



NOVA

IMS

Information
Management
School

MGI

Mestrado em Gestão de Informação

Master Program in Information Management

**A Client Focused Business Intelligence &
Analytics Solution for the Hospitality Sector**

Ana Rita Albuquerque Santos

Project Work presented as the partial requirement for
obtaining a Master's degree in Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

**A CLIENT FOCUSED BUSINESS INTELLIGENCE & ANALYTICS
SOLUTION FOR THE HOSPITALITY SECTOR**

Ana Rita Albuquerque Santos

Project Work presented as the partial requirement for obtaining a Master's degree in Information Management, Specialization in Knowledge Management and Business Intelligence

Co Advisor: Miguel de Castro Simões Ferreira Neto

Co Advisor: Manuel Inácio Veladas Dias

May 2020

ACKNOWLEDGEMENTS

First of all, I would like to thank Pestana Hotel Group and my NOVA IMS advisors for their collaboration in this project and for allowing me to learn even more than what I have learned during this master.

A special appreciation to my colleagues and friends of the Business Intelligence team, Edgar and André that helped me throughout all my difficulties and that have proven to be the best coworkers that I could possibly get. A special thanks to them is not enough. As well, I would like to thank the Business Intelligence team director, João, for his help and time spent on this project.

To my family, especially to my father, mother and brother that were and always will be my greatest strength. For never allowing me to give up and always showing me the path even through the darkest times, I don't know what I would do without you.

To my boyfriend, David, for his love and patience with me, for the encouragement to always be better and for the unconditional support, thank you. As well, to my best friends Catarina, Inês, Miguel and Miguel for their availability to listen to my doubts and for all the times they distracted me with their good mood, joy and their capability to show me the best side of life but especially to Catarina for helping me being on track even when everything was blurred, she was always one of my biggest supporters not only in this thesis but in everything in life.

Also, I would like to thank everyone not mentioned above that allowed the conclusion of this project, whether friends, master classmates, my teachers or my coworkers.

This is truly an accomplishment in my life. I would like to finish these acknowledgments with a phrase by Misty Copeland that perfectly describes this life phase: "Anything is possible when you have the right people there to support you".

ABSTRACT

One of the greatest needs of today's business is to know the customer or the type of customer it wants to reach, which makes a customer database a strategic weapon and one of the most important investments a company can make. The business world is becoming more competitive every day, we are constantly overwhelmed with advertisements of products we may like, product promotions we usually buy or discounts on the next purchase if we subscribe to the company's newsletter. All of this creates a client customization, and any company that is not able to do this cannot keep up with its competition. This report details the project developed at Pestana Hotel Group, which consisted of a Business Intelligence solution, more specifically the development of a customer database with the creation of two tabular models using SQL Server tools, one specific for loyal customers and another, more general, with information about all Pestana customers, and two Power BI reports that allow the visualization of the information obtained in an effective and simplified way. This report contains a literature review that situates the reader on the subject addressed in this project, a chapter dedicated to the data modeling used to create the tabular models, and another on the creation of the reports.

KEYWORDS

Business Intelligence; Business Analytics; Tabular models; Data Visualization; Power BI;

RESUMO

Uma das maiores necessidades dos negócios atuais é conhecer o seu cliente ou o tipo de cliente que quer atingir, o que torna uma base de dados de cliente uma arma estratégica e um dos mais importantes investimentos. O mundo empresarial está cada dia mais competitivo, somos constantemente assoberbados com anúncios de produtos que podemos gostar, promoções de produtos que costumamos comprar ou descontos na próxima compra caso subscrevamos a newsletter. Tudo isto cria uma personalização para o cliente, e qualquer empresa que não o consiga fazer não conseguirá acompanhar a concorrência. Este relatório detalha o projeto feito no *Pestana Hotel Group*, que consistiu numa solução de *Business Intelligence*, mais especificamente na construção de uma base de dados do cliente com a criação de dois modelos tabulares através de ferramentas do *SQL Server*, um específico para clientes fidelizados e outro mais geral com informação sobre todos os clientes Pestana, e dois relatórios em *Power BI* que permitem a visualização da informação obtida de uma forma eficaz e simplificada. O relatório contém uma revisão de literatura que situa o leitor sobre os assuntos abordados neste projeto, um capítulo dedicado à modelação dos dados de forma a criar os modelos tabulares e outro sobre a criação dos relatórios.

PALAVRAS-CHAVE

Business Intelligence; Análise de Negócios; Modelos tabulares; Visualização de dados; *Power BI*;

INDEX

| | |
|---|----|
| 1. Introduction..... | 1 |
| 1.1. Problem Definition and Objectives..... | 3 |
| 1.2. Methodology | 4 |
| 2. Literature review | 5 |
| 2.1. Business Intelligence..... | 5 |
| 2.2. Business Intelligence Architecture..... | 5 |
| 2.3. Self-Service Business Intelligence | 6 |
| 2.4. Data Warehouse | 6 |
| 2.5. Data Warehouse Architecture | 7 |
| 2.5.1. Types of Data Tables | 7 |
| 2.5.2. Schemas | 8 |
| 2.6. Extract, Transform, Load | 9 |
| 2.7. Cubes..... | 10 |
| 2.7.1. Multidimensional Structure..... | 11 |
| 2.7.2. Tabular Structure | 12 |
| 2.7.3. Views and Materialized Views | 13 |
| 2.8. Dashboards | 14 |
| 2.9. Technologies and Tools | 14 |
| 2.9.1. Technologies | 14 |
| 2.9.2. Tools | 15 |
| 2.10. Relevance of Cubes and Power BI | 16 |
| 3. Tabular Modeling..... | 17 |
| 3.1. Models..... | 17 |
| 3.2. Source Tables and Tables Description | 18 |
| 3.3. Views | 21 |
| 3.4. Data Import..... | 24 |
| 3.5. Tables Relationship | 24 |
| 3.6. Data Models Diagrams | 27 |
| 3.7. Tables Personalization..... | 29 |
| 3.8. Partitions | 29 |
| 3.9. Measures And Calculated Columns | 30 |
| 3.10. DAX Calculated Tables..... | 44 |
| 3.11. Perspectives..... | 44 |

| | |
|--|----|
| 3.12. Roles..... | 45 |
| 3.13. Deploy | 45 |
| 3.14. SSIS Processing Package | 46 |
| 3.15. Publishing In Production | 47 |
| 4. Report designing..... | 48 |
| 4.1. Data Loading Strategy | 48 |
| 4.2. Model Import to Power BI..... | 48 |
| 4.3. Requirements of the reports | 49 |
| 4.4. Report Design and Wireframing | 50 |
| 4.5. Metadata | 51 |
| 4.6. Client Report | 53 |
| 4.6.1. Report Explanation..... | 53 |
| 4.6.2. Bookmarks and Buttons | 54 |
| 4.6.3. Tooltips | 54 |
| 4.6.4. Mobile View | 56 |
| 4.7. PPG Report..... | 57 |
| 4.7.1. Report Explanation..... | 57 |
| 4.7.2. Dynamically Display of Images..... | 57 |
| 4.7.3. Bookmarks and Buttons | 58 |
| 4.7.4. Mobile View | 58 |
| 4.8. Final Considerations..... | 59 |
| 5. Conclusions..... | 60 |
| 6. Limitations and recommendations for future work | 62 |
| 7. Bibliography..... | 63 |
| 8. Annexes | 66 |
| 8.1. Annex A - Client Report | 66 |
| 8.2. Annex B – PPG Report | 71 |

LIST OF FIGURES

| | |
|---|----|
| Figure 1 - Star Schema Template. (Adapted from: (MicrosoftDocs, 2019)) | 8 |
| Figure 2 - Snowflake Schema Template. (Adapted from: (Urbanek, 2013)). | 9 |
| Figure 3 – 2020 Gartner Magic Quadrant for Analytics and Business Intelligence Platforms (Adapted from: (Microsoft, Info Microsoft, 2020)) | 16 |
| Figure 4 – Client model; | 18 |
| Figure 5 – PPG model; | 18 |
| Figure 6 – Client View representation; | 23 |
| Figure 7 – Fact Reservation View representation; | 23 |
| Figure 8 – Card View representation; | 23 |
| Figure 9 – Client PPG View representation; | 23 |
| Figure 10 – Table Import Wizard from SSAS; | 24 |
| Figure 11 – Client Data Model Diagram; | 28 |
| Figure 12 – PPG Data Model Diagram; | 29 |
| Figure 13 – Analysis Services Deployment Wizard details; | 45 |
| Figure 14 – Connections of the SSIS package; | 46 |
| Figure 15 – Package to process the model; | 46 |
| Figure 16 – Data Import in Power BI; | 48 |
| Figure 17 - Wireframe of reports; | 51 |
| Figure 18 – Client Report Metadata; | 52 |
| Figure 19 – PPG Report Metadata; | 52 |
| Figure 20 – Trip Type Tooltip; | 55 |
| Figure 21 – Repeaters Tooltip; | 55 |
| Figure 22 – Phone Layout example of each page of Client Report; | 56 |
| Figure 23 – Example of each PPG report page in mobile view; | 59 |
| Figure 24 - General page of the Client Report; | 66 |
| Figure 25 – Non PPG vs PPG page of the Client Report; | 67 |
| Figure 26 – Email Marketing page of the Client Report; | 68 |
| Figure 27 - Resorts List page of the Client Report; | 69 |
| Figure 28 – Clients List page of the Client Report; | 70 |
| Figure 29 - General page of the PPG Report; | 71 |
| Figure 30 - Tiers page of the PPG Report; | 72 |

LIST OF TABLES

| | |
|--|----|
| Table 1 – Existing Views in the Data Warehouse;..... | 22 |
| Table 2 – Views created AS-IS; | 22 |
| Table 3 – Relationships of the Client Model;..... | 25 |
| Table 4 – Relationships of PPG Model;..... | 27 |
| Table 5 – Shared calculated column between Client and PPG models of Card table;..... | 30 |
| Table 6 - Shared calculated columns between Client and PPG models of Client table; | 31 |
| Table 7 - Shared calculated column between Client and PPG models of Date table; | 31 |
| Table 8 - Shared calculated columns between Client and PPG models of Fact Event table; ... | 31 |
| Table 9 - Shared calculated columns between Client and PPG models of Fact Reservation table; | 32 |
| Table 10 - Shared measures between Client and PPG models of Card table;..... | 35 |
| Table 11 – Shared measures between Client and PPG models of Client table; | 35 |
| Table 12 - Shared measures between Client and PPG models of Fact Event table; | 35 |
| Table 13 - Shared measures between Client and PPG models of Fact Reservation table; | 36 |
| Table 14 - Unique calculated column of Client table – Client model;..... | 38 |
| Table 15 - Unique calculated column of Fact Campaign Interaction table – Client model;..... | 39 |
| Table 16 - Unique calculated column of Fact Reservation table – Client model;..... | 39 |
| Table 17 - Unique calculated column of Market table – Client model; | 39 |
| Table 18 - Unique calculated column of Meal Type table – Client model; | 40 |
| Table 19 - Unique measure of Client table – Client model;..... | 40 |
| Table 20 - Unique measure of Fact Campaign Interaction table – Client model;..... | 40 |
| Table 21 - Unique measure of Fact Food & Beverage table – Client model; | 41 |
| Table 22 - Unique measure of Fact Reservation table – Client model;..... | 41 |
| Table 23 - Unique measure of Fact Review table – Client model;..... | 41 |
| Table 24 – Unique calculated column of Card table – PPG model; | 42 |
| Table 25 - Unique calculated column of Client table – PPG model; | 43 |
| Table 26 - Unique measure of Card table – PPG model; | 43 |
| Table 27 - Unique measure of Fact Event table – PPG model; | 43 |

LIST OF ABBREVIATIONS AND ACRONYMS

| | |
|-------------|---|
| BI | Business Intelligence; |
| SSIS | SQL Server Integration Services; |
| SSAS | SQL Server Analysis Services; |
| SSRS | SQL Server Reporting Services; |
| ETL | Extract, Transform and Load; |
| DAX | Data Analysis Expressions; |
| GUID | Globally Unique Identifier; |
| PHG | Pestana Hotel Group; |
| PPG | Pestana Priority Guest – Loyalty program; |
| RN | Room Nights – Amount of nights a client stays in a hotel; |
| LOS | Length of Stay – Number of nights of a stay; |
| GSS | Guest Satisfaction Survey – Classification given in internal surveys. |

1. INTRODUCTION

With the evolution of the technological world, there is data everywhere. Some of the main challenges' companies have to face is taking advantages of data, turning it into information and insights with the main goal of being better than the competition, and that is why Business Intelligence is becoming each time more necessary.

This project will be conducted in the scope of hospitality, in particular in Pestana Hotel Group (PHG) and will have as main objective the construction of a customer database and reports using the information. The database and the reports will be constructed using Business Intelligence tools that will allow to analyze Pestana Hotel Group clients and members of the loyalty program.

Pestana Hotel Group is the biggest hotel chain in Portugal, and one of the 31 biggest hotel groups in Europe. It was established in 1972 and has been growing throughout the years. Nowadays, it has 100 hotels and pousadas located in 16 countries.

Pestana Hotel Group has a loyalty program named Pestana Priority Guest (PPG) which is divided into 6 categories: diamond, platinum, platinum corporate, gold, gold corporate and silver. This program allows the guest to earn points according to the amount of money spent at the hotels. Then, when the card reaches a certain quantity of points, the client receives advantages.

Specially in the case of PHG, due to having an extensive amount of hotels and with all the data stemming from the loyalty program there's a bigger necessity to keep track of it, in order to improve the experience of the customers and to keep being competitive, sustainable and to grow.

As such, BI tools will be used because BI turns data into functional information, permitting extended analysis of data and therefore the identification of tendencies, market behavior and organizational patterns. These analyses can be used to modify strategic plans, best assign resources and comply with budget. (Analytiks, 2019) & (Colavito, 2017) & (Lath, 2018)

Another great advantage of using business intelligence is to gain competitive advantages since it helps gain insights about the competition and market trends present in the industry. (Lath, 2018) Also, the BI systems allow to share important information across departments, as for example the use of cubes with Excel and the sharing of the Power BI reports that already happens in Pestana. (Analytiks, 2019)

Nevertheless, since the information available flows from many different sources (CRM, the reservation system, the platform used to track and view the answers given to the surveys, among others), it must be standardized with the aim of being effectively used. This standardization is done through an extraction, transformation and loading (ETL) process and it allows the creation of analysis and reports. (Desk, 2016)

Currently, Pestana Hotel Group already has consolidated data structures and several Power BI reports that help to better know the business and take insights from the data. There are 4 most-used cubes: Reservation, Reputation, Reservation Snapshots and F&B (food and beverages).

The Reservation cube contains all the data regarding the reservation (for example booked hotel, price, room type, quantity of nights, among others). The Reputation cube has the information of the opinion of the customers related to the hotels, it contains data from online reviews and the internal

surveys sent to each client post stay. The Reservation Snapshots cube contains the snapshot of each day of every hotel and 365 days ahead, which means that this cube contains both past information as well as reservations on the books (reservations to future dates). Besides, there's a food and beverages cube that contains all the information related to the consumption of food and beverages in Pestana Hotel Group outlets (restaurants and bars). Reservation, Reputation and Reservation Snapshot cubes are built with a multidimensional structure while F&B cube has a tabular structure.

However, PHG has not yet a structure dedicated fully to the customer – a customer database. A customer database can be recognized as one of the most important tools in an organization. (UK Essays, 2016) Its importance has been long discussed and several advantages are known.

Amongst them are the ability to communicate in a more efficient way by trying to reach the right audience and therefore building a brand affinity. (Stirista, 2020) Also, it can help in the mission of keeping the customers satisfied which can result in both an increase of the income and can help broaden the value of current customers. (Stirista, 2020) & (The Art of Service, 2020)

Likewise, it is much needed when analyzing different patterns of purchases of a client and to find out key tendencies. (The Art of Service, 2020) Another great advantage stemming from a client database is the capacity to enhance the marketing efficiency (in terms of campaigns) (The Art of Service, 2020) since it will allow to not waste money in sending campaigns to clients that will probably not respond. (Bonnie, 2019) Besides, a customer database can be linked with better customer service once that information like buying behavior, post-stay feedback and the answers to internal surveys allow to better know the client and attempt to meet its expectations. (Strategic Marketing, 2019) Specially in hospitality, this knowledge of the client can create a direct channel to them because knowing the customers preferences enables to characterize them. (Strategic Marketing, 2019) All this information will help to build an even more effective loyalty program which can result in repeated purchases, more memberships and therefore boost the revenue. (Bonnie, 2019)

When relating to customers, BI helps understanding the client buying tendency and having visibility of sales information allowing, thus, to tune and target the marketing campaigns but also to track these campaigns in order to assure they are profitable and efficient. (Richardson, 2018) & (Analytik, 2019) Furthermore, and since the market is becoming fiercely, it's never been more important to understand how the customers are interacting with Pestana Hotel Group and the best way to achieve them (Richardson, 2018).

Thus, there was a necessity to create a data structure especially dedicated to the customer in PHG. This database will be done by building two tabular models: one to know all the customers and the other to analyze the loyalty program members in particular.

The client model will have into account the customer behavior into 5 main perspectives: the reservations, the interaction with the campaigns, the satisfaction (measured by the responses given to internal surveys), the consumptions of food & beverages and, at last, the events associated with the loyalty card. The PPG (loyalty program – Pestana Priority Guest) model will have a more focused vision in the loyalty program, specifically the events associated with the card.

Then, these models will serve as base to construct two Power BI reports that will help to characterize Pestana clients, help the marketing and CRM teams to perform its analysis, and then to customize the campaigns sent to them by email.

This report is organized in seven chapters. The first chapter is the introduction where the objectives of this project are described, as well as a framework about Pestana Hotel Group and the intent of what the final products will be. The second chapter contains a literature review where all the main themes of this project are discussed, giving the reader a theoretical context. The third section holds all the details and steps done when building the tabular models (Client and PPG) and then, in the fourth chapter, the explanation of the Power BI built reports is done aside from other details about the construction of these reports. The fifth part contains the conclusion, furthermore, the sixth section comprises the limitations of this project along with future work that will be the next step. Lastly, in the seventh chapter a list of the books, articles and websites which served as base for the report, is presented.

1.1. PROBLEM DEFINITION AND OBJECTIVES

With market and competition evolution, it is crucial that the companies know their customers. As acknowledged, there are several problems that can come from not having a customer database, as the inability to personalize and fit marketing campaigns or emails, lack of context about clients to the sales representatives which can lead to direct sales to the wrong type of clients or being ineffective in engaging customers.

Currently, PHG already has client information however it is scattered across different systems and information sources which makes it difficult to access and often causes mismatches in information of the same customer. Furthermore, there are information collected by teams, for example the customer relationship management team, that may not be accessible to other teams, either because there is no knowledge of the platform used or even because there is no access to it by those teams.

There are different issues associated with not having a centralized view of the Pestana client: it is sometimes impossible to work on projects in collaboration with other teams because the results will depend on the customer data that the teams have access to, lower levels of security once that instead of having the information stored in a single source there are diverse sources, and the speed at which the projects are done considering that it is much slower to access several different sources that access a centralized view of the information.

Therefore, by developing this project, it will be possible to create a client database for Pestana Hotel Group that will be the centralized view, at which all the teams customer-related have access and where all the customer information is gathered and organized. Furthermore, the building of two Power BI reports will allow the teams to make analysis with the information from the database.

1.2. METHODOLOGY

This project will have as methodology a new approach to research designated Design Science Research.

The methodology of design science research consists in answering questions relevant to business problems by creating new artifacts which lead to new knowledge. The artifacts built are valuable and essential to understand the problem (Hevner & Chatterjee, 2010)

Likewise, according to (Iivary & Venable, 2009), Design Science Research creates resources to solve problems or accomplish enhancements which generates a new reality, instead of clarifying and expounding an already existing reality. This methodology can be understood as a mixture between sciences as mathematics or statistics and the industry development. Hence, knowledge of sciences is abstract while knowledge in development of solutions to industries is specific and theoretically coherent. (Horvath, 2007). Thus, Design Science Research has two principal objectives: applying the already acquired knowledge of sciences to solve problems and produce new knowledge by creating solutions that solve those problems. (Horvath, 2007)

As stated by (Pello, 2018), there are seven steps to create design solutions:

1. Perform end-user investigation: this investigation will allow to attain knowledge and find out the necessities of the users and what are the future needs.
2. Outline goals and limitations to the problem: after the end-user investigation, it will be possible to delineate the objectives of the design solution as well as the limitations that this solutions will have.
3. Collect ideas for the solution: By having brainstorming or meetings with users, it possible to collect ideas that can help in the creation of a better solution design.
4. Evaluate the ideas gathered: Sometimes the end-user has a vision that is not possible to perform due to restrictions associated with the platforms used to design the solutions, therefore, a process of evaluation needs to be done in order to select the best ideas.
5. Test: After the selection of the ideas and the building of the solution, it needs to be tested to see if it is comprehensible and easy to use.
6. After getting the first tests done, refine the solution and retest.
7. When reaching the final solution, share it with the suitable audience.

These steps illustrate that in Design Science Research, it needs to have an initial contextualization of the problem, afterwards building the solution to the problem and then test it until the end result is suited to the problem.

This project will follow this methodology considering the new reality (i.e. problem solving) that will be created is the customer database. Applying the methodology into this project, the existing problem was the lack of a customer database. Therefore, after the requirements gathered, solution building, tests made and the fine tuning of the data, the customer database is going to be shared across teams.

2. LITERATURE REVIEW

2.1. BUSINESS INTELLIGENCE

Business Intelligence does not have a single definition, the general definition was given by Van-Hau Trieu in 2016. She defined it as an umbrella term to describe processes, or concepts and methods, that improve decision making by using fact-based support systems. (Trieu, 2016)

Some years before, in 2009, BI was defined as a collection of mathematical models and methods to perform analysis which give the ability to retrieve information and knowledge from data in order to be helpful in complex decision-making processes. (Vercellis, 2009)

Furthermore, in 2004, Negash described business intelligence as a natural outgrowth of prior systems that helped decision making. There were several factors that helped build a finer environment for business intelligence, among them are the following: the development of the data warehouse as a repository, the progresses in data cleansing that lead to a single truth, the improved capacities of hardware and software and the technologies that enhanced the user interface. (Negash, 2004)

This means that business intelligence is each day more important to the enterprises since it provides the capacity to analyze past behaviors and actions, with the purpose of understanding where the company was, its present condition and predict or modify what will happen in the future. (Imhoff, Galemno, & Geiger, Mastering Data Warehouse Design. Relational and Dimensional Techniques, 2003)

As stated by Howson, BI is “a window to the dynamics of a business”. It was characterized like that since it grant people in all corporate levels the ability to accede data, interact and study it to manage, enhance performance, find opportunities and work in a more efficient way. (Howson, 2008)

BI systems offer “actionable information delivered at the right time, at the right location, and in the right form to assist decision makers.” The goal is to boost the timeliness and excellence of inputs to the decision-making process, facilitating executives and analysts work. (Negash, 2004)

The main benefit of business intelligence is that it allows to transform data into information. It gives organizations the ability to visualize the bigger picture by incorporate data from different sources and lodge it in an way that enables employees to make better decisions. (Burrows, 2015)

2.2. BUSINESS INTELLIGENCE ARCHITECTURE

A Business Intelligence system is divided in three main components: Data Sources, Data Warehouses & Data Marts and Business Intelligence Methodologies.

- **Data Sources:** Data sources are essential to collect and consolidate the data stored in the different sources (this data is heterogeneous in origin and type). The data could be stemming from operational systems, but can also come from unstructured documents, for instance emails or data from external providers. (Vercellis, 2009)
- **Data Warehouses & Data Marts:** applying ETL, “the data originating from the different sources are stored in databases intended to support business intelligence analyses.” (Vercellis, 2009)

- **Business Intelligence Methodologies:** Consists in mathematical models and analysis methodologies applied to support decisions. (Vercellis, 2009)

2.3. SELF-SERVICE BUSINESS INTELLIGENCE

The four main goals of self-service business intelligence are to make it easier to access source data for reporting and analysis, make BI tools easy to use and results easy to consume, make data warehouse solutions fast to deploy, easy to manage, customizable and simpler along with collaborative end-user interfaces. (Imhoff & White, Self-Service Business Intelligence. Empowering Users to Generate Insights, 2011)

When using SSBI tools, the users have access to diverse decision support features and functions to create their own ad-hoc BI solutions. Among these features are the creation, search, exploration, modelling, analysis and visualization of this solutions. With the lower complexity of this type of BI features it is more powerful and reachable to users since they can use SSBI technologies within their desktops or spreadsheet applications. (Poonnawat & Lehmann, 2014)

2.4. DATA WAREHOUSE

The overall accepted definition of a data warehouse was declared by Inmon: “a data warehouse is a subject-oriented, integrated, nonvolatile, and time-variant collection of data in support of management’s decisions” (Inmon, 2002). Even though this definition is the most accepted, there are several others which define a DW, among them is the one given by Howson. “A data warehouse is the collection of data extracted from various operational systems, transformed to make the data consistent, and loaded for analysis.” (Howson, 2008)

A data warehouse is a repository of the data utilized for “developing business intelligence architectures and decision support systems. The term data warehousing indicates the whole set of interrelated activities involved in designing, implementing and using a data warehouse”. (Vercellis, 2009)

Also, it is a database which stocks integrated, historical and aggregated information that is extracted from different information sources. A data warehouse is utilized in strategic decision making. (Husemann, Lechtenborger, & Vossen, 2000).

This type of data structure offers a unified vision of the company’s data, no matter how it is meant to be used afterwards by the users. It sustains the flexibility of data interpretation and delivers a stable basis of historical information which is constant, consistent and reliable. (Inmon, 2002)

Another great advantage of a DW is related to its size, that is to say that, as the information related to an enterprise is growing each day, a DW is able to grow at the same proportion. Since the beginning of the design, this structure is planned to adjust to the growth of the information in the most efficient way applying the rules defined by the enterprise. Furthermore, the data warehouse is used as support for any type of technology related to analysis within the enterprise. (Inmon, 2002)

The main goal of a DW is to give access to information which allows to manage the business. This management is possible by providing insights that could help improve revenues, reduce costs, enhance the customer service quality and align the strategic goals. (Howson, 2008)

2.5. DATA WAREHOUSE ARCHITECTURE

2.5.1. Types of Data Tables

2.5.1.1. Fact Tables

Fact tables typically refer to transactions and have links to dimension tables (these links are essential to guarantee the proper reference of the information), as well as several rows that correspond to observed facts, which means, values of the attributes that describe the corresponding transactions and represent the actual target of the subsequent OLAP analysis. This type of tables contains the descriptive attributes needed to perform decision analysis and query reporting. (Vercellis, 2009)

There are several attributes related to decision analysis, such as performance measures, operational metrics, aggregated metrics and all the others related to performance conduct. "In other words, the fact table primarily addresses what the data warehouse supports for decision analysis. Surrounding the central fact tables (and linked via foreign keys) are dimension tables.". (Sharda, Delen, & Turban, 2015)

2.5.1.2. Dimension Tables

Dimension tables usually correspond to the entities related to the processes of the organization. This means that this type of tables is associated with primary entities contained in the data warehouse. Regularly, the dimension tables are structured according to hierarchical relationships. (Vercellis, 2009)

These tables comprehend classification and aggregated information about the fact rows, besides, they contain attributes describing the data and address how data will be analyzed and summarized. In querying, and when performing an ad hoc analysis, the dimensions will be used to slice and dice the numerical values contained in the fact table to meet the requirements. (Sharda, Delen, & Turban, 2015)

Conformed Dimensions

According to Kimball, conformed dimensions exist when characteristics of different dimensions have equal column names and domain contents. (Kimball & Ross, 2013) For example, a single date dimension is needed even though it can be used with different filters (check-in, check-out, booking date, among others).

2.5.2. Schemas

2.5.2.1. Star Schema

The star schema consists on a central fact table which contains the facts of interest to an OLAP application surrounded by, and connected to, several dimension tables through referential integrity constraints based on the various dimension keys.

This schema is the current practice in data warehousing, most commonly used and the simplest style of dimensional modeling. Furthermore, the star schema is designed to allow fast query-response time, simplicity and easier maintenance.

In this non-normalized schema, the dimension tables have a one-to-many relationship with rows in the central fact table. The star schema is considered a special case of the snowflake schema. (Sharda, Delen, & Turban, 2015)

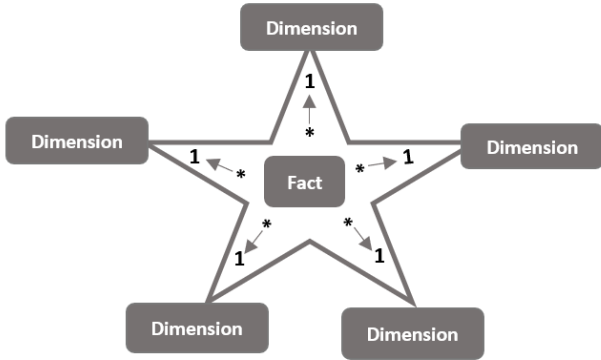


Figure 1 - Star Schema Template. (Adapted from: (MicrosoftDocs, 2019))

2.5.2.2. Snowflake Schema

The snowflake schema is similar to the star schema, however, it consists on centralized fact tables that are connected to multiple dimensions. This signifies that dimensions are normalized into multiple related tables in contrast to the denormalized dimensions in the star schema, where each dimension is represented by a single table.

The snowflake schema was named like this since the arrangement of tables in the multidimensional database, specifically the entity-relationship diagram appear the shape of a snowflake. This is due to the fact that dimensions can be composed of attribute hierarchies, that are unnormalized, their normalization results in this specific shape. (Howson, 2008)

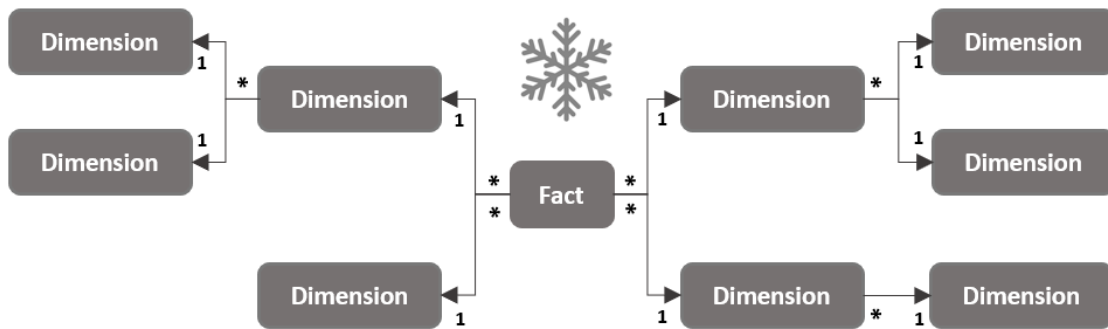


Figure 2 - Snowflake Schema Template. (Adapted from: (Urbanek, 2013)).

2.6. EXTRACT, TRANSFORM, LOAD

In an enterprise, there is a huge volume of data stemming from different applications and data sources, yet all this data has to be integrated in the same repository. While OLTP (Online Transaction Processing) is efficient for short update transactions covering a small part of the database, data warehousing requires long queries that cover a large part of the database.

The ETL process extracts the data from source systems, transforms the data according to business rules, and loads the results into the target data warehouse. However, the quality of data in the DW is directly related to the efficiency of the process, which signifies that it greatly affects the capacity of enterprises to make decisions.

The extract, transformation, and load system of the DW/BI environment consists of a work area, data structures, and a set of processes. (Kimball & Ross, 2013) & (Kakish & Kraft, 2012)

The ETL process is divided in three steps, as the name suggests. The three steps are the following:

- **Extract:** The extraction of data is done through internal and external sources, also, each system can use a different data format. The usual data source formats are flat files and relational databases, however, may also be non-relational database structures like Information Management System or alternative data structures. The extraction from these different types of sources create an additional difficulty to the process. The determination of which data is relevant is a challenging task since the data contained in the source system is often very complex. This choice is based on the design of the data warehouse that, in turn, relies on the information necessary to the business intelligence analysis and the decision support systems. The conversion of data into a single format appropriate for the next step of the ETL process is the main goal of the extraction phase, nonetheless, the design and creation of the extraction processes is a time-consuming effort. In order to keep the data updated in the data warehouse, it has to be extracted several times in a routine manner. (Kakish & Kraft, 2012) & (Vercellis, 2009)
- **Transform:** The next step in the ETL process is the transformation stage. In this stage, a set of rules and/or functions are implemented to the data extracted from the source in order to prepare the data to be loaded in the end target. The manipulation of data

depends on its source, this means that some sources will require few transformations whereas others require plenty.

The improvement of the quality of data extracted from the different sources is the main objective of the cleaning and transformation phase, this is done as a consequence of the correction of discrepancies, missing values and inaccuracies.

(Kakish & Kraft, 2012) & (Vercellis, 2009)

- **Load:** Finally, the last step in the process is the Loading phase. At this stage, already transformed is loaded into the tables of the data warehouse. The loading process depends widely on the requirements of the organization, meaning that the process may differ depending on it.

“There are some data warehouses that overwrite the existing information with cumulative information; frequently updates with extracted data are performed on hourly, daily, weekly, or monthly basis.” (Kakish & Kraft, 2012)

The restrictions delineated in the database schema and the triggers initiated when loading the data, like for example uniqueness, will contribute to the global data quality performance of the whole process. After the three steps, the main goal is to make data available to analysts and decision support applications.

(Kakish & Kraft, 2012) & (Vercellis, 2009)

The ETL process typically uses three layers to the information flow: staging, integration and access.

- **Staging Area:** This area stores raw data from each one of the sources.
- **Integration Layer:** This layer integrates the data sets transforming the data that flows from the staging area and, often, stores the data in an operational data store (ODS) database.

Then, the integrated data is moved to the data warehouse and the arrangement of the data into hierarchies (dimensions) and facts is completed.

This whole process has a great significance in data warehousing architecture because ETL processes transport the data from the source systems to the staging areas and from staging areas to data warehouses. (Kakish & Kraft, 2012)

2.7. CUBES

A cube in OLAP (Online Analytical Processing) is a multidimensional data structure which allows efficient manipulation and fast analysis of data from multiple perspectives.

By having a fact table connected to n dimension tables, it is possible to obtain a grid of cuboids, where each of them corresponds to a different level of detail on one or more dimensions. This kind of aggregation is comparable with a sum resultant from a “group by” in a SQL query.

The design of a data structure like a cube pretends to surpass the biggest limitation of relational databases, this type of database is not adapted to agile analysis of large amounts of data since it is better fit for adding, deleting and updating data.

The two main advantages the layout of data marts and data warehouses - the structures behind the cubes - offer are the fast response time when performing complex queries and the match of the dimensions with the criteria that the users apply when performing analysis.

By using OLAP, one can operate the database and choose a specific set of data by modifying the orientation of data and determine the calculations. This type of navigation through data is called slicing (via rotation) and drilling up/ down (via aggregation and disaggregation).

In essence, a cube is a portion of linked data organized in a way which allows to create two dimensional views, or slices by crossing attributes with metrics. (Sharda, Delen, & Turban, 2015) & (Vercellis, 2009)

A cube can have a multidimensional or tabular structure, these structures will be discussed in the section below.

2.7.1. Multidimensional Structure

The multidimensional structure of a cube is constituted with measures and dimensions originated from a relational database. The multidimensional model is used by the online analytical processing tool to aggregate data in large volume with the objective of allowing fast query response times.

There are three main OLAP architectures:

- **Multidimensional OLAP (MOLAP):** In this architecture, data is stored in a multidimensional database. Data from this type of database is stocked in a smaller space than the one used in a relational database. Data is kept in array data structures to provide better performance. Besides, this database has a rich and complex set of functions to analyze the data.
- **Relational OLAP (ROLAP):** It is considered a recreation of OLAP in a relational database that has the benefit of not narrowing the volume of data storage. As the user elaborates the query, the tool accesses the data instead of having already pre calculated aggregations. It allows users to analyze any data without restricting the extension of data.
- **Hybrid OLAP (HOLAP):** This architecture combines the capacities of the ones mentioned above. It has high performance and high scalability however possesses a complex architecture.

The key characteristics of multidimensional modeling are:

- **Rich Data Model:** The modeling of the measures and dimensions can be both simple or complex. Complexity can be originated from many-to-many relationships, for example.
- **Sophisticated Analysis:** In this multidimensional architecture, it is possible to use MDX (this query language is explained ahead). By the use of it, a series of complex calculations can be done.

Even though this type of data modeling has many advantages, it is more complex to develop and it is less able to quickly adapt to the alterations of business circumstances. Besides, it requires advanced modeling and MDX skills. (Vitt & Cameron, 2012) & (Prabhuswamy, 2015)

Query performance: Query performance impacts the quality of the experience the user has because of the amount of time it takes to process queries and present results. Aggregations, caching and indexed data retrieval are some of the forms used to speed up query performance.

An aggregation consists on a precalculated summary of data. Each time a query is performed, the system tries to recover the result from cached memory, if no data is in cache, it retrieves data from an aggregation, and if there's no aggregation present it recovers data from a measure group's partitions. Caching helps to improve the performance of a query since, throughout querying, memory is used to store results in the storage and query processor caches. Another way to improve query performance is optimize the design of dimensions, cubes and MDX queries. (Vitt & Cameron, 2012)

Processing Performance: Processing is the refresh of data in Analysis Services database, the quicker the processing of data, the sooner the users have access to up-to-date data. By applying query and processing optimization techniques, the database can grow and still be efficient. (Vitt & Cameron, 2012)

2.7.1.1. Pros and Cons of Multidimensional Structure

The biggest disadvantages found on the multidimensional structure are its increased complexity compared to the tabular structure, the fact that it is a consolidated product therefore no major innovations are expected, and finally, it cannot be used with Power View.

On the other hand, it is a mature technology able to deal with advanced modeling and computations requirements. Additionally, it is scalable and capable of dealing with large quantities of data. (Element 61, 2012)

2.7.2. Tabular Structure

The tabular modeling structures the data into related tables. The data can be loaded in two ways: xVelocity and Direct Query. xVelocity in-memory engine stores tabular data in memory to allow fast query response, on the other hand, Direct Query queries directly the database.

The key characteristics of tabular modeling are the following:

- **Familiarity:** Since the handling of data in tabular model is similar to many stored in table relational databases like Excel or Access, it is familiar to many users. Also, it is easier to manage and apply changes since the DAX language (explained in a section below) is considered an extension of the Excel formula language.
- **Flexibility:** This modeling can accelerate the development of the model because data preparation and design rigor are not as strict as in the multidimensional model. This is due to the fact that data is not organized into measures and dimensions. The tabular

model is also more prompt to data modifications or changes in relations between tables and calculations complaint with business needs modifications.

Nonetheless of the advantages this model presents, it may not be adequate for sophisticated business needs or very complex datasets. Furthermore, if advanced calculations need to be created, multidimensional capabilities are probably more appropriate. (Vitt & Cameron, 2012)

Querying Performance: When querying a database with the tabular structure, the in-memory engine browses the memory to collect the data and to compute the aggregations. This leads to high performance rates without the need of special tuning. Even if the memory is restricted, this tool affords basic support in agreement with physical memory. (Vitt & Cameron, 2012)

Processing Performance: Processing in tabular model is more flexible, mainly because this type of model does not classify data into dimensions and measure groups, besides, it loads data directly into memory without the need of writing into disk. Also, the processing time decreases. Even though it is possible to process various tables in parallel, it is not possible to process individual tables' partitions at the same time. (Vitt & Cameron, 2012)

Direct Query: Direct Query allows to avoid data processing by writing DAX queries and calculations to the source database. This feature is particularly advantageous when dealing with frequent refreshing large data volumes. Regardless, there are some limitations since calculated columns and part of DAX functions are not allowed. (Vitt & Cameron, 2012)

2.7.2.1. Pros and Cons of Tabular Structure

Tabular structure is easier to develop than multidimensional model, it has integration with Power Pivot, is more efficient than OLAP in most cases and it is considered a technology for the future. However, there are considerable enhancements to do, it misses advanced computations available with MDX and it is limited to the RAM memory accessible. (Element 61, 2012)

2.7.3. Views and Materialized Views

A view creates a virtual table by logically combining tables and fields when executing a query obeying to rational rules. It defines a function from a combined group of tables to a single derived table, usually, this function is recomputed every time a reference of the view is performed.

A specific type of view will also be used in this project, this type of view is named materialized view. A materialized view is a physical structure that enables better performance since database accesses to the materialized views are much quicker than recomputing an entire view. Thus, a materialized view is a copy of data that can be accessed promptly. The materialization of the view stores the tuples of it in the database, it can be defined on multiple tables and can have selections and group by clauses on different columns. (Agrawal, Chaudhuri, & Narasayya, 2000) & (Gupta & Mumick, 2011)

2.8. DASHBOARDS

A dashboard is a response to the growing complexity of data. It is a visual representation of crucial data which aids executives monitor important points and analyze them in short time. The dashboards have key performance indicators (KPI) with fundamental drivers to broadcast performance across the organization.

The necessity to create dashboards arise from some factors like the demand for integration of metrics related to all departments to analyze performance and resource allocation, the influence of managers in decision making, the goal of enterprises to grow while keeping the costs low and mainly the current poor use of data that is potentially very helpful to make important decisions.

The characteristics a dashboard must possess are the following: having consistent measures and measurement procedures along all departments, be a support in performance monitoring (evaluative and developmental), ability to be used to plan goals and strategies to the future; and the ability to communicate its performance to the users. Also, what an organization values as performance by the choice of metrics on the dashboard. (Sharda, Delen, & Turban, 2015) & (Pauwels, 2009)

Malik states that a dashboard should be SMART by including these elements: **S** Synergetic – be visually efficient for a user to understand information about different features; **M** Monitor KPIs - should present critical KPIs, these indicators will help decision making; **A** Accurate – the information presented in the dashboard must be precise; **R** Responsive - should answer predefined questions; **T** Timely – information must be current in order to improve decision making. (Malik, 2005)

Furthermore, and besides the elements mentioned before, there are some advanced elements that should be a part of a dashboard that Malik wrapped in the acronym IMPACT: **I** Interactive – should let users drill down to get detailed view; **M** More data history – a dashboard should have history in order to be possible to analyze trends; **P** Personalized – the dashboard should be particular to the users area of responsibility; **A** Analytical – should simplify understanding of the business; **C** Collaborative – the notes regarding characteristics of the dashboard should be shared with the analysts; **T** Trackability – the metrics each department is interested in tracking should be customizable. (Malik, 2005)

2.9. TECHNOLOGIES AND TOOLS

2.9.1. Technologies

- **Structured Query Language**

SQL is a computational language for relational database management and data manipulation used to query, insert, update and modify data. Most relational databases support SQL. (Techopedia, Techopedia, s.d.)

- **Multidimensional Expressions**

MDX is a query language which deals with multidimensional data. MDX is used when traditional SQL no longer serves as a sufficient extractor. (Techopedia, Techopedia, s.d.)

- **Data Analysis Expressions**

DAX are a collection of functions utilized to perform a task and return one or more values. DAX is a formula or query language, not a computer language. It is extremely popular since it is the base for Power BI and it is also used in Tabular SSAS. (Vijalapuram, 2018)

2.9.2. Tools

- **Microsoft SQL Server**

This tool is a relational database management system which stores and retrieves data solicited by other applications.

It provides support to business intelligence and analytics operations and also to a diversity of transaction processes. As the name suggests, the language used in this tool is SQL. (Rouse, 2019) & (Computer Business Review, s.d.)

SQL Server has four main elements:

- **Database Engine:** This element performs as the traditional relational database. (Withee, 2010)
- **Reporting Services:** SSRS is a Business Intelligence component that allows to create reports through the transformed data and handle the access given to the reports. (Withee, 2010)
- **Integration Services:** SSIS manages the ETL process, which means that it aids in the process of transforming data into a standard format to be used by all elements in the organization. (Withee, 2010)
- **Analysis Services:** SSAS incorporates an OLAP tool to analyze the data online as well as a data-mining engine that helps to find the data requested by the queries. (Withee, 2010)

- **Visual Studio**

Visual Studio is used for several Business Intelligence processes such as building cubes, SSRS reports or ETL processes. It is the main application to develop this type of task. (Withee, 2010)

- **Power BI**

Power BI is a business intelligence tool and a business analytics solution. It connects to several sources of data, transforms it and allows the users to create interactive reports and dashboards which give insights about the company. Power BI is rooted on several known programs such as SSAS, Power Pivot, Power View and Power Query. That is the reason why, even though Power BI is fairly new, it is familiar to the users. (Vijalapuram, 2018) & (Microsoft, 2019)

Microsoft Power BI has been recognized as a leader in Gartner Magic Quadrant for analytics and business intelligence platforms for thirteen consecutive years as can be seen in the figure below. (Microsoft, Info Microsoft, 2020)



Figure 3 – 2020 Gartner Magic Quadrant for Analytics and Business Intelligence Platforms (Adapted from: (Microsoft, Info Microsoft, 2020))

2.10. RELEVANCE OF CUBES AND POWER BI

The major advantages of having cubes as data repositories is that they can assure faster response times despite complex queries and dimensions matching requirements of the analysts when performing their analysis. (Vercellis, 2009)

Moreover, the tool used to create dashboards – Power BI – also has many advantages: it is powerful and allows interactive guided analysis, it has monthly releases with several information available online, it is simple to coordinate with Excel, gives a good report perception, holds extensive database availability since it can merge information from different sources and also has a mobile app which permits users to see their reports and dashboards through their phones. (DataFlairTeam, 2019)

3. TABULAR MODELING

Before starting modeling, an extensive data quality project was performed to permit the joining of customers which are the same person with the creation of a master id. In the data quality project, several rules were followed to lump the same customer. The first one was a combination of email, first and last names, the second was a combination of the passport number with first and last names and the last rule was the grouping of telephone number, first and last names. All these combinations had into account the punctuation and the capital and lower-case letters.

To emphasize that the information stemming from the CRM system was considered more reliable than the one coming from the registrations done at the reservation since the information of the CRM is filled by the client directly on Pestana site. In case of records from the same source matching the rules above, the one considered master is the one more complete. However, if the criteria described before is not enough to differentiate between two records, the most recent one is the one which is kept.

After the creation of this master ids (each corresponding to a client), every record which is added to the table is compared to the masters and if the match succeeded it is considered the same person, otherwise a new master id is created.

This data quality project was an essential step before the creation of the models which will be detailed later in this chapter since it allowed to aggregate information corresponding to the same client providing a better data quality than the one previously existing.

After this stage, the project dwell in the construction of two tabular models: Client and PPG. The first model will have all the information regarding a Pestana Hotel Group client, including reservations, surveys, food and beverage consumptions, campaign interactions and events associated with the loyalty card. The PPG model will have information regarding Pestana Hotel Group loyalty program named Pestana Priority Guest and will incorporate information about the events associated with the card and the specific clients that are a part of the program.

The choice of having two models instead of one was due to the direction of the relationships since having a two-way relationship would mean that the Client could filter both the Card and Fact Reservation which leads to a circular reference. Also, having two models instead of one will allow to comply with the requirements of some analysis requested: relate reservations with points and analysis of the points that are not associated with reservations (expired points or points associated to a client, for example). Besides, a Pestana priority guest is necessarily a client however, to be a client it is not necessary to be a part of the loyalty program.

3.1. MODELS

The **Client model** will have the subsequent structure:

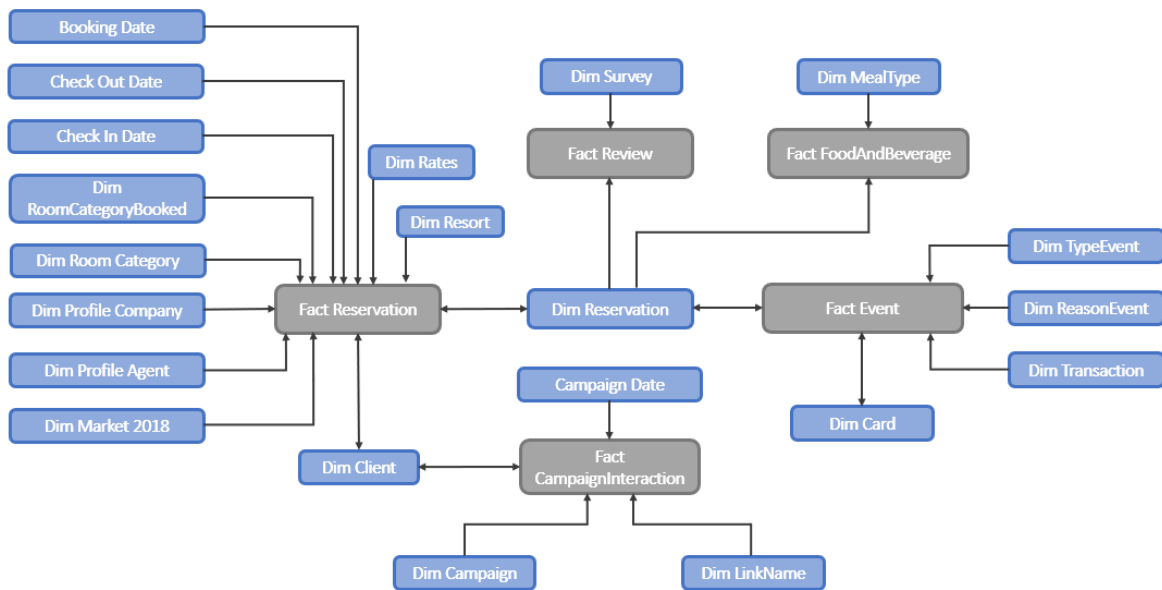


Figure 4 – Client model;

The **PPG model** will have the following scheme:

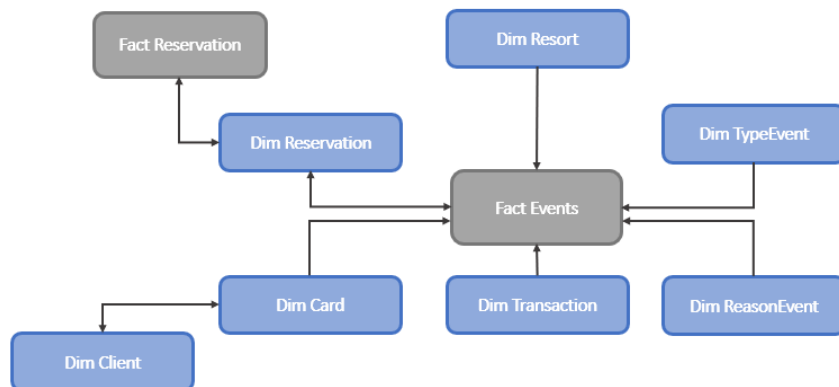


Figure 5 – PPG model;

3.2. SOURCE TABLES AND TABLES DESCRIPTION

The **Client model** will have 5 factual tables and 21 dimension tables (including 4 date tables: check out date, check in date, booking date and campaign date). The first three date tables mentioned are all connected to the Fact Reservation once the main objective is that this table can distribute date information through all the model. The other date table concerns campaigns.

The factual tables are:

- **Fact Reservation**

Contains the details of the reservation such as number of week and weekend days of the reservation, number of adults and children, quantity of room nights, number of points earned with the reservation, booking window, length of stay, average room rate, among others measures and details that are used to classify a reservation.

- **Fact Review**

This table contains the details stemming from the responses given to the internal surveys each client receives after a stay.

It contains the score given by the guest overall and in the departments related to restaurants, bars and spa. This factual table only contains the details about the internal surveys and not about the online reviews due to the impossibility to link an online review with a reservation of a client.

- **Fact Campaign Interaction**

This table aims at analyzing the campaigns created by Pestana site and email marketing teams. It contains the campaign date, the amount of sent emails, the number of these which were open and the number of clicks in the link sent within the emails.

- **Fact Event**

This table contains the details related to the card of the loyalty program (Pestana priority guest). There are several card tiers and different amounts of points that can be earned. The details of this table include the amount of total points, room, food and other (related to other hotel services as for example spa).

- **Fact Food Beverage**

This table comprises the facts of the expenses related to food and beverages, in detail, the net and gross revenue.

The dimension tables are:

- **Check Out Date**

Date table containing the details concerning the check out date.

- **Check In Date**

Date table containing the details concerning the check in date.

- **Booking Date**

Date table containing the details concerning the booking date.

- **Room Category Booked**

Table containing the specifics of the type of room the guest booked at the time of the reservation.

- **Room Category**

Table comprising the specificities of the type of room the guest actually stayed in (there can be overbooking or upgrade situations, for example).

- **Profile Company**

Table with the information about the company which made the reservation. This table is only filled when the client reserved through a company, for example in a business trip.

- **Profile Travel Agent**

Table with the information about the travel agent which made the reservation. Contains information about travel agents both online (for example, booking) and offline (for example, viagens Abreu).

- **Market**

Contains the details about the market that the client belongs to (direct, contracted, complimentary, etc.).

- **Rate**

The rate at which the client reserved the night (for example, BAR – best available rate).

- **Client**

Personal information about the client such as name, email, address, phone number or birth date.

- **Campaign**

Details about each campaign.

- **Link Name**

Name of the campaign.

- **Reservation**

Specific details about the reservation: reservation number, confirmation number and CRS number.

- **Survey**

Contains the identification of the survey because there are 4, one for each Pestana brand.

- **Resort**

Details about each resort: location, name, brand among others.

- **Meal Type**

Specific aspect of the meal (breakfast, lunch, dinner) and if it is a part of a package or not.

- **Card**

Details about the card. Card number, date of the beginning the membership in Pestana priority guest program, tier of the card, whether the card is active, its expiration date and if it is the card last used (in case there is more than one card associated to a guest).

- **Campaign Date**

Date table containing the details concerning the campaign date.

- **Transaction**

Table containing the transaction details related to the event.

- **Reason Event**

Table covering the reason of the event, if it is a consumption of points, redemption, purchase, expiration or cancellations.

- **Type Event**

Table that contains the type of event, if it is a credit or debit.

All these tables will constitute a model that will allow to analyze the client with detail.

The **PPG model** will have 2 factual tables and 7 dimension tables as well as a date table. The model will be named PPG as it is focused on the loyalty program and the events associated with it.

The Fact Event table is similar to the Events table used in the client model, however, since this model has as goal the analysis of the customers associated with the loyalty program, the table was filtered to contain only clients with a CRM number associated. The other factual table present in this model is the Fact Reservation. All the tables in this model are also present in the Client model though the connections are different.

The dimensions present in this second model are: Client, Card, Transaction, Reason Event, Type Event, Resort, Date or Reservation (a brief description of each table was done above).

In this model, the card and the client dimensions are connected to each other since that if the client is a part of the PPG program then it has a card. A client can have more than one card, but each card is associated with only one person.

The connection between these two tables is bidirectional as an analyst could want to analyze through the client point of view or through the card.

3.3. VIEWS

As stated by Russo, M. and Ferrari, A., the tabular model built should be established with views of the database tables instead of the tables itself to halt the dependency between the data model and the report.

There are several advantages in importing the views rather than the tables, among them are the following: possibility to add details to the view creation query that will improve efficiency (as NOLOCK), ability to change the column names into friendly or user-known names, capacity to include in the view business logic needed ahead or data transformations and possibility to remove columns that will be useless in the model. (Russo & Ferrari, 2017)

Since some of the tables which are going to be a part of this model are already a part of other existing models and cubes, some views were already created. Thus, the already existing views comprise:

Table 1 – Existing Views in the Data Warehouse;

| Dimensions | View Name |
|--------------------------|-------------------------|
| Dim Market | VwDimMarket |
| Dim Meal Type | VwDimMealType |
| Dim Profile Agent | VwDimProfileTravelAgent |
| Dim Profile Company | VwDimProfileCompany |
| Dim Rates | VwDimRate |
| Dim Reason Event | VwDimReasonEvent |
| Dim Reservation | VWDimReservation |
| Dim Resort | VwDimResort |
| Dim Room Category | VwDimRoomCategory |
| Dim Room Category Booked | VwDimRoomCategory |
| Dim Survey | VwDimSurvey |
| Dim Transaction | VwDimTransaction |
| Dim Type Event | VwDimTypeEvent |

The remaining views were created AS-IS based on the tables already present in the data warehouse with a exception of Dimension Client and Fact Reservation in the **Client Model**.

The views created AS-IS are the following:

Table 2 – Views created AS-IS;

| Dimensions / Facts | View Name |
|---------------------------|---------------------------------|
| Fact Event | VwFactEventsClient |
| Fact Campaign Interaction | VwFactCampaignInteractionClient |
| Fact Food And Beverage | VwFactFoodAndBeverageClient |
| Fact Review | VwFactReviewClient |
| Dim Campaign | VwDimCampaignClient |
| Dim Card | VwDimCardClient |
| Dim Link Name | VwDimLinkNameClient |

The Client View was created by filtering the master registrations which were defined in the data quality process as well as filtering the dates in an invalid format. The data warehouse and some column names were hidden to protect the privacy.

```
USE [ DW]
GO

SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

CREATE VIEW [Analysis].[VwDimClientClient]
AS
SELECT [GUID]
...
FROM [Client].[DimClient] WITH(NOLOCK)
WHERE IsMestre = 1 AND ISDATE(CONVERT(varchar, [BirthDate]))=1
GO
```

Figure 6 – Client View representation;

The Fact Reservation View was created by joining the fact reservation table with the client dimension table through IdProfileClient, the client dimension was filtered by Code Source, filtering only the ones coming from Opera. This was done to extract the GUID code which will allow the join of the two tables in the tabular model.

```
USE [ DW]
GO

SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

CREATE VIEW [Analysis].[VwFactReservationClient]
AS
SELECT [IdReservation]
...
FROM [Client].[FactReservation] AS R WITH(NOLOCK)
INNER JOIN (
    SELECT IdClientSourceDWH, [GUID]
    FROM [Client].[DimClient] WITH(NOLOCK)
    WHERE CodSource = 'OPE') AS CL
ON R.IdProfileClient = CL.IdClientSourceDWH
WHERE [CodCheckInDate]> 20100101
GO
```

Figure 7 – Fact Reservation View representation;

In the **PPG Model**, there were two views that were not created AS-IS: Card and Client.

The Card view was created by joining the Card table with the Client table by CodCustomer, allowing to attain the list of clients who are the same person and extracting the GUID which will permit to establish the relationship in the model.

```
USE [ DW]
GO

SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

CREATE VIEW [Analysis].[VwDimCardPPGClient]
AS
SELECT [IdCard]
...
FROM [Client].[DimCard] AS CRD WITH (NOLOCK)
INNER JOIN [Client].[DimClient] AS CL WITH(NOLOCK)
ON CRD.CodCustomer = CL.CodCustomer
WHERE CL.CodCustomer IS NOT NULL
AND LTRIM(RTRIM(CL.CodCustomer)) != ''
GO
```

Figure 8 – Card View representation;

The Client view was created by adding a filter to the ones previously applied in the client view in the client model (filtering master ids and valid birth dates). The filter applied allowed to differentiate the clients which have a CRM number, meaning they are a part of the loyalty program. Thus, the Client view in the PPG model will only have loyal clients.

```
USE [ DW]
GO

SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

CREATE VIEW [Analysis].[VwDimClientPPGClient]
AS
SELECT [GUID]
...
FROM [Client].[DimClient] WITH(NOLOCK)
WHERE IsMestre = 1 AND ISDATE(CONVERT(varchar, [BirthDate]))=1
AND [CodCustomer] <> ''
GO
```

Figure 9 – Client PPG View representation;

3.4. DATA IMPORT

After the creation of the views, the data import to the model was performed. A connection to the group data warehouse was defined, using impersonation. Impersonation is used to assume the identity of a client application (Microsoft, Microsoft Docs, 2020), in this case, since the processing of the models will be a scheduled task the best practice is to use a server account. Besides, this prevents possible problems if the database creator leaves the group.

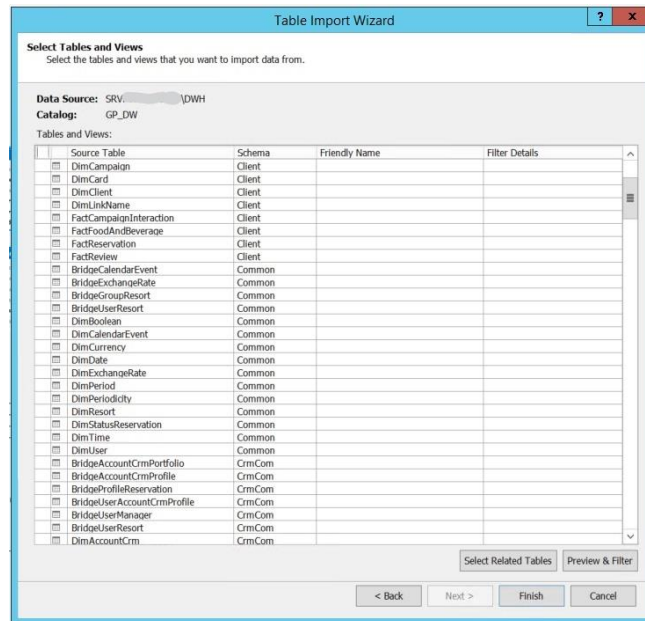


Figure 10 – Table Import Wizard from SSAS;

The views imported were the ones referred above, corresponding to each table of the **Client Model**: Fact Reservation, Fact Review, Fact Campaign Interaction, Fact Event, Fact Food Beverage, Check Out Date, Check In Date, Booking Date, Room Category Booked, Room Category, Profile Company, Profile Travel Agent, Market, Rate, Client, Campaign, Link Name, Reservation, Survey, Resort, Meal Type, Card, Campaign Date, Transaction, Reason Event and Type Event.

As stated in literature, there are conformed dimensions that have the same content. This is the case of the date tables of the model (Campaign Date, Check In Date, Check Out Date and Booking Date) and the room category tables (Room Category and Room Category Booked). These tables have the same amount of details about the dates and the room category and they are needed because there is different type of dates regarding hospitality and it is important to know if the room booked by the customer was indeed the one he stayed in.

Regarding the **PPG Model**, the process was similar to the one described above, the tables imported are: Card, Client, Date, Reason Event, Reservation, Resort, Transaction, Type Event, Fact Reservation and Fact Event. There were no conformed dimensions in this model.

3.5. TABLES RELATIONSHIP

After importing the tables to the **Client Model**, the next step was to create the relationships between the different tables. The usual cardinality of a relationship is “many-to-one” or “one-to-many”.

Regarding filter direction it can be to any of the two tables or, in the tabular model, to both tables which means that both tables can filter each other.

There are four relationships in this client model that have a bidirectional filter associated:

- **Fact Event [Id Card] to Card [Id Card]**

The user can decide to filter all the events associated with a card, in this case, the filter will be applied in the Card table and the Fact Event will have only the events associated with this specific card. However, the analyst may want to analyze all the cards that have a consumption as an event, in this instance, the events will be filtered in the Fact Event and then proceed to filter only the cards with this specific event.

- **Fact Event [Id Reservation] to Reservation [Id Reservation]**

The analyst may see which event or events are associated with a specific reservation or can decide to filter and see details of the reservations associated with a type or reason of events.

- **Fact Reservation [GUID] to Client [GUID]**

The analysis could be to perceive the client who made a reservation (the Fact Reservation filters the Client) or to see all the reservations associated with a client (the Client filters the Fact Reservation).

- **Fact Reservation [Id Reservation] to Reservation [Id Reservation]**

The reservation table works as a conductor since it allows to distribute the reservation filter through three factual tables (Review, Event and Food Beverage).

All the relationships of the model are listed below.

Table 3 – Relationships of the Client Model;

| Table | Cardinality | Filter direction | Table |
|--------------------------------------|-------------------|-------------------------------|-----------------------------|
| FactCampaignInteraction [EventDate] | Many to One (*:1) | << To FactCampaignInteraction | CampaignDate [CodDate] |
| FactCampaignInteraction [IdClient] | Many to One (*:1) | << To FactCampaignInteraction | Client [IdClient] |
| FactCampaignInteraction [IdLinkName] | Many to One (*:1) | << To FactCampaignInteraction | LinkName [IdLinkName] |
| FactCampaignInteraction [JobID] | Many to One (*:1) | << To FactCampaignInteraction | Campaign [JobID] |
| FactEvent [IdCard] | Many to One (*:1) | << To Both Tables >> | Card [IdCard] |
| FactEvent [IdReasonEvent] | Many to One (*:1) | << To FactEvent | ReasonEvent [IdReasonEvent] |
| FactEvent [IdReservation] | Many to One (*:1) | << To Both Tables >> | Reservation [IdReservation] |
| FactEvent [IdTransaction] | Many to One (*:1) | << To FactEvent | Transaction [IdTransaction] |
| FactEvent [IdTypeEvent] | Many to One (*:1) | << To FactEvent | TypeEvent [IdTypeEvent] |

| | | | |
|--|-------------------|------------------------|---|
| FactFoodBeverage [IdMealType] | Many to One (*:1) | << To FactFoodBeverage | MealType [IdMealType] |
| FactFoodBeverage [IdReservation] | Many to One (*:1) | << To FactFoodBeverage | Reservation [IdReservation] |
| FactReservation [CodBookingDate] | Many to One (*:1) | << To FactReservation | BookingDate [CodDate] |
| FactReservation [CodCheckInDate] | Many to One (*:1) | << To FactReservation | CheckInDate [CodDate] |
| FactReservation [CodCheckOutDate] | Many to One (*:1) | << To FactReservation | CheckOutDate [CodDate] |
| FactReservation [GUID] | Many to One (*:1) | << To Both Tables >> | Client [GUID] |
| FactReservation [IdMarket2018] | Many to One (*:1) | << To FactReservation | Market [IdMarket2018] |
| FactReservation [IdProfileAgent] | Many to One (*:1) | << To FactReservation | ProfileTravelAgent [IdProfileAgent] |
| FactReservation [IdProfileCompany] | Many to One (*:1) | << To FactReservation | ProfileCompany [IdProfileCompany] |
| FactReservation [IdRate] | Many to One (*:1) | << To FactReservation | Rate [IdRate] |
| FactReservation [IdReservation] | Many to One (*:1) | << To Both Tables >> | Reservation [IdReservation] |
| FactReservation [IdResort] | Many to One (*:1) | << To FactReservation | Resort [IdResort] |
| FactReservation [IdRoomCategory] | Many to One (*:1) | << To FactReservation | RoomCategory [IdRoomCategory] |
| FactReservation [IdRoomCategoryBooked] | Many to One (*:1) | << To FactReservation | RoomCategoryBooked [IdRoomCategoryBooked] |
| FactReview [IdSurvey] | Many to One (*:1) | << To FactReview | Survey [IdSurvey] |
| FactReview [IdReservation] | Many to One (*:1) | << To FactReview | Reservation [IdReservation] |

Regarding the **PPG Model**, there are also three bidirectional relationships:

- **Client [GUID] to Card [GUID]**

This two-way relationship is justified by two sorts of analysis. An analysis targeted to the client, and in this example, the client will filter the card and consequently the events associated with it or, if the goal is to find the events that a specific card has, the card table will filter the client associated with it and it will be possible to analyze the behavior of the guest.

- **Reservation [Id Reservation] to Fact Event [Id Reservation]**

The justification to this bidirectional relationship is the same to the one gave above regarding the client model.

- **Fact Reservation [Id Reservation] to Reservation [Id Reservation]**

The reservation dimension where works as a filter distributor between the Fact Reservation and the Fact Event.

The relationships of the model are:

Table 4 – Relationships of PPG Model;

| Table | Cardinality | Filter direction | Table |
|---------------------------------|-------------------|----------------------|-----------------------------|
| Card [GUID] | Many to One (*:1) | << To Both Tables >> | Client [GUID] |
| FactEvent [CodBusinessDate] | Many to One (*:1) | << To FactEvent | Date [CodDate] |
| FactEvent [IdCard] | Many to One (*:1) | << To FactEvent | Card [IdCard] |
| FactEvent [IdReasonEvent] | Many to One (*:1) | << To FactEvent | ReasonEvent [IdReasonEvent] |
| FactEvent [IdReservation] | Many to One (*:1) | << To Both Tables >> | Reservation [IdReservation] |
| FactEvent [IdResort] | Many to One (*:1) | << To FactEvent | Resort [IdResort] |
| FactEvent [IdTransaction] | Many to One (*:1) | << To FactEvent | Transaction [IdTransaction] |
| FactEvent [IdTypeEvent] | Many to One (*:1) | << To FactEvent | TypeEvent [IdTypeEvent] |
| FactReservation [IdReservation] | One to One (1:1) | << To Both Tables >> | Reservation [IdReservation] |

3.6. DATA MODELS DIAGRAMS

The diagram of the **Client Model** is shown in the next page, so it can be seen with more detail.



Figure 11 – Client Data Model Diagram;

The diagram of the **PPG Model** is the presented below:

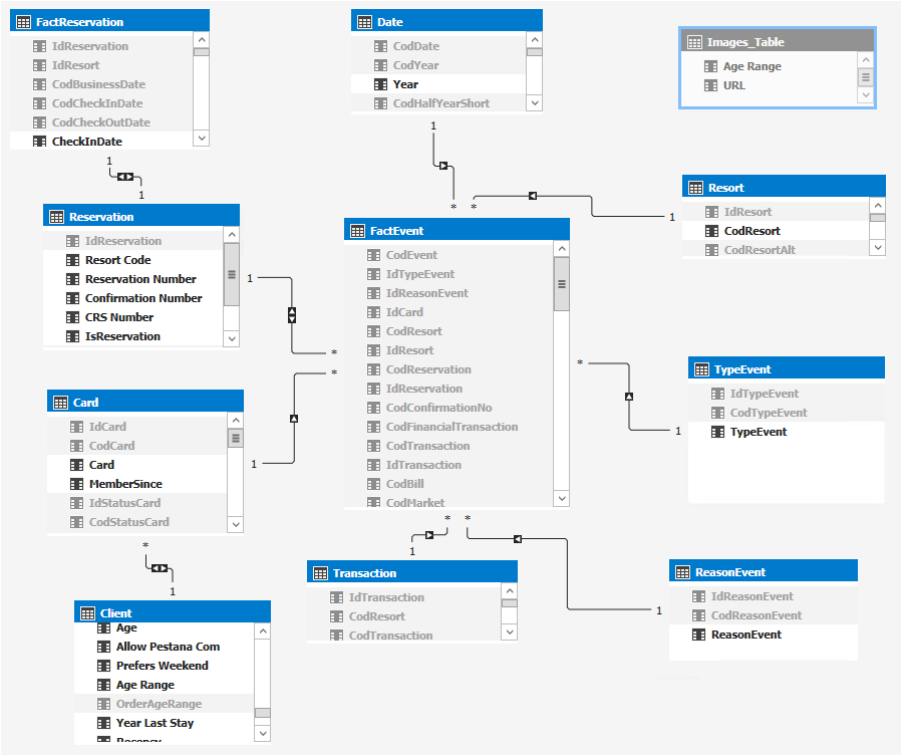


Figure 12 – PPG Data Model Diagram;

3.7. TABLES PERSONALIZATION

After the model was settled, the fields of each table were examined to see which ones needed to be hid from the client and the ones which needed to change names. This “cleaning” process took into account the cubes that already exist at Pestana Hotel Group, therefore, the field names of dimensions that already existed in other cubes were modified to match the names already given and the hierarchies created were also equal to the ones that exist.

In the dimension and fact tables that are new to this project, the followed criteria were similar to what was previously stated: the Id and Code fields were hid since the final user will not need them and the remaining fields name was changed to a more friendly name.

Both models had the same type of personalization.

3.8. PARTITIONS

The following step would be the creation of partitions. The partitions are used to create a more efficient model since a single partition can be processed at a time. This way, the cube can have refreshed data without the need to fully process.

However, in Tabular models, partitions do not significantly improve the performance of the model and the only reason to partition would be to help incremental loading or model administration. (Sirmon, Galloway, Gross, & Gulati, 2013)

Nonetheless, some tests were performed to verify if it was advantageous to partition any of the models (Client or PPG). The conclusion was that if it was to partition, the best choice would be to partition by year, however there was almost no gain in performance when partitioning, so the models will be full instead of partitioned.

3.9. MEASURES AND CALCULATED COLUMNS

Measures and calculated columns are calculations built in the tables, using DAX, to analyze the information. Note that in this project it was used a 2015 version of Visual Studio available in Pestana internal sever, which poses a difficulty, namely in accessing to more recent functions.

The measures are mainly created in fact tables. However, there are two tables (client and card) that have measures associated, even though they are dimensions. The creation of measures in dimension tables is considered a bad practice, nevertheless, and although there was a deliberation about this action and it was decided that it would be necessary due to business needs. The process of creating another two factual tables (one for the Client and another for the Card) was weighted, however, the final result for the user would have been the same since there is no difference between a dimension with measures or a factual table to the final user. Besides, in the Card table, the fields regarding balance and available balance would demand extra calculations that would affect the processing of the model and reduce the fast answer when querying.

This step is especially important using tabular model because, as it is going to be explained in the next chapter, by connecting live to Power BI it is discouraged to create measures in the report since they will be exclusive to it.

There are identical tables in both **Client** and **PPG** models, therefore, there are calculated columns and measures that are the same. Hence, the columns and measures tables presented first are the ones belonging to both models. Following these tables, details about measures and calculated columns that are unique to each model will be described.

Calculated Columns:

- **Card**

Table 5 – Shared calculated column between Client and PPG models of Card table;

| Column Name | DAX Formula | Description |
|-------------------|--|--|
| Year Member Since | IF(MONTH(Card[MemberSince]) <10, YEAR(Card[MemberSince])&"0"& MONTH(Card[MemberSince]), YEAR(Card[MemberSince]) &MONTH(Card[MemberSince])) | Year Month since when the client is a member of the loyalty program. |

- **Client**

Table 6 - Shared calculated columns between Client and PPG models of Client table;

| Column Name | DAX Formula | Description |
|-----------------------------|---|--|
| Age | <code>YEAR(TODAY())-YEAR(Client[BirthDate])</code> | Client's age. |
| Allow Pestana Com | <code>IF(Client[IsAllowedPestCom] = 1, "True", IF(Client[IsAllowedPestCom]=0, "False", "Unknown"))</code> | Binary variable turned into "Yes" or "No" regarding the authorization of being contacted by PHG. |
| Prefers Weekends | <code>IF(Client[PrefersWeekends] = 1, "Yes", IF(Client[PrefersWeekends]=0, "No", "Unknown"))</code> | Binary variable turned into "Yes" or "No" regarding if the client prefers weekends. |
| Age Range | <code>IF(Client[Age] >=18 && Client[Age]<=30, "18-30", IF(Client[Age]>30 && Client[Age]<=45, "31-45", IF(Client[Age]>45 && Client[Age]<=60, "46-60", IF(Client[Age]>60 && Client[Age]<=75, "61-75", IF(Client[Age]>75 && Client[Age]<=100, "+75", "Out of Range"))))</code> | Client's age in range. |
| Order Age Range (Hidden) | <code>IF(Client[Age] >=18 && Client[Age]<=30, 1, IF(Client[Age]>30 && Client[Age]<=45, 2, IF(Client[Age]>45 && Client[Age]<=60, 3, IF(Client[Age]>60 && Client[Age]<=75, 4, IF(Client[Age]>75 && Client[Age]<=100, 5, 0))))</code> | Column used to order the "Age Range". |

- **Date**

Table 7 - Shared calculated column between Client and PPG models of Date table;

| Column Name | DAX Formula | Description |
|------------------------------|---|---------------------------------------|
| Order Year Month (Hidden) | <code>LEFT('Date'[CodMonth],4) & RIGHT('Date'[CodMonth],2)</code> | Column created to order "Year Month". |

- **Fact Event**

Table 8 - Shared calculated columns between Client and PPG models of Fact Event table;

| Column Name | DAX Formula | Description |
|----------------------|--|---|
| ConsCredAux (Hidden) | <code>IF(AND(FactEvent[IdReasonEvent]=1, FactEvent[IdTypeEvent]=1), FactEvent[QtyPoints],0)</code> | Column referring to a credit consumption. |
| RedemDebAux (Hidden) | <code>IF(AND(FactEvent[IdReasonEvent]=2, FactEvent[IdTypeEvent]=2), FactEvent[QtyPoints],0)</code> | Column referring to a debit redemption. |
| RoomCredAux (Hidden) | <code>IF(SIGN (FactEvent[QtyRoomPoints])=-1, FactEvent[QtyRoomPoints],0)</code> | Event of credit room points. |

| | | |
|-------------------------------------|---|--|
| RoomDebAux (Hidden) | IF(SIGN (FactEvent[QtyRoomPoints])=1, FactEvent[QtyRoomPoints],0) | Event of debit room points. |
| FoodCredAux (Hidden) | IF(SIGN (FactEvent[QtyFoodPoints])=-1, FactEvent[QtyFoodPoints],0) | Event of credit food points. |
| FoodDebAux (Hidden) | IF(SIGN (FactEvent[QtyFoodPoints])=1, FactEvent[QtyFoodPoints],0) | Event of debit food points. |
| OtherCredAux (Hidden) | IF(SIGN (FactEvent[QtyOtherPoints])=-1, FactEvent[QtyOtherPoints],0) | Event of credit other points. |
| OtherDebAux (Hidden) | IF(SIGN (FactEvent[QtyOtherPoints])=1, FactEvent[QtyOtherPoints],0) | Event of debit other points. |
| YearBusinessDate (Hidden) | YEAR(TODAY()) – YEAR (FactEvent[BusinessDate]) | Quantity of years since the last event on the card. |
| Reason Of Event | IF(FactEvent[IdReasonEvent]=2 && FactEvent[CodRate]="PPG", "Full Points Redemption", IF(FactEvent[IdReasonEvent]=2 && FactEvent[CodRate]="PPGHB", "Full Points Redemption", IF(FactEvent[CodRate]="PartnershipPoints", "Partners Redemption", IF(FactEvent[IdReasonEvent]=2 && FactEvent[CodRate]<>"PPG" && FactEvent[CodRate]<>"PPGHB", "Cash & Points", SELECTCOLUMNS (RELATEDTABLE (ReasonEvent), "REASON", ReasonEvent [ReasonEvent]))) | Reason of Event with detail. |

- **Fact Reservation**

Table 9 - Shared calculated columns between Client and PPG models of Fact Reservation table;

| Column Name | DAX Formula | Description |
|-----------------|---|--|
| Cancelled? | IF(FactReservation[FlgIsCancelled] = 1, "Yes", IF(FactReservation[FlgIsCancelled]=0, "No" , "Unknown")) | Binary variable turned into "Yes" or "No" to identify if the reservation was cancelled. |
| Only Weekend? | IF(FactReservation[FlgOnlyWeekend] = 1, "Yes", IF(FactReservation [FlgOnlyWeekend]=0, "No" , "Unknown")) | Binary variable turned into "Yes" or "No" to identify if the reservation is only in weekend. |
| Only Week? | IF(FactReservation[FlgOnlyWeek] = 1, "Yes", IF(FactReservation[FlgOnlyWeek]=0, "No" , "Unknown")) | Binary variable turned into "Yes" or "No" to identify if the reservation is only in week. |
| Posting Master? | IF(FactReservation[FlgIsPostingMaster] = 1, "Yes", IF(FactReservation [FlgIsPostingMaster]=0, "No" , "Unknown")) | Binary variable turned into "Yes" or "No" to identify if the room is posting master. |
| Upgrade? | IF(FactReservation[FlgIsUpgrade] = 1, "Yes", IF(FactReservation[FlgIsUpgrade]=0, "No" , "Unknown")) | Binary variable turned into "Yes" or "No" to identify if the reservation had a room upgrade. |

| | | |
|---------------------|---|---|
| Time Share? | <pre>IF(FactReservation[FlgIsTimeShare] = 1, "Yes", IF(FactReservation[FlgIsTimeShare]=0, "No", "Unknown"))</pre> | Binary variable turned into "Yes" or "No" to identify if the reservation is time share. |
| OTA? | <pre>IF(FactReservation[FlgIsOTA] = 1, "Yes", IF(FactReservation[FlgIsOTA]=0, "No", "Unknown"))</pre> | Binary variable turned into "Yes" or "No" to identify if the was made through an online travel agent. |
| BookingWindow Range | <pre>IF(FactReservation[BookingWindow Reservation] <=3, "1-3", IF(FactReservation[BookingWindow Reservation]>3 && FactReservation[BookingWindow Reservation]<=7, "4-7", IF(FactReservation[BookingWindow Reservation]>7 && FactReservation[BookingWindow Reservation]<=15, "8-15", IF(FactReservation[BookingWindow Reservation]>15 && FactReservation[BookingWindow Reservation]<=30, "16-30", IF(FactReservation[BookingWindow Reservation]>30 && FactReservation[BookingWindow Reservation]<=50, "31-50", IF(FactReservation[BookingWindow Reservation]>50 && FactReservation[BookingWindow Reservation]<=100, "51-100", IF(FactReservation[BookingWindow Reservation]>100 && FactReservation[BookingWindow Reservation]<=150, "101-150", IF(FactReservation[BookingWindow Reservation]>150 && FactReservation[BookingWindow Reservation]<=240, "151-240", IF(FactReservation[BookingWindow Reservation]>240 && FactReservation[BookingWindow Reservation]<=365, "241-365", ">365"))))))))</pre> | Time difference between booking and reservation organized into ranges. |
| Order BW (Hidden) | <pre>IF(FactReservation[BookingWindow Reservation]<=3, 1, IF(FactReservation[BookingWindow Reservation]>3 && FactReservation[BookingWindow Reservation]<=7, 2, IF(FactReservation[BookingWindow Reservation]>7 && FactReservation[BookingWindow Reservation]<=15, 3, IF(FactReservation[BookingWindow Reservation]>15 && FactReservation[BookingWindow Reservation]<=30, 4, IF(FactReservation[BookingWindow</pre> | Column to sort the Booking Window range. |

| | | |
|-------------------|---|--|
| | <pre> Reservation]>30 && FactReservation[BookingWindow Reservation]<=50, 5, IF(FactReservation[BookingWindow Reservation]>50 && FactReservation[BookingWindow Reservation]<=100, 6, IF(FactReservation[BookingWindow Reservation]>100 && FactReservation[BookingWindow Reservation]<=150, 7, IF(FactReservation[BookingWindow Reservation]>150 && FactReservation[BookingWindow Reservation]<=240, 8, IF(FactReservation[BookingWindow Reservation]>240 && FactReservation[BookingWindow Reservation]<=365, 9, 10))))))))) </pre> | |
| LOS Range | <pre> IF(FactReservation[LOS Reservation] =1, "1", IF(FactReservation[LOS Reservation] =2, "2", IF(FactReservation[LOS Reservation]>2 && FactReservation[LOS Reservation]<=5, "3-5", IF(FactReservation[LOS Reservation]>5 && FactReservation[LOS Reservation]<=7, "6-7", IF(FactReservation[LOS Reservation]>10 && FactReservation[LOS Reservation]<=20, "8- 14", ">15"))))))) </pre> | Time difference between check in date and check out date organized into ranges. |
| OrderLOS (Hidden) | <pre> IF(FactReservation[LOS Reservation] =1, 1, IF(FactReservation[LOS Reservation] =2, 2, IF(FactReservation[LOS Reservation]>2 && FactReservation[LOS Reservation]<=5, 3, IF(FactReservation[LOS Reservation]>5 && FactReservation[LOS Reservation]<=7, 4, IF(FactReservation[LOS Reservation]>10 && FactReservation[LOS Reservation]<=20, 5,6)))))) </pre> | Column to sort the Length of Stay range. |
| Trip Type | <pre> IF(FactReservation[QtyAdults]=1 && FactReservation[QtyChildren]=0, "Individual", IF(FactReservation[QtyAdults] = 2 && FactReservation[QtyChildren]=0, "Couple", IF(FactReservation[QtyAdults]>2,"Family", IF(FactReservation[QtyAdults]=1 && FactReservation[QtyChildren]>=1, "Family", IF(FactReservation[QtyAdults]=2 && FactReservation[QtyChildren]>=1, "Family"))))))) </pre> | Association to classify the trip into three categories: Individual, Couple or Family. |
| Type Record | <pre> IF(TODAY())>FactReservation[CheckOutDate], "Actuals", IF(TODAY())>FactReservation[CheckInDate] && TODAY()<FactReservation[CheckOutDate], "Occurring", IF(FactReservation[CodCancelDate]<>-1, "Cancelled" , "On the Books")) </pre> | Association to classify the record into three types: Actuals, Occurring or On the Books. |

Measures:

- **Card**

Table 10 - Shared measures between Client and PPG models of Card table;

| Measure Name | DAX Formula | Description |
|-------------------|---|-----------------------------|
| Available Balance | <code>AVERAGE(Card[QtyAvailableBalance])</code> | Card available balance. |
| Balance | <code>AVERAGE(Card[QtyBalance])</code> | Card balance. |
| Card Quantity | <code>SUM(Card[QtyCard])</code> | Quantity of loyalty cards. |
| Expiring Points | <code>SUM(Card[QtyExpiring])</code> | Quantity of expired points. |

- **Client**

Table 11 – Shared measures between Client and PPG models of Client table;

| Measure Name | DAX Formula | Description |
|--------------------------|--|--|
| Average Nr Stays | <code>DIVIDE(SUM(FactReservation[QtyRealAux]), DISTINCTCOUNT(Client[GUID]),BLANK())</code> | Average number of stays of the clients. |
| Median Age | <code>MEDIAN(Client[Age])</code> | Median age of the clients. |
| Nights In Hotel Per Year | <code>AVERAGE(Client[QtyNightsInHotelPerYear])</code> | Average number of nights spent in the hotels per year. |

- **Fact Event**

Table 12 - Shared measures between Client and PPG models of Fact Event table;

| Measure Name | DAX Formula | Description |
|-----------------------------|---|--|
| Average Qty€ per Event | <code>DIVIDE(FactEvent[Event Amount Transaction], CALCULATE(COUNTROWS(FactEvent), FactEvent[AmtTransaction]>=0), BLANK())</code> | Average amount of money spent per event. |
| Average QtyPoints per Event | <code>DIVIDE(SUM(FactEvent[QtyPoints]), CALCULATE(COUNTROWS(FactEvent), FactEvent[QtyPoints]<>0), BLANK())</code> | Average amount of points spent per event. |
| Event Amount Transaction | <code>SUM([AmtTransaction])</code> | Amount of money in an event. |
| Food Points Credit | <code>SUM(FactEvent[FoodCredAux])</code> | Points spent in F&B purchase. |
| Food Points Debit | <code>SUM(FactEvent[FoodDebAux])</code> | Points earned in F&B purchase. |
| Other Points Credit | <code>SUM(FactEvent[OtherCredAux])</code> | Points spent in Other purchase (for example, Spa). |
| Other Points Debit | <code>SUM(FactEvent[OtherDebAux])</code> | Points earned in Other purchase. |
| Points Consum Credit | <code>ABS(SUM(FactEvent[ConsCredAux]))</code> | Points associated with a consumption. |
| Points Consumption € | <code>[Points Consum Credit]*0.003</code> | Money amount of consumed points. |

| | | |
|-------------------------|---|---------------------------------------|
| Points Consumption LY | <code>CALCULATE([Points Consum Credit], DATEADD(FactEvent[BusinessDate], -1, YEAR))</code> | Points consumed in the previous year. |
| Points Credit € | <code>ABS([Room Points Credit]+[Food Points Credit]+[Other Points Credit])*0.003</code> | Amount of money of credit events. |
| Points Debit € | <code>ABS([Room Points Debit]+[Food Points Debit]+[Other Points Debit])*0.003</code> | Amount of money of debit events. |
| Points Redemption € | <code>[Points Redemption Debit]*0.003*2</code> | Money amount of redeemed points. |
| Points Redemption Debit | <code>ABS(SUM(FactEvent[RedemDebAux]))</code> | Points associated with a redemption. |
| Points Redemption LY | <code>CALCULATE([Points Redemption Debit], DATEADD(FactEvent[BusinessDate],-1,YEAR))</code> | Points redeemed in the previous year. |
| Room Points Credit | <code>SUM(FactEvent[RoomCredAux])</code> | Points spent in Room purchase. |
| Room Points Debit | <code>SUM(FactEvent[RoomDebAux])</code> | Points earned in Room purchase. |

- **Fact Reservation**

Table 13 - Shared measures between Client and PPG models of Fact Reservation table;

| Measure Name | DAX Formula | Description |
|--------------------------|---|---|
| % Clients Only Weekend | <code>DIVIDE(CALCULATE(DISTINCTCOUNT(FactReservation[GUID])*100, FactReservation [Only Weekend]=1, FactReservation [QtyRealAux]=1), CALCULATE (DISTINCTCOUNT (FactReservation[GUID]), FactReservation [QtyRealAux]=1), BLANK())</code> | Percentage of clients that stays only in weekends. |
| % Clients RoomUpgrade | <code>DIVIDE(CALCULATE(DISTINCTCOUNT(FactReservation[GUID])*100, FactReservation [Is Upgrade]=1, FactReservation [QtyRealAux]=1), CALCULATE(DISTINCTCOUNT(FactReservation [GUID]), FactReservation [QtyRealAux]=1), BLANK())</code> | Percentage of clients that had a room upgrade during the stay (room category of the stay is different than the one booked). |
| % Reservations Cancelled | <code>DIVIDE(CALCULATE(COUNT(FactReservation [IdReservation]), FactReservation [Is Cancelled]=1, FactReservation [QtyRealAux]=1)*100, CALCULATE (COUNT(FactReservation [IdReservation]), FactReservation [QtyRealAux]=1), BLANK())</code> | Percentage of reservations that were cancelled. |
| %FB Expenses | <code>ROUND(DIVIDE([Food Rev Eur], [Total Rev Eur],0)*100,2)</code> | Percentage of F&B expenses among all the stay expenses. |
| %Other Expenses | <code>ROUND(DIVIDE([Other Rev Eur], [Total Rev Eur],0)*100,2)</code> | Percentage of Other expenses (for example, spa) among all the stay expenses. |
| Adults | <code>CALCULATE(SUM ([QtyAdults]), FactReservation[QtyRealAux]=1)</code> | Quantity of adults that stayed in the hotel. |

| | | |
|-----------------------|---|---|
| ARR | $\text{ROUND}(\text{DIVIDE}([\text{Room Rev Eur}], [\text{RN}], 0), 2)$ | Average Room Rate. |
| ATR | $\text{DIVIDE}([\text{Total Rev Eur}], [\text{RN}], \text{BLANK}())$ | Average Total Rate. |
| Avg Price Reservation | $\text{DIVIDE}([\text{Total Rev Eur}], \text{SUM}(\text{FactReservation}[\text{QtyRealAux}], \text{BLANK}()))$ | Average price of the reservation. |
| Booking Window | $\text{ROUND}(\text{AVERAGE}([\text{BookingWindow Reservation}], 2)$ | Time between the booking and the stay. |
| Children | $\text{CALCULATE}(\text{SUM}([\text{QtyChildren}], \text{FactReservation}[\text{QtyRealAux}=1])$ | Quantity of children that stayed in the hotel. |
| Clients | $\text{CALCULATE}(\text{SUM}([\text{QtyClients}], \text{FactReservation}[\text{QtyRealAux}=1])$ | Quantity of clients that stayed in the hotel. |
| F&B Points | $\text{SUM}([\text{FBPoints}])$ | F&B points of the reservation. |
| Food Rev Eur | $\text{SUM}([\text{AmtFoodRevenueEUR}])$ | F&B revenue of the reservation. |
| Food Rev Gross Eur | $\text{SUM}([\text{AmtFoodRevenueGrossEUR}])$ | F&B gross revenue of the reservation. |
| LOS | $\text{ROUND}(\text{CALCULATE}(\text{AVERAGE}(\text{FactReservation}[\text{LOS Reservation}], \text{FactReservation}[\text{QtyRealAux}=1]), 2)$ | Length of stay. |
| Other Points | $\text{SUM}([\text{OtherPoints}])$ | Other points of the reservation. |
| Other Rev Eur | $\text{SUM}([\text{AmtOtherRevenueEUR}])$ | Other revenue of the reservation. |
| Other Rev Gross Eur | $\text{SUM}([\text{AmtOtherRevenueGrossEUR}])$ | Other gross revenue of the reservation. |
| Physical | $\text{ROUND}(\text{CALCULATE}(\text{SUM}([\text{QtyPhysical}], \text{FactReservation}[\text{QtyRealAux}=1]), 2)$ | Physical amount of rooms. |
| Reserv % of Week | $\text{ROUND}(\text{DIVIDE}([\text{Weekdays}], ([\text{Weekdays}] + [\text{Weekends}]), 0) * 100, 2)$ | Percentage of the reservation corresponding to week. |
| Reserv % of Weekends | $\text{ROUND}(\text{DIVIDE}([\text{Weekends}], ([\text{Weekdays}] + [\text{Weekends}]), 0) * 100, 2)$ | Percentage of the reservation corresponding to weekend. |
| RN | $\text{SUM}([\text{QtyNights}])$ | Quantity of nights of a stay. |
| Room Points | $\text{SUM}([\text{RoomPoints}])$ | Room points of the reservation. |
| Room Rev Eur | $\text{SUM}([\text{AmtRoomRevenueEUR}])$ | Room revenue of the reservation. |
| Room Rev Gross Eur | $\text{SUM}([\text{AmtRoomRevenueGrossEUR}])$ | Room gross revenue of the reservation. |
| Total Points | $\text{SUM}([\text{TotalPoints}])$ | Total points of the reservation. |
| Total Rev Eur | $\text{SUM}([\text{AmtTotalRevenueEUR}])$ | Total revenue of the reservation. |
| Total Rev Gross Eur | $\text{SUM}([\text{AmtTotalRevenueGrossEUR}])$ | Total gross revenue of the reservation. |
| Weekdays | $\text{ROUND}(\text{CALCULATE}(\text{AVERAGE}([\text{QtyWeekdays}], \text{FactReservation}[\text{QtyRealAux}=1]), 2)$ | Amount of weekdays of a reservation. |
| Weekends | $\text{ROUND}(\text{CALCULATE}(\text{AVERAGE}([\text{QtyWeekends}], \text{FactReservation}[\text{QtyRealAux}=1]), 2)$ | Amount of weekends of a reservation. |

As mentioned, besides the calculated columns and measures created in both models, there were also ones created in only one of these two models.

In the **Client model**, the unique columns created are:

- Client

Table 14 - Unique calculated column of Client table – Client model;

| Column Name | DAX Formula | Description |
|-----------------------------|--|---|
| Is PPG | <code>IF(Client[CRMNumber] <> BLANK(), "Yes", "No")</code> | Flag to identify if the client is a part of the loyalty program. |
| Recency | <code>IF(Client[LastStayedDate]=0 Client[LastStayedDate]=-1, BLANK(), DATEDIFF(FORMAT(Client[LastStayedDate], "Short Date"), TODAY(), MONTH))</code> | Quantity of months since the last purchase. |
| Is Repeater | <code>IF(Client[Qty Reservations Client]>1, 1,0)</code> | Flag to identify if the client stayed in a PHG hotel more than once. |
| Qty Reservations Client | <code>CALCULATE(COUNTROWS(FactReservation), FILTER(FactReservation, FactReservation [GUID]= Client[GUID]), FactReservation[CodCancelDate]=-1, FactReservation[QtyRealAux]=1)</code> | Quantity of reservations that each client has. |
| Has OTB Reservation | <code>IF(CALCULATE(COUNT(FactEvent[IdReservation]), FILTER(FactReservation, FactReservation [GUID]=Client[GUID] && FactReservation[Type Record] = "On the Books"))>=1, "Yes", "No")</code> | Flag to identify if the client already has a future reservation in a PHG hotel. |
| Reservations Range | <code>IF(ISBLANK(Client[Qty Reservations Client])=TRUE(), BLANK(), IF(Client[Qty Reservations Client]=1, "1", IF(Client[Qty Reservations Client]=2, "2", IF(Client[Qty Reservations Client]=3, "3", IF(Client[Qty Reservations Client]>3 && Client[Qty Reservations Client]<=6, "4-6", IF(Client[Qty Reservations Client]>6 && Client[Qty Reservations Client]<=10, "7-10", IF(Client[Qty Reservations Client]>10 && Client[Qty Reservations Client]<=20, "10-20", ">20"))))))))</code> | Column that classifies the quantity of reservations into a range. |
| Reservations Order (Hidden) | <code>IF(ISBLANK(Client[Qty Reservations Client])=TRUE(), 0, IF(Client[Qty Reservations Client]=1, 1, IF(Client[Qty Reservations Client]=2, 2, IF(Client[Qty Reservations Client]=3, 3, IF(Client[Qty Reservations Client]>3 && Client[Qty Reservations Client]<=6, 4, IF(Client[Qty Reservations Client]>6 && Client[Qty Reservations Client]<=10, 5, IF(Client[Qty Reservations Client]>10 && Client[Qty Reservations Client]<=20, 6, 7))))))</code> | Column that allows to sort the Reservations Range. |

- **Fact Campaign Interaction**

Table 15 - Unique calculated column of Fact Campaign Interaction table – Client model;

| Column Name | DAX Formula | Description |
|-------------|---|--|
| Opened? | <code>IF(FactCampaignInteraction[QtdOpens] >=1, "Opened", "Not Opened")</code> | Flag to identify if the campaign was opened by the client. |
| Clicked? | <code>IF(FactCampaignInteraction[QtdClicks]>=1, "Clicked", "Not Clicked")</code> | Flag to identify if the link of the campaign was clicked. |

- **Fact Reservation**

Table 16 - Unique calculated column of Fact Reservation table – Client model;

| Column Name | DAX Formula | Description |
|--------------------|--|--|
| Age At Reservation | <code>IF(ISBLANK(SELECTCOLUMNS(RELATEDTABLE(Client), "ClientBirth", Client[BirthDate])), BLANK(), DATEDIFF(SELECTCOLUMNS(RELATEDTABLE(Client), "ClientBirth", Client[BirthDate]), FactReservation[CheckInDate], YEAR))</code> | Age the customer has at the time of the reservation. |
| QtyRealAux | <code>IF(FactReservation[QtyNights]=0, 0, 1)</code> | Flag to identify if the reservation is real or not. |

- **Market**

Table 17 - Unique calculated column of Market table – Client model;

| Column Name | DAX Formula | Description |
|-------------|--|--|
| Segment | <code>IF(Market[Market]= "TO - Web Based" , "OTA B2B", IF(Market[Market]="TO - YIELD" , "OTA B2C", IF(Market[Market Master]="Contracted" && Market[Market]<>"TO - Web Based" && Market[Market]<>"TO - YIELD", "Contracted Offline", IF(Market[Market]="TD - Pestana Site", "Site", IF(Market[Market]="TD - Central Reservations Direct Contact Center", "Call Center", IF(Market[Market Master]="Business", "Groups", IF(Market[Market Master]="Corporate", "Corporate", IF(Market[Market Master]="Direct" && Market[Market]<>"TD - Pestana Site" && Market[Market]<>"TD - Central Reservations Direct Contact Center", "Other Direct", Market[Market Master]))))))))</code> | Market segment at which the reservation belongs. |

- Meal Type

Table 18 - Unique calculated column of Meal Type table – Client model;

| Column Name | DAX Formula | Description |
|--------------|--|---------------------------------|
| Type of Meal | <pre>IF(MealType[F&B Service]="Breakfast" MealType[F&B Service]="Breakfast Overbooking", "Breakfast", IF(MealType[Meal]="Dinner", "Dinner", IF(MealType[Meal]="Lunch", "Lunch", IF(MealType[Package]="Full Board" MealType[Package]="All Inclusive", "Full Board", IF(MealType[Package]="Half Board", "Half Board", "Other F&B")))))</pre> | Type of meal of the reservation |

The measures created exclusively to the **Client** model are:

- Client

Table 19 - Unique measure of Client table – Client model;

| Measure Name | DAX Formula | Description |
|------------------------------|---|---|
| Qty Reservations CI Final | <pre>SWITCH(SUM(Vision[Code]), 1, SUMX(ADDCOLUMNS (SUMMARIZE (Client,Client[GUID]), "A", CALCULATE(SUM(Client[Qty Reservations Client]), ALL (CheckInDate), ALL (Resort))), [A]), 2, SUMX (ADDCOLUMNS (SUMMARIZE(Client, Client[GUID]), "A", CALCULATE (SUM(Client[Qty Reservations Client]), ALL (CheckInDate))), [A]), SUM(Client[Qty Reservations Client]))</pre> | Measure that allow to show the quantity of reservations according to one of the views: standard, date independent or date & resort independent. |

- Fact Campaign Interaction

Table 20 - Unique measure of Fact Campaign Interaction table – Client model;

| Measure Name | DAX Formula | Description |
|---------------------|--|---|
| Clicks | <pre>SUM([QtdClicks])</pre> | Quantity of clicks in the link of the campaign. |
| Click-to-open Ratio | <pre>ROUND(DIVIDE([Clicks]*100, [Opens], BLANK()),1)</pre> | Quantity of link clicks per quantity of opens. |
| Open Rate | <pre>ROUND(DIVIDE ([Opens]*100,[Sents], BLANK()),1)</pre> | Rate of opening of the campaign. |
| Opens | <pre>SUM([QtdOpens])</pre> | Quantity of campaign openings. |
| Sents | <pre>SUM([QtdSents])</pre> | Quantity of campaign emails sent. |

- **Fact FoodBeverage**

Table 21 - Unique measure of Fact Food & Beverage table – Client model;

| Measure Name | DAX Formula | Description |
|---------------|-------------------------------------|--------------------------------|
| Gross Revenue | <code>SUM([AmtGrossRevenue])</code> | Revenue without taxes applied. |
| Net Revenue | <code>SUM([AmtNetRevenue])</code> | Revenue with taxes applied. |

- **Fact Reservation**

Table 22 - Unique measure of Fact Reservation table – Client model;

| Measure Name | DAX Formula | Description |
|----------------------|--|---|
| Avg Qty Reservations | <code>DIVIDE(COUNTA(FactReservation[GUID]), DISTINCTCOUNT(FactReservation[GUID]), BLANK())</code> | Average quantity of reservations. |
| Repeaters Final | <code>SWITCH(SUM(Vision[Code]) ,1, SUMX(ADDCOLUMNS (SUMMARIZE (Client,Client[GUID]), "A", CALCULATE([Qty Repeaters], ALL (CheckInDate), ALL (Resort))), [A])), 2, SUMX (ADDCOLUMNS (SUMMARIZE(Client, Client[GUID]), "A", CALCULATE ([Qty Repeaters], ALL (CheckInDate))), [A]), [Qty Repeaters]))</code> | Measure that allow to show the quantity of clients that stayed more than once (repeaters) according to one of the views: standard, date independent or date & resort independent. |

- **Fact Review**

Table 23 - Unique measure of Fact Review table – Client model;

| Measure Name | DAX Formula | Description |
|--------------|---|--|
| CSAT | <code>DIVIDE(CALCULATE(COUNT (FactReview [ScrOverall]), OR(FactReview [ScrOverall] = 4, FactReview [ScrOverall]=5)), SUM(FactReview [QtyScrOverall]), BLANK()) * 100</code> | Customer Satisfaction Score. |
| GSS | <code>DIVIDE((SUM(FactReview[ScrOverall]) / 5), SUM(FactReview[QtyScrOverall]), BLANK())*100</code> | Guest satisfaction survey – score obtained in the internal survey. |
| GSS Bar | <code>DIVIDE((SUM(FactReview[ScrBar]) / 5), SUM(FactReview[QtyScrBar]), BLANK())*100</code> | Guest satisfaction survey in the bar. |
| GSS Bar Qty | <code>SUM([QtyScrBar])</code> | Quantity of responses to the bar question. |

| | | |
|--------------------|--|---|
| GSS Breakfast | $\text{DIVIDE}((\text{SUM}(\text{FactReview} [\text{ScrBreakfast}]) / 5), \text{SUM}(\text{FactReview} [\text{QtyScrBreakfast}]), \text{BLANK}()) * 100$ | Guest satisfaction survey in the breakfast. |
| GSS Breakfast Qty | $\text{SUM}([\text{QtyScrBreakfast}])$ | Quantity of responses to breakfast bar question. |
| GSS FB | $\text{DIVIDE}((\text{SUM} (\text{FactReview} [\text{ScrFoodAndBeverages}]) / 5), \text{SUM}(\text{FactReview} [\text{QtyScrFoodAndBeverages}]), \text{BLANK}()) * 100$ | Guest satisfaction survey in the F&B. |
| GSS FB Qty | $\text{SUM}([\text{QtyScrFoodAndBeverages}])$ | Quantity of responses to the F&B question. |
| GSS NPS | $\text{DIVIDE} (\text{SUM} (\text{FactReview} [\text{QtyPromoter}]) * 100, \text{SUM} (\text{FactReview} [\text{QtyScrRecommend}]), \text{BLANK}()) - \text{DIVIDE} (\text{SUM} (\text{FactReview} [\text{QtyDetractor}]) * 100, \text{SUM} (\text{FactReview} [\text{QtyScrRecommend}]), \text{BLANK}())$ | Net Promoter Score, measures how much a customer recommend a brand. |
| GSS Restaurant | $\text{DIVIDE}((\text{SUM}(\text{FactReview} [\text{ScrRestaurant}]) / 5), \text{SUM}(\text{FactReview} [\text{QtyScrRestaurant}]), \text{BLANK}()) * 100$ | Guest satisfaction survey in the Restaurant. |
| GSS Restaurant Qty | $\text{SUM}([\text{QtyScrRestaurant}])$ | Quantity of responses to the restaurant question. |
| GSS Spa | $\text{DIVIDE}((\text{SUM}(\text{FactReview} [\text{ScrSpa}]) / 5), \text{SUM}(\text{FactReview}[\text{QtyScrSpa}]), \text{BLANK}()) * 100$ | Guest satisfaction survey in the Spa. |
| GSS Spa Qty | $\text{SUM}([\text{QtyScrSpa}])$ | Quantity of responses to the spa question. |

In the **PPG model**, the unique columns created are:

- **Card**

Table 24 – Unique calculated column of Card table – PPG model;

| Column Name | DAX Formula | Description |
|---------------------|---|--|
| Expiring Year Month | $\text{IF}(\text{MONTH}(\text{Card}[\text{ExpiringDate}]) < 10, \text{YEAR}(\text{Card}[\text{ExpiringDate}]) \& "0" \& \text{MONTH}(\text{Card}[\text{ExpiringDate}]), \text{YEAR}(\text{Card}[\text{ExpiringDate}]) \& \text{MONTH}(\text{Card}[\text{ExpiringDate}]))$ | Year Month at which the card will have expired points. |

- **Client**

Table 25 - Unique calculated column of Client table – PPG model;

| Column Name | DAX Formula | Description |
|----------------|--|--|
| Year Last Stay | <code>IF(Client[LastStayedDate]=0, BLANK(), YEAR(Client[LastStayedDate]))</code> | Last year that the client stayed at a Pestana hotel. |

The measures created only in this **PPG model** are:

- **Card**

Table 26 - Unique measure of Card table – PPG model;

| Measure Name | DAX Formula | Description |
|---------------------------|--|--|
| % New Clients | <code>DIVIDE([Qty New Clients], DISTINCTCOUNT(Card[IdCard]), BLANK())</code> | Percentage of new clients in the loyalty program. |
| Cards With Expired Points | <code>CALCULATE(COUNTROWS(Card), Card[QtyExpiring] >0)</code> | Amount of loyalty cards which have expired points. |
| Expiring Points € | <code>[Expiring Points]*0.003</code> | Points worth in money. |
| New Clients LY | <code>CALCULATE([Qty New Clients], DATEADD(Card[MemberSince],-1,YEAR))</code> | Amount of clients that entered the loyalty program in the previous year. |
| Qty New Clients | <code>CALCULATE(DISTINCTCOUNT(Card[IdCard]), YEAR(Card[MemberSince]) = YEAR(TODAY()))</code> | Amount of clients that entered the loyalty program in the current year. |

- **Fact Event**

Table 27 - Unique measure of Fact Event table – PPG model;

| Measure Name | DAX Formula | Description |
|-------------------------|--|--|
| % Cash Consumptions | <code>DIVIDE([Points Consumption €]*100, ([Points Redemption €]+[Points Consumption €]), BLANK())</code> | Percentage of money spent in consumptions among all the money spent. |
| % Cash Redeemed | <code>DIVIDE([Points Redemption €]*100, ([Points Redemption €]+[Points Consumption €]), BLANK())</code> | Percentage of money spent in redemptions among all the money spent. |
| % Consumption in Events | <code>DIVIDE(CALCULATE(COUNTA(FactEvent[CodEvent]), FactEvent[IdReasonEvent]=1)*100, COUNTROWS (FactEvent), BLANK())</code> | Percentage of events that are a consumption among all the events. |
| %Redemption in Events | <code>DIVIDE(CALCULATE(COUNTA(FactEvent [CodEvent]), FactEvent [IdReasonEvent] =2)*100, COUNTROWS (FactEvent), BLANK())</code> | Percentage of events that are a redemption among all the events. |

| | | |
|--------------------------|--|--|
| Average € spent per Card | <code>DIVIDE(FactEvent[Event Amount Transaction], CALCULATE(COUNT (Card[IdCard]), FactEvent [AmtTransaction]<>0), BLANK()))</code> | Average amount of money spent per card. |
| Card Active Last 2y | <code>CALCULATE(DISTINCTCOUNT(FactEvent[IdCard]), FactEvent[YearBusinessDate]<=2)</code> | Amount of cards that have an event in the last two years. |
| CRM Sponsorship | <code>[Points Redemption €]/2</code> | Each time a customer earns a point, the CRM team sponsors another. |

3.10. DAX CALCULATED TABLES

Both models had a table creation through a DAX formula. The **Client model** has a table called “View” that will also be used in the report and that has as goal the switch of a measure value according to the analysis made, whether it is a standard view, a view independent of date or even independent of date & resort. The formula used was:

```
View = DATATABLE ("Code", INTEGER, "View", STRING, { { 1, "Date & Resort Independent" }, { 2, "Date Independent"}, { 3, "Standard"} })
```

This formula created a table with two columns and three rows in which the values of the first column are integers and the values in the second columns are strings.

The **PPG model** has a table called “Images_Table” that will be used in the report and that is meant to change the image showed according to the clients age range, this detail will be mentioned in a section ahead called “dynamically display of images”. The formula used was:

```
Images_Table = DATATABLE ("Age Range", STRING, "URL", STRING, { { "18-30", "https://image.freepik.com/vetores-gratis/casal-feliz-de-turistas-com-mala-de-viagem-pessoas-de-ferias-de-verao_167524-56.jpg" }, { "31-45", "https://image.freepik.com/vetores-gratis/casal-jovem-mapa-turistico-mochila-cap-viajante-urbano-fundo_18591-5154.jpg" }, { "46-60", "https://previews.123rf.com/images/yupiramos/yupiramos1701/yupiramos170109688/70084206-people-tourists-avatars-characters-vector-illustration-design.jpg" }, { "61-75", "https://image.freepik.com/vetores-gratis/pessoas-idosas-viajantes-de-ferias-sorrindo-avos-de-ferias-conceito-de-viagem-de-viagem-do-veterano-idoso-feliz-velho-viagem-homem-e-mulher-avos-com-bagagem-para-ilustracao-de-ferias_53562-4671.jpg" }, { "+75", "https://image.freepik.com/vetores-gratis/viagens-e-turismo-no-mundo_24640-55993.jpg" } })
```

This formula created a table with two columns and five rows in which the values of the columns are strings and the second columns was set as “ImageUrl” in data category.

3.11. PERSPECTIVES

Perspectives are a part of the model usually related to a specific part of the business and particular points of view.

In these models, perspectives will not be created as the teams (Customer Relationship Management, Email Marketing and Business Intelligence) that will use them are in need for all the information.

3.12. ROLES

The users that will have access to both the models need to be embedded into a role. The roles can have different permissions: administrator, process, read and process, read or none.

Two types of roles will be added to the models: **Administrators** and **Readers**. The Administrators and Readers have a specific group already used in other models and cubes, the lists were imported from an AD (active directory).

3.13. DEPLOY

To allow users to browse the models, they need to be deployed and afterwards accessed in SQL Server Management Studio, Excel or Power BI.

Before deploying the model, it was built. Subsequently, the model will be deployed. Even though there is an option to deploy in the SQL Server Data Tools, the choice was to utilize Analysis Services Deployment Wizard.

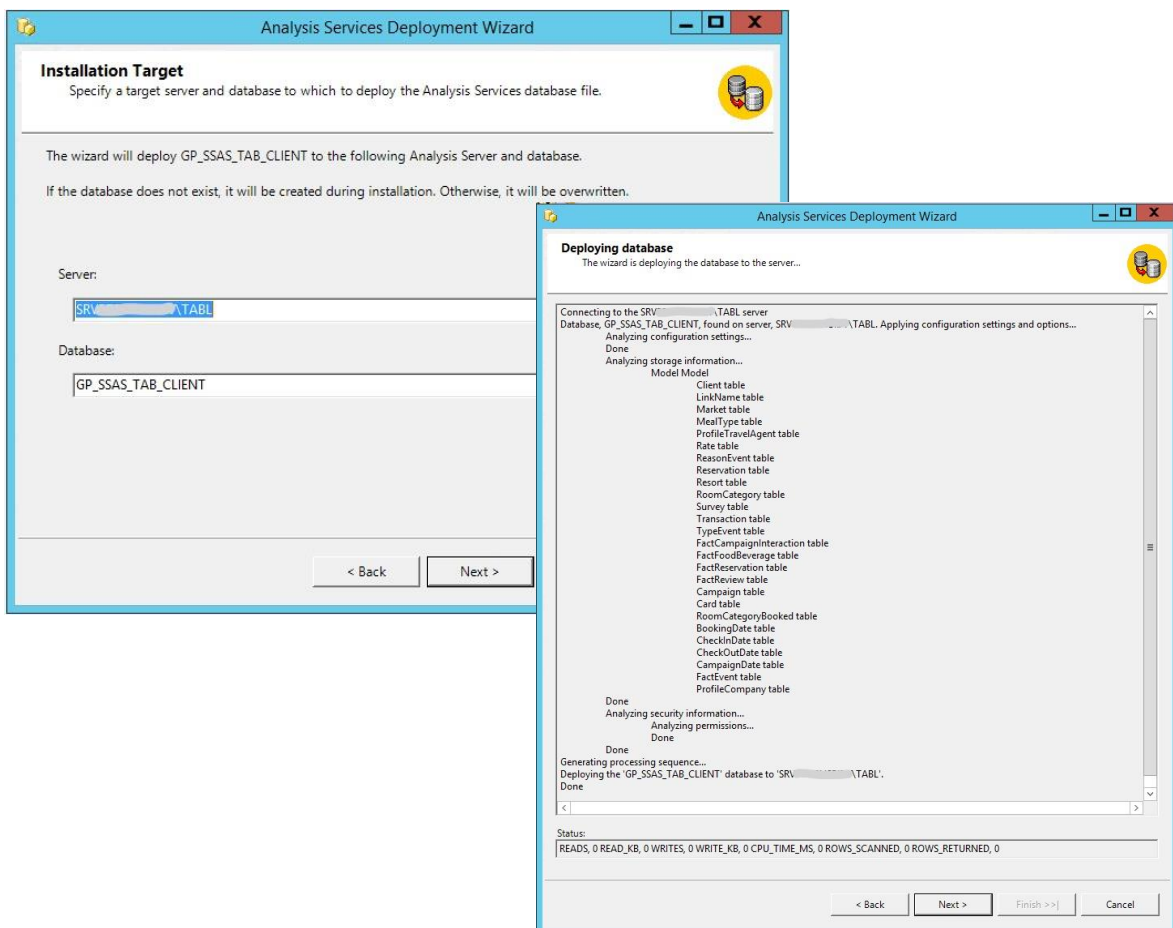


Figure 13 – Analysis Services Deployment Wizard details;

3.14. SSIS PROCESSING PACKAGE

There are different techniques to process the model, the one chose in this project was to create a SQL Server Integration Services package to do this. This form will allow to run the package remotely, without the need of intervention and scheduled by afterwards setting a SQL Server Agent. This package is integrated following the daily information processing.

Since the decision was not to create partitions, then the SSIS (SQL Server Integration Services) package that is going to be created is simpler. However, there was a consideration between choosing the “Analysis Services Execute DDL Task” or “Analysis Services Processing Task”.

After some performance tests, the decision was to choose the “Analysis Services Processing Task” since the processing time with this task took almost half the time it took when using the DDL task.

Subsequently, three connections were set to the Data Warehouse, another to the Operational Data Store and to the SQL Server Analysis Services referring to the client.



Figure 14 – Connections of the SSIS package;

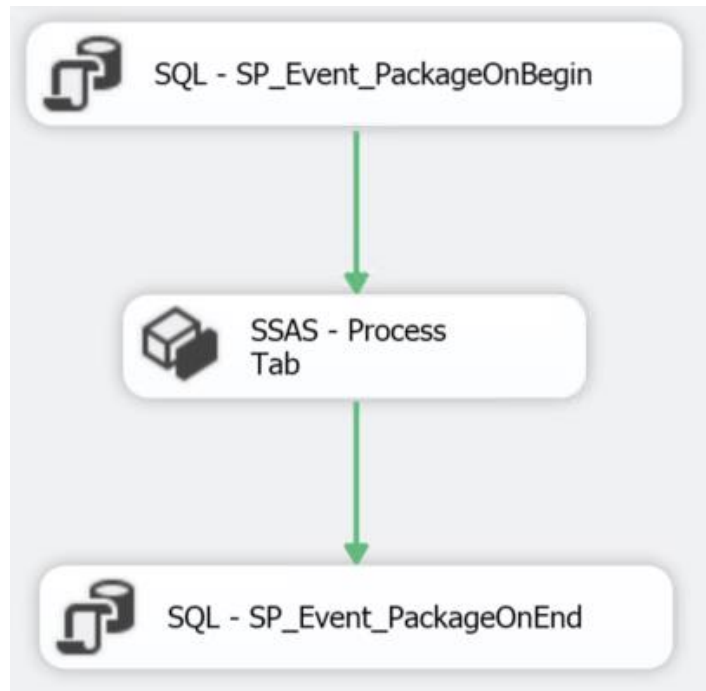


Figure 15 – Package to process the model;

3.15. PUBLISHING IN PRODUCTION

After every stage above is done in the development environment, it was time to publish the models in the production environment. The development environment is a safe space to work since that if any errors are made the consequences are not significant, in contrast to the huge loss of information it could happen if the same mistakes were done in live and up-to-date data that is available in the production server.

As such, four main steps were pursued:

1. Create the necessary views in the data warehouse of the production environment;

The views created in the data warehouse are the ones mentioned previously that are exclusive to these two models since that all the others were previously created and therefore the views were reused.

2. Models (Client and PPG) deployment in production environment;

In order to be in the production environment, the models need to be deployed. As before, it was performed through Analysis Services Deployment Wizard, only changing the connection to the production server.

3. Creation of the processing package that will allow the process of the models (Client and PPG);

Since the goal is to automate the processing of these models, the two packages were created which will allow the joining of the packages to the ones previously created.

4. Full Process both models;

By full processing the models, the data becomes up-to-date and the tables populated. After the packages creation, a SQL Server Agent Job is set which will permit the run of these packages in the automatic process that runs daily with data from the previous day.

4. REPORT DESIGNING

4.1. DATA LOADING STRATEGY

Considering that the industry in question is hospitality, it is important to mention that the usual behavior of the clients is to book few reservations per year, not the case in other industries such as retail where customers purchase on a weekly or monthly basis. The strategy followed to load the data into the model was “Live Connection”, this was done because the models were based on customer data and it makes no sense to build or have sliding windows of dates since the interest is in all the data and history.

Connection live has several advantages when linked to a tabular model, such as having analytical capabilities of SSAS and not stocking data in memory. For these reasons, visualizations built on Power BI query the data source directly and allow large-scale models. This is explained by faster calculations since data is warehoused in memory of the machine that runs SSAS Tabular. (Aspin, 2018) & (Rad, 2018)

Besides, importing data directly from a tabular model does not allow to change neither the architecture of the model nor the relationships, only allowing to create report level measures, this way there is a single source of the data.

4.2. MODEL IMPORT TO POWER BI

The first step when importing the **Client Model** to Power BI was to select the choice regarding the Analysis Services. Then, after filling the server and database names, a “Connect Live” option was chose to import the model.

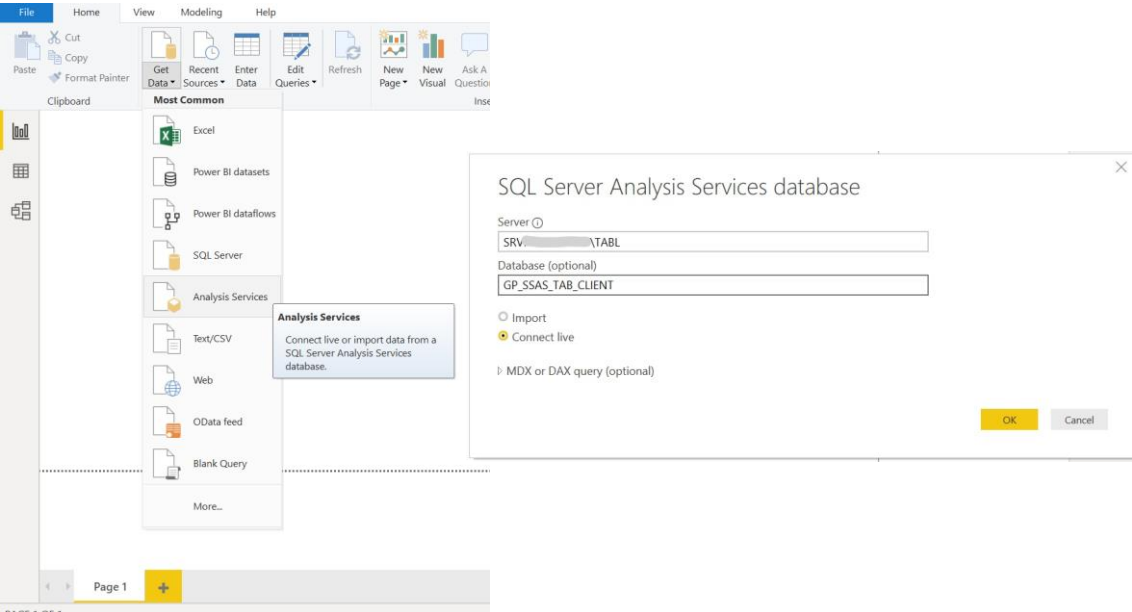


Figure 16 – Data Import in Power BI;

The same strategy was pursued when importing data to **PPG model** changing only the connection of the database to “GP_SSAS_TAB_PPG”.

4.3. REQUIREMENTS OF THE REPORTS

As known, there are several goals and advantages in knowing the customers of a business, particularly their characteristics and consumer behavior. As stated by (Butscher, 2002) there are five main goals in having loyalty programs such as:

1. Turning the customers into lasting loyal customers by creating a relationship with them;
2. Attracting new clients through giving appealing benefits to the current ones and satisfying the existing clients enough so they can “free-advertise”;
3. Acting as support to other departments such as product marketing or marketing itself;
4. Creating communication opportunities which can be used to intensify contact with program members. These opportunities can help the email marketing team;
5. Creating a customer database.

The last goal mentioned is the most important one following the scope of this project and it is the lack that this project proposes address. The 3rd and 4th goals are a consequence of having a rich database and also easier to reach by having both models and reports.

A customer database is considered the most powerful marketing tool a company can have and it can also be used for several marketing activities. Likewise, it is stated that taking advantage of this database is using a strategic weapon in terms of marketing. (Butscher, 2002) Even though this information is challenging to attain, once it depends on the customer filling the personal information in the forms when joining the program or the information provided to the front-office in the check-in, it is very important to not only having socio-demographic data but also comprehensive information on purchase compartment and preference data. (Butscher, 2002)

As previously mentioned, the main purpose of these models is helping to better know the Pestana client. By achieving this goal, it will also be possible to channel campaigns specific to a type of client helping the marketing and email marketing teams or understand the profile of a client associated with the loyalty program. Therefore, both models and dashboards will have a strong track to these teams which triggered that the requirements of their usual reports were enquired.

Among the requirements that the CRM, Marketing and Email Marketing asked to be present in the reports are the following:

- **CRM**

1. **Program Growth**

Evolution of the enrollments in the program; Enrollments by unit, country or brand.

2. **Program Demographics**

Split by top 10 markets (nationality of the client), age ranges and gender, overview of clients by card tier.

3. Program Reachability

Evolution of the number of active and valid emails; Evolution of allowed communications compared to last year and year-to-date.

4. Program Engagement

Activity rate evolution of PPG client (clients with events on the last two years); Top 10 units where there are more consumed points; Top 10 units where there are more redeemed points; Total of expired, consumed and redeemed points, compared both with last year and year-to-date.

a. Program Details

Average money spent by client and average points by reservation.

- **Email Marketing**

- 1. **Campaign Analysis**

- Analysis of emails sent, opened and clicks on the campaigns in the email using metrics as click-to-open rate or open rate.

- **Marketing**

- A better knowledge of the client will allow to personalize the communications allowing to satisfy clients' necessities and wills.

These models will provide the following client details: demographics (age, gender, nationality...), reservations by year, booking window, consumptions during the reservation, length of stay, type of booked room or type of reservation (week days or weekends).

In addition to the requirements deepened above, other details might be learned while browsing tabular models that are not a part of the requirements asked by the teams.

4.4. REPORT DESIGN AND WIREFRAMING

Some details were taken into consideration when designing these two reports. Among them:

1. All the pages are identified with PHG or PPG logo;
2. There was an attempt to keep the report visually soft by maintaining a color scheme (orange and blue in the Client report, green and yellow in the PPG report);
3. Defined structure meaning that the filters are always at the top of the page;
4. KPIs visuals are at the left or at a top panel;

5. Instead of having a Y axis choosing the data labels since it helps in reading the graphics;
6. All the visuals that make sense have percentage of values as an alternative to absolute values to ease comparisons;
7. When possible, the values have units (as thousands or millions) to facilitate the numbers reading.

Wireframing is the structure behind a report. In the Power BI built reports, not all the pages have the same wireframe even though all of them follow the seven details mentioned before. The wireframe of the reports are represented below.

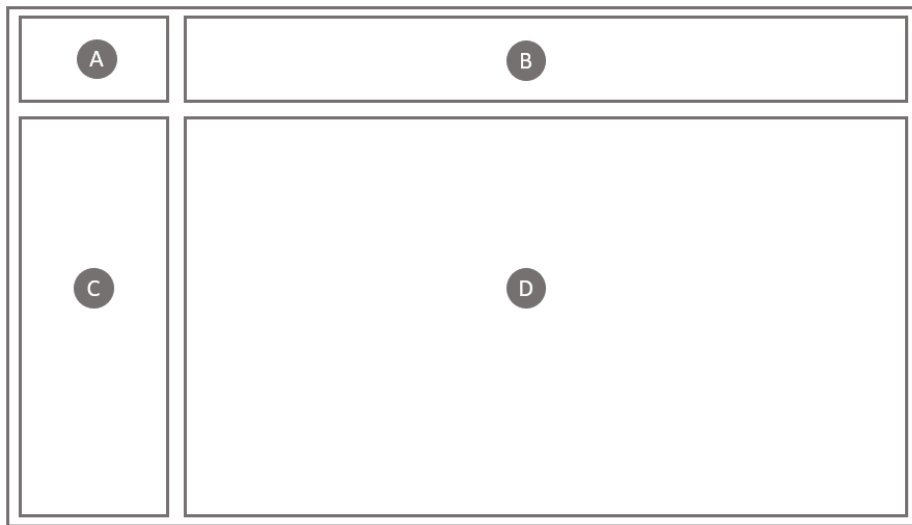


Figure 17 - Wireframe of reports;

- A. Company and loyalty program (when applicable) logotype;
- B. Filters that improve flexibility to the reports;
- C. Key Performance Indicators associated with the report theme. When *B* does not occupy all available space, KPIs can be added at the top instead of at left.
- D. Data Visualization. All the graphics of the report are in this area which can be extended if the report does not have the C section filled.

4.5. METADATA

It is considered a good practice to have a metadata page at the beginning of the reports, this way, anyone that accesses the reports will know where the data came from, some specificities of it and a little summary of what the user can find. Hence, both reports have this metadata page.

In the **Client report**, this page has more information since the report contains more pages and details. The user can click the buttons on the left panel to navigate through the report pages.



Client Report:

- This report has as main goal the profiling of the Pestana guests
- The source of all the data present in this reports is the Client tabular model
- The reservation data is comprehended between 01/01/2010 and the current day
- The report has 5 pages:
 - General**
General profiling of a Pestana guest with several filters available. It has personal information of the client and pre and in stay details.
 - Non PPG vs PPG**
Comparison of a non PPG and a PPG client characteristics.
 - Email Marketing**
Detailed information about the campaigns of the email marketing team, including the campaigns sent, opened and clicked.
 - Resorts Table**
Table with the regions and the resorts as rows and different characteristics as columns.
 - Clients List**
Table that allows filtering according to user's choice, the result is a table that can be downloaded to excel to help filter the clients that are targeted to certain campaigns. Due to GDPR, this page needs access restrictions.
- One can navigate through the report by clicking the bookmarks available in the left panel

Figure 18 – Client Report Metadata;

In the **PPG report**, and considering it only has two pages, the metadata page is simpler and has a button that allows the user to start the navigation.

PPG Report

- This report has as main goal the profiling of the PPG clients (Pestana Priority Guest)
- The source of all the data present in this report is the PPG tabular model
- The reservation data is comprehended between 01/01/2010 and the current day
- The report has 2 pages: one with a more general view about the client and other with details about the card tier (that is only visible to the user by clicking the bookmark).
- The General page has two main focus: points redeemed and consumed. One can reach these views by clicking the buttons at the top, it also has different filters.
- To start analyzing the report, please click here: [→](#)

Figure 19 – PPG Report Metadata;

4.6. CLIENT REPORT

4.6.1. Report Explanation

The **Client report** has 5 pages: the first one has a general view of all clients, the second one is a comparison between a non PPG client and a PPG client, the third page is specific to the email marketing team with a focus on the campaigns and its metrics, the fourth page is a table that profiles the client of each region and resort and the fifth page is meant to be used in order to get a list of clients with certain characteristics which can be chosen according to several filters.

The logic followed to build the first page (**Annex A: Figure 24 - General page of the Client Report;**) of the report was the same as the one to build the PPG report, therefore in the top left corner there is a Pestana Hotel Group logo, at the left panel there are several KPIs and at the top panel there are a set of filters to add flexibility to the analysis.

Then, this page is divided into three invisible pieces. The first set of three charts are about the client characteristics: the age distribution, the top 10 of nationalities and a histogram of the amount of reservations by client. The second set is a “pre-stay” group of charts, it has booking window distribution (how long before the stay a person has booked the room), the market segment at which the client belongs and a length of stay range (how many nights has a person stayed in the hotel). The third set is a “at-stay” group of graphics: a pie chart with the classification of the trip (individual, couple or individual), a waterfall graph with the type of meal of the client and a table showing the most booked room types.

The second page (**Annex A: Figure 25 – Non PPG vs PPG page of the Client Report;**) has the same charts in the left and the right side with the left side regarding to non PPG clients and right side is about PPG clients (the filter is set to “Is PPG” = “No” or “Is PPG” = “Yes”). Also, it has a identification logo to ease the recognition of each side and a card that shows the total quantity of clients.

At the top, on both sides, it has 5 KPIs: the number of clients belonging to this category, the median age, average number of stays, average length of stay (nights per stay) and guest satisfaction. The page has one pie chart regarding the trip type (individual, couple or family). It also has three donut charts: percentage of clients that are a repeater, in other words, clients that stayed more than once (1 - yes or 0 - no), the percentage of reservations that are only weekend (weekends in hospitality are nights of Friday and Saturday) and the reservation type (only weekend or only week). Besides, it has three clustered bar charts (two horizontal and two vertical): number of reservations, age, booking window and length of stay distribution. The other chart that figures in this report page is a tree map with the top 10 nationalities code.

The third page (**Annex A: Figure 26 – Email Marketing page of the Client Report;**) is about the email marketing, as mentioned before. It has two filters that allow to choose a campaign by its id or date and two KPIs that represent the open rate and click-to-open ratio (when no campaign is selected these cards show the average of these metrics) at the top. Then, it has a line and clustered chart which illustrates the amount of sent campaigns as well as a line which represents the open rate. At

the middle of the page, there is a tree map with the campaign types and a table with the top 5 categories of campaigns, furthermore it has a distribution over time line chart that will enable the analysis of the number of campaigns sent by month. At last, it has a word cloud that will be helpful to see the most used words in the campaigns.

The fourth page (**Annex A: Figure 27 - Resorts List page of the Client Report**) contains a table with the region and the resort name in the rows and several KPIs as columns, such as: average number of stays, median age, nights in hotel per year, guest satisfaction survey, percentage of clients who choose only weekends, percentage of clients that get a room upgrade, average booking window, percentage of food and beverages expenses, length of stay and average price of reservation. Besides seeing this values aggregated by hotel, the user can filter the check in date, the segment at which the client belongs, the nationality, if it belongs to the priority guest program or even if the reservation was in the past, is occurring or is in the future.

At last, the fifth page (**Annex A: Figure 28 – Clients List page of the Client Report;**) aims at filtering of the clients list according to the available slicers: if the client is a part of the priority guest program, allowance of communication, does only weekend reservation, age range, last booked date, if it has a future reservation, its market segment, the country where it comes from, quantity of reservations range, the package type (half board, full board, dinner, etcetera) and if it has opened the campaign. Using all these slicers a list with the clients' email will be attained and can then be downloaded into an excel sheet to facilitate the sending of campaigns. This page has GDPR (general data protection regulation) restrictions due to the presence of personal information, therefore the page print screen was covered.

4.6.2. Bookmarks and Buttons

This report has 5 bookmarks each one corresponding to a page. This is used to allow the user to navigate through the metadata page.

Each created button (in this case, each button is an image) was set to change the page to the one required. This is done through the action of the image setting it to a bookmark and then select the correspondent.

4.6.3. Tooltips

Tooltips are used to contextualize the user about report details. They are a way to deepen the knowledge of the user without the need to be visually exhausting by adding text or visuals to an already full-page report.

In this report there are two different types of tooltip, one of them intends to educate the user about what it was presumed to create a certain chart and the other aims to show more information about specific measure without being too confuse.

The first type was used to share more details about the Trip Type chart, in other words, this is an information that is assumed and therefore there was a need to explain it. To add this information to the chart, a simple measure having a question mark was done. This will allow to turn on the setting of the tooltip, select “Report Page” and then select the “Trip Type Info” page that was created having the details about the assumption made.

When hovering the mouse above the question mark, the report page created appears showing the details mentioned above.

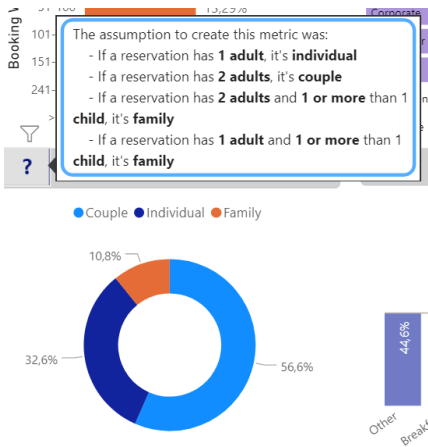


Figure 20 – Trip Type Tooltip;

The other type of tooltip created is related to the DAX table called “View” and that has the objective to show a measure according to different views. Since the measures will have three different values (one that is filtered by every table, one that is filtered by every table except the date and resort tables and the other that is filtered by everything except the date) the decision fell on showing always the one that is filtered and to create a tooltip that shows the other two. Once again a simple report-level measure containing a question mark was created and then added to the KPI corresponding to the measure.

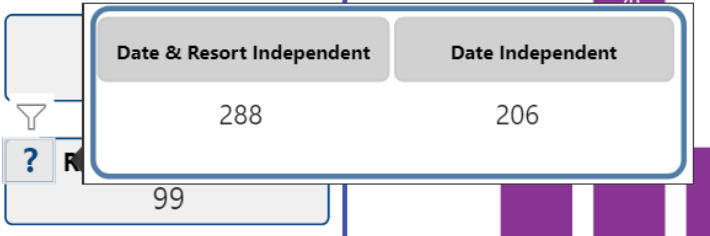


Figure 21 – Repeaters Tooltip;

The measure tooltip was set and the result can be seen by hovering the mouse. The use of the tooltip in this case permitted to have the information without filling even more the report.

4.6.4. Mobile View

Client report has a mobile view available. However, not all the pages in this report will be available in the app, since in my perspective it does not make sense to create this type of view when the page is constituted by a table with several columns. These tables would be hard to see and analyze in the phone which is exactly the opposite point of making a mobile view.

Therefore, the Client report phone layout will have four pages including the metadata. The metadata page will have three buttons that allow navigation through the report. The other pages available are: General, Non PPG vs PPG Clients and Email Marketing. All the filters available in the desktop view are displayed at the top of the report as well as some of the most important KPIs since maintaining all of them in this perspective will get too confusing.

A representation of each page of the report mobile view can be seen below, nevertheless it should be pointed out that all the pages contain more information as they can be scrolled down or up. Not all the pages are at the top, as said before, the top of the page contains always the page identification and filters.

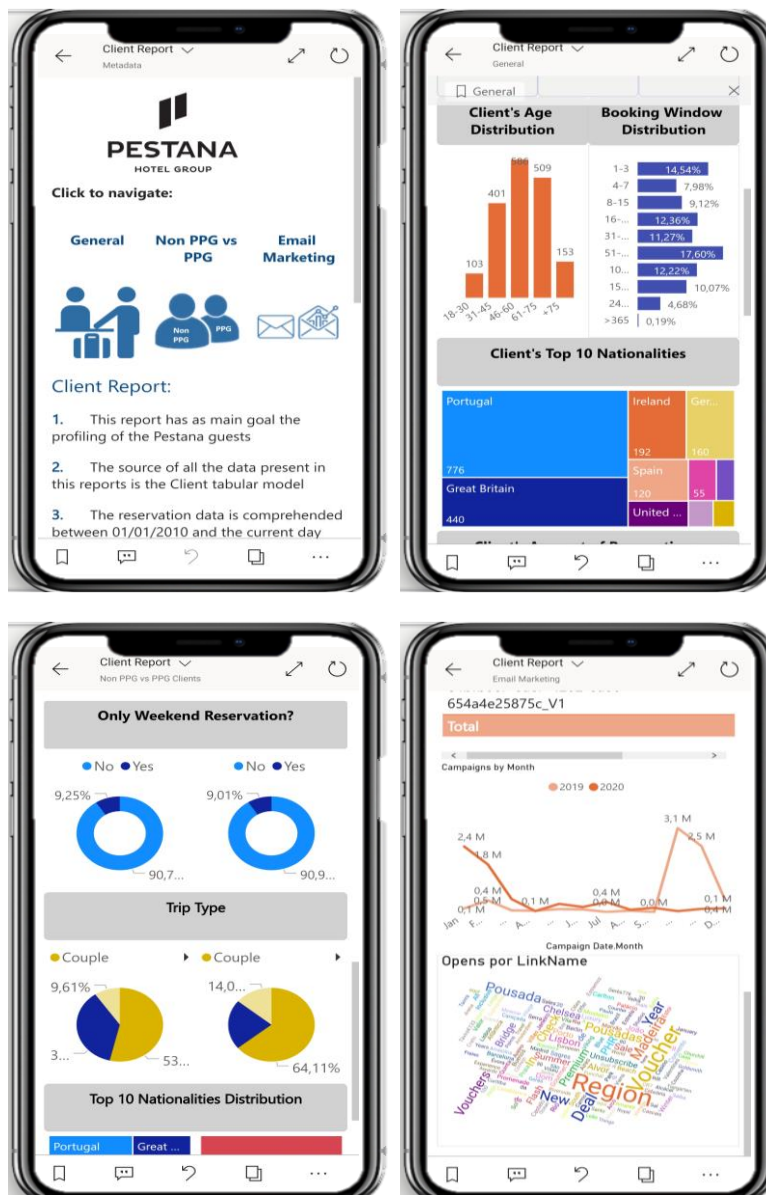


Figure 22 – Phone Layout example of each page of Client Report;

4.7. PPG REPORT

4.7.1. Report Explanation

The **PPG focused report** has 2 pages: one with a more general view about the client and other with details about the card tier (that is only visible to the user by clicking the bookmark).

The General page (**Annex B: Figure 29 - General page of the PPG Report;**) has in the top left corner, the company logo and the loyalty program logo frame the first page of the report. In the left side, there is a panel with the main key performance indicators (KPI), and in the top there is a panel with two buttons which allow to navigate through a consumption or redemption views. Besides, it has a few filters to facilitate the analysis.

There are 4 graphics which characterize the client: gender distribution, allowance of Pestana communication, age distribution and top 10 of nationalities. In addition, a funnel chart portray the number of clients in each card tier, this graphic has a bookmark button at the bottom which permits the navigation to the tiers page. To facilitate the navigation back to the general page of the report, a “back” button was inserted.

Also, there are 3 graphics that change accordingly to the view chosen: a bar chart showing the amount of points consumed or redeemed in each brand and the comparison to the previous year, a table with the top 10 units in which the clients consume or redeem more points and a line chart that illustrates the distribution of points over time. The x-axis in this last graphic is referring to the months and an extra year filter was added to improve flexibility. At the bottom of the page there are two buttons that allow the analyst to navigate through the distribution of expired points or the quantity of new clients. This page is very interactive and has several buttons that lead to bookmarks allowing a variety of analysis in the same page by changing the selected bookmarks.

The Tiers page (**Annex B: Figure 30 - Tiers page of the PPG Report;**) objective is profiling clients of each tier. Once again, the frame is done by both Pestana and PPG logos. At the top there are 6 buttons so the analyst can choose the tier, this choice will cause the change in all the indicators since the page will be filtered according to the tier. At the left panel, there is a box that shows how many customers each tier has and an image that differs corresponding to the age range presented. A card visual was used to show the average number of stays, average points per event, average money in euros per event and the average nights in hotel per year. Besides, a top 5 nationalities table and two pie charts show the gender and the preference of weekend distribution.

4.7.2. Dynamically Display of Images

Besides, the Tiers page have an image that changes dynamically accordingly to the age range where the tier clients are most represented. This dynamism is possible because of the DAX table “Images_table” and the two measures created at report level. The table has a row for each age range and a corresponding link to an image that is available online, that way it is possible to visualize the most represented age range in a more visual way.

The first measure was done to get a top 1 of the age ranges:

```
Top1 = TOPN(1;ALLSELECTED(Client[Age Range]);CALCULATE(COUNTA(Client[Age Range])))
```

After getting this top, the next step was to associate each image to the age range that is the top 1 through another measure:

```
URL_Final = FIRSTNONBLANK(SELECTCOLUMNS(FILTER( Images_Table; Images_Table [Age Range] = [Top1]);  
"URL"; Images_Table[URL]);0)
```

This last measure combined with a “Image” visual, allows the dynamical display of the images.

4.7.3. Bookmarks and Buttons

According to Microsoft, “Bookmarks capture the currently configured view of a report page, including filters, slicers, and the state of visuals.”

The **PPG report** has 11 buttons: General, Consumption, Redemption, Expiring Points, New Clients and one for each card tier: Diamond, Platinum, Platinum Corporate, Gold, Gold Corporate and Silver.

Excluding General button, which is used in the metadata, the first four buttons are linked to bookmarks that only change some selected visuals, namely, consumption and redemption switches accordingly the “Value by Brand”, “Top 10 Units” and the “Distribution Over Time”. The expiring points and new clients are also only linked to these two specific charts since it changes to a line graphic if “Expiring Points” is selected or to a bar chart if “New Clients” is selected.

In the tier page, the bookmarks are linked to all the data in that page and a filter is applied according to the user choice.

4.7.4. Mobile View

To allow an even bigger flexibility, the report was built not only in a desktop view but also in a mobile view which permits the access to the reports in the Power BI app.

The mobile view also allows the navigation through bookmarks and the visualization of all the graphics, however it is restricted to the screen side and therefore some charts may not be perfectly visualized.

Every page of the report was optimized into the mobile view. The metadata page has an arrow that enables the user to start the navigation. On the report pages, the rationale followed was to maintain the filters at the top and the graphics in the order they appear at the desktop view as the user scrolls down. Also, in the Tiers page the user is also requested to select a tier since the charts only appear when this is selected. The figure below depicts an example of each page as they appear in the Power BI app.

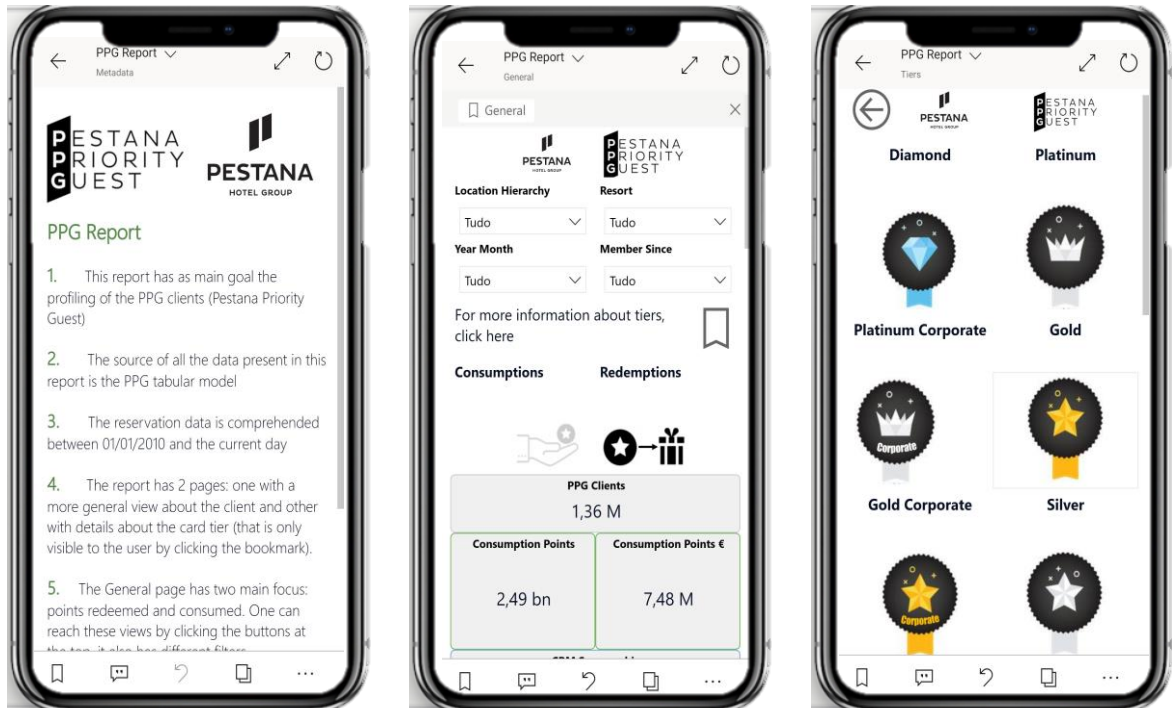


Figure 23 – Example of each PPG report page in mobile view;

4.8. FINAL CONSIDERATIONS

The intent when constructing these two reports was to have all the information required in the same place, however, as it happens in any project, not all the initial requirements were met even though the great majority were.

For example, one of the requirements asked by the CRM team was to have in Card table, the hotel where the adhesion took place and/ or the source system (knowing that the person can join online). However, this source system information will need to be a step subsequent to this project because this information was not a part of any of the data sources of this project requiring a new data integration. Also, one of the intentions was to be able to see, for example, the gender distribution of all the clients which was only possible in the PPG report since that information was not filled when the customer is not a part of the loyalty program.

With the more intense usage of the reports, some other details will presumably be found, however, some tests were made to see the interactivity and easiness of use and the reports seem to have a high response to interactions, several analysis could be made with the data available in the reports, also some important insights could be gathered.

5. CONCLUSIONS

This project was focused on addressing the lack of a client database in a single and recognized-by-all location in Pestana Hotel Group. As mentioned throughout this report, a customer database can be considered a strategic weapon as the knowledge of the audience a company reaches is a main asset in all industries but especially in hospitality.

As stated, the hospitality type audience is quite different than the one in industries such as retail, once the clients' purchases (hotel nights) are not as frequent. Thus, and considering the broad number of existing hotels, PHG hotels must provide "the time of your life" during a stay. That is the only way to influence the client to coming back and the main reason why customer knowledge is so important.

If a client receives focused promotions, goes to a hotel and feels that employees know his tastes, then the probability of coming back grows exponentially which consequently brings more revenue and reputation to the Pestana Hotel Group hotels. A happy customer, besides returning to the hotel, will more probably free advertise which provides Pestana an obvious leverage over the competition.

To frame this conclusion, a little recap will be made about the main objectives of this project as well as the final products achieved. The main objectives were the construction of a customer database along with an interface which allows the non-technical user to perform analysis and navigate throughout the database.

The first step of this project was to select the type of information that should be a part of the database. Furthermore, there was a need to analyze the sources of information in order to ascertain if some information was already being used, what was new to this project and the extra sources of it. After gathering the requirements needed to figure in this database the conclusion was that the majority of the information necessary was already a part of the enterprise data warehouse and some information needed to come from other sources.

Even though the reservation factual table already existed (there is a multidimensional cube focused on the reservations), the data needed to be aggregated into a row for reservation and all the characteristics associated with it combined. Besides, the campaigns details were new to the data warehouse and therefore the new data source needed to be incorporated. In addition to these prior stages, a much more extensive one was performed: the grouping of clients. As said in the beginning of the chapter about the tabular modeling, there were different client records that are in reality the same person which can lead to wrong conclusions so the step was to create a master id which assembles all the information into a single client.

Afterwards, the decision was to construct two tabular models. Even though the PPG model contains some of the tables existing in the Client model, the relationships between the tables are different. The PPG model is focused on the Pestana loyalty program, the program card, the events related to points and the clients that are a part of the program. The Client model is about all Pestana customers, not only the ones in the loyalty program, its characteristics and behavior. The tabular models were constructed using SSAS which required using DAX to build the several measures and calculated columns added to the original tables. These measures and columns extend the flexibility of the models by adding new perspectives and allowing different type of analysis.

Subsequently, and after the two models were deployed to the production environment, the necessity was to create an easy interface to the user. Therefore, the choice was creating two Power BI reports, these reports allow great adaptability, have great interactivity and enable a variety of analysis.

The PPG report has two pages related to points consumption, redemption and card tiers. The Client report is more complete since the model contains more data. It has five pages: a general view of the clients characteristics, a comparison between a client which is a part of the loyalty program and a client that is not a part of it, an email marketing dedicated page that contains details about the campaigns and two pages containing tables, one covering client characteristics of each resort and the other one having several filters that limit the list of clients that results of the filtering.

Thus, there are four main results of this project:

- PPG Tabular Model designed in SQL Server Analysis Services;
- PPG focused Power BI report;
- Client Tabular Model designed in SQL Server Analysis Services;
- Client focused Power BI report.

Most of the information required to be a part of the models is present but future work should include some more details.

The results intend to, besides creating a single source where all the customers data is located, to help the teams communicate to clients in a more efficient and customized way. Hereafter, the goal is to take advantage of this database and discover strategies to enrich it since with more information available, it will be possible to fine tune client related strategies and get more customer dedicated.

Some of the applications this project will have on Pestana Hotel Group are:

- Marketing: personalize and enhance campaign efficiency;
- Operations: Better understanding of customers preferences, as well as a tighter connection with the customers;
- Executives/ Leadership centers: Knowledge of clients characteristics can help making decisions related to value proposition and positioning of a given hotel

Besides, in the special context of the COVID-19 that the world is living on, the customer database will be helpful when deciding what hotels to reopen, since the knowledge of information such as nationality, repeating of stays and email can be decisive to communicate campaigns to customers.

The bigger expectation is that, some time from now, this client knowledge will enable Pestana Hotel Group to gain advantage over competition and grow.

6. LIMITATIONS AND RECOMMENDATIONS FOR FUTURE WORK

The central limitations found throughout this project were associated with the data quality. As previously mentioned, personal data information is very difficult to obtain since it depends on the customer will to share it with the hotels or at the time of registration in the PPG.

Likewise, the check-in in the hotels can be done by different family members which becomes impossible to link to previous reservations made by a different member. In addition, other problem found is the information given when filling the forms, since there are some fields like birth date filled with "01/01/1900" or email fields with "1234@gmail.com" causing that even when there is data completion, this data is wrong and can lead to erroneous conclusions. Moreover, there is always a problem of typos or handwriting legibility when the form is filled during the check in. Another main challenge is the data validity once every time a customer changes their address, changes their name (for example in a marriage situation), modifies its email address or alters another personal information, the already gathered data becomes outdated.

For future work, my recommendations fall into reinforcing the efforts already made by Pestana Hotel Group regarding the intensified attention when filling the customers data whether it is by installing identification readers which will allow to have birth date and names correct or to set rules in the loyalty program adhesion form to get more valid fields. Also, another step in the future work is to migrate these two models into the Azure cloud following the migration currently happening to other models and cubes.

7. BIBLIOGRAPHY

- (2016, December 5). Retrieved from UK Essays: <https://www.ukessays.com/essays/information-technology/the-importance-of-having-a-customer-database-information-technology-essay.php>
- (2019, November 17). Retrieved from Analytiks: <https://www.analytiks.co/post/why-business-intelligence-is-important/>
- (2019, February 18). Retrieved from Strategic Marketing: <https://thinkstrategic.com/five-benefits-of-a-customer-database/>
- (2020, March 10). Retrieved from Stirista: <https://www.stirista.com/blogs/benefits-of-database-marketing>
- (2020). Retrieved from The Art of Service: <https://theartofservice.com/the-importance-of-getting-customer-database.html>
- Agrawal, S., Chaudhuri, S., & Narasayya, V. (2000). Automated Selection of Materialized Views and Indexes for. *Proceedings of the 26th International Conference on Very*.
- Aspin, A. (2018). *Pro Power BI Desktop*. Staffordshire: Apress.
- Bonnie, E. (2019, September 3). Retrieved from Clever Tap: <https://clevertap.com/blog/database-marketing/>
- Burrows, D. (2015). *The Intelligent Business and Business Intelligence (White Paper)*.
- Butscher, S. (2002). *Customer Loyalty Programmes and Clubs - Second Edition*. England: Gower.
- Colavito, J. (2017, October 27). Retrieved from Smart Data Collective: <https://www.smartdatacollective.com/importance-business-intelligence-organizations/>
- Computer Business Review*. (n.d.). Retrieved from What is SQL Server?: <https://www.cbronline.com/what-is/what-is-sql-server-4914415/>
- DataFlairTeam. (2019, September 5). *DataFlair*. Retrieved from "8 Important Power BI Advantages and Disadvantages – Must Learn For 2019.": <https://data-flair.training/blogs/power-bi-advantages-and-disadvantages/>
- Desk, C. (2016, March 2). Retrieved from IT Toolbox: <https://it.toolbox.com/blogs/johndoe/crm-and-the-role-of-business-intelligence-030216>
- Element 61*. (2012, May 26). Retrieved from "The choice between Tabular or Multidimensional models in SQL Server Analysis Services 2012": <https://www.element61.be/en/resource/choice-between-tabular-or-multidimensional-models-sql-server-analysis-services-2012>
- Gupta, A., & Mumick, I. S. (2011). Maintenance of Materialized Views: Problems, Techniques, and Applications. *Data Engineering Bulletin*.

- Hevner, A., & Chatterjee, S. (2010). *Design Research in Information Systems*. New York: Springer.
- Horvath, I. (2007). Comparison of Three Methodology Approaches of Design Research. *International Conference on Engineering Design*.
- Howson, C. (2008). *Successful Business Intelligence: Secrets to Making BI a Killer App*. USA: The McGraw-Hill Companies.
- Husemann, B., Lechtenborger, J., & Vossen, G. (2000). Conceptual Data Warehouse Design. *International Workshop on Design and Management of Data Warehouses*.
- Iivry, J., & Venable, J. (2009). Action research and design science research - Seemingly similar but decisively dissimilar. *ECIS 2009 Proceedings 73*.
- Imhoff, C., & White, C. (2011). *Self-Service Business Intelligence. Empowering Users to Generate Insights*. TDWI Research.
- Imhoff, C., Gallemmo, N., & Geiger, J. G. (2003). *Mastering Data Warehouse Design. Relational and Dimensional Techniques*. Indianapolis: Wiley Publishing, Inc.
- Inmon, W. (2002). *Building the Data Warehouse. Third Edition*. USA: Wiley Computer Publishing.
- Kakish, K., & Kraft, T. A. (2012). ETL Evolution for Real-Time Data Warehousing. *Conference on Information Systems Applied Research*. New Orleans Louisiana.
- Kimball, R., & Ross, M. (2013). *The Data Warehouse Toolkit. Third Edition*. Indianapolis, Indiana: by John Wiley & Sons.
- Lath, A. (2018, May 3). Retrieved from BBN Times: <https://www.bbntimes.com/companies/the-importance-of-business-intelligence>
- Malik, S. (2005). *Enterprise Dashboards. Design and Best Practices for IT*. Hoboken, New Jersey: John Wiley & Sons, Inc.
- Microsoft. (2019). *Microsoft Docs*. Retrieved from Star Schema: <https://docs.microsoft.com/en-us/power-bi/guidance/star-schema>
- Microsoft. (2019). *Power BI*. Retrieved from <https://powerbi.microsoft.com/en-us/what-is-power-bi/>
- Microsoft. (2020). *Info Microsoft*. Retrieved from <https://info.microsoft.com/ww-landing-2020-gartner-magic-quadrant-for-analytics-and-business-intelligence.html?LCID=EN-US>
- Microsoft. (2020). *Microsoft Docs*. Retrieved from <https://docs.microsoft.com/en-us/analysis-services/tabular-models/impersonation-ssas-tabular?view=asallproducts-allversions>
- Negash, S. (2004). Business Intelligence. *Communications of the Association for Information Systems*, 177-195.
- Pauwels, e. a. (2009). Dashboards as a Service: Why, What, How, and What Research Is Needed? *Journal of Service Research*, 175-189.

- Pello, R. (2018, October 31). *Medium*. Retrieved from <https://medium.com/@pello/design-science-research-a-summary-bb538a40f669>
- Poonnawat, W., & Lehmann, P. (2014). Using self-service business intelligence for learning decision making with business simulation games. *Proceedings of the 6th International Conference on Computer Supported Education*.
- Prabhuswamy, S. (2015). *A Case Study For Data Warehousing Courseware*. Thesis, Computer Science, California State University, Sacramento.
- Rad, R. (2018). *Pro Power BI Architecture*. Meadowbank: Apress.
- Richardson, M. (2018, September 25). Retrieved from Maximizer CRM: <https://www.maximizer.com/7-reasons-why-business-intelligence-is-vital-to-business-success/>
- Rouse, M. (2019, June). *SearchSQLServer*. Retrieved from <https://searchsqlserver.techtarget.com/definition/SQL-Server>
- Russo, M., & Ferrari, A. (2017). In *Tabular Modeling in Microsoft SQL Server Analysis Services*. Redmond, Washington: Microsoft Press.
- Sharda, R., Delen, D., & Turban, E. (2015). *Business Intelligence and Analytics. Systems for Decision Support. Tenth Edition*. Pearson.
- Sirmon, J., Galloway, G., Gross, C., & Gulati, K. (2013). Performance Tuning of Tabular Models in SQL Server 2012 Analysis Services. *Microsoft*.
- Techopedia. (n.d.). *Techopedia*. Retrieved from Structured Query Language (SQL): <https://www.techopedia.com/definition/1245/structured-query-language-sql>
- Techopedia. (n.d.). *Techopedia*. Retrieved from Multidimensional Expressions (MDX): <https://www.techopedia.com/definition/8279/multidimensional-expressions-mdx>
- Trieu, V.-H. (2016). Getting Value from Business Intelligence Systems: A Review and Research Agenda. *Decision Support Systems*.
- Urbanek, S. (2013). *SQL Backend*. Retrieved from <http://cubes.databrewery.org/doc/backends/sql.html>
- Vercellis, C. (2009). *Business Intelligence*. West Sussex, United Kingdom: John Wiley & Sons Ltd.
- Vijalapuram, S. (2018, October 26). *Medium*. Retrieved from Towards Data Science: <https://towardsdatascience.com/introducing-dax-data-analysis-expressions-dca49cb68501>
- Vitt, L., & Cameron, S. (2012). Choosing a Tabular or Multidimensional Modeling Experience in SQL Server 2012 Analysis Services. *Microsoft Business Intelligence Technical Article*.
- Withee, K. (2010). *Microsoft Business Intelligence For Dummies*. Indianapolis, Indiana: Wiley Publishing, Inc.

8. ANNEXES

8.1. ANNEX A - CLIENT REPORT

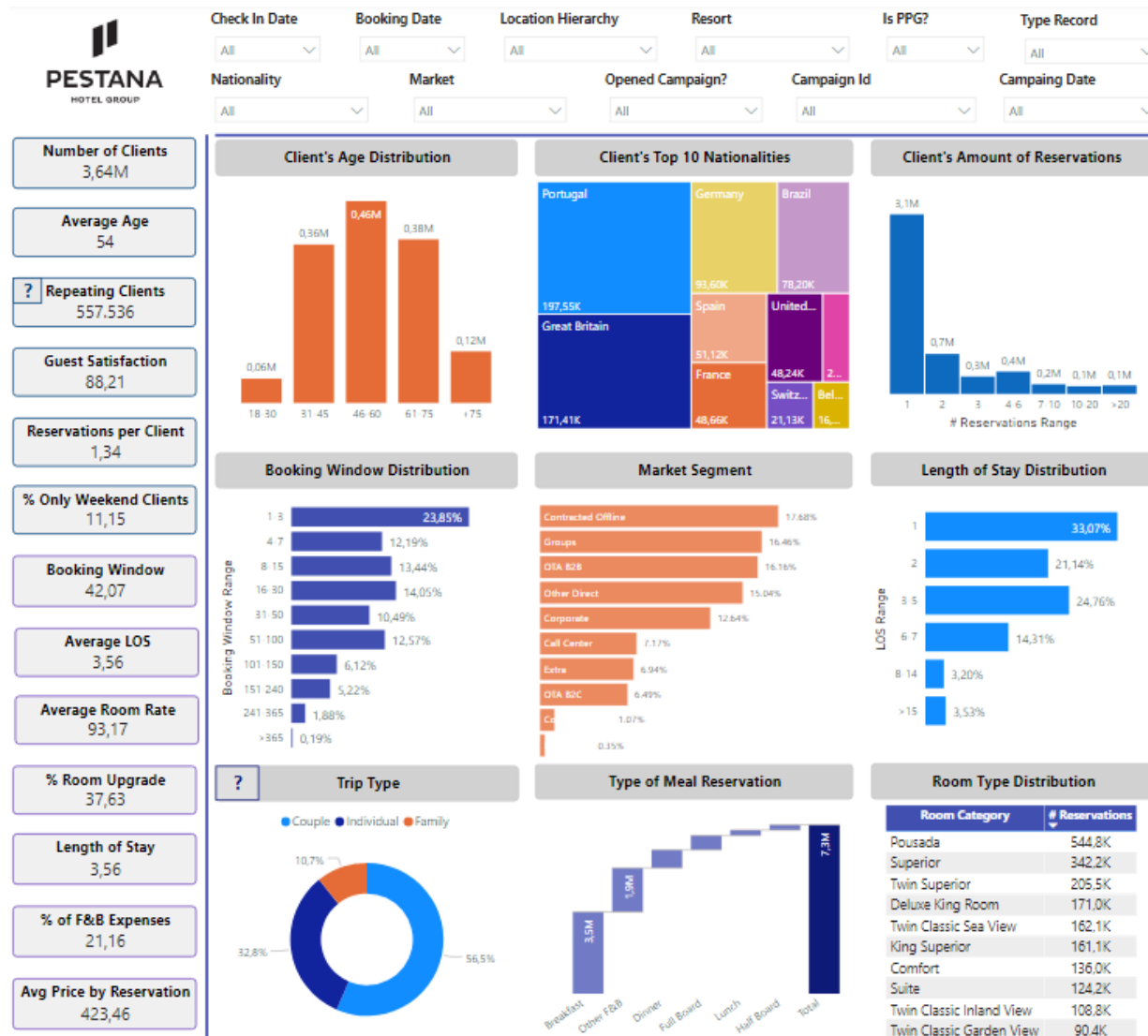


Figure 24 - General page of the Client Report;



Total Number of Clients
3.64M



Non PPG Clients
2.73M

PPG Clients
912.84K

| | |
|---------------------|-----------------------------|
| Median Age 53 | Average Nr of Stays 1.25 |
| Average LOS 3.39 | Guest Satisfaction 86.85 |

| | |
|---------------------|-----------------------------|
| Median Age 56 | Average Nr of Stays 1.63 |
| Average LOS 3.94 | Guest Satisfaction 89.17 |

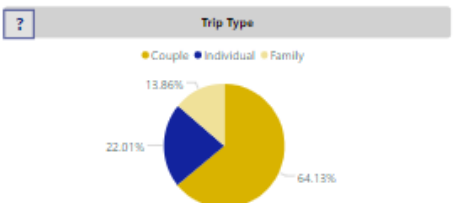
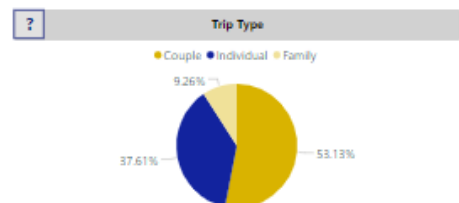
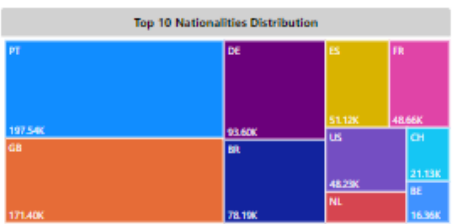
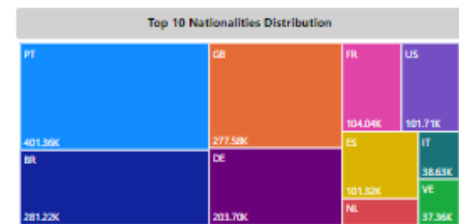
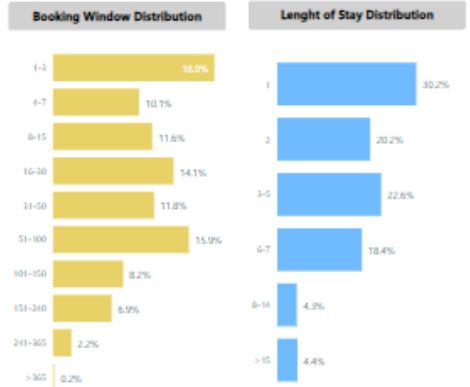
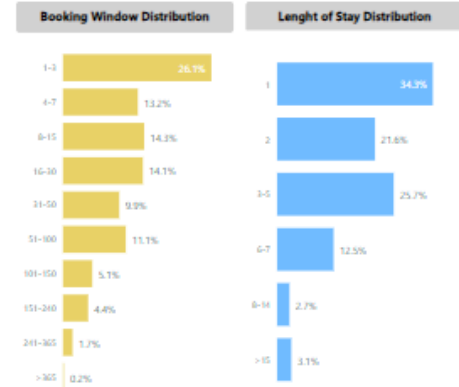
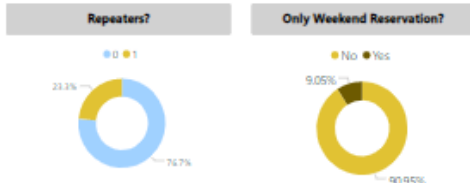
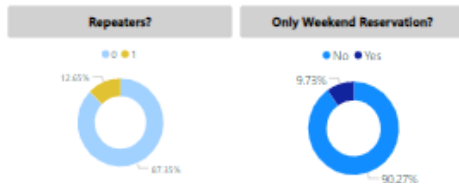
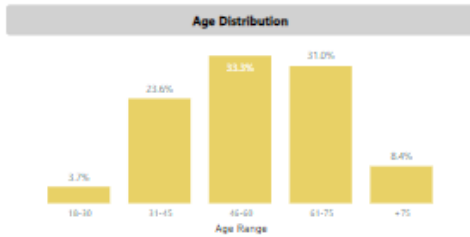
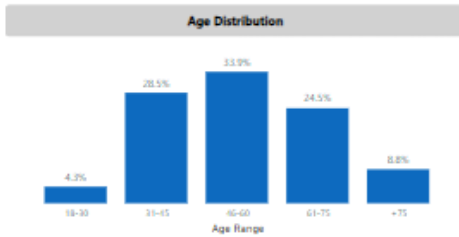
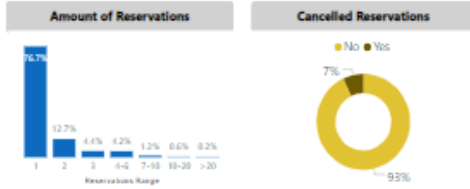
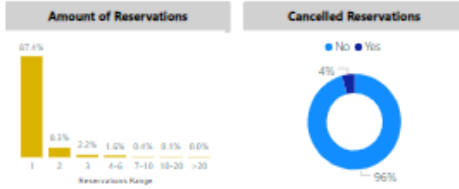


Figure 25 – Non PPG vs PPG page of the Client Report;

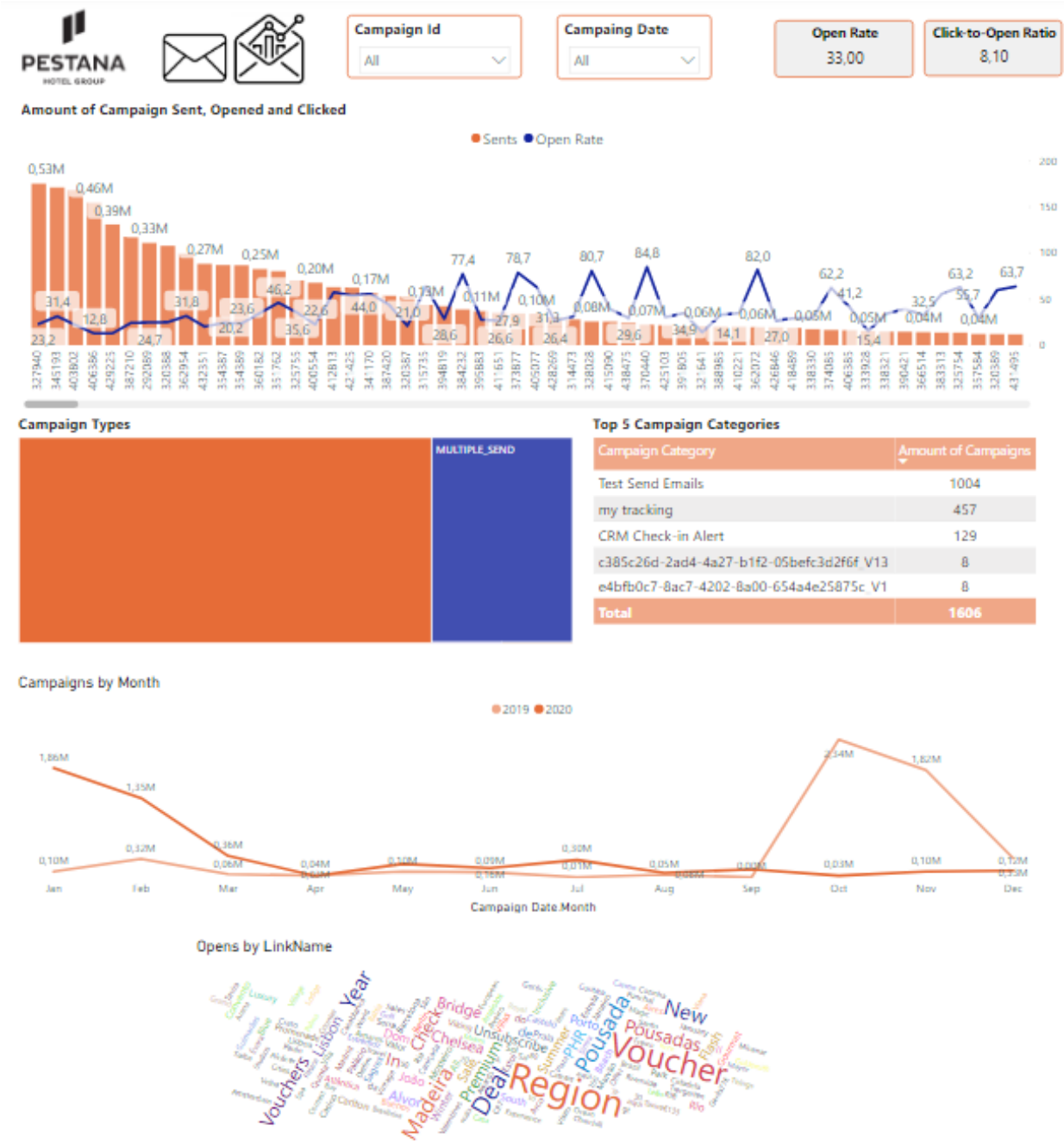


Figure 26 – Email Marketing page of the Client Report;



Check In Date

All ▼

Segment

All ▼

Nationality

All ▼

Is PPG?

All ▼

Type Record

All ▼

| Region | Avg Nr Stays | Median Age | GSS | % Clients Only Weekend | % Clients RoomUpgrade | Avg Booking Window | %FB Expenses | LOS | Avg Price Reservation |
|---------------------------------|--------------|------------|-------------|------------------------|-----------------------|--------------------|--------------|----------|-----------------------|
| Africa | 1.4 | 120 | 82.3 | 10.4 | 46.3 | 25 | 22.5 | 3 | 366.2 |
| Pestana Casablanca | 1.6 | 51 | 83.5 | 11.1 | 38.8 | 11 | 5.8 | 4 | 384.2 |
| Pestana Equador Ilhéu das Rolas | 1.1 | 120 | 78.4 | 15.3 | 22.2 | 37 | 56.2 | 3 | 538.7 |
| Pestana Kruger Lodge | 1.2 | 120 | 82.9 | 12.2 | 17.5 | 43 | 31.8 | 2 | 263.5 |
| Pestana Miramar São Tomé | 1.2 | 120 | 74.3 | 8.0 | 22.5 | 33 | 24.9 | 4 | 367.8 |
| Pestana Rovuma | 1.5 | 58 | 82.0 | 9.5 | 51.5 | 15 | 15.1 | 4 | 300.1 |
| Pestana São Tomé | 1.2 | 120 | 81.3 | 7.9 | 68.7 | 26 | 26.7 | 4 | 523.6 |
| Pestana Trópico | 1.4 | 60 | 88.3 | 8.3 | 73.5 | 24 | 19.0 | 4 | 347.8 |
| Europe | 1.3 | 120 | 89.4 | 11.1 | 36.0 | 48 | 22.6 | 4 | 444.0 |
| Pestana Alvor Atlântico | 1.2 | 120 | | 5.1 | 78.7 | 65 | 3.9 | 9 | 477.5 |
| Pestana Alvor Blue | 1.1 | 120 | 82.5 | 7.4 | 9.5 | 127 | 28.7 | 6 | 1281.1 |
| Pestana Alvor Park | 1.5 | 115 | 93.7 | 5.2 | 33.8 | 164 | 2.2 | 8 | 464.9 |
| Pestana Alvor Praia | 1.2 | 115 | 83.9 | 4.7 | 53.6 | 58 | 22.8 | 6 | 904.3 |
| Pestana Alvor South Beach | 1.1 | 120 | 90.2 | 5.6 | 28.6 | 96 | 16.6 | 6 | 943.0 |
| Pestana Amsterdam Riverside | 1.1 | 120 | 86.4 | 13.4 | 26.8 | 33 | 8.9 | 2 | 495.5 |
| Pestana Arena Barcelona | 1.2 | 120 | 88.5 | 9.2 | 36.7 | 36 | 9.9 | 3 | 271.3 |
| Pestana Bahia Praia | 1.1 | 120 | 89.4 | 7.6 | 50.0 | 67 | 23.3 | 5 | 610.8 |
| Pestana Berlin Tiergarten | 1.2 | 120 | 89.6 | 8.1 | 11.5 | 34 | 11.9 | 2 | 245.2 |
| Pestana Carlton Madeira | 1.1 | 79 | 89.0 | 1.7 | 62.2 | 55 | 24.6 | 7 | 703.6 |
| Pestana Carvoeiro Golf | 1.3 | 120 | | 5.0 | 16.2 | 117 | 3.3 | 6 | 1005.5 |
| Pestana Cascais | 1.1 | 82 | 84.7 | 8.2 | 25.2 | 39 | 20.1 | 5 | 400.4 |
| Pestana Casino Park | 1.2 | 115 | 88.5 | 4.2 | 50.5 | 39 | 25.2 | 5 | 465.5 |
| Pestana Casino Studios | 1.1 | 120 | 93.6 | 7.5 | 8.1 | 34 | 0.0 | 5 | 318.5 |
| Total | 1.3 | 120 | 88.2 | 11.3 | 37.2 | 42 | 21.0 | 4 | 423.3 |

Figure 27 - Resorts List page of the Client Report;



Choose the filters to get a list of clients with these characteristics

Is PPG?

Only Weekend?

Last Booked Date

Market Segment

Qty Reservations

Campaign Id

Allow Pestana Communication?

Age Range

Has OTB Reservation?

Country

Package Type

Opened Campaign?

| GUID | Email |
|---------|-------|
| {97...} | |
| {0E...} | |
| {70...} | |
| {BA...} | |
| {12...} | |
| {88...} | |
| {36...} | |
| {03...} | |
| {84...} | |
| {98...} | |
| {93...} | |
| {D0...} | |
| {3B...} | |
| {F1...} | |
| {78...} | |
| {DC...} | |
| {EC...} | |
| {EE...} | |
| {B1...} | |
| {81...} | |
| {1B...} | |
| {A4...} | |
| {1C...} | |
| {F1...} | |
| {57...} | |

Figure 28 – Clients List page of the Client Report;

8.2. ANNEX B – PPG REPORT



Figure 29 - General page of the PPG Report;



Choose a Card Tier:



Diamond



Platinum



Platinum Corporate



Gold



Gold Corporate



Silver

Number of Customers

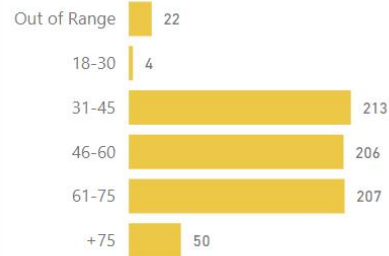
702



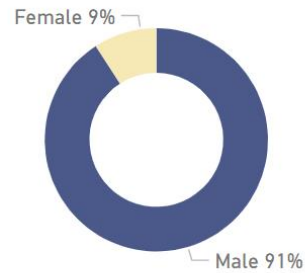
Most Represented Age Range

31-45

Age Range



Gender



Average Number of Stays

2.84

Average Points per Event

985.56

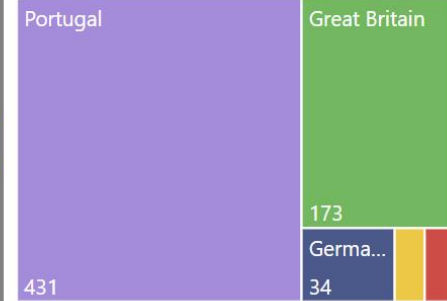
Average € per Event

49.89

Average Nights in Hotel

0.28

Top 5 Nationalities



Only reserves Weekends?

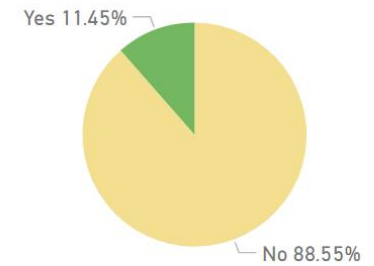


Figure 30 - Tiers page of the PPG Report;

