

Data Driven KPI (Key Performance Indicators) Insight & Prediction

Case Studies: Deutsche Bahn AG

Cevi Herdian, B. Sc, SFC (Scrum Fundamental Certified)

2019-03-02

Contents

| | | |
|----------|---|------------|
| 1 | Introduction | 5 |
| 1.1 | Master's thesis statement of originality | 5 |
| 1.2 | Acknowledgments | 5 |
| 1.3 | Abstract | 5 |
| 2 | Data is everywhere | 7 |
| 2.1 | Big data definition | 15 |
| 2.2 | T & L industry: Deutsche Bahn | 21 |
| 2.3 | Another industry | 29 |
| 3 | Big Data Analytics | 45 |
| 3.1 | Data Analytics Lifecycle | 45 |
| 3.2 | Modern Data Analytics | 47 |
| 3.3 | Analytics Paradigm | 47 |
| 3.4 | Deutsche Bahn Tools | 48 |
| 4 | Terminology | 57 |
| 4.1 | Balance Sheet | 57 |
| 4.2 | Data Cleaning | 57 |
| 4.3 | Data Model | 58 |
| 4.4 | Data Analysis Expression (DAX) | 58 |
| 5 | KPIs for big data | 59 |
| 5.1 | Financial Perspective | 62 |
| 5.2 | Customer Perspective | 70 |
| 5.3 | Process Perspective | 74 |
| 5.4 | Human Resource and Innovation Perspective | 76 |
| 5.5 | By Industry | 77 |
| 5.6 | By Business Goal | 89 |
| 5.7 | By Department | 89 |
| 6 | Deutsche Bahn Use Case | 99 |
| 6.1 | IT Spend Analysis KPI | 99 |
| 6.2 | Human Resources KPI | 100 |
| 6.3 | Opportunity Analysis KPI | 102 |
| 6.4 | Data Science HR Turnover | 104 |
| 6.5 | Data Science Sales Prediction | 104 |
| 7 | Appendix | 105 |

Chapter 1

Introduction

1.1 Master's thesis statement of originality

I hereby confirm that I have written the accompanying thesis by myself, without contributions from any sources other than those cited and acknowledgements. This applies also to all graphics, drawings, maps and images included in this thesis.

Berlin, 18. March 2019

Cevi Herdian, B. Sc, SFC (Scrum Fundamental Certified)

1.2 Acknowledgments

After fast 8 years of hard work with the study at Berlin University of applied Sciences (HTW Berlin). From Studentkolleg till this Master, I am now ready to face new challenges in the real world in my country Indonesia. Writing this thesis has been interesting experience an also my las work in university life. I have faced some difficulties but none that have stopped me to complete this work. This thesis would not be possible without help from the amazing people at that i meet in University and at works.

This thesis based on my experience as internship and working student in different company from 2015 till February 2019. And I took the use case for this thesis in my last works in Deutsche Bahn from September 2019 till February 2019.

I would like to announce special thanks to Fathimah Dzakiyyah, my beloved wife. And all of my family. I would also like to thank Prof. Dr. Christians and Prof. Dr. Beate Berger who has been inspriring me.

Last but not least I would like to thank my friends, Christ, Mathel, Jabr, Ouafaa, and Hourya. They has been a continuous source of inspiration and always been very helpful.

1.3 Abstract

Background: Big Data can be defined as high Volume and Variety of data that can be brought together and analyzed at high Velocity to discover patterns and make better decisions. Deutsche Bahn as a multinational company have also data growth from the company activities. The KPI (Key Performances Indicator) helps a big data to more understande in easily visualization. A right KPI should act as a gage, helping company understand whether the company is taking the right path toward the strategic goals. I have decided to investigate further of how Deutsche Bahn operates with KPIs from they Big Data.

Purpose: The purpose of this thesis is to examine the reception of Big Data Key Performance Indicators used at Deutsche Bahn Headquarters. Investigate how Big Data KPIs evaluate their way of daily monitoring.

See if there is room for further improvements and also if the findings are applicable in other company (another Deutsche Bahn company).

Method: I decided to do only quantitative-approach when collecting data. The quantitative data were retrieved from SAP Business Object Web Intelligence Database. Because of data privacy at Deutsche Bahn, the data that I used in this thesis isn't data from Deutsche Bahn, but I tried to get similar manner data from another sources. In Use Case part I created the analysis for problem solving from scratch (from the very beginning, especially without making use of or relying on any previous work for assistance).

Conclusions:

Deutsche Bahn uses KPIs in order to Increase sales, profit and to get useful information from their big data that can be analyzed. It is up to each and every sub departments to decide if and how they want to work with the KPIs. Deutsche Bahn Headquarters is successful when operating with big data KPIs. The teams get motivated by working with big data KPI and feel that they can affect the outcome to at least a sufficient extent. There are not many negative things to say about how Deutsche Bahn Headquarters operates with KPIs but there is room for improvements. We believe that other departments might be inspired of how Deutsche Bahn Headquarters operates with KPIs.

Chapter 2

Data is everywhere

Do you have smartphone? Of course, this kind of Ask is not relevant in this era. From our Laptop, Tablet, PC or Handphone. This devices creates more data nowadays. Data is more bigger and bigger. The world contains an astronomical amount of data, an amount that grows larger and larger each day. This Big Data has changed the way the world interacts, uncovered breakthroughs in fast all of our life, from ecommerce, medicine, genetic, financial etc.

The principle of Big Data works that the more you have data about anything or any situation, the more accurately you get new **insights** and make **predictions** about what will happen in the future. Also revealed new ways to understand trends in business and in our daily lives. By comparing more and more data and creating relationships that were previously hidden enable us to learn and make more smarter decisions on targeting business values such as sales, production, or financial situations.

Based on forbes.com only **53%** have big data strategies around the world. The top use case are retail, social analysis, and predictive maintenance.

Key points include the following:

- **Reporting, dashboards, advanced visualization end-user “self-service” and data warehousing are the top five technologies and initiatives strategic to business intelligence (Figure 1.1).**
- **53% of companies are using big data analytics today, up from 17% in 2015 with Telecom and Financial Services industries fueling the fastest adoption (Figure 1.2).**
- **Data warehouse optimization is considered the most important big data analytics use case in 2017, followed by customer/social analysis and predictive maintenance (Figure 1.3).**
- **Big data analytics use cases vary significantly by industry with data warehouse optimization dominating Financial Services, Healthcare, and Customer/social analysis is the leading use case in Technology-based companies (Figure 1.4).**
- **Spark, MapReduce, and Yarn are the three most popular software frameworks today (Figure 1.5).**
- **The big data access methods most preferred by respondents include Spark SQL, Hive, HDFS and Amazon S3 (Figure 1.6).**
- **Machine learning continues to gain more industry support and investment plans with Spark Machine Learning Library (MLib) adoption projected to grow by 60% in the next 12 months (Figure 1.7).**

The big four (Deloitte, Ernst & Young (EY), KPMG and PricewaterhouseCoopers (PwC)) are the four biggest professional services networks in the world. They did also so many research about this theme. One



Figure 2.1:

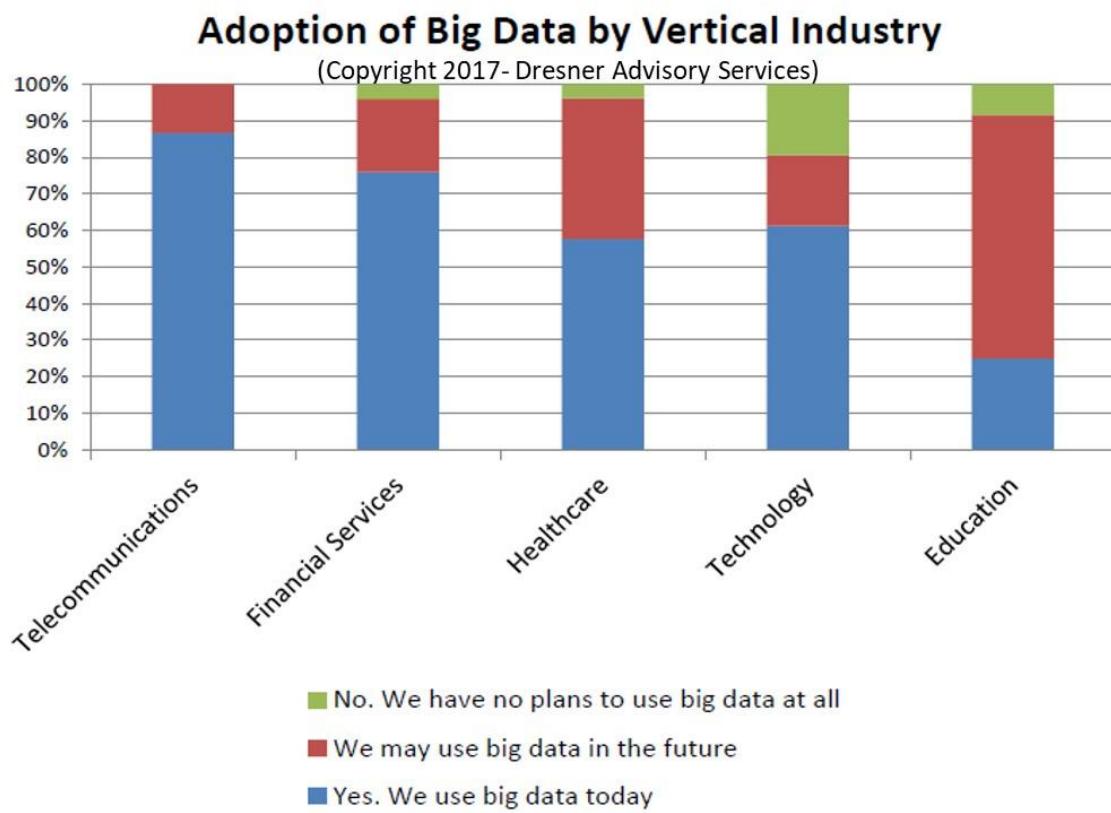
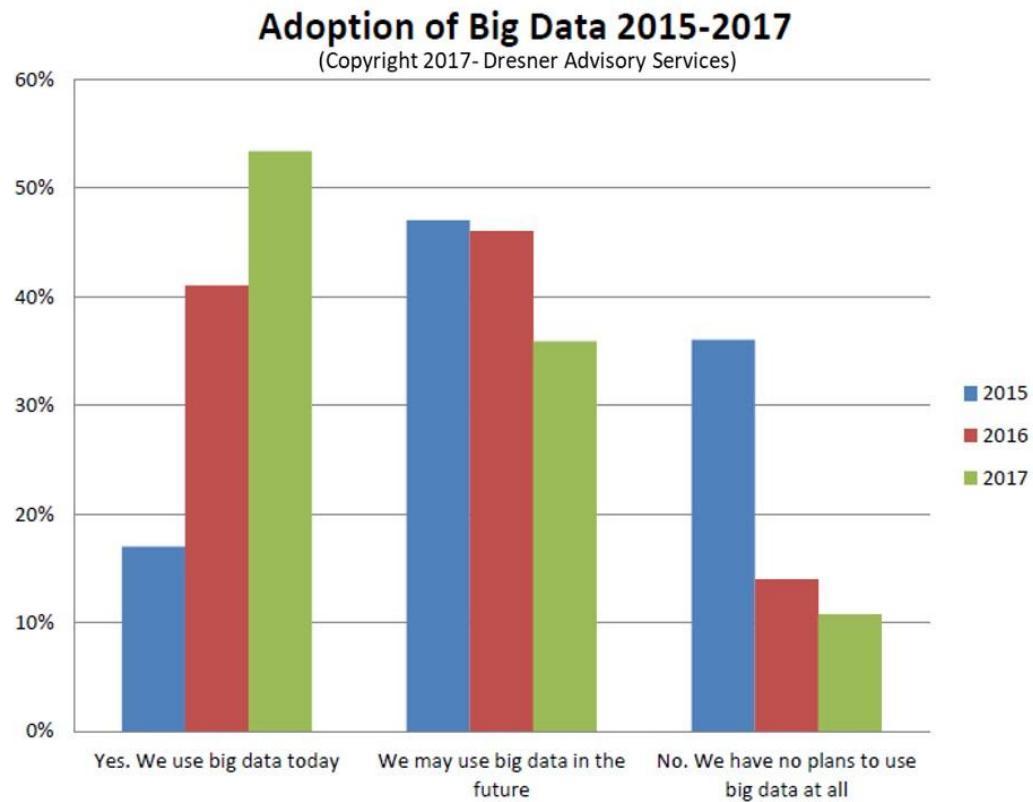


Figure 2.2:

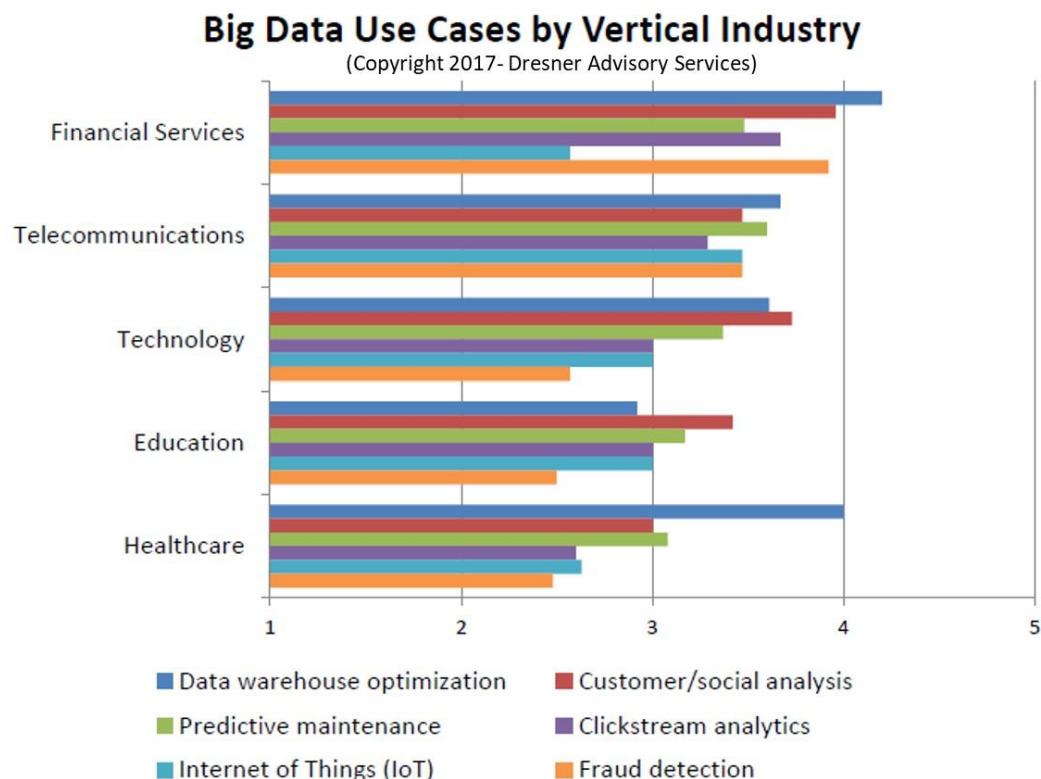


Figure 2.3:

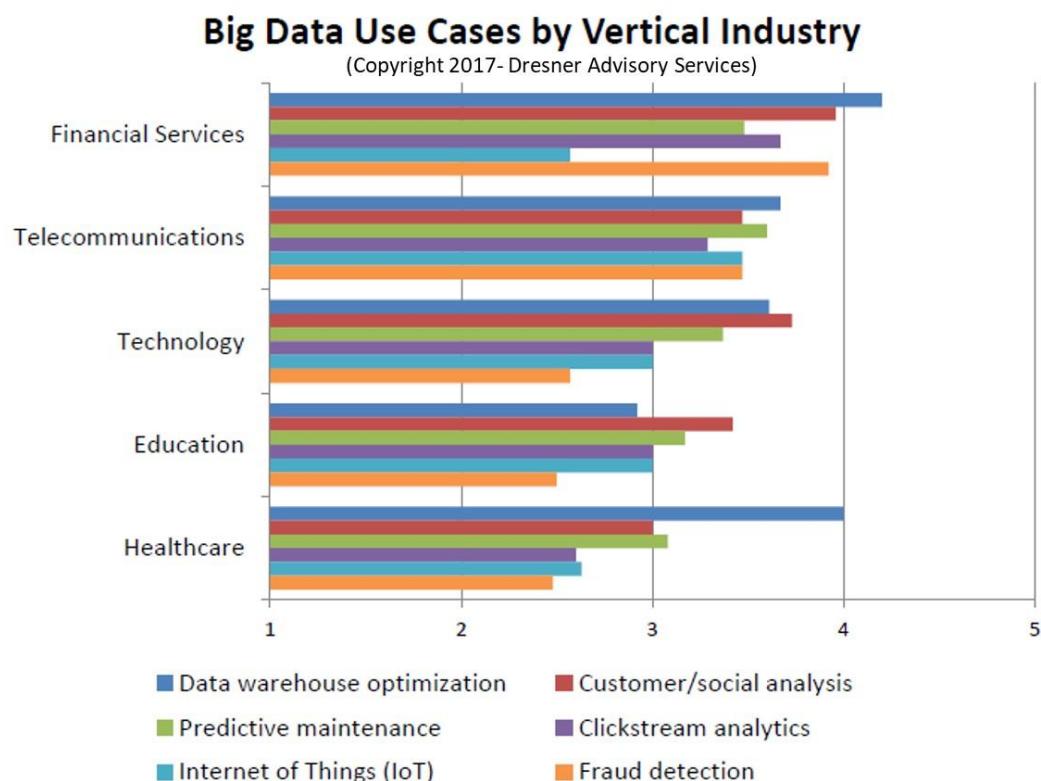


Figure 2.4:

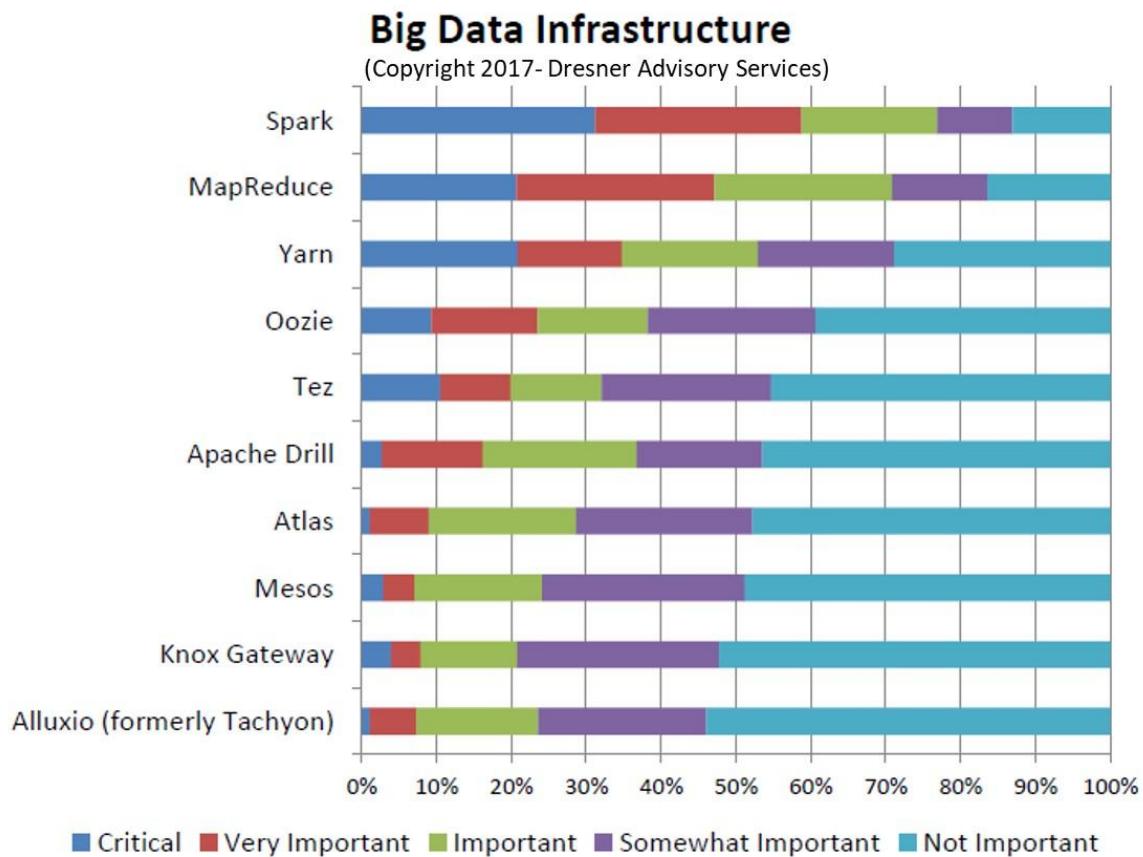


Figure 2.5:

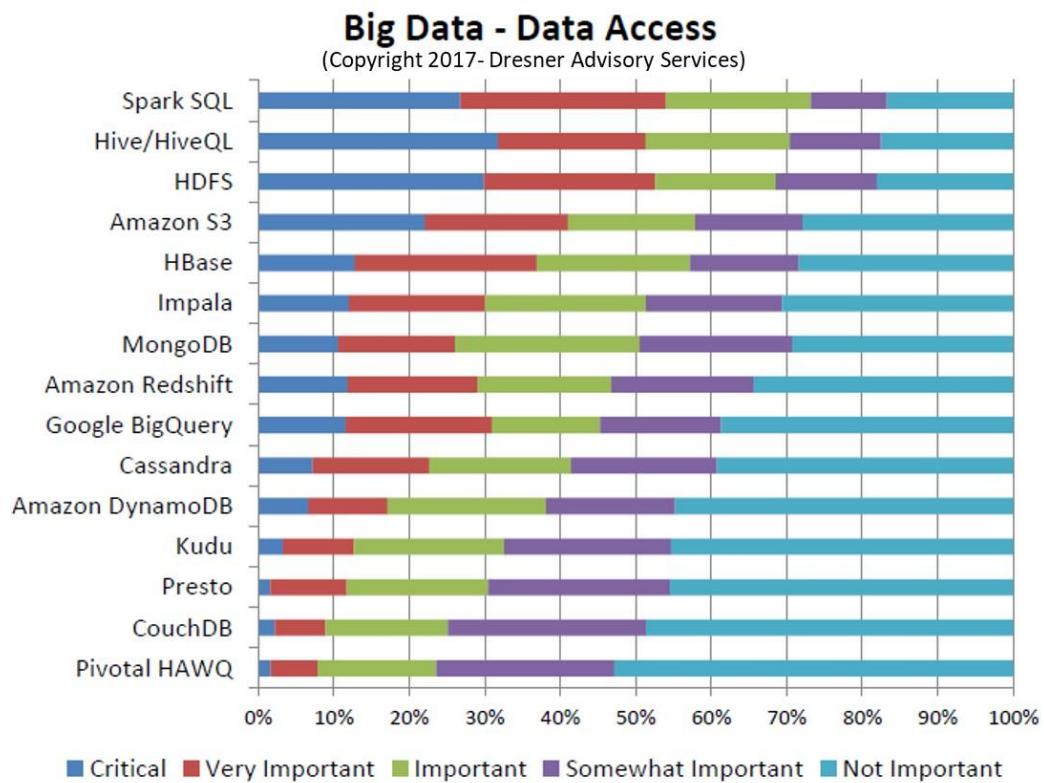


Figure 2.6:

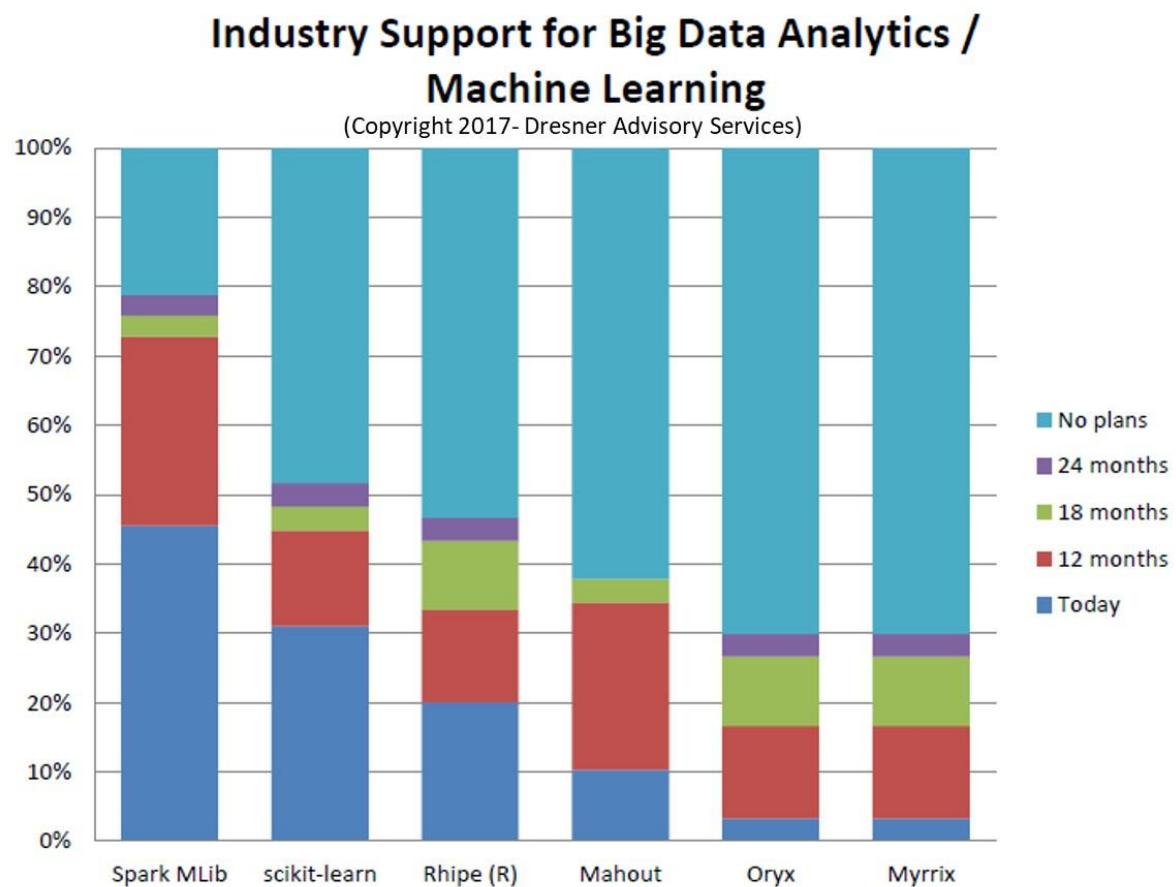


Figure 2.7:

of the interesting information are in the paper “Gut & gigabytes” from PricewaterhouseCoopers (PwC).

The point is that executives rely more on experience and advice than data to make business-defining choices but the data driven cultures could be impact more for decision making.

The summary of this paper:

Highly data-driven companies are three times more likely to report significant improvement in making big decisions, but only 1 in 3 executives say their organisation is highly data-driven.

More big decisions are made opportunistically than deliberately, and big decisions have big impact on future profitability; nearly 1 in 3 executives value those decisions at \$1 billion+

Many executives sceptical or frustrated by the practical application of data and analytics for big decisions, especially in emerging markets

And the five important issues in this survey are growing the existing business, collaborating with competitors, shrinking the existing business, entering a new industry or starting a new business, and corporate financing.

The last questions of this introduction are how data analytics can improve your businesses? Below are the answered.

1. Make data-driven business decisions

Making evidence-based rather than intuition-based decisions

2. Grow your business – discover new opportunities

Quickly identify future markets and the best areas for new investments Boost growth through strategic pricing models and data-driven marketing

3. Create a more efficient and smarter organization

Predict and anticipate the impacts of economic, market, and regulatory forces on business strategy and results Use automation and advanced statistical software to handle and analyze huge volumes of data

4. Manage risk and regulatory

Minimize compliance risks by ensuring the completeness, accuracy, and availability of data sources

2.1 Big data definition

As data becomes more bigger, manipulating of available data to get insights and make business decisions can be a useful. Statistical methods, artificial intelligence, machine learning, and data manipulation are some of new term that every business leaders at every level need to become data literate and be able to understand data and analytical concepts.

Before going further, what is actually Big data?. Big data term comes from John Mashey in 1990, a computer scientist from Pennsylvania State University. It all starts with the big bang explosion in the amount of data we have created since the rise of the digital era. Side effects of the rise of computers, the Internet and technology that capable of capturing data from the all kind of electronic processing.

Big Data have 3 defining properties call 3V. This terms introduces by **Gartner analyst Doug Laney** in a 2001 MetaGroup research publication. He published that publication with the title “3D data management: Controlling data volume, variety and velocity”. Also not all data is big data except they have 3 properties. Some others data expert added another 2V, value and veracity, but the main term is always 3V.

1. Volume: Volume refers to the huge amounts of data generated each time from all of electronics device such as website, social media, cell phones, cars, credit cards, sensors, photographs, video, etc. Volume is the V most associated with the term of big data because, volume can be incredible exploding big. Below is the explanation of how data growing rapidly from sisense.com (Figure 1.8)

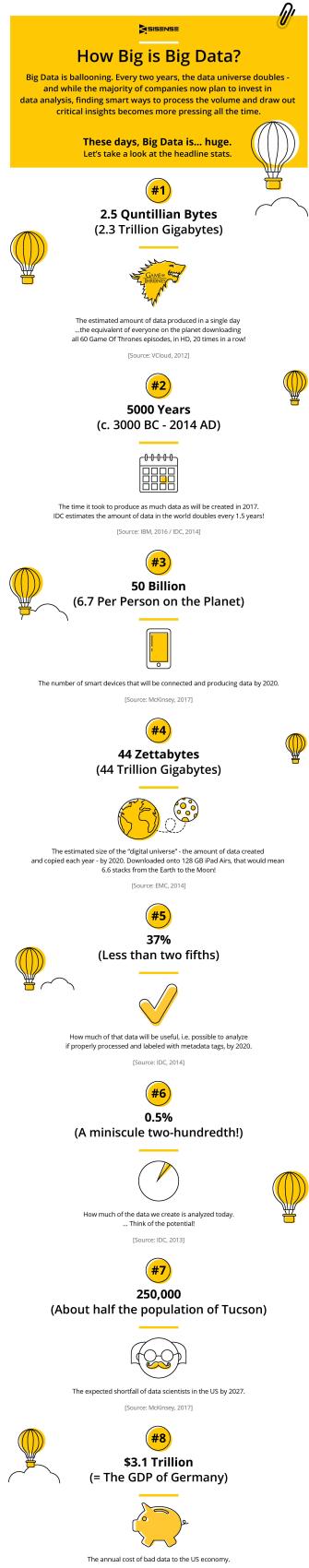


Figure 2.8:

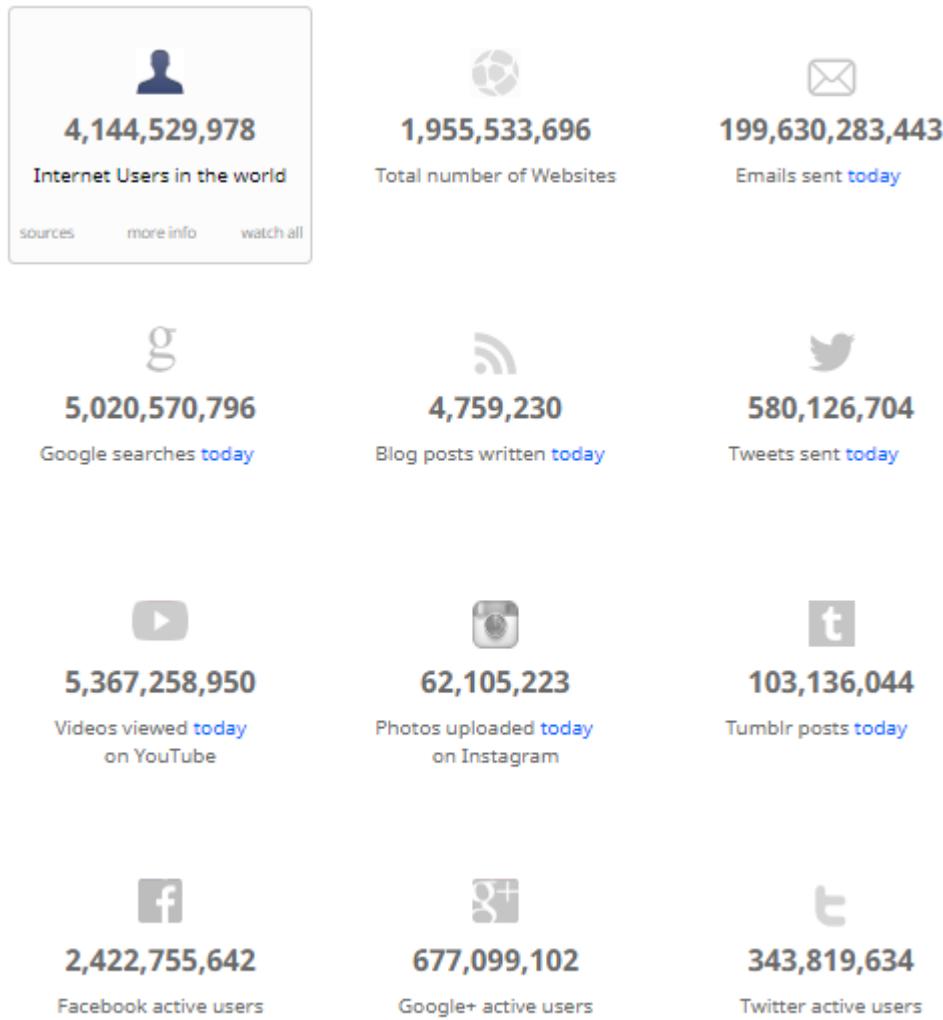


Figure 2.9:

For Example in internet world. Facebook is storing 250 billion images As far back as 2016, Facebook had 2.5 trillion posts. Can you imagine for the others online media such as twitter, blog, instagram and also our searching in google? It is a extreme amount of data. This link show the live statistics of some online area such as facebook, twitter., internet users, etc.

<http://www.internetlivestats.com/>

(Figure 1.9 and Figure 1.10)

Or

<http://www.worldometers.info/>

(Figure 1.11 & Figure 1.12)

For world parameter infos such as population, health, economics, food, water, energy, etc.

2. Variety: It refers to many sources and data types (structured and unstructured). Not only spreadsheet

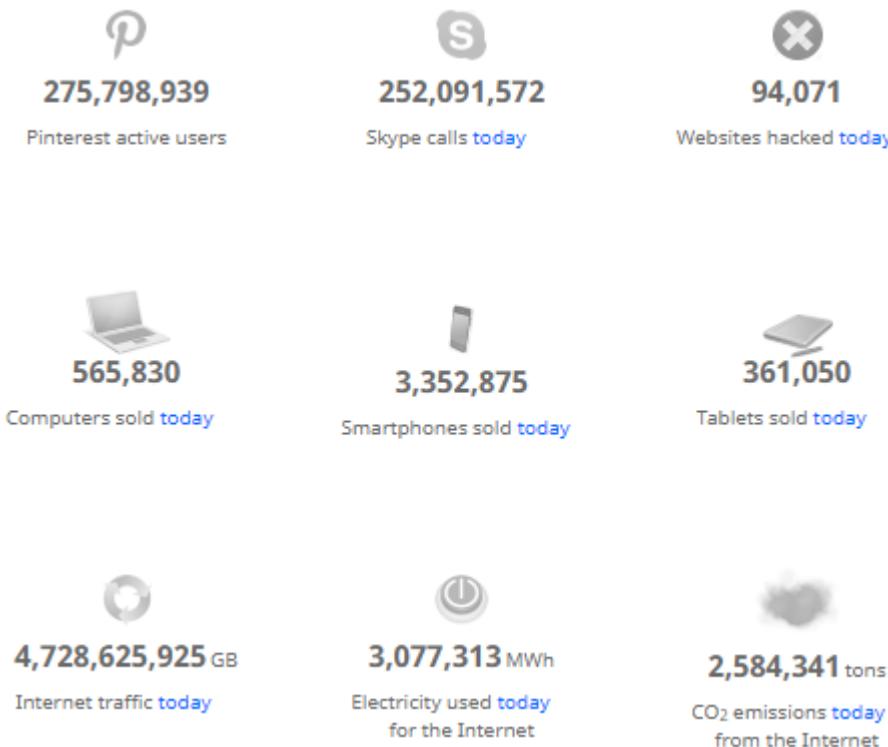


Figure 2.10:

WORLD POPULATION

| | | |
|---------------|---------------------------------|-----|
| 7,682,682,057 | Current World Population | [+] |
| 14,964,951 | Births this year | [+] |
| 310,716 | Births today | [+] |
| 6,278,815 | Deaths this year | |
| 130,366 | Deaths today | |
| 8,686,136 | Net population growth this year | [+] |
| 180,350 | Net population growth today | |

Figure 2.11:

HEALTH

| | | | |
|-------------------|----------------------------------|---------------------------|-----|
| 1,380,025 | Communicable disease deaths | this year | [+] |
| 808,032 | Deaths of children under 5 | this year | [+] |
| 4,465,971 | Abortions | this year | [+] |
| 32,858 | Deaths of mothers during birth | this year | [+] |
| 40,625,770 | HIV/AIDS infected people | | [+] |
| 178,706 | Deaths caused by HIV/AIDS | this year | [+] |
| 873,077 | Deaths caused by cancer | this year | [+] |
| 104,273 | Deaths caused by malaria | this year | [+] |
| 12,221,185,146 | Cigarettes smoked | today | [+] |
| 531,424 | Deaths caused by smoking | this year | [+] |
| 265,879 | Deaths caused by alcohol | this year | [+] |
| 113,996 | Suicides | this year | [+] |
| \$ 42,527,306,222 | Money spent on illegal drugs | this year | [+] |
| 143,501 | Road traffic accident fatalities | this year | [+] |

Figure 2.12:

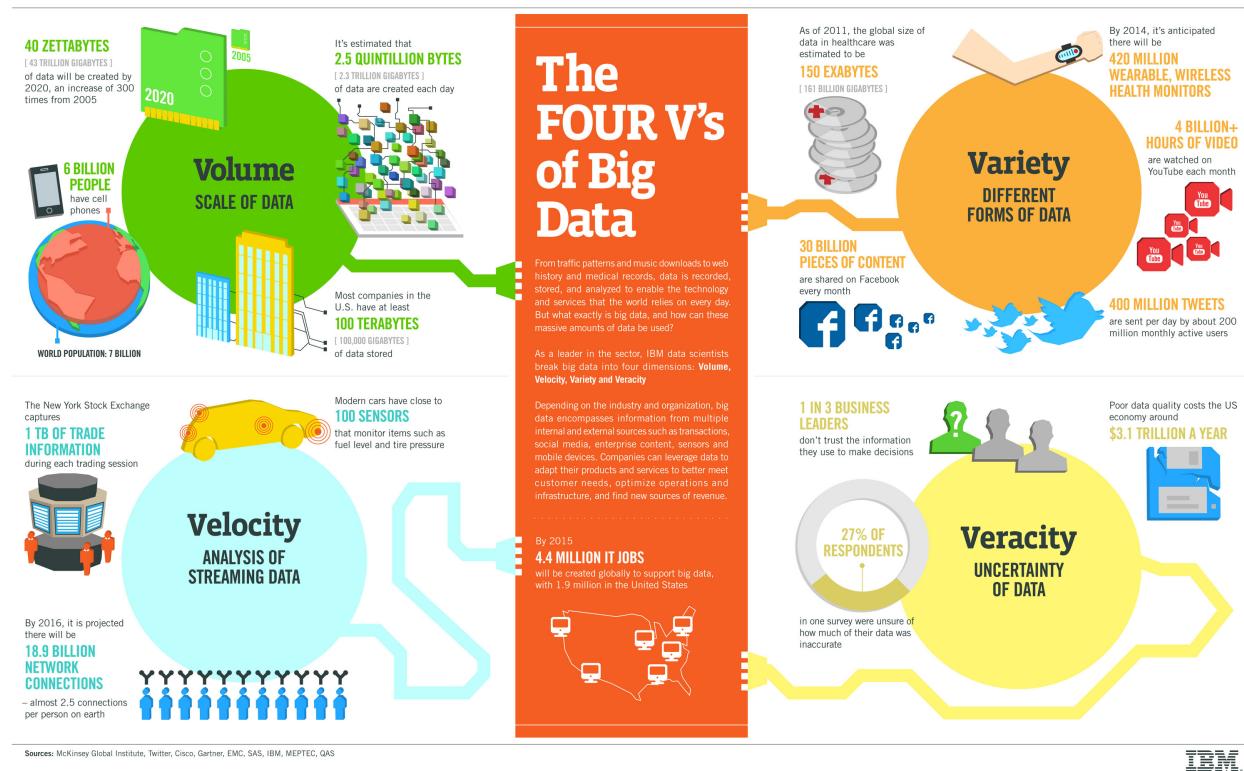


Figure 2.13:

and database but also emails, photos, videos, monitoring devices, audio, etc.

3. velocity: There are four velocity type in big data.

- Batch
- Periodic
- Near real time
- Real time

“Firehose” data sources such as social media and E-commerce need quickly analytics. Most importantly, use it at a faster rate than ever before. Using real-time alerting, Walmart company was able to find a particular Halloween novelty cookie of its stores where it wasn’t selling at all. More than realtime but predictive analytics.

The infographics & animations to understand more of 3V of big data from IBM are below (Figure 1.13).

There has been an rapidly growth in the field of Big Data with the benefits of the rising the new technology. This availability of big data has led to the use of big data in multiple industries ranging from

- Banking
- Healthcare
- Energy
- Technology
- Consumer
- Manufacturing
- etc

The next section are example use case big data in some of industries (our case studies: transportation and logistics industries and another industry).

2.2 T & L industry: Deutsche Bahn

Deutsche Bahn AG is a German railway company as the second-largest transport company in the world, after the German postal and logistics company Deutsche Post / DHL, and is the largest railway operator and infrastructure owner in Europe. Headquartered in Berlin, it is a private joint-stock company (AG), with the Federal Republic of Germany being its single shareholder.

The Deutsche Bahn Group is divided with some subsidiaries, such as:

- Personenverkehr
- Arriva
- DB Fernverkehr
- DB Regio
- DB Netze
- DB Engineering & Consulting
- Logistics
- ect

Deutsche Bahn Group (DB Group) is an international provider of mobility and logistics services operating globally in more than 130 countries. DB Group has more than 310,000 employees, with almost 40% employed outside Germany.

In passenger transport, we transport more than 12.5 million people each day on our trains and buses throughout Europe. In freight transport and logistics, our European network transports over 270 million t of goods per year by rail, and over 100 million shipments by road. Our global networks move about 1.3 million t of freight by air and nearly 2.2 million TEU by sea. At about 33,000 km, our rail network in Germany is Europe's longest. We are also the fifth-largest energy provider in Germany. The main components of our integrated rail system are our passenger transport activities in Germany, our rail freight transport activities, the operating service units and the rail infrastructure companies (RIC).

As an operator of networks and provider of services in passenger transport, freight transport and logistics, as well as track infrastructure, our economic success is influenced by the general economic environment and the specific development of the various relevant markets.

There are four key success factors in the development of DB Group, which are a central component of DB Group's business model:

1. Entrepreneurial approach to business: in the course of the German rail reform DB Group has established itself as a commercial enterprise. Particularly worth mentioning in this context are the establishment of a modern and efficient organization and a value-based management approach with capital market viability as a target.
2. Integrated Group: as a system integrator in Germany, DB Group optimizes the integrated rail system. In doing so, it serves as an important driving force for technological innovation. The integrated Group structure enables us to achieve positive synergies and align our infrastructure to support efficiency, market orientation and profitability.
3. International position: due to our focus on Europe in passenger transport as well as our European and global orientation in the areas of freight transport and logistics activities, DB Group has an excellent position in the relevant markets. As a result, we are responding to the increasing demand for cross-border solutions. At the same time, we are best positioned to take advantage of growth opportunities.
4. Cross-modal transport solutions: we offer our customers door-to-door mobility and logistics solutions from a single source. We use digital technologies to intelligently link various modes of transport in an economical and environmentally friendly way. In addition, we offer complementary products and services in the freight transport and logistics market.

One of the subsidiary company that focusing on big data analytics to achieve that key success factors is DB Systel GmbH (<https://www.dbsystel.de/dbsystel>). The key point of DB Systel are:

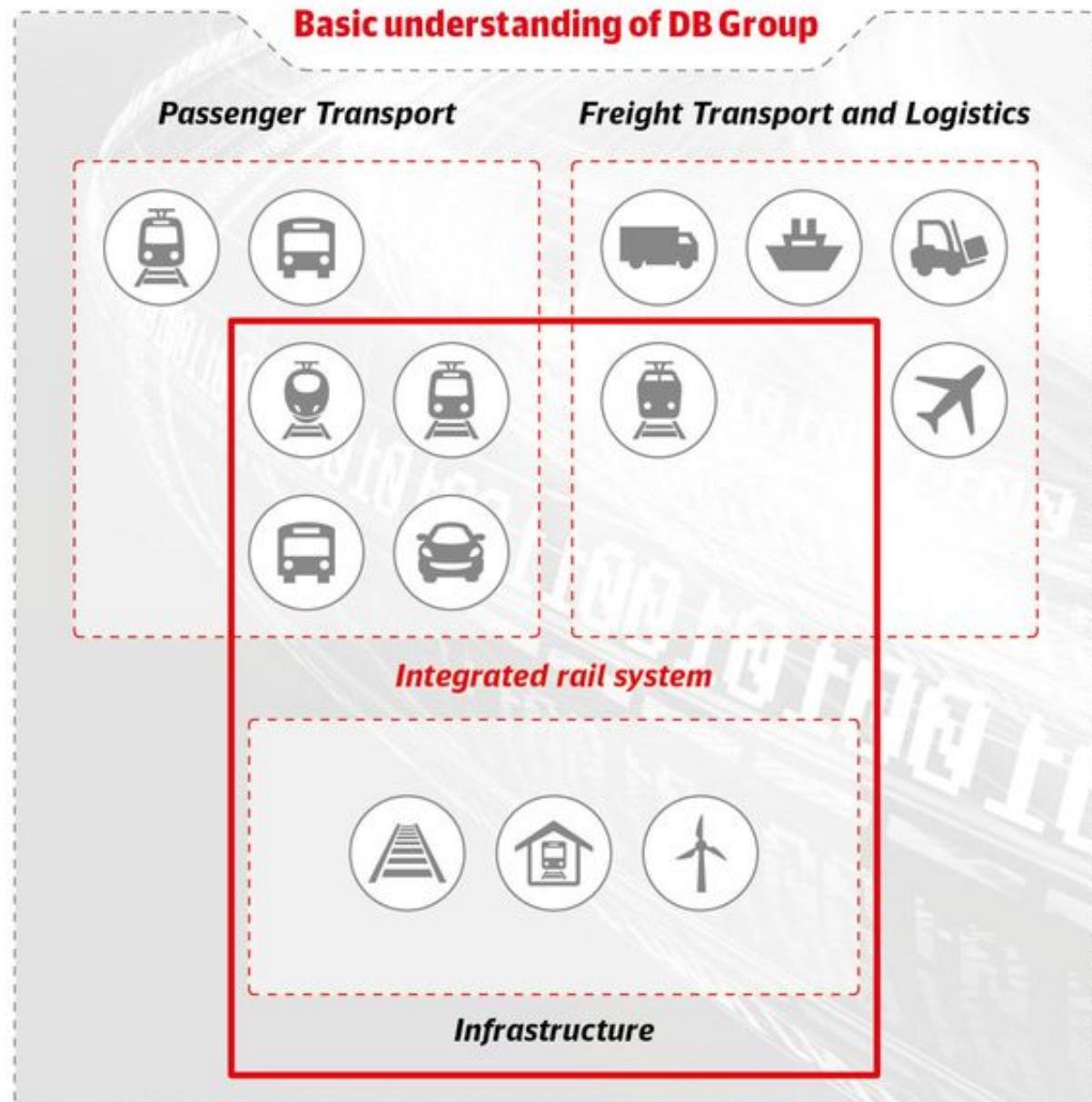


Figure 2.14:

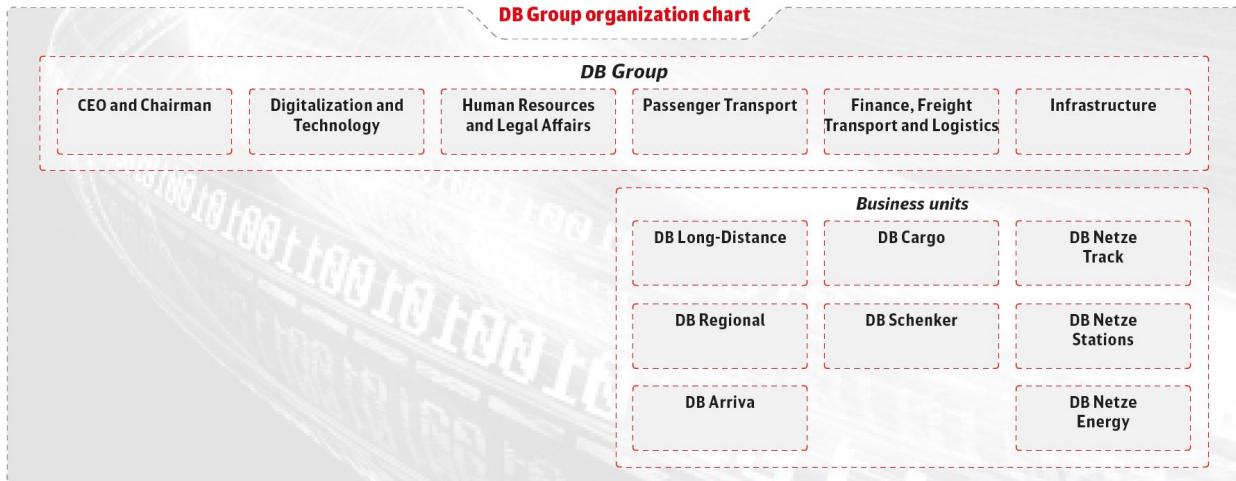


Figure 2.15:

Big data

If the data can no longer be stored by conventional databases, it is known as big data. We need mathematical methods, statistical, machine learning, and algorithms to get insight from the big data.

Business intelligence/business analytics

The process to collect, analyse, and visualize data, either as bar graph, pie chart or another form. The KPI (Key Performance Indicators) are the main end results from this process. And support decision making in the present and also future

Data mining

The three main concepts of data mining are statistics, mathematics, and algorithms. To detect the hidden patterns and connections in big data we need some methods, such as clustering (formation of groups of similar data), regression analysis (what depends on which other factors?) and association (if one thing happens, another does too)

Data lake

Every day, individual generates around 650 megabytes of data. There is a lake of data. Creating the data lake is a logical way of helping company to get flexible access to all of the data sources. A data lake is a central location where every department can store external and internal data (structured and unstructured data at any scale).

Data Lakes compared to Data Warehouses.

Smart data

While big data and the data lake consist of structured data and unstructured data, smart data have a specific purpose using certain algorithms and other tools. Smart data developed and improving a business processes and decision-making. The smart data consist also the analytics Edge, such as:

- Descriptive analytics
- Predictive analytics
- Prescriptive analytics

The description of this three analytics edge you can find in the next chapter 3. Big Data Analytics

Like most other industries, transportation and logistics (T&L) is currently confronting immense change and like all change, this brings both risk and opportunity. This risk and opportunity have new technology, new

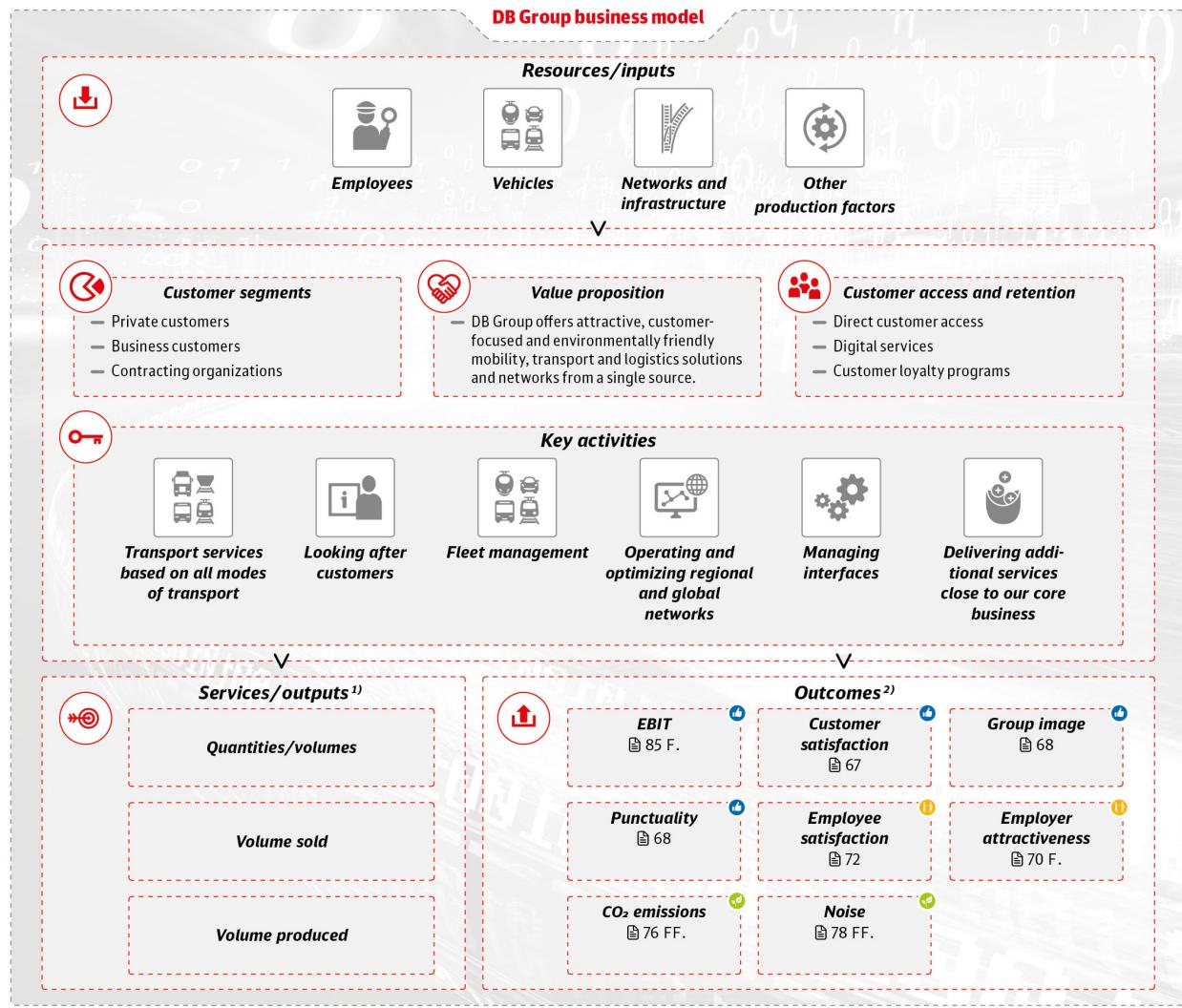


Figure 2.16:

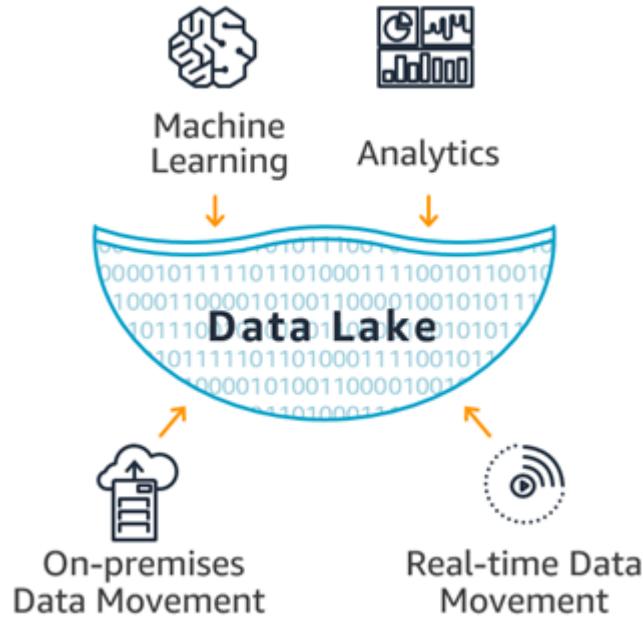


Figure 2.17:

| Characteristics | | Data Warehouse | Data Lake |
|--------------------------|---|--|-----------|
| Data | Relational from transactional systems, operational databases, and line of business applications | Non-relational and relational from IoT devices, web sites, mobile apps, social media, and corporate applications | |
| Schema | Designed prior to the DW implementation (schema-on-write) | Written at the time of analysis (schema-on-read) | |
| Price/Performance | Fastest query results using higher cost storage | Query results getting faster using low-cost storage | |
| Data Quality | Highly curated data that serves as the central version of the truth | Any data that may or may not be curated (ie. raw data) | |
| Users | Business analysts | Data scientists, Data developers, and Business analysts (using curated data) | |
| Analytics | Batch reporting, BI and visualizations | Machine Learning, Predictive analytics, data discovery and profiling | |

Figure 2.18:

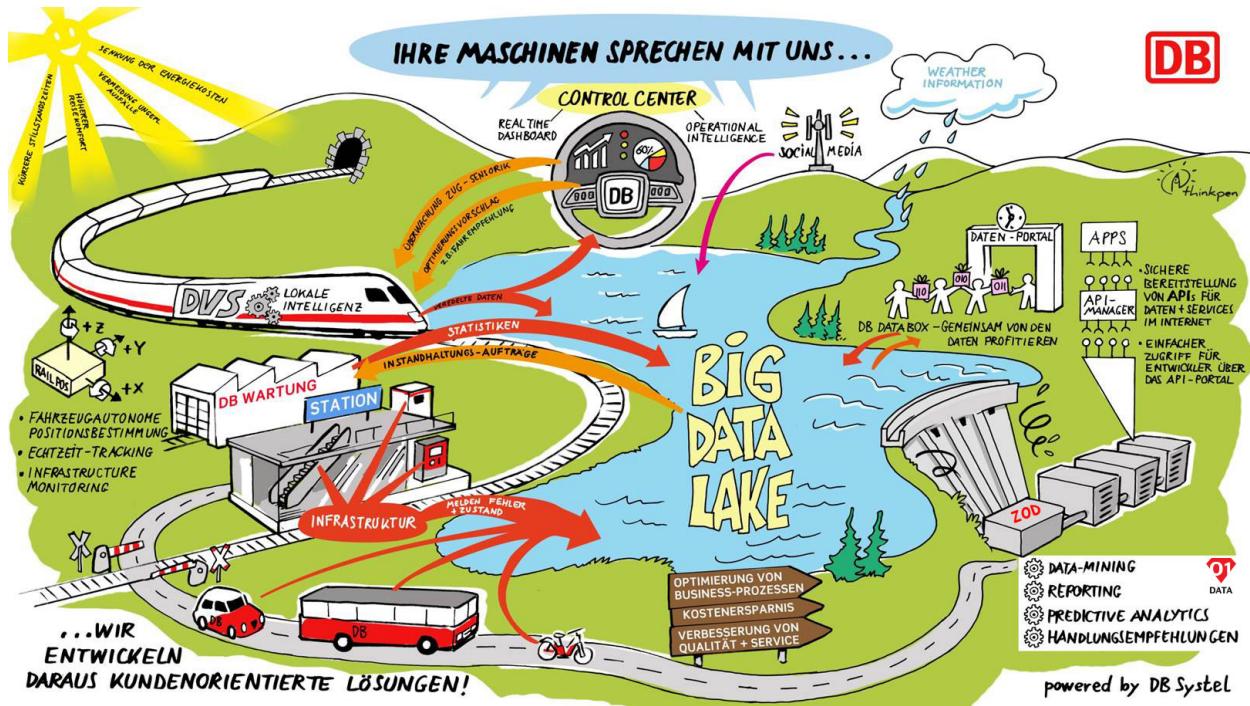


Figure 2.19:

market, new customers expectation, and new business modell. Based on pwc (PricewaterhouseCoopers) research, there are four disruption and uncertainty and five forces transforming in the era of big data for T&L industry (Transportation & Logistik).

Four areas of disruption and uncertainty:

1. Changing customer expectations

Like individual consumers, industrial customers now expect to get shipments faster, more flexibly, and with more transparency at a lower price. No surprise that across the industry, both operating models and profitability are under strain. And the pace of transformation for large manufacturing and retail customers may turn out to be even faster than for private final consumers.

2. Technological breakthroughs

Technology is changing every aspect of how logistics companies operate. ‘Digital fitness’ will be a prerequisite for success: the winners will be those who understand how to exploit a whole range of new technologies, from data analytics to automation and platform solutions. Those who don’t, risk obsolescence. But with so many technologies competing for management attention and investment, defining a clear digital strategy that’s integrated into business strategy will be critical.

3. New entrants to the industry

Platform technology has given rise to new business models, often driven by start-ups that enter the logistics industry. New ‘sharing’ business models could have as much of an impact on the sector as new technology. And the industry’s current customers and suppliers may end up being the biggest new entrants.

4. Redefining collaboration

Horizontal collaboration is already happening, especially in last-mile delivery, but it’s hampered by inconsistencies. Higher levels of efficiency could be achieved by more consistent standards, defined through the Physical Internet and increased collaboration, whether in the form of alliances, joint ventures or M&A.

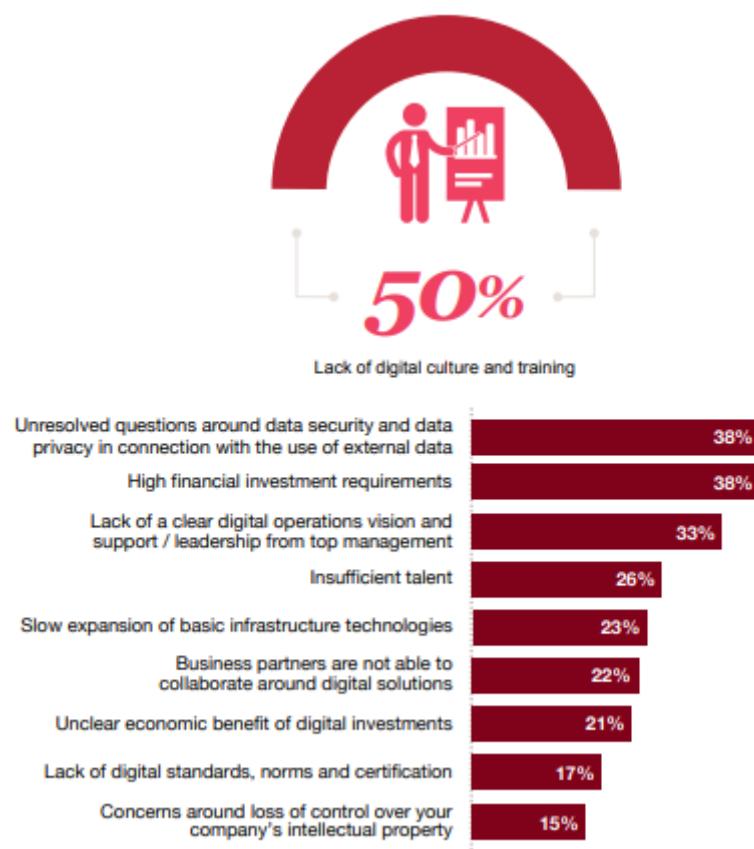


Figure 2.20:

| The technology ¹⁰ | The impact | The uncertainties |
|---|---|---|
| Physical Internet (based on the IoT) | <ul style="list-style-type: none"> Improved supply chain transparency, safety and efficiency Improved environmental sustainability (more efficient resource planning) | <ul style="list-style-type: none"> Social expectations around data privacy and security may change Regulation around data security and privacy may increase or be enforced more stringently The sector's willingness and ability to invest in collaboration Whether international bodies will drive standardisation |
| IT standards | <ul style="list-style-type: none"> Enabling collaboration horizontally More efficiency and transparency | <ul style="list-style-type: none"> Companies' willingness to adopt is uncertain due to data security concerns |
| Data analytics | <ul style="list-style-type: none"> Improvements in customer experience and operational efficiency in operations Greater inventory visibility and management Improved 'predictive maintenance' | <ul style="list-style-type: none"> Rate of development of data processing capacity is unclear Question marks around data security Social expectations around data privacy and security may change Regulation of data security and privacy may increase or be enforced more stringently |
| Cloud | <ul style="list-style-type: none"> Enabling new platform-based business models and increasing efficiency | <ul style="list-style-type: none"> Development of costs unclear (once a certain scale is reached physical data centres still tend to be cheaper) Uncertainties around data security |
| Blockchain | <ul style="list-style-type: none"> Enhanced supply chain security (reduction of fraud) Reduction in bottlenecks (certification by 3rd parties) Reduction of errors (no more paper-based documentation) Increased efficiency | <ul style="list-style-type: none"> Rate of adoption uncertain Unclear whether one or two dominant solutions will emerge or multiple competing solutions |
| Robotics & automation | <ul style="list-style-type: none"> Reduction in human workforce and increased efficiency in delivery and warehousing (including sorting and distribution centres) Lower costs | <ul style="list-style-type: none"> Speed of technology development unclear |
| Autonomous vehicles | <ul style="list-style-type: none"> Reduction in human workforce Increased efficiency in delivery processes | <ul style="list-style-type: none"> Regulatory environments not currently in place in most countries Liability issues not yet clear Ethical questions remain especially in relation to emergency situations |
| UAVs / Drones | <ul style="list-style-type: none"> Increased cost efficiency (use cases: inventory, surveillance, delivery) Workforce reduction | <ul style="list-style-type: none"> Regulation in most countries not sufficient for commercial use in public areas like delivery Safety and privacy concerns may hamper market acceptance |
| 3-d printing | <ul style="list-style-type: none"> Lower transportation demand Transported goods would mostly be raw materials | <ul style="list-style-type: none"> Speed, scale, and scope of uptake by customer industries still unclear |

Figure 2.21:

The five forces transforming transport and logistics and their key driving trends

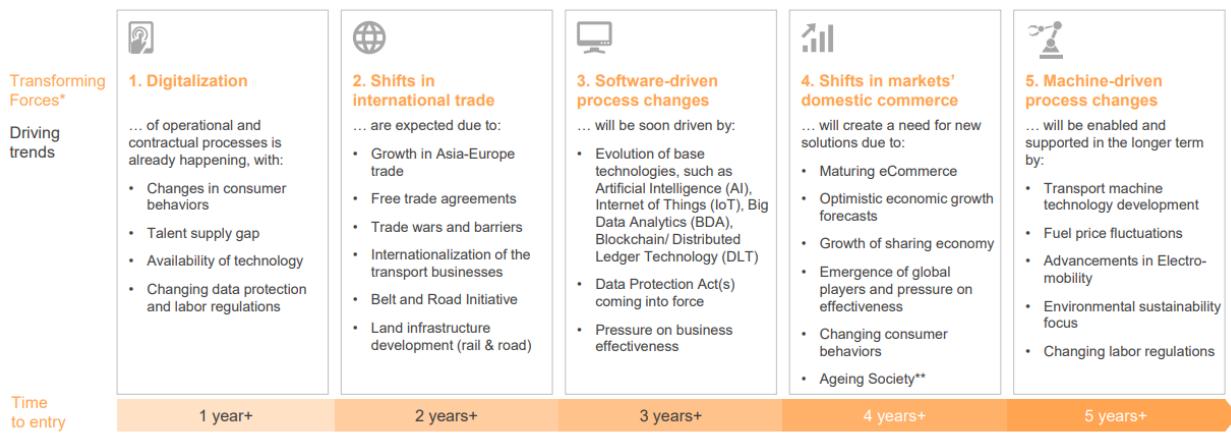


Figure 2.22:

Five Forces Transforming Transport & Logistics:

As the complexity of modern transport and logistics grows, it is increasingly difficult to understand what to focus on, so PWC have identified five key forces transforming the T&L segment:

1. Digitalization
2. Shifts in international trade
3. Software-driven process changes
4. Changes in markets' domestic commerce
5. Machine-driven process changes

2.3 Another industry

2.3.1 Agriculture

Agriculture businesses are using big data to improve production and yields and improve forecasting to better optimize supply chains. Getting deep insight through big data is important to highing product quality or rising sales and market share.

As agribusinesses growth and more diverse, the growing of data that must be managed are also becoming more huge and complex. The data comes from not only social media outlets and supplier network channels but also agricultural devices sensor and machine equipment. Today these data sources come from:

- Traditional enterprise data from operational systems
- Farm field sensor data (e.g. temperature, humidity, rainfall, sunlight)
- Farm equipment sensor data (from tractors, plows, and harvesters)
- Harvested goods and livestock delivery vehicles (from farms to processing facilities) sensor data
- Commodities trade data
- Financial data
- Weather data
- Animal and plant genomics research data
- Social media

Big Data can help improve forecasting and efficiency and lead precisely decision making. Big data technologies enable agricultural sectors to analyze a variety of data sources, in turn leading to better outcomes. Agribusinesses have long lists of needed metrics. But which data that we need and where the data come from

and also what is the outcome or beneficially result. Following is a list of areas where Big Data technologies can impact in agricultural sectors:

- Weather data from weather institution for better understanding time to plants
- Improved forecasting of yields and production data from Farm field sensor data (e.g. temperature, humidity, rainfall, sunlight)
- Better optimized seeds and livestock and new methodologies that improve yields and production with Farm equipment sensor data such as warehouse sensor data
- Faster delivery of goods produced to distribution centers and consumers with street data or vehicle (car) sensor data
- Real-time decisions and alerts based on data from fields and equipment with Farm equipment sensor data or Farm field sensor data
- Integrated production and business performance data for improved decision making with ERP system from integrated all sensor devices and sales data.
- Rationalized performance data across multiple geographies with sales data
- Accurate crop predictions with weather and crop data
- Stronger seeds and less hunger with plant data with purpose developing crops that can grow in any environment -Automated agriculture for better production with different system

2.3.2 Automotive

The automotive industry face a compleks challenges. One thing that has the opportunity to deliver solution is analytics. Shifting marketing conditions, volatility, competition, and cost pressure are the problem to change in the automotive industry.

As the automobile is being transformed by big data technologies (everything from sensors to artifical intelligence to big data analysis). The ecosystem are change and the analytics allows the data not only structured but also unstructured data such as videos, sound recording, or texts to be analyze- the results are impressive.

Below are the key points to deep undestand the automative industry:

- **Use the data for actionable customer segments and individualised offers and to boost sales and also improve customer retention.**

Customer behaviour analytics: Customer need a consistent personalised experience across their access channels. Automakers are having to sell by offering 24/7 connectivity and reliance on social media and internet as communication tool. There is a huge of data available to support automakers for understand their customers but automakers have limits and ability to collect, analyse, and act on it.

Automakers need to fully understand customers needs and behaviours to develop a bird view of customer segmentation. For example, the combination of online retailing and physical stores to build brand awareness, attract new target and transformational retailing experience for shopping anytime and anywhere. At least but not last, customer retention needs to be placed high on the strategy, understanding the customer experience with the lifecycle and journey not only just during the purchase.

The example of real strategy:

Inner city mobility: Pay for use customer mobility solution, covering central city locations, offering ease, speed and flexibility of commute

Urban lifestyle convenience: Taking direct retailing and servicing business to the consumer

Regional retailing: Providing the complete digitised brand experience through existing dealers

- **Applying big data analysis to a mass historical data to identify the impact of fixed and variable marketing parameters and supprt auto industry with a more precise and effective approach to quantity and composition of marketing cost.**

The growth in data and information available on customers is allowing automakers and dealers to focus on spesific customers. To make this happens, automakers bring togethers internal and external information

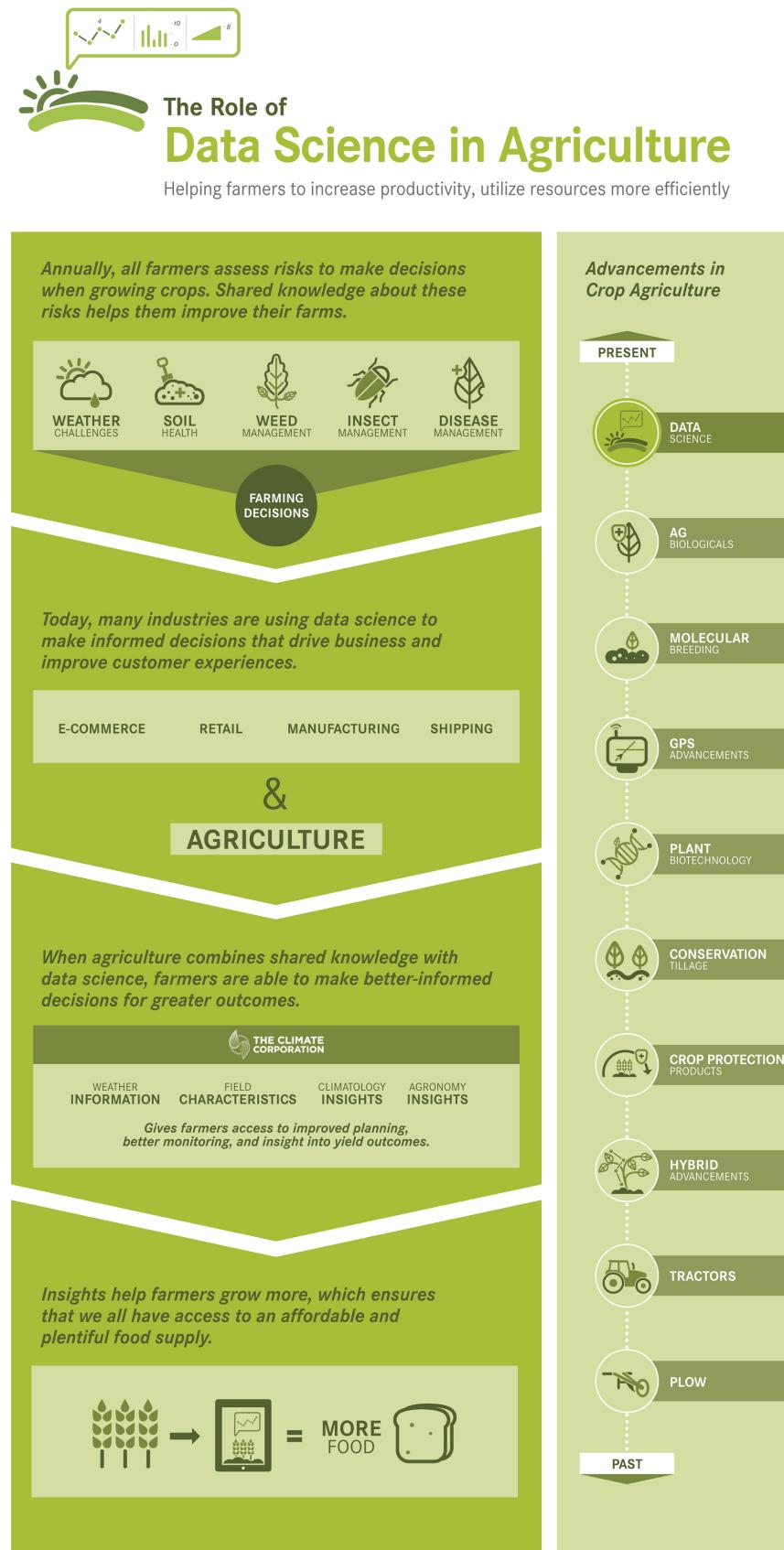


Figure 2.23:

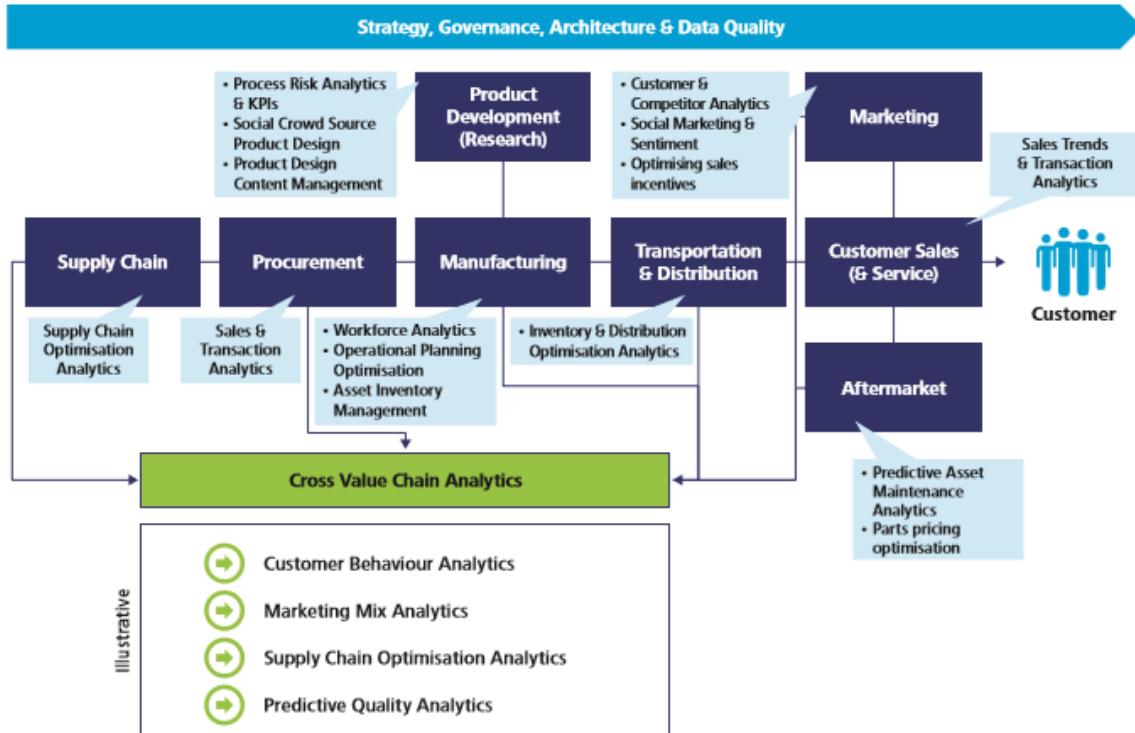


Figure 2.24:

sources relating to variable marketing spend to better understand what's working and what isn't. Besides that, there are some issues that must be overcome, including:

- Lack of available data on transaction for internal and competitor brands.
- Less understanding of competitor marketing and strategies
- Differences in customer behaviours across different regions and customer segments
- Less ability to understand the impact of trends on market and buying behaviours
- Supply chain data can reveal which links in chain could weaken allowing for proactive and timely countermeasure before real problems manifest.

Globalising operations to take advantage of high-growth markets, driving innovative strategies to optimise manufacturing process. Get it right and gain a competitive advantage and drive growth. Get it wrong and automakers can get difficult scenarios from parts shortages, government security, or lost growth opportunities. Moreover, advanced supply chain analytics can help automakers analyse large datasets using analytical and mathematical techniques. An example of this is the use of product configuration and web interactions which allow automakers to get early of emerging trends and forecast.

- Predictive analytics for forecasting efficiency as well as operations and performance.

The good news is that predictive quality analytics capabilities available today allow quality issues to be detected and taken preventively. This news provided improved ability to better manage customer satisfaction and cost control concerns.

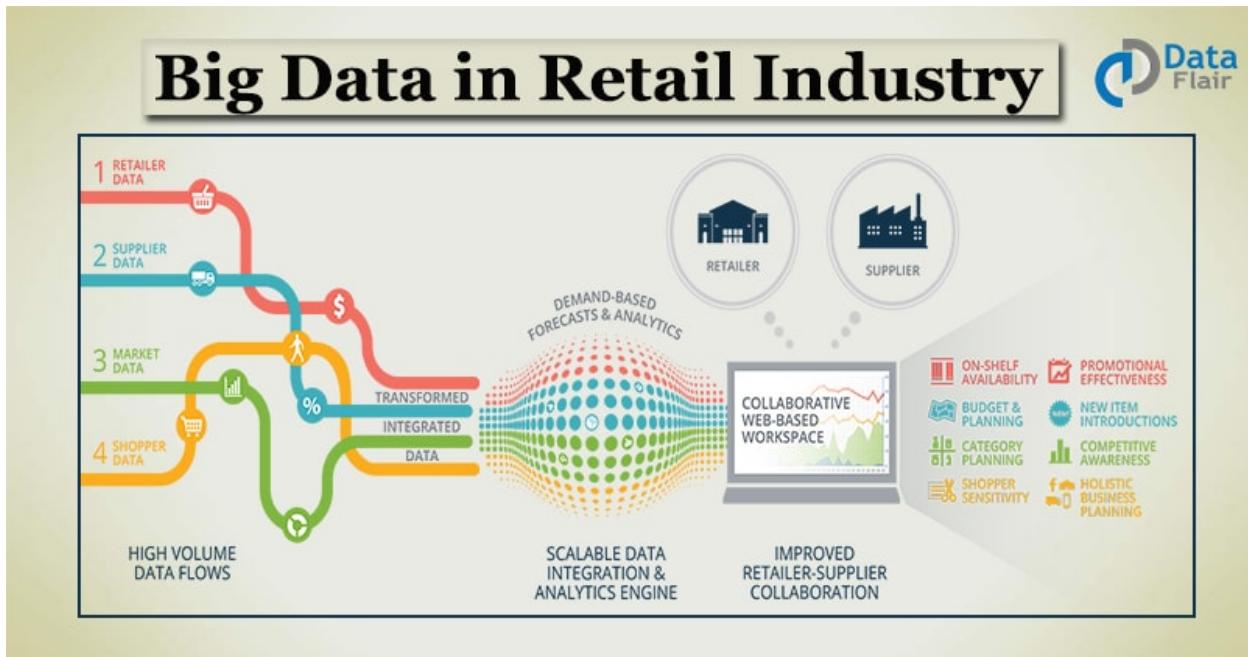


Figure 2.25:

2.3.3 Retail

The sphere of the retail industry develop rapidly. Besides that, retail industry is the top use case of big data. The retailers manage to analyze data from different sources and develop overview of a customer to learn targeting sales. Besides that, a customer need to be easily influenced by the strategy develop by another retailer. New sources of data, from machine data (log files) and transaction, to sensor and social media, present opportunities for retailers to get impacted value and competitive advanced. To achieve this is to make best plans and decisions, understand customer more deeply, uncover hidden trends, and becoming data driven organisation.

This examples presents top big data analytics use cases in the retail, created for retailers to be aware about trends and tendencies using big data.

1. Customer Behavior Analytics for Retail: Retailers can combine, integrate and analyze all of your external and internal data to generate the deeply or hidden insights.
2. Customer Journey Analytics: Today's customers are more connected than before. Using mobile devices, social media and e-commerce, customers can access all of information in seconds. With big data technologies, retailers can bring data into powerful insights. Retailers will be able to get deeply understanding this questions such as:
 - What is s really happening across every step in the customer journey?
 - Who are your high-value customers and how they behave?
 - How and when is it best to reach them?
3. Personalizing the In-Store Experience With Big Data

Big data analytics can turn data sources into advantage for retailers. These data sources can be gathered from:

- Websites
- Point of sales systems
- Mobile apps
- Supply chain systems

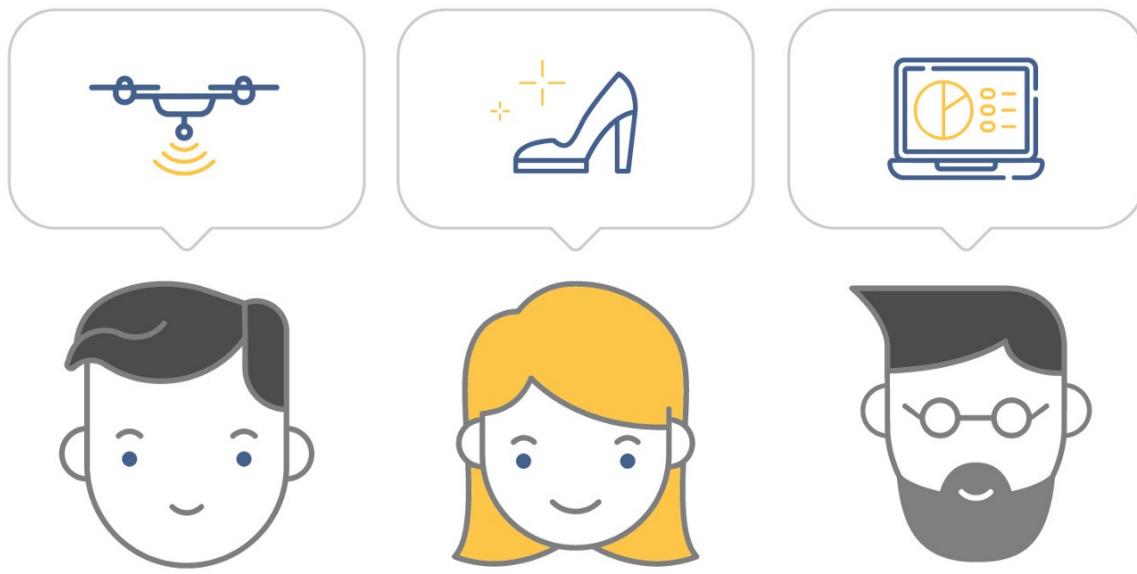


Figure 2.26:

- In store sensors
 - Cameras and more
4. Increasing conversion rates through predictive analytics and targeted promotions

To reduce costs and high customer acquisition, retailers need targeting promotions. This requires a 360-degree view of customers situation.

In the past, customer information has been limited to demographic data during sales transactions. But today, customers interact more than they transact, as example is social media. Using data from internal or external sources, omni-channel retailers can:

- Test and quantify the impact of different promotional tactics on customer behavior and conversion
 - Use a customer purchase and browsing history to identify needs and interests and then personalize promotions for customers
 - Monitor customer purchasing behavior and social media activity to drive timely offers to customers to incent online purchases with a specific retailer
5. Recommendation engines

Recommendation engines is one of tools for customers behaviour prediction. Recommendation engines provided retailers to increase sales and to dictate trends. Recommendation engines depend on the choices by the customers. Usually, this engines use either collaborative filtering or content based recommendation.

- Collaborative Filtering:

This system based on how similar users liked the item. As example, Cevi and Herdian have similar interests in games. Cevi buyed and played “Game of Thrones 4”. Herdian has not played this game. But the system recommend this game because the system has learned that Cevi and Herdian have similar interests. In simple words, users who liked this game also like too.

- Content-based Recommendations:

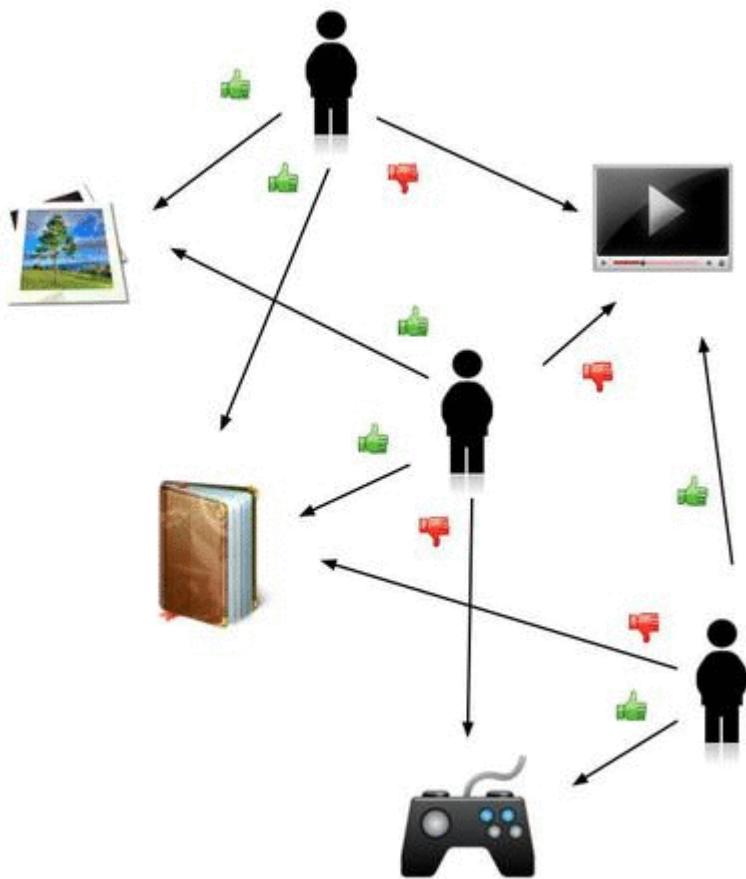


Figure 2.27:

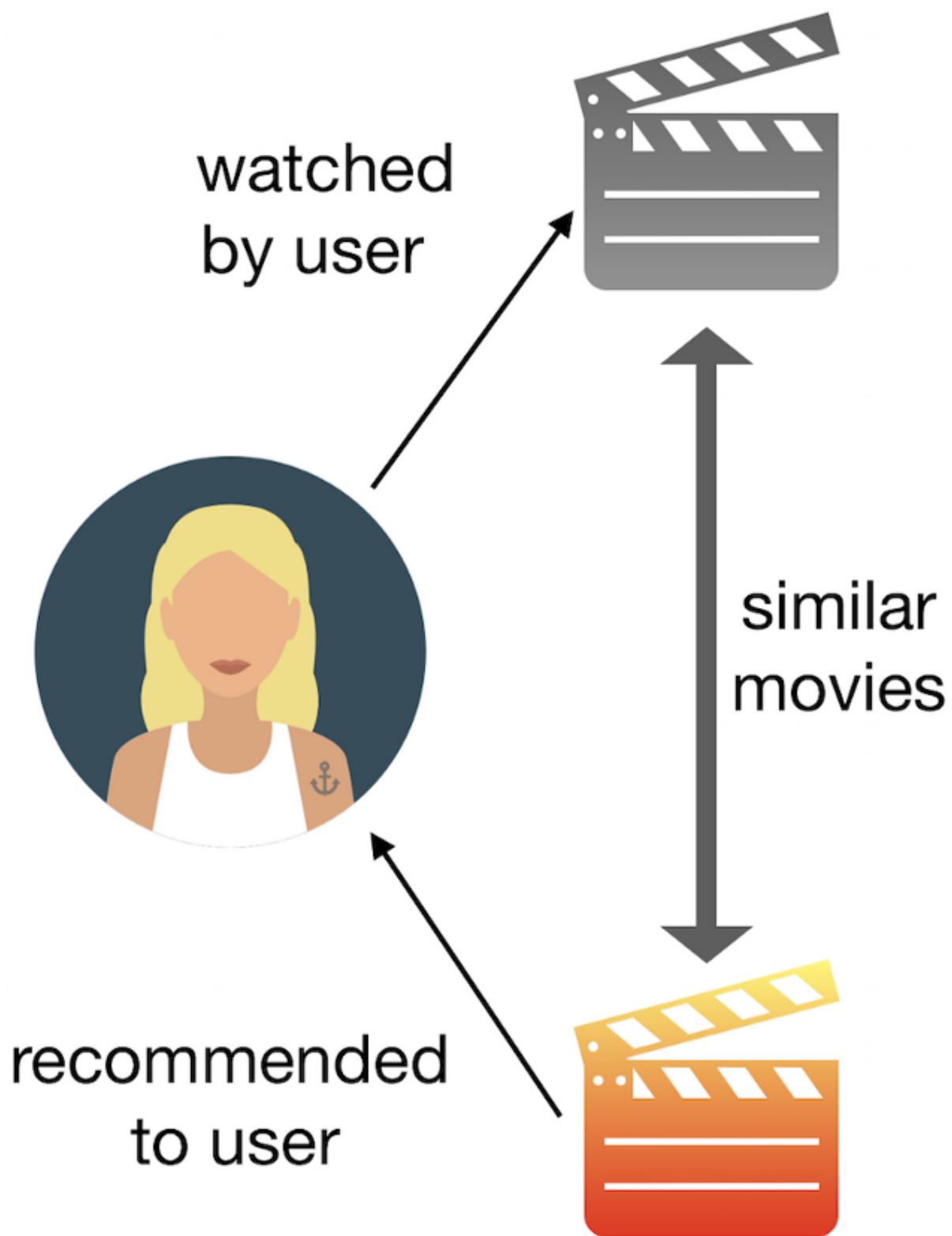


Figure 2.28:

The company must have detailed metadata (details data explanation about the data) about each of buyers. Retailers can recommend buyers with similar tags. For example, Fathimah watch the movies “Game of Thrones 1” on Netflix. This movie may have metadata tags of “War”, so Netflix recommends to Fathimah another movies with metadata tags “Ancient”.

6. Market basket analysis (transaction data)

Market basket analysis may be regarded as traditional analysis in the retail. This is a modelling method based on the theory that if you buy a certain items, you are more (or less) likely to buy another items. This process mainly depend on the amount of data collected via customers transactions. It works by looking for combinations of items in transactions. Association Rules are widely used to analyze retail basket or transaction data. An example of step by step Association Rules are:

- Assume there are 1000 customers
- 100 of them bought only CD games, 850 bought Playstation and 50 bought both of them.
- Bought CD games => bought Playstation
- Support = $P(\text{CD games} \& \text{Playstation}) = 50/1000 = 0.05$
- Confidence = support/ $P(\text{Playstation}) = 0.05/(850/1000) = 0.05/0.85=0.06$
- Lift = confidence/ $P(\text{Milk}) = 0.75/0.10 = 7.5$

Note:

Support:= the number of transactions that include items in the {A} and {B} parts of the rule as a percentage of the total number of transactions.

Confidence:= the ratio of the number of transactions that include all items in {B} as well as the number of transactions that include all items in {A} to the number of transactions that include all items in {A}

Lift:= the ratio of confidence to expected confidence. Expected confidence is the confidence divided by the frequency of B

7. Fraud detection

Fraud can encompass abuse and waste, improper payments money laundering, terrorist financing, cybersecurity and public security. The detection of fraud is important and challenging activity of retailers. The only good way to protect retailers is to be one step ahead of fraudsters.

To identify of fraud-while improving customers experienceces-retailers should follow four steps:

1. Extract and transform all available data types from across departments or channels and incorporate them into the analytical process.
2. Continually monitor transactions, social networks, high-risk anomalies, etc., and apply behavioral analytics to enable real-time decision making.
3. Install analytics culture through data visualization at all levels, including investigative workflow optimization.
4. Employ layered security techniques (another company who experts in this area).
5. Warranty analytics

With warranty issues costing retailers billions and indirect cost factors. To reduce warranty costs, build loyalty customer service, smart manufactures, and focus on post-sales service. The methods of detecting are complicated. They concentrate on the detecting anomalies in the warranty claims.

The warranty analytics have some of challenges:

- Increasing warranty costs and poor customer satisfaction ratings.
- Cannot track warranty claims and resolutions in a timely manner.
- New issues grow in the field for months before they are detected.



Figure 2.29:

- Hard to integrate and analyze historical/supplier warranty information, so it is difficult to anticipate and prepare for future issues.
- Business analysts and executives do not have ready access to warranty data and reports.
- Difficulty integrating quality and warranty data.
- Root cause analysis takes too long.
- Feeding warranty claim information back to design and manufacturing takes too long.
- Ineffective, manually intensive processes.

9. Customer sentiment analysis

Social networks (social media) and online services feedbacks can perform the sentiment analysis. Because social media data are available, also much easier to implement analytics on social platforms. This analysis uses language processing to know a positive or negative attitude of customer. The results are as a insight for services improvement.

The examples are:

- I do not dislike this games. (Negation handling)
- Disliking a games is not really my thing. (Negation, inverted word order)
- Sometimes I really hate the Gun. (Adverbial modifies the sentiment)
- I'd really truly love going out in this weather! (Possibly sarcastic)

10. Price optimization

Pricing optimization is the setting a price to maximize profits. It is driven by the supply and demand data, behavioural data, competition data, etc. The right price for customers and retailers is significant advantage brought by optimization method. The price depends not only on the costs to produce but on the competition. The big data analysis bring this issue to the next level. The data come from many sources and define buying attitude, seasoning, competitor pricing, etc. The algorithm define the response the changes in prices. Using the real time optimization the retailers have an advantage to attract more customers and to build personal pricing schemes.

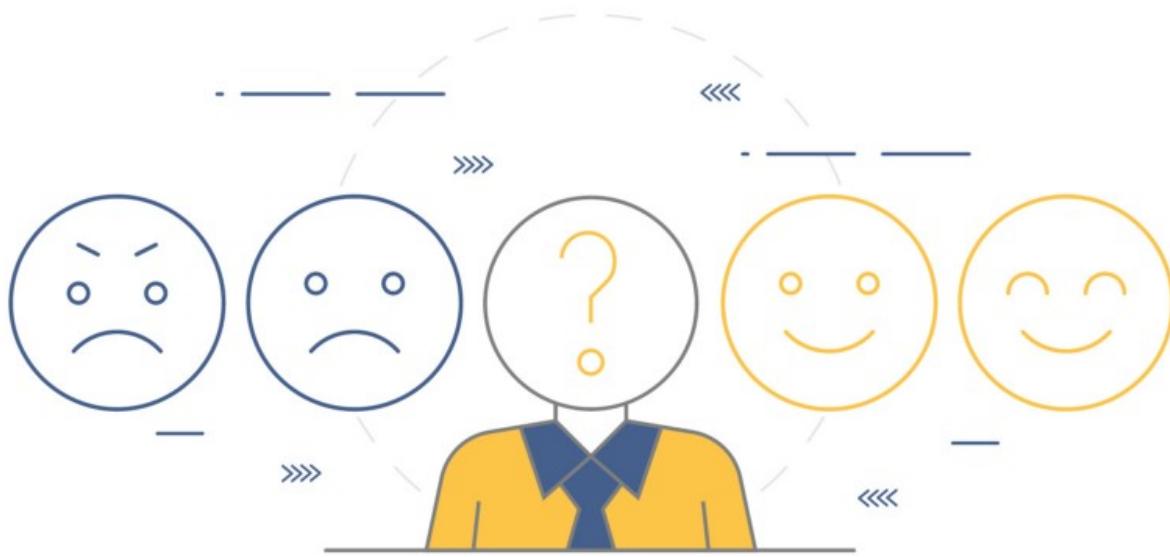


Figure 2.30:

Below are overview four 6 pricing strategies with big data:

1. Premium pricing: Retailers set costs higher than their competitors with premium services or items.
 2. Pricing for Market Penetration: Retailers offer lower prices on goods and services for first time opened.
 3. Bundle Pricing: Buy 1 get 1 pricing.
 4. Psychology Pricing: For example, setting the price of a watch at \$499 is proven to attract more consumers than setting it at \$500.
 5. Price Skimming: The retailers offer lowers prices as competitor goods appear on the market not only for the first time opened.
 6. Economy Pricing: An example of economy pricing is generic food sold at grocery stores. They need little marketing and promotion expenses.is incredibly effective for large retailers companies.
11. Location of new stores

Data analytics is efficient about this issue. The method is simple and also useful. The data are customers data, ZIP code and location for understanding the market potential. The Analyst find the solution by connection all the data source.

12. Lifetime value prediction

The customer lifetime value (LTV) is one of the most important parameter in retail industry. CLV is a total value of customers profit to the retailer over the entire customer business relationship. Retail can measure how long to it takes to retargeting the investment required to earn the customers.

The forecast are made on the past data to the most recent transactions data. The application of the statistical methodology and machine learning helps retailer to identify customers buying pattern. It is a important business process for the retailers. More customers understanding, more high LTV value.

2.3.4 Financial Services

Big data analytics can be the main driver of innovation in the financial industry. Using big data analytics in the financial industry is more than a trend, it has become the core analytics. The Banks and Insurances company have to realize that big data can help them more efficient, smarter decisions, and improve performance. This can be applied to the following activities:

- Banking and Insurance Industry:
 - Discovering the spending patterns of the customers
 - Identifying the main channels of transactions (ATM withdrawal, credit/debit card payments)
 - Splitting the customers into segments according to their profiles
 - Product cross-selling based on the customers' segmentation
 - Fraud management & prevention
 - Risk assessment, compliance & reporting
 - Customer feedback analysis and application
- Only Insurance Industry:
 - Better product design and marketing
 - More accurate risk assessment, underwriting, and pricing processes
 - Stronger commitment to helping customers
- Only Banking Industry:
 - Better claims management
 - Better customer targeting and ensuring growth
 - Enhancing risk assessment
 - Improving productivity and decision-making
 - More business opportunities
 - Digital banks – internet-based banks

Here is a list of data science use cases in banking area:

1. Fraud detection

Machine learning is effective detection and prevention of fraud include credit cards, accounting, insurance, and more. The key steps to fraud detection include: Obtaining data samplings for model estimation and preliminary testing Model estimation Testing stage and deployment.

The fraud detection need the deep theoretical knowledge into practical applications. Several algorithms are needed, such as association, clustering, forecasting, and classification. The simple example of efficient fraud detection is outlier data. When some unusually high transactions finded or it can investigate unusually high purchase of popular items and multiple accounts.

Palmer (2014) suggests that the following stages of enterprise counter fraud measures:

- **Detect:** By applying advanced analytics to all key fraud data to aid in predicting if an action is potentially fraudulent.
- **Respond:** Applying fraud insights to take action in real time.
- **Investigate:** By performing and managing inquiries into suspicious activity that are supported by data analysis and collaborative sophisticated case management.
- **Discover:** Using new big data analytics capabilities that help in the identification of suspicious activity by analyzing historical data to search for patterns of fraud and financial crimes.

2. Marketing

Xerago (2015) defines marketing analytics as the practice of measuring, managing and analyzing market performance to maximize the effectiveness of and return on investment (ROI) from the marketing activities. The marketing analytics will help banks through data to increase profitability. Pramanick (2013) states that banks are always at risk of losing customers and need strategies that are dependent in identifying the right action to the right customer. Thus banks need customer analytics to segment the customers. This marketing will assist in determining products and services, pricing, and strategy of the banks.

Morabito (2015) adds that big data enabled marketing automation will assist banks in servicing individual customer needs while keeping the marketing costs low, enabling a personalized experience at a good ROI.

3. Credit Risk Management

Sas (sas.com: one of leading analytics software company) defines credit risk as the probability of loss due to a borrower's failure to make payments on any type of debt. Better credit risk presents an opportunity and improve performance and also competitive advantage.

Due to sas institute (sas.com), credit risk have also challenges, such as:

- **Inefficient data management:** An inability to access the right data when it's needed causes problematic delays.
- **No groupwide risk modeling framework:** Without it, banks can't generate complex, meaningful risk measures and get a big picture of groupwide risk.
- **Constant rework:** Analysts can't change model parameters easily, which results in too much duplication of effort and negatively affects a bank's efficiency ratio.
- **Insufficient risk tools:** Without a robust risk solution, banks can't identify portfolio concentrations or re-grade portfolios often enough to effectively manage risk.
- **Cumbersome reporting:** Manual, spreadsheet-based reporting processes overburden analysts and IT.

The solution should include:

- Better model management that spans the entire modeling life cycle.
- Real-time scoring and limits monitoring.
- Robust stress-testing capabilities.
- Data visualization capabilities and business intelligence tools that get important information into the hands of those who need it, when they need it.

2.3.5 Healthcare

The industry has taken advantage of big data and analytics to make strategic business decisions. Healthcare big data refers to the quantities of data that is now available to healthcare providers. The big data in the healthcare industry is changing the way patients and doctors handle care. The more big data involved, the more efficient healthcare can be.

Benefits of Healthcare Big Data:

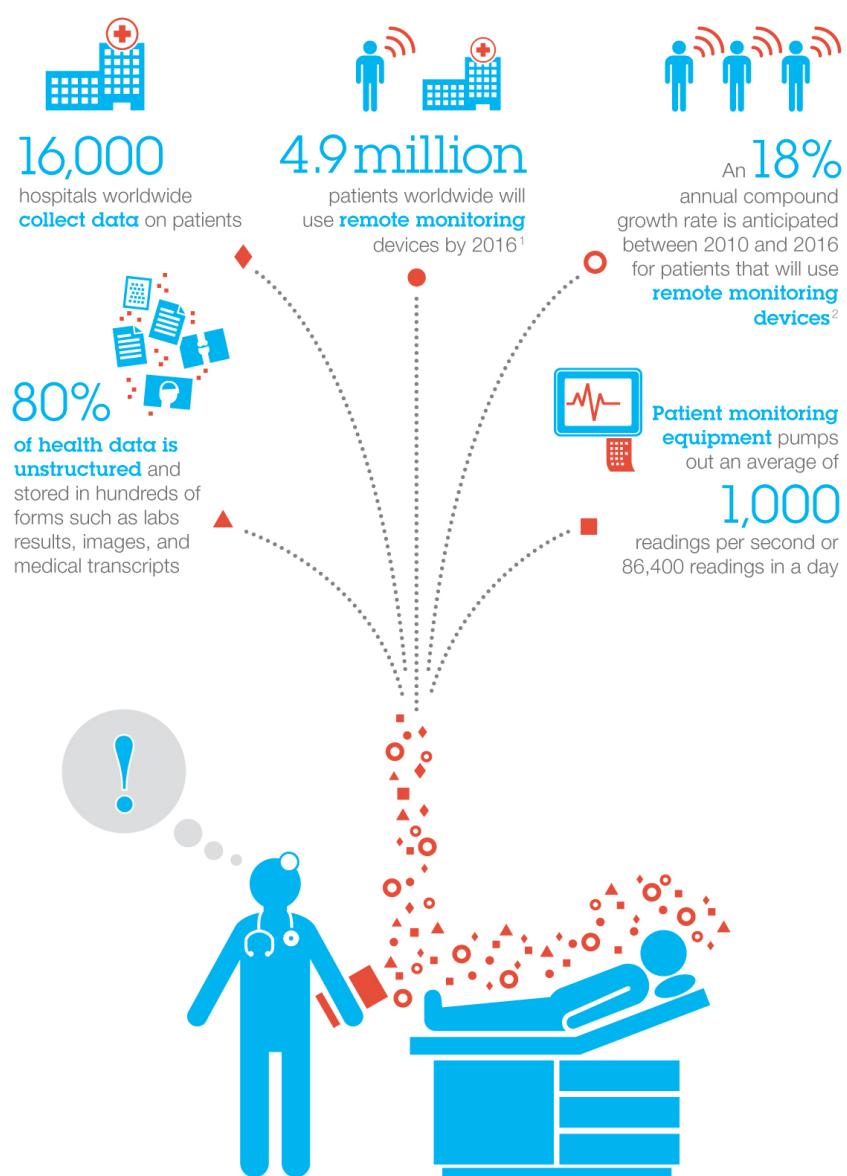
- Create holistic, 360-degree views of consumers, patients, and physicians.
- Improve care personalization and efficiency with comprehensive patient profiles.
- Inform physician relationship management efforts by tracking physician preferences, referrals, and clinical appointment data.
- Boost healthcare marketing efforts with information about consumer, patient, and physician needs and preferences.
- Analyze trends within a single hospital or greater a healthcare network to benefit research and care procedures for enhanced population health outcomes.
- Identify patterns in health outcomes, patient satisfaction, and hospital organization.
- Predict health outcomes and create preventive care strategies with data analysis.
- Optimize growth by improving care efficiency, effectiveness, and personalization.

Below are common questions about Healthcare Big Data:

- What is Driving the Adoption of Big Data in Healthcare?
 - Volume of Healthcare Data
 - Government Regulations
 - Desire for Personalized Care
- How Can Marketing Take Advantage of Healthcare Big Data?
 - Propensity models (score potential targets)

Big Data in Healthcare: Tapping New Insight to Save Lives

Healthcare is challenged by large amounts of data in motion that is diverse, unstructured and growing exponentially. Data constantly streams in through interconnected sensors, monitors and instruments in real-time faster than a physician or nurse can keep up.



The ability to analyze big data in motion in real-time as it streams in can help predict the onset of illness and respond instantly from new insight that will help transform healthcare.

Figure 2.31:

- Communication personalization (continue an ongoing relationship with the healthcare organization)
- Integrated communication (holistic customer experiences throughout the care continuum)
- What Challenges Arise with Healthcare Big Data?
 - Sorting and prioritizing of information
 - Ensuring that the right access to big data insights and analysis (data security issues)
- What is the Future of Healthcare Big Data?
 - Big data are more crucial for company success
 - Big data healthcare will also smarter and more integrated
 - Big data will grow with Internet of Things (IoT)
 - IoT will become standard methods

2.3.6 Crime & Terrorism

Some police are starting to use big data to predict crime. The use of fingerprints, DNA, ballistic analysis, CCTV and other types of technology have also played to improve big data analytics in crime. Top use case is predictive analytics. This helps in this aspect:

- Police can make compelling cases to get emergency resources to fight recent crime waves
- Police can identify the likelihood that they are dealing with serial offenders
- Police can look for precipitating factors that cause crime epidemics and pass that information along to policymakers to take preventive measures

The success story of crime prediction are come from RUSI (Royal United Services Institute for Defence and Security Studies). They published a paper “Big Data and Policing An Assessment of Law Enforcement Requirements, Expectations and Priorities”. One to the success project is **Predictive Crime Mapping**

How Predictive Crime Mapping Works:

Big data analytics has been used for a decade for predictive crime mapping, and the technology continues to be developed and advanced. On the basic level, this technology is very simple and that are based on:

- Crime type
- Date
- Time
- Crime locations

Police officers are printing out their predictive maps to use on patrol to combat crime in high mapping crime.

Issues Presented by Using Past Crime Data:

These big data solutions are also learning from the past, so the information can be discriminatory or bias as a result. Example:

100 crimes occurred in the area by males in the past two years. The same can be inputted for age, religion and race. So, it does present a system that may create a bias and reinforce this in the police force. If John Doe is convicted of rape, and all of his rapes occurred on High Street, algorithms may be able to suggest John Doe as a rapist if:

- Rapes occurred on High Street
- Rapes occurred at around the same time as his previous convictions
- Similar women were raped

Of course, it may not be John Doe, but this form of predictive modeling is an option with big data. It may be possible to scour records, looking for similar crimes with similar circumstances and tying them together.

Chapter 3

Big Data Analytics

3.1 Data Analytics Lifecycle

One of the most method for data analytics is **CRISP-DM methodology**. The method was found by five companies: SPSS, Teradata, Daimler AG, NCR Corporation, and OHRA (an insurance company). The CRISP-DM methodology stands for Cross Industry Standard Process for Data Mining. It is a robust and well-proven methodology. This model is an idealised iterative method. Take a look the following illustration.

Stage One- Determine Business Objectives

This initial phase focuses on understanding the project objectives and requirements from a business perspective. The goal of this stage is to uncover important factors for improving business. Neglecting this step can mean that a great deal of effort is put into producing the right answers to the wrong questions.

What are the desired outputs of the project?

1. **Set objectives** -It means describing the primary objective from a business perspective. For example, your primary goal might be to keep current customers by predicting when they are prone to move to a competitor. The questions are “Does the channel used affect whether customers stay or go?” or “Will lower ATM fees significantly reduce the number of high-value customers who leave?”.
2. **Produce project plan** -It describe the plan for achieving business needs (improving sales, reduce cost, ect). This step include the initial selection of tools and techniques.
3. **Business success criteria** -The successful criteria from the business point of view. The criteria should be specific and measurable, such as cost reduction or significant sales growth.

Assess the current situation

This involves more details about another factors that company will need when determining goals.

1. **Inventory of resources** : The resources available to determining business objects including:
 - Personnel (business experts, data experts, technical support, data mining experts)
 - Data (fixed extracts, access to live, warehoused, or operational data)
 - Computing resources (hardware platforms and software)
2. **Requirements, assumptions and constraints** : The list of requirements objects that company need. From scheduling, quality parameter, data security, and legal issues. Including the assumption of the the factors such as verification data and also constraints like availability of resources.
3. **Risks and contingencies** : The list of delay project and fail probability. what company take if the risk take place?.
4. **Terminology** -A glossary of relevant business and data mining terminology.

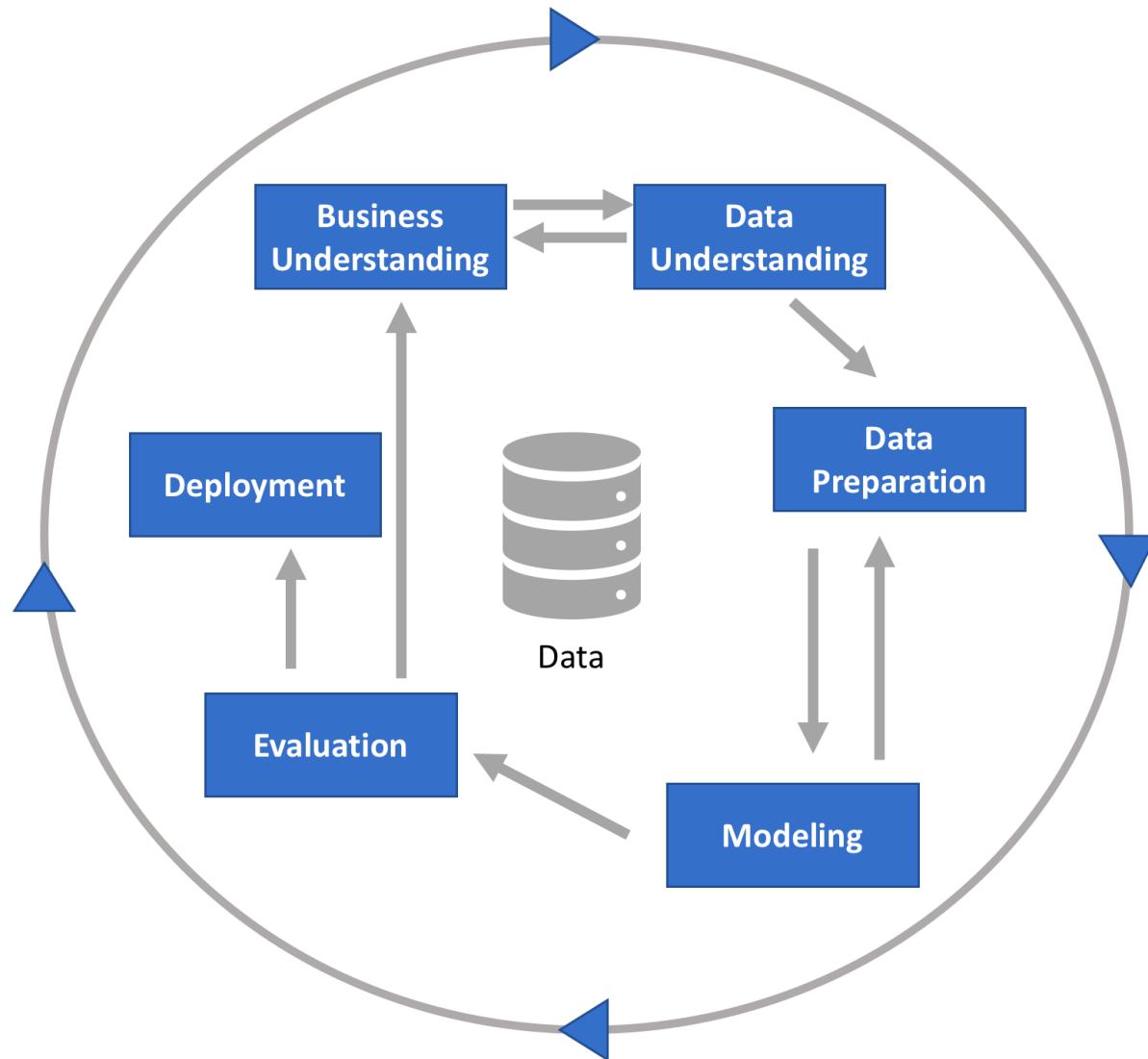


Figure 3.1:

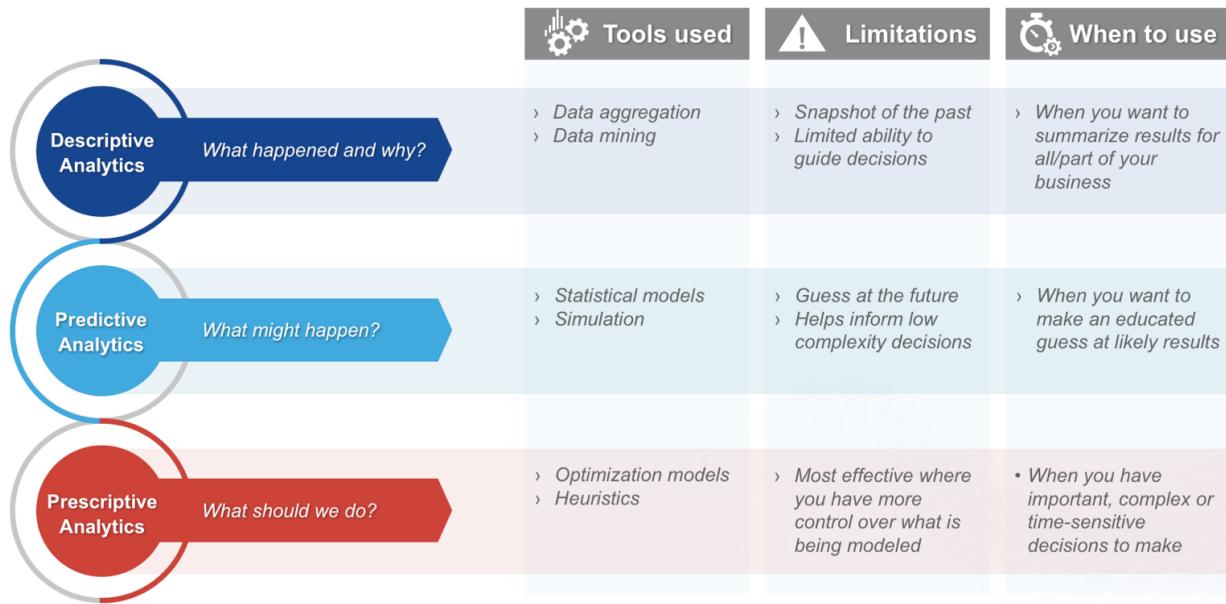


Figure 3.2:

5. Costs and benefits

Determine data analytics goals

Produce the plan

3.2 Modern Data Analytics

3.3 Analytics Paradigm

Analytics helps company answer the business questions. What happened in the past, what happen in the future, and what decisions company can make to improving future performance. Nowadays, data analytics is very important and more and more universities and further education programmes are creating analytics learning environments. Tom Davenport from CIO (Chief Information Officer) Magazine divided analytics area into three main types of analytics. The three main components of analytics is as follows:

Descriptive Analytics:

Descriptive Analytics helps company to get insight from the past and current state of business. Most every business function in the company (Production, Sales, Finance, etc.) used Descriptive Analytics to create custom reports.

Overview:

- Question being answered: What happened and why?
- Examples of tools used: Data aggregation and data mining
- Limitation to be aware of: Snapshot of the past, often with limited ability to help guide decisions
- When to use: When you want to summarize results for all or part of your business

Predictive Analytics:

Predictive Analytics predicts the business data in the future. It need applies statistical techniques (often machine learning) to get what the future may be happen.

Overview:

- Question being answered: What might happen?
- Examples of tools used: Machine learning, statistical models, and simulation
- Limitation to be aware of: Guess at the future, helps inform low complexity decisions
- When you want to use: When you want to make an educated guess at likely results

Prescriptive Analytics:

It optimized the set of decisions. Company can quickly evaluate trillions or more possible combinations of choices (for example: what products in which manufacturing, what product lines and in what quantities), minimum production of a given product, manufacturing time and cost, machine, raw material inventory, finished goods inventory capacity, and maximize or minimize your objectives (for example: total product costs).

Overview:

- Question being answered: What should we do?
- Examples of tools used: Optimization and heuristics
- Limitation to be aware of: Most effective where you have some control over what is being modeled
- When you want to use: When you have important, interdependent, complex or time-sensitive decisions to make

3.4 Deutsche Bahn Tools

3.4.1 SAP Business Object Web Intelligence

SAP BO Web Intelligence is databases system created by SAP. The reason for choosing this databases system is **robust software** (from a few user to tens of thousand of users) and **excellent service** from SAP company. This SAP system have some another advantages, such as:

- On-premise deployment (offline)
- Real-time business intelligence (offline and offline)
- Increased user autonomy
- Make information consumption simple, personalized, and dynamic

3.4.2 Open Refine

Openrefine is a data manipulation tool which cleans, reshapes and intelligently edit batch messy, and unstructured data. It is an open source tool and its code can be reused in other projects too. OpenRefine provides the flexibility to choose from a variety of data set functionalities, which makes it even more user friendly. Users can use this tool to get a big view of their data in terms of statistically curved graphs. They can play with messy data without worrying about risks.

What -A messy, unstructured, inconsistent dataset can be explored using open refine. In general, it will be very difficult to explore data through redundancies and inconsistencies. But, OpenRefine gives several functions through which one can filter the data, edit the inconsistencies, and view the data. It's a tool to clean the data.

Why - Spreadsheets can also refine a dataset but they are not the best tool for it as Openrefine cleans data in a more systematic controlled manner. While using historical data, we come across issues like blank fields, duplicate records, inconsistent formats and using Openrefine tool can help to resolve such issues.

When -Now data analysis play an important role in business. Data analysts improve decision making, cut costs and identify new business opportunities. Analysis of data is a process of inspecting, cleaning, transforming, and modelling data with the goal of discovering useful information, suggesting conclusions, and supporting decision making. So, to ensure the accuracy of our analysis, we have to clean our data.

Why OpenRefine is a better tool for data cleaning and modeling:

| Google refine | Spreadsheets | Databases |
|--|---|---|
| Batch editing of rows and columns possible | Editing of one cell at a time | Schema and programming language required for editing |
| Used for exploring and transforming data | Used for entering data and performing calculations, functions | Data is out of sight unless script is run to view it. |
| No Schema Required | No Schema Required | |
| Data is always visible at each step of editing | Data is always visible | |
| More interactive and visual | Visual is not impressive. | |

3.4.3 R Programming

R language is an open source program. It maintained by the R core volunteer developers from across the globe. R is a language and environment for statistical computing and graphics. It is a GNU project which is similar to the S language and environment which was developed at Bell Laboratories (formerly AT&T, now Lucent Technologies) by John Chambers and colleagues, founded by Ross Ihaka and Robert Gentleman.

Key points of R programming:

- R is open source.
- R programming language is best for statistical, data analysis and machine learning.
- R, SAS, and SPSS are three statistical languages. Only R is open source
- It supports procedural programming with functions and object-oriented programming.
- Packages are part of R programming. Thus, they are useful in collecting sets of R functions into a single unit.
- R have database input, exporting data, viewing data, variable labels, missing data, etc.
- R is an interpreted language. Hence, we can access it through command line interpreter.
- R supports matrix arithmetic.
- It has effective data handling and storage facilities.
- R supports a large pool of operators for performing operations on arrays and matrices.
- It has facilities to print the reports for the analysis performed in the form of graphs either on-screen or on hardcopy.

Comparison Of R With Other Technologies:

- Data handling Capabilities – Good data handling capabilities and options for parallel computation.
- Availability / Cost – R is an open source and we can use it anywhere.
- Advancement in Tool – If you are working on latest technologies, R gets latest features.
- Ease of Learning – R has a learning curve. R is a low-level programming language. As a result, simple procedures can take long codes.
- Job Scenario – It is a better option for start-ups and companies looking for cost efficiency.
- Graphical capabilities – R is having the most advanced graphical capabilities. Hence, it provides you with advanced graphical capabilities.
- Customer Service support and community – R is the biggest online growing community.

More info about R: <https://www.r-project.org/>

3.4.4 RStudio

RStudio is a free and open-source integrated development environment (IDE) for R. RStudio was founded by JJ Allaire, creator of the programming language ColdFusion. RStudio is available in open source and commercial editions and runs on the desktop (Windows, macOS, and Linux) or in a browser connected to RStudio Server or RStudio Server Pro (Debian, Ubuntu, Red Hat Linux, CentOS, openSUSE and SLES). ([wikipedia.com](https://en.wikipedia.org/wiki/RStudio))

RStudio IDE features:

- An IDE that was built just for R
 - Syntax highlighting, code completion, and smart indentation
 - Execute R code directly from the source editor
 - Quickly jump to function definitions
- Bring your workflow together
 - Integrated R help and documentation
 - Easily manage multiple working directories using projects
 - Workspace browser and data viewer
- Powerful authoring & Debugging
 - Interactive debugger to diagnose and fix errors quickly
 - Extensive package development tools
 - Authoring with Sweave and R Markdown

More info about RStudio: <https://www.rstudio.com/>

3.4.5 Microsoft Power BI

1. What Is Power BI?
 2. Why Power BI?
 3. Who Use Power BI?
 4. Components Of Power BI
-

1. What Is Power BI?

Power BI is a Data Visualization and Business Intelligence tool that converts data from different data sources to interactive dashboards and reports. For making best business decisions, it is important to get relevant information from the data sources and present it as visualization. Power BI has several products, such as:

- Power BI Desktop: Create rich, interactive reports with visual analytics at your fingertips—for free in desktop application.
- Power BI Pro: Connect to hundreds of data sources, and visualize all your data with live dashboards and reports. Then share insights across your organization to fuel intelligent action.
- Power BI Premium: Power BI Premium offers advanced, self-service data preparation that allows every user—from business analyst to data scientist—to accelerate the delivery of insights and collaborate with ease.
- Power BI Mobile: Stay connected to your data wherever business takes you. Mobile business intelligence is just a touch away.
- Power BI Embedded: Embed interactive Power BI visuals in your applications, websites, or portals to bring world-class analytics directly to your customers
- Power BI Report Server: Power BI Report Server is the on-premises solution for reporting today, with the flexibility to move to the cloud tomorrow. It's included with Power BI Premium so you have the ability to move to the cloud on your terms.

For more info: <https://powerbi.microsoft.com/en-us/>

2. Why Power BI?

- Power BI is built on the convention of best BI products available – SQL Server Analysis Services (SSAS) and Microsoft Excel. Power BI gives you the option to connect with these products (heavy users Microsoft products).
- Power BI is Being built/rebuilt using the latest technologies like HTML 5.0, cloud computing, column store databases & smartphone mobile apps.
- Microsoft has opened the custom visuals gallery for open source contributions which adds value to the community (from users and another company as adds in).
 - <https://community.powerbi.com/t5/Data-Stories-Gallery/bd-p/DataStoriesGallery>
 - <https://appsource.microsoft.com/en-us/marketplace/apps?product=power-bi-visuals>

- Trend towards self-service business intelligence indicates Microsoft's leading position in this space. Based on Gartner Research (the world's leading research and advisory company) Power BI get high position.

3. Who Use Power BI?

- Developers
- IT Professional
- Subject Experts (Financial, Insurance, IT Industry, etc)
- Business Analyst
- Data Analyst
- Data Scientist
- Business Intelligence Developer

4. Components Of Power BI

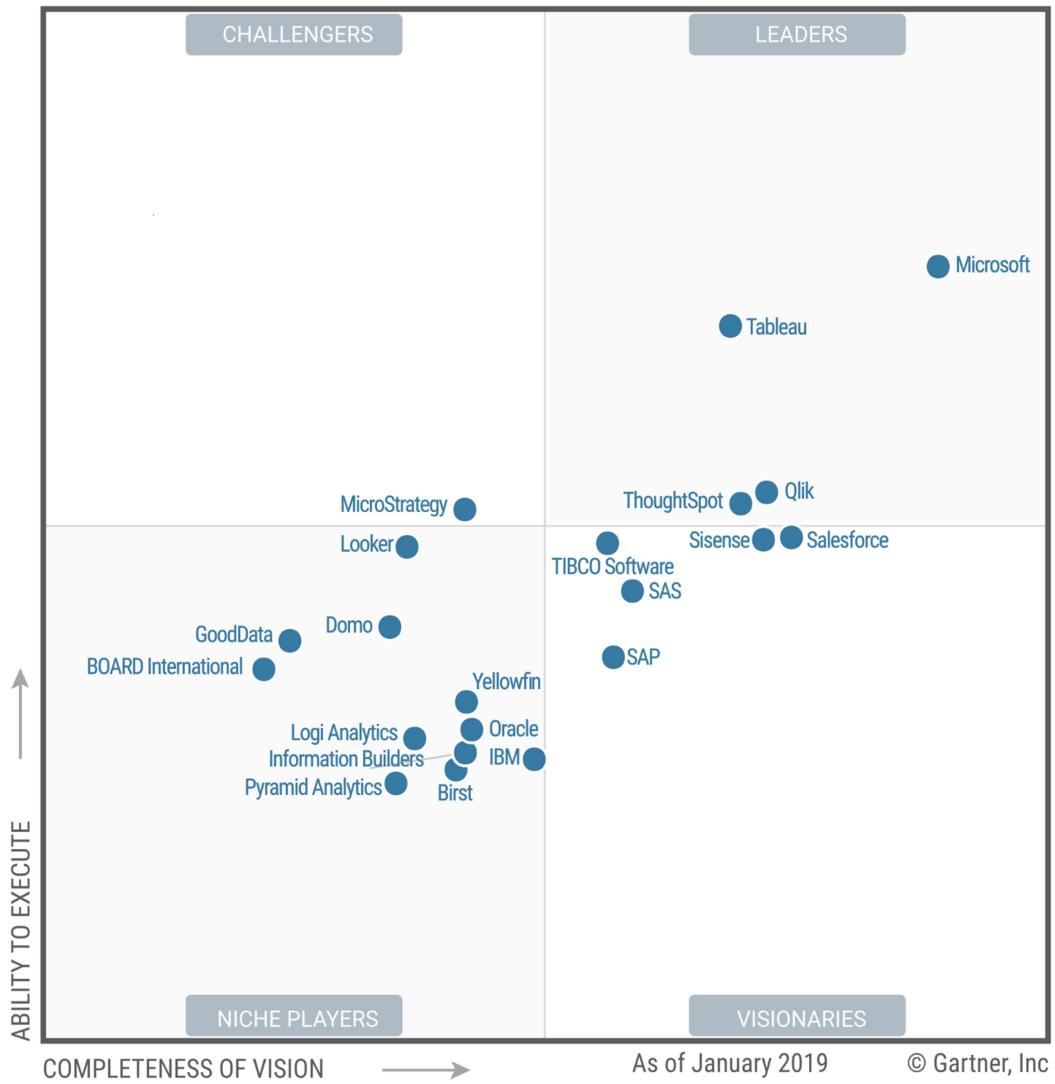
- Power Query: ETL (Extract, Transform, Loading) data from different sources. Power Query Data Sources are:
 - Web page
 - Excel or CSV file
 - XML file
 - Text file
 - Folder
 - SQL Server database
 - Microsoft Azure SQL Database
 - Access database
 - Oracle database
 - IBM DB2 database
 - MySQL database
 - PostgreSQL Database
 - Sybase Database
 - Teradata Database
 - SharePoint List
 - OData feed
 - Microsoft Azure Marketplace
 - Hadoop File (HDFS)
 - Microsoft Azure HDInsight
 - Microsoft Azure Table Storage
 - Active Directory
 - Microsoft Exchange
 - Facebook

More info about Power Query: <https://docs.microsoft.com/de-de/powerquery-m/power-query-m-reference>

- Power Pivot: Used in data modeling strategy (Snow flakes, Relational, Star Schema, etc). Power Pivot have DAX (Data Analysis Expressions) as the function programming to help create data model from different sources. DAX is a collection of functions, operators, and constants that can be used in a formula, or expression, to calculate and return one or more values.

More info about Power Pivot and DAX:

- <https://docs.microsoft.com/en-us/power-bi/desktop-quickstart-learn-dax-basics>
- <https://docs.microsoft.com/en-us/dax/dax-function-reference>
- Power View: Analyze, visualize and display data as an interactive data visualization.
- Power Map: Interactive geographical visualization.
- Power BI Service: Share data visualization with web browser (powerbi.com)



Source: Gartner (February 2019)

Figure 3.3:

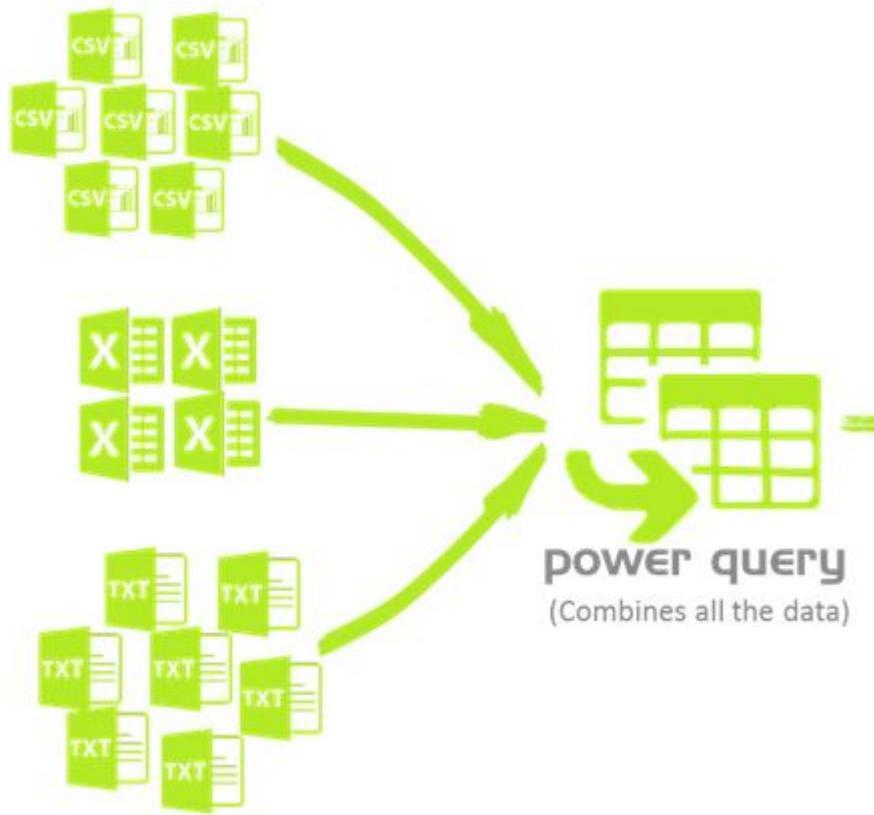


Figure 3.4:

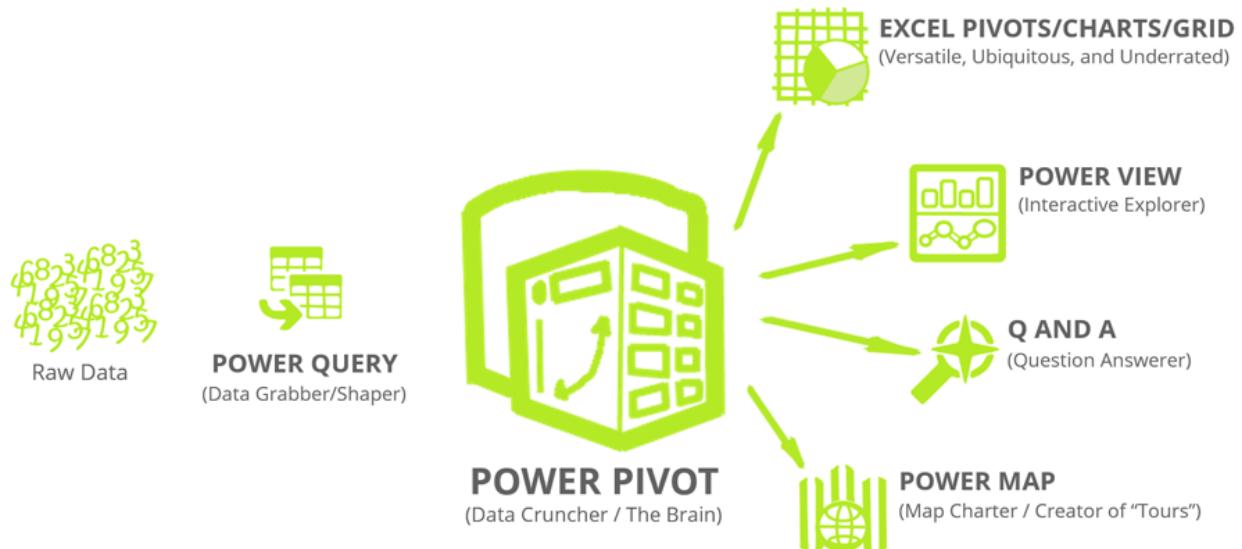


Figure 3.5:

- Power BI Q&A: Ask questions and get immediate answers with natural language query.
- Cortana for Power BI: Cortana is a voice virtual assistant created by Microsoft for Power BI.

3.4.6 Confluence (Software)

Confluence is a documentation collaboration software program developed and published by Australian software company Atlassian (atlassian.com). Confluence is content collaboration tool used to help company to collaborate and share knowledge efficiently. With Confluence, teams can create pages and blogs which can be commented on and edited by all members of the team. It is like microsoft word with additional features. For mor info: <https://www.atlassian.com/software/confluence>

Confluence have about 180 features as documentation cloud software. Here the complete list: <https://info.seibert-media.net/display/Atlassian/Overview+of+all+180+Confluence+features>

3.4.7 Github

Github is open source version control. Version control is a system that records changes to a file or set of files over time so that you can recall specific versions later. See the github site: <https://github.com/>

Key Points of Github as Distributed Version Control:

- Keep tracks of changes over time
- Allows the progress and projects to track
- Allows company to revert to earliner versions
- It easier to collaborate with teams
- Track changes on various files (more than one file)
- Track changes on a directory (Track ID)
- Allows savings non-text files (i.e. images, sheet file, ect)
- Various teams working on the same file
- Use of remote repositories to collaborate
-

DB use Github with RStudio. For more info about how to using Github with R: <https://happygitwithr.com/>

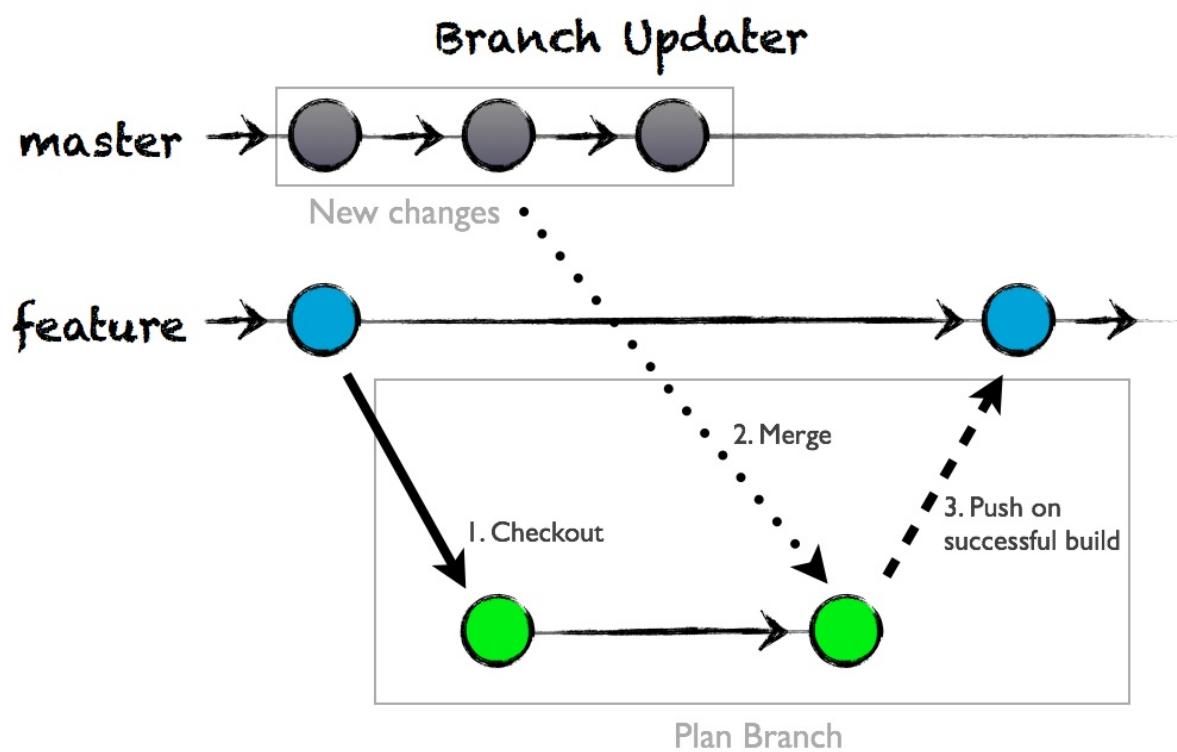


Figure 3.6:

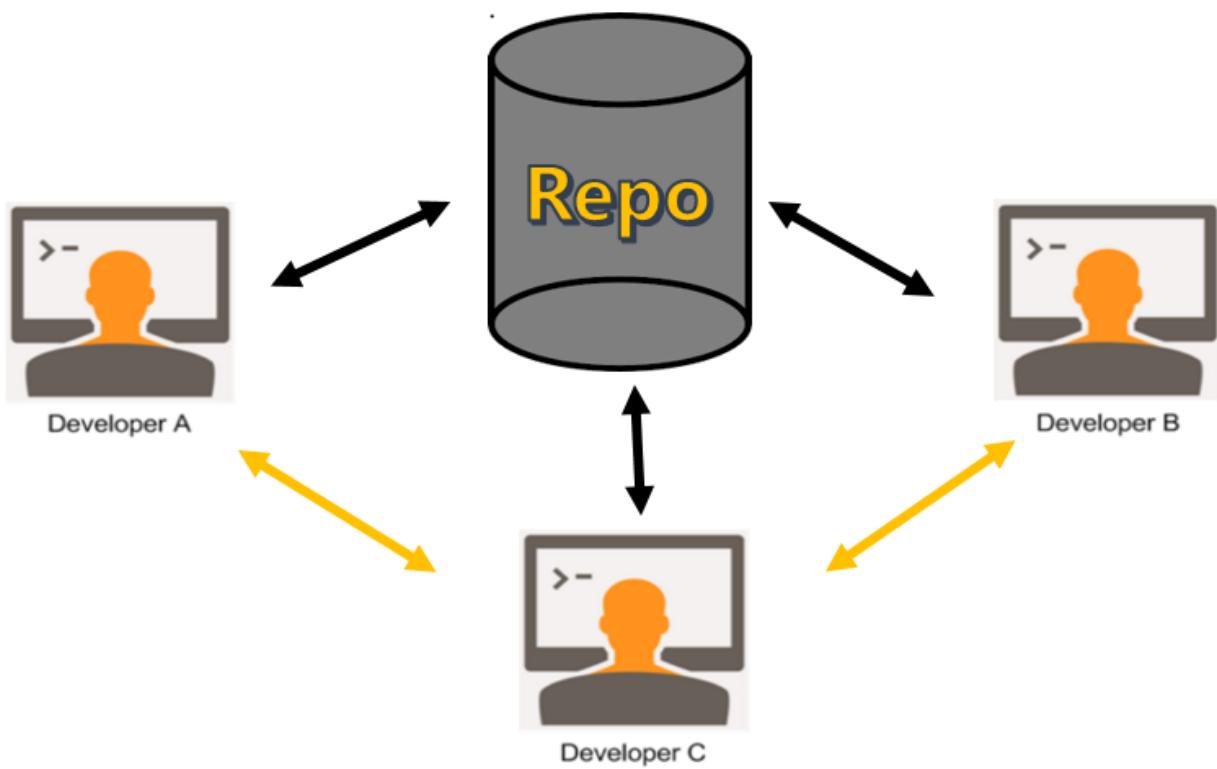


Figure 3.7:

Chapter 4

Terminology

4.1 Balance Sheet

TeX is the best way to typeset mathematics. Donald Knuth designed TeX when he got frustrated at how long it was taking the typesetters to finish his book, which contained a lot of mathematics. One nice feature of *R Markdown* is its ability to read LaTeX code directly.

If you are doing a thesis that will involve lots of math, you will want to read the following section which has been commented out. If you're not going to use math, skip over or delete this next commented section.

4.2 Data Cleaning

Chemical formulas will look best if they are not italicized. Get around math mode's automatic italicizing in LaTeX by using the argument `\mathrm{formula here}`, with your formula inside the curly brackets. (Notice the use of the backticks here which enclose text that acts as code.)

So, $\text{Fe}_2^{2+}\text{Cr}_2\text{O}_4$ is written `\mathrm{Fe_2^{2+}Cr_20_4}`.

Exponent or Superscript: O^-

Subscript: CH_4

To stack numbers or letters as in Fe_2^{2+} , the subscript is defined first, and then the superscript is defined.

Bullet: $\text{CuCl} \bullet 7\text{H}_2\text{O}$

Delta: Δ

Reaction Arrows: \longrightarrow or $\xrightarrow{\text{solution}}$

Resonance Arrows: \leftrightarrow

Reversible Reaction Arrows: \rightleftharpoons

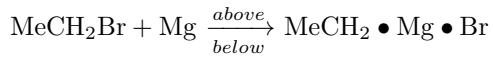
4.2.1 Information Technology Terms

You may wish to put your reaction in an equation environment, which means that LaTeX will place the reaction where it fits and will number the equations for you.



We can reference this combustion of glucose reaction via Equation (4.1).

4.2.2 Other examples of reactions



4.3 Data Model

Many of the symbols you will need can be found on the math page <http://web.reed.edu/cis/help/latex/math.html> and the Comprehensive LaTeX Symbol Guide (<http://mirror.utexas.edu/ctan/info/symbols/comprehensive/symbols-letter.pdf>).

4.4 Data Analysis Expression (DAX)

You will probably find the resources at <http://www.lecb.ncifcrf.gov/~toms/latex.html> helpful, particularly the links to bsts for various journals. You may also be interested in TeXShade for nucleotide typesetting (<http://homepages.uni-tuebingen.de/beitz/txe.html>). Be sure to read the proceeding chapter on graphics and tables.

Chapter 5

KPIs for big data

What is KPI?

KPI (Key Performance Indicators) are the response to company fear of big data, ugly spreadsheets and uncertain applications with unstructured data. The idea of KPI is that company presenting big data easily and also using business relevant language.

Key Performance Indicators are:

- Use rates, ratios, percentages and averages instead of raw numbers
- Leverage tachometers and thermometers and stoplights instead of pie charts and bar graphs
- Provide temporal context and highlight change instead of presenting tables of data
- Drive business-critical action

Measuring the right KPI is vital to the health and success of the company. Here are 5 reasons why You need KPIs.

1. Monitor company health
2. Measuring progress over time
3. To make adjustments & stay on track
4. Solving problems and getting more opportunities
5. Analyzing patterns over time

Examples of KPIs are:

- Net Profit Margin
- ROI (Return of Investment)
- Operating Cash Flow (OCF)
- Sales Growth
- BEP (Break Event Point)
- Cash Ratio

Types of KPI

Depending on your organization's goals and company objectives, they can track various Key Performance Indicators. The right KPIs right is crucial for getting actionable and insightful information about the company's performance and situation.

Each business department measures different KPIs because they all have different tasks and goals.

There are five main types of KPIs:

1. Business KPI
2. Financial KPI

3. Sales KPI
4. Marketing KPI
5. Project Management KPI
6. Business KPI

Business KPIs help to measure long-term business goals. By tracking business KPI, companies are able to understand important business processes and identify better decision.

Examples of popular business KPIs are:

- Revenue Growth Rate
- Churn Rate
- Acquisition Rate
- Relative Market Share
- Return on Equity

2. Financial KPI

This KPI provide an assessment of business performance. It is usually used by an organization's leader and financial department. Financial KPI indicate how good a company is doing in terms of generating revenue and profits.

Examples of popular financial KPIs are:

- MRR (Monthly Recurring Revenue)
- Net profit margin
- Operating cash flow (OCF)
- Working Capital
- Current Ratio
- Budget Variance

3. Sales KPI

Sales KPIs are measurable values that indicates the performance of various sales processes, used by the sales team to monitor the achievement of their key objectives and goals. Sales metrics help to keep and attain sustainable sales.

Example of Popular sales KPI are:

- Monthly New Leads/Sales
- Lead-to-Customer Conversion Rate
- Cost per Acquisition
- Sales Qualified Leads (SQL)
- Customer Lifetime Value (LTV)

4. Marketing KPI

Marketing KPIs help marketing department to monitor their success across all marketing channels. A quick overview of marketing metrics shows how well the marketing team's doing in terms of acquiring new leads.

Examples of popular marketing KPI are:

- Website Traffic Per Source
- Cost Per Acquisition CPA
- Marketing Qualified Leads (MQL)
- Net Promoter Score
- Conversion Rate

5. Project Management KPI

A project managers used this KPI to monitor project progress. Company use project KPI to identify successful projects and meet important deadlines.

Popular Project management KPI are:

- Planned Value (PV)
- Actual Cost (AC)
- Earned Value (EV)
- Cost Variance (CV)
- Schedule Variance (SV)

Before we talks about use case in Deutsche Bahn. We need to understand the types and classification of KPI terms based on perspektive, industry, and etc.

High level vs Low level KPI

High-level KPI demonstrate the company's overall performance. Examples of high-level KPIs include Annual Growth, Annual Recurring Revenue (ARR), and Relative Market Share.

Single individuals have no impact on these KPI. They're the result of teamwork across multiple departments and subsidiary company.

Low-level KPI indicate the performance of specific departments or individuals. Low-level business metrics are tied to people's day-to-day work and more actionable.

Five essential questions for creating a KPI

1. What are the business results (goals)?
2. How can the KPI values be improved by taking action?
3. Do we have all the relevant data?
4. Who is going to use the KPI?
5. How to visualize specific KPI (graphs, metrics, diagrams, etc.)?

How to choose the right KPI

The most important thing is define a business goals. It is better to focus on a few key metrics instead of many irrelevant ones. Ensure that every single one of your business metrics meets the SMART criteria:

SMART KPIs are:

Specific

Measurable

Attainable

Relevant

Time-Bound

Note

- **Understand that Key Performance Indicators** are different for each industry, growth stage, and project phase.
- **A KPI (Key Performance Indicator)** is a measurable value that indicates whether a team/company is reaching its targets (benchmarks).
- **The five main types of KPIs** are business KPIs, financial KPIs, sales KPIs, marketing KPIs, and project management KPIs.
- **KPIs are frequently monitored** with a real-time reporting tool – KPI dashboard.
- **Each KPI your monitor should meet the SMART criteria**, i.e. be Specific, Measurable, Attainable, Relevant, and Time-Bound.
- **Only track the metrics that are relevant** to your organization and business goals.

5.1 Financial Perspective

Based on the size, age, and industry, each and every company needs to be conscious of their financial situation. The fastest and most efficient way to keep track of a company business performance is to set up a KPI that displays financial concern. Let's started with the most widely used financial metrics that have the full spectrum of important budget.

Firstly, we need to understand the concept of balance sheet. Below is the example of complete balance sheet.

5.1.1 Operating Cash Flow (OCF)

How much money generated by the regular operating activities of a business in a specific time period.

Formula (short form): Operating Cash Flow = Net Income + Non-Cash Expenses – Increase in Working Capital

Formula (long form) : Operating Cash Flow = Net Income + Depreciation + Stock Based Compensation + Deferred Tax + Other Non Cash Items – Increase in Accounts Receivable – Increase in Inventory + Increase in Accounts Payable + Increase in Accrued Expenses + Increase in Deferred Revenue

Below is an example of operating cash flow (OCF) using Amazon's 2017 annual report.

Let's analyze how the operating section counts:

- Net income from the bottom of the income statement is used as the starting point
- All non-cash items are “added back” meaning any accruals are reversed, including:
 - Depreciation, which an accounting method for expensing property, plant, and equipment (PP&E) purchases
 - Stock-based compensation is not paid out with actual cash, but instead with the issuance of shares
 - Other expense/income could include various items such as unrealized gains or losses or accrued items
 - Deferred taxes arise from the difference between accounting methods companies use when filing their taxes vs their financial statements

5.1.2 Current Ratio

The ability to pay all the financial obligations in one year. A healthy Current Ratio is between 1.5 and 3

Formula: Current Ratio = Current Assets / Current Liabilities

Example of the Current Ratio Formula:

A business has:

- Cash = \$15 million
- Marketable securities = \$20 million
- Inventory = \$25 million
- Short-term Debt = \$15 million
- Accounts Payables = \$15 million

then,

- Current assets = $15 + 20 + 25 = 60$ million
- Current liabilities = $15 + 15 = 30$ million
- Current ratio = $60 \text{ million} / 30 \text{ million} = 2.0x$

The company can easily settle each accounts payable twice.

| Example Company Balance Sheet December 31, 2018 | | | |
|--|-------------------|--|-------------------|
| ASSETS | | LIABILITIES | |
| Current assets | | Current liabilities | |
| Cash | \$ 2,100 | Notes payable | \$ 5,000 |
| Petty cash | 100 | Accounts payable | 35,900 |
| Temporary investments | 10,000 | Wages payable | 8,500 |
| Accounts receivable - net | 40,500 | Interest payable | 2,900 |
| Inventory | 31,000 | Taxes payable | 6,100 |
| Supplies | 3,800 | Warranty liability | 1,100 |
| Prepaid insurance | 1,500 | Unearned revenues | 1,500 |
| Total current assets | <u>89,000</u> | Total current liabilities | <u>61,000</u> |
| Investments | <u>36,000</u> | Long-term liabilities | |
| Property, plant & equipment | | Notes payable | 20,000 |
| Land | 5,500 | Bonds payable | 400,000 |
| Land improvements | 6,500 | Total long-term liabilities | <u>420,000</u> |
| Buildings | 180,000 | | |
| Equipment | 201,000 | Total liabilities | <u>481,000</u> |
| Less: accum depreciation | (56,000) | | |
| Prop, plant & equip - net | <u>337,000</u> | | |
| Intangible assets | | STOCKHOLDERS' EQUITY | |
| Goodwill | 105,000 | Common stock | 110,000 |
| Trade names | 200,000 | Retained earnings | 220,000 |
| Total intangible assets | <u>305,000</u> | Accum other comprehensive income | 9,000 |
| Other assets | <u>3,000</u> | Less: Treasury stock | (50,000) |
| Total assets | <u>\$ 770,000</u> | Total stockholders' equity | <u>289,000</u> |
| | | Total liabilities & stockholders' equity | <u>\$ 770,000</u> |

The notes to the sample balance sheet have been omitted.

Figure 5.1:

| AMAZON.COM, INC. | | | |
|--|--------------------------------|-------------|-------------|
| CONSOLIDATED STATEMENTS OF CASH FLOWS | | | |
| (in millions) | | | |
| | Year Ended December 31, | 2015 | 2016 |
| | | 2016 | 2017 |
| CASH AND CASH EQUIVALENTS, BEGINNING OF PERIOD | | \$ 14,557 | \$ 15,890 |
| OPERATING ACTIVITIES: | | | \$ 19,334 |
| Net income | | 596 | 2,371 |
| Adjustments to reconcile net income to net cash from operating activities: | | | |
| Depreciation of property and equipment, including internal-use software and website development, and other amortization, including capitalized content costs | | 6,281 | 8,116 |
| Stock-based compensation | | 2,119 | 2,975 |
| Other operating expense, net | | 155 | 160 |
| Other expense (income), net | | 250 | (20) |
| Deferred income taxes | | 81 | (246) |
| Changes in operating assets and liabilities: | | | |
| Inventories | | (2,187) | (1,426) |
| Accounts receivable, net and other | | (1,755) | (3,367) |
| Accounts payable | | 4,294 | 5,030 |
| Accrued expenses and other | | 913 | 1,724 |
| Unearned revenue | | 1,292 | 1,955 |
| Net cash provided by (used in) operating activities | | 12,039 | 17,272 |
| INVESTING ACTIVITIES: | | | 18,434 |
| Purchases of property and equipment, including internal-use software and website development | | (5,387) | (7,804) |
| Proceeds from property and equipment incentives | | 798 | 1,067 |
| Acquisitions, net of cash acquired, and other | | (795) | (116) |
| Sales and maturities of marketable securities | | 3,025 | 4,733 |
| Purchases of marketable securities | | (4,091) | (7,756) |
| Net cash provided by (used in) investing activities | | (6,450) | (9,876) |
| FINANCING ACTIVITIES: | | | (27,819) |
| Proceeds from long-term debt and other | | 353 | 621 |
| Repayments of long-term debt and other | | (1,652) | (354) |
| Principal repayments of capital lease obligations | | (2,462) | (3,860) |
| Principal repayments of finance lease obligations | | (121) | (147) |
| Net cash provided by (used in) financing activities | | (3,882) | (3,740) |
| Foreign currency effect on cash and cash equivalents | | (374) | (212) |
| Net increase (decrease) in cash and cash equivalents | | 1,333 | 3,444 |
| CASH AND CASH EQUIVALENTS, END OF PERIOD | | \$ 15,890 | \$ 19,334 |
| | | | \$ 20,522 |

Figure 5.2:

5.1.3 Ratio / Acid Test

Indicates whether a business has sufficient short-term assets to cover its near-future liabilities. The quick ratio deleted liquid assets such as inventories.

A common alternative quick ratio formula is: (Current assets – Inventory) / Current Liabilities

Example:

| | | | |
|-----------------------|-------|---------------------------|------|
| Cash | \$60 | Accounts Payable | \$30 |
| Marketable Securities | \$10 | Expenses | \$20 |
| Account Receivables | \$40 | Notes Payable | \$5 |
| Inventory | \$50 | Long Term Debt | \$10 |
| Total Current Assets | \$160 | Total Current Liabilities | \$65 |

The Company quick ratio as follows:

$$(\$60,000 + \$10,000 + \$40,000) / \$65,000 = 1.7$$

This means that for every money of Company current liabilities, the firm has \$1.70 of very liquid assets to cover those immediate obligations.

5.1.4 Burn Rate

Burn rate is the amount of time it will take a company to exhaust the capital. Burn Rate is also in terms of cash burned per month, year , or quarterly.

Formula: Burn Rate= Cash / Expenses

There are two types of burn rates: net burn and gross burn.

- gross burn: the total amount of operating costs it include in expenses each month.
- net burn: the total amount of money a company loses each month.

Example:

A company spends \$5,000 monthly on office space, \$10,000 on monthly server costs and \$15,000 on salaries and wages for its engineers, its gross burn rate would be \$30,000.

However, if the company was already producing revenue, its net burn would be different. Even if the company operates at a loss, with revenues of \$20,000 a month and costs of goods sold (COGS) of \$10,000, it would still work to reduce its overall burn. In this scenario, the company's net burn would be \$20,000, derived as: \$20,000 - \$10,000 - \$30,000 = \$20,000.

5.1.5 Net Profit Margin

Indicates how efficient a company at generating profit compared to its revenue.

Formula: Net profit margin = net profit / revenue

Check this example:

Company ABC and DEF both operate in the same industry. Which company has a higher net profit margin?

| Company ABC | Income Statement | Company DEF | Income Statement |
|--------------------|------------------|--------------------|------------------|
| Revenue | \$100 | Revenue | \$225 |
| Cost of Goods Sold | \$20 ____- | Cost of Goods Sold | \$35 ____- |
| Gross Profit | \$80 | Gross Profit | \$195 |
| Operating Expenses | \$20 ____- | Operating Expenses | \$40 ____- |
| Operating Profit | \$60 | Operating Profit | \$150 |

| Company ABC | Income Statement | Company DEF | Income Statement |
|-----------------------|------------------|-----------------------|------------------|
| Interest Expense | \$5 __- | Interest Expense | \$10 __- |
| Earnings Before Taxes | \$55 | Earnings Before Taxes | \$140 |
| Tax Expense | \$25 __- | Tax Expense | \$60 __- |
| Net Income | \$30 | Net Income | \$80 |

Step 1 : Write out the formula

$$\text{Net Profit Margin} = \text{Net Profit}/\text{Revenue}$$

Step 2 : Calculate the net profit margin for each company

Company XYZ:

$$\text{Net Profit Margin} = \text{Net Profit}/\text{Revenue} = \$30/\$100 = 30\%$$

Company ABC:

$$\text{Net Profit Margin} = \text{Net Profit}/\text{Revenue} = \$80/\$225 = 35.56\%$$

Company ABC has a higher net profit margin.

5.1.6 Working Capital

The difference between a current assets (like cash and goods) and current liabilities (like debts or obligations).

Formula: Working Capital = Current Assets - Current Liabilities

Example:

| | | | |
|-----------------------|-------|---------------------------|------|
| Cash | \$60 | Accounts Payable | \$30 |
| Marketable Securities | \$10 | Expenses | \$20 |
| Account Receivables | \$40 | Notes Payable | \$5 |
| Inventory | \$50 | Long Term Debt | \$10 |
| Total Current Assets | \$160 | Total Current Liabilities | \$65 |

Using the formula, the working capital ist:

$$\$160 - \$65,000 = \$95$$

5.1.7 Current Accounts Receivable

How much money owed by its debtors. This account receivable helps to estimate the upcoming income.

Let's take a look the part of balance sheets.

| | | | |
|----------------------------|-------------|---------------------------|------|
| Cash | \$60 | Accounts Payable | \$30 |
| Marketable Securities | \$10 | Expenses | \$20 |
| Account Receivables | \$40 | Notes Payable | \$5 |
| Inventory | \$50 | Long Term Debt | \$10 |
| Total Current Assets | \$160 | Total Current Liabilities | \$65 |

The account receivable ist \$40.

5.1.8 Current Accounts Payable

How much money owed by its creditors (bank, suppliers, . This account receivable helps to estimate the upcoming expenses.

Let's take a look the part of balance sheets.

| | | | |
|-----------------------|-------|---------------------------|------|
| Cash | \$60 | Accounts Payable | \$30 |
| Marketable Securities | \$10 | Expenses | \$20 |
| Account Receivables | \$40 | Notes Payable | \$5 |
| Inventory | \$50 | Long Term Debt | \$10 |
| Total Current Assets | \$160 | Total Current Liabilities | \$65 |

The account payable ist \$5.

5.1.9 Inventory Turnover

- How efficiently a company sells and replaces its inventory during period of time (daily, monthly, or yearly).
- The bility to generate sales and quickly re-stock.

Formula: Inventory Turnover = Sales / Inventory or Inventory Turnover = Cost of Goods Sold / Average Inventory

Example:

Berlin's Paper Company sells office paper. During the current year, Berlin reported cost of goods sold on its income statement of \$1,000,000. Berlin's beginning inventory was \$3,000,000 and its ending inventory was \$4,000,000. Berlin's turnover is calculated like this:

$$\text{Inventory Turnover} = 1,000,000 / [(3,000,000 + 4,000,000) / 2] = 0,29$$

It means that Berlin Paper Company only sold a third (fast 30%) of its inventory during the year.

5.1.10 Budget Variance

A budget variance is a different between the predicted cost or revenue in a given account. (how projected budgets vary compared to actual budget totals). Optimistic forecasting or poor leadership decisions caused significant variance (big variance). This KPI used predictive analytics with mathematical and statistical methoden.

Formula: Predictived Cost - Actual Cost

Not only the cost but also sales, projects budget, ect.

5.1.11 Sales Growth

The growth of sales over a certain period (daily, weekly, quarterly, or yearly).

Formula: Sales Growth = (Current Period Net Sales - Prior Period Net Sales) / Prior Period Net Sales * 100

Example:

Sales 2018: \$10 Mio Sales 2017: \$5 Mio

So, the sales growth is $(10-5)/5 \cdot 100\% = 100\%$

A Sales hat increased by 100% than last years.

5.1.12 Days Sales Outstanding (DSO)

The average number of days required for clients to pay a company. If the DSO is lower, company can focus ordering additional supplies.

Formula: $DSO = [\text{Accounts Receivable} / \text{Net Credit Sales}] * 365$

- Accounts Receivable: \$25,000
- Net Credit Sales: \$200,000

So, the DSO is $[25/200] * 365 = 45.63$

It takes 46 days to collect cash from the customers.

5.1.13 Payment Error Rate

Uncompleted payments due to a lack of approval, poor documentation or a missing reference.

5.1.14 Complete List

- Profit Indicators
 - Earnings before Taxes (EBT)
 - Earnings before Interest and Taxes (EBIT)
 - Earnings before Interest, Taxes and Amortization (EBITA)
 - Profit or Loss from Ordinary Business Operations
 - Profit or Loss from Extraordinary Operations
 - Operating income from Ordinary Business
 - Non-operating income from Ordinary Business
 - Result from Discontinued Operations
 - Non-periodic Income
 - Net Operating Profit After Taxes (NOPAT)
- Profitability Indicators
 - EBIT-Turnover-Yield
 - Return On Sales (ROS)
 - Return On Equity (ROE)
 - Return On Assets (ROA)
 - Earnings Per Share (EPS)
 - Return On Investment (ROI)
 - Return On Invested Capital (ROIC)
 - Return On Capital Employed (ROCE)
 - Return On Net Assets (RONA)
 - Risk Adjusted Return On Capital (RAROC)
 - Cost-Income Ratio (CIR)
- Liquidity Indicators
 - Cash Ratio
 - Quick Ratio
 - Current Ratio
 - Working Capital
 - Cash-burn Rate
- Tests of Solvency

- Debt-to-Equity Ratio
- Debt-to-Cash Ratio
- Interest Coverage Ratio
- Cash Flow Measures
 - Cash Flow
 - Gross or Net Cash Flow
 - Free Cash Flow (FCF)
 - Operating Cash Flow (OCF)
 - Earnings before Interest, Taxes, Depreciation and Amortization (EBITDA)
- Cash Flow Ratios
 - Cash Flow Return on Investment (CFROI)
 - Cash Flow Return On Equity (CFROE)
 - EBITDA-Turnover-Yield
 - Income-Tax Burden Ratio
- Financial Structure Indicators
 - Equity-To-Fixed-Assets Ratio (Level I)
 - Equity-To-Fixed-Assets Ratio (Level II)
 - Equity-To-Fixed-Assets Ratio (Level III)
 - Equity Ratio
 - Financial Leverage Index
- Efficiency Ratios
 - Average Collection Period
 - Average Payment Period
 - Cash-to-Cash Cycle
 - Asset Turnover Ratio
 - Asset Coverage Period
- Value Based Management (VBM)
 - Cash Value Added (CVA)
 - Economic Profit (EP)
 - Economic Value Added (EVA)
 - Weighted Average Cost of Capital (WACC)
- Capital Market Tests
 - Market-to-Book Ratio
 - Stock Yield
 - Dividend Yield
 - Price-Earnings Ratio (P/E Ratio)
 - Price-To-Cash Flow Ratio
 - Cash Flow per Share

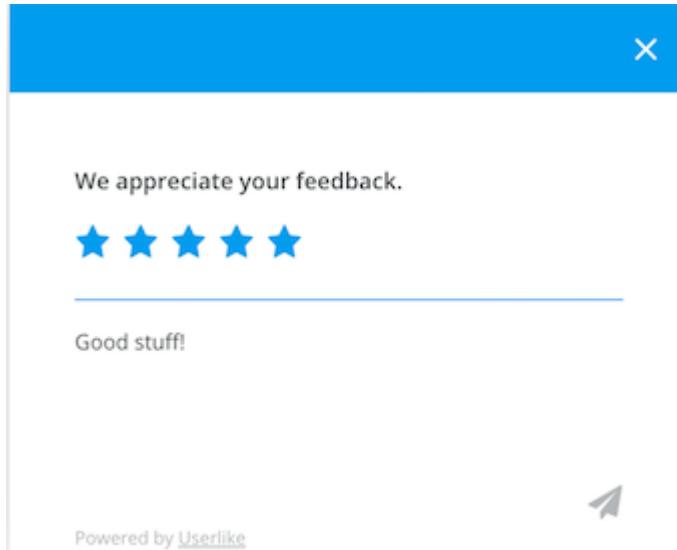


Figure 5.3:

- Capital Budgeting Tests
 - Payback Period
 - Time Adjusted or Discounted Payback Period

5.2 Customer Perspective

According to research from walkerinfo.com (customer research consulting), customer satisfaction will overtake as the key success. Company require volumes of customers, number contacts, and number employees, and another data as a key factors. It depend on the situation and company business model.

Below are example of the most using KPI for Customer Perspektive:

5.2.1 Customer Satisfaction Score (CSAT)

Measuring CSAT is very hard. Customers need to express an emotion, and emotions are harder to indentify than objective facts. CSAT can consist of regular numbers, some icon (like stars, smiley faces, etc).

Example: Users uses a 5 stars scale for ratings

5.2.2 Net Promoter Score (NPS)

How likely customers are referring the products to someone else.

Example: Company ask the customers how likely they are to recommend the products on a scale from 1 to 10.

5.2.3 First Response Time

Customers are changed like the Spice Girls: "If you wanna get with me, better make it fast!". The Customers expect an excellent shopping experience. A Salesforce study found that a third of the respondents felt positive about companies that offered a quick first response. First Response Time is calculated by subtracting the time of the request from the time of the initial reply.

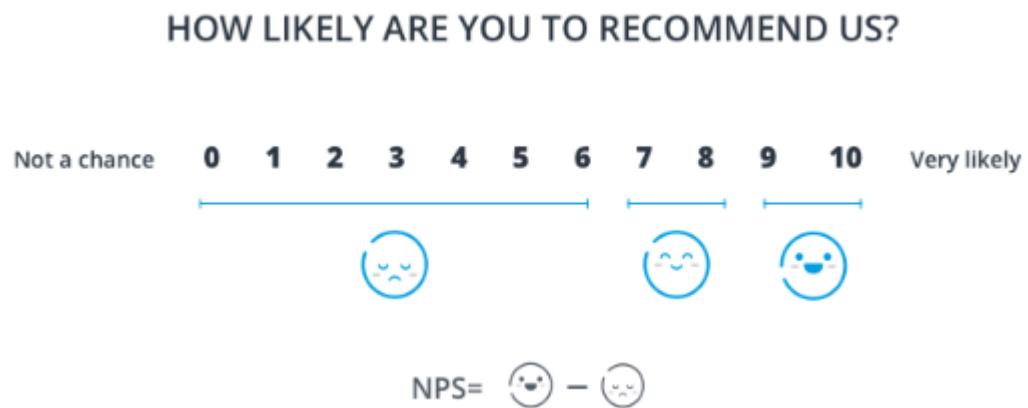


Figure 5.4:



Figure 5.5:

5.2.4 Customer Retention Rate

The ability to keep a customer over a set period of time (daily, monthly, yearly).

Formula: Customer Retention Rate = $((CE - CN) / CS) \times 100$

- CE = Number of customers at end of period
- CN = Number of new customers acquired during period
- CS = Number of customers at start of period

5.2.5 Employee Engagement

The KPI about employees satisfaction. The standards approach is direct ask. Another method is survey. The survey should consist of the following engagement levels:

- Management quality and time investment
- Influence from colleagues
- Relationships
- Work schedule

Another example of customer perspective are:

- Annual sales/customers(\$)
- Average customer size(\$)
- Customer rating(%)
- Average time from customer contact to sales response(No)
- Average time spent on customer relations(No)
- Customers/employee (No or %)
- Satisfied-customer index(%)
- Customer-loyalty index(%)
- Market Share
- No of Customer Complaints
- Return Rates
- Response Time
- Cost/customer(\$)
- Customers Lost(No or %)
- Customer retention
- Number of customers
- Annual sales per customer
- Marketing cost as a % of sales(%)
- Marketing expenses(\$)
- Number of proposals made
- Brand-image index (%)
- Response rate
- Sales volume
- Sales per channel
- Average customer size
- Customers per employee
- Frequency of sales transactions
- Sales closed/sales contacts(%)
- Number of visits to customers(No)
- Service expense/customer/year(\$)

5.2.6 Complete List

- Customer Relationship Management (CRM)
 - Customer Acquisition Rate 133 3.1.2

- Customer Churn Rate 134 3.1.3
- Customer Retention Period 136 3.1.4
- Customer Significance Level 138 3.1.5
- Cross-Selling Ratio 140 3.1.6
- Customer Lifetime Value (CLV) 141 3.1.7
- Customer Satisfaction Index 143 3.1.8
- Customer Complaint Ratio 146 3.1.9
- Flop Rate
- Marketing Communication Indicators
 - Media Coverage Level 150 3.2.2
 - Click Through Rate (CTR) 152 3.2.3
 - Conversion Rate 153 3.2.4
 - Cost per Thousand 155 3.2.5
 - Brand Awareness Level
- Product Pricing
 - Profit Margin
 - Gross Margin
 - Absolute Contribution Margin
 - Percentage Contribution Margin
 - Price Reduction Rate
 - Direct Product Profit (DPP)
 - Price Elasticity of Demand (PEoD)
 - Purchasing Power Index
- Cost-Profit-Volume Analysis
 - Break-Even Point (BEP)
 - Margin of Safety
 - Margin of Safety-Factor
 - Cash Point
- Market Coverage Indicators
 - Internationalization Level
 - Distribution Coverage Level
 - Customer Coverage Ratio
 - Market Saturation Level
- Market Position Indicators
 - Absolute Market Share
 - Relative Market Share
 - Bid Acceptance Rate
- Sales Efficiency Indicators
 - Sales per Reference Parameter
 - Contribution Margin per Reference Parameter
 - Sales Space Productivity

- Capacity Coverage Ratio
- Book-to-Bill Ratio
- Finished Goods Turnover Period

5.3 Process Perspective

5.3.1 Cost performance index (CPI)

The financial effectiveness and efficiency of a project. For example, if a project has a earned value of £30,000 but actual costs were £12,000.

$$\text{CPI} = \text{EV} / \text{AC} = 30,000 / 12,000 = 2.5$$

If the ratio has a value higher than 1 then it indicates the project is performing good.

5.3.2 Schedule Performance Index (SPI)

Indicates how efficiently the company actually progressing compared to the planned project schedule. It is the efficiency of the time on the project.

Formula: Schedule Performance Index = (Earned Value) / (Planned Value)

Example:

A Company have a project to be completed in 12 months, and the budget of the project is 100,000 USD. Six months have passed, and 60,000 USD has been spent, but on closer review, you find that only 40% of the work has been completed so far.

Given in the question:

$$\begin{aligned} \text{Actual Cost (AC)} &= \$60,000 \\ \text{Planned Value (PV)} &= 50\% \text{ of } \$100,000 \\ &= \$50,000 \\ \text{Earned Value (EV)} &= 40\% \text{ of } \$100,000 \\ &= \$40,000 \\ \text{And then,} \\ \text{Schedule Performance Index (SPI)} &= \text{EV} / \text{PV} \\ &= 40,000 / 50,000 \\ &= 0.8 \end{aligned}$$

The Schedule Performance Index is 0.8

This Company is behind schedule since the Schedule Performance Index is less than one.

5.3.3 Complete List

- Project Controlling
 - Schedule Performance Index (SPI)
 - Cost Performance Index (CPI)
 - Time Estimation at Completion (TEAC)
 - Estimate at Completion (EAC)
 - To-Complete-Performance Index (TCPI)

- Process Acceleration Costs
- Quality Controlling
 - Quality Rate
 - Rejection Rate
 - Follow-up Costs Ratio
- Conformity Costs Ratio
- Non-Conformity Costs Ratio
- Supply Chain Management
 - Procurement Efficiency Ratio
- Supply Chain Cycle Time
- Faulty Incoming Delivery Rate
- Faulty Outgoing Delivery Rate
- Vertical Integration Level
- Supplier's Service Level
- Cooperation Index
- Production Capacity Management
 - Plant Availability Time
- Plant Downtime Rate 236
- Maintenance Cost Intensity
- Capacity Utilization Level
- Contribution Margin per Unit of the Constrained Resource
- Process Controlling
 - Throughput Time (TPT)
 - Days Inventory Outstanding (DIO)
- Inventory Turnover Ratio
- Material Coverage Period
- Process Cost Rate
- Expected Process-based Loss
- Machine Hour Rate
- Bottleneck-induced Incremental Costs
- Sustainability Management
 - Resource Consumption Level
- Resource Consumption Efficiency Level
- Sustainable Value
- Emission Volume of Production-related Pollutants
- Emission Volume of Usage-related Pollutants
- Disposal Costs Ratio

- Recycling Ratio

5.4 Human Resource and Innovation Perspective

Retention

The basic formula for employee retention is the following:

$$(\# \text{ of employees who stayed at the company for the whole time period}) / (\# \text{ employees at start of the time period}) \times 100$$

Example : If 90 people were working at my company as of June 1st and 80 of those same people were still working at my company as of June 30th, my retention rate for the month of June would be the following:
 $80/90 * 100 = 88.9\%$

Time in The Position How long are employees in the same position

Absenteeism

- Delays
- Sick leave
- Excused or unexcused absences

Recruitment Time

Time between employee leaving and another candidate selected to replace him.

Education

The courses for employee that has direct impact on the company performances.

Time to Achieve Goals

The efficiency of the workforce to see how long it takes to finish the tasks.

Accidents at Work

The number of accidents in the workplace.

Complete List:

- Personnel Cost Management
 - Personnel Costs Ratio 283 5.1.2
 - Supplementary Personnel Costs Ratio 285 5.1.3
 - Personnel Costs per Employee 287 5.1.4
 - Unit Labor Cost
- Human Resource Controlling
 - Labor Productivity
 - Overtime Quota
 - Workforce Composition Ratios
 - Internally-staffed Executive Positions Ratio
 - Staff Recruitment Period
- Human Resource Development Indicators
 - Apprenticeship Quota
 - Trainee Absorption Rate
 - Professional Development Training Time per Employee
 - Professional Development Training Costs Ratio
- Organizational Behavior Indicators

- Labor Turnover Rate
- Employees Satisfaction Index
- Sickness-Absenteeism Rate
- Accident Occurrence Rate
- Participation Rate in Ideas Management
- Innovativeness
 - Innovation Rate
 - Research and Development Intensity (R&D Intensity)
 - Research and Development Costs Ratio (R&D Costs Ratio)
 - Break-Even Time

5.5 By Industry

5.5.1 T & L Industry

Example KPIs for the Transportation and Logistic (Warehousing) Industry:

- Annualized inventory turns: a ratio showing how many times a company has sold and replaced inventory during a given period (years, months, 10 years).
- Annualized cost of goods sold (COGS)/average daily inventory value
- Backlog value
- Value of open, not yet fulfilled, booked order lines
- Book to fulfill ratio
- Booked order value/fulfilled value
- Book to ship days
- Average of shipped date - Firm date (booked date used if no firmed date)
- Booked order value
- Booked order line value (not including returns)
- Claims percentage for freight costs
- Customer order promised cycle time
- Defects per million opportunities
- Inventory months of supply
- On-time line count
- On-time pickups
- Pick exceptions rate
- Percentage of picks with exceptions
- Pick release to ship
- Planned inventory turns
- Planned cost of goods sold/planned inventory value
- Planned margin
- Planned revenue - Planned costs
- Planned margin percentage
- Planned margin/planned revenue
- Planned on-time shipment
- Planned service level (percentage of shipments shipped on time)
- Planned resource utilization
- Planned resource usage
- Product revenue
- Product sales revenue (not including service) recognized in selected period (based on AR invoice lines)

- Product revenue backlog
- Value of booked order lines less returns plus deferred revenue backlog (invoiced but not recognized)
- Production value
- Value of work-in-process (WIP) completions into inventory
- Production to plan rate
- Production standard value/planned standard value
- Receipt to put-away
- Time elapsed from pick release to ship confirm
- Time elapsed from receipt
- Transit time

5.5.2 Wholesale Trade

Example KPIs for the Wholesale Trade Industry:

- Dock turnaround time
- Freight costs (minimize costs without affecting deliveries)
- Inventory accuracy, stockouts
- Inventory carrying costs
- Inventory turns per year
- Logistics costs per year
- Low-velocity inventory comparison through sectors
- Order fill rate and accuracy
- Technology used to execute inventory strategies
- Warehouse flow-through (or some measure of yard or warehouse productivity)
- Wholesale revenue
- Total factor productivity
- Labor productivity
- Return on assets
- Profit margin
- Debt to equity
- Inventory turnover
- Asset utilization
- Collection efficiency

5.5.3 Utilities Industry

Example KPIs for the Utilities Industry:

- Annual labor cost per device
- Average cost per job category
- Average cost per megawatt produced
- Average labor hours per device per year
- Average maintenance cost per mile of pipe/line/cable
- Average number of days each work order is past due
- Average number of labor hours to complete a maintenance task
- Average response time to fix breaks
- Average revenue per megawatt produced
- Average time to settle a rate case
- Consumption analyzed by units consumed and target reduction achieved
- Crew productivity
- Drinking water quality - Percentage of water tests that meet regulatory standards
- Electrical grid load
- Equipment failure rate
- Equipment unavailability, hours per year - Planned maintenance

- Equipment unavailability, hours per year - Sustained fault
- Equipment unavailability, hours per year - Temporary fault
- Equipment unavailability, hours per year - Unplanned maintenance
- Maintenance backlog
- Maintenance cost as a percentage of manufacturing cost
- Maintenance technician's skill level improvement, year over-year
- Mean time to repair
- Number of complaints received by type
- Number of customers who were cut off due to violations of regulations
- Number of disconnections
- Number of pending work orders
- Number of power failures per year
- Number of reported gas leakages per 1,000 households
- Number of sewage blockages per month/year
- Number of staff per 1,000 customer connections
- Number of uncontrolled sewage overflows affecting private properties
- Outage time per event
- Percentage of customers that would characterize their bills as accurate and timely
- Percentage of possible power revenue billed
- Percentage reduction in number of complaints to the local regulatory body
- Percentage reduction in number of employee injuries
- Percentage reduction in number of equipment failures
- Percentage of maintenance work orders requiring rework
- Percentage of man-hours used for proactive work
- Percentage of scheduled man-hours to total man-hours
- Profit redistribution (rural electric coops)
- Reduction in hazardous liquid spill notification time
- Reduction or stabilization in rates (municipally owned utilities)
- Response time to gas or water leaks
- Sewage system reliability
- Station unavailability - Planned maintenance
- Station unavailability - Sustained fault
- Station unavailability - Temporary fault
- Total shareholder returns (investor-owned utilities)
- Total time to complete new customer connections
- Transformer/pump station reliability
- Voltage deviations per year
- Water system reliability

5.5.4 Retail Industry

Example KPIs for the Retail Trade Industry:

PRODUCT SALES

- Average inventory
- Cost of goods sold
- Gross profit budget percentage
- Sales budget percentage
- Discount
- Gross profit
- Gross profit and prognostics
- Gross profit and prognostics percentage
- Gross profit budget
- Gross profit campaign

- Gross profit percentage KPI
- Gross profit prognostics
- Gross profit prognostics percentage
- Gross profit standard
- Gross profit year to date
- Number of stores
- Product quantity
- Sales
- Sales and prognostics
- Sales campaign
- Sales growth period
- Sales growth year
- Sales growth year by week
- Sales prognostics
- Sales standard
- Sales trend percentage KPI
- Sales value-added tax (VAT)
- Sales view
- Sales view year-to-date
- Share prognostics
- Time range

FINANCE AND ACCOUNTING

- Accounts payable turnover
- Accounts receivable turnover days
- Acid test ratio
- Administrative cost percentage
- Break-even (dollars)
- Cash conversion cycle
- Contribution margin
- Cost of goods
- Cost of goods sold
- Current ratio
- Ending inventory at retail
- Gross margin
- Gross margin return on investment
- Initial mark-up
- Interest cost percentage
- Inventory turnover
- Maintained mark-up (dollars)
- Margin percentage
- Mark-up percentage
- Net receipts
- Net sales
- Retail price
- Return on capital invested
- Sales per square foot
- Stock turnover days
- Total asset sales ratio
- Turnover

SALARY

- Real absence hours
- Real absence share

- Real GPWH
- Real overtime hours
- Real overtime share
- Real TWH
- Real working hours
- Salary
- Salary amount
- Salary amount exchange currency
- Salary hours
- Salary turnover share

SALARY TARGETS

- Real absence hours
- Real GP work hours
- Real total work hours
- Salary absence percentage
- Salary GP work hour
- Salary overtime percentage
- Salary target absence percentage
- Salary target GP work hour
- Salary target overtime percentage
- Salary target turnover percentage
- Salary target work hour
- Salary turnover percentage

HOURLY SALES

- Customers per hour
- Discount
- Gross profit
- Items
- Margin per customer
- Number of customers
- Sales growth year
- Sales growth year percentage
- Sales last year
- Sales per customer
- Sales trend percentage
- Sales view
- Total number of stores

BUDGET SALES

- Budget gross profit
- Budget number of customers
- Budget sales
- Customers
- Discount
- Gross profit
- Items
- Sales
- Sales exchange currency
- Sales VAT

PAYMENT WITH POINT-OF-SALE (POS) STATISTICS

- Amount

- Amount exchange currency
- Items
- Number of customers
- Number of items
- Refund amount
- Refund count
- Sales income VAT
- Time range
- Transaction cancel amount
- Transaction cancel count
- Transaction cancel percentage
- Void amount
- Void count
- Void percentage
- Zero sale count

HOURLY PRODUCT SALES

- Gross profit percentage
- Item discount
- Item gross profit
- Item quantity
- Item sales
- Item sales exchange currency
- Item sales VAT
- Items sold

5.5.5 Real Estate and Rental and Leasing

Example KPIs for the Real Estate and Rental and Leasing Industry:

REALTOR WEBSITE

- Conversation rate (i.e., take rate) - Number of conversations over number of website visits
- Top conversion page exit - The page where website visitors change their minds and exit your website.
- Traffic source percentage - Website visits referred by

REAL ESTATE OFFICE

- Advertising and promotion
- Average commission per sale
- Average commission per salesperson
- Commission margin
- Net profit
- Office cost (telephone, fax, and other office cost)
- Rent cost of premises
- Sold homes per available inventory ratio
- Total income
- Wages and salaries (including commissions and vehicle allowances)
- Year-to-year variance on average sold price
- Year-to-year variance on dollar volume of sold listings
- Year-to-year variance on sold average dollar per square foot

COMMERCIAL PROPERTY MANAGEMENT

- Annual return on investment in percentage
- Construction/purchaser rate - New constructed or purchased units over time
- Cost per square foot

- Equity value growth in percentage
- Lease events coverage ratio - Number of lease inquiries over number of available units
- Management efficiency - Number of leased spaces over number of staff
- Market share growth
- Monthly return on investment as percentage
- Occupancy cost - Cost per occupied unit
- Operation cost to rent income ratio
- Percentage of rent collected
- Price to income as percentage
- Profitability per square foot
- Real estate demand growth - Market rental demands
- Rented space usage quality - Average number of tenant visits over rented space
- Renting cost - Renting cost per square foot
- Renting return on investment - Rent income over cost
- Revenue per square foot
- Risk metrics as percentage
- Total property management income per property manager
- Usage efficiency - Available renting square feet over number of staff
- Utilization (vacancy) rate - Rented square feet over total square feet, or rented units over total units

REAL ESTATE INVESTOR

- Average gross multiplier for portfolio
- Cost per square foot to value per square foot ratio
- Equity to value ratio
- Gross multiplier per commercial property
- LTV (loan to value) ratio per property
- Mortgage rate index
- Overall LTV (loan to value) ratio for portfolio
- Price per square foot to value per square foot ratio
- Profitability per square foot
- Property value growth (market trend)
- Purchase price-to-appraisal value ratio
- Rental value growth rate ROI (return on investment)

5.5.6 Manufacturing Industry

Example KPIs for the Manufacturing Industry:

- Asset utilization
- Availability
- Avoided cost
- Capacity utilization
- Comparative analytics for products, plants, divisions, companies
- Compliance rates (for government regulations, etc.)
- Customer complaints
- Customer satisfaction
- Cycle time
- Demand forecasting
- Faults detected prior to failure
- First aid visits
- First time through
- Forecasts of production quantities, etc.
- Increase/decrease in plant downtime
- Industry benchmark performance
- Integration capabilities

- Interaction level Inventory
- Job, product costing
- Labor as a percentage of cost
- Labor usage, costs-direct and indirect
- Machine modules reuse
- Maintenance cost per unit
- Manufacturing cost per unit
- Material costing, usage
- Mean time between failure (MTBF)
- Mean time to repair
- Number of production assignments completed in time
- On-time orders
- On-time shipping
- Open orders
- Overall equipment effectiveness
- Overall production efficiency of a department, plant, or division
- Overtime as a percentage of total hours
- Percentage decrease in inventory carrying costs
- Percentage decrease in production-to-market lead-time
- Percentage decrease in scrap and rework costs
- Percentage decrease in standard production hours
- Percentage increase in productivity
- Percentage increase in revenues
- Percentage material cost reduction
- Percentage reduction in defect rates
- Percentage reduction in downtime
- Percentage reduction in inventory levels
- Percentage reduction in manufacturing lead times
- Percentage savings in costs
- Percentage savings in inventory costs
- Percentage savings in labor costs
- Percentage savings in transportation costs
- Planned work to total work ratio
- Predictive maintenance monitoring (maintenance events per cycle)
- Process capability
- Productivity
- Quality improvement (first-pass yield)
- Quality tracking-six sigma
- Reduced time to productivity
- Reduction in penalties
- Savings in inventory carrying costs
- Scheduled production
- Spend analytics
- Storehouse stock effectiveness
- Supplier trending
- Time from order to shipment
- Time on floor to be packed
- Unplanned capacity expenditure
- Unused capacity expenditures
- Utilization
- Waste ration reduction
- Work-in-process (WIP)

INSURANCE

- Average insurance policy size
- Claims
- Combined cost and claims ratio
- Combined ratio
- Current premium versus loss
- Earned premium
- Expense ratio
- Expenses
- Exposure
- Loss adjustment expenses (LAE)
- Loss ratio
- Number of days open of insurance claims
- Number of new insurance policies
- Previous premium versus loss
- Underwriting speed of insurances
- Written premium

5.5.7 Finance and Insurance

Example KPIs for the Finance and Insurance Industry:

FINANCE

- Accounting costs
- Accounts payable
- Accounts payable turnover
- Asset turnover rate
- Average sum deposited in new deposit accounts
- Average value of past due loans
- Cash conversion cycle (CCC)
- Cash dividends paid
- Cash flow return on investments (CFROI)
- Common stock equity
- Cost of goods sold (COGS)
- Cost per hour per lawyer (in-house)
- Creditor days
- Cumulative annual growth rate (CAGR)
- Cycle time to perform periodic close
- Cycle time to resolve an invoice error
- Days payable
- Debt-to-asset ratio
- Debtor days
- Direct costs
- Earnings per share (EPS)
- EBIT
- EBITDA
- Economic value added
- Enterprise value/takeover value
- Fixed costs
- Gross margin on managed assets
- Gross profit
- Gross profit margin
- Indirect costs
- Interest expense
- Interest on net worth

- Invoicing processing costs
- Labor and management cost
- Labor and management earnings
- Legal staff per size of revenue
- Long-term debt
- Marginal costs
- Market share
- Net change in cash
- Net interest margin
- Net new money
- Net profit
- Net profit margin
- Number of budget deviations
- Number of invoices outstanding
- Number of past due loans
- Operating income
- Operating leverage
- Operating margin
- Operating profit margin
- Other current liabilities
- Other noncurrent liabilities
- Percentage of accuracy of periodic financial reports
- Percentage of effectiveness in payables management
- Percentage of budget deviation relative to total budget
- Percentage of electronic invoices
- Percentage of financial reports issued on time
- Percentage of invoices requiring special payment
- Percentage of invoices under query
- Percentage of legal budget spent outside
- Percentage of low-value invoices
- Percentage of payable invoices without purchase order
- Preferred stock equity
- Product turnover ratio
- Profit
- Profit loss due to theft
- Profit margin
- Profit per product
- Quick ratio
- Rate of return on assets
- Rate of return on equity
- Return on assets
- Return on capital employed (ROCE)
- Return on investment (ROI)
- Return to equity
- Revenue
- Revenue per employee
- Sales per share
- Same store sales
- Selling general and administrative (SG&A) expenses
- Share price
- Shares outstanding
- Sharpe ratio
- Short-term debt
- Sortino ratio

- Systems cost of payroll process as a percentage of total payroll cost
- Tier 1 capital
- Total assets
- Total current liabilities
- Total equity
- Total legal spending as a percentage of revenue
- Total liabilities
- Total of uninvested funds
- Total quantity of new deposit accounts
- Total sum deposited in new deposit accounts
- Total value of past due loans
- Variable costs

5.5.8 Construction

Example KPIs for the Construction Industry:

- Number of accidents
- Number of accidents per supplier
- Actual working days versus available working days
- Cash balance - Actual versus baseline
- Change orders - Clients
- Change orders - Project manager
- Client satisfaction - Client-specified criteria
- Client satisfaction product - Standard criteria
- Client satisfaction service - Standard criteria
- Cost for construction
- Cost predictability - Construction
- Cost predictability - Construction (client change orders)
- Cost predictability; Construction (project leader change orders)
- Cost predictability - Design
- Cost predictability - Design and construction cost to rectify defects
- Customer satisfaction level
- Day to day project completion ratio - Actual versus baseline
- Fatalities
- Interest cover (company)
- Labor cost - Actual versus baseline
- Labor cost over project timeline
- Liability ratio (over asset) on current versus completion comparison
- Number of defects
- Outstanding money (project)
- Percentage of equipment downtime
- Percentage of labor downtime
- Percentage of backlogs over project timeline
- Percentage of unapproved change orders
- Productivity (company)
- Profit margin - Actual versus baseline profit margin over project timeline
- Profit predictability (project)
- Profitability (company)
- Quality issues at available for use
- Quality issues at end of defect rectification period
- Ratio of value added (company)
- Repeat business (company)
- Reportable accidents (including fatalities)
- Reportable accidents (non-fatal)

- Return on capital employed (company)
- Return on investment (client)
- Return on value added (company)
- Time for construction
- Time predictability - Construction
- Time predictability - Construction (client change orders)
- Time predictability - Construction (project leader change orders)
- Time predictability - Design
- Time predictability - Design and construction
- Time taken to reach final account (project)
- Time to rectify defects

5.5.9 IT Industry

Example KPIs for the Information Technology Industry:

- Annual cost per reading
- Average cost per article
- Average cost per subscription
- Average dollars per email sent or delivered
- Average order size
- Average quarter-hour audience
- Average revenue per subscription
- Average time spent listening per user (day/week/month/year)
- Bounce rate
- Click to open rate (number of unique clicks/ number of unique opens)
- Click-through rate
- Click-through rate (CTR)
- Conversion rate
- Conversion rate (number of actions/unique click-throughs)
- Conversion rates
- Cost per broadcast hour
- Cost per consumed (by viewers/listeners) hour
- Cost per customer
- Cost per lead, prospect, or referral
- Cost per production hour
- Cost per viewer/listener
- Cost per visitor
- Cost per action (CPA)
- Cumulative audience sessions
- Delivery rate (emails sent, bounces)
- Gross ratings points
- Life cycle cost per reading
- Local content as a percentage of all content
- Net subscribers (number of subscribers plus new subscribers) -(bounces + unsubscribes)
- Number of broadcast hours per day/week/month/year
- Number of or percentage of spam complaints
- Number of orders, transactions, downloads, or actions
- Open rate
- Output per employee (unique first run broadcast hours by employee for each medium)
- Pay per click (PPC)
- Pay per lead (PPL)
- Pay per sale (PPS)
- Percentage of broadcast hours by genre (news/sports/entertainment, etc.)
- Percentage of overhead (non-direct operating costs) against total expenditure

- Percentage of orders, transactions, downloads, or actions of emails sent or delivered
- Percentage unique clicks on a specific recurring link(s)
- Referral rate (“send-to-a-friend”)
- Site stickiness (number of pages visited per visit)
- Subscriber retention (number of subscribers, bounces, unsubscribes/number of subscribers)
- Total cost per subscription
- Total listener hours (day/week/month/year)
- Total revenue
- Total revenue per subscription
- Unique visitors (total number of unique visitors per day/week/month)
- Unsubscribe rate
- Utilization of production resources
- Value per visitor
- Viewers/listeners for each medium as a percentage of total population
- Website actions (number of visits to a specific web page or pages)
- Website traffic (total page impressions per day/week/month)

5.6 By Business Goal

When it comes to improving on specific business goals, KPI got you covered.

5.6.1 Improve response time

Goals: Improve representative overall efficiency.

5.6.2 Increase profit margins

Goals:

- Allow the CFO of the company to compare the net profit margin over time and relative to other companies in the same sector.
- Allow investors to compare the net profit margin across industries to identify the most profitable and attractive sectors and companies to invest in

5.6.3 Optimize campaigns

Goals: Increase sales originating from our online marketing activity.

5.7 By Department

5.7.1 Sales

Example KPIs for Sales Departments:

- Actual calls
- Actual sales value versus initial bid
- Age of sales forecast
- Average administrative time per sales person
- Average deal size
- Average number of activities (calls, meetings, etc.) to close a deal
- Average price discount per product
- Average price discount per sales person
- Average revenue per product
- Call quota
- Closed sales

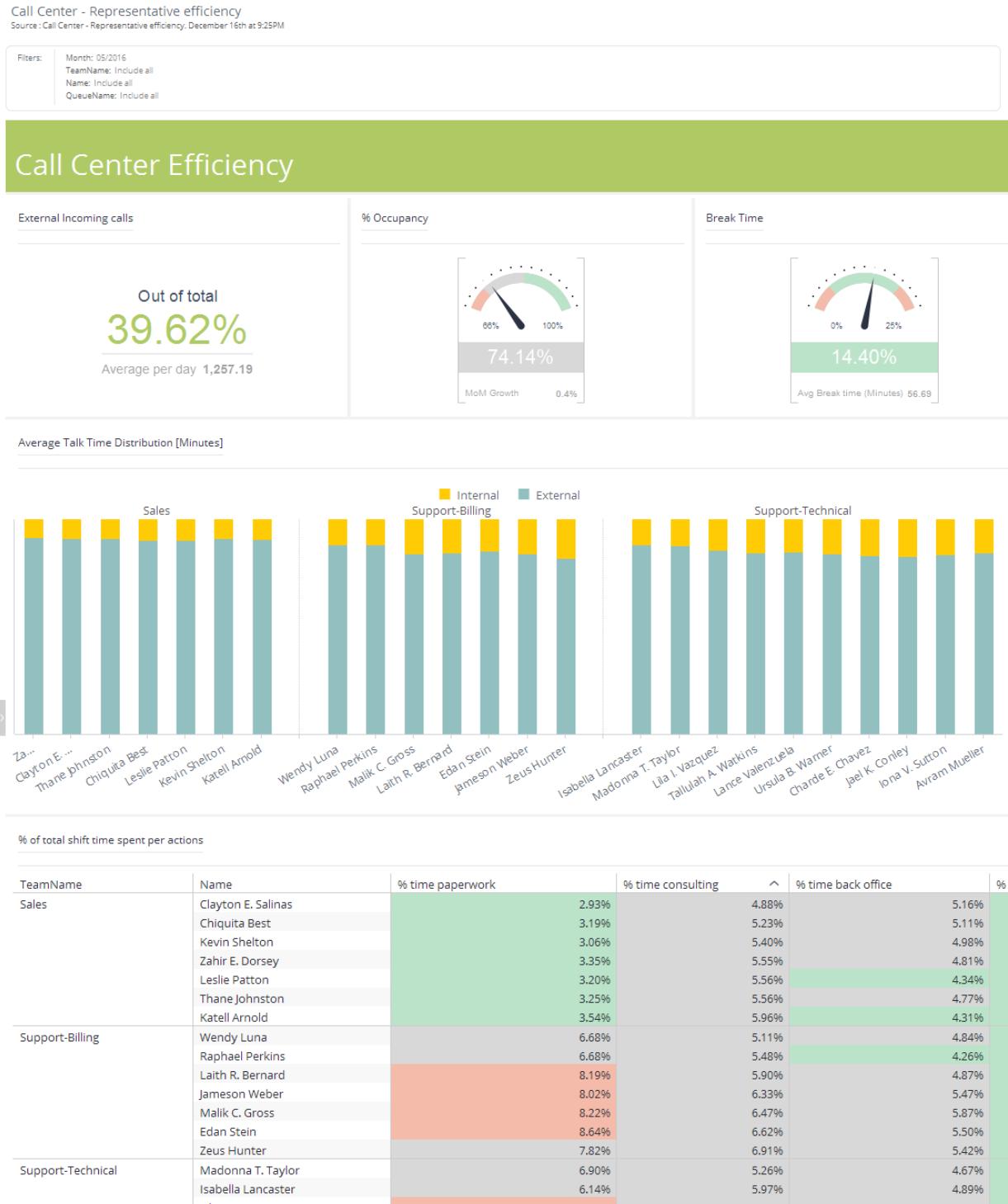


Figure 5.6:

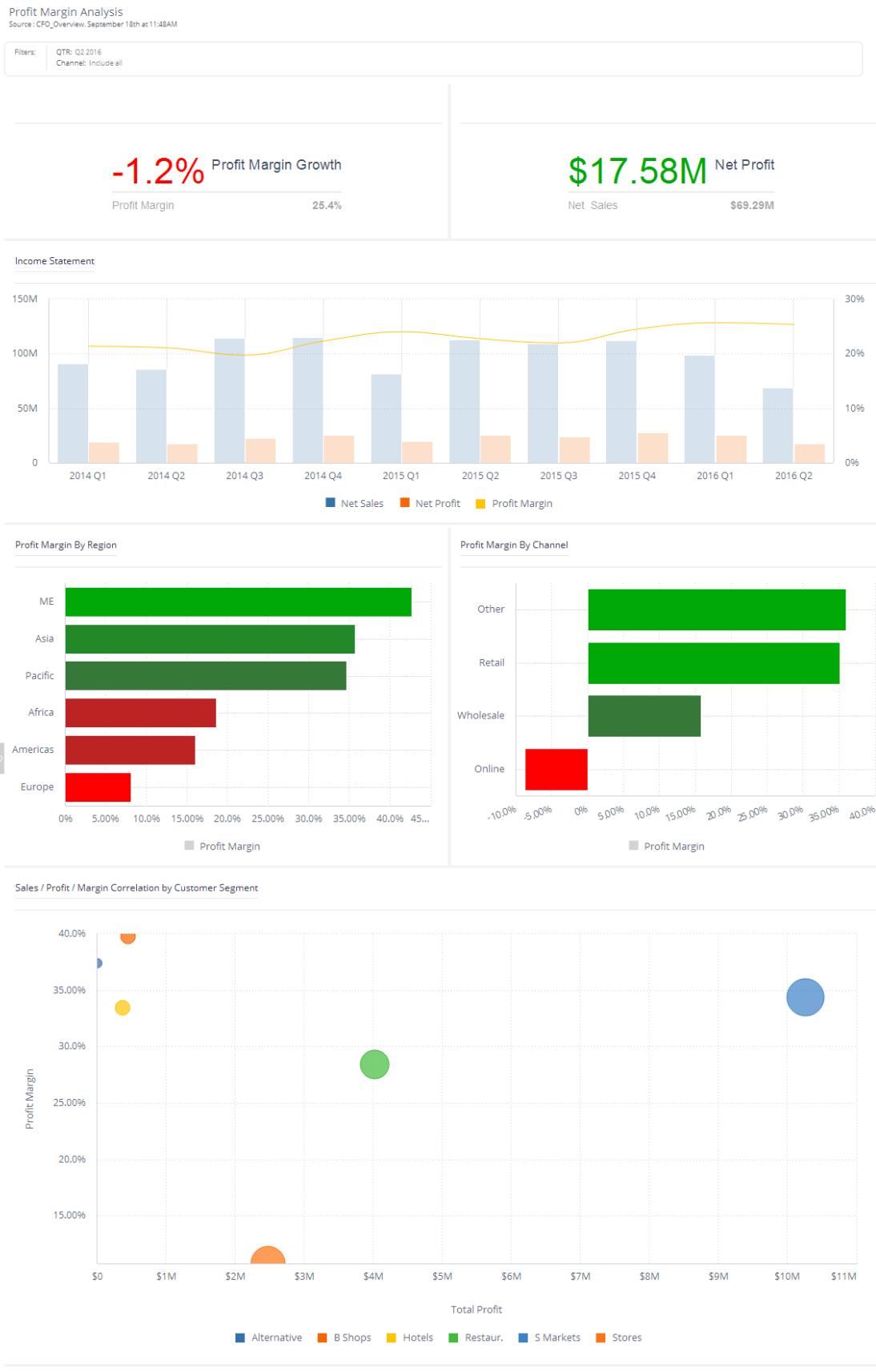


Figure 5.7:



Figure 5.8:

- Closing ratio
- Customer acquisitions costs as a percentage of sales value
- Customer churn ratio
- Customer loyalty
- Customer purchase frequency
- Customer satisfaction
- Frequency of sales transactions
- Gross margin per product
- Gross margin per sales person
- New sales person ramp-up time
- Number of certified partners
- Number of deals per partner
- Number of sales orders by FTE
- Number of sales people meeting their quota
- Number of units sold per day/week/month/quarter/year
- Partner churn ratio
- Partner profit margin
- Percentage of converted opportunities
- Percentage of online sales revenue
- Percentage of sales due to launched product/services
- Percentage of sales representatives to achieve quota
- Percentage of sales revenue via partner channel
- Pipeline by sales stage
- Qualified leads
- Qualified opportunities
- Revenue per sales person
- Sales capacity
- Sales cycle time
- Sales per department
- Sales person turnover
- Sales quota
- Time utilization
- Unweighted sum of deal size in sales pipeline
- Value of sales lost
- Win/loss ratio percentage

5.7.2 Marketing

Example KPIs for Marketing Departments:

- Ad click-through ratio (CTR)
- Average response rates of campaigns
- Brand awareness percentage
- Brand consideration
- Brand credibility
- Brand strength
- Column inches of media coverage
- Consumer awareness
- Contact rate (number of contacts effectively contacted / number of contacts in the target list)
- Cost per converted lead
- Cost per lead
- Cost per mille (CPM)
- Delivery of materials
- Effective reach
- Gross rating point (GRP)

- Growth sustainability rate of brand
- Leads generated
- Marketing budget awareness-demand ratio
- Marketing budget ratio (MER)
- Number of article placements in trade magazines
- Number of client visits
- Number of product focus groups conducted
- Number of customer satisfaction surveys administered
- Number of placements in trade magazines
- Number of trade shows attended / participated in
- Percentage of customers willing to promote your product/service
- Q score (a way to measure the familiarity and appeal of a brand, etc.)
- Response rate
- Return on investment (ROI) of brand
- Return on marketing investment (ROMI)
- Revenue generation capabilities of brand
- Staying in budget
- Target rating point
- Total cost of customer acquisition
- Transaction value of brand
- Website click-throughs
- Website hits
- Website leads generated

5.7.3 Finance

Example KPIs for Finance Departments:

- Accounting costs
- Accounts payable turnover
- Accounts receivable collection period
- Accounts receivable turnover
- Actual expenses
- Amount due (per customer)
- Average customer receivable
- Average monetary value of invoices outstanding
- Average monetary value of overdue invoices
- Average number of trackbacks per post
- Budget variance for each key metric
- Budgeted expenses
- Capital expenditures
- Cash conversion cycle (CCC)
- Cash flow return on investments (CFROI)
- Cost of goods sold (COGS)
- Cash dividends paid
- Cost per pay slip issued
- Creditor days
- Current receivables
- Cumulative annual growth rate (CAGR)
- Cycle time for expense reimbursements
- Cycle time to process payroll
- Cycle time to resolve an invoice error
- Cycle time to resolve payroll errors
- Days payable
- Debtor days

- Direct cost
- Discounted cash flow
- Earnings before interest and taxes (EBIT)
- Earnings before interest, taxes, depreciation (EBITDA)
- Economic value added (EVA)
- Employee available time
- Employee scheduled time
- Employee work center loading
- Enterprise value/ takeover value
- Expense account credit transactions
- Expense account debit transactions
- Expense account transactions
- Fixed costs
- Gross profit
- Gross profit margin
- Indirect costs
- Inventory turnover
- Inventory value
- Invoice processing costs
- Internal rate of return (IRR)
- Market share gain comparison percentage
- Net change in cash
- Net income
- Net present value (NPV)
- Number of invoices outstanding
- Number of unapplied receipts
- Number of past-due loans
- Open receivables
- Open receivables amount (per customer)
- Operating leverage
- Past-due receivables
- Payables turnover
- Payment errors as a percentage of total payroll disbursement
- Percentage accuracy of financial reports
- Percentage of bad debts against invoiced revenue
- Percentage of electronic invoices
- Percentage in dispute (per customer)
- Percentage of invoices being queried
- Percentage of invoices requiring special payment
- Percentage of low-value invoices
- Percentage of open receivables (per customer)
- Percentage of payable invoices without purchase order
- Percentage of service requests posted via web (self-help)
- Perfect order measure
- Quick ratio
- Receivables
- Receivables turnover
- Return on capital employed (ROCE)
- Sales growth
- Share price
- Systems cost of payroll process as a percentage of total payroll cost
- Total payables
- Total energy used per unit of production
- Total receivables

- Total sales
- Unapplied receipts
- Variable costs
- Weighted days delinquent sales outstanding
- Weighted days delinquent sales outstanding (per customer)
- Weighted terms outstanding
- Weighted terms outstanding (per customer)

5.7.4 Human Resources

Example KPIs for Human Resources (HR) Departments:

- Actual versus budgeted cost of hire
- Annualized voluntary employee turnover rate
- Annualized voluntary turnover rate
- Average headcount of employees each human resources (HR) employee working is caring for
- Average interviewing costs
- Average length of placement in months for the manager
- Average length of service of all current employees
- Average length of service of all employees who have separated
- Average months placement
- Average number of training hours per employee
- Average number of vacation days per employee
- Average performance scores of departing employees
- Average retirement age
- Average salary
- Average salary for all employees reporting to the selected manager
- Average sourcing cost per hire
- Average time employees are in same job/ function
- Average time to competence
- Average time to update employee records
- Average training costs per employee
- Compensation cost as a percentage of revenue
- Contingent workers
- Employee satisfaction with training
- End placements
- Female to male ratio
- Full-time employees (FTEs) per human resources (HR) department FTE
- Headcount of contingent workers for the manager
- HR average years of service (incumbents)
- HR average years of service (terminations)
- HR department cost per FTE
- HR headcount - Actual
- HR headcount - Available
- HR to employee staff ratio
- Job vacancies as a percentage of all positions
- New hire quality
- Time to fill
- Hiring manager satisfaction
- Cost per hire
- Staffing efficiency
- Internal, external, and total headcount recruiting costs and ratios
- Number of end placements made in the reporting period for the manager
- Part-time employees as a percentage of total employees
- Percentage of employees receiving regular performance reviews

- Percentage of employees that are near or at max for their vacation balances
- Percentage of HR budget spent on training
- Percentage of new hire retention
- Ratio of internal versus external training
- Ratio of standard level wage to local minimum wage
- Return on investment (ROI) of training
- Total overtime hours as a percentage of all work hours
- Training penetration rate (percentage of employees completing a course compared to all FTEs)
- Workforce stability

5.7.5 Information Technology

Example KPIs for Information Technology (IT) Departments:

- Account create success
- Account termination success
- Active directory performance index
- Alert-to-ticket ratio
- Average data center availability
- Call center PBX availability
- Campus PBX availability
- Customer connection effectiveness
- Data center capacity consumed
- Email client availability
- Exchange server availability
- Incidents from change
- Internet proxy performance
- Network availability - High availability sites
- Network availability - Standard sites
- Network manageability index
- No problem found/duplicate tickets
- Percentage of branch office backup success
- Percentage of circuits exceeding target utilization
- Percentage of IT managed servers patched at deadline
- Percentage of production servers meeting software configuration standards
- Percentage of security update restarts within maintenance window
- Percentage successful remote access server (RAS) connections
- Phone answer service level
- Priority 1 and priority 2 network incidents meeting SLA
- Product adoption status and compliance
- Restore success rate
- Server growth rate
- Server manageability index
- Service desk client satisfaction - Percentage dissatisfied
- Service desk tier 1 resolution rate
- Service desk time to escalate
- Service desk time to resolve
- Storage utility service availability
- Storage utility utilization
- Virtual machine provisioning interval
- Virtual server utility availability
- Web server availability

5.7.6 Customer Service

Example KPIs for Customer Service Departments

- Agent's full-time employees (FTEs) as percentage of total call center FTEs
- Answering percentage (number of sales calls answered/total number of sales calls offered)
- Average after-call work time
- Average number of calls/ service request per handler
- Average queue time of incoming phone calls
- Cost per minute of handle time
- Costs of operating call center/ service desk
- Email backlog
- Field service technician utilization
- Hit rate (products sold compared to total received sales calls)
- Inbound abandon rate
- Inbound agent dialed calls
- Inbound availability rate
- Inbound average talk time
- Inbound average wrap time
- Inbound call center leads created
- Inbound call center opportunities created
- Inbound calls handled
- Inbound calls handled per agent hour
- Inbound service level
- Number of complaints
- Percentage of customer service requests answered in given timeframe
- Percentage of calls transferred
- Total calling time per day/week/month

Chapter 6

Deutsche Bahn Use Case

DB Headquarters has the consulting function for all of DB Subdiaries. DB Headquarters works together with the all data sources to devise data usage strategies. By collaborating with the domain's experts, DB Headquarters develops matching mathematical models and realises the application environment. The data and corresponding analytical results are then prepared for visual presentation and made available for further examination.

At Deutsche Bahn the using of KPI have three functions:

1. General Overview
2. Warning system
3. Decisions making

Below are some of use case in DB:

6.1 IT Spend Analysis KPI

Overview:

The IT Spend Analysis Dashboard analyze the planned vs. actual costs of an IT department. This Dashboard comparison helps us understand how well the company planned for the year and investigate areas with huge deviations from the plan.

Glossary:

Year-to-date (YTD) is a period, starting from the beginning of the current year (either the calendar year or fiscal year) and continuing up to the present day.

Var: Variance or the deviation

Dataset:

This data is part of a series that illustrates how you can Dashboard BI with business-oriented data, reports and dashboards. This is real data from obviEnce (<http://obvience.com/>) that has been anonymized.

This analysis contains 8 entity (table) and every table has attribute.

1. business area
 - Attribute: Business Area, Business Area ID,
2. cost element
 - Attribute: Cost element name, Cost Element Group, Cost Element Sub Group, Cost Element ID
3. country region

- Attribute: Sales Region, Country/Region, Country/Region ID
- 4. date
 - Attribute: Date, Year, Period, Month
- 5. department
 - Attribute: VP, Department
- 6. it-area
 - Attribute: IT Area, IT Sub Area, IT Sub Area ID
- 7. scenario
 - Attribute: Scenario, Scenario ID, ScenarioDescription
- 8. fact_it
 - Attribute: Date, Value, Department, Cost Element ID, Country/Region ID, Business Area ID, IT Sub Area ID, Scenario ID

Solution:

IT Spend Trend Dashboard: YTD IT Spend Trend Analysis page

- Var Plan % by Sales Region
- Var Plan by Month
- Var Plan , Var Plan % and Actual by Business Area and Period
- IT Area

Spend by Cost Element Dashboard: YTD Spend by Cost Elements page

- Plan and Target
- Var Plan % and Var LE3 % by IT Area
- Var Plan % by Sales Region
- Amount by Month and Scenario

Plan Variance Analysis Dashboard

- Variance Latest Estimates
- Var Plan % by Business Area
- Var Plan by Sales Region
- Var Plan % by Month and Business Area

Technical documentation:

Let's make the data model from the 8 entity. Just drag every Primary Key of the entity in power pivot or power bi data model

After that, create a measure in power bi or power pivot

Note: we use the format of table and attribut with DAX language

Extras: Quick Guide DAX

Below are the explanation of every graph:

6.2 Human Resources KPI

Overview:

Human Resources (HR) dashboards aggregate and present employee data in a meaningful way. As a metrics, Human Resources dashboards simplify information gathering, and present data in a way that can be sorted,

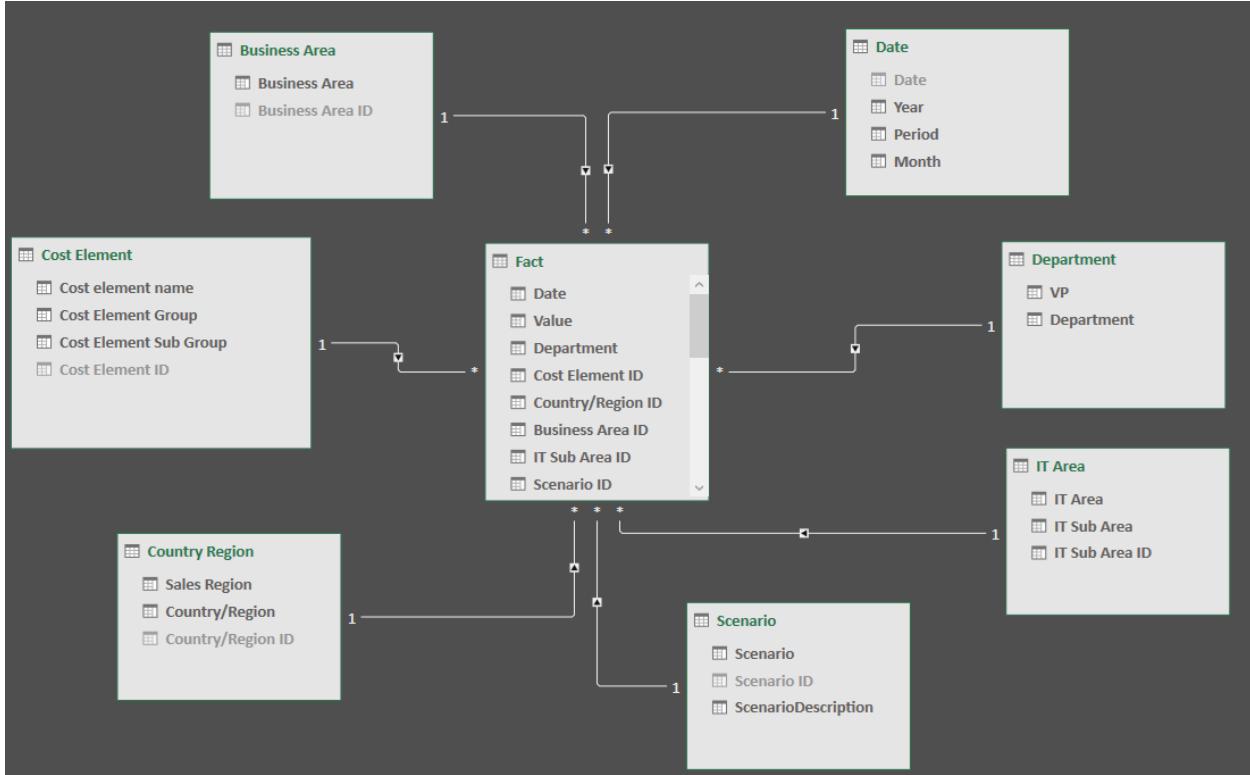


Figure 6.1:

analyzed, and presented. This use case dashboard looks at new hires, active employees, and employees who left and tries to uncover any trends in the hiring strategy. Our main objectives are to understand:

- Who we hire
- Biases in our hiring strategy
- Trends in voluntary separations

Glosarry:

SPLY: same periode last year. One year back in time from the context dates

Dataset:

This Human Resources data is part of a series that illustrates how you can Dashboard BI with business-oriented data, reports and dashboards. This is real data from obviEnce (<http://obvience.com/>) that has been anonymized.

This dataset contains 9 entity (table) and every table has attribute to.

1. age group
 - Attribute: AgeGroupID, AgeGroup
2. gender
 - Attribute: ID, Gender, Sort
3. ethnicity Attribute: Ethnic Group, Ethnicity
4. date Attribute: Date, Month, MonthNumber, Period, PeriodNumber, Qtr, QtrNumber, Year, Day, MonthStartDate, MonthEndDate, MonthIncrementNumber
5. separation reason Attribute: SeparationTypeID, SeparationReason

- 6. pay type Attribute: PayTypeID, PayType
- 7. FP
 - Attribute: FP, FPDesc
- 8. bu Attribute: BU, RegionSeq, VP, Region
 - employee_fact Attribute: date, EmplID, Gender, Age, EthnicGroup, FP, TermDate, isNewHire, BU, HireDate, PayTypeID, TermReason, AgeGroupID, TenureDays, TenureMonths, BadHires

Solution:

New hires Dashboard

- New Hires Vs New Hires SPLY
- New Hires by Region and FPDesc
- New Hires by Region and Ethnicity
- Gender
- AgeGroup

Active Employees vs Separations Dashboard

- YoY Change
- AgeG roup
- Gender
- Actives by Region
- Actives vs Actives SPLY
- Separations by Reason
- Seps vs Seps SPLY

Bad Hires Dashboard

Bad hires are defined as employees who didn't last for more than 60 days.

- Bad Hires by Region and Gender
- Bad Hires by Region and Ethnicity
- Bad Hires YoY % Change by Month and Age Group
- Bad Hire % of Actives and Bad Hires % of Active SPLY by Region

Technical Documentation:

6.3 Opportunity Analysis KPI

Overview :

Opportunity analysis is a detailed review of the prospects within a potential market. The company has 2 sales channels: direct and partner. The Sales Department created this dashboard to track opportunities and revenue by region, deal size, and channel.

The Sales Department relies on two measures of revenue:

- **Revenue** – this is a salesperson's estimate of what he believes the revenue will be.
- **Factored Revenue** – this is calculated as Revenue X Probability% and is generally accepted as being a more-accurate predictor of actual sales revenue. Probability is determined by the deal's current **Sales Stage**.
 - Lead – 10%
 - Qualify – 20%
 - Solution – 40%
 - Proposal – 60%
 - Finalize – 80%

Glossary:

AVG: Average

Dataset :

This data is part of a series that illustrates how you can Dashboard BI with business-oriented data, reports and dashboards. This is real data from obviEnce (<http://obvience.com/>) that has been anonymized.

This dataset contains 5 entity (table) and every table has attribute.

1. sales stage
 - Attribute: Probability, Sales Stage, Sales Stage ID
2. product
 - Attribute: Product Code, Product ID
3. partner
 - Attribute: Partner, Partner ID, Partner Driven
4. opportunity
 - Attribute: Name, Opportunity ID, Rank, SizeID, Opportunity Size
5. fact
 - Attribute: EstimatedCloseDate, Opportunity ID, Sales Stage ID, Account ID, Partner ID, Product ID, ProductRevenue, FactoredProductRevenue, Create Date, Opportunity Days, Year, Month_Number, Month

Solution :

Opportunity Counts Overview

- Opportunity Count
- Opportunity Count by Region
- Opportunity Count by Sales Stage
- Opportunity Count by Partner Driven and Opportunity Size
- Opportunity Count by Partner Driven and Sales Stage

Revenue Analysis

- Revenue by Region
- Revenue by Sales Stage and Partner Driven
- Revenue
- Factored Revenue
- Opportunity Count
- AVG Revenue by Partner Driven and Opportunity Size

Upcoming Opportunities by Month

- Opportunity Count
- Factored Revenue by Opportunity Size
- AVG Revenue by Partner Driven and Sales Stage
- Opportunity Count by Month and Sales Stage

Technical Documentation:

6.4 Data Science HR Turnover

6.5 Data Science Sales Prediction

- [1] Schmidt, Klaus D, 2005, Versicherungsmathematik, Springer Lehrbuch, Springer Verlag
- [2] Beate, Bergter, WS 2013, Formelsammlung Wahrscheinlichkeitsrechnung, HTW Berlin, Deutschland

Chapter 7

Appendix