University Institute of Lisbon

Department of Information Science and Technology

Predictive Analysis for sales

A B2B case

Nelito Cravid Calixto

Dissertation in partial fulfillment of the requirements for the degree of

M.Sc. in Computer Science and Business Management

Supervisor

João Carlos Ferreira

ISCTE-IUL

October 2019

Strategy is about setting yourself apart from the competition. It's not a matter of being better at what you do – it's a matter of being different at what you do.

Michael Porter

Abstract

Measuring salespeople's performance is a process that occurs multiple times per year on a company. During this process, the manager and the salesperson evaluate how the salesperson performed on numerous Key Performance Indicators (KPIs). To prepare the evaluation meeting, managers have to gather data from Customer Relationship Management, Financial Systems, Excel files, among others, leading to a very time-consuming process. The result of the performance evaluation is a classification followed by actions to improve the performance where it is needed. Nowadays, through predictive analytics technologies, it is possible to make classifications based on data. In this work, the author applied a Naive Bayes model to classify salespeople into pre-defined categories provided by the business, through the use of data mining techniques over a dataset of about three years of sales made by 566 salespeople of a global freight forwarder. The classification is done in 3 classes, being: Not Performing, Good and Outstanding, the classification was achieved based on KPI's like growth volume and percentage, sales variability along the year, opportunities created, customer baseline, target achievement among others. The author also identified the most critical factors for salesperson's success based on the dataset as Growth amount, Target achievement, Growth percentage, and the number of months with growth above 0. The author assessed the performance of the model with a confusion matrix and other techniques like True Positives, True Negatives, and F1 score. The results showed an accuracy of 92,10% for the whole model.

Keywords: Data Mining, Predictive Analytics, Sales, Performance Measurement, Human Resources

Resumo

Avaliar a performance de vendedores é um processo que ocorre várias vezes por ano numa empresa. Durante este processo, o gestor e o vendedor avaliam o desempenho do vendedor em vários Indicadores de Performance. Para a reunião de avaliação, os gestores recolhem dados do sistema de Gestão de Vendas, Sistemas Financeiros, ficheiros Excel, entre outros, levando a um processo longo e exaustivo. O resultado da avaliação de desempenho é uma classificação seguida por sugestões de melhoria. Atualmente, através das tecnologias de análise preditiva, é possível fazer classificações com base em dados. Neste trabalho, o autor aplicou um modelo Naive Bayes para classificar os vendedores em categorias predefinidas fornecidas pelo negócio, usando técnicas de data mining aplicados a um conjunto de dados, composto por cerca de três anos de vendas de um transitário global. A classificação é feita em 3 classes, sendo estas: Baixo desempenho, Bom e Fora de Série, a classificação foi alcançada com base em KPI's como a percentagem de crescimento, a variabilidade de vendas entre muitos outros. O autor também identificou os fatores críticos para o sucesso de um vendedor, de acordo com os dados, como sendo volume do crescimento da base de clientes, a capacidade de atingir os objetivos, a percentagem de crescimento e número de meses com crescimento positivo. O autor avaliou o desempenho do modelo com uma matriz de confusão e outras técnicas como True Positives, Negatives, e o score F1. Os resultados apresentaram uma precisão de 92,10 % para todo o modelo.

Palavras-chave: Extracção de Conhecimento em Dados, Análise Preditiva, Vendas, Avaliação de Desempenho, Recursos Humanos

Acknowledgements

During the development of this thesis and dissertation, I had the help and patience of many people. Without them, I would not be able to finish this work.

I want to start by expressing my gratitude to my supervisor Dr. João Ferreira, for all the valuable feedback and guidance through the incredible adventure that was the development of this work. I would also like to thank the management of the company who supported this work, for the data, the time needed, and knowledge shared through this research.

My eternal gratitude for my family and close friends, for the patience and support during all the times the work was challenging.

Nelito Cravid Calixto

Contents

List of Figures

List of Tables

Abbreviations

ETL: Extract Transform and Load

KPI: Key Performance Indicator

NB: Naive Bayes

DM: Data mining

KDD: Knowledge Discovery Database

BI: Business Intelligence

CRISP-DM: Cross Industry Process for Data Mining

ICMS: Incentive Management System

IV: Industry Vertical

SQL: Structured Query Language

CRM: Customer Relationship Management

ERP: Enterprise Resource Planning

IV: Industry Vertical

UOM: Unit of Measure

APAC: Asia Pacific

MEAC: Middle East and Africa

TSQL: Transact SQL

SSIS: SQL Server Integration Services

TMS: Transportation Management System

GPaPS: Gross Proffit After Proffit Share

NFR: Net Forwarding Revenue

HR: Human Resources

OOB: Out of Bag

SVR: Support Vector Regression

HRPA: Human Resource Predictive Analytics

FDMA: Fuzzy Data Mining Algorithm

GM: Grey Model

ELM: Extreme Learning Machine

3F: Fast Fashion Forecasting

ML: Machine Learning

Chapter 1

Introduction

1.1 Overview

Salesperson performance measurement is a process that occurs multiple times per year on a company. The performance evaluation is based on various Key Performance Indicators (KPI's) extracted from multiple systems like Customer Relationship Management (CRM) and Enterprise Resource Planning (ERP).

Evaluating these KPI's can be time-consuming as they require the analysis of figures with complex calculations and require a judgment based on the values, and the weight that each of the KPI contributes to the performance as a whole. The KPI's often include the amount of products/services sold by the salesperson, the number of opportunities created, the ability to sell multiple products/services, the variability of the sales along the year, among many others. When a company has dozens or hundreds of salespeople, this process transforms on a thorough process that may involve other departments like Human Resources (HR) and Operations.

The result of the performance evaluation is a classification followed by actions to improve the performance where it is needed. Technology, through Data Mining (DM), currently is capable of make classification based on data. DM is the process of exploration and analysis, by automatic or semiautomatic means, of large quantities of data to discover meaningful patterns and rules [Koti, 2013]. DM tasks are classified into two categories: descriptive and predictive [Koti, 2013].

In this work, the author proposes the use of DM techniques to allow sales leaders to make a better decision about salespeople performance measurement, by building a model in R that can classify a salesperson's performance based on metrics defined by the business.

1.2 Motivation

The performance evaluation is the process is of assessing various KPI's that reflect a person's performance along a period. The metrics to evaluate vary from function to function. For some functions, most of the parameters are objective; for others, most are subjective.

When measuring salesperson performance, there are objective data, such as total sales increase, sales costs or percent of quota gained, and subjective measures like the manager's assessment of the salesperson [Reday et al., 2009]. The result of the evaluation is a classification based on data that reflect the past. Technology is now capable of make classifications based on data through DM. Algorithms like Naive Bayes (NB) and Random Forest can evaluate several KPI's and return a classification.

In this dissertation, it's evaluated the possibility of classifying salespeople's performance based on KPI's that reflect the past under assessment. The result of the classification can reduce the amount of time spent by management, sales support functions, and the salesperson evaluating all the KPIs to conclude how the salesperson performed.

1.3 Objectives

Predictive analytics is an area increasingly entering the business and academic fields [Waller and Fawcett, 2013]. Companies more and more have been using DM to improve their internal processes and automate not only repetitive, but complex tasks nowadays completed by humans [Ryu and Siegel, 2013] and [Kawas et al., 2013]. Predictive Analytics abilities like classification can read several chunks of data and apply a label to it, and this opens the road to many different abilities. This research seeks to deploy DM techniques to a dataset provided by a freight forwarding company, that contains sales made to their customers for about three years by 566 salespeople and classify how these salespeople performed. The performance evaluation should not only base on their sales amount, but also on other areas where a judgment from a person would be critical, to evaluate the combination of multiple KPI's and decide if the salesperson performed or not.

To achieve this goal, the following are the objectives of this research:

1. To study the salesperson performance evaluation and, predictive analytics for sales and HR in the academic area
2. To examine the current status of the DM platforms available in the market
3. To evaluate and select the methodology and tools to use in the research
4. Through descriptive analysis evaluate the KPIs used to assess salesperson performance
5. To classify salespeople based on many levels selected by the company using predictive analytics

The author will now on this dissertation describe the work done to achieve the above objectives, where the first three are related to the literature revision, the forth is an extensive analysis over the KPIs used to evaluate the salesperson's performance, and the fifth is the classification using predictive analytics.

1.4 Outline of the Dissertation

This dissertation is composed of 8 chapters, including Introduction structured in the following way:

Chapter 2 has a literature review. In this chapter, the author describes the research made in the academic area for salesperson performance measurement and predictive analytics for sales and HR. At the end of the chapter, the author closes with a conclusion, where he wraps the literature review comparing what is available in the academic area with the goal in this work.

Chapter 3 has the background, where DM tools and methodologies to work with data are covered.

Chapter 4 describes the work methodology used, the steps to prepare the data for reporting, and then the author makes an extensive descriptive analytics process to asses the KPIs used to classify salespeople's performance. This chapter end's with a classification model and a report with all the KPI's ready to be assessed.

Chapter 5 describes the steps made to build the reports and to evaluate the metrics used to classify the salespeople.

Chapter 6 describes the modeling phase. In this chapter, the author details all the steps executed to clean a new dataset generated only with KPIs prepared to build the model. The author also describes the model results and the impact it may have on the predicted data.

Chapter 7 has a description of all the evaluation steps executed to assess the model performance and the identification of the most critical factors for a salesperson to succeed.

Chapter 8 has the conclusions of the work, starting with the work done and goals, going then to the discussion of the results and future work suggested by the author.

Chapter 2

Literature Review

Here, the author evaluates academic researches about the subjects used in this research. Starting with the salesperson performance evaluation, then what have been doing with predictive analytics for sales, moving later to predictive analytics in human resources, closing the chapter with a conclusion of the literature review.

2.1 Review Protocol

2.1.1 Background

Salesperson performance measurement is a process that occurs multiple times per year on a company. The performance evaluation is based on various KPI's extracted from various systems like CRM and ERP. Evaluating these KPI's can be time-consuming as they require the analysis of figures with complex calculations and require a judgment based on the values, and the weight that each of the KPI contributes to the performance as a whole. Although these might be a reliable way of doing performance evaluation, this process has been evolving in the last years, and PA has been used in many areas on the companies, from predicting sales to recruiting and performance evaluation in human resources. Therefore, there is a need to undertake a systematic review of available methods in the academic world about how PA is being utilized.

This protocol goal is to present the scope of our literature review, and here we will specify all the criteria needed for literature to be included in our research.

2.1.2 Objectives

The objectives of the review are to:

- Undertake a systematic review of empirical research on DM applied to sales and human resources.
- Select a sub-set of studies to review in-depth.
- Synthesize the evidence from these studies about how PA is used in the companies.
- Identify the research methods used.
- Identify any gaps in current research to suggest areas for further investigation.
- Provide a framework/background in order to position new research activities appropriately.

More specifically, we want to find out:

1. What is currently known in the sales based on ML
2. What is currently known in the HR based on ML

In addition to producing substantive findings regarding sales and HR using Machine Learning, the review also aims to advance a methodology for integrating diverse study types, including 'qualitative' research, within systematic reviews DM.

2.1.3 Criteria for considering studies for this review

Types of studies:

Articles to be included in the review must present empirical data and pass the minimum quality threshold described below.

Types of participants:

Studies of both students and from professional data scientists will be included.

Types of intervention:

Inclusion of studies will not be restricted to any specific type of intervention.

Types of outcome measures:

Studies must include detailed specs and steps of how to replicate the results with similar data.

2.1.4 Search strategy for identification of studies

The search strategies will include electronic databases and hand searches of conference proceedings.

Electronic databases:

The following electronic databases will be searched: IEEE Xplore, ACM Digital Library, Science Direct – Elsevier, Springer, Emerald Publishing, and International Journal of Scientific & Technology Research.

Conference proceedings:

The following conference proceedings will be hand searched for research papers: IEEE International Conference on Data Mining (ICDM), Transforming Data With Intelligence (TWDI), Association for the Advancement of Artificial Intelligence (AAAI)

Search strategy:

The title, abstract, and keywords of the articles in the included electronic databases and conference proceedings will be searched according to the following search strategy:

1. Data AND Mining
2. Predictive AND Analytics
3. Machine AND Learning AND Human AND Resources
4. Predictive AND Analytics AND People
5. Predictive AND Analytics AND Classification
6. Predictive AND Analytics AND Sales
7. Predictive AND Analytics AND B2B
8. Performance AND Evaluation AND Human AND Mining
9. Companies AND Predictive AND Analytics AND Competitive AND Advantage
10. Predictive AND Analytics AND Sales
11. Machine AND Learning
12. Predictive AND Analytics AND HR Mining

2.1.5 Methods of the review

Selection of studies:

The search strategy will identify all relevant articles. Only studies written in the English language are considered. Studies will be excluded if:

(i) the study's focus, or main focus, is NOT Data Mining or NOT Sales or NOT HR-related;

> If it is unclear from the title, abstract, and keywords, whether a study conforms to these screening criteria, it will be included for a detailed quality assessment.

The nine criteria cover three main quality issues that need to be considered when appraising the studies of this research:

- rigor: has a thorough and appropriate approach been applied to key research methods in the study?
- credibility: are the findings well presented and meaningful?
- relevance: how useful are the findings to the software development industry and the research community?

Two screening criteria relate to the quality of the reporting of a study's rationale, aims, and context. Each study will be assessed according to whether:

(i) the aims and objectives are reported (including a rationale for why the study was undertaken);

(ii) there is an adequate description of the context in which the research was carried out.

A further five criteria related to the rigor of the research methods employed to establish the validity of data collection tools and the analysis methods, and hence the trustworthiness of the findings. Each study will be assessed according to whether:

(iii) the research design was appropriate to address the aims of the research

(iv) there is an adequate description of the sample used and the methods for how the sample was identified and collected;

(v) any control groups were used to compare treatments

(vi) appropriate data collection methods were used and described;

(vii) there is an adequate description of the methods used to analyze data and whether appropriate methods for ensuring the data analysis was grounded in the data.

Also, two criteria relate to the assessment of the credibility of the study methods for ensuring that the findings are valid and meaningful. In relation to this, the reviewers will judge studies according to whether:

(viii) the study provides clearly stated findings with credible results and justified conclusions.

The final criterion relates to the assessment of the relevance of the study for the software industry at large, the Freight Forwarding industry and the research community. Thus, the reviewers will judge studies according to whether:

(ix) they provide value for research or practice.

Taken together, these nine criteria provide a measure of the extent to which we can be confident that a particular study's findings can make a valuable contribution to this review.

Now, with the rules defined above, the author describes the findings in the areas of Salesperson Performance and Predictive Analytics applied to sales and HR.

2.2 Salesperson Performance

Academic studies demonstrate that the success of a salesperson normally has a direct relationship with company performance, [Rich et al., 1999] states that: "When salespeople do well, the organization is likely doing well, and the contrary is normally true as well.". When measuring salesperson performance, there are objective data, such as total sales increase, sales commissions or percent of quota, and subjective measures like manager's or peer's assessment of the salesperson [Reday et al., 2009]. Many companies use a combination of objective and subjective KPI's to make the assessment. A meta-analysis of objective and subjective sales indicators suggests that there is a low correlation identified between objective and subjective sales success indicators, which suggests that these indicators are not necessarily interchangeable and, the choice of the most appropriate may require trade-off [Rich et al., 1999].

2.3 Predictive Analytics for Sales

Authors in the academic area refer that Predictive Analytics (PA) has been used for several years in by companies to get a competitive advantage, [Domingos and Van de Merckt, 2010], [Bose, 2009]. At first, by companies acting in the B2C with a large customer base and capacity to collect and store transactional data from customers, and only then by companies acting in the B2B area [Domingos and Van de Merckt, 2010].

B2B selling companies are hiring cloud-based PA providers to draw on both inside and outside data sources to identify new leads so that they can take advantage of PA [Lilien, 2016].

[Mirzaei and Iyer, 2014] did a comprehensive study on the application of PA over CRM data. The results show 57 articles found in 4 databases, where the studies focused on dimensions like Customer Acquisition, Attraction, Retention, Development, and Equity Growth. The results show that PA techniques between 2003 and 2013 gained a lot of popularity not only in the banking area but also on: casinos, retailers, telecommunications, manufacturing, insurance, and healthcare [Mirzaei and Iyer, 2014].

To understand what has been studied in the academic area in terms of predictive analytics, the author hereunder describes some success cases of PA applied in sales forecasting.

2.3.1 Sales forecasting of computer products based on variable selection scheme and SVR

Like many other industries, sales forecasting is also a challenge for computer product retailers. Wrong forecasts can cause product backlog or inventory shortages, incorrect customer demands, and decrease customer satisfaction [Lu, 2014].

[Lu, 2014] combined Multi Variable Adaptive Regression Spines (MARS) with SVR to make a sales forecasting model for computer products. The combination of MARS scheme to forecast computer products called (MARS VR Scheme). The main idea over the scheme was first to use MARS to select the essential forecasting variables and then use the identified key forecasting variables as the input variables for SVR. The data used was a compilation of the weekly sales data of five computer products from a computer retailer in Taiwan. Products like Notebooks, LCDs, Main Board, Hardrives, and Display cards.

2.3.2 Fast fashion sales forecasting with limited data and time

Another case of success found is applied to fast fashion, which is an industrial practice, where the main idea is to offer a continuous stream of new merchandise to the market [Choi et al., 2014]. With this practice, some fashion companies are even capable of having the products from the conceptual design to the final product in just two weeks. Companies working with this practice have to make their inventory decisions based on a forecast, with short lead time and a tight schedule. The result is companies making a forecast on a near real-time basis and with a minimal amount of data. [Choi et al., 2014] proposed an algorithm called

Fast Fashion Forecasting (3F) that give the companies the ability to make forecasts with limited data and time. This algorithm uses two artificial intelligence methods: Extreme Learning Machine (ELM) and the Grey Model (GM). The data used belonged to a knitwear fashion company using a fast-fashion concept. The data included real sales collected used in the analysis. Products are categorized into different types, and there are multiple styles and colors. [Choi et al., 2014] tested the algorithm with real and artificial data, and results revealed an acceptable forecasting accuracy with the 3F algorithm.

2.3.3 Support Vector Regression for Newspaper/Magazine Sales Forecasting

The next case is in the media area, where due to the constant transformations that information technologies are bringing to the world, new generations are more and more used to grow up used to browse the internet for news and exciting stories [Yu et al., 2013]. With that in mind, the media industry also has to evolve to keep up with the evolution. Because of that, it is more urgent for traditional media companies to make an accurate forecast on printing newspapers and magazines, to avoid excessive printing or not meet the expected demand [Yu et al., 2013]. The study [Yu et al., 2013] made used SVR in a media company with printed newspaper/magazines to create a sales forecast that estimate and prepare the prints plan and distribution. The results of the study showed that demographic characteristics as gender, age, income, education, and occupation distributions affect the newspaper and magazine sales. In particular, the percentage of people with income between 100.000 and 124.000 has a stronger influence on sales. This experiment also showed that SVR is a superior method in forecasting sales for the news/magazines industry [Yu et al., 2013].

With these scientific articles about success cases of PA in the B2C, we move next to success cases in the B2B area.

2.3.4 On Machine Learning towards Predictive Sales Pipeline Analytics

On companies operating in B2B, new sales are often identified as Leads. These leads move then into the Sales Opportunity Pipeline Management System. Later on, some of these Leads are qualified into opportunities. A sales opportunity is a set of none, one, or several products or services that the salesperson is trying to convert into a purchase. All the Opportunities are tracked, ideally ending on a won business that generates revenue for the company [Yan et al., 2015].

A fundamental part of the pipeline quality assessment is the lead-level win-propensity score, identified as the win-propensity. The salesperson usually enters these scores, to avoid noise inserted by the salesperson for various reasons and biased scores, [Yan et al., 2015] proposed a model to calculate the win-propensity using the Hawkes process model [Yan et al., 2015]. The proposed method was deployed on a multinational Fortune 500 B2B-selling company with success in 2013.

2.3.5 Prescriptive Analytics for Allocating Sales Teams to Opportunities

Still, in the Opportunities,[Kawas et al., 2013] used Predictive and Prescriptive Analytics to increase the revenue of a company by 15%. By automating the allocation of sales resources to opportunities to maximize opportunities revenue in B2B selling for the company.

For Predictive [Kawas et al., 2013], mined the historical selling data to learn sales response functions that have the behavioral relationship between the size and composition of a sales team, and the revenue earned for the different types of customers and opportunities using multiple linear regression.

For Prescriptive, [Kawas et al., 2013] used the sales response functions to determine the allocation of salespeople's effort to the customer's opportunities that maximize the overall revenue earned by the salespeople, using a piece-wise linear approximation.

With this base for PA for sales, the author now moves to the application of PA to HR. HR is essential in this work because the goal is to return the salesperson performance evaluation as a form of classification.

2.4 Predictive Analytics in HR Management

2.4.1 In General

[Malisetty et al., 2017] on an article published in 2017, propose the use of PA in HR for:

- Employee Profiting and segmentation, by anticipating the standing of every employee to profit from learning opportunities or capitalize on new undertakings;
- Employee Attrition and Loyalty Analysis, using predictive risk models to predict potential loss of employee and, by combining attrition risk score with worker performance info, HR can distinguish high-performing employees and also reduce potential attrition;
- Forecasting of HR Capacity and Recruitment Needs, using PA to anticipate the recruiting needs by combining the gap between people to recruit and people already employed, allowing HR to avoid under and over employment;

The author also proposes research in Appropriate Recruitment Profile Selection, Employee Sentiment Analysis, and Employee Fraud Risk management.

[Sujeet N. Mishra, 2019] propose the use of Human Resource Predictive Analytics (HRPA) for decision making by presenting two cases of success: One in a US wind turbine maker that changed the recruitment and retaining policies based on HRPA;

Another is at Cisco, which used IBM SPSS to transform the relationship between its HR analysis and executive leaders. [Kessler et al., 2007] presents the categorization module of E-Gen, a modular system to treat job listings automatically. Through SVM [Kessler et al., 2007] managed to rank candidate responses based on several information. [Menon and Rahulnath, 2016] used machine learning techniques to rank candidates on a recruiting process by analyzing the candidate adaptability to a job position based on the candidate tweets. [Faliagka et al., 2012] proposed an approach to evaluate job applications in online recruitment systems, to solve the candidate ranking issue. By analyzing the candidate's Linkedin profile and infer their personality characteristics using linguistic analysis on the candidate blog profile. This was achieved by using training data provided by human recruiters and applied in a large-scale recruitment scenario with three different positions and 100 applicants using Regression Tree and SVR.

2.4.2 For performance evaluation and analysis

[Zhao, 2008] proposed a method of DM for performance evaluation, for that they gathered information about Ability, Attitude, Performance, Harvest, and Spirit in a dataset. Then they used the K-Expectation algorithm to classify employees into the same group. After that, a Decision tree is used to train a model based on rules that can be used by managers to classify and select the best employees from the applicants.

[Jing, 2009] applied Fuzzy Data Mining Algorithm (FDMA) for performance evaluation of human resources. For that, the author used evaluation records with four features: innovation ability, learning level, work efficiency, independence, and workability, and each of these had four levels, which are the corresponding score of each feature. Then [Jing, 2009] used the maximal tree to cluster the human resource, with that the next step was to compare the data from management with each cluster and calculate the proximal values based on the FDMA, the last step referred to determine the evaluation. The evaluation, in this case, was a result closer to each of the 4 clusters that are named as Best, Better, General, and worse.

[Xiaofan and Fengbin, 2010] applied Decision Trees on Performance Analysis of human resources to make classification Analysis. The results show that there are mutual restraint and influence between performance results and working quality, tasks, skills, and attitude. Concluding that if the enterprise in the future cultivates employee working skills and quality, the employees consciously improve themselves in these areas.

2.5 Conclusion

The success of a salesperson usually has a direct relationship with company performance [Rich et al., 1999]. To evaluate a salesperson's performance, there are objective and subjective measures, like total sales increase, sales commissions, or percent of quota and subjective measures like managers' assessment [Reday et al., 2009]. Many have been made in the academic area in PA for sales. During this Literature Review, several cases of success in PA were covered, using different algorithms and in industries like computer sales, fashion, or media [Lu, 2014], [Choi et al., 2014], [Yu et al., 2013]. The author was also able to cover some cases using CRM data, specifically in the B2B area, where Opportunities scores and allocation were automated using PA [Yan et al., 2015], [Yan et al., 2015] and [Kawas et al., 2013]. In the HR field, PA is proposed in multiple cases for the recruiting area, data analytics, and performance evaluation [Malisetty et al., 2017], [Sujeet N. Mishra, 2019], [Kessler et al., 2007], [Menon and Rahulnath, 2016], [Zhao, 2008], [Jing, 2009], [Xiaofan and Fen and [Faliagka et al., 2012]. As in this work, the goal is to classify salespeople using PA the author now follow confident that the goals should be achievable, using DM and PA.

In this research, the author has to apply DM to a dataset and classify salespeople. Only with the difference that most of the metrics are objective and not subjective as used in some of the works covered in this literature review, and the result should be a classification text and not a number.

Chapter 3

Background

3.1 Data Mining Platform

3.1.1 Machine Learning

Machine Learning (ML) is the study of statistical models and algorithm's, that computer systems use to improve their performance on a task. The algorithms used in ML build mathematical models of sample data that allow the computer to make predictions or take decisions without explicit programming. These algorithms are nowadays used on various applications like: email spam filtering, computer vision, fraud detection, user purchase suggestions among others. ML is also highly related to computational statistics focused on making predictions [Kotsiantis et al., 2006].

Machine Learning (ML) can be made Supervised or Unsupervised. According to [Kotsiantis et al., 2006], every instance in a dataset that is used by a ML algorithm is represented using the same set of features. These features can be categorical or binary. [Kotsiantis et al., 2006] also states that "If instances are given with known labels (the corresponding correct outputs) then the learning is called supervised, in contrast to unsupervised learning, where instances are unlabeled". [Kotsiantis et al., 2006] also states that "There are several applications for Machine Learning (ML), the most significant of which is data mining."

3.1.2 Data Mining

Researchers define DM as an interdisciplinary sub-field of computer science that involves the computational process of large data sets patterns discovery [Koti, 2013]. With the goal of extract information from data sets and put it on forms, a human can extract knowledge from it. The same researcher states that the methods used are at the juncture of artificial intelligence, machine learning, statistics, database systems, and business intelligence [Koti, 2013].

3.1.3 Knowledge discovery in Database KDD

To extract knowledge from the data, [Fayyad et al., 1996] proposes the use of the knowledge discovery in databases (KDD) process. In the KDD process, the user follows a process that [Fayyad et al., 1996] describes as an interactive and iterative process, involving numerous steps with many decisions made by the user. The steps and sequence without the potential multitude of iterations and loops

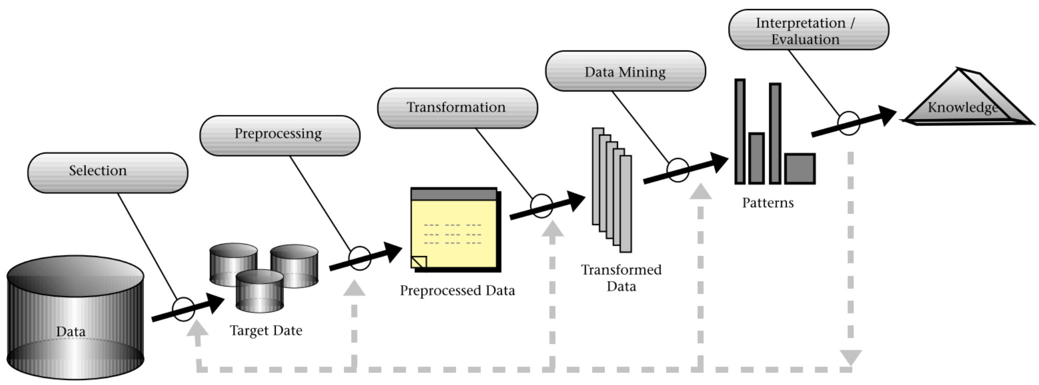

FIGURE 3.1: An overview of the steps that compose the KDD process extracted from [Fayyad et al., 1996]

illustrated in Figure 3.1 shows that multitude. This author also mentions that most of the previous steps focus on the DM step, but has the same importance and sometimes may have more [Fayyad et al., 1996].

The steps for the KDD process described by [Fayyad et al., 1996] are:

The first step is to understand the application domain, the previous knowledge, and identify the goal of the KDD process from the customer's point of view.

The second step is to create a target dataset, by selecting a dataset, or focus on a subset of variables samples of data, where the findings should happen.

The third step is composed of cleansing and pre-processing tasks. These tasks include the removal of noise if needed, and gather the necessary info to model. It is also on this step steps that missing data is processed, and accounting for time-sequence information is set.

The fourth step is where the project reduces data by finding useful features to represent the data. On this step, the final number of variables can reduce.

The fifth step is where a match of the KDD goals to a specific DM method happens. These can be summarization, classification, regression, clustering, and so on [Pitre et al., 2014].

The sixth step is where exploratory analysis, model, and hypothesis selection is selected. In this step, the DM algorithm and methods are chosen to find patterns in the data [Fayyad et al., 1996].

The seventh step is where DM happens. The DM steps include the use of classification rules or trees, regression, and clustering. It is possible to significantly aid the DM method by correctly performing the preceding steps.

The eighth step is where the interpretation of the patterns mined is made. In this stage, the user can return to any of the previous actions. Another thing that can be necessary to do on this step involves visualization of the extracted pattern, models, or visualize the data provided by the derived model.

The ninth and last step is where the user acts on the discovered knowledge — using the extracted knowledge on another system to take actions, document, or report to the interested parties.

3.2 Data Mining Tools

By part of this work, the author queried the internet for DM tools and found a total of 89 tools where 68 are proprietary and 21 Open Source. Gartner that is a consultancy company known for their magic quadrant, that provides a wide-angle of the relative positions of the market's competitor. The magic quadrant helps a person quickly identify who are the challengers, leaders, niche players, and visionaries. They built a quadrant to identify the most used data science tools, which include DM tools. In the quadrant, it's possible to see that the leaders for 2018 are: Alteryx, SAS, Rapid Miner, KNIME and H20.ai [Idoine et al., 2018].

FIGURE 3.2: Gartner 2018 Magic Quadrant for Data Science and Machine Learning Platforms, extracted from: [Idoine et al., 2018]

KNIME is an open source analytics platform with more than 100k users around the world. This platform also offers commercial support and extension for security, collaboration, and performance for enterprise deployments. In 2017 this platform added cloud versions for its platform for Amazon Web Services, Microsoft Azure, and also deep-learning capabilities [Idoine et al., 2018].

Alteryx, is a platform that allows users to build models on a single workflow. It includes functionalities for ETL, data quality scoring, data governance, enrichment with spatial and web resources among other features [Idoine et al., 2018].

SAS Enterprise Miner, allows users to create descriptive and predictive modeling with an interactive self-documenting process flow diagram environment [Idoine et al., 2018].

RapidMiner platform includes the RapidMiner Studio development tool. The Rapid Miner studio is a visual design environment that allows the user to build predictive analytic workflows rapidly. It provides a big library of machine learning algorithms, data preparation and exploration features and model validation tools [Idoine et al., 2018].

H20.ai, H20, is an open source application for big data analysis. H20.ai can automate feature engineering, model building, visualization, and interpretability of data. The platform works with R, Python, Scala on Hadoop/Yarn and Spark [Idoine et al., 2018].

Other major applications in the markets are:

IBM bluemix, IBM® Bluemix® Data, and Analytics are a PaaS. This platform provides a big set of features for Data preparation, data warehousing, and analytics. It also allows fully managed databases (NoSQL and open source, and last fully managed apache add hoop development) [IBM, 2018].

Amazon Web Services (AWS) delivers a set of tools that allow users to Import data from on-premises to the cloud, store data securely, from gigabytes to terabytes. It also includes the ability to extract, transform, and load to prepare data for analysis. For predictive analytics, AWS can use Machine Learning services and run tools on data stored on AWS. AWS provides Deep Learning capabilities to build deep learning models and build clusters; it also supports machine learning frameworks like Apache MXNet, Tensor Flow, and Caffe2 [Amazon, 2019].

Weka is an application that contains a collection of visualization tools and algorithms for data analysis and predictive modeling. Built over Java and open source, which allows it to run on multiple platforms and free of cost. Weka supports several DM tasks like clustering, classification, regression, and visualization [Roulston et al., 2017].

Another tool massively used in the Market is RStudio. This tool is an Integrated development environment, for R, that supports direct code execution and tools for plotting, history, debugging and workspace management. This tool is available in open source and commercials edition. The free version has features like, local access, syntax highlight, execution of R code directly from source editor among other features [RStudio.com, 2019].

3.2.1 Power BI

Power Bi is an analytics service provided by Microsoft. With this service, a user can easily create interactive visualizations that he can then incorporate on reports or dashboards — all these without being dependent on technology staff.

Power BI supports Self Service BI (SSBI) Concept. According to [Dw I R E S E et al., 2011] the definition for Self Service BI is: "The facilities within the BI environment that enable BI users to become more self-reliant and less dependent on the IT organization. These facilities focus on four main objectives: easier access to source data for reporting and analysis, easier and improved support for data analysis features, faster deployment options such as appliances and cloud computing, and simpler, customizable, and collaborative end-user interfaces".

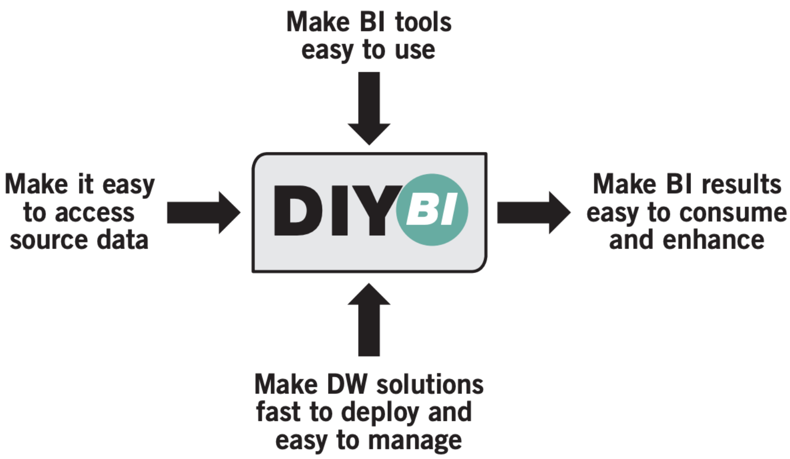

FIGURE 3.3: The four objectives of do-it-yourself or self-service BI extracted from [Dw I R E S E et al., 2011]

According to [Dw I R E S E et al., 2011] as represented in Figure 3.3 the 4 objectives of the Self-Service Business Intelligence are:

- Make BI Results Easy to Consume, as stated in [Dw I R E S E et al., 2011], SSBI must be an environment in which it is easy to discover, access, and share information, reports, and analytics. The users have to be able to create personalized dashboards or have automated BI capabilities.
- Make BI tools easy to use, meaning that not only the data need to be easy to consume, but the tools to generate the dashboards and reports also need to be easy to use.
- Make Data Warehouse Solutions Fast to Deploy and Easy to Manage. These may include the use of alternative mechanisms to reduce costs and increase time to value and data processing. For that, users may have to use Agile methodologies, Software as a service, cloud offerings, and analytical Database Management Systems. Another concept is the opening of BI to the business community.
- Make Data Sources Easy to Access, one of the main differences from traditional BI and SSBI is that in conventional BI, all the data is stored in the data warehouse. In SSBI data from other sources can and may be necessary to be available for access by the business community without the need of IT staff. The job of the BI implementation staff is to create an infrastructure that permits the flow of data from these sources. According to [Dw I R E S E et al., 2011] the BI team: monitor access and utilization of the data, ensure the environment's optimal performance, implement appropriate security and privacy procedures, and provide support to the business community where needed in the construction or publication of BI reports, analytics.

3.3 Methodology

3.3.1 CRISP-DM

CRISP-DM means the cross-industry process for DM. This methodology provides a structured way of planning a DM project. The author describes below the CRISP-DM process according to the: [SMART Vision, 2018]. The Figure: 3.4 illustrates the process.

The first stage of the methodology is understanding the goals from a business perspective. At this stage, important factors that can influence the outcome of the project may be uncovered. The project outputs and followings are defined:

- objectives from the business perspective
- project plan
- business success criteria

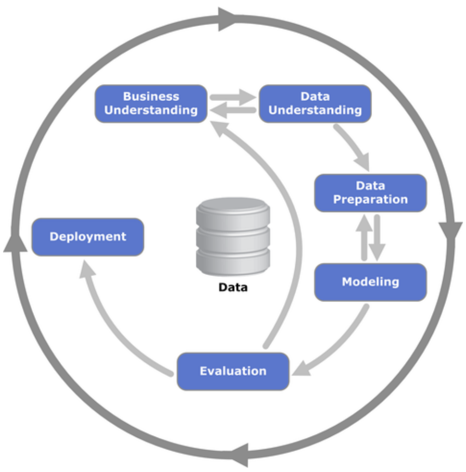

Figure 3.4: CRISP-DM Methodology [Chris Manna, 2019]

The second stage of the methodology requires data acquisition and, if necessary, data loading.

The third stage is data preparation; here, tasks are: select the data, attributes, and records that meet the criteria needed to achieve project goals. Another job on this stage is cleaning the data, where it's possible to raise the data quality to the required by the analysis techniques level, data construct, and integration is also done in this stage.

The fourth stage is Modelling. Here the modeling technique an test design are selected. The next task is to build the model, where the modeling tool is executed to prepare the dataset for one or more models. After making the model, the next task is to assess the model, where the model is interpreted according to the domain knowledge by business analysts or domain experts.

The fifth stage is to evaluate the results and assess factors like accuracy and generality of the model. This step is where the author evaluates the degree to which the produced model meets the business objectives. The next phase of this stage is reviewing the process, where a more thorough review of the DM engagement is made, in order to determine if there are important factor or task that might be overlooked. The last task on this stage is determining the next steps, where depending on the results of the assessments and revision, the decision of how to proceed forward is made, like project finish and move for deployment [SMART Vision, 2018].

The last stage is the deployment stage, where the author starts by evaluating the results and determine the strategy for the deployment. The next task is to plan monitoring and maintenance, where a careful preparation of a maintenance strategy is placed to avoid unnecessarily long periods of incorrect usage of the results. The next task is to produce a final report, where, depending on the deployment plan, the report can be only a summary of the project or a final and comprehensive presentation of the DM results. The last task of the project is Review Project, where an assessment of what went wrong, right, what was done well, and needed improvement [SMART Vision, 2018].

As the background chapter is concluded, the selected methodology is CRISP-DM, because it's a proven methodology being used in research and in the marked on multiple research projects. The tools selected are, Power BI, because the company has the tool, the license and the business people has the knowledge to work with this tool. The other selected tool is RStudio, this tools is the selected one for modeling, for it's free open source and has many packages online with all the required models for this research. The author now proceed to concluding the remaining objectives.

Chapter 4

Work Methodology

In the previous chapter, the author concluded objectives 1, 2, and 3. These objectives were related to the literature review regarding salesperson performance evaluation, Data Mining Platforms, and methodology. The author follows now to the objective: Through descriptive analysis, evaluate the KPIs used to assess salesperson performance. For that, the author applied a data analytics process to the dataset provided by the company with multiple data treatment steps. After concluding these steps, the author creates the requested reports. Hereunder are now described all the steps executed to achieve the goal.

The work methodology used in this research was based on CRISP-DM. As presented in Figure: 3.4, the author starts with the Business Understanding, describing the Business Objectives, moving then to the Data Understanding, where the author describes and explore the data, the next stage is Data Preparation where data is selected, cleaned and integrated.

Before the modeling stage, the author created a set of Power BI reports so that together with the business, the author can find the 10 best salespeople worldwide and for each region. The last step of this analysis is modeling the data with the NB algorithm to build a model capable of classifying salespeople on a dataset. This leads us to our Research Steps displayed in Figure: 4.1.

As presented in the Figure: 4.1 the research steps are:

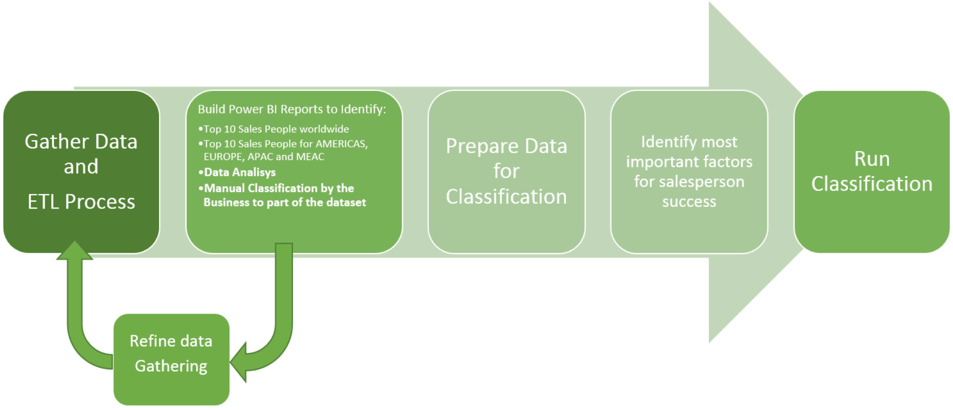

FIGURE 4.1: Research steps

1. Gather Data and ETL Process, where most of the steps of CRISP-DM are done, on this phase the author developed a SQL Integration Services Project to Extract, Transform and Load all data provided by the company to a new database prepared to be loaded to Power BI

2. Build Power BI Reports to Identify best salespeople and classify them, on this stage, the business with the help of the author classified a portion of the salespeople in the dataset, this was made by:

    - First the author together with the business build the reports
    - Second the reports are evaluated and analyzed by the business
    - Third the Business classified a subset of salespeople
    - Last the classified list of salespeople is returned to the author so that he can classify remaining dataset with predictive analytics

3. Data is then prepared for Classification

4. The most important factors for salesperson success is identified

5. Classification is done, and it's evaluated by the algorithm metrics

4.1 Business Understanding

4.1.1 Objectives

With the main goal of classifying salespeople, and build a model that can tell if a salesperson is successful or not, this research project has the following business objectives:

- Through data analysis and discovery find the best salespeople in the company
- Identify the factors that contribute to the success of salespeople
- Use predictive analytics process to classify salespeople

4.1.2 Business success criteria

The main success criteria for this project is the ability to achieve the specific goals defined previously on the objectives. To evaluate these goals, the author used the metrics provided by algorithms that measure the accuracy of the classifications.

4.2 Data Understanding

The data used in this research refers to sales between Jan 2017 and June 2019, from a freight forwarding company that operates worldwide on Air, Ocean, and land. The data refers to shipments for the customers grouped by month. There are 1.982.969 rows and 49 columns on the main table. Each row represents the shipments made to one customer of one product moved from one place to another in one month of one year.

As presented in Figure 4.2, every year, the company establish contracts with their customers, where both companies agree on the shipments sources, destinations,

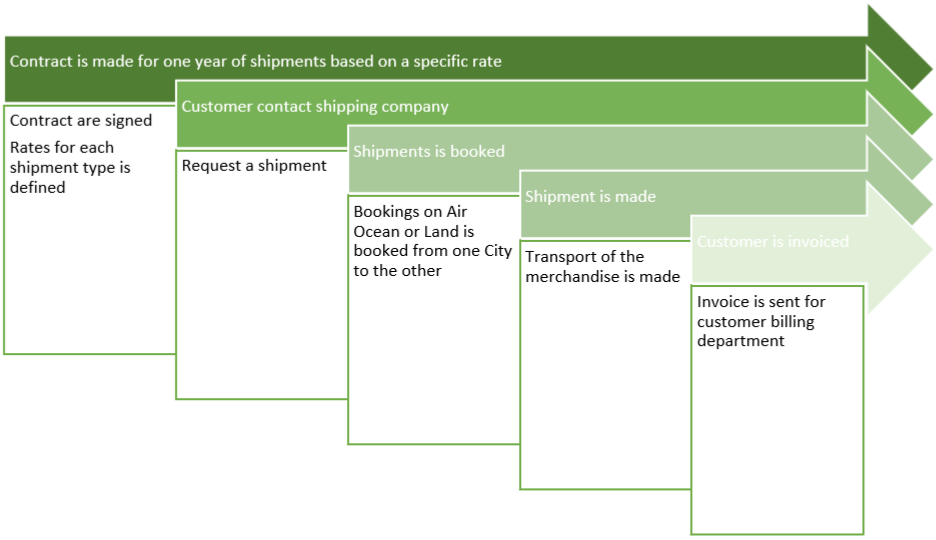

FIGURE 4.2: Year base overall sales process

and transport type. After this contract is established, Shipments can be requested by the customer to the freight forward company.

One shipment is the act of shipping merchandise from one place to the other, and these can be by Air, Ocean, and Land. Air shipments are made through charted plain or by hiring the charter service from another company.

To make the shipments by the ocean, this company hires third-party freight carriers that operate on the ocean, but this company makes most or all the transport documentation required by law for the shipment. The shipments by land can be made by hiring external companies or by using the company assets, both systems are possible.

All the customers on the dataset are companies, and no shipments are made to singular persons.

The data goes from the ERP to the ICMS as displayed in Figure: 4.3, going through the following path: whenever a customer requests a shipment, it is recorded on the SAP TM by the freight forwarding operator. An ETL process transfers it then to the data warehouse platform. After all the validations occur, the shipment goes to the ICMS. The dataset used in this research is a subset of this database.

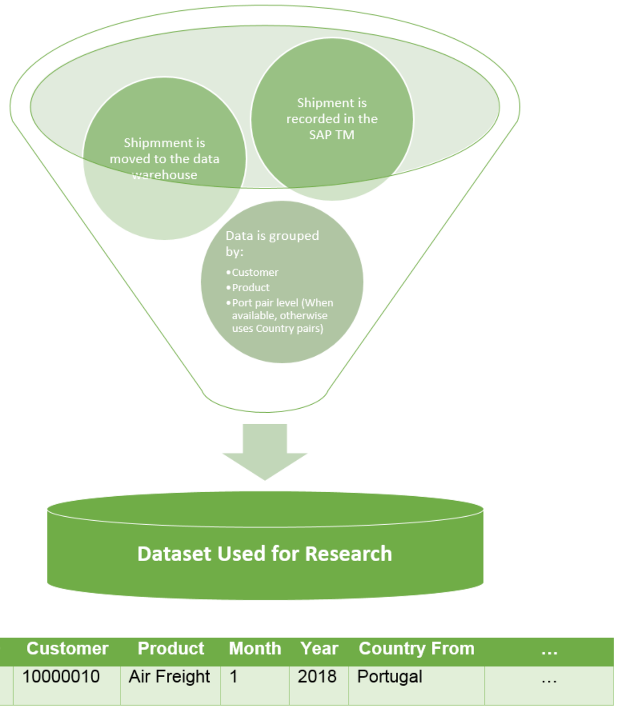

FIGURE 4.3: One row representation

According to the defined Research Steps, the author hereunder describes the first stage "Gather Data and ETL Process":

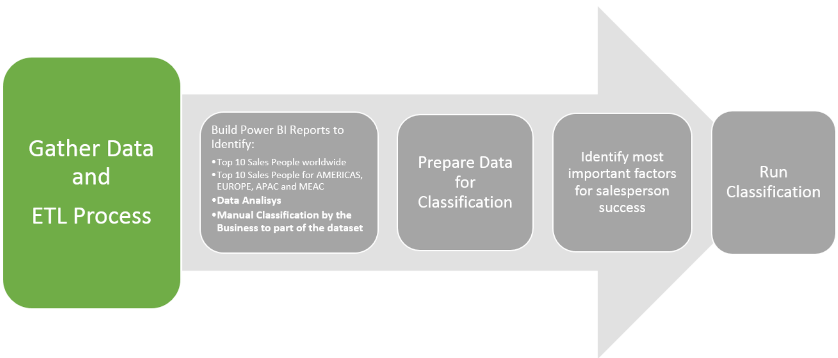

FIGURE 4.4: Research steps ETL

4.2.1 Data description

The database engine used for this analysis is: Microsoft SQL Server 2014 (SP3-CU4), the database provided by the company contains the following nine tables:

- Sales_Person_Details (2 tables, 1 table for Current and 1 table for History data)
- Payout_Exceptions (2 tables,1 table for Current and 1 table for History)
- Products
- CRM_Opportunities
- Targets
- Geography
- Currency

A demonstration of the database structure is provided in Figure: 4.5.

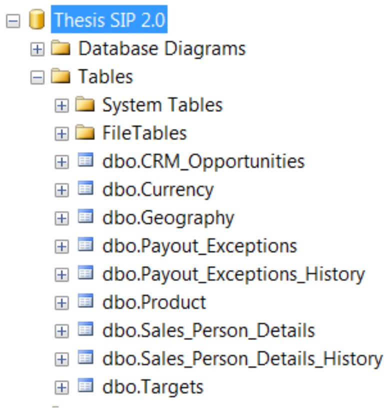

FIGURE 4.5: Source tables

For simplicity, the author provides all details regarding the functional meaning of the tables and columns in the Appendix: .1.

Salesperson Details (Current and History) has a lower level of our data. One row of these tables refers to all the shipments made in one month, for one product, shipped to one customer, to one city pair (source and destination country and city), and sold by one salesperson. This table is composed of 49 columns and has 1.982.969 rows, more information in Appendix: .1.1.

Payout Exception is the table that holds the exceptions. Due to several factors like errors in operations, some shipments are not recorded correctly in the Transportation Management Systems (TMS). When these cases occur, the Incentive

Management System (ICS) is ready to accept what is called exceptions, where the missed volume is recorded. This table is composed of 19 columns and holds 3.661 Rows, more details in Appendix: .1.2.

CRM Products is a table that holds the services provided by the company as products. This table is composed of 5 columns and contains 54 Rows, more information in Appendix: .1.4.

The CRM Opportunities table holds the opportunities created by the salespeople for the sales made to their customers. One Opportunity can only refer to 1 customer but may have multiple city pairs. This table is composed of 13 columns and holds 142.016 Rows, more information in Appendix: .1.3.

The Targets table stores the targets for each product, salesperson, year, and month. This table is composed of 13 columns and holds 127.932 Rows, more information in Appendix: .1.5.

The geography table holds all the geographies where the company has offices and also contains the destinations of the shipments. This table is composed of 13 columns and holds 701 Rows, more information in Appendix: .1.6.

The currency table holds the currency code, country, and exchange rate of every month for all the countries where the company has offices. This table is composed of 3 columns and holds 3.954 Rows, more information in Appendix: .1.7.

4.3 Data Preparation

To have the data prepared for the analysis, a SQL Server Integration Services (SSIS) Project was created to extract, transform, and load the data to a new database. Below is a short description of the work made in the SSIS project. The complete details of the applied steps are available in the Appendix: .2.

As presented in Figure: 4.6, the data was provided by the company in a database called Thesis SIP 2.0. This database holds only a subset of the tables contained

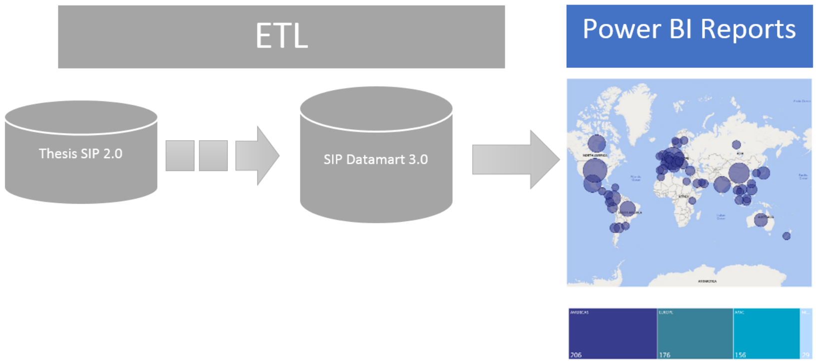

FIGURE 4.6: ETL diagram

in the main application. A SSIS Project was created to Extract, Transform, and Load the data to a new database.

As presented in the Figure: 4.7, the process starts by cleaning the staging tables and then copy the data from the Thesis SIP 2.0 database to the staging tables in the SIP Datamart 3.0 database. The second stage of the project is where the data is prepared to be transformed into dimensions and facts. Some transformations are made in this stage to have the data prepared for further analysis. The next phase of the process is where all the dimension tables are filled, by copying the data transformed in the previous stages to the dimension tables. The last step of the process is where the fact tables are filled, and the last transformations regarding the fact tables are made.

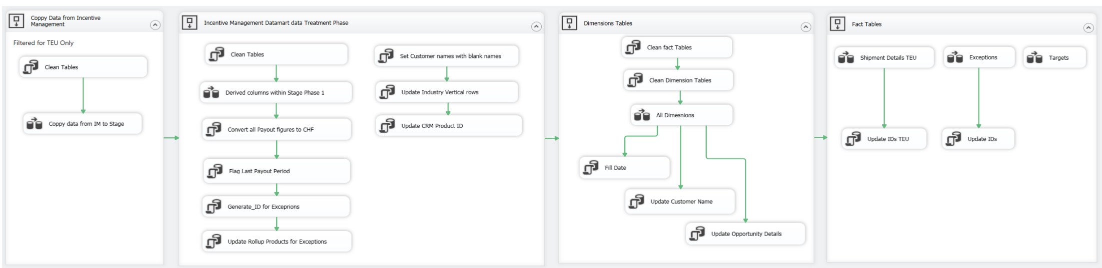

FIGURE 4.7: Integration Services overall

4.3.1 Cleaning data

The SSIS project included the following tasks to clean the data:

- All the rows where column Industry_Vertical had empty values were defined as "NA"
- Customer_ACR_Name, every empty row was filled with the names extracted from the master data system and filled manually on the database

For this research, the business requested to focus only on the TEU Unit of Measure (UOM). In the SSIS project, a filter was defined to exclude all the rows where the UOM is not TEU. This action reduced the data in the tables of sales_person_details, from the 1.982.969 to 180.875 rows.

After Extract Transform and Load all the data to the SIP Datamart 2.0, the data was loaded to Power BI so that a set of Power BI Reports can be created, and support the business assessing salesperson performance. The author describes the analysis in the next chapter.

Chapter 5

Salespeople assessment and classification

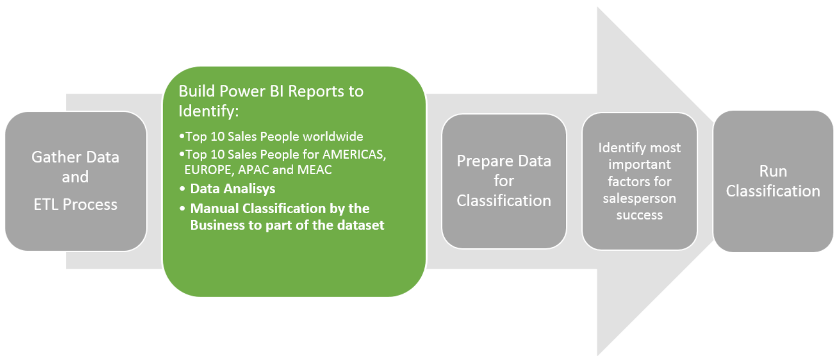

FIGURE 5.1: Research steps, Power BI reports stage

According to our research steps, its now time to work on the fourth objective for this research. The author now describe the evaluation of various Power BI reports created for the business to assess sales performance. These reports should allow the business to:

- Make a general assessment of global sales performance
- Assess salespeople performance, where it should be possible to identify:
  - The top 10 salespeople globally, by region and what volumes were sold by these salespeople

- In what countries are these top 40 salespeople working
- Assess other relevant KPIs for the salesperson performance
- Classify salespeople based on selected KPIs

The goal of these reports is to provide all the KPIs required to evaluate a salesperson's performance. Based on the performance evaluation a classification representing the performance of each salesperson is done. The provided classification is what the author will replicate through the predictive analytics process.

In the next chapters, the author describes the analysis and findings made on the reports together with the business. Due to data protection, the author replaced the salesperson's names on this dissertation but delivered the reports with the real salesperson's names to the business.

5.1 General assessment of global sales performance

As mentioned before, this company is a freight forward that transport merchandise from location A to location B, one transport is called a shipment. The shipments are made by Air Ocean and Land. In this analysis, the focus is on Ocean's shipments. Therefore the unit of measure is TEU's, and all shipments figures refer to 20 foot TEU containers like the one displayed in the Figure: 5.2.

FIGURE 5.2: 20 foot container = 1 TEU

The company has a global sales structure that has salespeople working at a global, regional, country level, and field sales. The hierarchy is presented in Figure: 5.3

for reference. The salespeople data reflected in this research belongs to the field sales, highlighted in the image in green in Figure: 5.3. These people represent about 50% of the company sales force.

FIGURE 5.3: Sales structure

In the dataset, there are 566 salespersons, distributed by 79 countries in 4 regions.

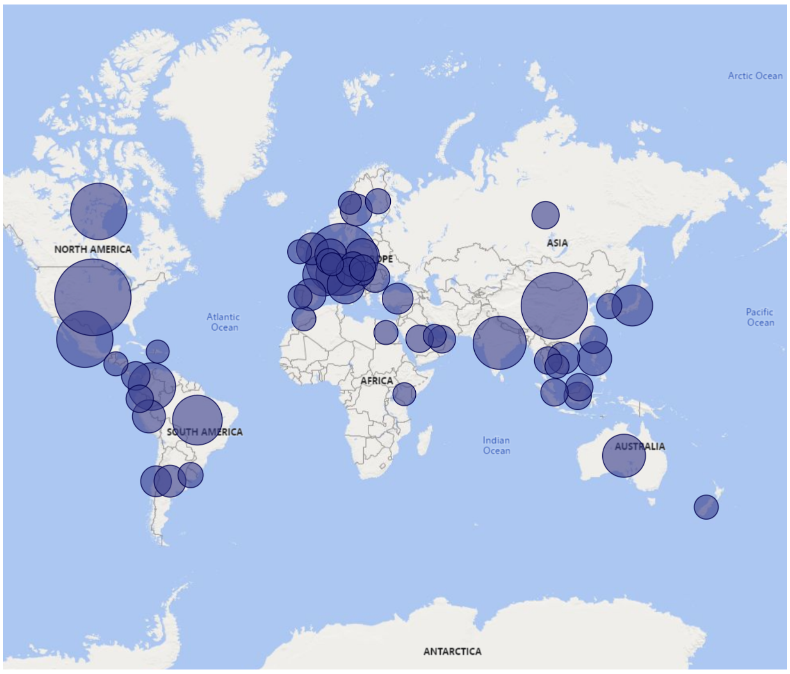

FIGURE 5.4: Salespersons distribution worldwide

The distribution of the salespeople by region is respectively Americas with 206, Europe with 176, APAC with 156, and MEAC with 29.

FIGURE 5.5: Salespersons distribution by region

As displayed in Figure: 5.4 from the 79 countries, the ones with the higher concentration of salespeople are the United States, Brazil, China, Mexico, Canada, Germany, and Australia.

For reference the countries where each region is present are displayed on figure: 5.6.

FIGURE 5.6: Countries where company is established

According to the business, the main KPI's used to evaluate a salesperson performance are:

- Customer Base which is all the customers assigned to a salesperson in the current year
- Customer Baseline that is the Sum of volume sold to the Customer Base on the previous year (0 is assumed for new customers)

- Growth is the difference between the Sum of the volume sold in the current year and the Base Line, taking the example in Figure: 5.7, the Growth is 2 TEU's because in 2019 2 more TEU's were shipped to the customer

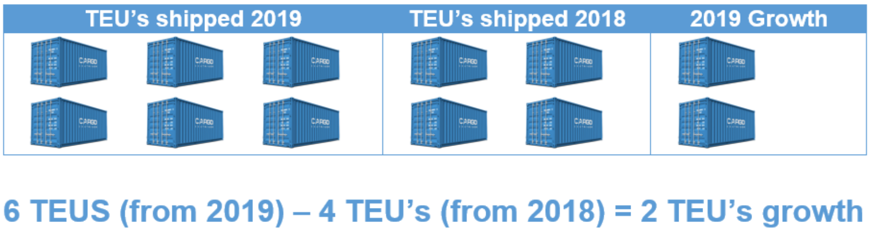

FIGURE 5.7: Growth example

Another concept widely used on the company are the exceptions, here called adjustments. Due to several factors that sometimes include human error, some shipments can be registered incorrectly in the TMS. When these cases occur, the IMS allows the creation of exceptions. These Exceptions have recorded: Salesperson, Date, Customer, Volume, and sometimes a respective payment. For this analysis, we recorded the exceptions and called them Adjustments, as the business name it (Payout Exceptions is the IT technical concept).

The data is gathered from Jan 2017 to June 2019. In 2,5 years between growth and adjustments, the 566 salespeople managed to grow the company volume by 150.478 TEU's, from this, 27.797 TEU's are from adjustments. As it's possible to verify in Figure: 5.8, in the year 2017, there were about 28K, in 2018 about 92K, and in 2019 so far there are about 32K TEU's.

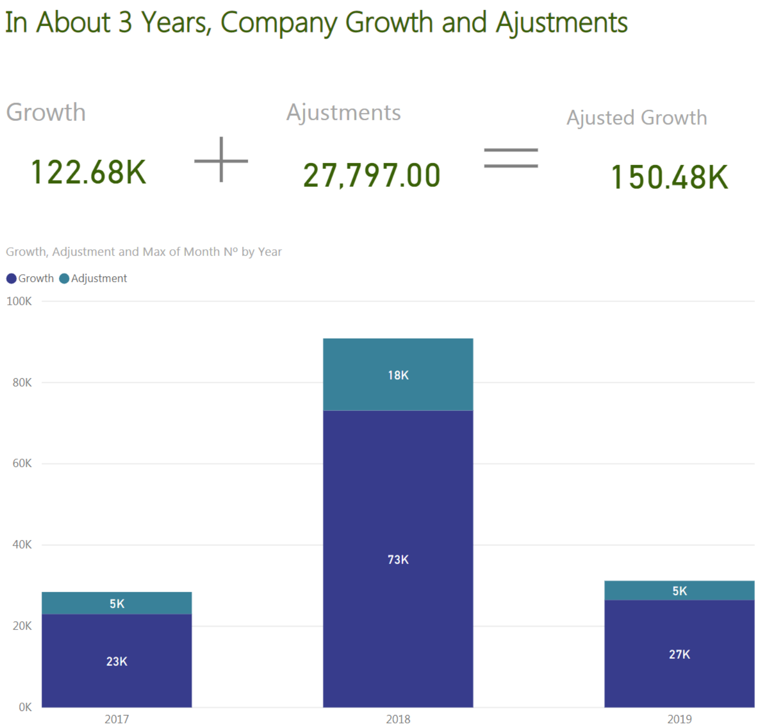

FIGURE 5.8: Growth by year

The reason for the smaller growth in 2017 and 2019 is due to the fact's that in 2017, the IMS was launched worldwide, and therefore not all countries were entirely using the system, and in 2019, the data available in the dataset belongs to only half year. We could still argue that the growth in 2019 could be higher since there are about 6 months of performance, but the business also informed that the higher volumes come at the end of the year as it is possible to verify in Figure: 5.9 in the last 2 months of the year the growth usually increases.

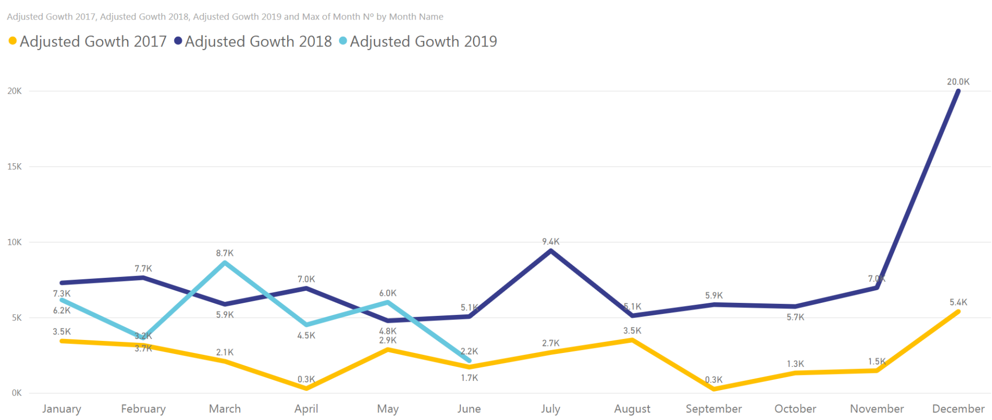

FIGURE 5.9: 3 years Growth by month

Analyzing region by region, the Americas have higher growth (81,45K). Europe comes next with about half (40,73K). The remaining two areas are more or less equal-sized APAC (15,05K) and MEAC (13,25K). Please refer to figure: 5.10 for better insight.

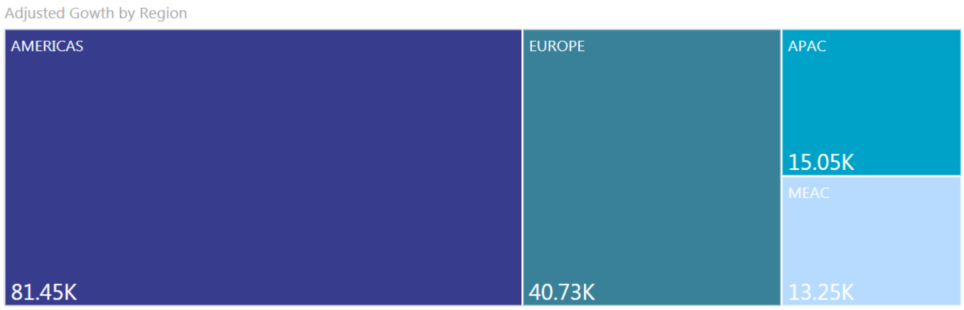

FIGURE 5.10: 3 years Growth by region

The TEU's shipment is divided into categories and products. This company Oceans products are divided into 3 Categories:

- Ocean FCL which refers to Ocean Full Container Load (Import and Export), is the act of buying the shipment of the complete container
- Ocean Other, is composed mainly by Freight Management, which is the act of buying the service of people's knowledge and coordination of the carriers and shippers, so in this case, the company providing the service don't move any container, only manages the shipments documentation and information
- Ocean Buyers Console is the consolidation of multiple shipments on the same container

As we can verify in Figure: 5.11, the main products shipped by this company are Oceans FCL Import and Export, there is also a significant amount in Freight Management, the remaining products are residual shipments wit lower TEU's.

With this analysis, the author concludes the goal of assessing global sales. The author delivered the reports to the business, and the moved forward to the next step, assess salesperson performance, where the top 10 salespeople worldwide and by region are identified, the countries where these people work, and other relevant KPIs are assessed.

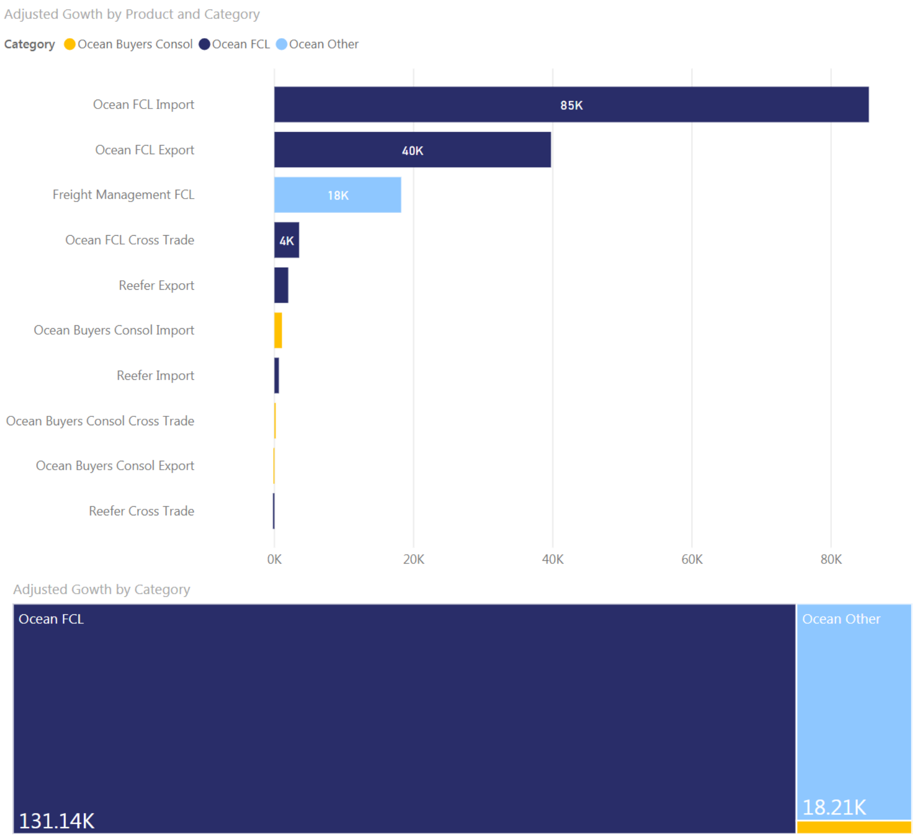

FIGURE 5.11: Growth by categories and products

5.2 Assess salespeople performance

To find the top 10 salespeople worldwide, by region, and identify other relevant KPIs to evaluate the salesperson's performance. The author made a set of queries suggested by the business to the dataset. These queries are:

- Who are the people with higher growth?
- Do these people achieved their defined Targets?
- Do the assigned targets to these salespeople follow the company guidelines?
- Other relevant KPIs, on this stage we will make a number of queries that goes from an in-depth analysis of the sales fluctuation, Customer Base, and to the ratio of opportunities created for each customer

As displayed in Figure: 5.12, the first level to verify is the higher growth, then check if the targets were achieved and finally if the targets follow the company guidelines. Other relevant KPIs that contribute to salesperson performance is also assessed, but these are the most important ones.

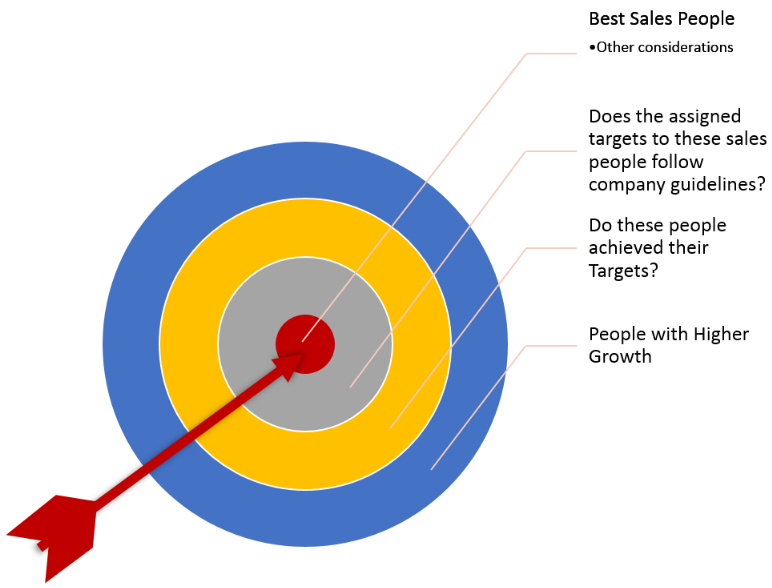

FIGURE 5.12: Queries to find the best salespeople

5.2.1 Who are the people with higher growth?

Starting with the first query: "Who are the people with higher growth." As explained before, a salesperson is assigned to an Account Base that has, on average, 70 customers, the base for analysis is the growth, which is the difference between the number of TEU's sold between the current and previous year. The base in the analysis is the Sum of the growth for each year. In order not to publish the names of the salespeople, a random name generator was used to replace all the names in the dataset.

With this procedure, the author Identified Top 10 salespeople worldwide and the top 10 salespeople for each region the company operates on.

The 566 salespeople grew the company Base Line by 150.478 TEU's in 2,5 Years, supported on a Customer Base of 9.261 customers. From these 150K, about 53K that is 35,4% of the total belongs to the top 10 salespeople worldwide, and these shipments are broth by only 343 customers who refer to about 3,5% of all the customers in the dataset.

5.2.1.1 Top 10 Salespeople worldwide

The countries with people that bring more TEUs are the USA, Uruguay Germany, Colombia, United Arab Emirates, France, and Canada. As we can also verify in Figure: 5.13, the USA is the country with higher growth with about 430% more growth than the next country, which is Germany.

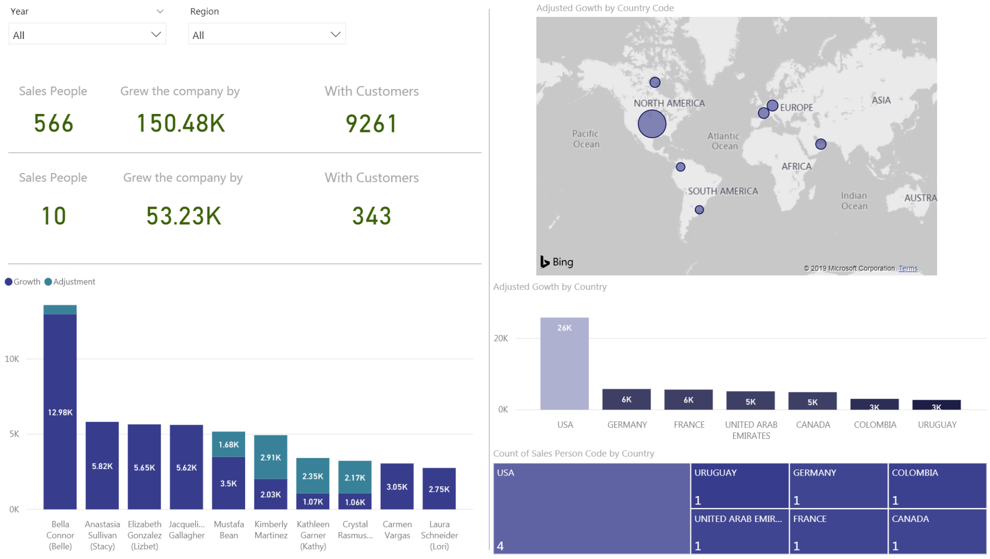

FIGURE 5.13: Top 10 salespeople worldwide

Also according to figure: 5.13, the top 10 salespersons worldwide are:

1. Bella Connor with: 13.586
2. Anastasia Sullivan with: 5.818
3. Elizabeth Gonzalez with: 5.652
4. Jacqueline Gallagher with: 5.616
5. Mustafa Bean with: 5.170
6. Kimberly Martinez with: 4.934

7. Kathleen Garner with: 3.417

8. Crystal Rasmussen with: 3.228

9. Carmen Vargas with: 3.055

10. Laura Schneider with: 2.753

5.2.1.2 Top 10 Salespeople Americas

Analyzing the top 10 salespeople for the Americas region, the total growth is 81,45K. The number of salespeople is 206, and the count of customers goes to 3.338. These top 10 salespeople now bring 43,01K TEU's which refers to about 53% of the total growth in the Americas region. The USA remains the country with more salespeople having 5 salespeople in the top 10. From this 10, the remaining salespeople are from Colombia, Canada, and Uruguay.

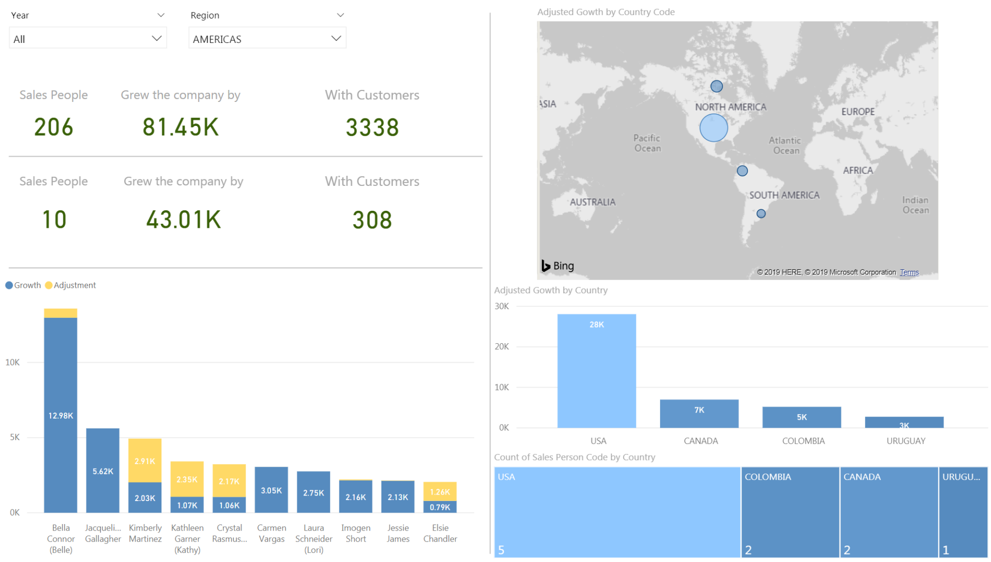

FIGURE 5.14: Top 10 Salespeople for region Americas

The top 10 salespeople in the Americas region are:

1. Bella Connor with: 13.586

2. Jacqueline Gallagher with: 5.616

3. Kimberly Martinez with: 4.934

4. Kathleen Garner with: 3.417

5. Crystal Rasmussen with: 3.228

6. Carmen Vargas with: 3.055

7. Laura Schneider with: 2.753

8. Imogen Short with: 2.219

9. Jessie James with: 2.150

10. Elsie Chandler with: 2.052

Worth of highlight on Figure: 5.14 , is that 54% of the total growth of the company worldwide comes from America, 7 of the top 10 salespeople worldwide are also from this region. Another information to highlight is that the USA continues to be the country with higher growth with about 400% more volume than the next country, which is Canada.

5.2.1.3 Top 10 Salespeople Europe

Analyzing the growth for Europe, 176 salespeople grew the company Base Line by 40,73K TEU's with 3.406 Customers, as presented in Figure: 5.15, from this growth, 10 salespeople grew the Base Line by about 24,64K TEU's which represents 60% of the total growth for Europe. In this region, Germany is the country with a higher growth concerning about 8,6K TEU's but with only about 30% more than the second country that is the Netherlands. Both countries, Germany and Netherlands, have 3 salespeople each in the top 10 salespeople in Europe, followed by Sweden, Poland, Hungary, and France, all with 1 salesperson each in the top 10 salespeople in Europe.

The top 10 salespeople in Europe are:

1. Anastacia Sulivan with: 5.818

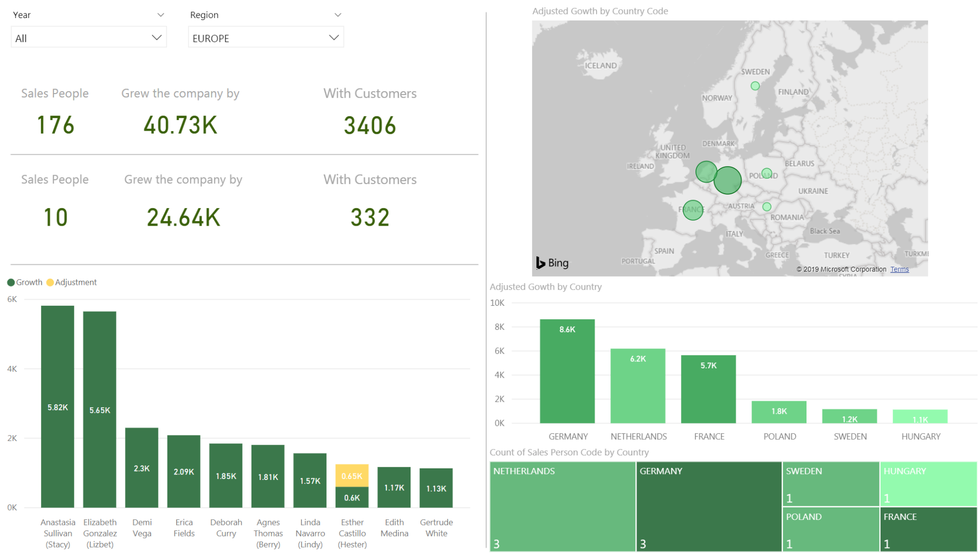

FIGURE 5.15: Top 10 Salespeople for region Europe

2. Elizabeth Gonzales with: 5.6552
3. Demi Vega with: 2.301
4. Erica Fields with: 2.087
5. Deborah Curry with: 1.847
6. Agnes Thomas with: 1.808
7. Linda Navarro with: 1.566
8. Esther Castillo with: 1.251
9. Edith Medina with: 1.171
10. Gertrude White with: 1.134

Refer to figure: 5.15 for a visual reference.

From all these salespeople, only Anastacia Sulivan and Elizabeth Gonzales are part of the top 10 worldwide as second and third.

5.2.1.4 Top 10 Salespeople APAC

In APAC 156 salespeople managed to grow the company during the 2,5 years by 15K TEU's based on 1.981 customers, from this, 9,8K TEU's are from the top 10 salespeople, which represents about 65% of the total TEU's in this region, all this with 274 customers. China is the country that contributes more to the growth with 5K TEU's, which is more than 66% than the next country that is India. In terms of salespeople contributing to the top 10 of the region, China is the country with more, contributing with 5 salespeople, the next is India with 2, and remaining countries are Thailand, Japan, and Australia, all with only 1 salesperson each.

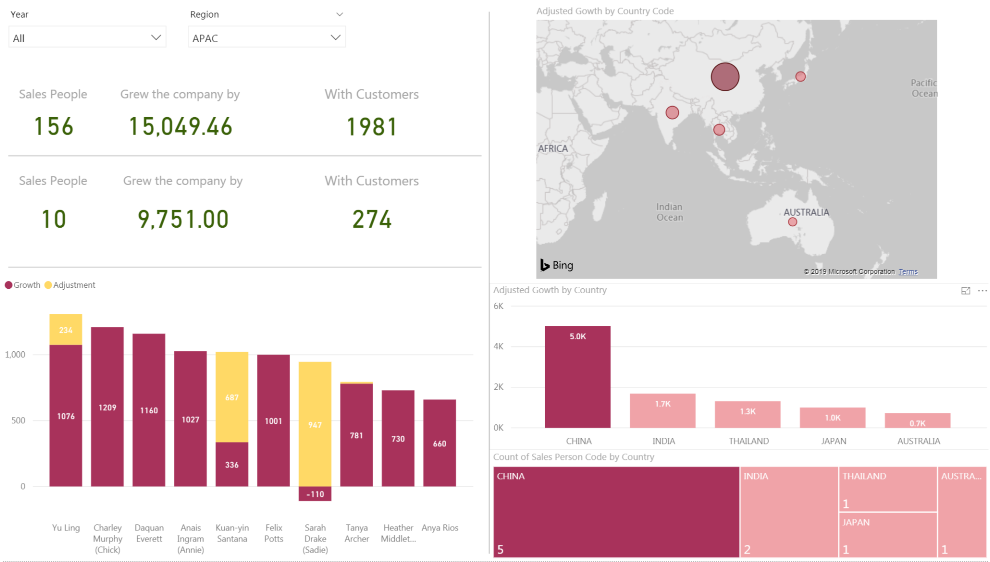

FIGURE 5.16: Top 10 Salespeople for region APAC

According the report displayed in the Figure: 5.16, the top 10 salespeople in APAC are:

1. Yu Ling with: 1.310
2. Charley Murphy with: 1.209
3. Daquan Everett with: 1.160
4. Anais Ingram with: 1.027
5. Kuan-yin Santana with: 1.023

6. Felix Potts with: 1.001
7. Sarah Drake with: 837
8. Tanya Archer with: 781
9. Heather Middleton with: 730
10. Anya Rios with: 660

Worth of notice is that none of the top 10 salespeople in APAC is part of the top 10 salespeople worldwide, and the difference between last salesperson worldwide and first in APAC is about 47% less TEU's (2,75K versus 1,3K). Also, according to Figure: 5.16, 2 salespeople have a very high value of exceptions, which compared to Europe and the Americas is not normal.

5.2.1.5 Top 10 Salespeople MEAC

The region with fewer salespeople is MEAC, only have 29 salespeople working in this region. These salespeople grew the company by 13,25K TEU's with 538 customers. From this growth, 10 salespeople broth 91% of the growth, respectively 12,06K TEU's with 206 customers. When it comes to the countries, the United Arab Emirates is the country with more growth 7,1K TEU's, which is about 355% more than the following country, that is Egypt with 2K TEU. United Arab Emirates is also the country with more salespeople in the TOP 10 of MEAC contributing with 4 salespeople, followed by Saudi Arabia and Egypt with 2 salespeople each and then Turkey and Morocco with one each.

As the report displayed in the Figure: 5.17 the top 10 salespeople are:

1. Mustafa Bean with: 5.170
2. Zora Simon with: 1.187
3. Radoslava Sandoval with: 1.213

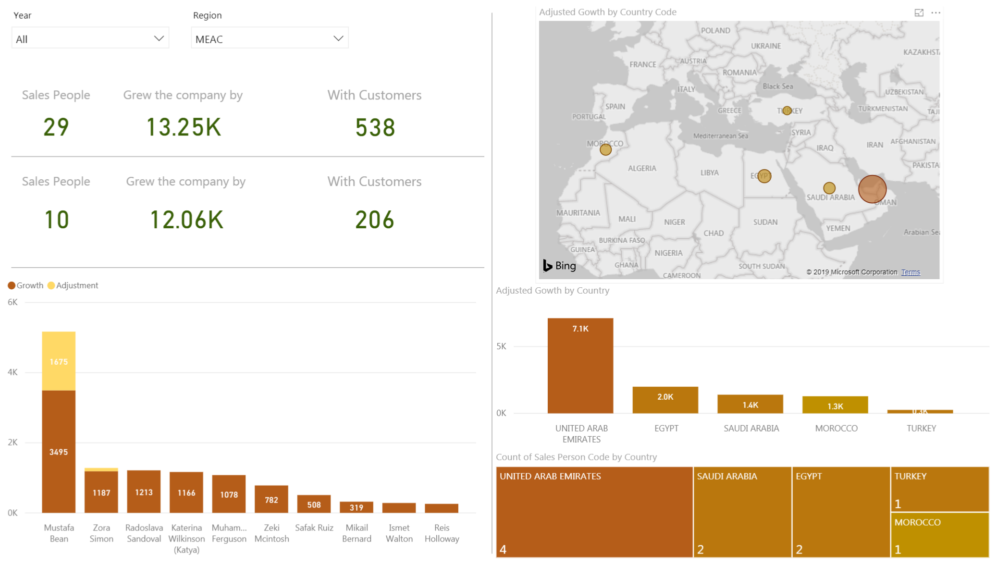

FIGURE 5.17: Top 10 Salespeople for region MEAC

4. Katerina Wilkinson (Katya) with 1.166
5. Muhammet Ferguson 1.078
6. Zeki Mcintosh with 782
7. Safak Ruiz with: 508
8. Mikail Bernard with 319
9. Ismet Walton with 282
10. Reis Holloway with 257

Worth noticing that according to the report in the Figure: 5.13 1 salesperson from MEAC is part of the top 10 salespeople worldwide.

The author evaluated these results with the business, which validated the provided numbers and reports. After the validation, the business confirmed the KPI's, Growth, Customer Baseline, and Account Base could be used to evaluate the salesperson's performance, but more KPI's are needed. The next stage is to evaluate the Targets and target achievements.

5.2.1.6 Do these people achieved their defined Targets?

Now that we found the top 40 salespeople in the company in terms of growth, the second stage defined by the business to find the best salespeople is to verify if these people achieved their targets.

The target definition in this company is supported on a top/down process. Targets are based on a roadmap that is defined globally by the sales controlling department, these targets are then assigned for each region and then distributed by the regional managers to the countries, and the process continues until it reaches the salesperson. As exemplified in Figure: 5.18, we have a global roadmap of 10.000 TEU's globally, that is shared among all the regions and ends on salesperson x and y in Lisbon with 30 TEU's each.

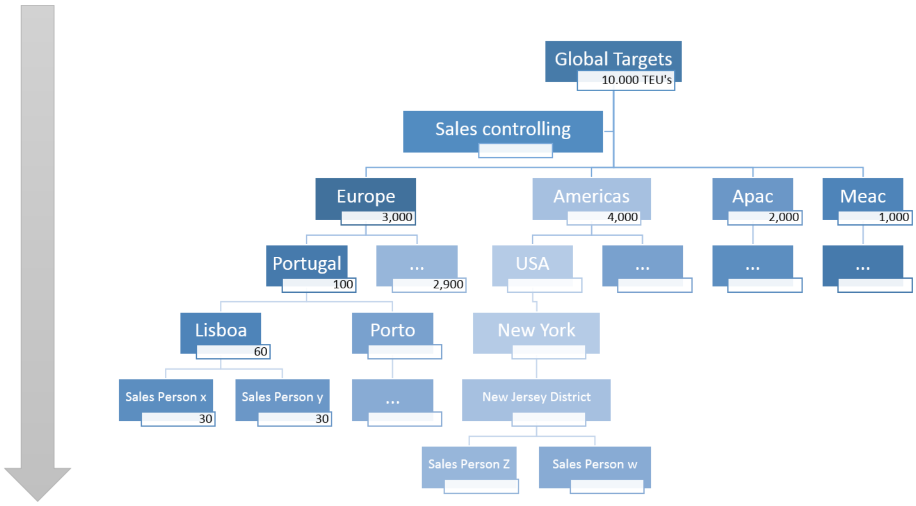

FIGURE 5.18: Target definition Top Down

Although the company has implemented this process, not always the salesperson gets a reasonable target, because this will depend on the strategy defined by the local sales management, and on this company, part of the strategy is defined locally. For instance, in Figure: 5.18, all Portugal's targets are assigned to Lisbon and none to Porto, if sales management in Portugal believe it's possible to achieve all targets with the 2 salespeople in Lisbon, they don't have to assign targets to salespeople in Porto.

Other than the number of TEU's assigned for a region/country, there is also a target definition at the product level. This is another way of strategically redirect the sales team to target a specific product. For instance, if a country has a higher market for Import, the sales manager should set Targets on Import to boost Import sales.

Analyzing our top 40 salespeople in terms of growth filtered only by-products with targets defined, we realize that the Top 1 salesperson: Bella Connor is not the first salesperson anymore, but 36$^{\text{th}}$ (4$^{\text{th}}$ counting from the bottom). With only 541 TEU's of growth in the 2,5 years, because the product she has more TEU's sold is not part of the products strategically targeted by the company, as displayed in Figure: 5.21.

In order to get a visually glimpse on the salespeople target achievement, the table presented in figure: 5.19, have the following color code applied in the achievement: Below 0 is red, between 0 and 70 is orange and above 70 is green.

Year	2017			2018			2019			Total		
Sales Person Name	Target	Adjusted Gowth	% Achieved	Target	Adjusted Gowth	% Achieved	Target	Adjusted Gowth	% Achieved	Target	Adjusted Gowth	% Achieved
Anastasia Sullivan (Stacy)	560.00	3,791.00	676.96 %	650.00	860.00	132.31 %	650.00	1,167.00	179.54 %	1,860.00	5,818.00	312.80 %
Ocean FCL Import	224.00	3,423.00	1528.13 %	350.00	510.00	145.71 %	650.00	1,188.00	182.77 %	1,224.00	5,121.00	418.38 %
Ocean FCL Export	336.00	368.00	109.52 %	300.00	350.00	116.67 %	0.00	-21.00	0.00 %	636.00	697.00	109.59 %
Mustafa Bean	884.00	2,383.00	269.57 %	796.00	2,169.00	272.49 %	500.00	618.00	123.60 %	2,180.00	5,170.00	237.16 %
Ocean FCL Cross Trade		1,434.00	0.00 %	204.00	952.00	466.67 %		361.00	0.00 %	204.00	2,747.00	1346.57 %
Ocean FCL Import	884.00	949.00	107.35 %	592.00	1,217.00	205.57 %	500.00	257.00	51.40 %	1,976.00	2,423.00	122.62 %
Jacqueline Gallagher			0.00 %	406.00	3,106.00	765.02 %	1,402.00	1,929.00	137.59 %	1,808.00	5,035.00	278.48 %
Ocean FCL Export			0.00 %	406.00	3,106.00	765.02 %	2.00	6.00	300.00 %	408.00	3,112.00	762.75 %
Ocean FCL Import			0.00 %			0.00 %	1,400.00	1,923.00	137.36 %	1,400.00	1,923.00	137.36 %
Kimberly Martinez	1,000.00	2,130.00	213.00 %	1,250.00	2,639.00	211.12 %	624.00	165.00	26.44 %	2,874.00	4,934.00	171.68 %
Ocean FCL Import	1,000.00	2,130.00	213.00 %	1,250.00	2,639.00	211.12 %	624.00	165.00	26.44 %	2,874.00	4,934.00	171.68 %
Kathleen Garner (Kathy)			0.00 %	900.00	2,823.00	313.67 %	600.00	603.00	100.50 %	1,500.00	3,426.00	228.40 %
Ocean FCL Export			0.00 %	900.00	2,823.00	313.67 %	50.00	258.00	516.00 %	950.00	3,081.00	324.32 %
Ocean FCL Import			0.00 %			0.00 %	550.00	345.00	62.73 %	550.00	345.00	62.73 %
Elizabeth Gonzalez (Lizbet)	6,200.00	479.00	7.73 %	399.00	3,224.00	808.02 %	304.00	-616.00	-202.63 %	6,903.00	3,087.00	44.72 %
Ocean FCL Export	6,200.00	479.00	7.73 %	399.00	3,224.00	808.02 %	304.00	-616.00	-202.63 %	6,903.00	3,087.00	44.72 %
Carmen Vargas	872.00	2,131.00	244.38 %	350.00	49.00	14.00 %	248.00	814.00	328.23 %	1,470.00	2,994.00	203.67 %
Ocean FCL Import	792.00	2,561.00	323.36 %	200.00	-238.00	-119.00 %	124.00	383.00	308.87 %	1,116.00	2,706.00	242.47 %
Ocean FCL Export	80.00	-430.00	-537.50 %	150.00	287.00	191.33 %	124.00	431.00	347.58 %	354.00	288.00	81.36 %
Crystal Rasmussen			0.00 %	750.00	1,864.00	248.53 %	1,025.00	1,062.00	103.61 %	1,775.00	2,926.00	164.85 %
Ocean FCL Export			0.00 %	750.00	1,864.00	248.53 %	150.00	-2.00	-1.33 %	900.00	1,862.00	206.89 %
Ocean FCL Import			0.00 %			0.00 %	875.00	1,064.00	121.60 %	875.00	1,064.00	121.60 %
Laura Schneider (Lori)	380.00	37.00	9.74 %	300.00	1,314.00	438.00 %	1,110.00	1,490.00	134.23 %	1,790.00	2,841.00	158.72 %
Ocean FCL Export	243.00	-109.00	-44.86 %	100.00	1,406.00	1406.00 %	1,110.00	1,490.00	134.23 %	1,453.00	2,787.00	191.81 %
Ocean FCL Import	137.00	146.00	106.57 %	200.00	-92.00	-46.00 %			0.00 %	337.00	54.00	16.02 %
Demi Vega	1,500.00	-82.00	-5.47 %	1,750.00	1,605.00	91.71 %	700.00	767.00	109.57 %	3,950.00	2,290.00	57.97 %
Ocean FCL Import	1,500.00	-82.00	-5.47 %	1,750.00	1,605.00	91.71 %	700.00	767.00	109.57 %	3,950.00	2,290.00	57.97 %
Erica Fields	1,500.00	876.00	58.40 %	1,250.00	1,277.00	102.16 %	624.00	110.00	17.63 %	3,374.00	2,263.00	67.07 %
Ocean FCL Import	1,500.00	876.00	58.40 %	1,250.00	1,277.00	102.16 %	624.00	110.00	17.63 %	3,374.00	2,263.00	67.07 %
Jessie James	200.00	-33.00	-16.50 %	768.00	1,886.00	245.57 %	274.00	290.00	105.84 %	1,242.00	2,143.00	172.54 %
Ocean FCL Import	140.00	-38.00	-27.14 %	548.00	1,638.00	298.91 %	174.00	31.00	17.82 %	862.00	1,631.00	189.21 %
Ocean FCL Export	60.00	5.00	8.33 %	220.00	248.00	112.73 %	100.00	259.00	259.00 %	380.00	512.00	134.74 %
Elsie Chandler	1,000.00	1,317.00	131.70 %	1,250.00	923.00	73.84 %	500.00	-158.00	-31.60 %	2,750.00	2,082.00	75.71 %
Ocean FCL Import	1,000.00	1,317.00	131.70 %	1,250.00	923.00	73.84 %	500.00	-158.00	-31.60 %	2,750.00	2,082.00	75.71 %
Deborah Curry	120.00	929.00	774.17 %	280.00	630.00	225.00 %	150.00	283.00	188.67 %	550.00	1,842.00	334.91 %
Ocean FCL Import	120.00	933.00	777.50 %	200.00	606.00	303.00 %	100.00	98.00	98.00 %	420.00	1,637.00	389.76 %
Ocean FCL Export		-4.00	0.00 %	80.00	24.00	30.00 %	50.00	185.00	370.00 %	130.00	205.00	157.69 %
Agnes Thomas (Berry)	2,250.00	674.00	29.96 %	1,550.00	1,022.00	65.94 %	274.00	106.00	38.69 %	4,074.00	1,802.00	44.23 %
Ocean FCL Import	2,250.00	674.00	29.96 %	1,550.00	1,022.00	65.94 %	274.00	106.00	38.69 %	4,074.00	1,802.00	44.23 %
Total	26,332.00	23,403.00	88.88 %	24,879.00	36,362.00	146.16 %	12,922.00	10,625.00	82.22 %	64,133.00	70,390.00	109.76 %

FIGURE 5.19: Top 40 Salespeople worldwide Part 1

The first salesperson is now Anastasia Sullivan (Stacy), with the same 5.818 TEU's, followed by Mustafa Bean, the fifth salesperson worldwide, with the same 5.170

TEU's, Jacqueline Gallagher remains the same third position globally as presented in Figure: 5.19.

Year	2017			2018			2019			Total		
Sales Person Name	Target	Adjusted Gowth	% Achieved	Target	Adjusted Gowth	% Achieved	Target	Adjusted Gowth	% Achieved	Target	Adjusted Gowth	% Achieved
Linda Navarro (Lindy)	**560.00**	**369.00**	**65.89 %**	**500.00**	**185.00**	**37.00 %**	**350.00**	**1,012.00**	**289.14 %**	**1,410.00**	**1,566.00**	**111.06 %**
Ocean FCL Import	224.00	318.00	141.96 %	300.00	258.00	86.00 %	300.00	1,001.00	333.67 %	824.00	1,577.00	191.38 %
Ocean FCL Export	336.00	51.00	15.18 %	200.00	-73.00	-36.50 %	50.00	11.00	22.00 %	586.00	-11.00	-1.88 %
Zora Simon	**600.00**	**287.00**	**47.83 %**	**1,409.00**	**994.00**	**70.55 %**	**70.00**		**0.00 %**	**2,079.00**	**1,281.00**	**61.62 %**
Ocean FCL Export	300.00	74.00	24.67 %	798.00	655.00	82.08 %	34.00		0.00 %	1,132.00	729.00	64.40 %
Ocean FCL Import	300.00	213.00	71.00 %	611.00	339.00	55.48 %	36.00		0.00 %	947.00	552.00	58.29 %
Esther Castillo (Hester)			**0.00 %**	**820.00**	**724.00**	**88.29 %**	**260.00**	**527.00**	**202.69 %**	**1,080.00**	**1,251.00**	**115.83 %**
Ocean FCL Export			0.00 %	800.00	461.00	57.63 %	250.00	529.00	211.60 %	1,050.00	990.00	94.29 %
Ocean FCL Import			0.00 %	20.00	263.00	1315.00 %	10.00	-2.00	-20.00 %	30.00	261.00	870.00 %
Radoslava Sandoval			**0.00 %**	**275.00**	**1,213.00**	**441.09 %**			**0.00 %**	**275.00**	**1,213.00**	**441.09 %**
Ocean FCL Import			0.00 %	138.00	1,217.00	881.88 %			0.00 %	138.00	1,217.00	881.88 %
Ocean FCL Export			0.00 %	137.00	-4.00	-2.92 %			0.00 %	137.00	-4.00	-2.92 %
Charley Murphy (Chick)	**1,550.00**	**1,694.00**	**109.29 %**	**100.00**	**109.00**	**109.00 %**	**50.00**	**-594.00**	**-1188.00 %**	**1,700.00**	**1,209.00**	**71.12 %**
Ocean FCL Export	1,550.00	1,694.00	109.29 %	100.00	109.00	109.00 %	50.00	-594.00	-1188.00 %	1,700.00	1,209.00	71.12 %
Yu Ling	**12.00**	**41.00**	**341.67 %**	**450.00**	**650.00**	**144.44 %**	**292.00**	**508.00**	**173.97 %**	**754.00**	**1,199.00**	**159.02 %**
Ocean FCL Export	6.00	-14.00	-233.33 %	225.00	595.00	264.44 %	146.00	621.00	425.34 %	377.00	1,202.00	318.83 %
Ocean FCL Import	6.00	55.00	916.67 %	225.00	55.00	24.44 %	146.00	-113.00	-77.40 %	377.00	-3.00	-0.80 %
Edith Medina	**960.00**	**336.00**	**35.00 %**	**1,000.00**	**835.00**	**83.50 %**			**0.00 %**	**1,960.00**	**1,171.00**	**59.74 %**
Ocean FCL Import	480.00	210.00	43.75 %	400.00	590.00	147.50 %			0.00 %	880.00	800.00	90.91 %
Ocean FCL Export	480.00	126.00	26.25 %	600.00	245.00	40.83 %			0.00 %	1,080.00	371.00	34.35 %
Katerina Wilkinson (Katya)	**884.00**	**93.00**	**10.52 %**	**1,070.00**	**599.00**	**55.98 %**	**500.00**	**474.00**	**94.80 %**	**2,454.00**	**1,166.00**	**47.51 %**
Ocean FCL Import	884.00	80.00	9.05 %	841.00	175.00	20.81 %	500.00	389.00	77.80 %	2,225.00	644.00	28.94 %
Ocean FCL Cross Trade		13.00	0.00 %	229.00	424.00	185.15 %	0.00	85.00	0.00 %	229.00	522.00	227.95 %
Daquan Everett	**500.00**	**627.00**	**125.40 %**	**700.00**	**734.00**	**104.86 %**	**250.00**	**-201.00**	**-80.40 %**	**1,450.00**	**1,160.00**	**80.00 %**
Ocean FCL Export	500.00	627.00	125.40 %	700.00	734.00	104.86 %	250.00	-201.00	-80.40 %	1,450.00	1,160.00	80.00 %
Gertrude White	**300.00**	**467.00**	**155.67 %**	**300.00**	**552.00**	**184.00 %**	**0.00**	**114.00**	**0.00 %**	**600.00**	**1,133.00**	**188.83 %**
Ocean FCL Import	300.00	467.00	155.67 %	300.00	552.00	184.00 %	0.00	114.00	0.00 %	600.00	1,133.00	188.83 %
Muhammet Ferguson	**425.00**	**517.00**	**121.65 %**	**683.00**	**392.00**	**57.39 %**	**450.00**	**138.00**	**30.67 %**	**1,558.00**	**1,047.00**	**67.20 %**
Ocean FCL Import	425.00	517.00	121.65 %	683.00	392.00	57.39 %	450.00	138.00	30.67 %	1,558.00	1,047.00	67.20 %
Anais Ingram (Annie)	**1,000.00**	**777.00**	**77.70 %**	**515.00**	**250.00**	**48.54 %**			**0.00 %**	**1,515.00**	**1,027.00**	**67.79 %**
Ocean FCL Import	500.00	773.00	154.60 %	515.00	250.00	48.54 %			0.00 %	1,015.00	1,023.00	100.79 %
Ocean FCL Export	500.00	4.00	0.80 %			0.00 %			0.00 %	500.00	4.00	0.80 %
Felix Potts	**400.00**	**73.00**	**18.25 %**	**400.00**	**925.00**	**231.25 %**			**0.00 %**	**800.00**	**998.00**	**124.75 %**
Ocean FCL Import	240.00	58.00	24.17 %	280.00	928.00	331.43 %			0.00 %	520.00	986.00	189.62 %
Ocean FCL Export	160.00	15.00	9.38 %	120.00	-3.00	-2.50 %			0.00 %	280.00	12.00	4.29 %
Imogen Short			**0.00 %**	**901.00**	**863.00**	**95.78 %**	**350.00**	**122.00**	**34.86 %**	**1,251.00**	**985.00**	**78.74 %**
Ocean FCL Export			0.00 %	901.00	863.00	95.78 %	100.00	-8.00	-8.00 %	1,001.00	855.00	85.41 %
Ocean FCL Import			0.00 %			0.00 %	250.00	130.00	52.00 %	250.00	130.00	52.00 %
Sarah Drake (Sadie)	**15.00**	**784.00**	**5226.67 %**	**1.00**	**163.00**	**16300.00 %**	**24.00**	**-110.00**	**-458.33 %**	**40.00**	**837.00**	**2092.50 %**
Ocean FCL Export	15.00	784.00	5226.67 %	1.00	163.00	16300.00 %	24.00	-110.00	-458.33 %	40.00	837.00	2092.50 %
Total	**26,332.00**	**23,403.00**	**88.88 %**	**24,879.00**	**36,362.00**	**146.16 %**	**12,922.00**	**10,625.00**	**82.22 %**	**64,133.00**	**70,390.00**	**109.76 %**

FIGURE 5.20: Top 40 Salespeople worldwide Part 2

For this study, targets are analyzed on a yearly base, and not all salespeople have targets or performance for every year, as displayed on figures: 5.19. Jacqueline Gallagher and Kathleen Garner (Kathy) only has growth and targets for 2018 and 2019. Regardless of only having data for these 1,5 years, these salespeople are still on the top 40 worldwide.

Year	2017			2018			2019			Total		
Sales Person Name	Target	Adjusted Gowth	% Achieved	Target	Adjusted Gowth	% Achieved	Target	Adjusted Gowth	% Achieved	Target	Adjusted Gowth	% Achieved
Kuan-yin Santana	**220.00**	**865.00**	**393.18 %**	**1.00**	**96.00**	**9600.00 %**	**20.00**	**-146.00**	**-730.00 %**	**241.00**	**815.00**	**338.17 %**
Ocean FCL Export	220.00	865.00	393.18 %	1.00	96.00	9600.00 %	20.00	-146.00	-730.00 %	241.00	815.00	338.17 %
Tanya Archer	**107.00**	**-134.00**	**-125.23 %**	**228.00**	**741.00**	**325.00 %**	**70.00**	**185.00**	**264.29 %**	**405.00**	**792.00**	**195.56 %**
Ocean FCL Export	107.00	-134.00	-125.23 %	228.00	741.00	325.00 %	70.00	185.00	264.29 %	405.00	792.00	195.56 %
Zeki Mcintosh	**187.00**	**280.00**	**149.73 %**	**682.00**	**650.00**	**95.31 %**	**254.00**	**-148.00**	**-58.27 %**	**1,123.00**	**782.00**	**69.63 %**
Ocean FCL Import	187.00	60.00	32.09 %	152.00	378.00	248.68 %	62.00	31.00	50.00 %	401.00	469.00	116.96 %
Freight Management FCL		220.00	0.00 %	530.00	272.00	51.32 %	130.00	-130.00	-100.00 %	660.00	362.00	54.85 %
Ocean FCL Export			0.00 %			0.00 %	62.00	-49.00	-79.03 %	62.00	-49.00	-79.03 %
Anya Rios	**496.00**	**1,232.00**	**248.39 %**	**50.00**	**-567.00**	**-1134.00 %**	**6.00**	**-5.00**	**-83.33 %**	**552.00**	**660.00**	**119.57 %**
Ocean FCL Export	248.00	996.00	401.61 %			0.00 %			0.00 %	248.00	996.00	401.61 %
Ocean FCL Import	248.00	236.00	95.16 %	50.00	-567.00	-1134.00 %	6.00	-5.00	-83.33 %	304.00	-336.00	-110.53 %
Bella Connor (Belle)			**0.00 %**	**750.00**	**525.00**	**70.00 %**	**54.00**	**16.00**	**29.63 %**	**804.00**	**541.00**	**67.29 %**
Ocean FCL Export			0.00 %	750.00	525.00	70.00 %	18.00	-176.00	-977.78 %	768.00	349.00	45.44 %
Ocean FCL Import			0.00 %			0.00 %	36.00	192.00	533.33 %	36.00	192.00	533.33 %
Mikail Bernard	**900.00**	**319.00**	**35.44 %**			**0.00 %**			**0.00 %**	**900.00**	**319.00**	**35.44 %**
Ocean FCL Import	900.00	319.00	35.44 %			0.00 %			0.00 %	900.00	319.00	35.44 %
Ismet Walton			**0.00 %**	**991.00**	**167.00**	**16.85 %**	**500.00**	**115.00**	**23.00 %**	**1,491.00**	**282.00**	**18.91 %**
Ocean FCL Import			0.00 %	791.00	106.00	13.40 %	500.00	73.00	14.60 %	1,291.00	179.00	13.87 %
Ocean FCL Cross Trade			0.00 %	200.00	61.00	30.50 %		42.00	0.00 %	200.00	103.00	51.50 %
Reis Holloway	**100.00**	**47.00**	**47.00 %**	**404.00**	**171.00**	**42.33 %**	**87.00**	**39.00**	**44.83 %**	**591.00**	**257.00**	**43.49 %**
Ocean FCL Export	100.00	47.00	47.00 %	403.00	173.00	42.93 %	87.00	23.00	26.44 %	590.00	243.00	41.19 %
Reefer Export			0.00 %	1.00	-2.00	-200.00 %	0.00	16.00	0.00 %	1.00	14.00	1400.00 %
Safak Ruiz	**650.00**	**97.00**	**14.92 %**			**0.00 %**	**350.00**	**-51.00**	**-14.57 %**	**1,000.00**	**46.00**	**4.60 %**
Ocean FCL Import	650.00	97.00	14.92 %			0.00 %			0.00 %	650.00	97.00	14.92 %
Ocean FCL Export			0.00 %			0.00 %	350.00	-51.00	-14.57 %	350.00	-51.00	-14.57 %
Total	**26,332.00**	**23,403.00**	**88.88 %**	**24,879.00**	**36,362.00**	**146.16 %**	**12,922.00**	**10,625.00**	**82.22 %**	**64,133.00**	**70,390.00**	**109.76 %**

FIGURE 5.21: Top 40 Salespeople worldwide Part 3

The target achievement is calculated by applying the Formula: (5.1). As presented on the Figure: 5.20, the lower achievement is -1,188 %, what this means is that this salesperson probably did not sell any TEU's in 2018 in any product with target defined. The higher percengate Achieved is 16300% achieved by Sarah Drake in 2018, this high overachievement is caused by the deficient (1) Target when compared to the Growth the salesperson brought to the company.

$$\%Achieved = \frac{Gowth}{Target} \tag{5.1}$$

This concludes the evaluation of the target achievement, confirming that this KPI can be used to assess the salesperson's performance. Next is the assessment of the targets defined to the salesperson.

5.2.1.7 Do the assigned targets to these salespeople follow the company guidelines?

Now that we have an idea of how is target achievement for the salespeople, the next step is to verify if the targets applied to these salespeople are set based on the company's guideline.

In this company, targets are set to a salesperson mainly based on 3 pillars:

- Account Base
- Sales roadmap
- Salesperson seniority

As described previously, the Account Base is composed of the customers that are assigned to the salesperson, and it has a significant impact on the level of the target that can be assigned to the person. If a salesperson has a Customer Base composed by 10 customers and these customers have a possibility of purchase 100 TEU's along the year, the targets assigned to this salesperson should not be a value that is too far from the 100 TEU's, unless the person setting the targets have information's that may indicate that the customer will have an exponential increase.

Sales roadmap is the document that has the plan for the company sales growth for the long term. This document for the company in question is composed of the main product categories, regions, trade lanes, among other information. Often sales managers set targets just based on the sales roadmap, but this may lead to the definition of "unrealistic" targets if the Account Base does not provide the potential needed to achieve the targets. When this happens, CRM Pipeline figures is another ally to set the targets. Usually, to improve target setting, Pipeline figures are added to the Sales Planning process. This way, the salesperson and manager have not only the Customer Baseline but also the forecast (assuming good forecasting accuracy).

The salesperson seniority also has a significant role in how the salesperson works the Customer Base. A junior salesperson may not have the ability to manage complex accounts. Therefore the sales manager, when assigning the Customer Base, has to know salesperson seniority. Seniority in the company/products has also consequences on managing the Account Base, for instance, if somebody has just joined the company and is also junior (young), he/she will need "more" time to start generating results: new company, new products, the need to build an internal network, among other relevant tasks. To mitigate this issue, often sales managers give a new/junior salesperson lower targets in the beginning and then increase the targets year-by-year as the seniority increases.

In Figure: 5.22 is displayed an example of target definition for one salesperson (removed name for data protection), where it's possible to verify a 15% increase from the Account Base that is 283 TEU's, the increase has an impact of 42 more TEU's and is spitted across 4 quarters by 10 for Q1, 10 for Q2, 11 for Q3 and 11 for Q4.

Sales Person	Product	Unit	Annual Target 2018	FY Actual 2018	FY Adjustment 2018	FY Actual Adjusted 2018	Annual Target 2019		Q1 Target 2019	Q2 Target 2019	Q3 Target 2019	Q4 Target 2019
							% Increase	**Value Increase**				
		TEU										
	Ocean FCL Export	TEU	264	283	0	283	15	42	10	10	11	11
	Ocean FCL Cross Trade	TEU	0	10	0	10	10	1	0	0	0	1
	Ocean Buyers Consol Cross Trade	TEU	0	0	0	0	0	0	0	0	0	0

FIGURE 5.22: Example of target setting in Incentive Management System

Pipeline and seniority are entirely missing in this research, so to judge the targets, a validation is made comparing the targets directly with the Customer Baseline on a report.

This report makes the evaluation by dividing the targets with the Customer Baseline as displayed in the Formula: 5.2, highlighting in green, every target that is between 70% and 130% of the Customer Baseline.

$$Target\%evaluation = \frac{Target}{AccountBaseline} \tag{5.2}$$

Sales Person Name	2017	2018	2019	**Total**
Radoslava Sandoval	0.00 %	1718.75 %	0.00 %	**1718.75 %**
Ismet Walton	0.00 %	792.80 %	1851.85 %	**980.92 %**
Zora Simon	0.00 %	646.33 %	0.00 %	**953.67 %**
Jacqueline Gallagher	0.00 %	6766.67 %	485.12 %	**612.88 %**
Reis Holloway	0.00 %	859.57 %	150.00 %	**562.86 %**
Katerina Wilkinson (Katya)	1449.18 %	594.44 %	230.41 %	**535.81 %**
Imogen Short	0.00 %	600.67 %	105.11 %	**259.01 %**
Erica Fields	230.06 %	219.68 %	178.29 %	**214.77 %**
Muhammet Ferguson	0.00 %	137.15 %	171.76 %	**205.00 %**
Laura Schneider (Lori)	136.69 %	80.43 %	437.01 %	**197.79 %**
Safak Ruiz	833.33 %	0.00 %	68.90 %	**170.65 %**
Mustafa Bean	225.51 %	109.04 %	169.49 %	**153.85 %**
Anais Ingram (Annie)	14285.71 %	52.39 %	0.00 %	**153.03 %**
Zeki Mcintosh	9350.00 %	215.14 %	59.62 %	**150.74 %**
Edith Medina	125.33 %	125.16 %	0.00 %	**125.24 %**
Linda Navarro (Lindy)	96.22 %	53.48 %	213.41 %	**83.88 %**
Elizabeth Gonzalez (Lizbet)	1324.79 %	9.17 %	8.13 %	**80.68 %**
Agnes Thomas (Berry)	105.98 %	48.18 %	29.69 %	**65.05 %**
Tanya Archer	41.31 %	171.43 %	29.91 %	**64.70 %**
Felix Potts	64.62 %	64.41 %	0.00 %	**64.52 %**
Esther Castillo (Hester)	0.00 %	49.10 %	161.49 %	**58.98 %**
Demi Vega	77.16 %	37.23 %	113.64 %	**54.41 %**
Jessie James	66.67 %	96.12 %	21.73 %	**52.63 %**
Crystal Rasmussen	0.00 %	26.40 %	59.70 %	**38.94 %**
Gertrude White	254.24 %	38.76 %	0.00 %	**38.81 %**
Anya Rios	1271.79 %	3.84 %	1.52 %	**31.80 %**
Elsie Chandler	33.81 %	33.22 %	24.43 %	**31.36 %**
Kimberly Martinez	32.36 %	27.02 %	24.90 %	**28.11 %**
Daquan Everett	25.64 %	27.44 %	22.03 %	**25.73 %**
Deborah Curry	38.46 %	25.16 %	19.28 %	**24.97 %**
Kathleen Garner (Kathy)	0.00 %	21.24 %	31.70 %	**24.47 %**
Yu Ling	0.69 %	62.15 %	35.31 %	**22.85 %**
Charley Murphy (Chick)	62.50 %	2.23 %	2.68 %	**19.28 %**
Carmen Vargas	35.56 %	8.32 %	11.31 %	**16.60 %**
Anastasia Sullivan (Stacy)	19.13 %	9.77 %	25.54 %	**15.34 %**
Bella Connor (Belle)	0.00 %	18.07 %	4.58 %	**15.08 %**
Kuan-yin Santana	24.94 %	0.06 %	2.70 %	**7.36 %**
Sarah Drake (Sadie)	1.34 %	0.07 %	4.53 %	**1.32 %**
Mikail Bernard	0.00 %	0.00 %	0.00 %	**0.00 %**
Total	**92.05 %**	**37.76 %**	**41.53 %**	**51.06 %**

FIGURE 5.23: Top 40 salespeople highlight targets between 70% and 130% of Customer Baseline

In Figure: 5.23, the small number of fields in green reflects the unrealistic targets defined for the salespeople, as most of the targets are under 70% and above 130% of the Customer Baseline.

To validate if the target should be considered as a KPI to be used in the salesperson performance evaluation, the filter for the top 40 salespeople were removed from the previous report. The result is that most salespeople have their targets within this range. Therefore, the business accepted the inclusion of this metric also in the salesperson performance measurement. The report is provided in the Figure: 5.24 for reference. Next is the consideration of other KPIs that help in the measurement of the salesperson's performance.

Part 1

Sales Person Name	2017	2018	2019	**Total**
Mariam Moyer (Mitzi)	129.31 %	0.00 %	0.00 %	**129.31 %**
Kaitlyn Clements (Caitie)	11.39 %	526.09 %	0.00 %	**127.45 %**
Lacey Suarez	30.77 %	555.56 %	386.67 %	**126.57 %**
Isabelle Myers	166.67 %	318.07 %	12.27 %	**125.49 %**
Laura Schneider (Lori)	210.77 %	96.62 %	0.00 %	**123.90 %**
Katerina Wilkinson (Katya)		1090.48 %	0.00 %	**123.78 %**
Jose Berry	149.46 %	111.34 %	0.00 %	**121.68 %**
Liberty Pitts	310.81 %	74.30 %	124.48 %	**120.69 %**
Victoria Medina	95.24 %	397.73 %	65.57 %	**120.06 %**
Katy Medina	580.00 %	153.06 %	27.91 %	**119.59 %**
Sally Hooper	118.81 %	0.00 %	0.00 %	**118.81 %**
Robbie Evans	118.48 %	0.00 %	0.00 %	**118.48 %**
Kyra Peters	0.00 %	116.67 %	0.00 %	**116.67 %**
Adele Sampson	0.00 %	116.58 %	111.11 %	**115.91 %**
Fleur Lara	0.00 %	115.38 %	0.00 %	**115.38 %**
Bethany Bloggs	1625.00 %	48.86 %	0.00 %	**115.24 %**
Leia Peters	115.00 %	0.00 %	0.00 %	**115.00 %**
Bailey Vasquez	178.96 %	81.29 %	0.00 %	**114.57 %**
Esther Fischer (Hester)	0.00 %	122.90 %	84.03 %	**114.44 %**
Amina Richards	112.61 %	116.10 %	0.00 %	**114.08 %**
Imogen Henry	0.00 %	55.56 %	540.54 %	**114.01 %**
Shania Yang	0.00 %	114.29 %	111.54 %	**113.66 %**
Thea Green	112.68 %	0.00 %	0.00 %	**112.68 %**
Serena Osborn	0.00 %	172.41 %	26.10 %	**110.43 %**
Chloe Lopez	75.47 %	1111.11 %	94.59 %	**108.09 %**
Bethany Mcconnell	72.55 %	201.82 %	82.64 %	**108.05 %**
Beth Mitchell (Liza)	95.97 %	145.03 %	46.51 %	**104.74 %**
Alisha Warren	104.55 %	0.00 %	0.00 %	**104.55 %**
Aoife Bauer	44.01 %	191.49 %	693.75 %	**104.46 %**
Crystal Short	62.90 %	240.20 %	118.52 %	**104.40 %**
Aliyah Dominguez	103.81 %	0.00 %	0.00 %	**103.81 %**
Aysha Larson	117.92 %	91.91 %	0.00 %	**103.31 %**
Kayla Aguilar (Kaila)	0.00 %	62.50 %	0.00 %	**102.98 %**
Georgie Fowler	103.41 %	101.86 %	0.00 %	**102.62 %**
Anais Ingram (Annie)	7142.86 %	52.39 %	0.00 %	**102.53 %**
Poppy Hines	144.74 %	97.56 %	37.78 %	**101.97 %**
Julie Sullivan		1623.08 %	29.74 %	**101.66 %**
Maisie Gamble (Gretal)	227.27 %	65.18 %	181.25 %	**100.42 %**
Alisha Fisher	111.11 %	97.09 %	83.61 %	**100.07 %**
Faith Oneal	100.00 %	0.00 %	0.00 %	**100.00 %**
Total	**137.35 %**	**84.23 %**	**62.16 %**	**95.33 %**

Part 2

Sales Person Name	2017	2018	2019	**Total**
Kye Wright	0.00 %	0.00 %	100.00 %	**100.00 %**
Ria Cunningham	0.00 %	0.00 %	100.00 %	**100.00 %**
Marie Melton	82.76 %	76.92 %	192.31 %	**97.53 %**
Alina Murray	252.33 %	22.60 %	72.55 %	**97.01 %**
Courtney Stein	38.22 %	175.00 %	156.25 %	**96.49 %**
Claudia Mcgrath	95.73 %	0.00 %	0.00 %	**95.73 %**
Lee Harrell		500.00 %	37.50 %	**94.87 %**
Sophia Boyle (Sophie)	0.00 %	86.96 %	100.67 %	**94.70 %**
Leona Schneider (Loni)	0.00 %	140.00 %	80.00 %	**94.63 %**
Hafsa Sheppard	93.28 %	0.00 %	0.00 %	**93.28 %**
Marnie Olson	0.00 %	61.03 %	179.75 %	**93.15 %**
Isabel Hurst	103.90 %	115.34 %	56.49 %	**92.96 %**
Sabrina Finch	216.22 %	59.41 %	64.59 %	**92.88 %**
Imogen Short	0.00 %	0.00 %	92.59 %	**92.59 %**
Zeki Mcintosh	0.00 %	0.00 %	92.54 %	**92.54 %**
Tianna Carroll	0.00 %	77.92 %	114.89 %	**91.94 %**
Marnie Koch	53.00 %	227.42 %	0.00 %	**91.76 %**
Josephine Price (Josie)	0.00 %		94.29 %	**90.41 %**
Rhea King	93.40 %	71.43 %	124.44 %	**90.40 %**
Autumn Frost	90.03 %	0.00 %	0.00 %	**90.03 %**
Cory Morgan	0.00 %	0.00 %	89.82 %	**89.82 %**
Julia Collier (Jules)	0.00 %	110.53 %	49.50 %	**89.35 %**
Leila Proctor	0.00 %	95.24 %	78.57 %	**88.57 %**
Evangeline Webster	61.54 %	70.85 %	0.00 %	**88.40 %**
Leah Perez	390.06 %	36.90 %	34.48 %	**87.59 %**
Ava Hancock	86.96 %	0.00 %	0.00 %	**86.96 %**
Jerry Brady	0.00 %	37.15 %	1388.89 %	**86.91 %**
Elin Chan	112.32 %	44.57 %	110.29 %	**85.02 %**
Scarlett David	0.00 %	0.00 %	84.62 %	**84.62 %**
Katy Lamb	0.00 %	55.56 %	0.00 %	**84.44 %**
Anastasia Sullivan (Stacy)	186.67 %	55.15 %	0.00 %	**84.13 %**
Nora Brown	0.00 %	0.00 %	84.03 %	**84.03 %**
Callie Larsen (Cali)	106.38 %	88.13 %	60.61 %	**84.00 %**
Roxanne Rice	0.00 %	83.33 %	0.00 %	**83.33 %**
Alicia Lane	82.41 %	0.00 %	0.00 %	**82.41 %**
Saffron Ward	0.00 %	113.21 %	63.04 %	**81.38 %**
Freya Chapman	129.03 %	39.55 %	0.00 %	**81.33 %**
Anisa Mata	0.00 %	70.93 %	98.25 %	**81.26 %**
Bonnie Walters	81.06 %	0.00 %	0.00 %	**81.06 %**
Total	**137.35 %**	**84.23 %**	**62.16 %**	**95.33 %**

Part 3

Sales Person Name	2017	2018	2019	**Total**
Paula Mckenzie (Kenzie)	81.01 %	0.00 %	0.00 %	**81.01 %**
Tiana Gibbons		98.43 %	84.04 %	**80.85 %**
Elizabeth Gonzalez (Lizbet)	1324.79 %	9.17 %	8.13 %	**80.68 %**
Tabitha Dale	0.00 %		109.89 %	**72.46 %**
Francis Orozco		153.64 %	46.61 %	**69.94 %**
Syeda Bridges		173.91 %	46.30 %	**49.43 %**
Kiera Carson		151.72 %	10.00 %	**35.49 %**
Total	**137.35 %**	**84.23 %**	**62.16 %**	**95.33 %**

FIGURE 5.24: Targets between 70% and 130% of Customer Baseline all salespeople

5.2.1.8 Other relevant KPIs for the salesperson performance

There are other KPIs that need to be validated over the salesperson, to measure performance, these include:

- Customer Base, number of customers assigned to the salesperson
- Customer Baseline, TEU's sold in the previous year to all the customers from the Customer Base
- Number of Opportunities created by the salesperson
- Average number of different Opportunities per customer
- Growth variability along the year (Number of months with positive growth)
- Grow with Different Products
- Number of products with positive growth

The table available in the Figure: 5.25 provides all this information's for our top 40 salespeople worldwide. We can highlight from the table that number of opportunities is remarkable when compared to the second salesperson, another important information is the average number of months with growth above 0, in average she is able to grow the Customer Base for about 8 months each year, and she can also grow more than one product.

Row Labels	Adjusted Growth	Customer Base	Base Line	Nº Opportunities created	Average Nº of different Opportunities per customer	Average of Nº Months with growth above 0	Grow with Different Products	Number of products with positive growth
Bella Connor (Belle)	13,075	20	20,727	318	30	8	1	4
Anastasia Sullivan (Stacy)	5,818	182	12,124	170	3	9	2	5
Elizabeth Gonzalez (Lizbet)	5,652	79	9,420	122	3	7	1	4
Jacqueline Gallagher	5,616	23	315	22	2	9	2	7
Kimberly Martinez	5,349	73	10,223	112	5	5	-	2
Mustafa Bean	5,171	75	1,417	138	4	10	3	6
Elsie Chandler	4,707	126	9,021	174	3	6	1	4
Crystal Rasmussen	3,449	30	4,651	69	4	7	2	4
Kathleen Garner (Kathy)	3,417	21	6,345	76	7	7	2	4
Carmen Vargas	3,072	110	8,883	264	6	6	3	10
Demi Vega	2,936	50	7,261	36	3	5	2	7
Laura Schneider (Lori)	2,753	64	1,275	72	6	8	1	5
Agnes Thomas (Berry)	2,608	51	6,295	36	3	5	2	5
Imogen Short	2,235	29	1,103	33	2	9	2	6
Jessie James	2,130	101	2,397	275	7	8	2	7
Erica Fields	2,087	34	1,827	20	2	6	-	3
Deborah Curry	1,847	68	2,203	55	3	9	2	7
Linda Navarro (Lindy)	1,566	46	1,681	41	3	8	2	5
Yu Ling	1,310	84	3,302	426	15	7	2	5
Zora Simon	1,281	24	218	25	2	8	2	4
Esther Castillo (Hester)	1,251	31	1,831	29	2	6	-	2
Radoslava Sandoval	1,213	6	16	3	1	6	-	1
Charley Murphy (Chick)	1,209	51	8,819	225	13	7	-	2
Edith Medina	1,171	24	1,565	21	2	11	2	4
Katerina Wilkinson (Katya)	1,166	36	458	64	5	6	3	6
Daquan Everett	1,160	80	5,636	166	6	6	-	2
Gertrude White	1,134	44	1,553	19	3	8	1	4
Muhammet Ferguson	1,078	43	760	70	4	8	2	5
Anais Ingram (Annie)	1,027	12	990	20	4	8	1	3
Felix Potts	1,001	67	1,240	32	2	9	2	4
Kuan-yin Santana	959	95	3,277	262	7	5	-	2
Safak Ruiz	805	20	629	28	2	5	1	3
Tanya Archer	791	69	626	82	4	5	-	2
Zeki Mcintosh	782	23	745	19	3	6	2	5
Sarah Drake (Sadie)	753	43	3,027	101	6	5	-	1
Heather Middleton	730	2	1,725	-	-	1	-	1
Anya Rios	660	35	1,736	63	6	7	1	2
Mikail Bernard	319	5	-	-	-	5	-	1
Ismet Walton	282	40	152	41	2	6	2	4
Reis Holloway	257	51	105	82	3	7	1	4
Grand Total	93,827	2067	145,578	3,811	188	7	52	162

FIGURE 5.25: Other relevant KPIs for top 40 salespeople sorted by growth

5.2.1.9 Salespeople performance evaluation model

Based on previous data, the author, together with the business, was able to define a set of rules to evaluate salespeople's performance.

These rules are based on a set of 42 KPI's where, based on the accumulated performance of the salesperson on each of the measures, a classification is possible to define for the salesperson. These categories are: Not Performing, Good, and Outstanding.

As displayed in figure: 5.26 these Categories will be based on following KPIs: Customer Base: The Customer Base is the number of customers assigned to the salesperson, where Small is defined for a Customer Base with less than 30 customers, the Medium is between 30 and 49, and Large for 50 or greater.

Customer Baseline is the Sum of the amount of TEU's sold to all the accounts in the Customer Base for the previous year.

Forecasting is the ability to use other means of forecast like Sales Pipeline, as displayed in the Figure: 5.26 this one is in red because the dataset used for this research do not have sales pipeline data; therefore this will not be considered for the current study.

The target is calculated by comparing the Customer Baseline with the Target, using formula 5.2. Based on the result of Formula the considerations are: Unrealistic when Baseline is 0 or Target is more significant than 130% of the baseline, other values are: Very Low when is under 5%, low between 5% and 10%, average between 10% and 20%, Medium between 20% and 49%, high between 50% and 100%, and between 100% and 130% very high.

The Target Achievement is calculated based on before described Formula: 5.1, by dividing Target by Growth, the result is then evaluated by following ranges: Very Bad as under 10%, Bad by between 11% and 20% Average between 21% and 49% Somehow good by between 50% and 69% Good is the range between 70% and 99% and very good above 100%.

Each product will evaluate Customer Baseline, Target, Growth, Target achievement, and Growth percentage in separated, and an additional category for product growth will be added for remaining products.

The Nº of Opportunities created evaluates the number of opportunities the salesperson created, and it is categorized by Very Low less than 1, Low between 2 and 4, Average between 5 and 36, Medium between 37 and 49, Medium between 50 and 99, high by between 100 and 101 and Very High above 101.

The Average Nº of opportunities per customer is only the Sum of opportunities divided by the number of customers, and evaluate the number of opportunities over the Customer Base. It is grouped by Small that is less than 4, Medium between 4 and 9 and large above 10.

FIGURE 5.26: Salespeople performance measure model

Average Nº of months with growth above 0, evaluate the ability of the salesperson to keep the growth of he/she's Account Base over the year and also ability to spread the seasonality of the Account Base, this measure is grouped by: Small as less than 5, Medium between 5 and 7 and high above 8.

The Growth with different products evaluates the ability of the salesperson to sell more than one product, and this is graded by Yes or no and the Number of Products with Positive Growth is grouped by Small for at least 1, Medium 2, and Large more than 3.

5.2.2 Report to evaluate salesperson performance

Based on the rules discussed above, the author provided a new report composed of 45 fields with sales and other information to the business. So the business can evaluate the performance of the salespeople, classify manually with the levels defined, and return so classification can be made with PA. The report has the following requirements:

- Data should reflect two different years respectively 2017, and 2018
- Data should be grouped by, salesperson and year
- The volumes, baseline, growth, target, and achievement must be provided in separate columns for each of the six main products, and one extra group for the remaining products together
- Monthly variability must be part of the report
- Opportunity information must be included
- Target and achievements have to be included in volume and percentages

For the structure, the following fields compose the report:

- Sales_Person_Code, the internal identification of the salesperson
- Sales_Person_Name, the real name of the salesperson, so they could validate data with other internal systems
- Year, the year the data refers to
- Growth_All_Products, the growth the salesperson brought on that year for all the products together
- Customer_base, count of the number of customers assigned to the salesperson for each year

- Base_Line_All_Products, the TEU's sold for all the Customer Base of the salesperson in the previous year for all the products
- Growth_Percent_All_Products, the result of the Growth_All_Products divided by the Base_Line_All_Products
- Target_All_Products, the sum of all the targets defined for each salesperson in one year, for all the products together
- Target_Achievement_All_Products, the target achievement for all the products, this is achieved by dividing the Growth_All_Products by the Targets
- N_Opp_created, count of the number of won opportunities by the salesperson for each year
- Avg_Opp_per_Cust, the average number of won opportunities per customer for each year
- Nº_Months_with_growth_above_0, the number of months with growth above 0, for each year (for all the products together)
- N_Prod_With_Po_Gw, count of the number of products that the salesperson can grow on one year
- Grow_Df_Prod, Yes/No field identifying if the salesperson is able to grow more than one product on the year
- for each of the six main products, add the following fields with performance indicators:
  - Targ, the defined Target for the whole year
  - Gw, the sum of all the growth for the year
  - Baseline, the sum of all the Previous year TEU's for all the Customer Base of one year
  - the growth_Percent, the result of the column Gw divided by the Baseline

 - Target_Achiev, the calculation of the Target Achievement for the year

- Remaining_Products_Growth_Percentage same as Growth_Percentage_All_Products but filtered to exclude all the six main products

The a sample of the report is provided on this work in the Figure: 5.27 for better understanding (Salesperson Number removed for data protection and name is replaced by an invented name). The report was submitted for the business to evaluate the salesperson and classify them based on their performance.

Field	Sample 1	Sample 2
Sales_Person_Code	0000000000	0000000000
Sales_Person_Name	Bella Connor (Belle)	Anastasia Sullivan (Stacy)
Year	2018	2017
Growth_All_Products	12,617	3,791
Customer_Base_All_Products	12	61
Base_Line_All_Products	9,366	2,928
Growth_Percent_All_Products	134.71%	129.47%
Target_All_Products	750	560
Target_Achievement_All_Products	1682.27%	676.96%
N_Opp_created	202	46
Avg_Opp_per_Cust	16	1
Nº_Months_with_growth_above_0	11	12
N_Prod_With_Po_Gw	3	2
Grow_Df_Prod	Yes	Yes
Ocean_FCL_Import_Target	0	224
Ocean_FCL_Import_Growth	0	3,423
Ocean_FCL_Import_Base_Line	0	2,748
Ocean_FCL_Import_Growth_Percent	0.00%	124.56%
Ocean_FCL_Import_Target_Achievement	0.00%	1528.13%
Ocean_FCL_Export_Target	750	336
Ocean_FCL_Export_Growth	428	368
Ocean_FCL_Export_Base_Line	4,151	180
Ocean_FCL_Export_Growth_Percent	10.31%	204.44%
Ocean_FCL_Export_Target_Achievement	57.07%	109.52%

Field	Sample 1	Sample 2
Reefer_Export_Target	0	0
Reefer_Export_Growth	-1	0
Reefer_Export_Base_Line	25	0
Reefer_Export_Growth_Percent	-4.00%	0.00%
Reefer_Export_Target_Achievement	0.00%	0.00%
Ocean_FCL_Cross_Trade_Target	0	0
Ocean_FCL_Cross_Trade_Growth	0	0
Ocean_FCL_Cross_Trade_Base_Line	0	0
Ocean_FCL_Cross_Trade_Growth_Percent	0.00%	0.00%
Ocean_FCL_Cross_Trade_Target_Achievement	0.00%	0.00%
Freight_Management_FCL_Target	0	0
Freight_Management_FCL_Growth	12391	0
Freight_Management_FCL_Base_Line	4858	0
Freight_Management_FCL_Growth_Percent	255.06%	0.00%
Freight_Management_FCL_Target_Achievement	0.00%	0.00%
Reefer_Import_Target	0	0
Reefer_Import_Growth	4	0
Reefer_Import_Base_Line	0	0
Reefer_Import_Growth_Percent	100.00%	0.00%
Reefer_Import_Target_Achievement	0.00%	0.00%
Remaining_Products_Growth_Percent	0.00%	0.00%

FIGURE 5.27: Sample of report data

5.2.3 Classification by the business

The report was delivered in excel with all the salespeople and metrics for years 2017 and 2018 to be classified manually. The business did the manual classification over the report, and delivered the same excel with a new column filled with the categories Not Performing, Good and Outstanding, these categories for the business represented the following:

- Not Performing: as someone who has no growth, low or no Opportunities created, low target achievement and low growth over the months on one year

- Good: as someone who was able to grow the baseline on at least 2 products, have steady growth for at least 7 months, have some opportunities but not mandatory to have many and, a good target achievement
- Outstanding: as someone who has grown on more than 2 products or extremely high growth on one, a steady performance along 8 months or more, have a good or high target achievement based on a large baseline and large targets

The author now reaches the end of this chapter, where an ETL process was developed to extract, transform and load data to Power BI, reports were built, and several KPIs for salesperson performance assessment were evaluated. In the next stage of the research, the author describes the steps executed to prepare data for classification and then the use of predictive analytics to classify salespeople.

Chapter 6

Salespeople classification through predictive analytics

6.1 Naive Bayes

In the research, from the studied algorithms, the author selected the Naive Bayes (NB) because of ease of it's implementation. The NB algorithm is a probabilistic classifier that selects each independent variable, and then associate it to a conditional probability. The conditional probability is calculated based on the following formula: 6.1

$$P(C|A) = \frac{P(A|C) * P(C)}{P(A)} \tag{6.1}$$

The algorithm calculates the probability of an event occurs, based on another event that occurred in the past. For example, to predict if a salesperson may achieve his targets. In the formula, we can associate C to the probability of a salesperson achieve his targets, while A corresponds to the conditions that allowed the salesperson to achieve the targets, for instance, a customer base composed by customers that buy high volumes of TEU's.

6.2 Prepare data for classification

In this chapter, as defined in the research steps displayed in Figure: 6.1, the tasks executed to prepare data to be modeled are described and executed.

FIGURE 6.1: Research steps, prepare data for classification

6.2.1 The new dataset description

As described before, the new dataset is composed of 46 columns and 695 Rows. From the 46 columns, four have categorical data: these are Sales_Person_Code, Sales_Person_Name, Year, and Talent. The remaining columns have numerical data containing the salesperson's performance. A summary of the data available in the dataset is provided in the Table: 6.1 for reference. The columns Sales_Person_Code, Sales_Person_Name and Year, were removed from the dataset, leaving the dataset with 43 columns.

In the next sections, the author submits the dataset to several techniques that evaluates the importance that each column may have to the model, and eliminates all the ones that contributes little or none.

6.2.2 Near Zero Variance

Columns with low variance on the data, provide little or no knowledge to models, so to improve the performance of the model, these columns can be eliminated. To Identify the columns that provide low knowledge, the author used the function

Table 6.1: Table with classification statistics

Field	Sample Value	Min	1st Quartile	Median	Mean	3rd Quartil	Max
Talent	Good	Not Performing: 373, Good: 269 and Outstanding: 53					
Growth_All_Products	-46	-1.790	-26,5	37	138,4	205,5	12.617
Customer_Base_All_Products	4	1	7	14	15,97	22	83
Base_Line_All_Products	78	0	40,5	204	633,4	631,5	12.443
Growth_Percent_All_Products	-0,59	-1	0,17	0,18	3,17	1,08	566
Target_All_Products	0	0	120	272	405,7	560	6.200
Target_Achievement_All_Products	0	-293	0,10	0,11	1,07	0,66	88
Nº_Opportunities_created	0	0	2	10	17,83	24	202
Average_Nº_of_Opportunities_per_customer	0	0	1	1	1,13	1	16
Nº_Months_with_growth_above_0	3	0	3	6	5,87	9	12
Nº_Different_Products	0	0	0	0	0,35	1	1
Grow_with_Different_Products	0	0	1	1	1,30	2	6
Ocean_FCL_Import_Target	0	0	0	101	220,9	300	2.250
Ocean_FCL_Import_Growth	0	2.193	0	0	72,88	87	3.423
Ocean_FCL_Import_Base_Line	0	0	0	37	303.800	233	12.199
Ocean_FCL_Import_Growth_Percent	0	-1	0	0	1.853	1	110.429
Ocean_FCL_Import_Target_Achievement	0	0	0	0	0,03	0,41	88
Ocean_FCL_Export_Target	0	0	10	100	177,50	240	6.200
Ocean_FCL_Export_Growth	-46	1.790	-7,5	0	35,07	56	3,224
Ocean_FCL_Export_Base_Line	78	0	0	28	279,50	195,50	8.413
Ocean_FCL_Export_Growth_Percent	-0,59	-1	-0,18	0	2,27	1	517,67
Ocean_FCL_Export_Target_Achievement	0	-293	-0,03	0	-1,00	0,33	27
Reefer_Export_Target	0	0	0	0	4,94	0	2.000
Reefer_Export_Growth	0	-771	0	0	2,48	0	1.923
Reefer_Export_Base_Line	0	0	0	0	16,41	0	5.585
Reefer_Export_Growth_Percent	0	-1	0	0	0,102	0	46,8
Reefer_Export_Target_Achievement	0	-2	0	0	0,02	0	9,36
Ocean_FCL_Cross_Trade_Target	0	0	0	0	1,49	0	229
Ocean_FCL_Cross_Trade_Growth	0	-320	0	0	2,41	0	706
Ocean_FCL_Cross_Trade_Base_Line	0	0	0	0	4,07	0	1.270
Ocean_FCL_Cross_Trade_Growth_Percent	0	-1	0	0	0,24	0	45,56
Ocean_FCL_Cross_Trade_Target_Achievement	0	3,7	0	0	0,002	0	3,43
Freight_Management_FCL_Target	0	0	0	0	0,76	0	530
Freight_Management_FCL_Growth	0	-595	0	0	23,99	0	12.391
Freight_Management_FCL_Base_Line	0	0	0	0	24,07	0	4.858
Freight_Management_FCL_Growth_Percent	0	-1	0	0	0,24	0	55,5
Freight_Management_FCL_Target_Achievement	0	0	0	0	0,001	0	0,51
Reefer_Import_Target	0	0	0	0	0,11	0	24
Reefer_Import_Growth	0	-186	0	0	0,91	0	172
Reefer_Import_Base_Line	0	0	0	0	3,11	0	501
Reefer_Import_Growth_Percent	0	-1	0	0	0,34	0	155
Reefer_Import_Target_Achievement	0	-0,65	0	0	-0,001	0	0,58
Remaining_Products_Growth_Percent	0	-0,39	0	0	0,64	0	408

nearZeroVar from the carret package. This function diagnoses the predictors that have one unique value, or predictors that have few unique values relative to the number of samples and the ratio of the frequency, from the most common value to the frequency of the second most common value.

From the results provided by the function, the most importants are zeroVar that has TRUE when the column contains only one distinct value and nzv, which has TRUE when the column in question has a near-zero variance predictor, for reference, the results are provided in the Table: 6.2.

TABLE 6.2: Result of the nearZeroVar function

Column	**freqRatio**	**percentUnique**	**zeroVar**	**nzv**
Talent	1,39	0,43	FALSE	FALSE
Growth_All_Products	1,14	66,91	FALSE	FALSE
Customer_Base_All_Products	1,03	7,77	FALSE	FALSE
Base_Line_All_Products	3,91	64,89	FALSE	FALSE
Growth_Percent_All_Products	9,20	90,50	FALSE	FALSE
Target_All_Products	1,64	38,99	FALSE	FALSE
Target_Achievement_All_Products	14,67	89,93	FALSE	FALSE
Nº_Opportunities_created	3,25	12,52	FALSE	FALSE
Average_Nº_of_Opportunities_per_customer	3,83	1,29	FALSE	FALSE
Nº_Months_with_growth_above_0	1,00	1,87	FALSE	FALSE
Nº_Different_Products	1,97	1,01	FALSE	FALSE
Grow_with_Different_Products	1,87	0,29	FALSE	FALSE
Ocean_FCL_Import_Target	7,89	28,63	FALSE	FALSE
Ocean_FCL_Import_Growth	25,25	46,91	FALSE	FALSE
Ocean_FCL_Import_Base_Line	25,22	44,60	FALSE	FALSE
Ocean_FCL_Import_Growth_Percent	5,61	64,17	FALSE	FALSE
Ocean_FCL_Import_Target_Achievement	80,67	63,45	FALSE	FALSE
Ocean_FCL_Export_Target	5,23	28,78	FALSE	FALSE
Ocean_FCL_Export_Growth	12,83	44,75	FALSE	FALSE
Ocean_FCL_Export_Base_Line	14,07	41,29	FALSE	FALSE
Ocean_FCL_Export_Growth_Percent	2,41	62,88	FALSE	FALSE
Ocean_FCL_Export_Target_Achievement	48,50	65,61	FALSE	FALSE
Reefer_Export_Target	136,40	1,44	FALSE	TRUE
Reefer_Export_Growth	218,33	4,75	FALSE	TRUE
Reefer_Export_Base_Line	221,33	3,60	FALSE	TRUE
Reefer_Export_Growth_Percent	54,58	2,88	FALSE	TRUE
Reefer_Export_Target_Achievement	687,00	1,29	FALSE	TRUE
Ocean_FCL_Cross_Trade_Target	136,00	1,58	FALSE	TRUE
Ocean_FCL_Cross_Trade_Growth	214,00	5,47	FALSE	TRUE
Ocean_FCL_Cross_Trade_Base_Line	163,75	3,74	FALSE	TRUE
Ocean_FCL_Cross_Trade_Growth_Percent	49,38	4,60	FALSE	TRUE
Ocean_FCL_Cross_Trade_Target_Achievement	685,00	1,58	FALSE	TRUE
Freight_Management_FCL_Target	694,00	0,29	FALSE	TRUE
Freight_Management_FCL_Growth	25,33	10,94	FALSE	FALSE
Freight_Management_FCL_Base_Line	39,67	7,48	FALSE	TRUE
Freight_Management_FCL_Growth_Percent	7,82	8,20	FALSE	FALSE
Freight_Management_FCL_Target_Achievement	694,00	0,29	FALSE	TRUE
Reefer_Import_Target	344,50	0,72	FALSE	TRUE
Reefer_Import_Growth	39,00	4,17	FALSE	TRUE
Reefer_Import_Base_Line	64,80	3,45	FALSE	TRUE
Reefer_Import_Growth_Percent	20,13	4,32	FALSE	TRUE
Reefer_Import_Target_Achievement	691,00	0,72	FALSE	TRUE
Remaining_Products_Growth_Percent	42,86	5,61	FALSE	TRUE

There are 19 columns identified by the nearZeroVar function to be removed. After the removal of the 19 columns, the dataset still has 24 columns, 23 numerical +

the Talent column.

6.2.3 Correlation matrix

After the removal of the columns with low variance, the author made a correlation matrix between the remaining columns to find the ones that are highly correlated and remove at least one of them. For that, the author used the function cor from the caret package. The cor function computes the variance, and the covariance of x and y. The results are a percentage of correlation between columns.

FIGURE 6.2: Correlation matrix

The result of the correlation matrix, as presented in Figure: 6.2, shows that there are 6 columns highly correlated (above 0.8). The author eliminated three of the six columns, specifically: (Grow_with_Different_Products, Ocean_FCL_Export_Target_Achievement, and Ocean_FCL_Export_Growth_Percent). The dataset has now 21 columns, 20 numeric + the Talent.

6.2.4 Outliers treatment

After removing the columns that contribute less, and the columns that are highly correlated, the author did an outlier analysis to the remaining columns of the dataset. Currently, there are 21 columns in the dataset, including the Talent column, which is the column with the classification.

The dataset has a high number of outliers, as it's possible to verify in Figure: 6.3. To identify the outliers, the author used the boxplot.stats function of R. This function is typically called by another function to build the boxplot. With it, the author managed to identify the outliers for all the 20 numeric fields.

To not remove data from the small dataset (695 rows), the outlier treatment made by the author, focused on applying to every outlier, the values in the range limit. The lower and higher values applied are in Table 6.3 for reference, limits were applied to all columns except column: Nº_Months_with_growth_above_0.

TABLE 6.3: Outlier conversion table

Field	Min value	Max value	Applied lower-Value	Applied higher Value
Growth_All_Products	-1.790	12.617	-373	552
Customer_Base_All_Products	1	83	1	44
Base_Line_All_Products	0	12.443	0	1.468
Growth_Percent_All_Products	-1	566	-1	2,94
Target_All_Products	0	6.200	0	1.200
Target_Achievement_All_Products	-293	88	-1,18	1,76
Nº_Opportunities_created	0	202	0	57
Average_Nº_of_Opportunities_per_customer	0	16	1	1
Nº_Months_with_growth_above_0	0	12	0	12
Nº_Different_Products	0	6	0	3
Ocean_FCL_Import_Target	0	2.250	0	750
Ocean_FCL_Import_Growth	-2.193	3.423	-129	215
Ocean_FCL_Import_Base_Line	0	12.199	0	571
Ocean_FCL_Import_Growth_Percent	-1	110,43	-1	2,47
Ocean_FCL_Import_Target_Achievement	-185,5	88	-0,60	1,02
Ocean_FCL_Export_Target	0	6.200	0	583
Ocean_FCL_Export_Growth	-1.790	3.224	-102	148
Ocean_FCL_Export_Base_Line	0	8.413	0	482
Freight_Management_FCL_Growth	-595	12.391	-100	100
Freight_Management_FCL_Growth_Percent	-1	55,5	0	20

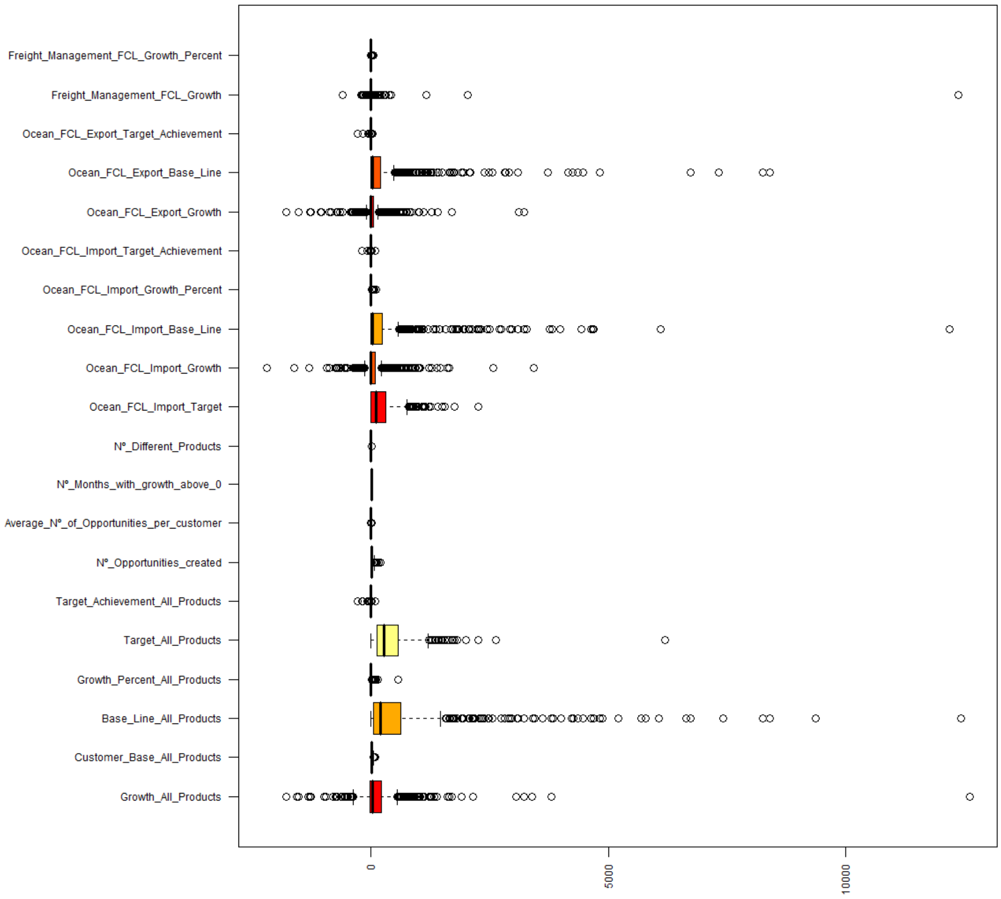

FIGURE 6.3: Outlier display

After all the evaluation made, the author discussed with the business the added value of the columns that refers to specific products, like Ocean FCL Export and Ocean FCL Import. The fact that these 2 products are the only ones in the model would bias the salespeople that succeed more on these 2 products over the other remaining products. Although the Overall Growth is still part of the module, the removal of all the columns specific for the products would produce similar results and with more value to the business. This lead to the removal of the other 10 columns. After the removal of these 10 columns, the dataset got reduced to 11 columns 10 numeric + 1 categorical.

6.2.5 Normalize data

After the completion of all the data treatment steps, and as the NB from R requires all the numeric columns to be Standardized. The author Standardized all the numeric columns using the function normalize of R from the BBmisc package.

6.3 The model

The data was split into 2 separate datasets using the sample function in R, the training dataset with 70% of the data, which corresponds to 481 observations and the test dataset with 214 observations.

The model was created using the package naivebayes of R. The values to be predicted are in the column Talent and has the following values:

- Not Performing, as someone who has no growth, low or no Opportunities created, low target achievement and low growth over the months on one year
- Good as someone who was able to grow the baseline on at least 2 products, have steady growth for at least 7 months, have some opportunities but not mandatory to have many, good target achievement
- Outstanding as someone who has grown on all products or extremely high growth on one, a steady performance along 8 months or more, have a good or high target achievement based on a large baseline

These classifications, as mentioned in previous chapters, were defined by the business and refers to salespeople classification for sales regarding 2017 and 2018.

The a priori probabilities of the model are:

- Not Performing: 0,52
- Good: 0,40

- Outstanding: 0,08

The Mean and Standard Deviation for each of the features in the model are displayed in Table 6.4. Analyzing the numbers that describe the model, the a priori probabilities reflect the amount of data that is classified. The amount of people Not Performing is more than half of the dataset, the Good is the next with about 40%, and the Outstanding is represented in a small number.

Analyzing the model details provided in the Table: 6.4, we realize that the column Growth_All_Products, for the Not Performing category have a very low mean and the standard deviation are higher between the classes. The Good has a similar distance provided by the standard deviation and a positive mean. The Outstanding has a high mean, but the distance between salespeople is minimal based on the standard deviation, this is also the column that has the most influence in the Outstanding classification.

Table 6.4: Model details

Feature	Values	Not Performing	Good	Outstanding
Growth_All_Products	mean	-0,75	0,58	1,91
	sd	0,55	0,63	0,03
Customer_Base_All_Products	mean	-0,21	0,24	0,92
	sd	0,94	0,96	1,20
Growth_Percent_All_Products	mean	-0,53	0,58	0,75
	sd	0,72	0,97	0,93
Target_Achievement_All_Products	mean	-0,67	0,49	1,48
	sd	0,56	0,75	0,97
Nº_Opportunities_created	mean	-0,22	0,17	0,94
	sd	0,92	0,96	1,12
Nº_Months_with_growth_above_0	mean	-0,68	0,69	1,36
	sd	0,66	0,70	0,45
Nº_Different_Products	mean	-0,43	0,45	0,63
	sd	0,92	0,80	0,89
Target_All_Products	mean	-0,1	0,06	1,02
	sd	0,94	0,78	2,19
Baseline_All_Products	mean	-0,03	-0,08	0,89
	sd	1,00	0,91	1,11
Average_No_of_Opportunities_per_customer	mean	-0,15	0,13	0,38
	sd	1,11	0,87	0,43

The column Customer_Base_All_Products has a very similar standard deviation for the Not Performing and Good, surprisingly for the Outstanding the Standard Deviation is high, which reflects the high variation on the number of customers assigned to a salesperson, in smaller countries a salesperson can have about 15 customers while in big countries a salesperson can have 70.

The standard deviation in Growth_Percent_All_Products and Target_Achievement_All_Products reflect that the Not Performing salespeople are more concentrated, the distance for categories like Good and Outstanding is bigger. No_Opportunities_created also has a standard deviation higher for Outstanding than the Good and the Not Performing. The means of Not Performing and Good are very close to each other. This is probably one of the columns that can cause one of the higher errors between these 2 classes.

No_Months_with_growth_above_0, when analyzing this feature we can realize that the means of Not performing and Good are very distant from each other, this feature provides a good separation between these 2 classes, but between the Good and Outstanding the difference is smaller which may help in the separation of these 2 classes (Not Performing and Good).

The Nº_Different_Products, the difference between the Good and Outstanding is small, which reflects that the salespeople classified as Good are already capable of selling more than one product.

The Target_All_Products and Baseline_All_Products, have both a big difference between Good and Outstanding, highlighting that to be an outstanding salesperson, the person has to work with extremely high targets and baseline.

The last feature analyzed is the Average_No_of_Opportunities_per_customer, this feature represents that although the average is low in order to be Good there has to be opportunities and to be Outstanding a salesperson have to have many more won opportunities.

The model is now created and ready to make the predictions. In the next step, the author describes the results and evaluation of the predictions.

Chapter 7

Evaluation

According to our methodology, as presented in Figure: 7.1, the next stage is the identification of the most important factors to salespeople success.

FIGURE 7.1: Most important factors for salesperson success

7.1 Identify most important factors for salesperson success

To achieve the goals: Identify the most important factors for salesperson success, the author built a Random Forest model with the same train dataset prepared for the NB model, but with the randomForest of R so that the function varImp could be used. The Random Forest model was created using the defaults of R, adding the following parameters: Type of random forest: classification, number of trees: 500 and No. of variables tried at each split: 2. The results were: Out of Bag (OOB) estimate of error rate: 2,91% and the confusion matrix provided in the Table: 7.1.

TABLE 7.1: Confusion matrix from Random Forest

	Good	Not Performing	Outstanding	class.error
Good	186	0	5	0,026
Not Performing	1	250	0	0,004
Outstanding	7	0	32	0,179

The results of the varImp function are provided in the Table: 7.2.

TABLE 7.2: Feature importance

Feature	**Overall**
Growth_All_Products	132,94
Target_Achievement_All_Products	54,46
Growth_Percent_All_Products	26,35
Nº_Months_with_growth_above_0	23,59
Target_All_Products	9,61
Baseline_All_Products	7,87
Customer_Base_All_Products	6,06
Nº_Opportunities_created	4,63
Nº_Different_Products	4,47
Average_Nº_of_Opportunities_per_customer	0,22

The results show that the most important features are: Growth_All_Products, Target_Achievement_All_Product, Growth_Percent_All_Products and Nº_Months_with_growth_above_0, the remaining columns have residual importance compared to the ones before mentioned. The results go in line with the business people's opinions. The salesperson to succeed, have to: focus on growing the customer base, work to achieve their targets and, have steady sales for as many months as possible.

7.2 Run the Classification

To validate if it's possible to increase the accuracy of the NB model, a 20 Fold Cross Validation NB model was created, there was an increase on the accuracy from 90,65% to 92.10%. Based on this new model, the testing dataset was loaded and the predictions was requested.

A confusion matrix was built to evaluate the performance of the predictions made over the test dataset. The results are displayed in the table: 7.3.

The Accuracy (average) of the model is 92,10%. Based on the Confusion Matrix provided in the Table: 7.3 it's possible to verify that the model only failed in 7,9% of the cases.

TABLE 7.3: Confusion matrix

	Good	**Not Performing**	**Outstanding**
Good	36,0	3,3	0,8
Not Performing	1,9	48,9	0,0
Outstanding	1,9	0,0	7,3

An evaluation of the Precision, Specificity, Sensitivity, and an F1 score was made to evaluate the model accuracy and the results. As it's possible to verify in the Table: 7.4, the Outstanding has a high Specificity but has a lower Sensitivity.

TABLE 7.4: Evaluation scores for NB model

	Good	**Not Performing**	**Outstanding**
Sensitivity	83%	97%	68%
Specificity	93%	87%	99%
Pos Pred Value	88%	89%	93%
Neg Pred Value	90%	97%	97%
Precision	88%	89%	93%
Recall	83%	97%	68%
F1	86%	93%	79%
Prevalence	39%	52%	9%
Detection Rate	32%	51%	6%
Detection Prevalence	36%	57%	7%
Balanced Accuracy	88%	92%	84%

The F1 score display that the precision of the Not Performing is the highest, but for the Outstanding and Good classes, the accuracy of the tests made are high, which is very important considering that the results of this model are to evaluate people performance.

As the example, in the Figures: 7.2, 7.3, and 7.4, it's possible to review the results of the assessment in Power BI on a dashboard created for salesperson assessment, the dashboard has all the metrics and a classification made by the Predictive Analytics as Not Performing, Good and Outstanding, with this all the objectives of the research are concluded successfully.

Above steps conclude the evaluation of the model performance. This was the last task in the research. In the next chapter, the author concludes the research with a summary of the work and suggestions for future work.

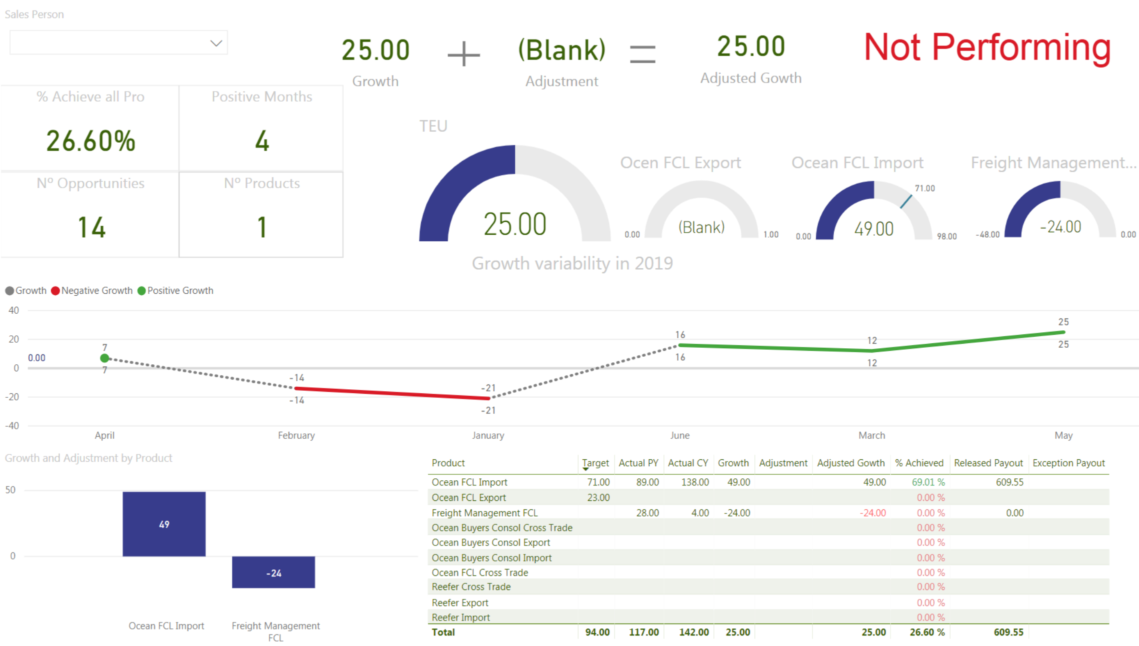

FIGURE 7.2: Dashboard for a salesperson classified as Not Performing

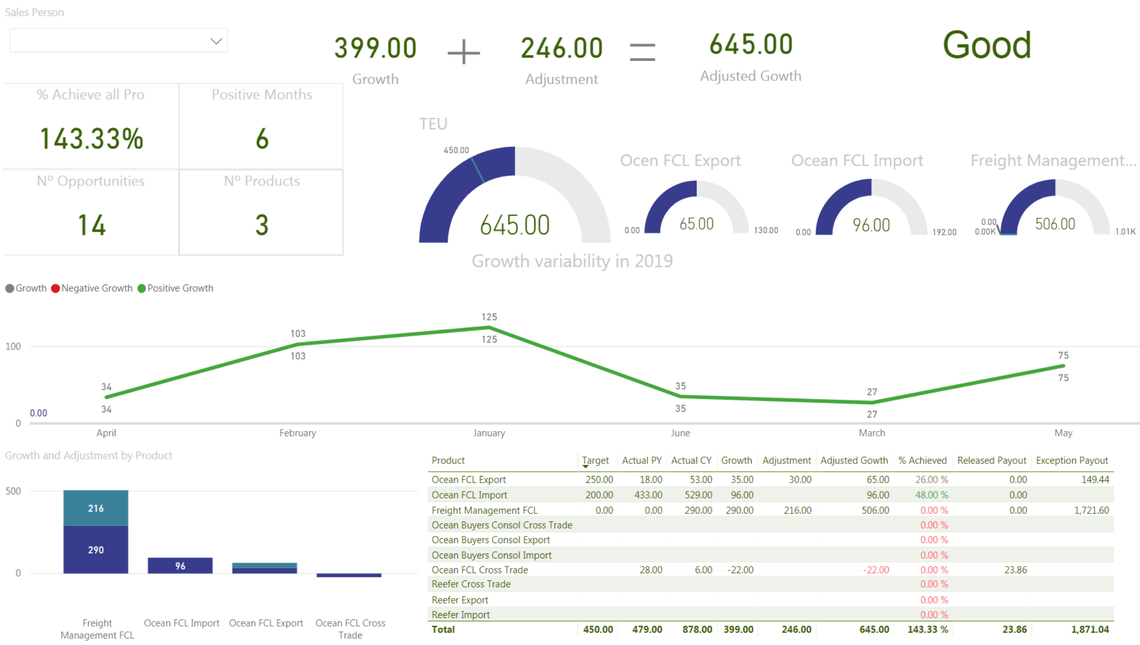

FIGURE 7.3: Dashboard for a salesperson classified as Good

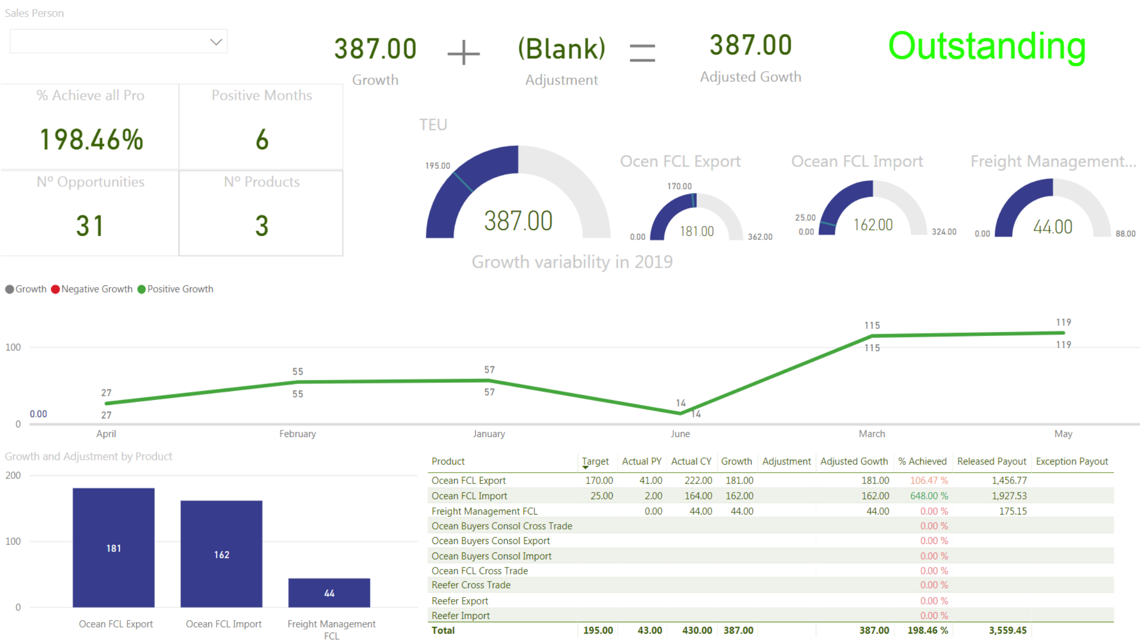

FIGURE 7.4: Dashboard for a salesperson classified as Outstanding

Chapter 8

Conclusions

8.1 Work done and goals

Salespeople's performance measurement is a process that occurs multiple times per year on a company. During this process, the manager and the salesperson evaluate how the salesperson performed on various KPI's.

Evaluating these KPI's may be time-consuming, depending on the performance measurement process of the company. The KPI's often include the amount of products/services sold by the salesperson, the number of opportunities created, the ability to sell multiple products/services, the variability of the sales along the year, among many others. To prepare the evaluation meeting, managers have to gather data from Customer Relationship Management (CRM), Financial Systems, Operational Systems, Business Intelligence, and sometimes excel files. The measurement of the performance often includes a compensation plan to motivate the salesperson to sell more products/services.

In this work, the author aimed to use descriptive analytics to evaluate the KPIs used for salesperson assessment, after that, recurring to with predictive analytics, classify salespersons into pre-defined categories, based on complex combinations of KPIs.

To select the KPIs, the author, together with the business, used multiple Power BI reports to assess different KPIs on the company data, and validate their ability to represent the salespeople's performance. The result was a report with all the KPIs and a manual classification completed by the business and provided to train the machine learning.

The classification is based on a report with 659 salespeople classified into three categories, Not Performing, Good and Outstanding, being these classes:

- Not Performing: as someone who has no growth, low or no Opportunities created, low target achievement and low growth over the months on one year
- Good: as someone who was able to grow the baseline on at least two products, have steady growth for at least seven months, have some opportunities but not mandatory to have many and, a good target achievement
- Outstanding: as someone who has grown on more than two products or extremely high growth on one, a steady performance along eight months or more, have a good or high target achievement based on a large baseline and large targets

The dataset used had in the beginning 46 columns. It was then reduced to 11 columns, based on several techniques to clean the data and evaluate the relevance of the columns to classify a salesperson's success. In this process, the author also identified the most critical factors to evaluate a salesperson's performance as Growth amount on all the products, Target achievement on all the products, Growth percentage on all the products, and the Number of Months with Growth above 0. The business agreed with these metrics and requested a new dashboard on Power BI, including these metrics and the classification to use as a base for future salesperson performance evaluation.

The author applied a Naive Bayes (NB) model to classify salespeople into three pre-defined categories, training a model with a portion of the 659 salespeople from the report provided by the business. The classification was possible, and the results delivered to the business, which agreed to use the classification model in the 2019 salespeople performance assessment in the company.

8.2 Discussion

The classification model was built using the NB package of R over R Studio. The model presented a priori probabilities of 0,52 to the Not Performing Class, 0,40 for the Good, and 0,08 for the Outstanding classes.

The author evaluated the performance of the model with a confusion matrix and other techniques like True Positives, True Negatives, and F1 score. The results showed an Accuracy (average) of 92,10% for the whole model. For each of the classes in terms of precision, Not Performing has 89%, Good 88%, and Outstanding 93%. The F1 scores for Not Performing were 93%, for good 86%, and Outstanding 79%.

These results were discussed with the business, and they agreed to use the classification provided by the model for the performance evaluation of 2019. This evaluation will be used for a small set of salespeople as a prototype. Based on the results and adoption, assess how to progress with the model in the future.

The accuracy results in this work are high because the size of the dataset and the variations of data have similar behavior for each of the classes. For instance, a salesperson not performing has in most of the time, low growth, low number of opportunities, and sales above 0 for a small number of months in one year. A good salesperson may have good performance in at least six months and one product with good performance, and the outstanding salesperson should have growth extremely high for at least one product and months growth above 0 for at least eight months.

8.3 Future work

As for future work, the author proposes the use of a NB model to evaluate salespeople's performance with more CRM information. By taking advantage of other information that is also part of the salesperson job, information like the number of activities (Leads, Visits, Calls), the other opportunity states, opportunities conversion rate, and the costs involved for each of the salespeople. The inclusion of subjective factors can also be part of the salesperson's performance. For instance, a more experienced salesperson may be training a junior salesperson, or taking several lost customers to recover, these facts can have an impact on the sales performance of the salesperson, the inclusion of flags that rate these can also be included.

All to aim towards a detailed and precise evaluation of salespeople's performance, increasing the fairness and reduce drastically the amount of work needed to make a performance evaluation for the salesperson.

Bibliography

[Amazon, 2019] Amazon (2019). Cloud computing with AWS.

[Bose, 2009] Bose, R. (2009). Advanced analytics: opportunities and challenges. *Industrial Management & Data Systems*, 109(2):155–172.

[Choi et al., 2014] Choi, T. M., Hui, C. L., Liu, N., Ng, S. F., and Yu, Y. (2014). Fast fashion sales forecasting with limited data and time. *Decision Support Systems*, 59(1):84–92.

[Chris Manna, 2019] Chris Manna (2019). An intro to the CRISP-DM Methodology.

[Domingos and Van de Merckt, 2010] Domingos, R. and Van de Merckt, T. (2010). Best practices for predictive analytics in b2b financial services.

[Dw I R E S E et al., 2011] Dw I R E S E, T., Imhoff, C., and White, C. (2011). Self-Service BuSineSS intelligence TDWI besT pracTIces reporT. *TDWI - Transforming Data With Inteligence.*

[Faliagka et al., 2012] Faliagka, E., Ramantas, K., Tsakalidis, A., and Tzimas, G. (2012). Application of machine learning algorithms to an online recruitment system. In *Proc. International Conference on Internet and Web Applications and Services*. Citeseer.

[Fayyad et al., 1996] Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17(3):37–37.

[IBM, 2018] IBM (2018). Get started on IBM Cloud.

[Idoine et al., 2018] Idoine, C. J., Krensky, P., Brethenoux, E., Hare, J., Sicular, S., and Vashisth, S. (2018). Magic Quadrant for Data Science and Machine-Learning Platforms.

[Jing, 2009] Jing, H. (2009). Application of fuzzy data mining algorithm in performance evaluation of human resource. In *2009 International Forum on Computer Science-Technology and Applications*, volume 1, pages 343–346.

[Kawas et al., 2013] Kawas, B., Squillante, M. S., Subramanian, D., and Varshney, K. R. (2013). Prescriptive Analytics for Allocating Sales Teams to Opportunities. In *2013 IEEE 13th International Conference on Data Mining Workshops*, pages 211–218. IEEE.

[Kessler et al., 2007] Kessler, R., Torres-Moreno, J. M., and El-Bèze, M. (2007). E-gen: Automatic job offer processing system for human resources. In Gelbukh, A. and Kuri Morales, Á. F., editors, *MICAI 2007: Advances in Artificial Intelligence*, pages 985–995, Berlin, Heidelberg. Springer Berlin Heidelberg.

[Koti, 2013] Koti (2013). IJRET_110211019. *IJRET: International Journal of Research in Engineering and Technology*, eISSN pISS:2319–1163.

[Kotsiantis et al., 2006] Kotsiantis, S. B., Zaharakis, I. D., and Pintelas, P. E. (2006). Paper 1 (20). *Artificial Intelligence Review*, 26:159–190.

[Lilien, 2016] Lilien, G. L. (2016). The b2b knowledge gap. *International Journal of Research in Marketing*, 33(3):543 – 556.

[Lu, 2014] Lu, C.-J. (2014). Sales forecasting of computer products based on variable selection scheme and support vector regression. *Neurocomputing*, 128:491–499.

[Malisetty et al., 2017] Malisetty, S., Archana, R., and Kumari, V. (2017). Predictive analytics in hr management. *Indian Journal of Public Health Research & Development*, 8:115.

[Menon and Rahulnath, 2016] Menon, V. M. and Rahulnath, H. A. (2016). A novel approach to evaluate and rank candidates in a recruitment process by estimating emotional intelligence through social media data. In *2016 International Conference on Next Generation Intelligent Systems (ICNGIS)*, pages 1–6.

[Mirzaei and Iyer, 2014] Mirzaei, T. and Iyer, L. (2014). Application of predictive analytics in customer relationship management: a literature review and classification. In *Proceedings of the Southern Association for Information Systems Conference*, pages 1–7.

[Pitre et al., 2014] Pitre, R., Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., Chavan, M. V., and Phursule, P. R. N. (2014). A Survey Paper on Data Mining With Big Data. *AI magazine*, 5(3):37–53.

[Reday et al., 2009] Reday, P. A., Marshall, R., and Parasuraman, A. (2009). An interdisciplinary approach to assessing the characteristics and sales potential of modern salespeople. *Industrial Marketing Management*, 38(7):838–844.

[Rich et al., 1999] Rich, G. A., Bommer, W. H., MacKenzie, S. B., Podsakoff, P. M., and Johnson, J. L. (1999). Methods in sales research : Apples and apples or apples and oranges ? A ... *Journal of Personal Selling and Sales Management*, 19(4):41–52.

[Roulston et al., 2017] Roulston, S., Hansson, U., Cook, S., and McKenzie, P. (2017). If you are not one of them you feel out of place: understanding divisions in a Northern Irish town. *Children's Geographies*, 15(4):452–465.

[RStudio.com, 2019] RStudio.com (2019). R Studio.

[Ryu and Siegel, 2013] Ryu, S. and Siegel, E. (2013). Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie or Die Book Review. In *Healthc Inform Res*, volume 19, pages 63–65. John Wiley & Sons, Inc.

[SMART Vision, 2018] SMART Vision (2018). What is the CRISP-DM methodology?

[Sujeet N. Mishra, 2019] Sujeet N. Mishra, Dev Raghvendra Lama, Y. P. (2019). Human resource predictive analytics (hrpa) for hr management in organizations. *International Journal of Scientific & technology Research*, 5:3.

[Waller and Fawcett, 2013] Waller, M. A. and Fawcett, S. E. (2013). Data science, predictive analytics, and big data: A revolution that will transform supply chain design and management. *Journal of Business Logistics*, 34(2):77–84.

[Xiaofan and Fengbin, 2010] Xiaofan, C. and Fengbin, W. (2010). Application of data mining on enterprise human resource performance management. In *2010 3rd International Conference on Information Management, Innovation Management and Industrial Engineering*, volume 2, pages 151–153.

[Yan et al., 2015] Yan, J., Zhang, C., Zha, H., Gong, M., Sun, C., Huang, J., Chu, S., and Yang, X. (2015). On machine learning towards predictive sales pipeline analytics.

[Yu et al., 2013] Yu, X., Qi, Z., and Zhao, Y. (2013). Support vector regression for newspaper/magazine sales forecasting. In *Procedia Computer Science*, volume 17, pages 1055–1062. Elsevier.

[Zhao, 2008] Zhao, X. (2008). A study of performance evaluation of hrm: Based on data mining. In *2008 International Seminar on Future Information Technology and Management Engineering*, pages 45–48.

Appendices

Data Description details

.1 Tables Details

In this Appendix, there are functional description of all the tables included in the dataset. Below descriptions are what each table and fields mean for the business.

.1.1 Salesperson Details

The Sales_Person_details (current and history) is the thinnest grain of the dataset. It refers to all the shipments made in one month, for one product, shipped to one customer, on one city pair (source and destination country and city), and sold by one salesperson. Another important information about these tables is that it has stored the current and historical data. The following fields are part of this dataset:

- Unique_Grid_ID, unique identification key of a record
- Business_Key, a key that identifies a product, a customer, and a city pair, the goal is to determine from when the shipments started for a customer to a specific source and destination
- Transaction_ID, the opportunity id in the CRM system
- Payout_Period, incentive payments are based on one period, which can be one month or one quarter. This field has stored the label of the payout period where a transaction belongs. The tags are for example "JAN-2014" or "Q1-2014"

- SIP_Eligible, specify if the purchase has a matching opportunity, if so, then the transaction is marked as eligible for commission calculation
- Sales_Person_Name, the name of the salesperson responsible for the customer
- Sales_Person_Code, this is a ten-digit code that identifies the salesperson in the company systems. It is not the HR employee number. It is only a number to identify the salesperson between systems like CRM, ERP, IMS, etc.
- Customer_ACR_Name, the fiscal customer name
- Customer_ACR_Code, a ten-digit code to identify the customer among the multiple company systems
- District, identify the district where the salesperson works
- House_Name, identify the branch where the salesperson works
- House_Code, code that identifies the office where the salesperson works
- Partner_Country_Code is the two-digit ISO code of the country where the shipment is delivered
- City_Pair, A eleven digit code composed by two-digit of the source country, followed by the three-digit code to identify the city from, a dash, and then the two-digit code that identifies the destination country followed by the three-digit code that identifies the destination city
- Customer_Product_Name, the internal name of the product sold to the customer
- Roll up_Product_Name is the name of the product that the volumes of this row will be rolled up to
- Industry_Vertical identifies the internal vertical industry the customer belongs to, and it can be, for instance, Automotive, Chemicals, etc.

- Starting_Date, the information here, identify when the customer in question started using this business key (please refer to business _Key field for more information)
- Year, the year where the transactions occurred on
- Month, the month where the transaction occurred on
- Country_Code, the country where the salesperson works
- SIP_Plan_Code, identify the incentive plan where the salesperson belongs. The incentive plan is what stores all the schemes for an incentive, like for example, the payout periods, incentive rates among other information
- SIP_Plan_Name, the name that identifies the incentive plan internally on the company
- Payout_Method_Name, identify the method used to calculate the Incentives. There are two methods, Actual and Variance, when using Actual, the incentive is calculated, based on all the volumes sold during the current year. When the method is variance, the incentive is calculated based on the variance between last year and current year growth
- Payout_Frequency_Name, identify the frequency of the payments which can be quarterly or monthly
- Business_Age, identify for how long the customer makes transactions with the company for this business key. This number is set to 0 every time the customer starts using another company as the supplier and come back
- Business_Status, each business key on one year and one month can have a different status; generally, it is "New Business" or "Retained Business," the main purpose is to identify when the salesperson is getting new shipments or only retaining current shipments, as these involve a different amount of work for the salesperson

- Customer_Cluster_Name, identify the cluster level where the company configured on, for example, Global Key Account, Country Strategic Account, etc.
- UOM_Name, identifies the unit of measure related to the product
- Actual_CHF, identifies the amount shipped for the referred month, product and city pair on the current year
- Actual_PY_CHF, identifies the amount shipped for the referred month, product and city pair on the previous year
- Variance_CHF identify the variance between the previous two described fields
- Base_Value_CHF, store the value used for calculation of the incentive when the product sold is based on currency products, for example, net margin
- Exchange_Rate, all the products that have are base on currency, like Gross Proffit After Proffit Share (GPaPS) of Net Forwarding Revenue (NFR), need an exchange rate to convert figures from the local currency to the corporate currency which is CHF, this is the exchange rate of the transactions
- Currency, store the currency code of the country where the customer is invoiced
- Base_Value, store the amount used to calculate incentive or performance, for all the products based on volume (TON, CBM, TEU)
- Standard_Value, to calculate the incentives, amounts are separated between standard and over achievement, the standard value is the one used to calculate incentives while the growth is below target
- Standard_Rate, when calculating incentives, the rates used to calculate incentives before the growth reaches the targets rate defined here
- Standard_Payout stores the result of the standard value multiplied by the standard rate

- Over_Achievement_Value, to calculate the incentives, amounts are separated between standard and over achievement, over achievement value, is the incentive calculated for the growth above targets
- Over_Achievement_Rate, when calculating incentives, all the volumes used to calculate incentives after the target is achieved uses the rate defined here
- Over_Achievement_Payout stores the result of the overachievement value multiplied by the over achievement rate
- Tradelane_Percent to allow the company to boost or retract the sales on certain trade lanes, the system allows the country to set the trade lane percent, this will increase or decrease the incentive released for the trade lane
- Cluster_Percent, similar to the trade lane percent, it is also possible to have the incentive increased or decreased based on the customer cluster
- Total_Payout, stores the sum of the standard and the overachievement payout
- Retained_Percent, stores the percentage of incentive retained by the company because the growth is lower than the defined targets
- Released_Percent, stores the percentage of incentive released to the salesperson, due to targets achieved
- Retained_Payout, stores the incentive retained due to targets not achieved
- Released_Payout, stores the released incentives, based on the target achievement levels

.1.2 Payout Exceptions

Due to several factors like, for example, errors in operations, some shipments are not recorded correctly in the Transportation Management Systems (TMS), and when these cases occur, the Incentive Management System (IMS) is ready to accept what is called exceptions. These exceptions are volumes and the corresponding incentive that the salesperson is entitled to. This table has these exceptions recorded. The following fields are part of this table:

- Fields explained before in the sales_person_details table: SIP_Plan_Id, SIP_Plan_Name, Payout_Period, Sales_Person_Code, Customer_ACR_Code, Customer_ACR_Name, Currency_Code, Adjustment_Product_Name (product for the adjustment), Adjustment_Volume_Adjustment, Country_Code, Customer_Cluster_Name, Industry_Vertical .1.1.
- Exception_ID The internal Identification for the exception
- Payout_Impact and Adjustment_Payout_Adjustment, both refer to the payment that the salesperson will receive (there are two fields due to application upgrades)
- Adjustment_Against_KPI_Target, identify if the volume on these exceptions should count for the salesperson performance
- Adjustment_Volume_Adjustment, the volume that this exception refers to

.1.3 CRM Opportunities

The CRM Opportunities table holds the opportunities created by the salesperson for the sales done to their customers. One Opportunity can be created for one customer but can have multiple city pairs. One opportunity can have the following statues: New, Open, Win, Win & Implemented, Lost, and Canceled. For the current table, we have only Win & Implemented opportunities. The Win & Implemented are sold and successfully implemented. Following fields are part of this table:

- Fields explained before in the sales_person_details table: TRANSACTION_ID, SALES_RESPONSIBLE_CODE, CUSTOMER_ACR_CODE, CRM_PRODUCT_CODE, .1.1

.

- CITY_FROM_CODE is the source of the shipment, explained above in the salesperson details table
- CITY_TO_CODE is the destination of the shipment, explained above in the salesperson details table
- INDUSTRY_VERTICAL is the Industry Vertical (IV) the customer belongs to, more details in the salesperson details table explanation
- STATUS_OPPORTUNITY_ITEM_CODE is the status code of the opportunity status, for the data available in the dataset it will be always Win & Implemented
- COUNTRY_TO_CODE is the destination country of the shipment
- COUNTRY_FROM_CODE is the shipment source country
- COUNTRY is the country where the salesperson works on
- TARGET_CUSTOMER_FLAG is a flag that identifies if the opportunity is identified to the customer getting the shipment or not
- TM_FLAG, identify if the customer is in the SAP TM system or not

.1.4 CRM Products

The CRM Products is a table that holds the services provided by the company, as products. The table only has five columns with the following purpose:

- CRM_Product_Code is an internal code to identify the service. This code is the same in CRM, data warehouse and Incentive Management System
- CRM_Product_Name is the short name of the service provided
- UOM_Name is the Unit of Measure (UOM) used by the service, can be for instance TEU

- CRM_Product_Id is the IMS ID for the product
- Category_Name is the grouping identification to group products by categories

.1.5 Targets

The Targets table stores the targets for each product, salesperson, and year. The targets are defined every year and some times reviewed throughout the year. The following columns are part of the table:

- Sales_Person_Code is the eleven digits identification of the salesperson
- CRM_Product_id is the IMS internal product
- Target_Year is the year that the refers to
- Target_Month, this field stores the month that the target belongs to
- Monthly_Target is the amount of the target for one month
- Target_Currency_Code is the code that identifies the currency for money base targets (GPaPS or NFR)
- Month_AC is an integer that identifies the month and year of the target

.1.6 Geography

The geography table holds all the geographies where the company is based on and sends shipments to. Following columns are part of the table:

- ID is an internal identifier for the geography
- Region_Code is an internal code to identify a Region
- Region_Description is the region description

- Country_Group_Code is an internal code to identify aggregation of the countries
- Country_Group_Description is the description of the country group used internally by the company
- Country_Code is the two digits ISO based on two digits to identify the country
- Country_Description is the country name
- District_Code is the internal identifier for the districts where the company are based
- District_Description is the name of the district
- BU_Code is the internal code for the business unit
- BU_Description is the internal name for the business unit
- Branch_Code is the internal code for the branch
- Branch_Description is the internal name for the branch

.1.7 Currency

The currency table holds the currency code, country, and exchange rate of every month for all the countries where the company has offices. The following columns are part of the table:

- Currency_Code the three-digit ISO code that identify the currency
- Exchange_Rate the exchange rate between the currency and the corporate currency
- BW_GDATU a internal code that identifies the year and month the exchange rate belongs to

Extract Transform and Load data

.2 ETL Process

The company provided the data in a database called Thesis SIP 2.0. This database holds only a subset of the tables contained in the main application. From this database, a SQL Server Integration Services (SSIS) project was created to Extract, Transform, and Load (ETL) the data to a new database.

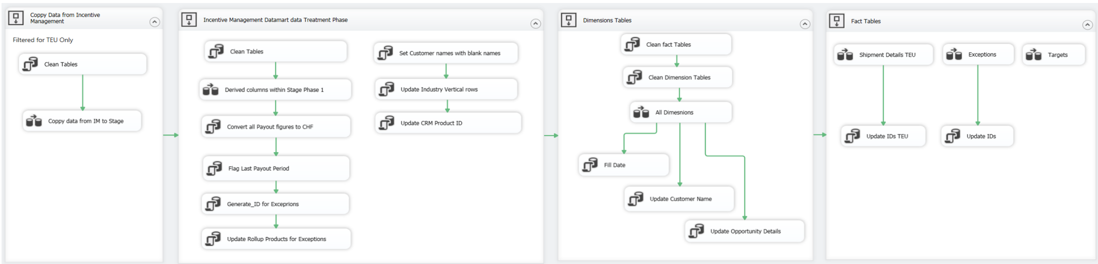

FIGURE 1: Integration Services overall

The Process starts by cleaning the stating tables and then copy the tables from the Thesis SIP 2.0 database to the staging tables in the SIP Datamart 3.0, as presented in Figure: 2

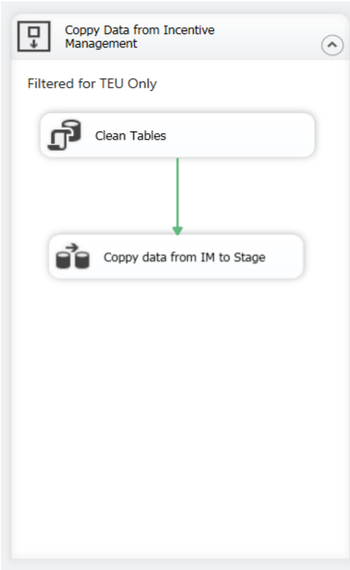

FIGURE 2: Copy data to staging

In the details of the Copy data from Incentive Management System (IMS) to stage, there are 9 data source and destination tasks, to copy the data from the salesperson details (current and history), products, opportunities, targets, geography, payout exceptions and currency as presented in Figure: 3 to the stage tables.

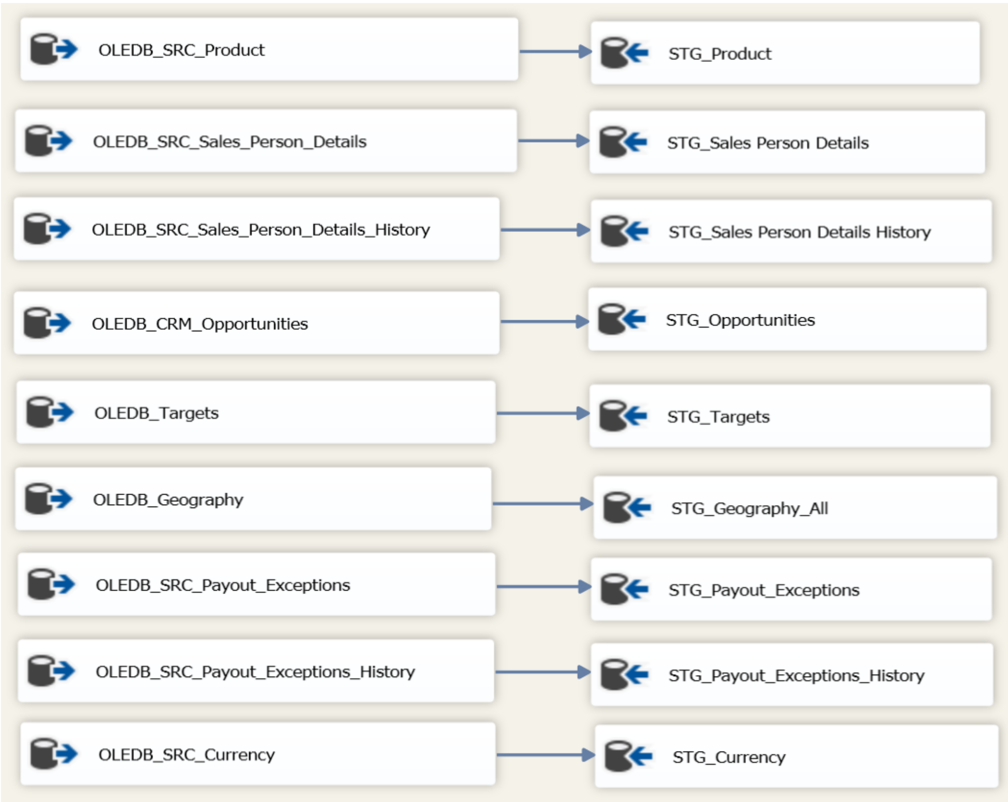

FIGURE 3: Copy data to staging details

The second stage of the project, as presented in Figure: 4, is where the data is prepared to be transformed into dimensions and facts. The process starts by cleaning the affected tables in this stage, then information from salespeople, customer, payout period, and exceptions is derived from the main tables, as presented in the Figure: 5.

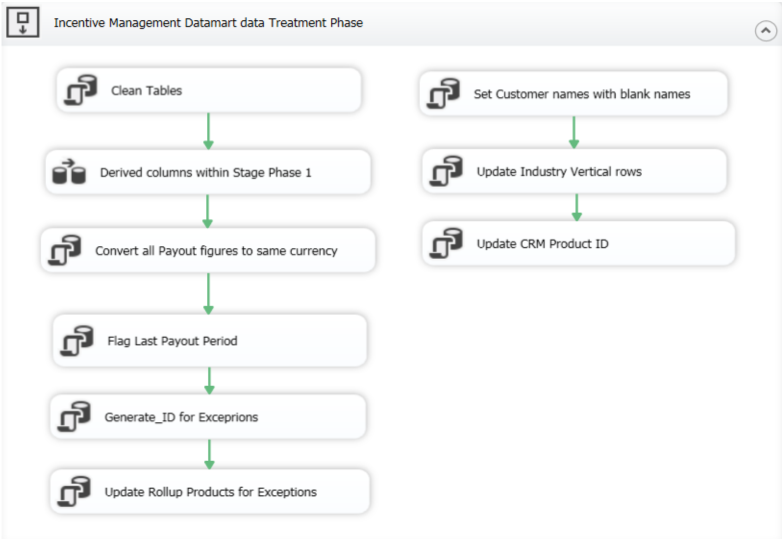

FIGURE 4: Incentive Management Datamart data Treatment Phase

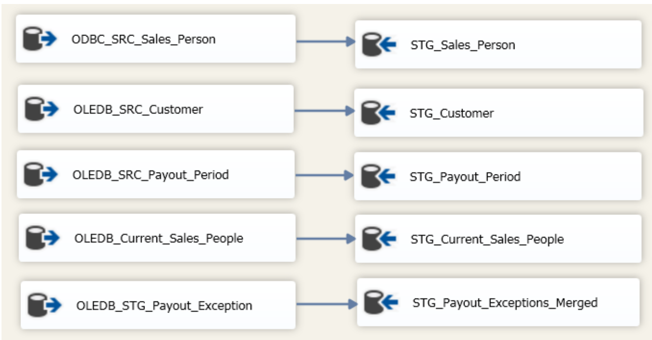

FIGURE 5: Derive data from staging tables

The process then converts all payout information to one currency only, and flag the payout periods that refers to latest payout periods, this step is required because for each payment period closed, the Incentive Management System keeps the history for auditing purposes, which leads to data multiplication. The process follows then by generating internal ids for exceptions. The last step of this flow transfers the volumes and payments from products being rolled-up to the main products.

In parallel, customers without names and IV's are updated and rollup products set with the leading products ids in the Sales_Person_Details table.

The process in the next stage fills all the dimension tables by copying data from the staging tables to the final dimension tables, as presented in Figure: 6.

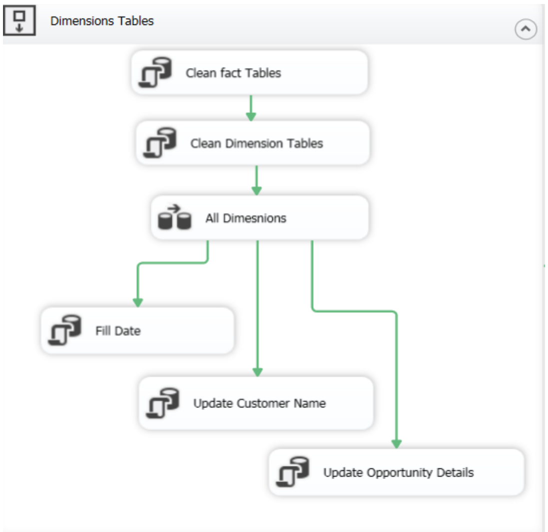

FIGURE 6: Dimensions stage

After filling all the dimensions, the process generates new internal ids. As mentioned before, each row of the salesperson details table refers to one month. A date column is generated using the month and year from the table, adding 1 as the day. This way, it's possible to create analysis with date filters based on the months.

In the last stage, the process fills all the fact tables. There are three fact tables, one with the details, another with exceptions, and the last one with targets.

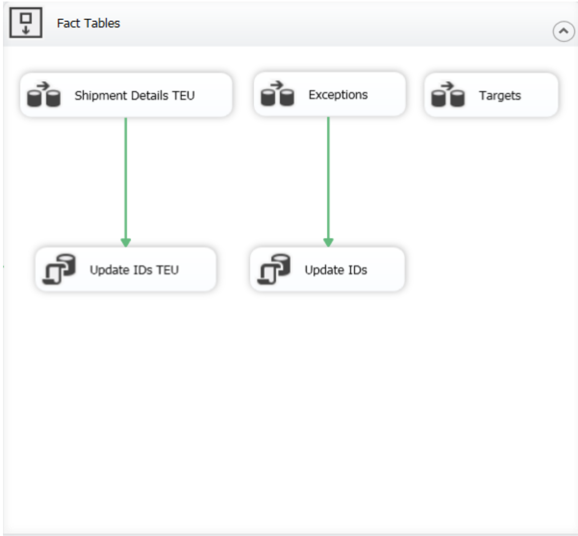

FIGURE 7: Fact phase overall

In the details of each data flow task, a copy of the data from the staging to the fact tables is made as displayed in figures 8, 9 and 10

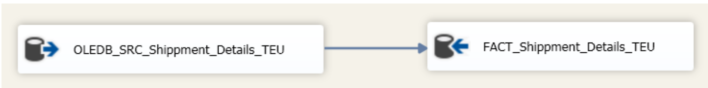

FIGURE 8: Fact phase for Shipment Details

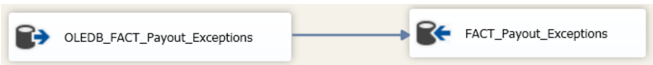

FIGURE 9: Fact details for Exceptions

FIGURE 10: Fact details for Targets

After completing these tasks, the process generates ids for all the fact tables.

.2.1 Selecting Data

For this research, the business requested to focus only on one Unit of Measure (UOM), and the selected UOM is TEU. In the SSIS project in the "Copy data to

Staging," a filter was defined to exclude all the rows where the UOM is not TEU. This action reduced the data in the tables of sales_person_details current and history, from the 1.982.969 to 180.875 rows. There was no need to drop columns.

.3 Data Structure

After defining the structure and process of all the cleaning phases. The dimensions and fact tables are created, and the data is loaded by the SSIS process, as displayed in 11.

The structure is the following:

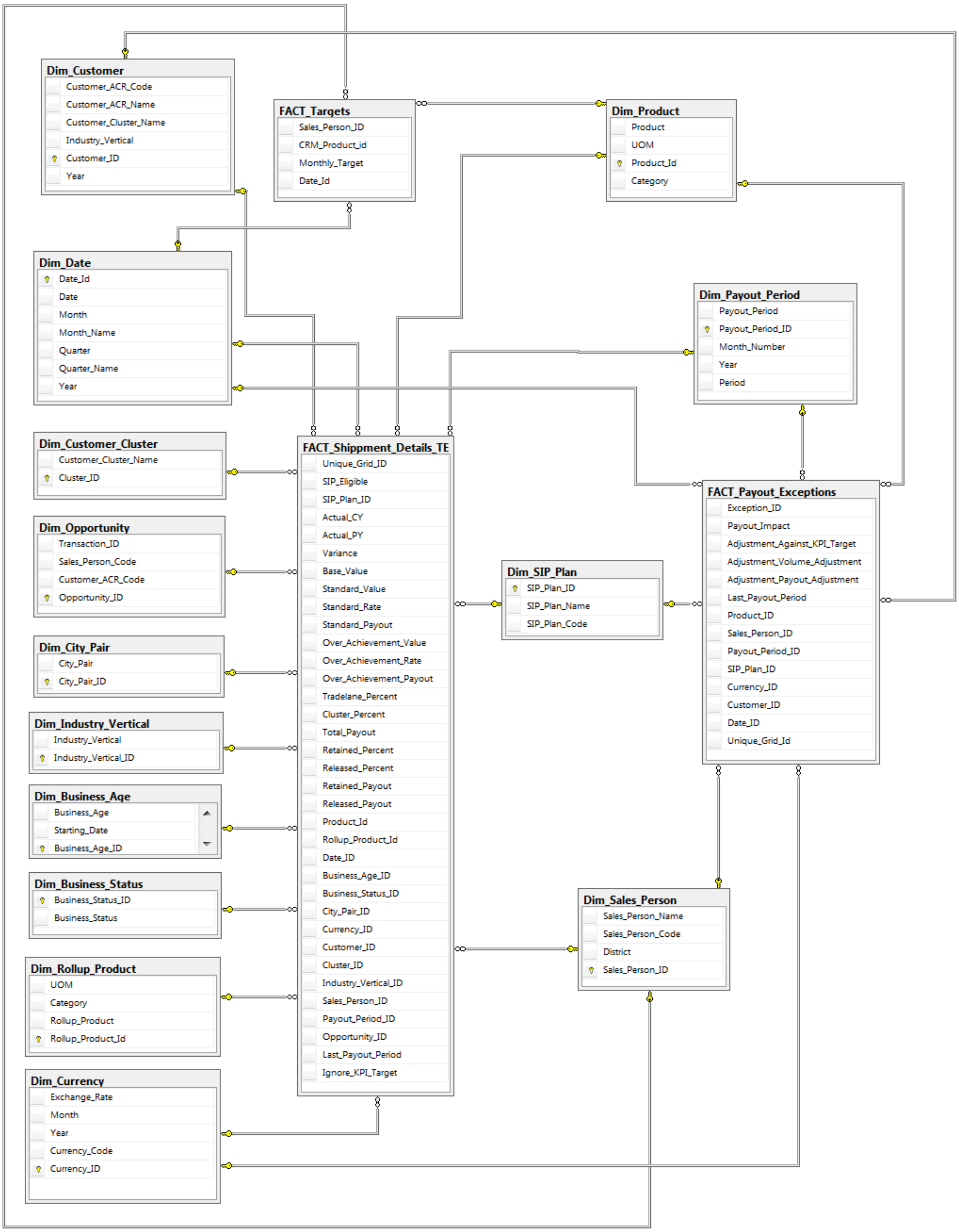

FIGURE 11: Data model

The model was then imported into Power BI for the analysis, and reports can be built as displayed in 12.

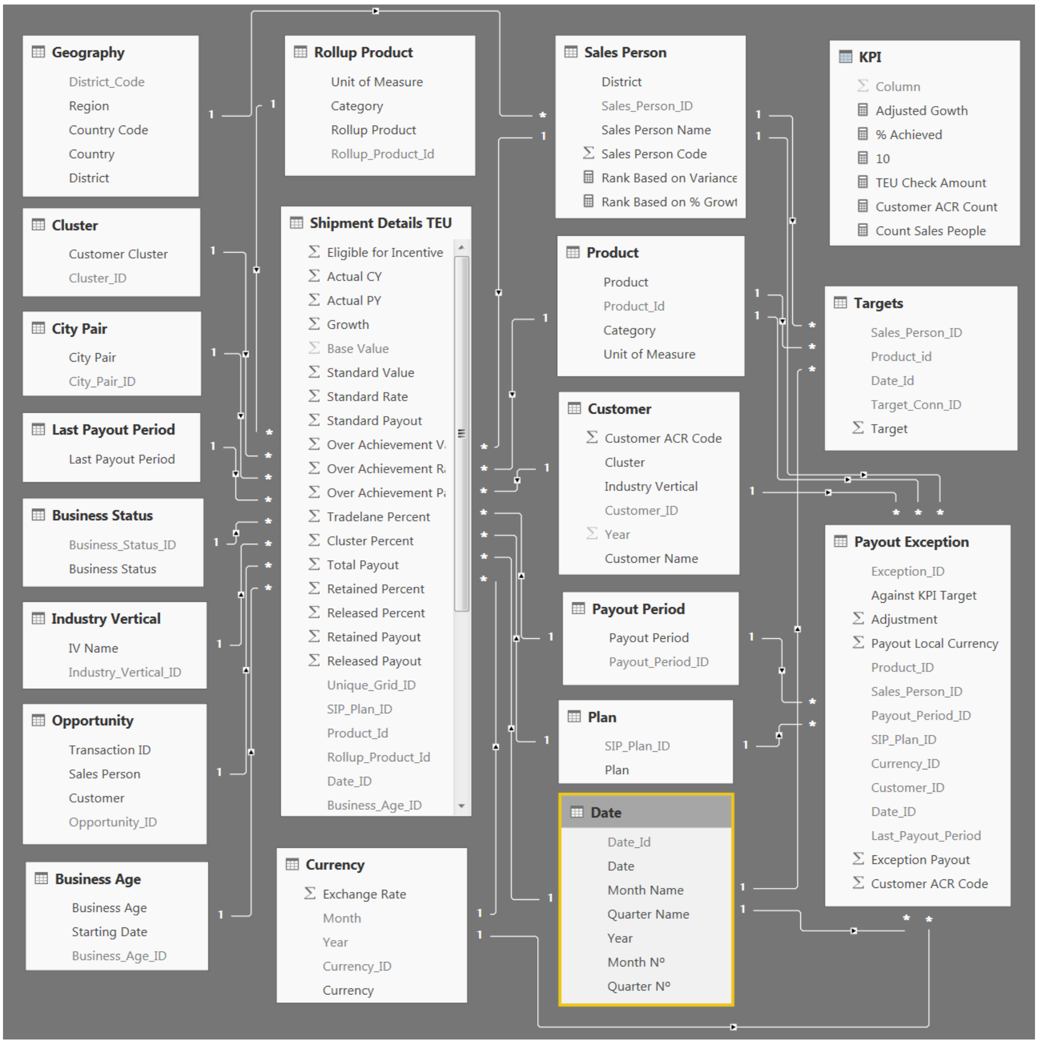

Figure 12: Power BI data model