (Uab) and one industry expert at Orbev; both have reviewed each item on the questionnaires, assessing its content, scope and purpose. To have a more comprehensive view on the subject, both from an IT and a business perspective, two different surveys were sent to two different groups in the company:

- Group 1: (IT Team) with technical expertise, included 8 IT managers and executives and their teams,
- Group 2: (BF Team) business function leaders with business domain expertise, with around four business functions executives.

The differences in two different surveys were that in the Survey for IT team (Group 1) there were additional questions related to technical aspects of data ecosystem, which were not included in the Survey to BF (Group 2) this allows more reliability by reducing ambiguity and simplifying interpretation. The surveys were released to 12 people within OrBev, both from IT and from BF. The survey instruments are presented in full in Annex A and Annex B.

Third, a bibliometric research has been conducted to identify existing data mining use-cases in the beverage industry, the more relevant use cases in terms of impact/utility are discussed as well as other general industry use-cases that have a successful track-record and that might find business application in the beverage industry.

The analysis methodology used includes mixed research methods (quantitative and qualitative) because it allows completeness, allowing a better overall framework of the phenomenon; Compensation as the mixed methods compensate the weakness of the use of only one of the methods quantitative/qualitative by using the other method; and Diversity, by gaining different insights on the same phenomenon (Venkatesh, Brown, & Bala, 2013).

Before initiating the results analysis, the survey questions were divided into open-end and closed-end questions. For the quantitative analysis the closed-end questions were used. Due to the reduced number of respondents it was only possible to perform basic descriptive statistics on these results; these are then be presented and interpreted in next section. For the qualitative analysis, the closed-end questions results were used and presented in bar-charts and pie charts. Additionally, NVivo, a qualitative data analysis (QDA) software was used. NVivo provides a platform to organize, store and retrieve data, allows to import any data source (e.g. text, audio, email, online surveys, etc.) and allows users to classify, sort, examine and analyze research data. NVivo Plus version also provides two algorithms that ca can automatically identify some themes and sentiment in the responses to open/ended questions and it was used to present the sentiment in the Group 1 and Group 2 surveys.

The research commences with an exploratory analysis looking at each of the individual variables and its components as suggested by Saunders et al (2016), which include:

1) Specific values represented by individual data values;
2) Relative values such as high/low;
3) Trends in data values, proportions and percentages or,
4) Values distribution (Kossly 2006 apud Saunders at al 2016).

The following constructs are part of research model: data sources and availability, technology compatibility, BDA usage, organizational readiness, decision-making, expected benefits, competitive advantage, and DM challenges.

The data collected in interviews allows a better understanding of the data mining tools and models currently used, which data type is more frequently used and in need of higher processing and which are the key data types that support the decision-making process. By analyzing the type of data used by OrBev and by doing a bibliographic review with key search words data mining, decision-making, actionable knowledge discovery it should be possible to define its main advantages, opportunities and possible uses.

The authors proposed a tool for the study of adoption and diffusion of IT innovations within organizations by focusing on measuring the potential adopters' perceptions of the technology (Moore & Benbasat, 1991). These perceptions were based on the five characteristics of innovations, which affect the rate of diffusion of an innovation, derived by Rogers (1983), and two additional (Ease of use and image) developed specifically by Moore& Benbasat (1991). The goal was to understand how potential users' perceptions of the information technology innovation influence its adoption (Moore & Benbasat, 1991).

The personal competences to produce this analysis are based on the knowledge acquired during the master's in management (2016-2018) at Uab with curricular units as Business Intelligence, which allows a broader understanding of Data Mining and other BI applications. In addition, the studies undergone in Corporate Finance, Corporate Strategy, Innovation and Knowledge Management provide a domain basis for the analysis of the survey results. During the analysis and interpretation of the survey results this research attempts to apply the models and theories related to data mining, AKD, BDA, information systems and decision-making process to understand primarily how data mining can assist businesses within the beverage industry.

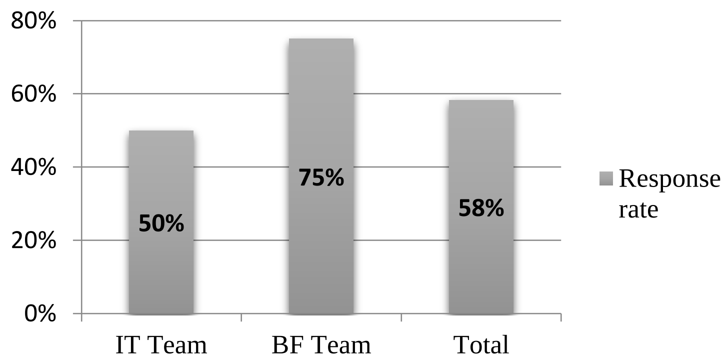
2.3. Data collection and surveys

The surveys were conducted during June-August 2018. To ensure the quality of the data, the respondent profile used the following criteria: (1) deep knowledge of the organization strategy; (2) experience in BI&A/BDA initiatives. The mailing database was provided by a Data Analytics Manager at OrBev and the initial sample of 12 firm executives and specialists receive an email to participate in the survey. One month was given, and after that time, a follow-up email was sent to increase the response rate. The surveys were closed on 15 August 2018 and the total response rate was 58.33%, comparatively, the survey Group 1 FH Team had a 50% response rate, while Group 2 FH

Team had 75 as shown in Figure 2.3. Additionally, two meetings were conducted with OrBev's Data Analytics Manager where it was given some more clarity on the business environment and current context.

Figure 2.3 Survey Instrument Response Rate 2018

	Sent	#Responses	Response rate
IT Team	8	4	50%
BF Team	4	3	75%
Total	12	7	58%



For the bibliometric research, the sources used were B-on, JSTOR, Elsevier, ScienceDirect, Scrib, books, industry reports and these were stored and managed using Mendely. Mendely is an academic software and a research workflow tool available in various platforms and it can be used to manage citations to articles, books, government documents, book chapters, or websites. The tool allowed to insert citations to articles previously marked and stored in Mendely library database. The library has been divided in folders, according to the research topics, e.g. Data Mining., DM applications, DSS, AKD, Big Data Analytics, Decision-making which allowed for better overview and in total 154 articles and books have been reviewed and stored over the last year. The search words used to develop the theoretical framework were data mining, decision making, DSS, actionable data mining and big data analytics.

To understand industry's environment and insights as well as current state of use and development of advanced analytics, several leading consulting firms' reports have been reviewed, namely McKinsey & Co, Deloitte, and KPMG.

2.4. Presentation and interpretation of results

It should be noted that this exploratory work, due to the lack of sufficient responses, cannot be representative of the whole industry or the company in study. The survey's results will be presented under each of the RQ question headings and will display the qualitative analysis followed by the quantitative analysis. Where relevant, the sentiment analysis performed in NVivo will be introduced and discussed.

2.4.1 RQ1: What is the type of data stored in the database or accessed by the company and which has influence in the decision making?

The data used at OrBev has been segmented into three verticals: by data sources, data type and data applications. This information has been collected from the survey results and during the two interviews with Data Analytics Manager at OrBev, and it is partially represented in Table 2.1. The company gathers information from various sources, but with the increase in data management tools, the Internet of Things (IoT), e-commerce, machine-learning and digitalization, many more data and metadata sources will be generated and will become available for mining.

Table 2.1 provides a brief overview of the some of the data sources and applications used and it demonstrates that Orbev is in line with the current CPG companies' trend is gathering high volumes of data, from various sources and types and then deploying data analytics tools. Some of the data sources are machine generated data, transactional data ERP data and external data, the data collected comes is different types such as structured data, unstructured data and semi-structured data and are analyzed using different tools and applications namely Enterprise Applications, collaborative analytics or predictive analytics and visualization tools.

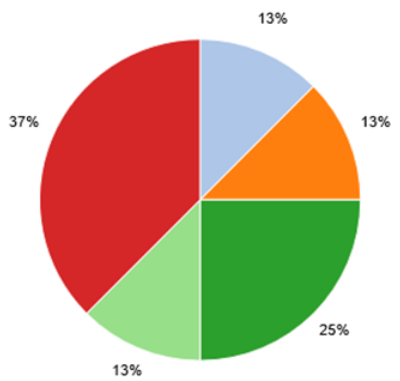
Table 2.1 Overview of the data used, segmented in three verticals, data sources, data types and data applications

Data Sources	**Data Type**	**Data Applications**
Machine-to-machine data	Structured Data	Enterprise Applications (ERP/CRM)
Transactional Data	Semi-structured data	Collaborative Analytics
Office Documentation	Unstructured Data	Visualization Tools
Social Media		Embedded analytics
Enterprise systems		Data Mining
External Feeds		Predictive Analytics

To find out if the company collects and deals with Big Data, we look into each of the 4Vs, as mentioned by Poleto et al (2015), namely: Volume (large amounts of data), Velocity (huge, steady stream of data), Variety (different types of data), and Value (hidden in the data). From the Survey Question in Figure 2.4, we can see what the primary data issues that OrBev is facing and that are driving the need for Big Data analytics.

Figure 2.4 Survey Question: What are the primary data issues driving you to consider Big Data?

Analyzing data sets < 1TB	0	0%
Analyzing data sets 1TB - 100TB	1	12.5%
Analyzing data sets 100TB - 1PB	1	12.5%
Analyzing data sets > 1PB	0	0%
Analyzing new data types (text, relationship, time-series)	2	25%
Analyzing streaming data	1	12.5%
Analyzing data from diverse sources	3	37.5%
Total answers:	8	



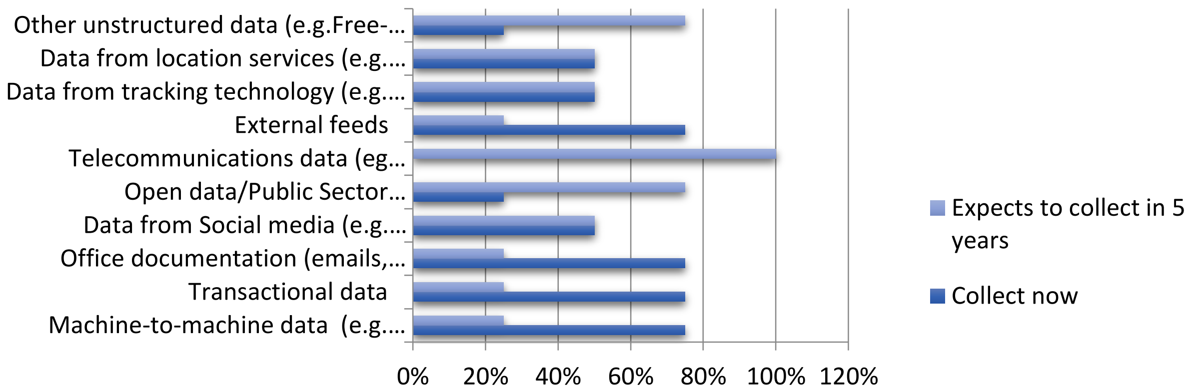
This question was only included in the IT Survey given the technical aspect; 37.5% point out that analyzing data from different sources is one of the main data issues driving BDA, and 25% refer to the analyzes of new data types, which corresponds to one of the 4 V's, Variety, since diverse sources can represent different data types. With regards to Volume, 12.5% refer analyzing data sets with 1TB-100 TB can be an issue and 12.5% refer analyzing data sets 100TB to 1PB, which represent huge volumes of data. Concerning to data Velocity, 12.5% mention that analyzing streaming data can also drive BDA use.

According to the definition of Big Data provided by Chen et al (2012) we could argue that OrBev is handling Big Data and it could benefit from the use of BDA, which are "analytical techniques in applications that are so large (from terabytes to exabytes) and complex (from sensor to social media data) that they require advanced and unique data storage, management, analysis, and visualization technologies"

From the survey results, it can be seen in more detail the data sources and which of them are the company collecting now and what it is excepting to collect in five years (Figure 2.5). OrBev is already gathering data from many the different data sources with exception from mobile data. However, there are plans to collect mobile data, PSI data (Public Sector Information) and other unstructured data in the next five years, in line with the market trend.

As more and more people use mobile for transactions, this could represent a key source for understanding consumer behavior and preferences. An AdWeek article also highlights that there is a current growing tendency to use the internet and mobile technologies for product searches: "81% of the shoppers online research before purchasing, and 61 % read product reviews as well" (Jaikumar, 2016:1).

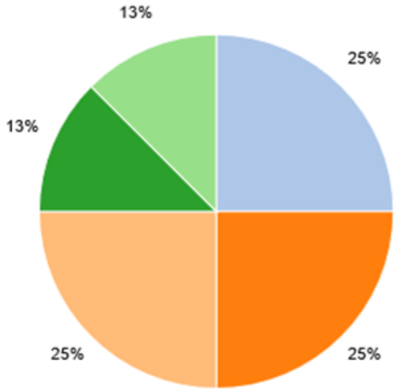
Figure 2.5 Survey Question: From what data sources does OrBev collect, or expects to collect data? (Survey Data Table in Appendix I)



To be able to process large and diverse sets of data, the company is using mainly SQL, scripting Languages and Open-source Libraries as shown in Figure 2.6, but it also considers Product-specific Languages and Proprietary Language.

Figure 2.6 Survey question: What programming languages/tools will you use for development?

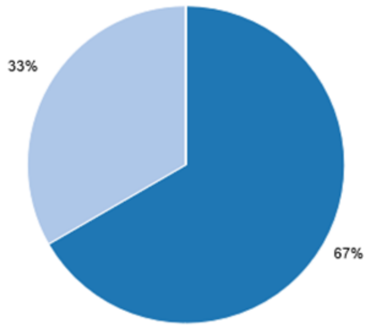
Traditional languages (e.g., Java, C, C++)	0	0%
SQL	2	25%
Scripting Languages (e.g., Python, Perl)	2	25%
Open-source Libraries	2	25%
Product-specific Languages / Libraries	1	12.5%
Proprietary Language / Libraries	1	12.5%
Total answers:	8	



These data analytics tools are usually integrated or embedded in major systems like ERP/CRM or BPM (business processes) as seen from the survey question in Figure 2.7 which shows the percentage of integration in major systems at Orbev.

Figure 2.7 Survey Question: Do your Big Data applications stand on their own or are they tightly integrated with any major systems?

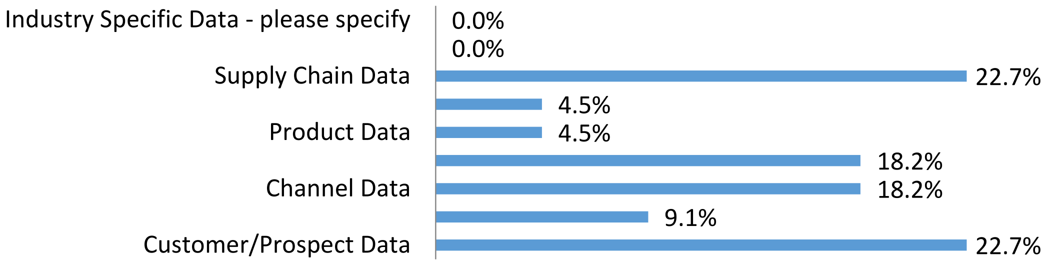
Enterprise applications (ERP, CRM)	4	66.67%
Business processes (BPM)	2	33.33%
Business rules (BRE)	0	0%
No other system	0	0%
Don´t know	0	0%
Total answers:	6	



The data domains that OrBev is mostly focused on is customer data and supply chain data, followed by channel data and market and competitive data. As referred by Power (2018) customer data can be mined for insights that drive new strategies for customer acquisition, retention, campaign optimization and next-best offers etc. (Power, 2013), while supply chain data could be used to analyze millions of stock keeping units (SKUs) to define optimal prices that maximize profit and clear inventory optimization.

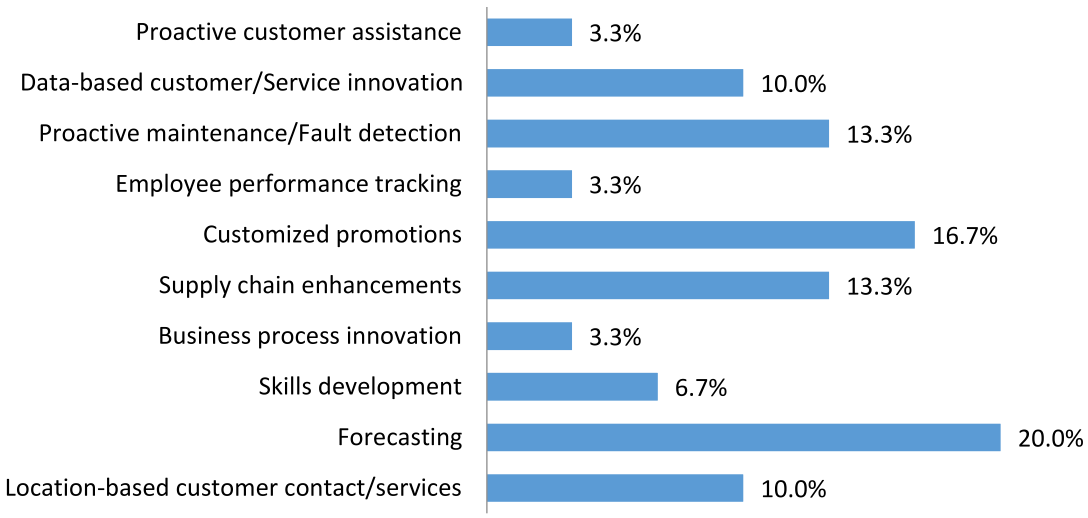
The channel data and market and competitive data could be used to generate retail coupons at the point of sale based on the customer's purchases to increase redemption rate, and data from social media could be analyzed to detect new market trends and changes in demand among many other possible uses cases (Power, 2013).

Figure 2.8 Survey question: What data domains is OrbBev most focused on in Big Data Initiatives?



Some of the data that could potentially have an influence on decision making within these business functions are shown be seen in Figure 2.10 and are mainly, in forecasting as for example, manufacturers could increase forecasting accuracy with big data analytics, minimizing their inventory cost (J. Lee, Lapira, Bagheri, & Kao, 2013) or in defining customized promotions, by doing proactive maintenance, in creating data-based customer and service innovations or for supply-chain enhancements among others.

Figure 2.10 Survey Question: How do you believe the results of data analysis will most likely be used?



Moreover, from the survey's results shown in Figure 2.11, the company OrBev has already the right data infrastructure to handle big volumes of data but need to develop analytical skills to be able to analyze and act on data. As referred by Berman (2013) Big Data analytics could support the decision-making process by using technology to rapidly analyze large amounts of data from different types and from a variety of sources, to produce a stream of actionable knowledge (Berman, 2013). After performing some basic descriptive statistics on the responses, we can note that the minimum and maximum show the lowest and highest numbered answer options; in this case, the min is 1(Strongly agree) and max is 2 (Agree). The average of all answers, shown by the median and mean with 1.5 and 1.43 respectively, show that most respondents agree with the statement done.

Figure 2.11 Survey Question: To what extend do you agree with the following statement: "The issue for us is not the growing volumes of data, but rather being able to analyze and act on data in real-time"

Answer Choices	BF	IT	Total	%
Strongly Agree (1)	1	1	2	28.6%
Agree (2)	2	3	5	71.4%
Disagree (3)	0	0	0	0.0%
Strongly Disagree (4)	0	0	0	0.0%
Total Answers	3	4	7	100.0%
Descriptive statistics				
Minimum	Maximum	Median	Mean	Standad Deviation
1	2	1.5	1.43	0.48

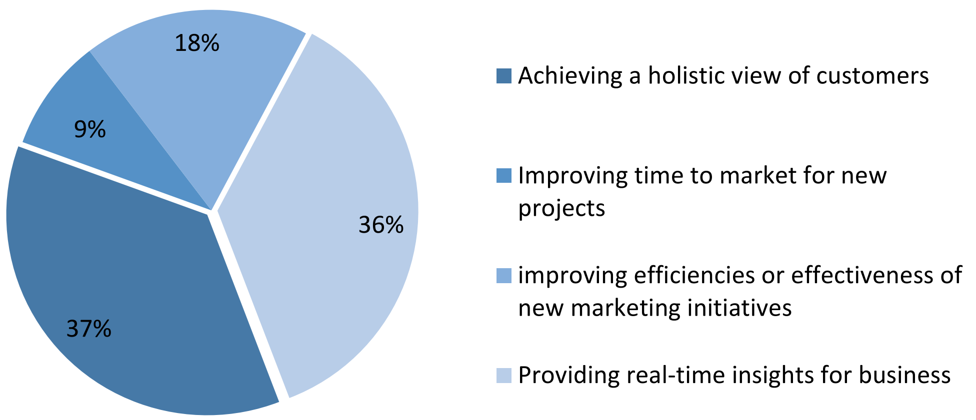
The standard deviation (SD), or the spread or variation of the responses, is 0.48 meaning that the variance is low. In general, numbers within one SD of the mean will correspond to 68% of the respondents and numbers within two SD of the mean will encompass 96% of the respondents. In our case, there is a small standard deviation, meaning that more of responses are clustered around the mean, a larger standard deviation would indicate that the responses were more spread.

Finally, it is important to highlight the importance of having data quality. Quality data means that the data is useful, accurate and relatively free from errors and ready for use in data ruining tasks and this is "a necessary precondition for extracting useful knowledge" (van Hulse, 2007:1). This is because DM techniques use automated or semi-automated procedures to process vast amounts of data looking for interesting patterns. Van Hulse

(2007) states that DM techniques do not create knowledge, but the assumption is that knowledge is present within the data and therefore, the quality of the data used is extremely crucial in ensuring successful analysis outputs.

For example, at OrBev, not having the right quality of data in seen as an impediment to achieving a holistic view of customers according to 37% of the respondents, for 36% the issue is in providing real-time insights and 18% say it restricts them from improving efficiencies or effectiveness of new marketing strategies. The complete survey data table is provided in Appendix I.

Figure 2.12 Survey question consolidated: What does the quality of data restrict you from doing?

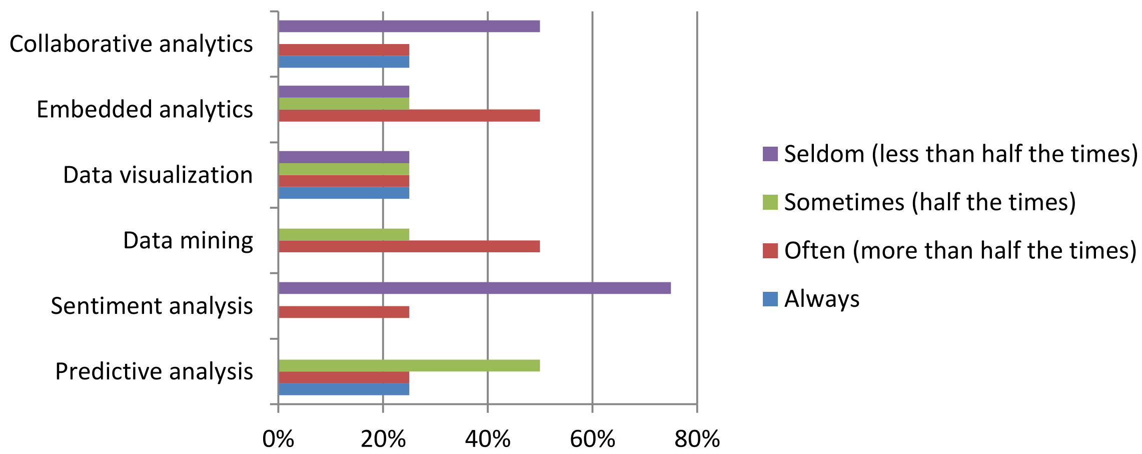


2.4.2 RQ2: Capture the insights from the Data & Analytics Team and IT as well as Business Functions on the use and application of data mining methods and techniques (association, visualization, regression, etc.);

Newton & Singh (2013) state that data mining methods could extend the possibilities of discovering information and patterns by using richer model representations better-suited for making the results more comprehensible to the non-technically oriented business users (Newton & Singh, 2013).

To understand the use of data mining methods and techniques at OrBev the survey instrument was used, and results presented in Figure 2.13, which show the main types of big data analysis tools. It can be observed that the analysis tools used more frequently are data visualization and predictive analysis, and that often (more than half of the times) the company uses data mining and embedded analytics. Collaborative analytics tools are also used, but with less frequency. There are some discrepancies in the responses to this question, which might denote the company is involved in different analytics projects, which may be using different tools.

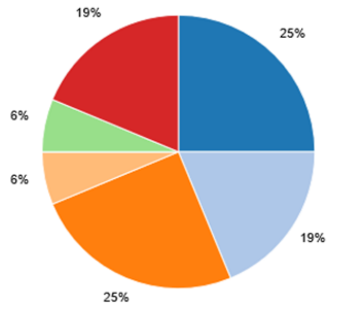
Figure 2.13 Survey question: To what extent does OrBev use the following types of big data analysis tools? (Full survey data table in Appendix I)



Besides the above types of analytics, it is relevant to know what analytics features are most important and/or useful for OrBev. To understand this, the results of the survey question from IT and BF Team Surveys have been consolidated and it is presented in Figure 2.14 below. There are four main features important to the company, where advanced analytics algorithms and Machine Learning come in first (25%), followed by data visualization and executing existing algorithms on much larger data sets (18.75%). Other analytic functions like executing existing algorithms faster and social network analytics come in last place with 6.25%.

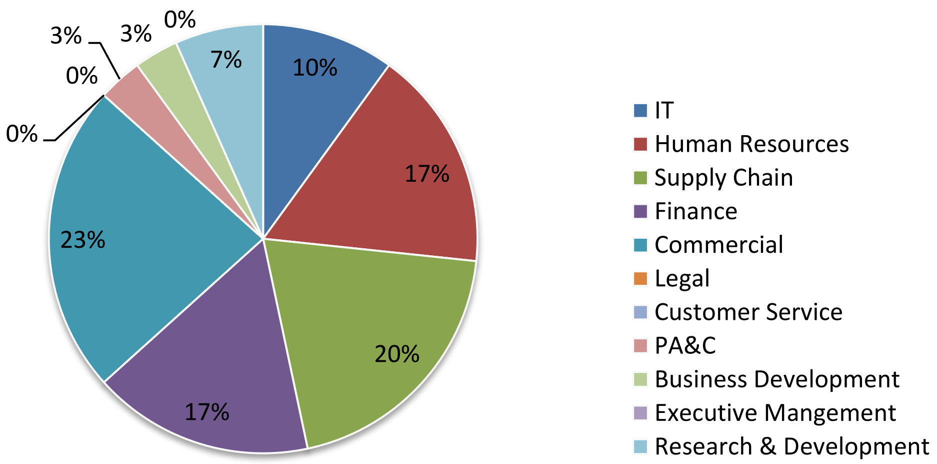
Figure 2.14 Survey question: In your opinion, what analytic functions/features are most important for OrBev?

Advanced analytics algorithms	4	25%
Data visualization	3	18.75%
Machine learning	4	25%
Executing existing algorithms faster	1	6.25%
Text analytics	0	0%
Social network analytics	1	6.25%
Executing existing algorithms on much larger data sets	3	18.75%
Total answers:	16	



Looking into the most important users of the data and analytics at OrBev (Figure 2.15), there is Commercial Function in first place, followed by Supply Chain in second and in third Human Resources and Finance. If the outputs from data mining and data analytics could be easier for business people to interpret and action and embedded into existing processes and systems, then Cao et al (2010) believe that data mining could support with productivity gain, smarter operation and decision making in business context.

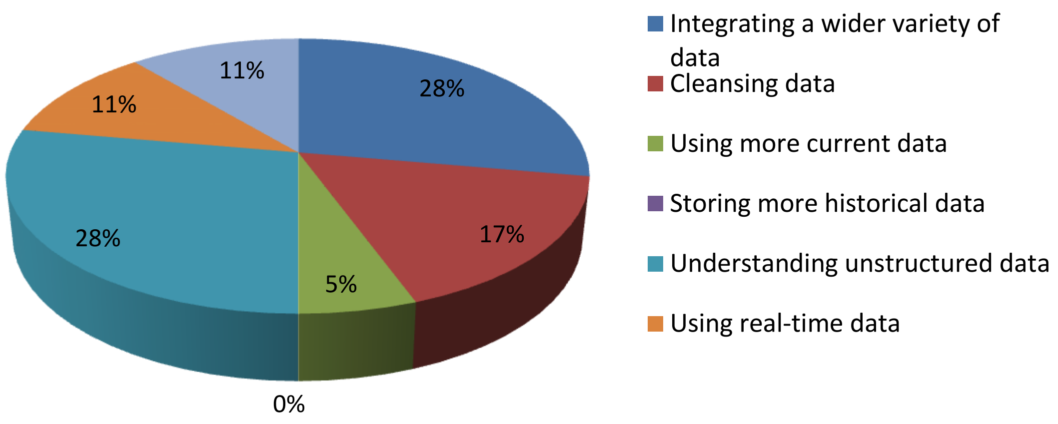
Figure 2.15 Survey question: Which business functions at Orbev are the most important users of data and data analytics? (Full survey data table in Appendix I)



However, for data mining to support decision making there are data challenges that need to be addressed to take advantage of data stored and managed by the company. Gantz & Reinsel (2011) suggest that companies should start by determining which big data projects will have the most impact, along with the requisite data sets and analytical tools. From there, the company should formulate an enterprise data strategy that overcomes the limitations of legacy data integration and layers in the new required tools and techniques (Gantz & Reinsel, 2011).

Looking into OrBev, the data challenges being addressed with data analytics are mainly concerned with integrating a wider variety of data and understanding unstructured data (28%), followed by issues with data cleaning (17%) and using real-time data and more granular data (11% each) shown in Figure 2.16. This result has been consolidated and the full survey data table presented in Appendix I. Kitchen et al (2018) developed a novel composite kernel SVM a method tailor-made for extracting insight and value from a rich variety of structured and unstructured relationship- oriented data. According to Van Rijmenam at al (2018) any company can benefit from big data analytics since it offers insights by extracting structured information from unstructured data using tools such as descriptive or predictive analytics (M. van Rijmenam et al., 2018).

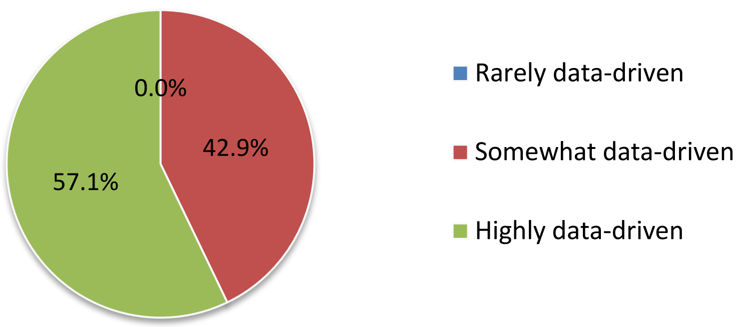
Figure 2.16 Survey question: What data challenges are you addressing with Data Analytics?



2.4.3 RQ3: What is the existing decision-making process? How can Data Mining support the tactical and strategic decision-making?

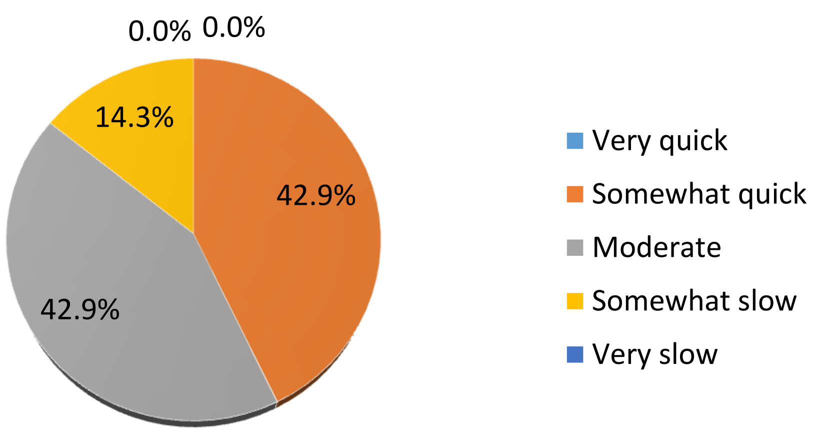
To have a better overview of some characteristics of the decision-making process is currently prevalent at OrBev, some questions have been selected to describe the type of decision, the speed of decision and the ability to use data. From Figure 2.17 which describes the decision making at the company, it is clear that most decisions are already highly data-drive (according to 57.1% of respondents) or somewhat data-driven (42.9%). This could

Figure 2.17 Survey question: Which of the following best describes decision-making at OrBev?



indicate that there is already a big focus in making informed and data-based decision and is leveraging its data assets. Concerning the speed of decisions within the business function, the general opinion of respondents is divided between moderate (42.9%) and somewhat quick (42.9%) demonstrating that there is still room for improvement in embedding data analytics with existing business functions systems to better capitalize on the data insights generated by BDA (Figure 2.18) and there were no responses to very slow or very fast decision making.

Figure 2.18 Survey question: How would you rate the speed of decision-making within your Function/BU when using big data analytics as a key resource?



Here it is relevant to do a descriptive statistic and compare between IT and BF team responses as shown in Figure 2.19.

Figure 2.19 Survey question: How would you rate the speed of decision making within your FH/BU when using big data analytics as a key resource? - Descriptive statistics table

How would you rate the speed of decision-making within your Function/BU when using big data analytics as a key resource?				
Answer Choices	**BF**	**IT**	**Total**	**%**
Very quick (1)	0	0	0	0.0%
Somewhat quick (2)	1	2	3	42.9%
Moderate (3)	1	2	3	42.9%
Somewhat slow (4)	1	0	1	14.3%
Very slow (5)	0	0	0	0.0%
Total Answers	3	4	7	100.0%

Descriptive Statistics				
Minimum	Maximum	Median	Mean	Standad Deviation
2	4	3	2.71	0.70

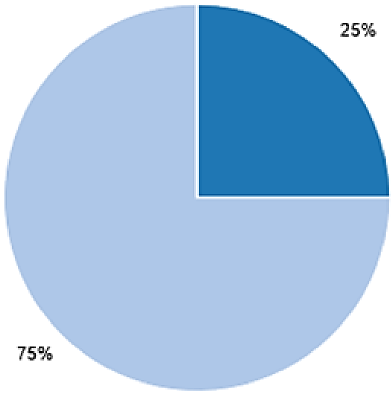
Descriptive Statistics+ Compare	Minimum	Maximum	Median	Mean	Standad Deviation
BF	2	4	3	3.00	0.82
IT	2	3	2.5	2.50	0.50

While for the BF team the median and mean are around 3 (moderate), from the point of view of IT team, the median and mean are situated around 2.5 (closer to somewhat quick), demonstrating that there are discrepancies in the way both functions access the big data analytics as key resource. This could be because for IT it might seem the information in readably available, but from the business the same information even if available might not be accessed on time probably it is why it is not seen as a key resource.

What is relevant to understand to answer our research question is what is the share of data collected by the company that can be used to improve decision-making, business processes, products, etc. It is challenging to understand the real impact of data in the daily operations, but we can have an estimate from the survey results in Figure 2.20.

Figure 2. 20 Survey Question: From all data collected by OrBev, what is approximately the share of data the company exploits to improve business processes, products, etc.?

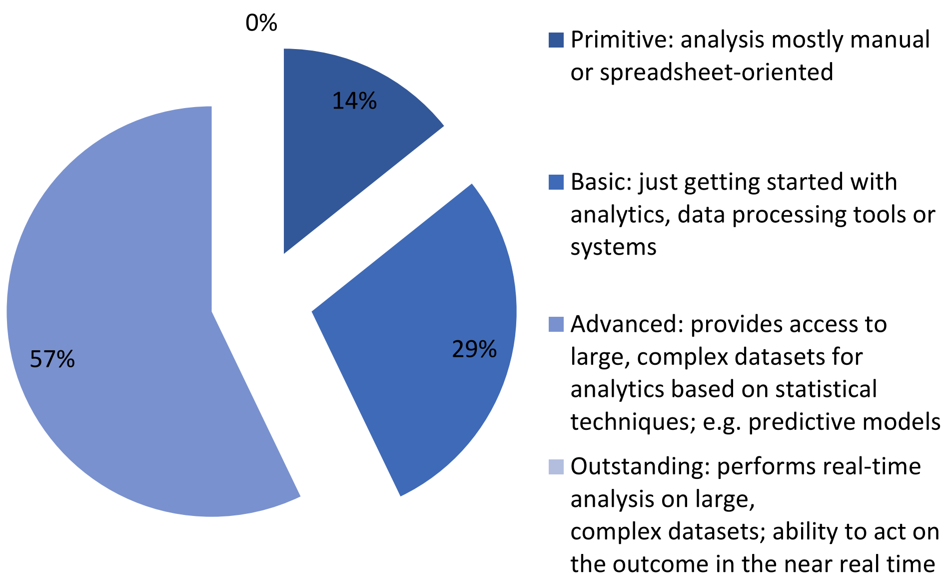
<10%	1	25%
10-40%	3	75%
41-60%	0	0%
61-90%	0	0%
>90%	0	0%
Total answers:	4	



We can see that 10-40% of the data collected is used improve business processes and showing there is still much room for improvement and better insights integration. This gap between the amount of data collected and the share of data exploited to improved business decisions and processes could be due to a lack of understanding on how to implement insights from data analytics. As mentioned by Cao et al (2010) the Actionable Knowledge Discovery deliverables should be business-friendly enough for business people to interpret, validate, action and be “seamlessly embedded” into existing business processes and systems (Cao et al., 2010:1300).

Concerning the ability of Orbev to drive decisions today leveraging on data and data analytics, the respondents agree that the company is already very advanced, providing access to large, complex datasets for analytics based on predictive models (57%). In Figure 2.21 the consolidate responses have been represented, 29% said that the company ability was basic, just getting started with analytics and 14% are the opinion that the data analytics capability is primitive. From the data observed so far, it cannot be said that OrBev s capabilities are primitive in terms of data analytics since they are already deploying data mining, customer segmentation and predictive analytics, which all seem more towards advance stage in its use to use available data.

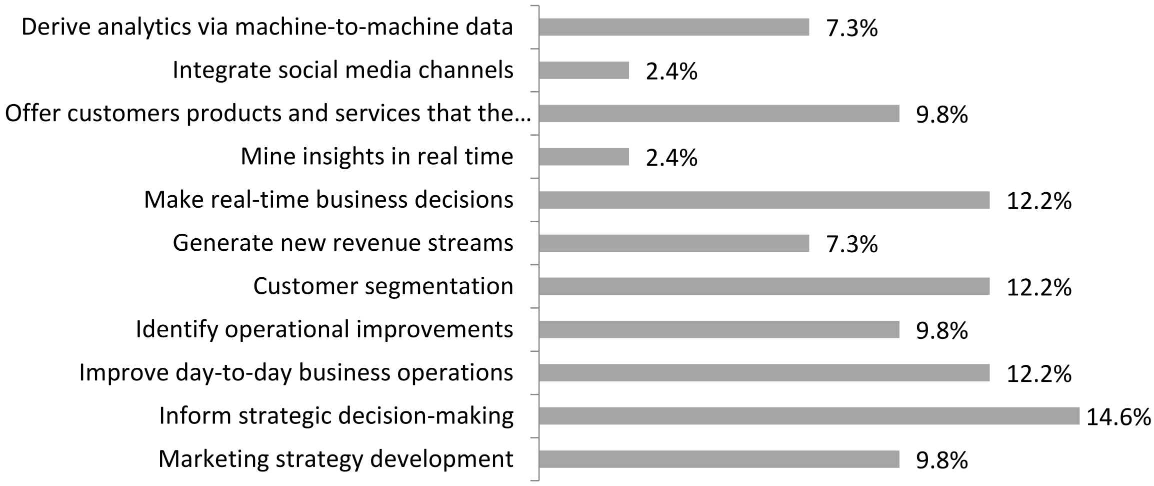
Figure 2. 21 Survey question: How would you rate Orbev’ s ability to use available data to drive executive decisions today? (Full Survey Data Table in Appendix I)



According to He et al (2005) there are two major issues in data mining research and applications: patterns and interest as mention in chapter 1.1.3. Interest refers to patterns that are useful or meaningful in business applications, e.g. an "important measure of interestingness is whether it can be used in the decision making process of a business to increase its profit" (He et al., 2005:1). In this context, the results shown in Figure 2.22 demonstrate some of the possible uses of data analytics to support decision making at OrBev, and which could potentially lead to increase in profit.

Most respondents believe that data analytics could be used to support informed strategic decision-making (14.6%), improve day-to-day business operations (12.2%), customer segmentation (12.2%), making real-time business decisions (12.2%), and marketing strategy development (9.8%) among others. For these type of decisions, more tactical and strategic, data mining could be used to predict future trends, to examine customer purchase habits, lower costs etc. Nonetheless, data mining also has some limitations such as privacy and security issues, costly to implement, it cannot guarantee perfect results and cannot explain why an outcome occurs (Pathak, Singh, & Oberoi, 2013).

Figure 2. 22 Survey question: How could data analytics be used to make decisions at OrBev? (Survey Data Table in Appendix I)

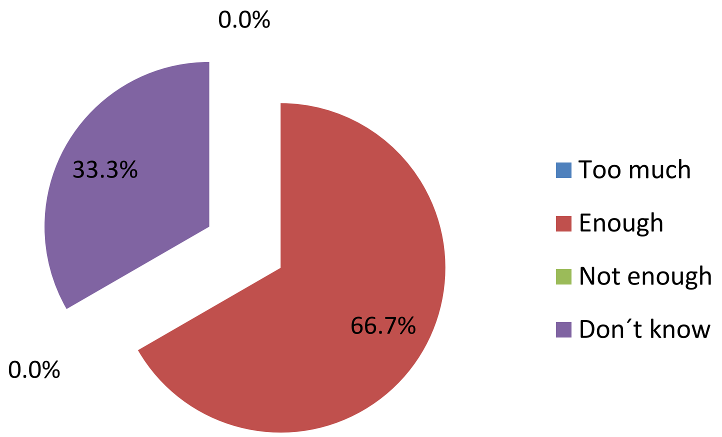


With the support of data mining technology the user or decision-maker in the corporate sector is able to process huge amounts of data and make decisions which are useful for whole organization (Pathak et al., 2013). According to Pathak, data mining is "becoming increasingly popular as a business information management tool" and it can be used in conjunction with a data warehouse to help with certain types of decisions. It does so by providing answers to many queries to the organization and the user and helps in decision making.

There are many types of queries, like tactical query or strategic query where a tactical query is a database operation that attempts to determine the best course of action today, tend to produce small result set, while the strategic query provides information necessary to make long term business decision and attempts to determined what happened, why and what will happen next producing vast amounts of data (Pathak et al., 2013).

The amount of data available for decision making at OrBev is represented in Figure 2.23. This survey question was only presented to the business function leaders as they are usually the end users of the data analytics insights. 66.7% of BF team believes they have enough amounts of data available to support decision-making, while there are 33.3% that did not know.

Figure 2. 23 Survey question: Looking specifically at your function, how would you characterize the amount of data available to support decision-making



If Orbev collects high amounts of data, from different sources and is deploying big data analytics tools, why is the speed of decision making seems as moderate or somewhat quick? What could be the main obstacles for effective data-based decision-making? The survey results for this question has been consolidated and the survey data table presented in Appendix I and shown in Figure 2.24. The main reason referred was the lack of accurate, timely or relevant data from across the business (27.8%), inadequate quantitative expertise among executives and support staff (22.2%) and in third with 16.7% is Inconsistent reporting of information among business units, geographies or functional operations and/or inadequate tools for gathering, integrating or analyzing operational information. The 2 less mentioned issues were overwhelmed by data or lack of support from top executives.

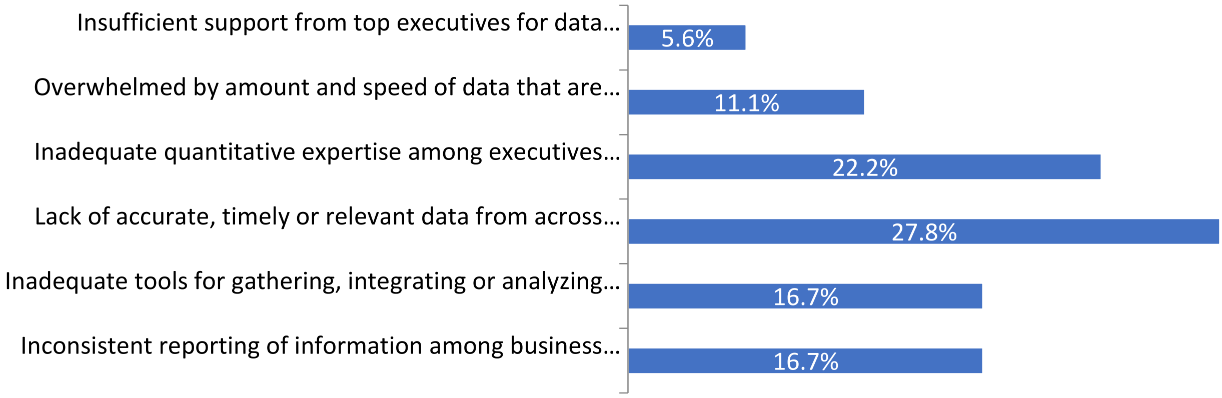
The ultimate value of a big data implementation will be judged based on one or more of “three criteria:

(1) Does it provide more useful information?

(2) Does it improve the fidelity of the information?

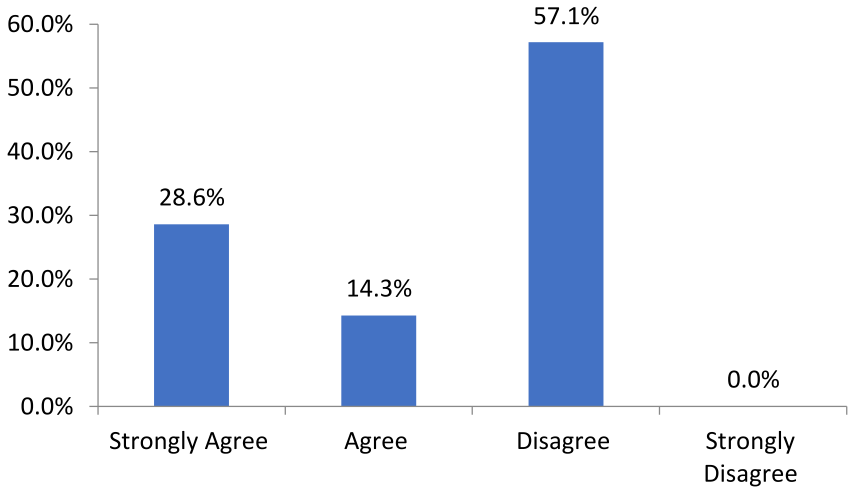
(3) Does it improve the timeliness of the response?” (Gantz & Reinsel, 2011:8).

Figure 2.24 Survey question: What are the biggest obstacles to successful data-based decision-making in your Function/BU?



Gantz & Reinsel (2011) state that the combination of post-recession business growth, a technology, and growth of the digital universe in this decade present a unique opportunity for CIOs and their staff to "drive change and growth for their organizations", which may be a challenge, but "it is also a propellant for new and exciting uses of data"(Gantz & Reinsel, 2011:12). In this context, and looking at Figure 2.25, the results show that there is still place to automate many of the operational and tactical decisions at OrBev as 57,1% disagree that most operational/tactical decision that could potentially be updated, were.

Figure 2. 25 Survey question: To what extent do you agree with the following statement: "Most operational/tactical decisions that can be automated, have been automated."



2.4.4 RQ4: What is the adoption rate of data mining models and tools in the company and in the beverage industry?

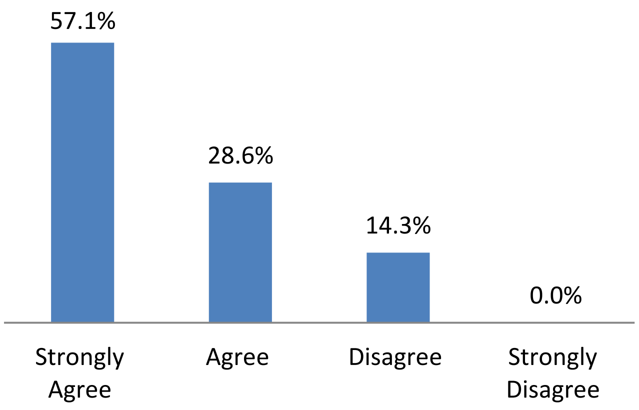
To understand the various perceptions that individuals at OrBev may have of adopting data mining for decision making, the instrument proposed by Moore & Benbasat (1991) will be partially used as reference. Due to the lack of sufficient responses it was not possible to use the instrument itself. The authors proposed a tool for the study of adoption and diffusion of IT innovations within organizations by focusing on measuring the potential adopters' perceptions of the technology (Moore & Benbasat, 1991). These perceptions were based on the five characteristics of innovations, which affect the rate of diffusion of an innovation, derived by Rogers (1983), and two additional (Ease of use and image) developed specifically by Moore& Benbasat (1991). The goal was to understand how potential users' perceptions of the information technology innovation influence its adoption (Moore & Benbasat, 1991).

The main constructs of interest in this research relates to the perceptions of using data mining and big data analytics rather than on perceptions of the innovation itself. In determining what characteristics to examine, Rogers (1983) identified five general attributes of innovations and defined them as follows:

1) Relative Advantage: the degree to which an innovation is perceived as being better than its precursor;
2) Compatibility: the "degree to which an innovation is perceived as being consistent with the existing values, needs, and past experiences of potential adopters";
3) Complexity: the "degree to which an innovation is perceived as being difficult to understand and use";
4) Observability: the degree to which the results of an innovation are observable to others; "the easier it is for individuals to see the results of an innovation, the more likely they are to adopt it";
5) Trialability: the degree to which an innovation may be experimented with before adoption(Rogers, 1983:15-16).

Looking at the first characteristic, relative advantage which refers to the degree an innovation is perceived as being better than its precursor, and looking into the survey results, it can be observed (Figure 2.26) that data mining is viewed as advantageous in the beverage industry by assisting in decision making, with 57.1% of responses saying they strongly agree with that statement and with 28,6% that agree with the statement. Other statements that could indicate relative advantage would be for example, if the use of data mining enables the respondent to accomplish tasks more quickly, it if makes it easier makes it easier to do my job and improves their performance or it increases the company's productivity. From the descriptive statistics below, it can be noted that the answer options are within min 1 and max 3 which mean, 1 for strongly agree and 3 for disagree and there is a higher standard deviation for BR team 0.94, than for IT team 0.35, demonstrating that the BF answers are more spread.

Figure 2. 26 Survey question: How far would you agree with the following statement? "Data mining techniques can support the business process of decision making in the beverage industry." (Survey Data Table in Appendix I)



How far would you agree with the following statement? "Data mining techniques can support the business process of decision making in the beverage industry."

Answer Choices	BF	IT	Total	%
Strongly Agree (1)	2	2	4	57.1%
Agree (2)	0	2	2	28.6%
Disagree (3)	1	0	1	14.3%
Strongly Disagree (4)	0	0	0	0.0%
Total Answers	3	4	7	100.0%

Descriptive Statistics

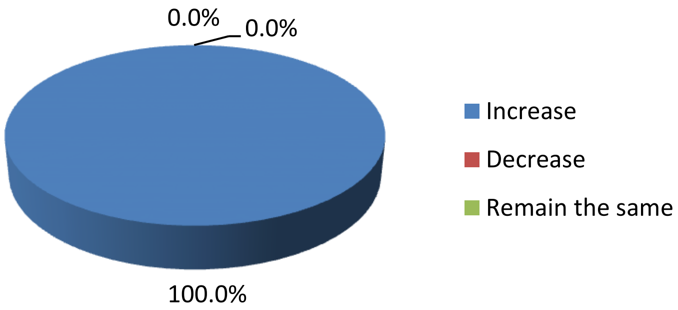
Minimum	Maximum	Median	Mean	Standad Deviation
1	3	2	1.57	0.73

Descriptive Statistics+ Compare

	Minimum	Maximum	Median	Mean	Standad Deviation
BF	1	3	2	1.67	0.94
IT	1	2	1.5	1.50	0.35

On this last point, the survey results in Figure 2.27 show that OrBev could increase its productivity by relying on data, and indirectly data mining, for decision-making.

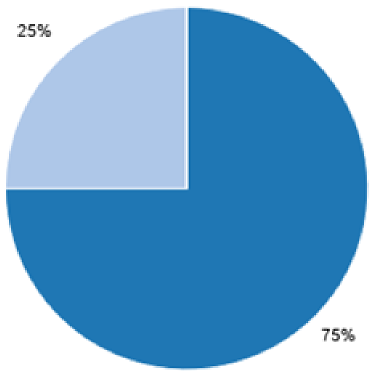
Figure 2. 27 Survey question: For the purpose of increasing effectiveness, should OrBev increase or decrease its reliance on data in the decision-making process? (Survey Data Table in Appendix I)



The second characteristic, compatibility looks at how an innovation is perceived as being consistent with the existing values, needs, and it could be perceived by statements like, the use of using advanced analytics incl. data mining is compatible with all aspects of respondent work and with their current situation or if the use of advanced analytics fits into their work style. Here the survey results displayed in Figure 2.28 show that 75% of the people agree on how advanced analytics and data mining can be integrated into the business structure and this could be indicative of compatibility.

Figure 2. 28 Survey question IT team: How are you thinking about Big Data capabilities with respect to Advanced Analytics (data mining, predictive modeling, etc.)

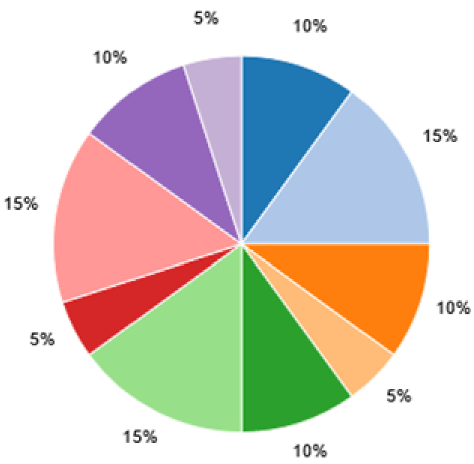
Big Data is a part of the Advanced Analytics toolbox	3	75%
Big Data and Advanced Analytics are separate things	1	25%
Big Data is unproven while Advanced Analytics is a part of core business operations	0	0%
Don't know	0	0%
Total answers:	4	



Moreover, Figure 2.29 presents some possibilities of how advanced analytics are most likely to be used, addressing how it could fit it into their work style or business decisions at OrBev. The results show that there would potential to incorporate data analytics is across several decisions, for example, it could be used to make decisions based on forecasting (15%), for the definition of customized promotions (15%), for skill development (10%) among others.

Figure 2.29 Survey question IT: How do you believe the results of data analysis will mostly be used?

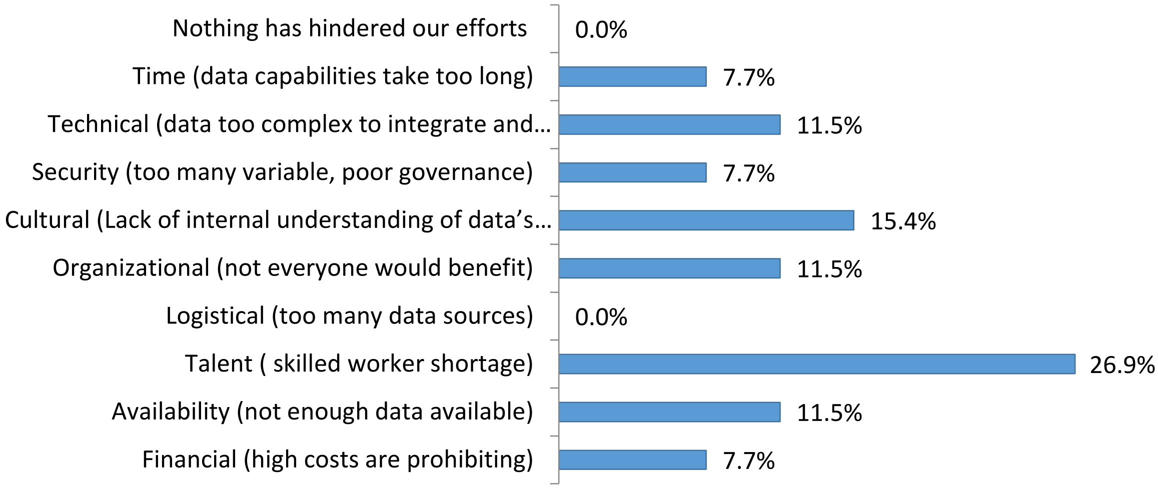
Location-based customer contact/services	2	10%
Forecasting	3	15%
Skills development	2	10%
Business process innovation	1	5%
Supply chain enhancements	2	10%
Customized promotions	3	15%
Employee performance tracking	1	5%
Proactive maintenance/Fault detection	3	15%
Data-based customer/Service innovation	2	10%
Proactive customer assistance	1	5%
Total answers:	20	



In terms of complexity and how data mining is perceived as being difficult to understand, the results show that there is some complexity in the adoption of advanced analytics. Some statements to indicate this could be, for example, how difficult it is telling others about the results of using advanced analytics, or to communicate the consequences of using advanced analytics or if these results would be apparent to the respondent.

In Figure 2.30 below, when asked what factors have hindered the increase of big data use, the three top factors referred were talent, and the lack of skilled workers (26.9%) stressing the complexity of the innovation, cultural factors due to lack of understanding of the value of data (15.4%) or technical due to data that is too complex to integrate and analyze (11.5%). These factors may be indicative that the data mining's adoption rate at OrBev and in the beverage industry could be slower than advocated due to constrains in talent and cultural aspects.

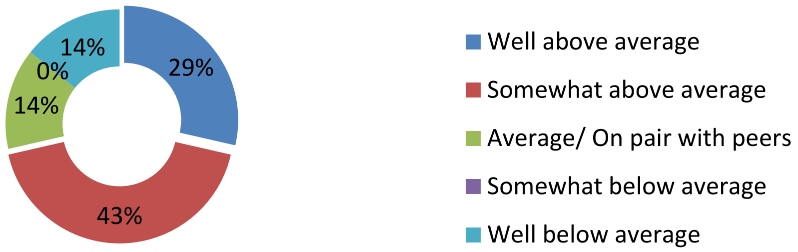
Figure 2.30 Survey question: What factors have hindered OrBev to increase its use of big data so far? (Survey Data Table in Appendix I)



The fourth perceived characteristic mention by Rogers was observability, that investigates the degree to which the results of an innovation are observable to others and some statements to infer this characteristic could be what others do use their advanced analytics including data mining or is data mining very visible in their company or is it used in their department. The survey results do not address this aspect concerning how visible data mining is within the organization, but they provide a view on how the respondents see their selves when compared with their industry peers.

Figure 2.31 present the result to this question, where the majority believes they are "somewhat above average" (43%) and 29% believes they are "well above the average". These results are not very clear nor conclusive since the descriptive statistics showed that there is a high Standard Deviation of 1.28 in the responses, especially among the Business Team (1.68 SD), meaning that there is high variance and the responses are far from the mean. For this reason, observability, as a perceived characteristic of innovation, indicates that the company is still not clear on the benefits of using data mining as an innovation. While for IT team the benefits of data mining help the company to gain competitive advantage placing it well above average when compared to its peers, for the BF team there is still unclear.

Figure 2.31 Survey question consolidated: How would you rate OrBev's success compared with that of your industry peers? (Survey Data Table in Appendix I)



How would you rate CCH's success compared with that of your industry peers?				
Answer Choices	BF	IT	Total	%
Well above average (1)	1	1	2	28.6%
Somewhat above average (2)	0	3	3	42.9%
Average/On pair with peers (3)	1	0	1	14.3%
Somewhat below average (4)	0	0	0	0.0%
Well below average (5)	1	0	1	14.3%
Total Answers	3	4	7	100.0%

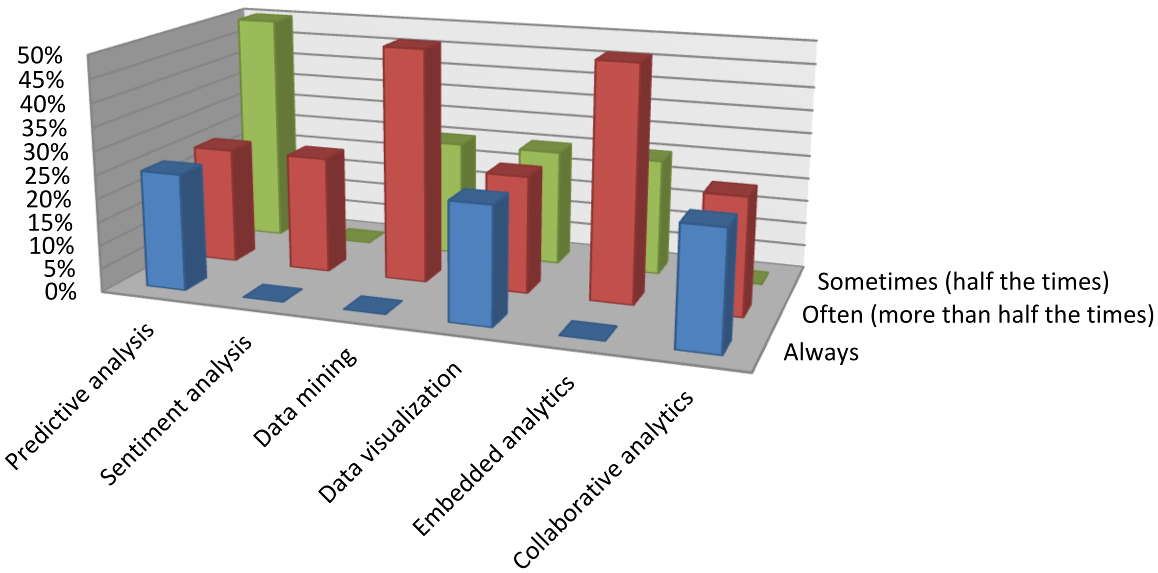
Basic Statistics				
Minimum	Maximum	Median	Mean	Standad Deviation
0	5	2.00	2.29	1.28

Descriptive Statistics+ Compare

	Minimum	Maximum	Median	Mean	Standad Deviation
BF	0	5	1	3.00	1.63
IT	1	2	0.8	1.75	0.43

The last innovation perceived characteristic referred is Trialability and it looks into the degree to which an innovation may be experimented with before adoption. Some possible statements to infer this could be opportunity that respondent to try various advanced analytics applications, or they are satisfactorily trying out various uses of advanced analytics. Looking to OrBev, the results show that there is trialability and space to explore advance analytics. From Figure 2.32 we can see some of the current big data analytics tools being used according to the frequency of its use, i.e. if they are used always, often or seldom. Among others, the most used data analytics tools are collaborative analytics, data visualization and predictive analytics, and the tools also used, but with less frequency are embedded analytics, data mining to sentiment analysis. This demonstrates the trialability of data mining since that there are different analytic tools being used by different departments or in different projects and Data Mining is used half of the times (not always) but could be available for others in the organization to access and use it.

Figure 2. 32 Survey Question: To what extent does OrBev use the following types of big data analysis tools? (Full Survey Data Table in Appendix 1)



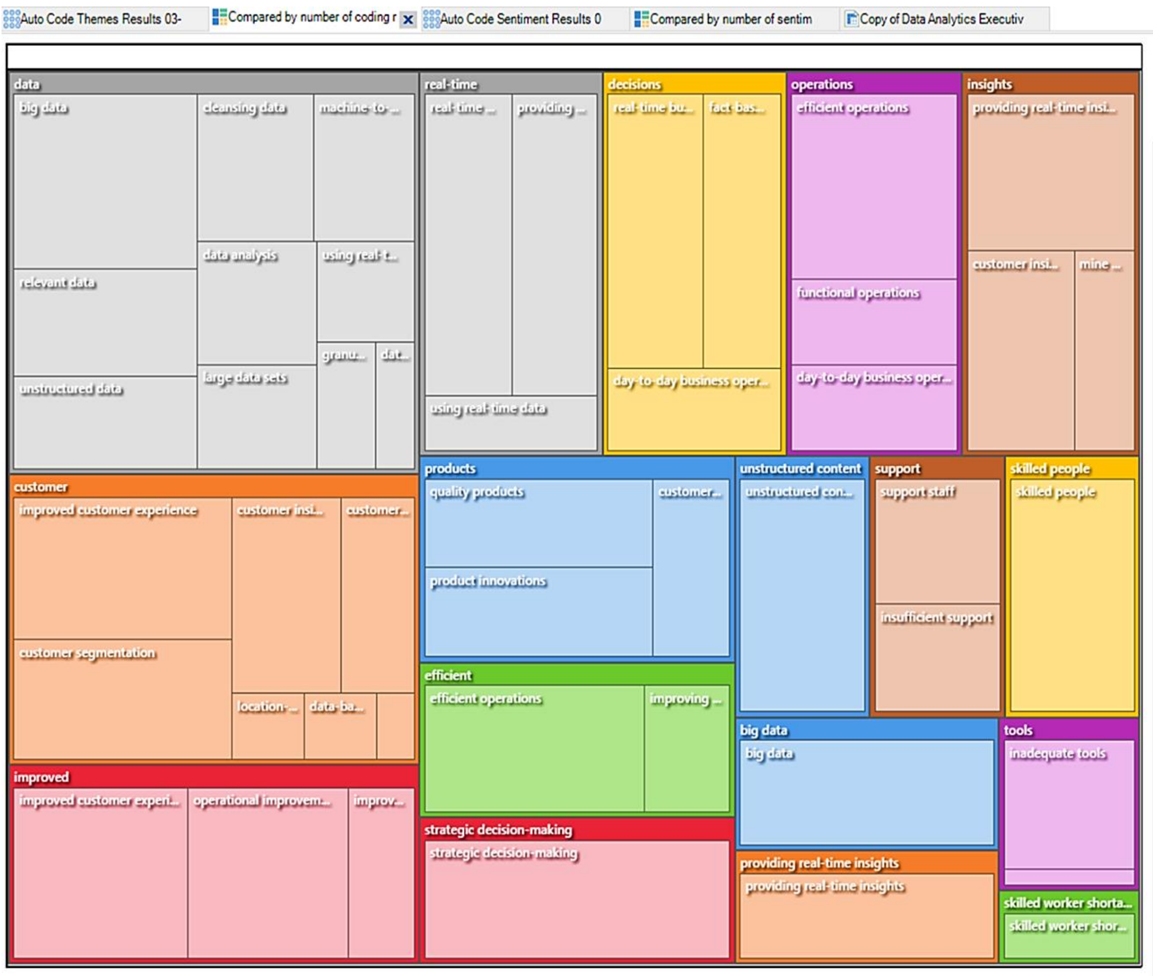
	Predictive analysis	Sentiment analysis	Data mining	Data visualization	Embedded analytics	Collaborative analytics
Always	25%	0%	0%	25%	0%	25%
Often (more than half the times)	25%	25%	50%	25%	50%	25%
Sometimes (half the times)	50%	0%	25%	25%	25%	0%

Using the software NVivo it was possible to perform text coding and sentiment analysis and the coded data visualized using tree. For the text coding a tree map was used, which is a diagram that shows hierarchical data as a group of nested rectangles of varying sizes, and this will be used for our data. The size of the rectangles indicates the amount of coding references. There are key words that might appear in all sentiment categories, but normally this happens at different frequencies and with different word senses. The subsetted data can be analyzed by looking into the text frequency count, text search, or matrix coding.

Figure 2.33 shows a tree map created with NVivo. The size of the region for a specific term equals the frequency of appearance of that word, which means that the more a term is mentioned, the more important that concept might be for the text corpus/data set. Larger areas are displayed at the top left of the chart while smaller rectangles are displayed toward the bottom right. Looking into the tree map for the survey instrument used we see on the left side the placement of the words most commonly used, which are “data” (incl.

Big Data), "real-time", "customer", "decisions" or "insights". The words at the bottom right on the tree map are the ones used less, namely "tools", "providing real-time insights" and "skilled people". These results reflect the focus of OrBev in leveraging data and data analysis for decision making, while denoting less interest in the technical side, like tools (e.g. data mining models, etc.) and capabilities (talented workers) to allow this analysis.

Figure 2. 33 Tree map of a Node (Coding) Structure of the survey instrument used. Extracted from NVivo Plus Trial version, September 2018



For the sentiment analysis, NVivo software runs a sentiment classifier, which codes the text into four categories: very negative, moderately negative, moderately positive and very positive. The idea behind this is that attitude leads to behavior, and the way people feel (if positive or negatively) towards something will affect how they think and their actions. Sentiment analysis, is conceptualized as a positive-negative polarity, with semantic-based terms indicating positive or negative leanings (QSR, 2016).

Some sentiment algorithms allow close analyses of sentiment on particular types of texts and others are generalist algorithms that try to capture sentiment in general. NVivo 11 Plus sentiment analysis tool enables the capturing of sentiment, but still has a lot of “noise” in the data. The results of the sentiment analysis are shown in Figure 2.34 and while it might enable early insights or leads on further exploration this sentiment analysis should not be used alone in a research context but only as a complement. The general sentiment observed is more neutral and positive which might indicate some apprehensiveness in the adoption of advance analytics (incl. data mining) at OrBev.

Figure 2. 34 Tree map by Number of Sentiment. Extracted from NVivo Plus Trial Version, September 2018

Furthermore, QSR International also refer that various types of expressions are not captured with the sentiment analysis tool, i.e. words of “sarcasm, double negatives, slang, dialect variations, idioms, (and) ambiguity”, and texts are analyzed in a quiet simplistic way, without tuning to cultural and other nuances (QSR, 2016).

2.4.5 RQ5: Which DM models could be additionally explored to solve business problems?"

According to McKinsey (2018) advanced analytics can create value when big data and advanced algorithms are applied to business problems to find a solution that is measurably better than before and it can do this by identifying, sizing, prioritizing, and phasing all applicable use cases, businesses can create an analytics strategy (McKinseyAnalytics, 2018). The SAS website briefly identified some possible uses of Big Data with appropriate analytics, namely : recalculate entire risk portfolios in minutes and understand future possibilities to mitigate risk, mine customer data for insights, quickly identify customers who matter the most, send tailored recommendations to mobile devices when customers based on location, analyze data from social media to detect new market trends and changes in demand, use clickstream analysis and data mining to detect fraudulent behavior, and identify root causes of failures and defects by investigating user sessions, network logs and machine sensors (Power, 2013).

From Figure 2.35 some uses cases/benefits of Big Data Analytics at OrBev can be inferred. The results show that 25.9% see an increase in sales as tangible benefit (by using clustering to send customers tailored promotions), 14.8% believe BDA could be used for improved customer experience (by identifying the customers who matter most and target these), more efficient operations (by using proactive maintenance and identify root causes of defects) and it could support informed decision making among others. In fact, Saggi and Jain (2018) also refer that some of the key benefits of BDA are understanding customers better; improving products and services; improving the management of existing data; creation of new revenue streams; better management of governance, risk and compliance and improving the detection and prevention of fraud (Saggi & Jain, 2018).

Data mining and BDA could allow beverage industry to improve the capacity of insights from data, support in effective marketing campaigns by using for example, segmentation or even in creating more innovative products. Because the demand of consumer is changing at a fast pace, companies need new technologies and strategies to keep up, and this could be done through data mining and BDA by providing patterns identification and insights from data on consumer changing behavior leading to improved sales and business efficiency.

Figure 2.35 Survey question consolidated: What tangible benefits do you hope to achieve through your Big Data initiatives? (Survey Data Table in Appendix 1)

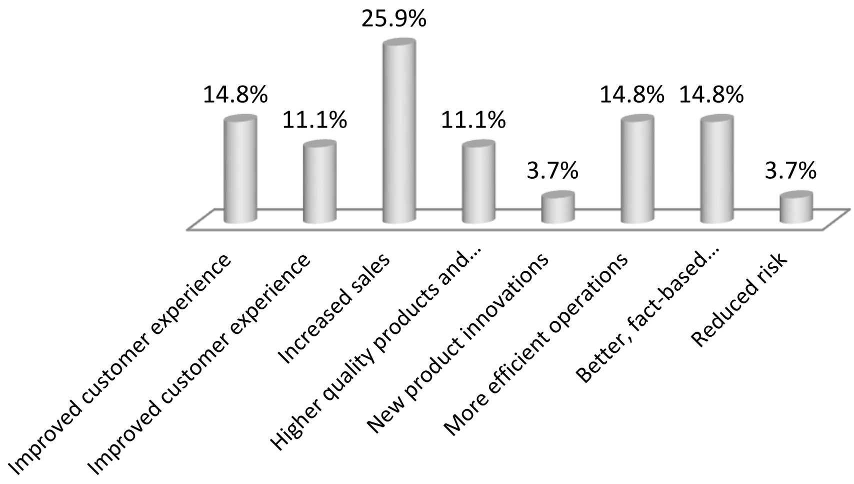
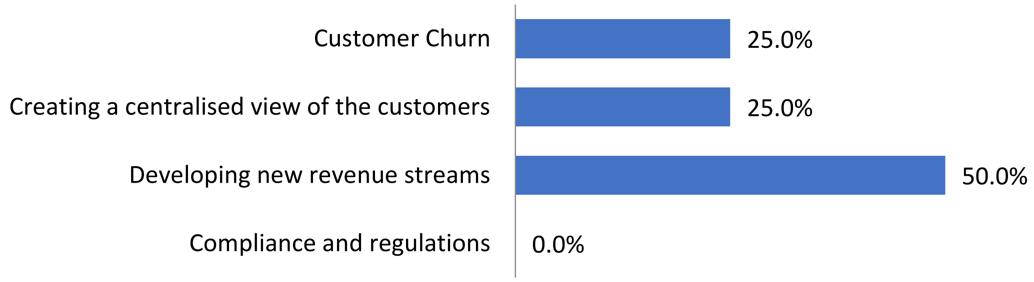


Figure 2.36 displays some of the potential focus areas for developing data mining models in the next years at Orbev. Some of the use cases could be creating new revenue streams as selected by 50% of respondents, and secondly by creating a centralized view of the customer (25%) and for customer churn (25%). DM models like market segmentation, could be developed to analyze the customer database in order to identify different groups and forecast their behavior (Fayyad et al., 1996) and potentially leading to new revenue streams. To create a better view of the customers, market basket analysis, which aim to understand what products or services are commonly purchased together, could also be used together with trend analysis uncover the difference between a typical customer present and future (Newton & Singh, 2013).

Figure 2. 36 Survey question consolidated: What do you see as the two main business challenges for the beverage industry over the next 5 years? (Survey Data Table in Appendix 1)



2.5. Existing use cases in the beverage industry and CPG

In the beverage industry, there are some companies already leveraging advanced analytics for a more data driven decision making. Banks (2012) states that while people are discussing about the "velocity" aspect of Big Data and how it might support organizations, CPG companies are early innovators who grabbed the opportunity to use Big Data and analytics tools to accelerate their business (Banks, 2012).

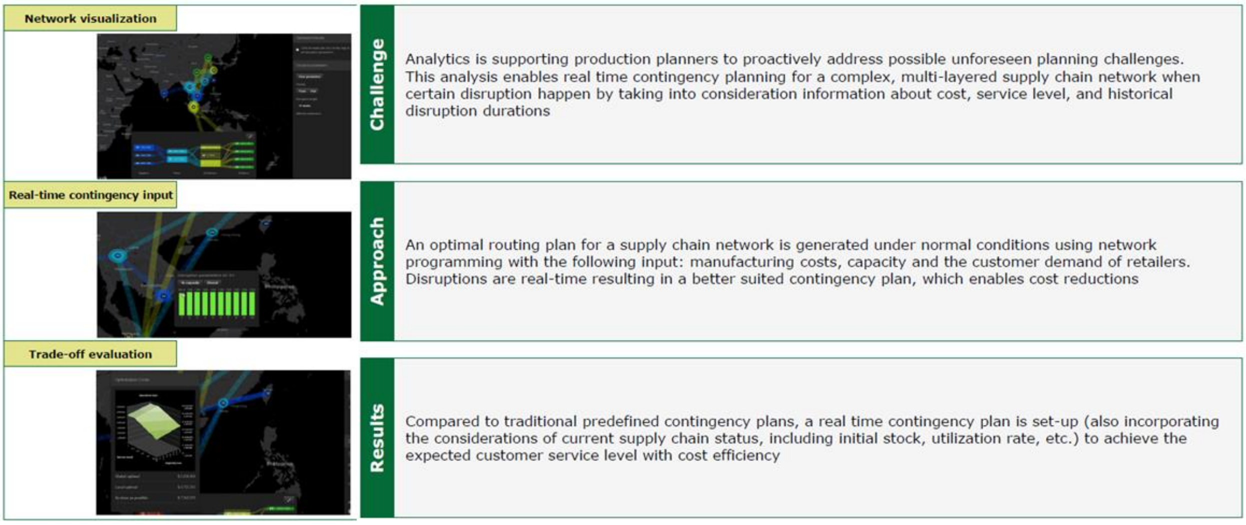
One of these companies is Nongfu Spring, one of the largest bottled water companies in China. In the past, Nongfu used to take two days to collect and produce reports on a very large volume of point-of-sale (POS) data from its retailers, and which later were used by executives to make their important business decisions. But Nongfu realized that two days was too long in the current competitive market and they start searching for options to deliver business reports at the speed of their business by applying up-to-date analytics on every transaction as it is happening (Banks, 2012).

To achieve this, the company started working with SAP and using new innovations like SAP HANA, allowing them to significantly reduce the lag time in producing insight on the retail data collected. The result was that the data preparation and loading, which previously took one day, was now being done in real-time. Banks (2012) refers that Nongfu currently "runs reports and queries 200 to 300 times faster", for example "one business process that took 24 hours to complete is now available in 37 seconds" (Banks, 2012:2). This provides the executives and sales force at Nongfu immediate insight supporting them in making smarter decisions depending on market dynamics.

According to Banks (2012) one of the reasons for this shift in speed was not only the technical side like in-memory computing and analytics software, but also the vision of their executives who facilitate these innovations, but who are also prepared to empower the end users to explore data and make decisions (e.g. production scheduling to increase the speed – and effectiveness - of the business) (Banks, 2012). Figure 2.37a/b shows two case studies presented by Deloitte in their 2017 FMCG Analytics Framework for production planning, where specific production planning variables are taking into account to allow real time contingency planning for a complex network in case of disruptions (Deloitte, 2017) and in Marketing/Sales for Omni channel voice of the customer, which performed analysis of the customer voice topics and sentiment across multiple channels

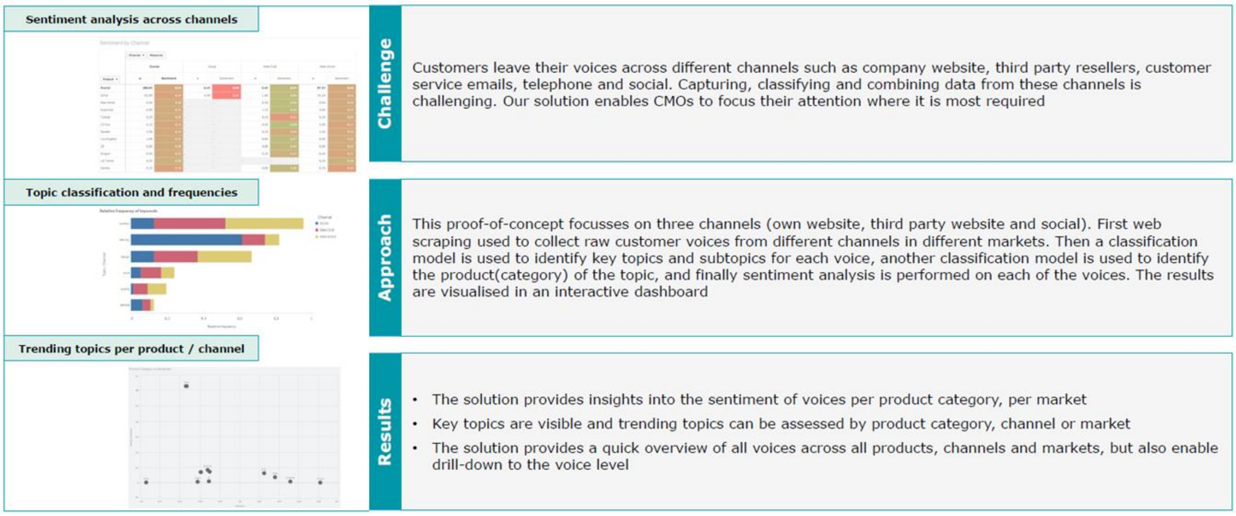
Figure 2. 37Case studies in: a) Production planning and b) Onmi channel voice of customer

a) Production planning. Deloitte (Deloitte, 2017)



Source: Deloitte (2017), p.21

b) Omni channel voice of the customer (Deloitte, 2017)



Source: Deloitte (2017), p. 10

Another example of a company leveraging on Big Data Analytics is Coca-Cola Company. In fact, the company is using BDA to produce orange juice with a consistent taste year-round. They have developed an algorithm, called the Black Book model that combines various data sets such as satellite imagery, weather date, expected crop yields, cost pressures, regional consumer preferences, and detailed data of over 600 different flavors that make up an orange. Those data are then matched to a profile with details on the acidity, sweetness and other attributes of each batch of raw juice and finally, the algorithm tells the company how to blend batches to replicate a certain taste and consistency (Stanford, 2013). According to a Bloomberg article Bippert, from Coca-Cola Company, stated that this algorithm allows better planning of supplies, because “If we have a hurricane or a freeze, we can quickly replan the business in 5 or 10 min just because we’ve mathematically modeled it” (Stanford, 2013).

In a post on BusinessWeek, Bob Cross, architect of Coke’s juice model Black Book, calls it “one of the most complex applications of business analytics. It requires analyzing up to 1 quintillion decision variables to consistently deliver the optimal blend, despite the whims of Mother Nature” (V. Rijmenam, 2018). Additionally, the company leverages POS data to build customer profiles, create centralized iPad reporting across the company and enable Collaborative Planning, Forecasting and Replenishment process within their supply chain using all data at hand (V. Rijmenam, 2018) and it is also on a mission to transform into a customer-centric and digitally-led brand by using data, artificial intelligence and connected devices to help customers make decisions in a more effective way (Swant, 2017).

Additionally, there are other CPG companies that have already started deriving value from advanced analytics, for example Nestlé was able to improve the sales precision in one of their business units in 9 percent by adopting SAS solutions (SAS, 2017):

> “Comparing semestral measures it is possible to notice an improvement of 9 percent in effectiveness of demand forecast. Consequently, it's already possible to note a 1 percent improvement in service customer levels, with less disruptive issues. These percentages represent a significant contribution to our business, compared to the previous scenario”,
>
> Coordinator of Planning at Nestlé Dolce Gusto, Sérgio Garnica

Another great example is Johnson & Johnson. The company is collaborating with Rest Devices, a global leader in infant sleep monitoring, to create and develop products that can help parents customize their baby’s care. As mentioned at the company’s website, one of the products is Nod™, which is “an app that can track the sleeping habits and patterns of infants and give recommendations on feedings and naps based on details like age and development” (Johnson & Johnson, 2018). Another product is Mimo Baby which is wearable tech that tracks data like body temperature and breathing. Nestlé is individualizing the AI to learn about each baby, with the goal that these products can help parents become much more informed about their baby's healthcare needs (Johnson & Johnson, 2018)

Bhatia (2017) refers that more and more CPG manufactures are looking for micro-targeting strategies (usually advanced predictive analytics) as applied by ecommerce businesses like Amazon and which are helping to unleash new and deeper insights into their targeted customer segments and allowing them to grow brands in a profitable way.by putting a laser focus on customers (Bhatia, 2017).

Another example is the international fashion retailer, GUESS? Inc, which has been using descriptive analytics to improve its decision- making capabilities and drive business actions by using a mobile analytics platform called GMobile. The company managed to turn fashion trends and customer data into insights by allowing buyers, planners and distributors to place 'the right apparel in the right store at the right time to appeal to its fashion-savvy shoppers' (M. van Rijmenam et al., 2018:11) and by visually displaying information (e.g. bestsellers or sales information) it has enabled employees to understand what markets to target (Teece, 2007). (M. van Rijmenam et al., 2018). Predictive analytics could therefore enable an organization to improve its processes (Kindström, Kowalkowski, & Sandberg, 2013) and respond to changes in their environment (Teece, 2007) by supporting more informed decisions.

CHAPTER III – Final Considerations

3.1. Deriving Value from Data Mining and Big Data Analytics

> "In an information society, information is money. The trick is to generate value by extracting the right information from the digital universe"
>
> (Gantz & Reinsel, 2011:2)

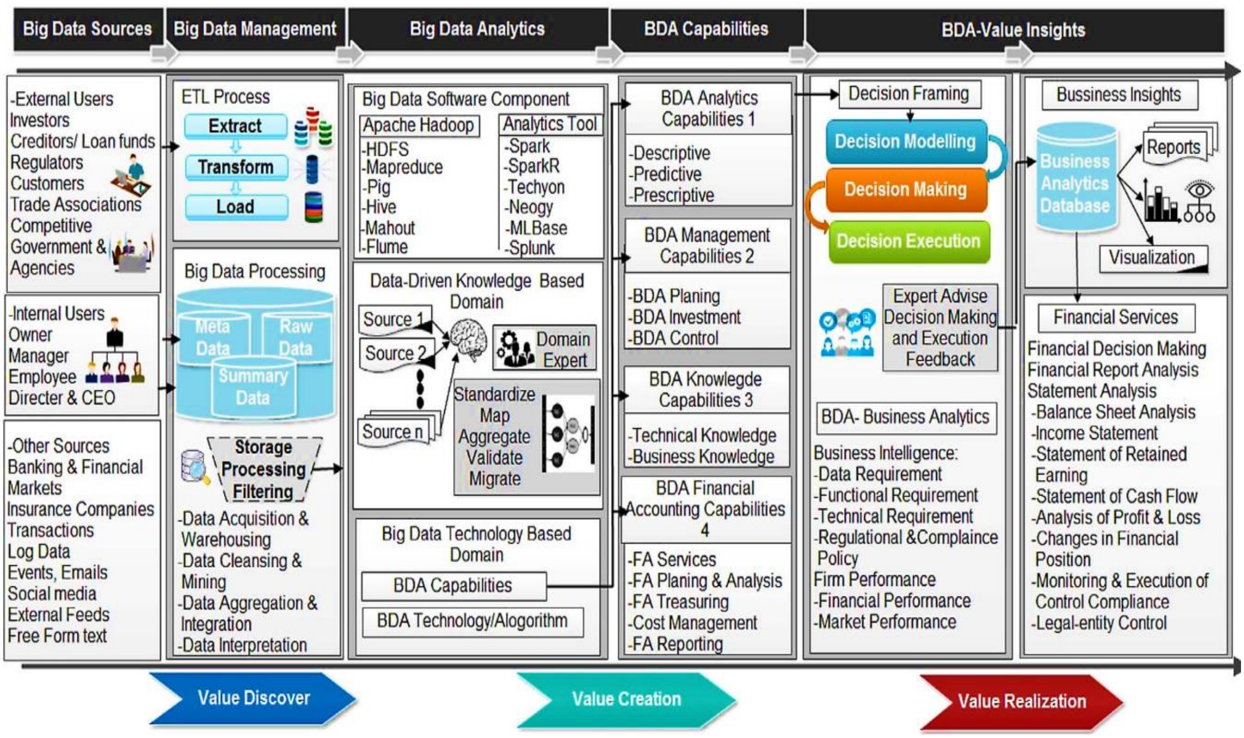
The value-creation is an important sustainability factor for companies in any industry, in addition to profit maximization, customer retention, business goals or revenue generation (Saggi & Jain, 2018). According to McKinsey (2018) there is a noticeable difference between companies that use data effectively and those that do not, which translate to a 1 percent margin improvement for those who use it (McKinseyAnalytics, 2018:58).

As seen in previous chapters, to extract knowledge from vast amounts of data, companies should have the right data ecosystem to run their models, be able to integrate the model insights into the business workflow by redesigning decision processes or developing user-friendly interfaces and ensure adoption by building front line and management capabilities and by looking into KPIs. To achieve this, companies should build their business around a common strategic vision, establishing the fundamentals, and generating momentum which typically takes two to three years according to McKinsey Analytics (2018). Therefore companies have only a narrow window in which to work or they will fall behind, as mentioned by one CEO working with McKinsey "It's no longer the big fish eating the small, but the fast ones eating the slow" (McKinseyAnalytics, 2018:41).

Gantz and Reinsel (2011) refer that datacenter architectures and organizational models will need to evolve because "big data applications are spreading through a company's infrastructure" (Gantz & Reinsel, 2011:7). In their view, cloud computing is facilitating the "consumption of IT as a service which could encourage organizations to increasingly consume IT as an external service" instead of internal infrastructure investments (Gantz & Reinsel, 2011:2). The authors also suggest that IT disciplines (from infrastructure to applications to governance) should be part of a single integrated team and work closely with users of big data in ways that are very distinct from traditional enterprise IT approaches (Gantz & Reinsel, 2011).

In order to help uncover value in the business ecosystem, Saggi and Jain (2018) proposed a Big Data Analytics and Decision-Making Framework (BDA-DMF) and they applied it to create value for financial & accounting companies. This framework could be a useful tool also in other industries, such as the beverage industry by providing support to the Finance Business Unit. The BDA-DMF is shown in Figure 3.1 where five components were defined – Big Data Source, Data Management, Big Data Analytics, BDA capabilities and BDA Value-Insights – and a three-phase method was proposed, by which big data analytics could create value for companies.

Figure 3.1 Framework for big data analytics and business insights for financial accounting proposed by Saggi & Jain (2018), p.777



Source: Saggi & Jain (2018), p.777

The first phase is "Value Discover" and it includes: (1) Big Data Sources, (2) Big Data Management, and in this phase BDA could "create new insights that improve business-driven decision-making" by for example, showing how the company "can improve customer satisfaction and the specific features of the service experience" (Saggi & Jain, 2018:778). Leavitt (2013) stresses that BDA aim at three major values to discover when implementing BDA technology: (1) minimize hardware costs, (2) check the value of big data before committing significant company resources, and (3) reduce processing costs (Leavitt, 2013). To achieve this, the author adds the importance of alignment between business objectives, the big data storage and analytics approach.

The second phase is "Value Creation" and it encompasses (3) Big Data Analytics and (4) Big Data Analytics Capabilities. McKinsey & Co. added 'Value' as the fourth 'V' to define one of the characteristics of big data, besides volume, velocity, variety (M. Chen, Mao, & Liu, 2014) where the value refers to the worth of hidden insights inside big data and represents the "transactional, strategic, and informational benefits of big data" (Saggi & Jain, 2018:778). Wamba et al (2015) add that value represents the extent to which big data generates economically relevant insights and benefits through extraction and transformation.

The third phase "Value Realization" includes (5) Big Data Analytics-Value Insights and it focus is on the development of big data analytics technology-based solutions to the customer and it usually requires a strategic transformation (Saggi & Jain, 2018). For example, in the case an accounting firm, it could support the collection and processing of vast amounts of raw accounting data, organizing it to allow an accurate assessment for identification of an adequate course of actions. Saggi et al (2018) believe that for companies to "effectively deliver an analytics capability, business customers need to be assisted with decision-framing, decision-making, and decision-execution"(Saggi & Jain, 2018:780).

Additionally, Ransbotham & Kiron (2017) mention that to extract strategic value from analytics, it is essential that firms innovate by moving from general-purpose to specialized analytics uniquely optimized to address specific business issues; and by eliminating or reducing organizational silos to coordinate data sharing and analytics across functional units (Ransbotham & Kiron, 2017). As Venture Capitalist Bryce Roberts (2012) states "Data, big, medium or small, has no value in and of itself. The value of data is unlocked through context and presentation" (Power, 2013:2).

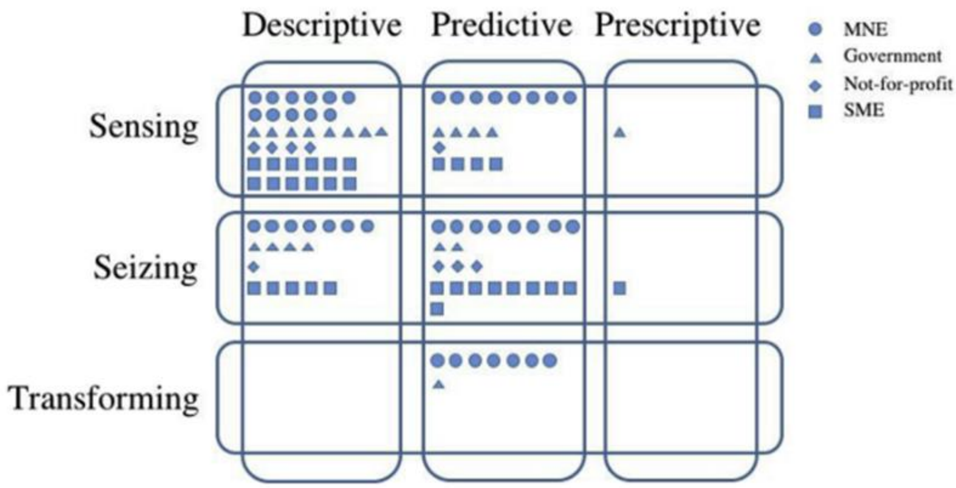
To achieve this, companies need to have the right capabilities and an agile business, where aagility is defined by the "capacity of an organization to efficiently and effectively redeploy/redirect its resources to value creating and value protecting (and capturing) higher-yield activities as internal and external circumstances warrant" (Teece, Peteraf, & Leih, 2016 apud Côrte-Real et al., 2017:382) For analytical purposes, dynamic capabilities can be disaggregated into three distinct activities:

1. sensing opportunities and threats (understand customers, market trends and technological changes);
2. seizing those opportunities, and
3. transforming in order to maintain competitiveness (Teece, 2007).

Kindström et al (2013) state that companies that have dynamic capabilities are the ones that align internal processes and routines, decision-making and culture to seize the opportunities that have been sensed (Kindström et al., 2013) and they achieve this by determining what technologies to use, what business models to apply and what specific market segments to target (Teece, 2007).

Van Rijmenam et al (2018) research on Big Data revealed that most organizations in their study (44 companies) were Multinational Enterprises (MNEs), which have applied big data analytics in different ways for different use cases as shown in Figure 3.2. Their research demonstrated that BDA can be seen as a dynamic capability that support in understanding the environment and which assist "managers to take action and provides organizations with sustained superior performance and competitive advantage in times of ambiguity and uncertainty" (M. van Rijmenam et al., 2018:15).

Figure 3.2 Analytics vs use case vs organization proposed by Van Rijmenam et al (2018)



Source: Van Rijmenam et al (2018), p.15

According to Teece (2007) and Wang and Ahmed (2007), sensing, seizing and transforming are critical for sustaining profitable growth (M. van Rijmenam et al., 2018).

In the end, the real value of a big data implementation will be judged based on "one or more of three criteria: Does it provide more useful information? ; Does it improve the fidelity of the information? ; Does it improve the timeliness of the response?" (Gantz & Reinsel, 2011:8) . Companies will need to have metrics that explicitly measure the value of analytics and communicate these metrics to the business users, which is not an easy task since analytics are usually used to support decisions and the value cannot always be isolated from other projects (McKinseyAnalytics, 2018). Therefore McKinsey recommends "using a case-driven approach, quantifying the impact of each potential use case, and balancing it with the level of effort required to implement it" (McKinseyAnalytics, 2018:70), for example, in the case of customer churn, for each analytics use case the associated outcome metrics could be reviewed by asking "If the use of analytics decreases customer churn by 2 percent, how much savings does that translate into?" (McKinseyAnalytics, 2018:39).

Studies from van Rijmenam et al have found that both descriptive and predictive analytics help organizations to sense and seize opportunities in changing environments and both enable them to turn data into information, thereby offering a competitive advantage (George et al., 2014; Gabel and Tokarski, 2014; Pigni et al., 2016; Fitzgerald, 2016a apud M. van Rijmenam et al., 2018). Moreover, their research suggested that predictive analytics allow companies to improve their decision-making processes by "providing the historical context, but also by recommending the best course of action to be taken based on the full context of the environment” (M. van Rijmenam et al., 2018:14).

3.2. Discussion

> “The growth of the digital universe may be a challenge, but it is also a propellant for new and exciting uses of data.”
>
> (Gantz & Reinsel, 2011:12)

The data generated is doubling in size every two years, and by 2020 the data created and copied annually, i.e. the digital universe, “will reach 44 zettabytes, or 44 trillion gigabytes (EMC DigitalUniverse, 2018). Moreover, IDC Corporate USA (2018) states that by 2021, “20% of G2000 manufacturers will depend on embedded intelligence, IoT, blockchain, and cognitive, to automate large-scale processes and speed execution times by up to 25%” (IDC, 2018) and machine-generated data seems to be a key driver for future data growth.

However, a 2012 IDC Digital Universe Study pointed out that there is a “Big Data gap since less than 1% of world data is analyzed; and less than 20% is protected” (KDnuggets, 2012). Martens et al (2016) also mentioned that companies have access to massive, fine-grained data on consumer behavior, but “only a few organizations have incorporated such data in a non-aggregated manner into their predictive analytics”(Martens, Provost, Clark, & de Fortuny, 2016:1).Another two gaps were identified by Gantz and Reinsel (2011) as they refer that the amount of information managed by companies’

> “datacenters will grow by a factor of 50, and the number of files the datacenter will have to deal with will grow by a factor of 75, while the number of IT professionals in the world will grow by less than a factor of 1.5”
>
> (Gantz & Reinsel, 2011:4).

This means that companies are lacking the skills and experience to manage the exponential growth of data collected to be able to extract new knowledge from data. The third gap identified by IDC is related to "the mismatch between the value of some data and the value of other data" (Gantz & Reinsel, 2011:11), where there is the need to identify which data and variables are relevant and can create value for the company, and the ones that don't create value.

It becomes evident that to generate business insights and new use case ideas there is a need to build the right organizational capabilities to uncover new use cases, new data features, run models or perform analysis. McKinsey Analytics states that the combination of people, novel digital technologies, and advanced analytics could produce a new breakthrough in productivity if companies learn how to blend them all together (Alldredge, Henry, Lowrie, & Rocha, 2016). Data Mining is a useful advanced analytic tool, but it is not a stand-alone tool, which means that it will only be effective if combined with available and accurate data, proper data engineering, and the talent to guide insights and insights implementation by integrating it in strategic decision making and in daily business operations.

Van Rijmenam et al (2018) studies contributed to the understanding of the "importance of big data analytics to obtain a better understanding of an organization's context, which improves an organization's decision-making" and, could be a factor of competitive advantage (M. van Rijmenam et al., 2018:14). The insights from advance analytics could support decision in different ways, as suggested by Deloitte, allowing to make same decisions faster, allowing to make the same decisions but cheaper, supporting better /informed decisions and assisting in innovations in products and services (Deloitte, 2017)

There is still much research needed in both technical and organizational components of big data analytics. To close the gaps identified by IDC over the last years, some further questions need to be addresses. Some of the aspects that could be used for further research within the beverage industry could be:

1. **How to close the data gap between data collected and data analyzed?** Which new tools could be used to create metadata? How could metadata be used efficiently for the discovery, interpretation and use of data? Could the application

of virtualization solutions be used to support BDA analytics? Does the company have a governance framework that provides consistent guidance, procedures and processes for data capture and management? Does the company have the necessary quality and monitoring parameters for big data to deal with unstructured data?

2. **How to close the gap between the data generated every day and the slow growth in data experts?** Should organizations invest in current talent or should they outsource analytics capabilities? What is the role of leadership in boosting the use of data for decision making? Can new technologies like cloud computing be used to reduce this gap? Does the company have a data culture that supports data analytics and data-driven decision making?

3. **How to close the gap between value of certain data and of other?** What is the financial value of a targeted business initiative? Which data projects and use cases have the most financial impact? What are the requisites for data sets and analytical tools for certain use case? How can companies develop and manage advanced storage management tools? How will the company ensure the privacy of personal information? And how will it implement information security strategies?

Conclusion

This research project on Data Mining aimed at adopting a more industry-oriented approach, focused on the beverage industry, following Pechenizkiy et al (2018) suggestion that DM research should be taking the needs of industry into account, because the analytics outputs differ significantly in different industries, "affecting the meaning and measurement of utility and performance" (Pechenizkiy et al., 2008:245)**.** In order to answer the research question on how data mining can support the decision-making process in the beverage industry, five sub-questions were defined. First sub-question aimed at understanding what type of data is generated by the company by looking into data sources, data types and data analytics applications, and which data influence in the decision making. The results showed the company deals with Big Data by identifying the 4 Vs Volume, Velocity, Variety, Value in the data generated and collected and named commercial, supply-chain, HR, Finance and IT as the main users of data and data analytics.

The second sub-question focused on the use and application of data mining methods and techniques at OrBev to understand usability and insights integration. Some of the big data analytics already used are predictive analytics and data visualization, and with less frequency data mining or embedded analytics. The third sub-question looked into the existing decision-making process and how could data support tactical and strategic decision-making. Results show that data analytics could be used to inform strategic decision-making, to make real-time decisions or to improve day-to-day business operations. The fourth sub-question was designed to understand the adoption rate of data mining models and tools in the company and in the beverage industry, showing the beverage industry in general is already advanced in term of BDA use and integration. The last sub-question investigated which DM models could be additionally explored to solve business problems.

CPG companies in general and the beverage sector, has a large amount of data available for analytics, but they struggle to turn data it into actionable information. Data mining was defined by Fayyad e Uthurusamy (2002) as the identification of interesting structure in data and its main goal is to produce new knowledge that the user can act upon by building a model of the business environment based on data collected from a variety of sources (including, corporate transactions, customer histories and demographic information, process control data, weather data, etc.) (Two Crows Corporation, 2005). This research showed that data mining applications still face significant issues with usability of the data mining and BDA deliverables which may explain the relatively low-medium usage in the beverage sector. Cao et al (2010) suggested that the next-generation of data mining methodologies and techniques should “aim for a paradigm shift from data-centered hidden pattern mining into a more domain-driven actionable knowledge delivery” (Cao et al., 2010:1310).

Data mining is a tool that can be source of competitive advantage by allowing companies to better understand their environment, take more informed decisions and predict future consumer behaviors (e.g. using predictive mining). In order to implement Data Mining projects, it is necessary not only an excellent collection of DM algorithms and talented data miners, but also that business use cases and opportunities are selected carefully and should support other company efforts. *During the application of data analytics it is critical to make sense of information across the businesses a whole to gain insights, and employees to make decisions based on these (Medeiros et al., 2018).* Davenport (2006) states that as business processes become major differentiators for organizations in many industries, organizations are increasingly using analytics to “wring every last drop” of value from those processes (Davenport, 2006). McKinsey also note the evident consequences of BDA across sectors as “competition is intensifying not just within industries, but also between them” (McKinseyAnalytics, 2018:72).

Data mining can bring many benefits in extracting insights from data, but it cannot be a stand-alone tool. In order to transform data into useful insights, data mining needs to be combined with business understanding and use case identification, data collection and data engineering, deployment of models and algorithms and proper interface for data visualization and data integration. The present study intents to give an overview of the use of data mining for decision-making in a specific industry, but one of the limitations was the low response rate of the survey instrument. For a more accurate view on the subject further research would be required using a bigger sample, either a bigger population or several companies within the sector. Moreover, it was challenging to find use-cases in the beverage industry across the academic review and many of the examples presented are retrieved from articles and reports. Furthermore, some topics were identified as crucial for data analytics and could be object of further as for example, how to build a data culture that enhances and facilitates the use of data and data insights, what is the more appropriate technological infrastructure to support data use within beverage industry, how to address the lack of human capabilities and data literacy among business users.

References

Abbasi, A., Sarker, S., & Chiang, R. (2016). Big Data Research in Information Systems: Toward an Inclusive Research Agenda. *Journal of the Association for Information Systems*, *17*(2), I–XXXII. https://doi.org/10.17705/1jais.00423

Alldredge, K., Henry, J., Lowrie, J., & Rocha, A. (2016). Wining in consumer packaged goods through data and analytics. Retrieved from https://www.mckinsey.com/industries/consumer-packaged-goods/our-insights/winning-in-consumer-packaged-goods-through-data-and-analytics

Ankerst, M. (2002). Report on the SIGKDD-2002 Panel the Perfect Data Mining Tool: Interactive or Automated?". *ACM SIGKDD Explorations Newsletter*, *4*(2), 110–111.

Banerjee, A., Bandyopadhyay, T., & Acharya, P. (2013). Data Analytics: Hyped Up Aspirations or True Potential? *Vikalpa: The Journal for Decision Makers*, *38*(4), 1–11. Retrieved from http://widgets.ebscohost.com/prod/customerspecific/ns000290/authentication/index.php?url=https%3A%2F%2Fsearch.ebscohost.com%2Flogin.aspx%3Fdirect%3Dtrue%26AuthType%3Dip%2Ccookie%2Cshib%2Cuid%26db%3Dbth%26AN%3D93404757%26lang%3Dpt-br%26site%3Deds-live%26sc

Banks, B. (2012). Why Consumert Packaged Goods Companies Love Big Data. Retrieved July 29, 2018, from https://www.forbes.com/sites/sap/2012/03/27/why-consumer-packaged-goods-companies-love-big-data/#18973e447fae

Benson-Armer, R., Noble, S., & Thiel, A. (2015). The consumer in 2030: trends and questions to consider. Retrieved August 19, 2018, from https://www.mckinsey.com/industries/consumer-packaged-goods/our-insights/the-consumer-sector-in-2030-trends-and-questions-to-consider

Berman, J. . (2013). *Principles of Big Data Prepararing, Sharing, and Analyzing Complex Information*. Waltham: Elsevier Ltd.

Bhatia, R. (2017, September). Can Micro-segmentation save CPG companies find niche customers in the age of personalization. *Analytics India Magazine*. Retrieved from https://www.analyticsindiamag.com/can-micro-segmentation-save-cpg-companies-find-niche-customers-age-personalization/

Breuer, P., & Moulton, J. (2013). Applying advanced analytics in consumer companies. Retrieved from https://www.mckinsey.com/industries/consumer-packaged-goods/our-insights/applying-advanced-analytics-in-consumer-companies

Cao, L., Yu, P., & Zhang, C. (2008). *Data Mining for Business Applications*. (Springler, Ed.).

Cao, L., Yu, P., & Zhang, Y. (2009). *Domain Driven Data Mining*. Springer.

Cao, L., & Zhang, C. (2007a). Knowledge Actionability: Satisfying Technical and Business Interestingness. *Int'l J. Business Intelligence and Data Mining*, *2*(4), 496–514.

Cao, L., & Zhang, C. (2007b). The Evolution of KDD: Towards Domain-Driven Data Mining. *Int'l J. Pattern Recognition and Artificial Intelligence*, *21*(4), 677–692.

Cao, L., Zhao, Y., Zhang, H., Luo, D., Zhang, C., & Park, E. K. (2010). Flexible Frameworks for Actionable Knowledge Discovery. *IEEE Transactions on Knowledge and Data Engineering, Knowledge and Data Engineering, IEEE Transactions on, IEEE Trans. Knowl. Data Eng.*, *22*(9), 1299–1312. https://doi.org/10.1109/TKDE.2009.143

Charest, M., Delisle, S., Cervantes, O., & Shen, Y. (2008). Bridging the gap between data mining and decision support: A case-based reasoning and ontology approach. *Intelligent Data Analysis*, *12*, 211–236. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=36155153&site=ehost-live

Chen, H., Chiang, R., & Storey, V. (2012). BUSINESS INTELLIGENCE AND ANALYTICS: FROM BIG DATA TO BIG IMPACT. *MIS Quarterly, 36*(4), 1165–1188. Retrieved from http://widgets.ebscohost.com/prod/customerspecific/ns000290/authentication/index.php?url=https%3A%2F%2Fsearch.ebscohost.com%2Flogin.aspx%3Fdirect%3Dtrue%26AuthType%3Dip%2Ccookie%2Cshib%2Cuid%26db%3Dheh%26AN%3D83466038%26lang%3Dpt-br%26site%3Deds-live%26sc

Chen, M., Mao, S., & Liu, Y. (2014). Big Data: A Survey. *Mobile Networks and Applications*, *19*(2), 171–209.

CNECT. (2015). Questionnaire on the European Data-Driven Economy. Retrieved September 6, 2018, from https://www.ihk-nuernberg.de/de/media/PDF/Innovation-Umwelt/e-commerce/umfragebogen-der-europaeischen-kommission-zu-big-data.pdf

Côrte-Real, N., Oliveira, T., & Ruivo, P. (2017). Assessing business value of Big Data

Analytics in European firms. *Journal of Business Research*, *70*, 379–390. https://doi.org/https://doi.org/10.1016/j.jbusres.2016.08.011

Daas, D., Overbeek, S., Bouwman, H., & Hurkmans, T. (2012). Developing a decision support system for business model design. *Electronis Markets*. Retrieved from http://widgets.ebscohost.com/prod/customerspecific/ns000290/authentication/index.php?url=https%3A%2F%2Fsearch.ebscohost.com%2Flogin.aspx%3Fdirect%3Dtrue%26AuthType%3Dip%2Ccookie%2Cshib%2Cuid%26db%3Dedswss%26AN%3D000324069400008%26lang%3Dpt-br%26site%3Deds

Davenport, T. H. (2013). Analytics 3.0. *Harvard Business Review*, *91*(12), 64–72. Retrieved from http://widgets.ebscohost.com/prod/customerspecific/ns000290/authentication/index.php?url=https%3A%2F%2Fsearch.ebscohost.com%2Flogin.aspx%3Fdirect%3Dtrue%26AuthType%3Dip%2Ccookie%2Cshib%2Cuid%26db%3Dheh%26AN%3D92545710%26lang%3Dpt-br%26site%3Deds-live%26sc

Deloitte. (2017). *FMCG Analytics Framework*. Retrieved from https://www2.deloitte.com/content/dam/Deloitte/nl/Documents/consumer-business/deloitte-nl-cip-fmcg-analytics-framework.pdf

Dong, C.-S. J., & Srinivasan, A. (2013). Agent-enabled service-oriented decision support systems. *Decision Support Systems*, *55*, 364–373. Retrieved from http://10.0.3.248/j.dss.2012.05.047

Dong, G. (2009). *Sequence Data Mining*. Berlin: Springer Science & Business Media B.V. Retrieved from https://dl.acm.org/citation.cfm?id=1822819

EMC DigitalUniverse. (2018). The digital universe of opportunities:rich data and increasing value of the Internet of Things.

Fayyad, Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *Social Science Research*, *17*(3), 37. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&cmd=Retrieve&dopt=AbstractPlus&list_uids=14728591308876110427related:W1p3j8J3ZswJ%5Cnpapers2://publication/uuid/ABD8C83F-C4AA-4694-8369-5651091CBA9E

Fayyad, U., & Uthurusamy, R. (2002). EVOLVING DATA MINING INTO SOLUTIONS FOR INSIGHTS. *Communications of the ACM*, *45*(8), 28–31. Retrieved from http://widgets.ebscohost.com/prod/customerspecific/ns000290/authentication/index.p hp?url=https%3A%2F%2Fsearch.ebscohost.com%2Flogin.aspx%3Fdirect%3Dtrue%26AuthType%3Dip%2Ccookie%2Cshib%2Cuid%26db%3Dbth%26AN%3D11863425%26lang%3Dpt-br%26site%3Deds-live%26sc

Gantz, J., & Reinsel, D. (2011). *Extracting Value from Chaos*.

Gartner. (2018). Gartner IT Glossary. Retrieved August 5, 2018, from https://www.gartner.com/it-glossary/data-mining

Gartner IT Glossary. (2018). Retrieved July 29, 2018, from https://www.gartner.com/it-glossary/big-data

Han, J. (2018). Sequantial Patterns and Sequential Pattern Mining. Retrieved September 10, 2018, from https://www.coursera.org/lecture/data-patterns/5-1-sequential-pattern-and-sequential-pattern-mining-REbEU

He, Z., Xu, X., & Deng, S. (2005). Data Mining for Actionable Knowledge: A Survey. Retrieved from http://arxiv.org/abs/cs/0501079

Hormozi, A. M., & Giles, S. (2004). Data mining: A competitive weapon for banking and retail industries. *Information Systems Management*, *21*(2), 62–71. https://doi.org/10.1201/1078/44118.21.2.20040301/80423.9

Ichijo, K., & Nonaka, I. (2006). *Knowledge Creation and Management* (1st ed.). Oxford University Press.

IDC. (2018). Product as a Service.

Jaikumar, V. (2016). Customer segmentation analysis is key to competing in the shifting CPG industry. Retrieved August 19, 2018, from http://www.ibmbigdatahub.com/blog/customer-segmentation-analysis-key-competing-shifting-cpg-industry

Johnson & Johnson. (2018). How Artificial Intelligence Is Revolutionizing Medical Technology. Retrieved November 9, 2018, from http://www.careers.jnj.com/careers/how-artificial-intelligence-is-revolutionizing-medical-technology

Kannimuthu, S., Premalatha, K., & Usha, G. (2015). Survey of recent developments in utility based Data mining 1, *2*(9), 894–901.

KantarWorldpanel. (2018). *Brand Footprint 2018*. Retrieved from https://www.kantarworldpanel.com/global/news/Brand-footprint-2018-most-chosen-brands#downloadThankyou

Kautkar, R. (2014). A comprehensive survey on data mining, 185–191.

KDnuggets. (2012). IDC Study: Digital Universe in 2020.

Kelly, G., Kopka, U., Küpper, J., & Moulton, J. (2018). The new model for consumer goods. Retrieved August 19, 2018, from https://www.mckinsey.com/industries/consumer-packaged-goods/our-insights/the-new-model-for-consumer-goods

Kesner, R. M. (2010). The Evolving Symbiosis between Decision Support and Knowledge Management Systems : A Study in Emerging Industry Practices, *3*(2), 1–28.

Kindström, D., Kowalkowski, C., & Sandberg, E. (2013). Enabling service innovation: A dynamic capabilities approach. *Journal of Business Research*, *66*(8), 1063–1073. https://doi.org/https://doi.org/10.1016/j.jbusres.2012.03.003

Kitchens, B., Dobolyi, D., Li, J., & Abbasi, A. (2018). Advanced Customer Analytics: Strategic Value Through Integration of Relationship-Oriented Big Data. *Journal of Management Information Systems*, *35*(2), 540–574. Retrieved from http://10.0.4.56/07421222.2018.1451957

Langley, P., & Simon, H. A. (1995). Applications of Machine Learning and Rule Induction. *Communications of the ACM*, *38*(11), 55–64.

Leavitt, N. (2013). Bringing big analytics to the masses. *Computer*, *46*(1), 20–23.

Ling, C. X., Chen, T., Yang, Q., & Chen, J. (2002). Mining optimal actions for intelligent CRM. In *Proc. of ICDM O2* (pp. 767–770).

Liu, S., Duffy, A. H. B., Whitfield, R. I., & Boyle, I. M. (2010). Integration of decision support systems to improve decision support performance. *Knowledge Information Systems*, *22*(3), 261–286.

Lodhi, O. S. (2012). Most commonly used techniques in data mining. *International Journal of Advances in Engineering Research (IJAER)*, *4*(III).

Martens, D., Provost, F., Clark, J., & de Fortuny, E. J. (2016). MINING MASSIVE FINE-GRAINED BEHAVIOR DATA TO IMPROVE PREDICTIVE ANALYTICS. *MIS Quarterly*, *40*(4), 869–888. Retrieved from http://widgets.ebscohost.com/prod/customerspecific/ns000290/authentication/index.php?url=https%3A%2F%2Fsearch.ebscohost.com%2Flogin.aspx%3Fdirect%3Dtrue%26AuthType%3Dip%2Ccookie%2Cshib%2Cuid%26db%3Dheh%26AN%3D119473688%26lang%3Dpt-br%26site%3Deds-live%26s

McKinseyAnalytics. (2018). *Analytics comes of age*. Retrieved from https://www.mckinsey.com/~/media/McKinsey/Business Functions/McKinsey Analytics/Our Insights/Analytics comes of age/Analytics-comes-of-age.ashx

Medeiros, A., Lauster, S., Veldhoen, S., & Soundararajan, R. (2018). *Organic Growth Barometer 2018*. Retrieved from https://assets.kpmg.com/content/dam/kpmg/uk/pdf/2018/06/kpmg-cpg-organic-growth-barometer-2018.pdf

Moore, G. C., & Benbasat, I. (1991). Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation. *Information Systems Research*, *2*(3), 192–222.

Newton, P. N., & Singh, S. (2013). Data Mining in Decision Support System. *Proceedings of National Conference on Emerging Trends: Innovations and Challenges in IT*, (April), 19–20.

NewVantagePartners. (2018). *Big Data Executive Survey*.

Pathak, M., Singh, S., & Oberoi, S. . (2013). Impact of Data Warehousing and Data Mining in Decision Making. *International Journal of Computer Science and Information Technologies*, *4*(6), 995–999.

Pechenizkiy, M., Puuronen, S., & Alexey, T. (2005). Competitive advantage from data mining: some lessons learnt in the information systems field. *16th International Workshop on Database and Expert Systems Applications (DEXA'05), Database and Expert Systems Applications, 2005. Proceedings. Sixteenth International Workshop on, Database and Expert Systems Applications*. Los Alamitos, CA, USA, USA: IEEE. https://doi.org/10.1109/DEXA.2005.64

Pechenizkiy, M., Puuronen, S., & Tsymbal, A. (2008). Towards more relevance-oriented data mining research. *Intelligent Data Analysis*, *12*, 237–249. https://doi.org/10.5465/AMLE.2006.21253789

Philip Chen, C. L., & Zhang, C.-Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences*, *275*, 314–347.

Poleto, T., Heuer de Carvalho, V. D., & Cabral Seixas Costa, A. P. (2015). The roles of Big Data in the Decision-Support Processes: An empirical Investigation. In *Proceedings of the First International Conference, ICDSST*, *10-21* (pp. 10–21).

Retrieved from https://pdfs.semanticscholar.org/c125/ee02945551332deceead0e32d3fdb0efb196.pdf

Pomerol, J.-C., & Adam, F. (2004). Practical Decision Making–From the Legacy of Herbert Simon to Decision Support Systems. In *IFIP WG 8.3 Conference, Prato, Italy* (pp. 647–657).

Power, D. (2013). Using 'Big Data' for analytics and decision support. In *MWAIS 2013 Proceedings* (pp. 1–5). Retrieved from https://pdfs.semanticscholar.org/17ed/6e8d8b35cb4422d0f959e4ccb808f7bf5a3a.pdf

Provost, F., & Fawcett, T. (2013). *Data Science for Business*. (L. Mike & B. Meghan, Eds.) (1st Editio). O'Reilly Media Inc.

QSR. (2016). How auto coding sentiment works.

Ransbotham, S., & Kiron, D. (2017). Analytics as a Source of Business Innovation. *MITSloan Management Review*, 1–19.

Reinsel, D., Gantz, J., & Rydning, J. (2017). *Data Age 2025: The Evolution of Data to Life-Critical*.

Rijmenam, V. (2018). How Coca-Cola Takes A Refreshing Approach On Big Data.

Rogers, E. (1983). *Diffusion of innovations* (3rd Editio). New York: The Free Press.

Saggi, M. K., & Jain, S. (2018). A survey towards an integration of big data analytics to big insights for value-creation. *Information Processing and Management*, *54*(In (Big) Data we trust: Value creation in knowledge organizations), 758–790. Retrieved from http://10.0.3.248/j.ipm.2018.01.010

SAS. (2017). Nestlé enhances demand forecast with SAS analysis solutions. Retrieved November 10, 2018, from https://www.sas.com/de_ch/news/press-releases/2017/oktober/2017-10-12-nestle-enhances-demand-forecast-with-sas-analysis-solutions.html

Schmidt, F. (2018). Finding CPG Growth in Emerging Megacities. Retrieved August 19, 2018, from https://www.retailnetgroup.com/research/post.aspx?id=105699

Shim, J. P., Warkentin, M., Courtney, J. F., Power, D. J., Sharda, R., & Carlsson, C. (2002). Past, present, and future of decision support technology. *Decision Support Systems*, *33*(Decision Support System: Directions for the Nest Decade), 111–126. Retrieved from http://10.0.3.248/S0167-9236(01)00139-7

Simon, H. A. (1977). *The New Science of Management Decision*. Prentice Hall College

Div.

Stanford, D. (2013). Coke Engineers Its Orange Juice—With an Algorithm.

Swant, M. (2017, December). Coca-Cola Is Embracing AI and Chatbots in Preparation for a Digital-First Future. *Adweek Magayine*. Retrieved from https://www.adweek.com/digital/coca-cola-is-embracing-ai-and-chatbots-in-preparation-for-a-digital-first-future/

Teece, D. J. (2007). Explicating dynamic capabilities: the nature and microfoundations of (sustainable) enterprise performance. Strategictle. *Strategic Management Journal*, *28*, 1319–1350. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.321.5849&rep=rep1&type=pdf

van Hulse, J. (2007). *Data quality in data mining and machine learning*. Florida Atlantic University Boca Raton, FL, USA. Retrieved from https://dl.acm.org/citation.cfm?id=1292676

van Rijmenam, M., Erekhinskaya, T., Schweitzer, J., & Williams, M.-A. (2018). Avoid being the Turkey: How big data analytics changes the game of strategy in times of ambiguity and uncertainty. *Long Range Planning*. https://doi.org/https://doi.org/10.1016/j.lrp.2018.05.007

Venkatesh, V., Brown, S., & Bala, H. (2013). Bridging the Qualitative–Quantitative Divide: Guidelines for Conducting Mixed Methods Research in Information Systems. *Management Information Systems Quarterly*, *37*(1), 21–54.

Yang, Q., Yin, J., Lin, C. X., & Chen, T. (2003). Postprocessing decision trees to extract actionabl knowldge. In *Proc. of ICDM’03*.

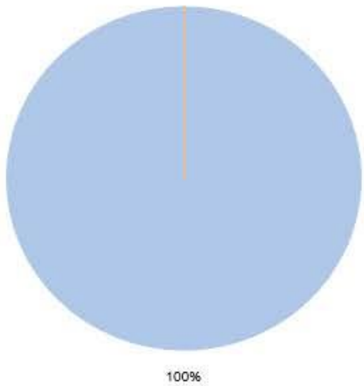
Annex I – Survey Instrument IT

15 people have viewed your survey
4 people have answered your survey

Data Analytics Survey

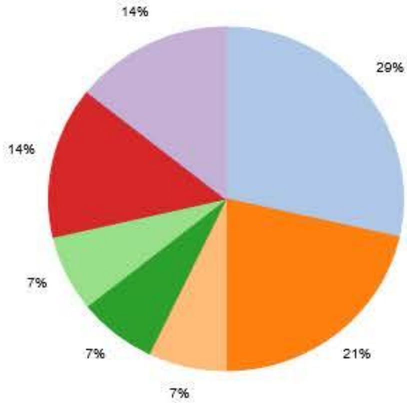
1. How far do you agree with the following statement? "The beverage industry has a clear understanding of the benefits of big data."

Strongly agree	0	0%
Agree	4	100%
Disagree	0	0%
Strongly disagree	0	0%
Total answers:	4	



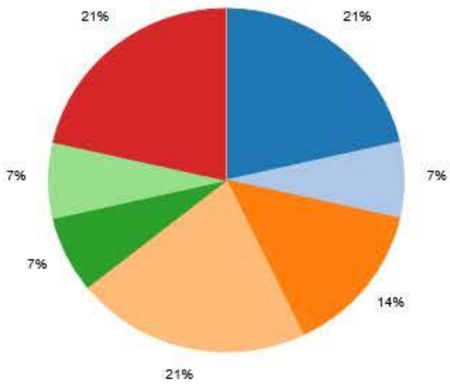
2. At OrBev, which functional groups rely most heavily on big data?

Finance	0	0%
Commercial	4	28.57%
Supply Chain	3	21.43%
Human Resources/ Talent Management	1	7.14%
Technology & Innovation	1	7.14%
Research and development	1	7.14%
Strategic planning	2	14.29%
PA&C	0	0%
Customer service	0	0%
BSS	2	14.29%
Total answers:	14	



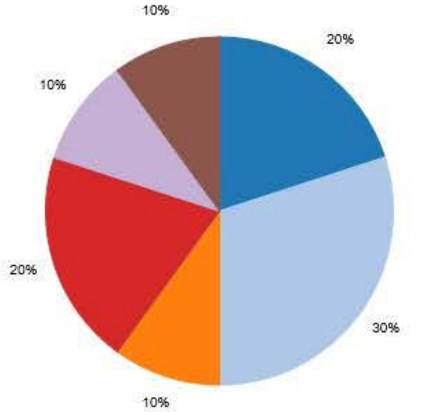
3. What data domains is OrBev most focused on in Big Data initiatives? (check all that apply)

Customer/Prospect Data	3	21.43%
Customer Transactions	1	7.14%
Channel Data	2	14.29%
Market and Competitive Data	3	21.43%
Product Data	1	7.14%
Service Data	1	7.14%
Supply Chain Data	3	21.43%
Fraud Detection	0	0%
Total answers:	14	



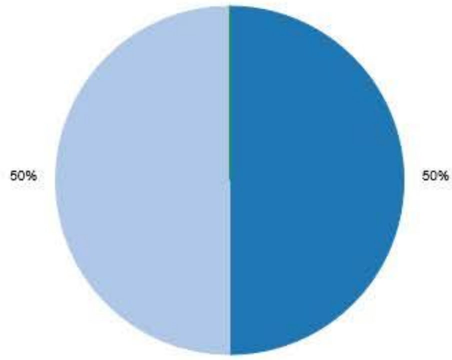
4. What do you see as the main obstacles for the adoption of big data at OrBev? (select up to three)

Low data quality	2	20%
Difficulty finding and retaining qualified staff to manage/ analyze data	3	30%
Inadequate analysis of data for business insights	1	10%
Poor organization of structured data	0	0%
Threats to data security	0	0%
Limited storage capacity	0	0%
Compliance/Regulatory issues	2	20%
Cost of investment	0	0%
Legacy systems	0	0%
Complexity of technology	1	10%
Limitations of current data integration capabilities	1	10%
Total answers:	10	



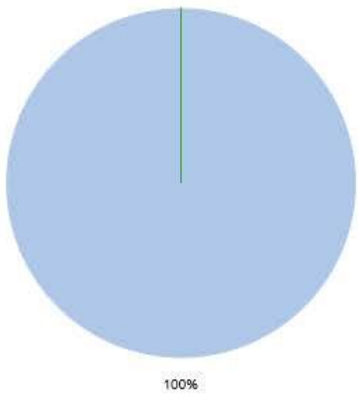
5 * How relevant are the following big data-related challenges for OrBev? - Timeliness

Very Important	2	50%
Important	2	50%
Moderately Important	0	0%
Of little importance	0	0%
Unimportant (not a challenge)	0	0%



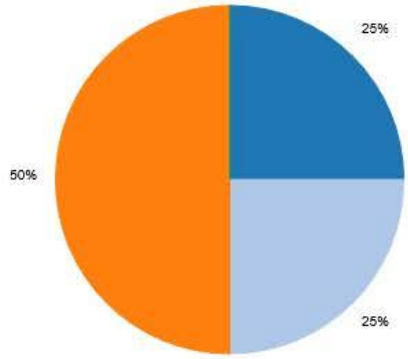
5 * How relevant are the following big data-related challenges for OrBev? - Overwhelming volume

Very Important	0	0%
Important	4	100%
Moderately Important	0	0%
Of little importance	0	0%
Unimportant (not a challenge)	0	0%



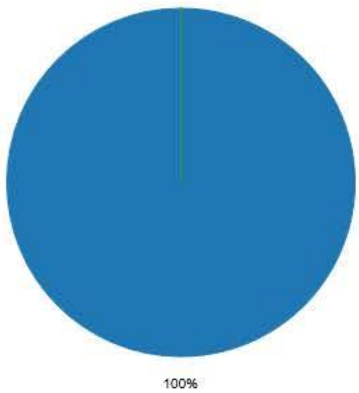
5 * How relevant are the following big data-related challenges for OrBev? - Managing unstructured data

Very Important	1	25%
Important	1	25%
Moderately Important	2	50%
Of little importance	0	0%
Unimportant (not a challenge)	0	0%



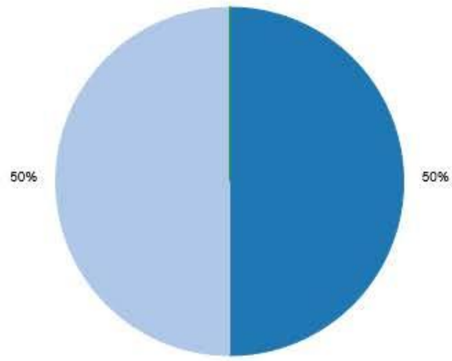
5 * How relevant are the following big data-related challenges for OrBev? - Data quality

Very Important	4	100%
Important	0	0%
Moderately Important	0	0%
Of little importance	0	0%
Unimportant (not a challenge)	0	0%



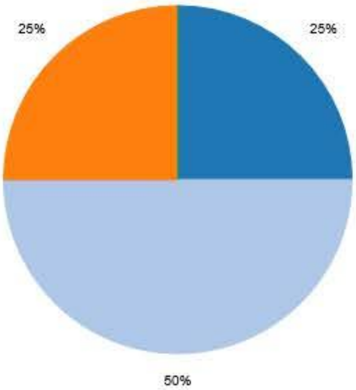
5 * How relevant are the following big data-related challenges for OrBev? - Availability of data

Very Important	2	50%
Important	2	50%
Moderately Important	0	0%
Of little importance	0	0%
Unimportant (not a challenge)	0	0%



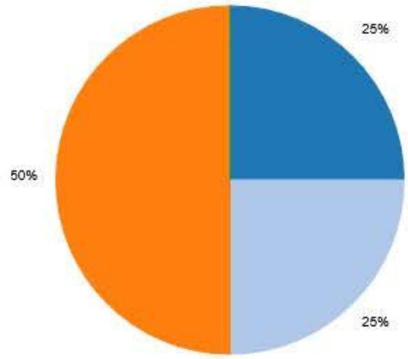
5 * How relevant are the following big data-related challenges for OrBev? - Access rights to data

Very Important	1	25%
Important	2	50%
Moderately Important	1	25%
Of little importance	0	0%
Unimportant (not a challenge)	0	0%



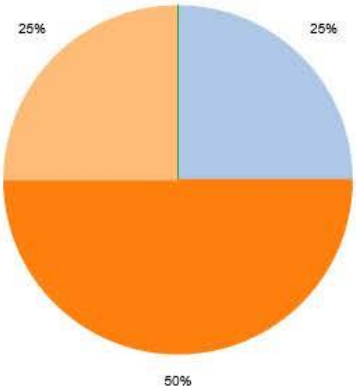
5 * How relevant are the following big data-related challenges for OrBev? - Data ownership issues

Very Important	1	25%
Important	1	25%
Moderately Important	2	50%
Of little importance	0	0%
Unimportant (not a challenge)	0	0%



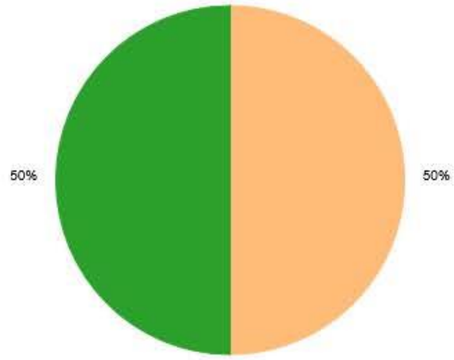
5 * How relevant are the following big data-related challenges for OrBev? - Cost of data

Very Important	0	0%
Important	1	25%
Moderately Important	2	50%
Of little importance	1	25%
Unimportant (not a challenge)	0	0%



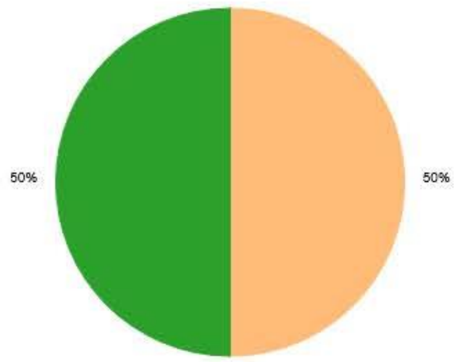
5 * How relevant are the following big data-related challenges for OrBev? - Lack of facilities, infrastructure

Very Important	0	0%
Important	0	0%
Moderately Important	0	0%
Of little importance	2	50%
Unimportant (not a challenge)	2	50%



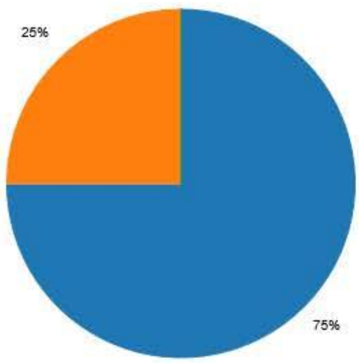
5 * How relevant are the following big data-related challenges for OrBev? - Lack of technology

Very Important	0	0%
Important	0	0%
Moderately Important	0	0%
Of little importance	2	50%
Unimportant (not a challenge)	2	50%



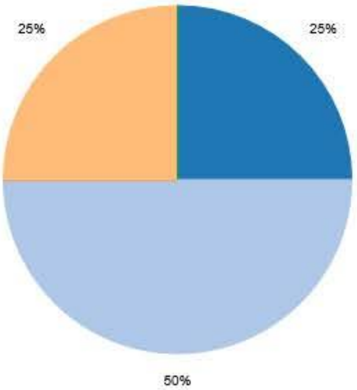
5 * How relevant are the following big data-related challenges for OrBev? - Shortage of talent/skills

Very Important	3	75%
Important	0	0%
Moderately Important	1	25%
Of little importance	0	0%
Unimportant (not a challenge)	0	0%



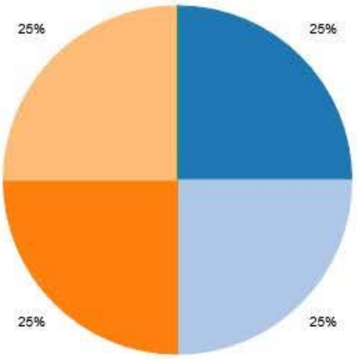
5 * How relevant are the following big data-related challenges for OrBev? - Privacy concerns and regulatory risks

Very Important	1	25%
Important	2	50%
Moderately Important	0	0%
Of little importance	1	25%
Unimportant (not a challenge)	0	0%



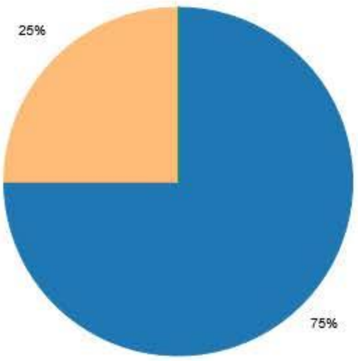
5 * How relevant are the following big data-related challenges for OrBev? - Security

Very Important	1	25%
Important	1	25%
Moderately Important	1	25%
Of little importance	1	25%
Unimportant (not a challenge)	0	0%



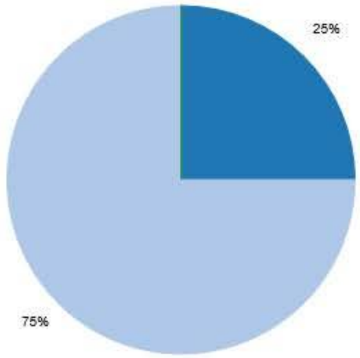
5 * How relevant are the following big data-related challenges for OrBev? - Corporate culture

Very Important	3	75%
Important	0	0%
Moderately Important	0	0%
Of little importance	1	25%
Unimportant (not a challenge)	0	0%



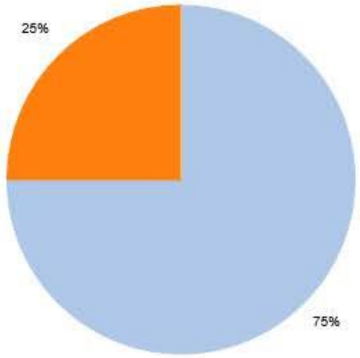
6. To what extent do you agree with the following statement: "The issue for us is now not the growing volumes of data, but rather being able to analyze and act on data in real-time."

Strongly Agree	1	25%
Agree	3	75%
Disagree	0	0%
Strongly Disagree	0	0%
Don't know/Not applicable	0	0%
Total answers:	4	



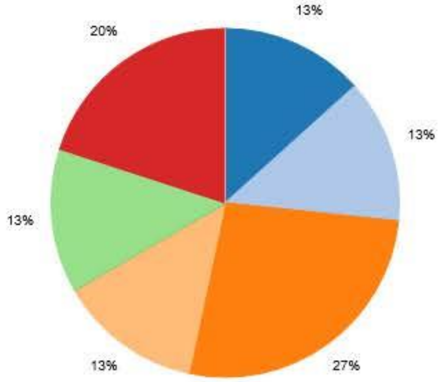
7. Is OrBev currently involved in any big data initiatives?

Well advanced in usage of big data	0	0%
Actively using big data	3	75%
Planning to invest in big data analytics	1	25%
Have no plans to use big data	0	0%
Total answers:	4	



8. What tangible benefits do you hope to achieve through your Big Data initiatives? (select up to three)

Targeted marketing and improved customer experience	2	13.33%
Improved customer insights	2	13.33%
Increased sales	4	26.67%
Higher quality products and services	2	13.33%
New product innovations	0	0%
More efficient operations	2	13.33%
Better, fact-based decision making	3	20%
Reduced risk	0	0%
Total answers:	15	

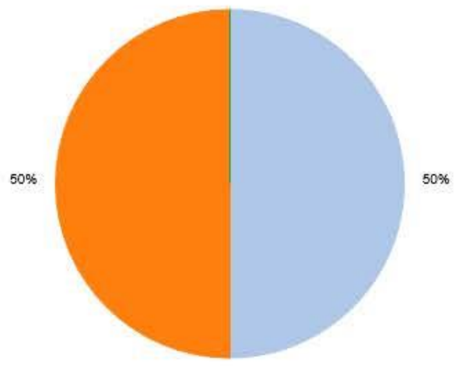


9. What do you think is the biggest opportunity for using Big Data at OrBev?

contribute to top line growth	1	25%
Decide to be AI-driven company and really do it. The main three factors are in place: 1)The infrastructure to execute AI and scale the insights. 2)Enterprise & External data available centrally. 3) Management commitment We need to ensure: - Secure but frictionless access to data! (pilot SAP Data Hub in Azure) - Work around DataScientist shortage, empower AI for CCH analysts. - A partner for AI operationalization and model management (ATOS)	1	25%
To understand how to improve in-store execution to capture shoppers' needs. To improve/actionalize the customer segmentations and introduce Dynamic Routing.	1	25%
Using automated business insights in objective decision making	1	25%
Total answers:	4	

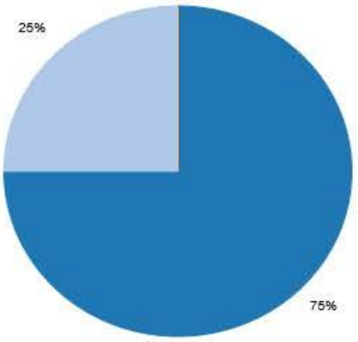
10. How would you rate the access to relevant, accurate and timely data at OrBev today?

minimal	0	0%
less than adequate	2	50%
adequate	2	50%
more than adequate	0	0%
world class	0	0%
Total answers:	4	



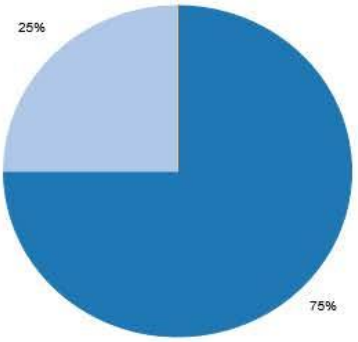
11 * From what sources does OrBev collect, or expects to collect, data? - Machine-to-machine data (e.g. data logged by machines)

Collect now	3	75%
Expects to collect in 5 years	1	25%
No plans to collect	0	0%
Do not know	0	0%



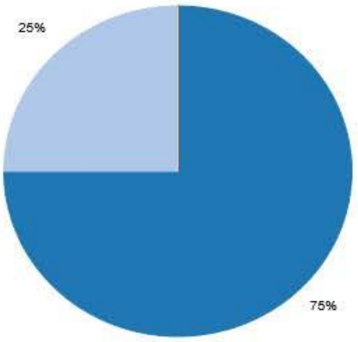
11 * From what sources does OrBev collect, or expects to collect, data? - Transactional data

Collect now	3	75%
Expects to collect in 5 years	1	25%
No plans to collect	0	0%
Do not know	0	0%



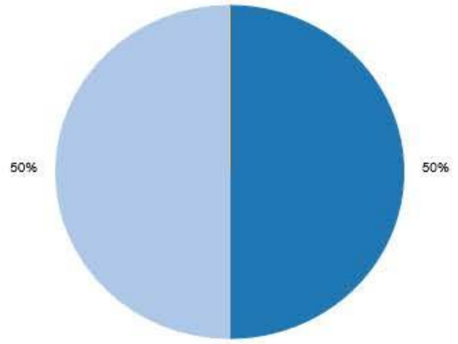
11 * From what sources does OrBev collect, or expects to collect, data? - Office documentation (emails, document stores)

Collect now	3	75%
Expects to collect in 5 years	1	25%
No plans to collect	0	0%
Do not know	0	0%



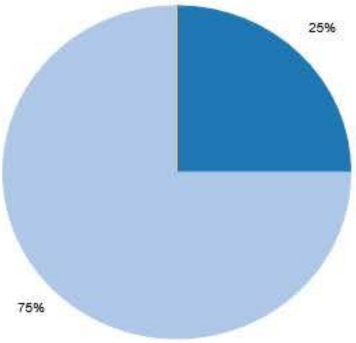
11 * From what sources does OrBev collect, or expects to collect, data? - Data from Social media (e.g. Facebook, twitter feeds)

Collect now	2	50%
Expects to collect in 5 years	2	50%
No plans to collect	0	0%
Do not know	0	0%



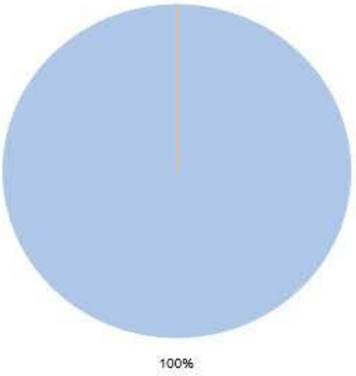
11 * From what sources does OrBev collect, or expects to collect, data? - Open data/Public Sector Information (PSI)

Collect now	1	25%
Expects to collect in 5 years	3	75%
No plans to collect	0	0%
Do not know	0	0%



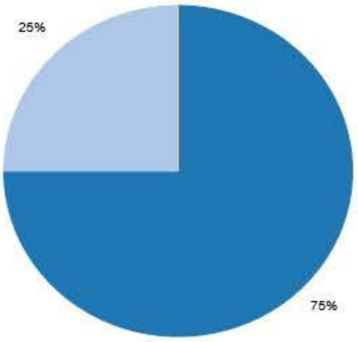
11 * From what sources does OrBev collect, or expects to collect, data? - Telecommunications data (eg phone or data traffic)

Collect now	0	0%
Expects to collect in 5 years	4	100%
No plans to collect	0	0%
Do not know	0	0%



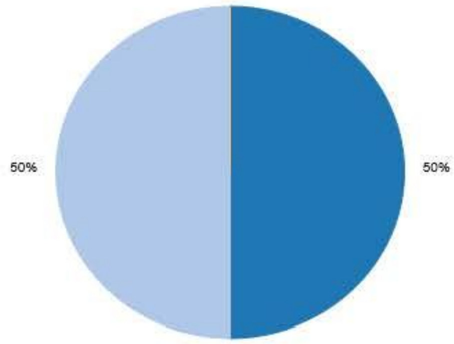
11 * From what sources does OrBev collect, or expects to collect, data? - External feeds

Collect now	3	75%
Expects to collect in 5 years	1	25%
No plans to collect	0	0%
Do not know	0	0%



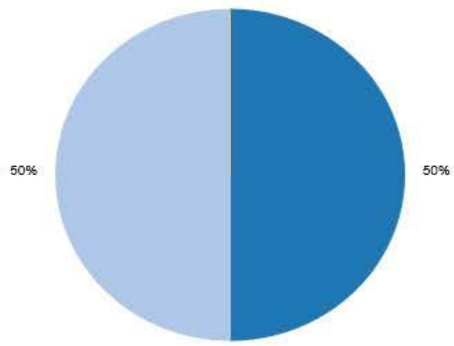
11 * From what sources does OrBev collect, or expects to collect, data? - Data from tracking technology (e.g. RFID scans or POS data, cookies, sensors)

Collect now	2	50%
Expects to collect in 5 years	2	50%
No plans to collect	0	0%
Do not know	0	0%



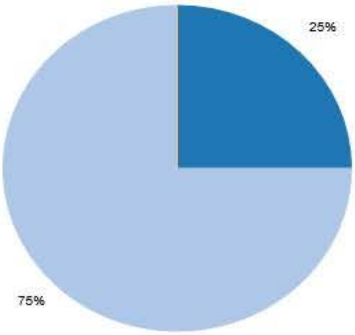
11 * From what sources does OrBev collect, or expects to collect, data? - Data from location services (e.g. GPS, mobile phones, wireless network recognition)

Collect now	2	50%
Expects to collect in 5 years	2	50%
No plans to collect	0	0%
Do not know	0	0%



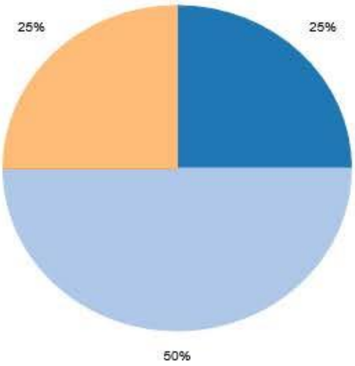
11 * From what sources does OrBev collect, or expects to collect, data? - Other unstructured data (e.g.Free-form text, still images/videos, audio)

Collect now	1	25%
Expects to collect in 5 years	3	75%
No plans to collect	0	0%
Do not know	0	0%



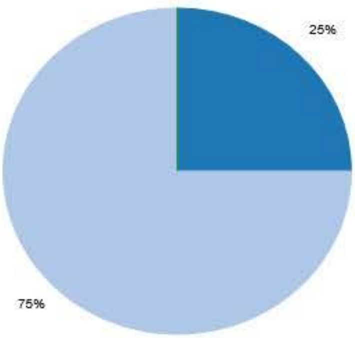
12. What does the quality of your data currently restrict you from doing?

Achieving a holistic view of customers	1	25%
Improving time to market for new projects	2	50%
Improving efficiencies or effectiveness of new marketing initiatives	0	0%
Providing real-time insights for business	1	25%
Total answers:	4	



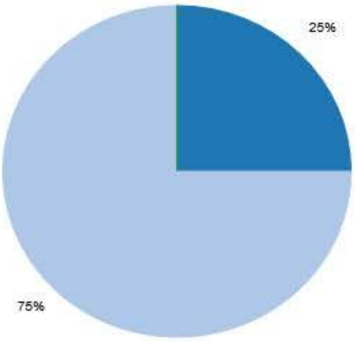
13. From all data collected by OrBev, what is approximately the share of external data (collected from external sources), compared to internal data (produced by your operations)?

<10%	1	25%
10-40%	3	75%
41-60%	0	0%
61-90%	0	0%
>90%	0	0%
Total answers:	4	



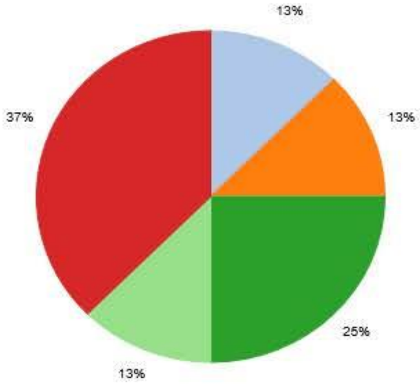
14. From all data collected by OrBev, what is approximately the share of data that the company exploits to improve business processes, products, etc.?

<10%	1	25%
10-40%	3	75%
41-60%	0	0%
61-90%	0	0%
>90%	0	0%
Total answers:	4	



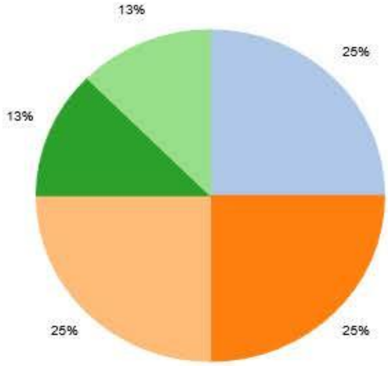
15. What are the primary data issues driving you to consider Big Data? (select all that apply)

Analyzing data sets < 1TB	0	0%
Analyzing data sets 1TB - 100TB	1	12.5%
Analyzing data sets 100TB - 1PB	1	12.5%
Analyzing data sets > 1PB	0	0%
Analyzing new data types (text, relationship, time-series)	2	25%
Analyzing streaming data	1	12.5%
Analyzing data from diverse sources	3	37.5%
Total answers:	8	



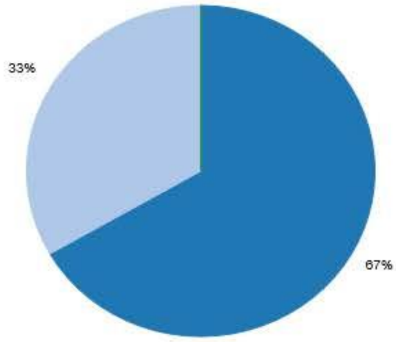
16. What programming languages/tools will you use for development? (check all that apply)

Traditional languages (e.g., Java, C, C++)	0	0%
SQL	2	25%
Scripting Languages (e.g., Python, Perl)	2	25%
Open-source Libraries	2	25%
Product-specific Languages / Libraries	1	12.5%
Proprietary Language / Libraries	1	12.5%
Total answers:	8	



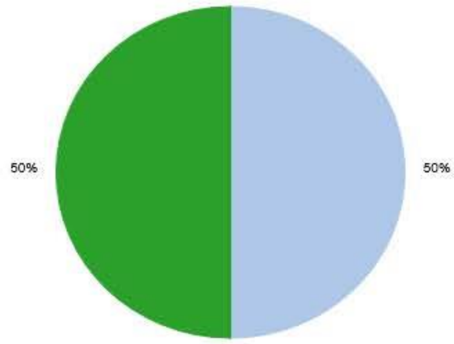
17. Do your Big Data applications stand on their own or are they tightly integrated or embedded with any major systems? (check one)

Enterprise applications (ERP, CRM)	4	66.67%
Business processes (BPM)	2	33.33%
Business rules (BRE)	0	0%
No other system	0	0%
Don´t know	0	0%
Total answers:	6	



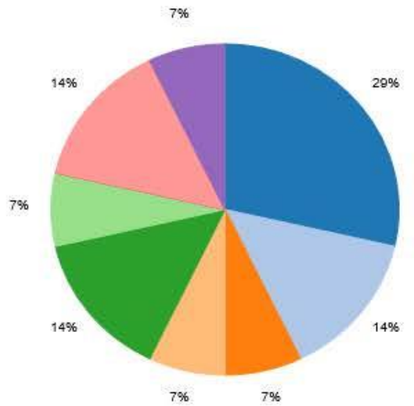
18. What are the key data integration challenges you face? (select all that apply)

Data held in legacy systems	0	0%
Lack of tools in place to integrate data from multiple systems	2	50%
Inability to integrate data effectively from silo sources	0	0%
Data management strategy is outdated	0	0%
No data integration challenges incurred	2	50%
Total answers:	4	



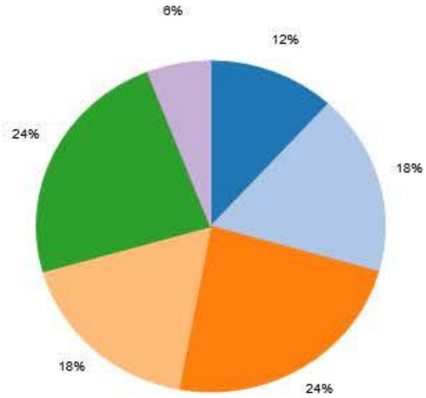
19. What data challenges are you addressing with Big Data? (select all that apply)

Integrating a wider variety of data	4	28.57%
Cleansing data	2	14.29%
Using more current data	1	7.14%
Storing more historical data	1	7.14%
Understanding unstructured data	2	14.29%
Using real-time data	1	7.14%
Using more granular data	0	0%
Using higher quality data	2	14.29%
Understanding streaming data	1	7.14%
Total answers:	14	



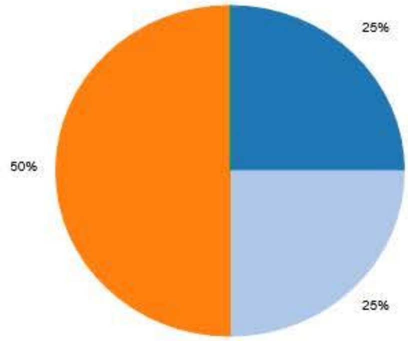
20. Which business functions at OrBev are the most important users of data and data analytics? (check all that apply)

BSS	2	11.76%
Human Resources	3	17.65%
Suppy Chain	4	23.53%
Finance	3	17.65%
Commercial	4	23.53%
Legal	0	0%
PA&C	0	0%
Business Development	0	0%
Executive Management	0	0%
Research & Development	1	5.88%
Total answers:	17	



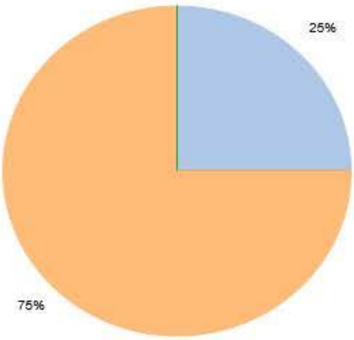
21 * To what extent does OrBev use the following types of big data analysis tools? - Predictive analysis

Always	1	25%
Often (more than half the times)	1	25%
Sometimes (half the times)	2	50%
Seldom (less than half the times)	0	0%
Never	0	0%



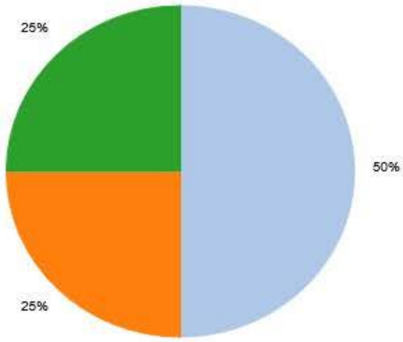
21 * To what extent does OrBev use the following types of big data analysis tools? - Sentiment analysis

Always	0	0%
Often (more than half the times)	1	25%
Sometimes (half the times)	0	0%
Seldom (less than half the times)	3	75%
Never	0	0%



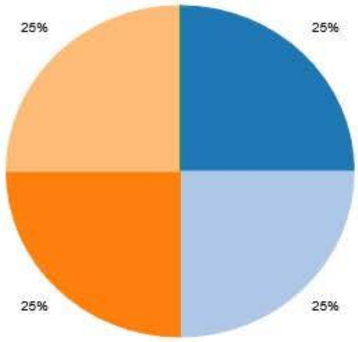
21 * To what extent does OrBev use the following types of big data analysis tools? - Data mining

Always	0	0%
Often (more than half the times)	2	50%
Sometimes (half the times)	1	25%
Seldom (less than half the times)	0	0%
Never	1	25%



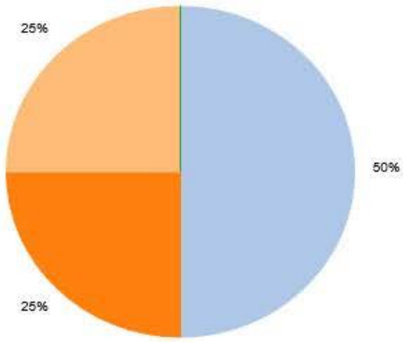
21 * To what extent does OrBev use the following types of big data analysis tools? - Data visualization

Always	1	25%
Often (more than half the times)	1	25%
Sometimes (half the times)	1	25%
Seldom (less than half the times)	1	25%
Never	0	0%



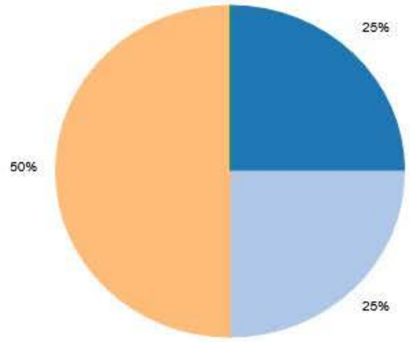
21 * To what extent does OrBev use the following types of big data analysis tools? - Embedded analytics

Always	0	0%
Often (more than half the times)	2	50%
Sometimes (half the times)	1	25%
Seldom (less than half the times)	1	25%
Never	0	0%



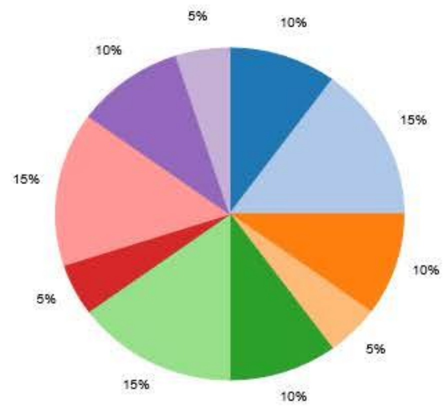
21 * To what extent does OrBev use the following types of big data analysis tools? - Collaborative analytics

Always	1	25%
Often (more than half the times)	1	25%
Sometimes (half the times)	0	0%
Seldom (less than half the times)	2	50%
Never	0	0%



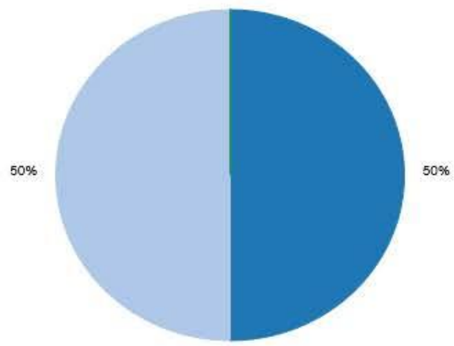
22. How do you believe the results of data analysis will most likely be used? (select up to three)

Location-based customer contact/services	2	10%
Forecasting	3	15%
Skills development	2	10%
Business process innovation	1	5%
Supply chain enhancements	2	10%
Customized promotions	3	15%
Employee performance tracking	1	5%
Proactive maintenance/Fault detection	3	15%
Data-based customer/Service innovation	2	10%
Proactive customer assistance	1	5%
Total answers:	20	



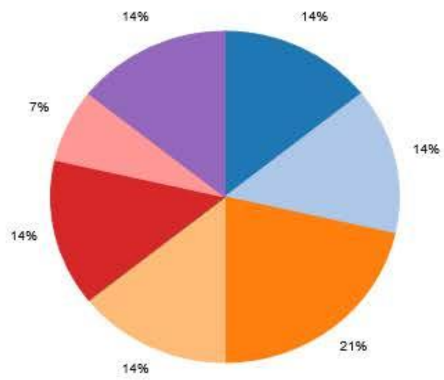
23. How do you plan to measure the success of your Big Data Analytics initiatives?

With quantitative metrics tied to business performance	2	50%
With qualitative metrics tied to business performance	2	50%
With quantitative metrics tied to IT performance	0	0%
With qualitative metrics tied to IT performance	0	0%
No specific measurement methodology in place	0	0%
Total answers:	4	



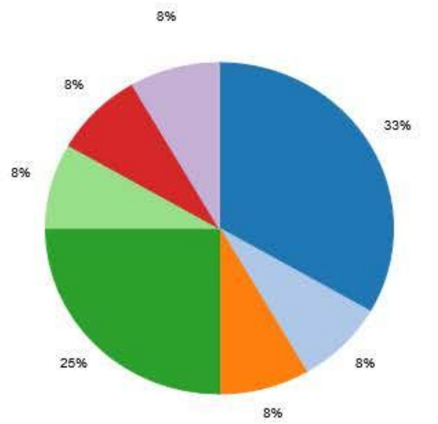
24. What data structures and standards are of particular interest in your Big Data initiatives? (select all that apply)

Flat-File	2	14.29%
Relational	2	14.29%
Unstructured text	3	21.43%
Time-series	2	14.29%
Graphs	0	0%
Semantic Web	0	0%
XML	2	14.29%
Multimedia	1	7.14%
Proprietary	2	14.29%
Total answers:	14	



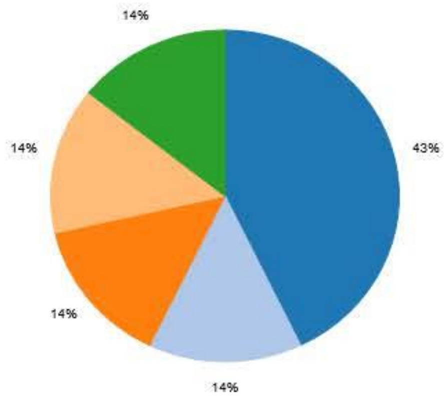
25. What data challenges are you addressing with Data Analytics? (select all that apply)

Integrating a wider variety of data	4	33.33%
Cleansing data	1	8.33%
Using more current data	1	8.33%
Storing more historical data	0	0%
Understanding unstructured data	3	25%
Using real-time data	1	8.33%
Using more granular data	1	8.33%
Using higher quality data	0	0%
Understanding streaming data	0	0%
Automated generation of insights, to be used in standard business processes	1	8.33%
Total answers:	12	



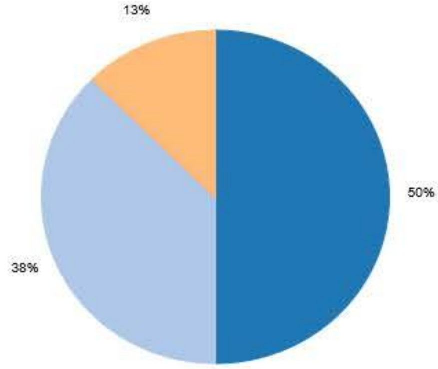
26. What does the quality of your data currently restrict you from doing? (select all that apply)

Achieving a holistic view of customers	3	42.86%
Improving time to market for new projects	1	14.29%
Improving efficiencies or effectiveness of new marketing initiatives	1	14.29%
Providing real-time insights for business	1	14.29%
The lack of Sell-out and for indirect cust. Sell-in data	1	14.29%
Total answers:	7	



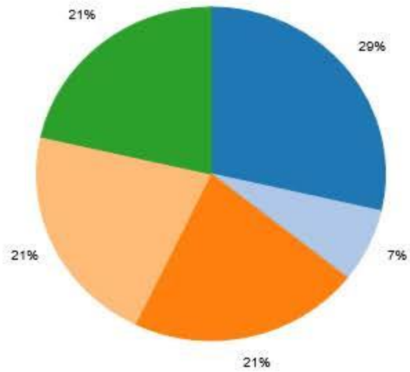
27. What data management approaches are you considering? (check all that apply)

Distributed File (e.g., Hadoop, Grid)	4 50%
Specialized Relational (e.g., Appliances, Columnar, In-Memory);	3 37.5%
Traditional Relational	0 0%
AI automation	1 12.5%
Total answers:	8



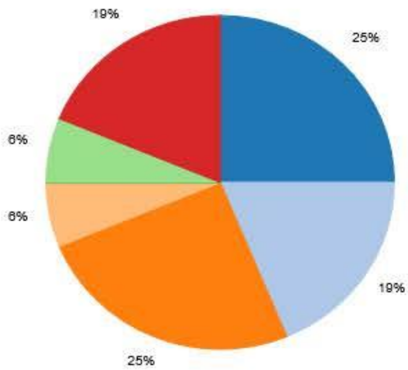
28. For what types of analyses do you want to use Big Data? (check all that apply)

Real time analytics and alerts	4 28.57%
Ability to analyze text	1 7.14%
Ability to analyze relationships	3 21.43%
Ability to analyze very large data sets D Ability to analyze disparate data sets D Ability to analyze external data sets	3 21.43%
Ability to evaluate new analytic algorithms	3 21.43%
Total answers:	14



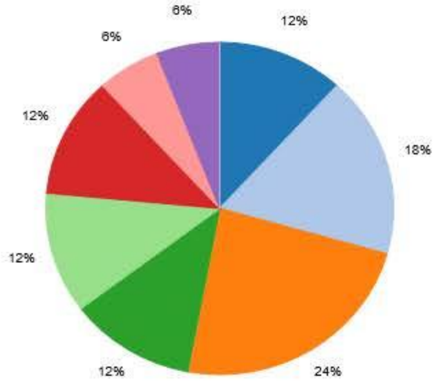
29. In your opinion, what analytic functions/features are most important for OrBev? (check all that apply)

Advanced analytics algorithms	4 25%
Data visualization	3 18.75%
Machine learning	4 25%
Executing existing algorithms faster	1 6.25%
Text analytics	0 0%
Social network analytics	1 6.25%
Executing existing algorithms on much larger data sets	3 18.75%
Total answers:	16



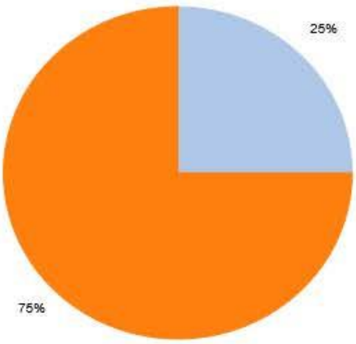
30. What factors have hindered OrBev to increase its use of big data so far? (select all that apply)

Financial (high costs are prohibiting)	2	11.76%
Availability (not enough data available)	3	17.65%
Talent (skilled worker shortage	4	23.53%
Logistical (too many data sources)	0	0%
Organizational (not everyone would benefit)	2	11.76%
Cultural (Lack of internal understanding of data's value)	2	11.76%
Security (too many variable, poor governance)	2	11.76%
Technical (data too complex to integrate and analyze)	1	5.88%
Time (data capabilities take too long	1	5.88%
Nothing has hindered our efforts to increase the use of big data	0	0%
Total answers:	17	



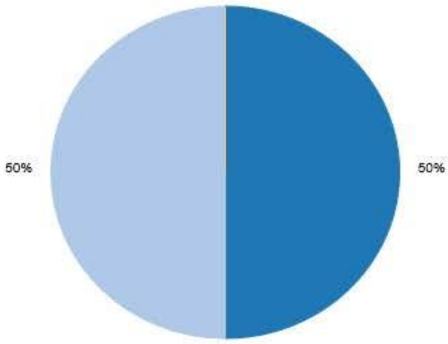
31. Which of the following best describes decision-making at OrBev?

Rarely data-driven	0	0%
Somewhat data-driven	1	25%
Highly data-driven	3	75%
Total answers:	4	



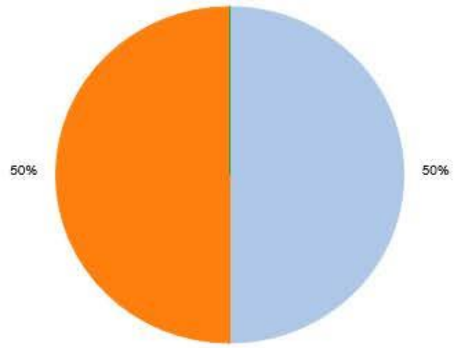
32. How far would you agree with the following statement? "Data mining techniques can support the business process of decision making in the beverage industry."

Strongly Agree	2	50%
Agree	2	50%
Disagree	0	0%
Strongly Disagree	0	0%
Total answers:	4	



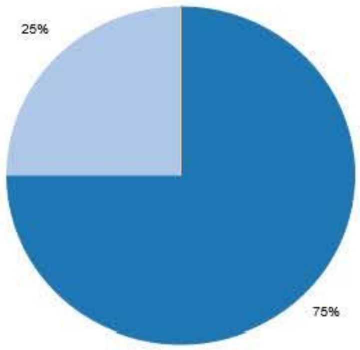
33. How would you rate the speed of decision-making within OrBev when using big data analytics as a key resource?

Very quick	0	0%
Somewhat quick	2	50%
Moderate	2	50%
Somewhat slow	0	0%
Very slow	0	0%
Total answers:	4	



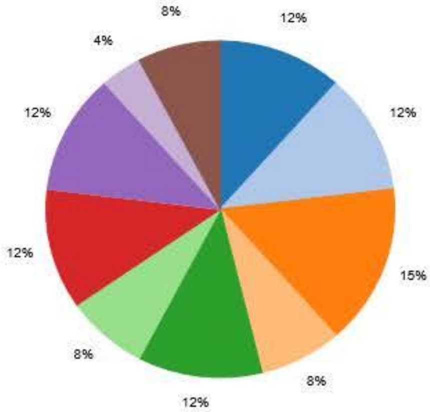
34. How are you thinking about Big Data capabilities with respect to Advanced Analytics (data mining, predictive modeling, etc.) initiatives? (check one)

Big Data is a part of the Advanced Analytics toolbox	3	75%
Big Data and Advanced Analytics are separate things	1	25%
Big Data is unproven while Advanced Analytics is a part of core business operations	0	0%
Don't know	0	0%
Total answers:	4	



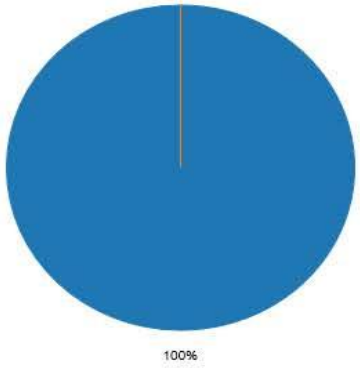
35. How could data analytics be used to make decisions at OrBev? (select all that apply)

Marketing strategy development	3	11.54%
Inform strategic decision-making	3	11.54%
Improve day-to-day business operations	4	15.38%
Identify operational improvements	2	7.69%
Customer segmentation	3	11.54%
Generate new revenue streams	2	7.69%
Make real-time business decisions	3	11.54%
Mine insights in real time	0	0%
Offer customers products and services that the data suggest that they will want to buy	3	11.54%
Integrate social media channels	1	3.85%
Derive analytics via machine-to-machine data	2	7.69%
Total answers:	26	



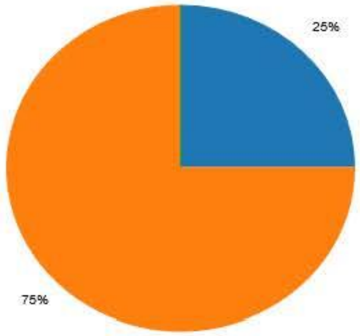
36. For the purpose of increasing effectiveness, should OrBev increase or decrease its reliance on data in the decision-making process?

Increase	4	100%
Decrease	0	0%
Remain the same	0	0%
Total answers:	4	



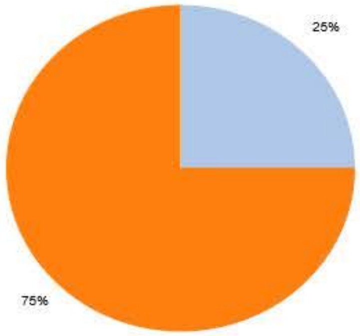
37. To what extent do you agree with the following statement: "Most operational/tactical decisions that can be automated, have been automated.

Strongly Agree	1	25%
Agree	0	0%
Disagree	3	75%
Strongly Disagree	0	0%
Don't know	0	0%
Total answers:	4	



38. How would you rate XXX's ability to use available data to drive executive decisions today?

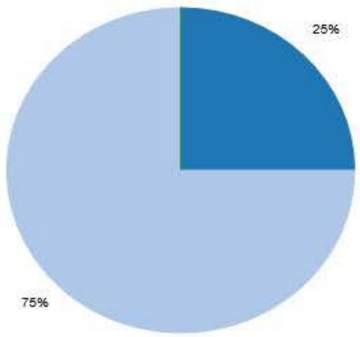
Primitive: most analysis is manual or spreadsheet-oriented	0	0%
Basic: just getting started with analytics, data processing tools or systems typically used to build value models	1	25%
Advanced: provides access to large, complex datasets for analytics based on statistical techniques; likely to include predictive models	3	75%
Outstanding: performs real-time analysis on large, complex datasets and has the ability to act on the outcome in the near real time	0	0%
Total answers:	4	



Comments:
using algorithms from XTel, McK, TD, Atos/eBest, Accenture
Far beyond Basic, but not yet Advanced.
only in Customer Segmentation and Demand Forecasting

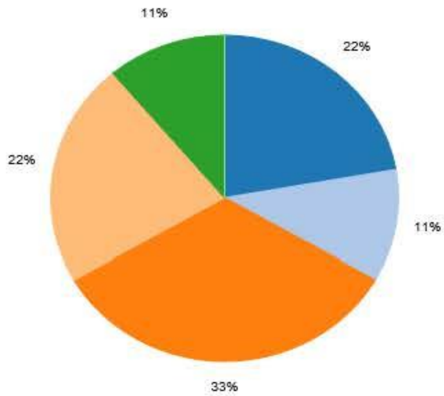
39. How would you rate XXX's success compared with that of your industry peers?

Well above average	1	25%
Somewhat above average	3	75%
Average/ On pair with peers	0	0%
Somewhat below average	0	0%
Well below average	0	0%
Total answers:	4	



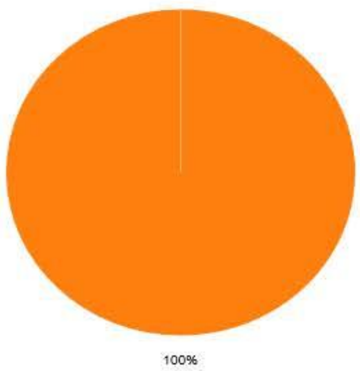
40. What are the biggest obstacles to successful data-based decision-making at OrBev? (select up to three)

Inconsistent reporting of information among business units, geographies or functional operations	2	22.22%
Inadequate tools for gathering, integrating or analyzing operational information	1	11.11%
Lack of accurate, timely or relevant data from across the business	3	33.33%
Inadequate quantitative expertise among executives and support staff	2	22.22%
Overwhelmed by amount and speed of data that are reaching the organization	1	11.11%
Insufficient support from top executives for data analysis as a key component of corporate strategy	0	0%
Total answers:	9	



41. Does OrBev have a governance framework in place that includes consistent guidance, procedures and processes for data capture and management?

Strongly disagree	0	0%
Disagree	0	0%
Agree	4	100%
Strongly Agree	0	0%
Total answers:	4	



42. What are OrBev´s three biggest impediments to using data analytics for effective decision-making? (select up to three)

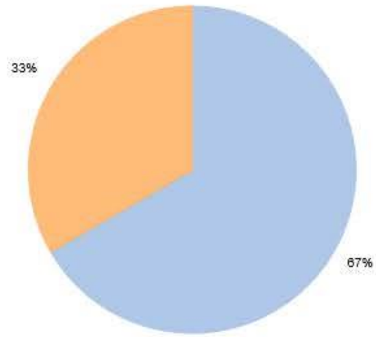
Annex II – Survey Instrument B

8 people have viewed your survey
3 people have answered your survey

Data Analytics Executive Survey

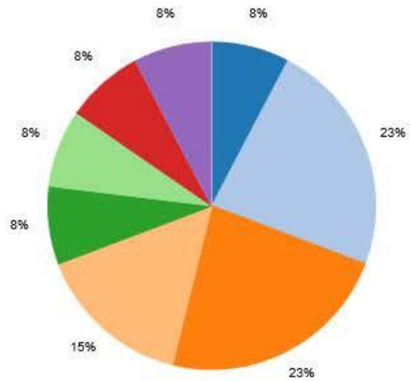
1. How far do you agree with the following statement? "The beverage industry has a clear understanding of the benefits of big data."

Strongly Agree	0	0%
Agree	2	66.67%
Disagree	0	0%
Strongly Disagree	1	33.33%
Total answers:	3	



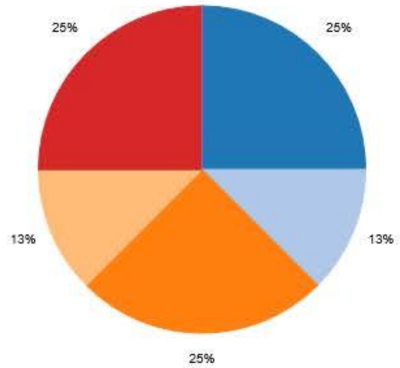
2. At OrBev, which functional groups rely most heavily on big data? (select up to three)

Finance	1	7.69%
Commercial	3	23.08%
Supply Chain	3	23.08%
Human Resources/ Talent Management	2	15.38%
Technology & Innovation	1	7.69%
Research and development	1	7.69%
Strategic planning	1	7.69%
PA&C	0	0%
Customer service	1	7.69%
BSS	0	0%
Total answers:	13	



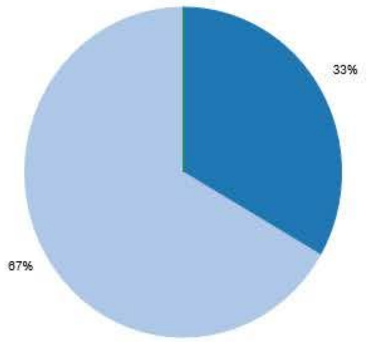
3. What data domains is OrBev most focused on in Big Data initiatives? (check all that apply)

Customer/Prospect Data	2	25%
Customer Transactions	1	12.5%
Channel Data	2	25%
Market and Competitive Data	1	12.5%
Product Data	0	0%
Service Data	0	0%
Supply Chain Data	2	25%
Fraud Detection	0	0%
Industry Specific Data - please specify	0	0%
Total answers:	8	



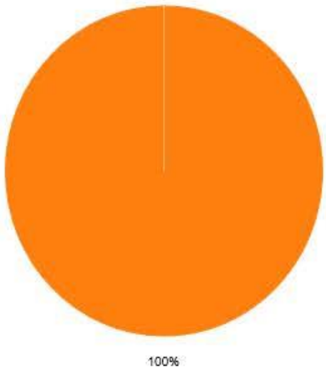
4. To what extent do you agree with the following statement: "The issue for us is now not the growing volumes of data, but rather being able to analyze and act on data in real-time."

Strongly Agree	1	33.33%
Agree	2	66.67%
Disagree	0	0%
Strongly Disagree	0	0%
Don´t Know/ Not Applicable	0	0%
Total answers:	3	



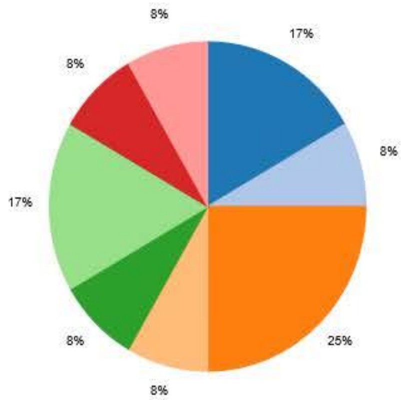
5. Is OrBev currently involved in any big data initiatives?

Well advanced in usage of big data	0	0%
Actively using big data	0	0%
Planning to invest in big data analytics	3	100%
Have no plans to use big data	0	0%
Total answers:	3	



6. What tangible benefits do you hope to achieve through your Big Data initiatives? (select up to three)

Targeted marketing and improved customer experience	2	16.67%
Improved customer insights	1	8.33%
Increased sales	3	25%
Higher quality products and services	1	8.33%
New product innovations	1	8.33%
More efficient operations	2	16.67%
Better, fact-based decision making	1	8.33%
Reduced risk	1	8.33%
Total answers:	12	

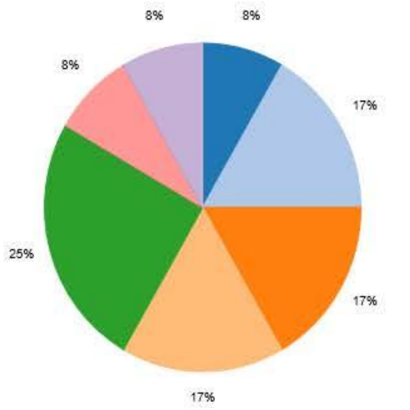


7. What do you think is the biggest opportunity for using Big Data at CCH?

RGM	1	33.33%
Customer segmentation and demand forecasting	1	33.33%
Start! And experiment	1	33.33%
Total answers:	3	

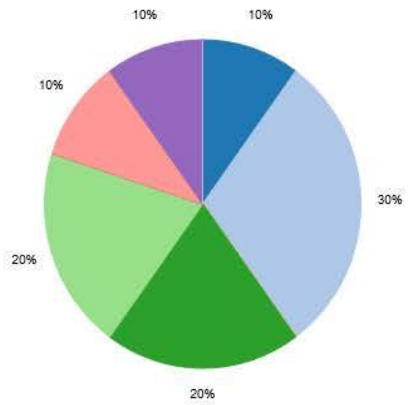
8. Which business functions at OrBev are the most important users of data and data analytics? (check all that apply)

BSS	1	8.33%
Human Resources	2	16.67%
Suppy Chain	2	16.67%
Finance	2	16.67%
Commercial	3	25%
Legal	0	0%
PA&C	0	0%
Business Development	1	8.33%
Executive Management	0	0%
Research & Development	1	8.33%
Total answers:	12	



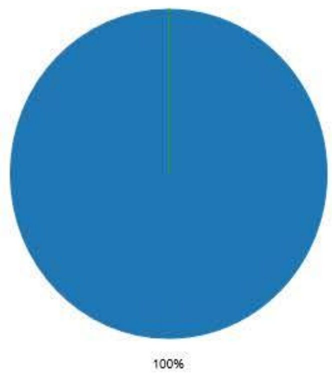
9. How do you believe the results of data analysis will most likely be used? (select up to three)

Location-based customer contact/services	1	10%
Forecasting	3	30%
Skills development	0	0%
Business process innovation	0	0%
Supply chain enhancements	2	20%
Customized promotions	2	20%
Employee performance tracking	0	0%
Proactive maintenance/Fault detection	1	10%
Data-based customer/Service innovation	1	10%
Proactive customer assistance	0	0%
Total answers:	10	



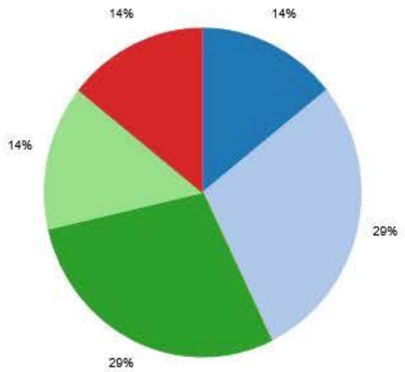
10. How do you plan to measure the success of your Big Data Analytics initiatives?

With quantitative metrics tied to business performance	3	100%
With qualitative metrics tied to business performance	0	0%
With quantitative metrics tied to IT performance	0	0%
With qualitative metrics tied to IT performance	0	0%
No specific measurement methodology in place	0	0%
Total answers:	3	



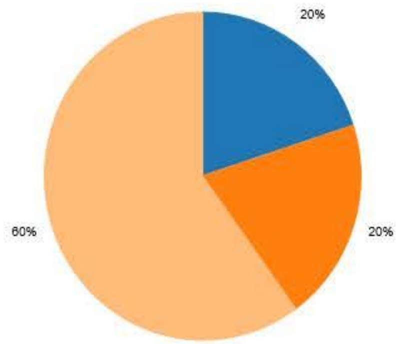
11. 6. What data challenges are you addressing with Data Analytics? (select all that apply)

Integrating a wider variety of data	1	14.29%
Cleansing data	2	28.57%
Using more current data	0	0%
Storing more historical data	0	0%
Understanding unstructured data	2	28.57%
Using real-time data	1	14.29%
Using more granular data	1	14.29%
Using higher quality data	0	0%
Understanding streaming data	0	0%
Total answers:	7	



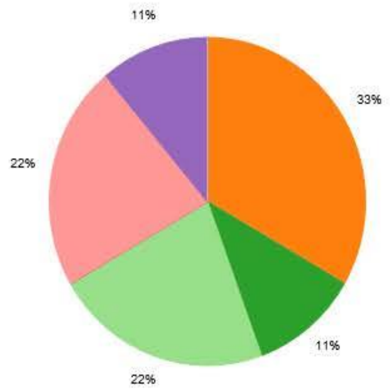
12. What does the quality of your data currently restrict you from doing? (check all that apply)

Achieving a holistic view of customers	1	20%
Improving time to market for new projects	0	0%
Improving efficiencies or effectiveness of new marketing initiatives	1	20%
Providing real-time insights for business	3	60%
Total answers:	5	



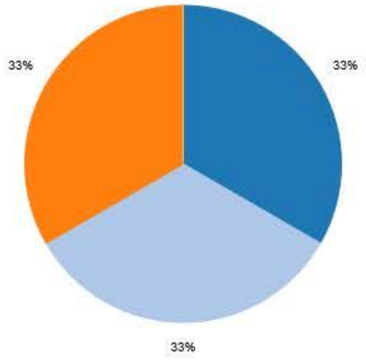
13. What factors have hindered OrBev to increase its use of big data so far? (select all that apply) Financial (high costs are prohibiting) 0 0%

Availability (not enough data available)	0	0%
Talent (skilled worker shortage	3	33.33%
Logistical (too many data sources)	0	0%
Organizational (not everyone would benefit)	1	11.11%
Cultural (Lack of internal understanding of data's value)	2	22.22%
Security (too many variable, poor governance)	0	0%
Technical (data too complex to integrate and analyze)	2	22.22%
Time (data capabilities take too long	1	11.11%
Nothing has hindered our efforts to increase the use of big data	0	0%
Total answers:	9	



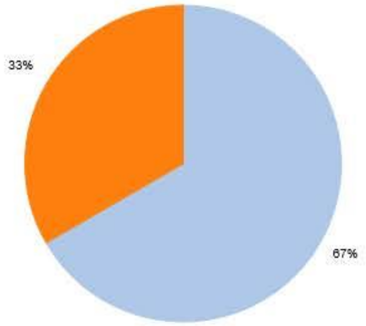
14. How would you rate XXX's ability to use available data to drive executive decisions today?

Primitive: most analysis is manual or spreadsheet-oriented	1	33.33%
Basic: just getting started with analytics, data processing tools or systems typically used to build value models	1	33.33%
Advanced: provides access to large, complex datasets for analytics based on statistical techniques; likely to include predictive models	1	33.33%
Outstanding: performs real-time analysis on large, complex datasets and has the ability to act on the outcome in the near real time	0	0%
Total answers:	3	



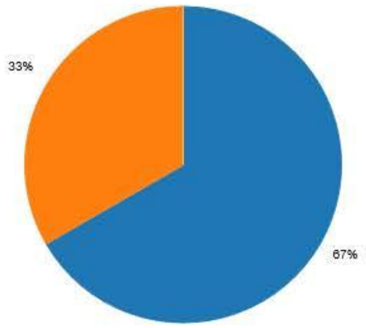
15. Which of the following best describes decision-making at OrBev?

Rarely data-driven	0	0%
Somewhat data-driven	2	66.67%
Highly data-driven	1	33.33%
Total answers:	3	



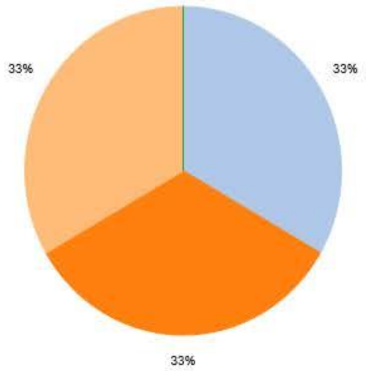
16. How far would you agree with the following statement? "Data mining techniques can support the business process of decision making in the beverage industry."

Strongly Agree	2	66.67%
Agree	0	0%
Disagree	1	33.33%
Strongly Disagree	0	0%
Total answers:	3	



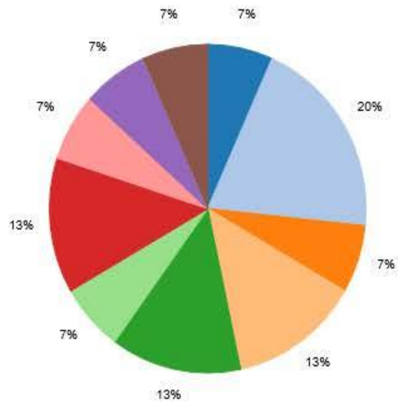
17. How would you rate the speed of decision-making within your Function/BU when using big data analytics as a key resource?

Very quick	0	0%
Somewhat quick	1	33.33%
Moderate	1	33.33%
Somewhat slow	1	33.33%
Very slow	0	0%
Total answers:	3	



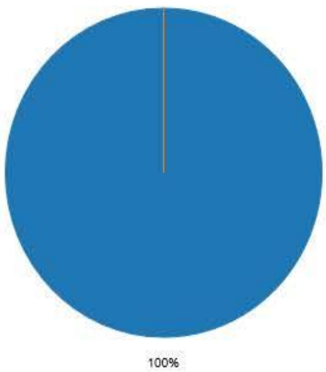
18. How could data analytics be used to make decisions at OrBev? (select all that apply)

Marketing strategy development	1	6.67%
Inform strategic decision-making	3	20%
Improve day-to-day business operations	1	6.67%
Identify operational improvements	2	13.33%
Customer segmentation	2	13.33%
Generate new revenue streams	1	6.67%
Make real-time business decisions	2	13.33%
Mine insights in real time	1	6.67%
Offer customers products and services that the data suggest that they will want to buy	1	6.67%
Integrate social media channels	0	0%
Derive analytics via machine-to-machine data	1	6.67%
Total answers:	15	



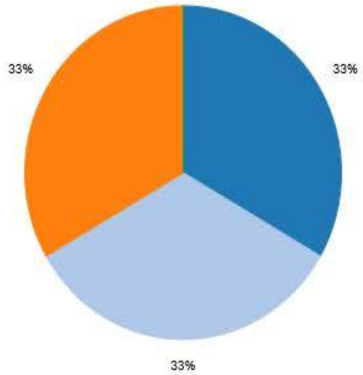
19. For the purpose of increasing effectiveness, should OrBev increase or decrease its reliance on data in the decision-making process?

Increase	3	100%
Decrease	0	0%
Remain the same	0	0%
Total answers:	3	



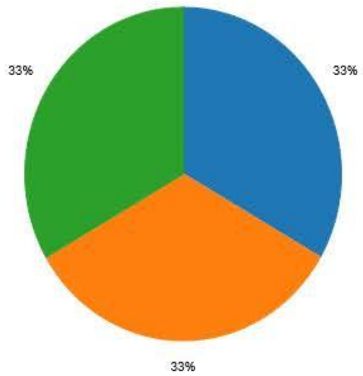
20. To what extent do you agree with the following statement: "Most operational/tactical decisions that can be automated, have been automated.

Strongly Agree	1	33.33%
Agree	1	33.33%
Disagree	1	33.33%
Strongly Disagree	0	0%
Don't know	0	0%
Total answers:	3	



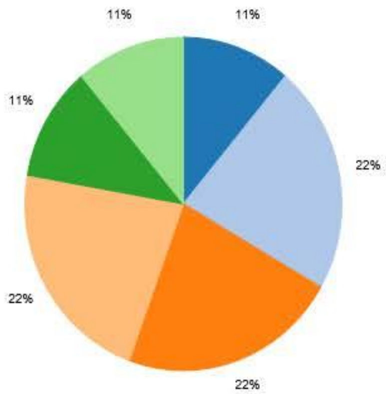
21. How would you rate XXX's success compared with that of your industry peers?

Well above average	1	33.33%
Somewhat above average	0	0%
Average/ On pair with peers	1	33.33%
Somewhat below average	0	0%
Well below average	1	33.33%
Total answers:	3	



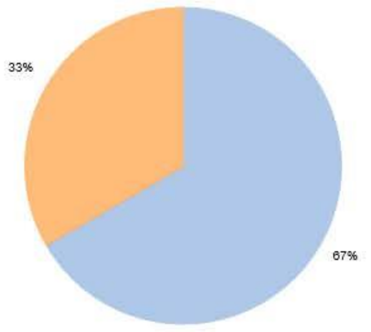
22. What are the biggest obstacles to successful data-based decision-making in your Function/BU? (check all that apply)

Inconsistent reporting of information among business units, geographies or functional operations	1	11.11%
Inadequate tools for gathering, integrating or analyzing operational information	2	22.22%
Lack of accurate, timely or relevant data from across the business	2	22.22%
Inadequate quantitative expertise among executives and support staff	2	22.22%
Overwhelmed by amount and speed of data that are reaching the organization	1	11.11%
Insufficient support from top executives for data analysis as a key component of corporate strategy	1	11.11%
Total answers:	9	



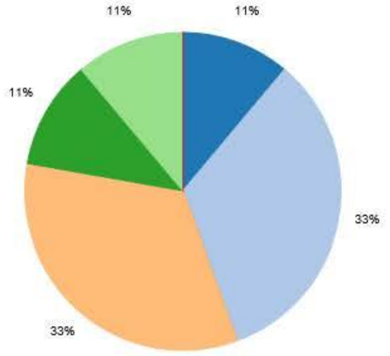
23. Does your Function/BU have a governance framework in place that includes consistent guidance, procedures and processes for data capture and management?

Strongly disagree	0	0%
Disagree	2	66.67%
Agreed	0	0%
Strongly Agreed	1	33.33%
Total answers:	3	



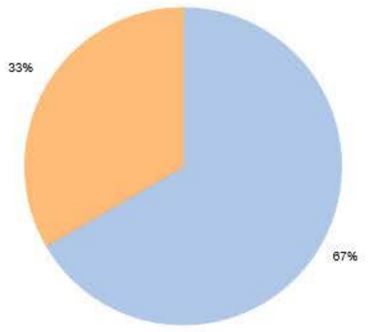
24. What are OrBev´s three biggest impediments to using data analytics for effective decision-making? (select up to three)

Too many "silos"—data is not pooled between functions	1	11.11%
Unstructured content in big data is too difficult to interpret	3	33.33%
Big data sets are too complex to collect and store	0	0%
Shortage of skilled people to analyze the data properly	3	33.33%
Big data is not viewed sufficiently strategically by senior management	1	11.11%
The time taken to analyze large data sets	1	11.11%
The high cost of storing and manipulating large data sets	0	0%
Total answers:	9	



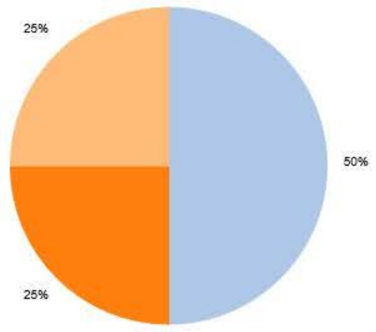
25. Looking specifically at your function, how would you characterize the amount of data available to support decision-making?

Too much	0	0%
Enough	2	66.67%
Not enough	0	0%
Don´t know	1	33.33%
Total answers:	3	



26. What do you see as the two main business challenges for the beverage industry over the next 5 years?

Compliance and regulations	0	0%
Developing new revenue streams	2	50%
Creating a centralized view of customers	1	25%
Customer churn	1	25%
Total answers:	4	



Appendix I

Figure 2.5

From what sources does OrBev collect, or expects to collect, data?	n	Collect now	Expects to collect in 5 years	No plans to collect	Do not know	Total
Machine-to-machine data (e.g. data logged by machines)	4	75%	25%	0%	0%	100%
Transactional data	4	75%	25%	0%	0%	100%
Office documentation (emails, document stores)	4	75%	25%	0%	0%	100%
Data from Social media (e.g. Facebook, twitter feeds)	4	50%	50%	0%	0%	100%
Open data/Public Sector Information (PSI)	4	25%	75%	0%	0%	100%
Telecommunications data (eg phone or data traffic)	4	0%	100%	0%	0%	100%
External feeds	4	75%	25%	0%	0%	100%
Data from tracking technology (e.g. RFID scans or POS data, s	4	50%	50%	0%	0%	100%
Data from location services (e.g. mobile phones, wireless ne	4	50%	50%	0%	0%	100%
Other unstructured data (e.g.Free-form text, still images/videos, audio)	4	25%	75%	0%	0%	100%

Figure 2.8

What data domains is OrBev most focused on in Big Data initiatives? (check all that apply)	BF	IT	Total	%
Customer/Prospect Data	2	3	5	22.7%
Customer Transactions	1	1	2	9.1%
Channel Data	2	2	4	18.2%
Market and Competitive Data	1	3	4	18.2%
Product Data	0	1	1	4.5%
Service Data	0	1	1	4.5%
Supply Chain Data	2	3	5	22.7%
Fraud Detection	0	0	0	0.0%
Industry Specific Data - please specify	0	0	0	0.0%
TOTAL	8	14	22	100.0%

Figure 2.10

How do you believe the results of data analysis will most likely be used? (select up to three)	BF	IT	Total	%
Location-based customer contact/services	1	2	3	10.0%
Forecasting	3	3	6	20.0%
Skills development	0	2	2	6.7%
Business process innovation	0	1	1	3.3%
Supply chain enhancements	2	2	4	13.3%
Customized promotions	2	3	5	16.7%
Employee performance tracking	0	1	1	3.3%
Proactive maintenance/Fault detection	1	3	4	13.3%
Data-based customer/Service innovation	1	2	3	10.0%
Proactive customer assistance	0	1	1	3.3%
TOTAL	10	20	30	100.0%

Figure 2.12

What does the quality of your data currently restrict you from doing? (check all that apply)	BF	IT	Total	%
Achieving a holistic view of customers	1	3	4	36.4%
Improving time to market for new projects	0	1	1	9.1%
improving efficiencies or effectiveness of new marketin	1	1	2	18.2%
Providing real-time insights for business	3	1	4	36.4%
Other: Lack of sell-out and for indirect cust sell-in data	0	1	1	9.1%
TOTAL	5	7	11	100.0%

Figure 2.13

To what extent does OrBev use the following types of big data analysis tools?	**n**	**Always**	**Often (more than half the times)**	**Sometimes (half the times)**	**Seldom (less than half the times)**	**Never**	**Total**
Predictive analysis	4	25%	25%	50%	0%	0%	100%
Sentiment analysis	4	0%	25%	0%	75%	0%	100%
Data mining	4	0%	50%	25%	0%	25%	100%
Data visualization	4	25%	25%	25%	25%	0%	100%
Embedded analytics	4	0%	50%	25%	25%	0%	100%
Collaborative analytics	4	25%	25%	0%	50%	0%	100%

Figure 2.15

Which business functions at OrBev are the most important users of data and data analytics? (check all that apply)	BF	IT	Total	%
IT	1	2	3	10.0%
Human Resources	2	3	5	16.7%
Supply Chain	2	4	6	20.0%
Finance	2	3	5	16.7%
Commercial	3	4	7	23.3%
Legal	0	0	0	0.0%
Customer Service	0	0	0	0.0%
PA&C	1	0	1	3.3%
Business Development	1	0	1	3.3%
Executive Mangement	0	0	0	0.0%
Research & Development	1	1	2	6.7%
TOTAL	13	17	30	100.0%

Figure 2.16

What data challenges are you addressing with Data Analytics? (select all that apply)	BF	IT	Total	%
Integrating a wider variety of data	1	4	5	27.8%
Cleansing data	2	1	3	16.7%
Storing more historical data	0	0	0	0.0%
Understanding unstructured data	2	3	5	27.8%
Using more current data	0	1	1	5.6%
Using real-time data	1	1	2	11.1%
Using more granular data	1	1	2	11.1%
Using higher quality data	0	0	0	0.0%
Understanding streaming data	0	0	0	0.0%
Automated generation of insights, to be used in standard business sprocess (*entered in others*)	0	1	1	5.6%
TOTAL	7	12	18	100.0%

Figure 2.17

Which of the following best describes decision-making at CCH?	BF	IT	Total	%
Rarely data-driven	0	0	0	0.0%
Somewhat data-driven	2	1	3	42.9%
Highly data-driven	1	3	4	57.1%
TOTAL	3	4	7	100.0%

Figure 2.18

How would you rate the speed of decision-making within your Function/BU when using big data analytics as a key resource?	BF	IT	Total	%
Very quick	0	0	0	0.0%
Somewhat quick	1	2	3	42.9%
Moderate	1	2	3	42.9%
Somewhat slow	1	0	1	14.3%
Very slow	0	0	0	0.0%
TOTAL	3	4	7	100.0%

Figure 2.21

How would you rate OrBev’s ability to use available data to drive executive decisions today?	BF	IT	Total	%
Primitive: analysis mostly manual or spreadsheet-oriented	1	0	1	14.3%
Basic: just getting started with analytics, data processing tools or systems	1	1	2	28.6%
Advanced: provides access to large, complex datasets for analytics based on statistical techniques; e.g. predictive models	1	3	4	57.1%
Outstanding: performs real-time analysis on large, complex datasets; ability to act on the outcome in the near real time	0	0	0	0.0%
TOTAL	3	4	7	100.0%

Figure 2.22

How could data analytics be used to make decisions at CCH? (select all that apply)	BF	IT	Total	%
Marketing strategy development	1	3	4	9.8%
Inform strategic decision-making	3	3	6	14.6%
Improve day-to-day business operations	1	4	5	12.2%
Identify operational improvements	2	2	4	9.8%
Customer segmentation	2	3	5	12.2%
Generate new revenue streams	1	2	3	7.3%
Make real-time business decisions	2	3	5	12.2%
Mine insights in real time	1	0	1	2.4%
Offer customers products and services that the data suggest that they will want to buy	1	3	4	9.8%
Integrate social media channels	0	1	1	2.4%
Derive analytics via machine-to-machine data	1	2	3	7.3%
TOTAL	15	26	41	100.0%

Figure 2.24

What are the biggest obstacles to successful data-based decision-making in your Function/BU? (check all that apply)	BF	IT	Total	%
Inconsistent reporting of information among business units, geographies or functional operations	1	2	3	16.7%
Inadequate tools for gathering, integrating or analyzing operational information	2	1	3	16.7%
Lack of accurate, timely or relevant data from across the business	2	3	5	27.8%
Inadequate quantitative expertise among executives and support staff	2	2	4	22.2%
Overwhelmed by amount and speed of data that are reaching the organization	1	1	2	11.1%
Insufficient support from top executives for data analysis as a key component of corporate strategy	1	0	1	5.6%
TOTAL	9	9	18	100.0%

Figure 2.25

To what extent do you agree with the following statement: "Most operational/tactical decisions that can be automated, have been automated.

Answer Choices	BF	IT	Total	%
Strongly Agree (1)	1	1	2	28.6%
Agree (2)	1	0	1	14.3%
Disagree (3)	1	3	4	57.1%
Strongly Disagree (4)	0	0	0	0.0%
Total Answers	3	4	7	100.0%

Descriptive Statistics

Minimum	Maximum	Median	Mean	Standad Deviation
0	3	1.50	2.29	0.88

Figure 2.27

For the purpose of increasing effectiveness, should OrBev increase or decrease its reliance on data in the decision-making process?

	BF	IT	Total	%
Increase	3	4	7	100.0%
Decrease	0	0	0	0.0%
Remain the same	0	0	0	0.0%
TOTAL	3	4	7	100.0%

Figure 2.30

What factors have hindered OrBev to increase its use of big data so far? (select all that apply)

	BF	IT	Total	%
Financial (high costs are prohibiting)	0	2	2	7.7%
Availability (not enough data available)	0	3	3	11.5%
Talent (skilled worker shortage)	3	4	7	26.9%
Logistical (too many data sources)	0	0	0	0.0%
Organizational (not everyone would benefit)	1	2	3	11.5%
Cultural (Lack of internal understanding of data's value)	2	2	4	15.4%
Security (too many variable, poor governance)	0	2	2	7.7%
Technical (data too complex to integrate and analyze)	2	1	3	11.5%
Time (data capabilities take too long)	1	1	2	7.7%
Nothing has hindered our efforts	0	0	0	0.0%
TOTAL	9	17	26	100.0%

Figure 2.31

How would you rate OrBev's success compared with that of your industry peers?

	BF	IT	Total	%
Well above average	1	1	2	20.0%
Well above average	1	1	2	20.0%
Well above average	1	1	2	20.0%
Well above average	1	1	2	20.0%
Well above average	1	1	2	20.0%
TOTAL	5	5	10	100.0%

Figure 2.35

What tangible benefits do you hope to achieve through your Big Data initiatives? (select up to three)

	BF	IT	Total	%
Improved customer experience	2	2	4	14.8%
Improved customer experience	1	2	3	11.1%
Increased sales	3	4	7	25.9%
Higher quality products and services	1	2	3	11.1%
New product innovations	1	0	1	3.7%
More efficient operations	2	2	4	14.8%
Better, fact-based decisionmaking	1	3	4	14.8%
Reduced risk	1	0	1	3.7%
TOTAL	12	15	27	100.0%

Figure 2.36

What do you see as the two main business challenges for the beverage industry over the next 5 years?				
	BF	**IT**	**Total**	**%**
Compliance and regulations	0	na		0.0%
Developing new revenue streams	2	na		50.0%
Creating a centralised view of the customers	1	na		25.0%
Customer Churn	1	na		25.0%
TOTAL	4	0	0	100.0%