UTM
UNIVERSITI TEKNOLOGI MALAYSIA

SCHOOL OF COMPUTING
Faculty of Engineering

ALTERNATIVE ASSESMENT

MCSD1013 – 01

BUSINESS INTELLIGENCE AND ANALYTICS

Lecturer's Name: Dr. Nor Erne Nazira Bazin

Student's Name:

Ahmed Hashim Taha Salim (MCS211041)

OLIST Dashboard

TABLE OF CONTENTS

1.0 Executive Summary

- **Data synopsis:** Brazilian E-commerce Public Dataset: Retail datasets of 100k orders placed on Olist spanning between October'2016 and September'2018 across several states. Information is trickled with price, orders, order status, payment, freight and user review along with many other parameters.

1.1 Introduction

Businesses have always tried to keep their customers base engaged and satisfied with the services provided by them. For remaining relevant in the industry, they need to incorporate the latest technological advances into their services. More than a decade back, it was the internet which was completely new and various industries tried to leverage the capabilities of this technology that effortlessly acted as a medium of communication between various businesses and their customers. In this decade, industries have started to provide services that are catered towards each client's individual needs. For such services, they are required to leverage the power of artificial intelligence.

1.2 Business problem

The Olist store is an e-commerce business headquartered in Sao Paulo, Brazil. This firm acts as a single point of contact between various small businesses and the customers who wish to buy their products. Recently, they uploaded a dataset on Kaggle that contains information about 100k orders made at multiple marketplaces between 2016 to 2018. What we purchase on e-commerce websites is affected by the reviews which we read about the product posted on that website. This firm can certainly leverage these reviews to remove those products which consistently receive negative reviews. It could also advertise those items which are popular amongst the customers.

1.3 Formulation of business problem

I am presented a dashboard that summarizes the overall satisfaction of the customer with the product which he or she had just purchased. As well as descriptive analysis and forecasting using Microsoft Power BI tool.

1.4 Organizational structure (Stakeholder)

Olist organizational structure data is layered. The CEO leads the company and is sometimes the chairman or owner. He has the most invested in the company and makes significant decisions, including as analyzing and approving team budgets and assuring subordinates can do their jobs well.

Second is the marketing team; whose role is to market the company's products using cutting-edge technology to compete commercially. The financial team helps the company manage its accounts, income before interest, taxes, depreciation, debt, loans, and cash according to the budget. This group will also provide clients conditional vouchers to increase sales or any marketing budget required to get this project up and running.

Information technology (IT) team is responsible for overseeing every aspect of the company's system, including its hardware and software. This team supports all other teams inside the organization. The e-commerce team, which comes in at number seven, is crucial to the online sales process. Customers will be dealt with personally by these personnel, who will also guarantee their happiness with every transaction There is more department that exist in this company however for our issue and project, these are the most impacted and involved in the project.

1.5 Dataset

I am going to make use of the dataset that was so generously provided by Olist, which is the largest department store in Brazilian marketplaces. Olist eliminates administrative burdens and consolidates legal obligations by linking small enterprises located anywhere in Brazil to various distribution channels. These retailers are allowed to sell their wares through the Olist Store and have Olist's logistics partners transport the products straight to the buyers after the sale has been completed.

The data consist of almost 100 000 customer id and order id. To summarize, the data consist of order detail, customer detail, product detail, seller detail, payment detail, geolocation detail and review detail. The entity-relationship model is shown in Figure 1 to help understand how the data interrelate with each other. [Kaggle]

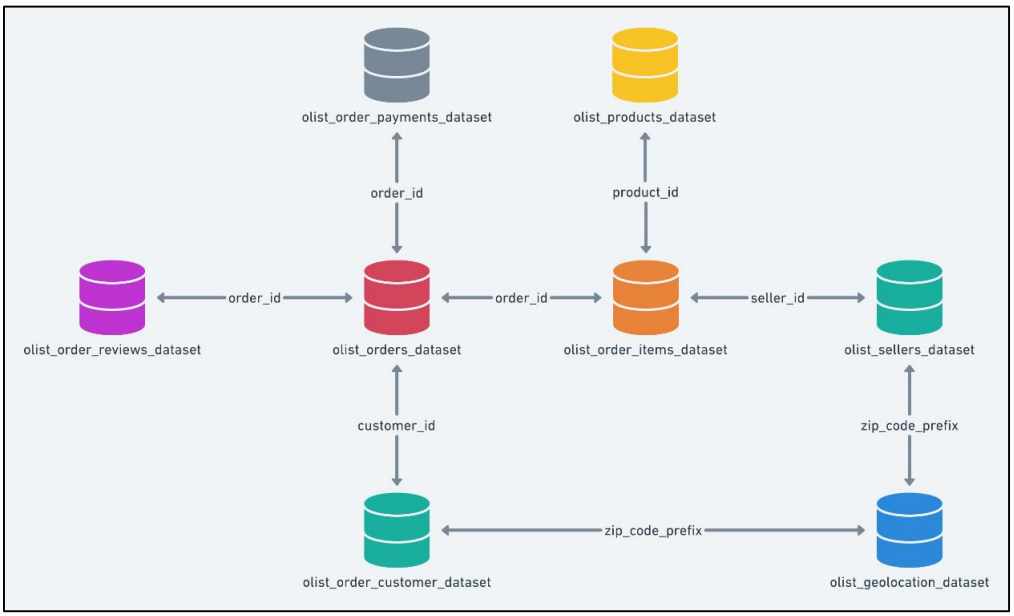

Figure 1 Entitiy-relationship model diagram of Olist Dataset

Table	Description
olist_orders_dataset	This table is connected to 4 other tables. It is used to connect all the details related to an order.
olist_order_items_dataset	It contains the details of an item that had been purchased such as shipping date, price and so on.
olist_order_reviews_dataset	It contains details related to any reviews posted by the customer on a particular product that he had purchased.
olist_products_dataset	It contains related to a product such as the ID, category name and measurements.
olist_order_payments_dataset	The information contained in this table is related to the payment details associated with a particular order.
olist_customers_dataset	Details the customer base information of this firm.
olist_sellers_dataset	This table contains the information related to all the sellers who have registered with this firm.
olist_geolocation_dataset	It contains geographical information related to both the sellers and customers.

Table 1: Descriptions of the tables

2.0 Storyline and choice of visualization

As my manager is asking me to critically analyse the provided datasets using Business Intelligence tools and provide some marketing findings / recommendations in a report format. The dataset has information of 100k orders made at multiple marketplaces in Brazil. Its features allow viewing an order from multiple dimensions: from order status, price, payment and freight performance to customer location, product attributes and finally reviews written by customers. A geolocation dataset that relates Brazilian zip codes to lat/lng coordinates is also integrated in the dataset.

After a customer purchases the product from Olist Store, a seller gets notified to fulfil that order. Once the customer receives the product, or the estimated delivery date is due, the customer gets a satisfaction survey by email where they can give a note for the purchase experience and write down some comments.

Eventually, I decided to choose PowerBI Desktop as my tool for visualization, due to its simplicity and multiple visuals and formatting.

2.1 Objectives of data stories

I. How many customers, orders, and orders per customer does the company have?

II. What is the number of customers by state?

III. What is the number of orders by month?

IV. What are the top and bottom 5 product categories?

V. Visualize the company's customers' demographics, sales trend, orders by categories, orders changes by year

3.0 Visualization Analytics

3.1 Data model

Once I have downloaded the dataset from Kaggle, it is possible to load it into Power BI. The data will have to be pre-processed in order to obtain relevant analytics as it only has the tables and keys referring to each file of the dataset. At first we obtain the data model visible in the following image.

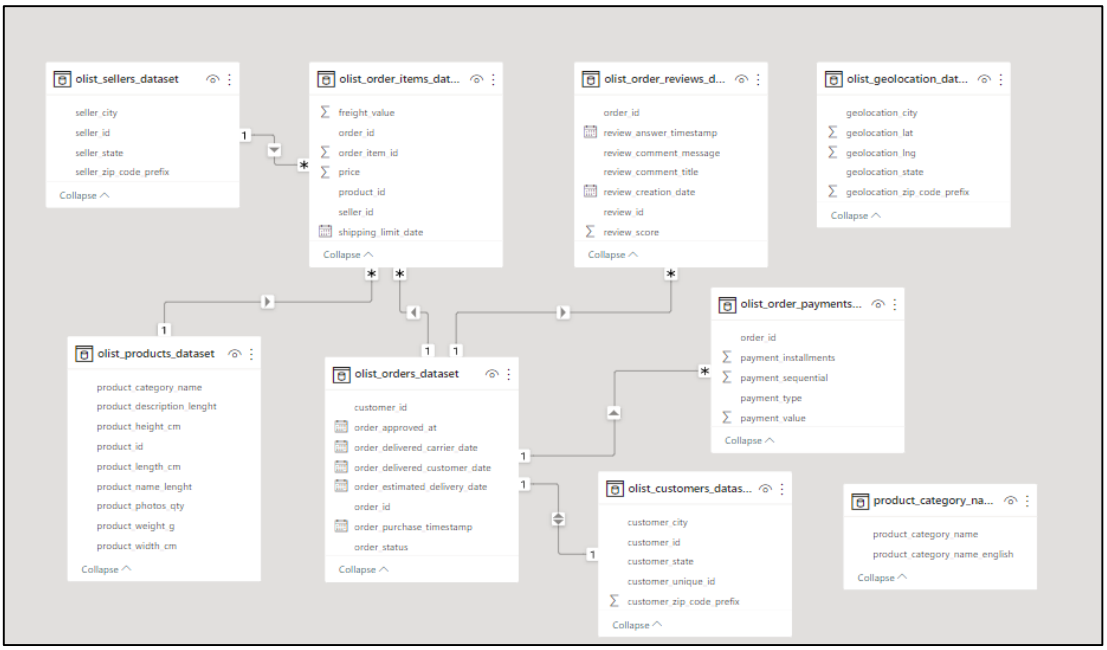

Figure 2: Model Before

Thus, it needs to make links between tables depending on how they interact with each other. The olist_geolocation and product_category tables are not linked to the model at all which means it is not possible to leverage this data. The temporal is also not explicitely indicated as Power BI needs rigorous time management to apply it to the other tables. Moreover, the variables have a type depending on whether they are numerical or textual data. Indeed, we can see that it contains data about transportation logistics, customers, sellers and their different products each with a different type. In the end, we can see the result on the following image.

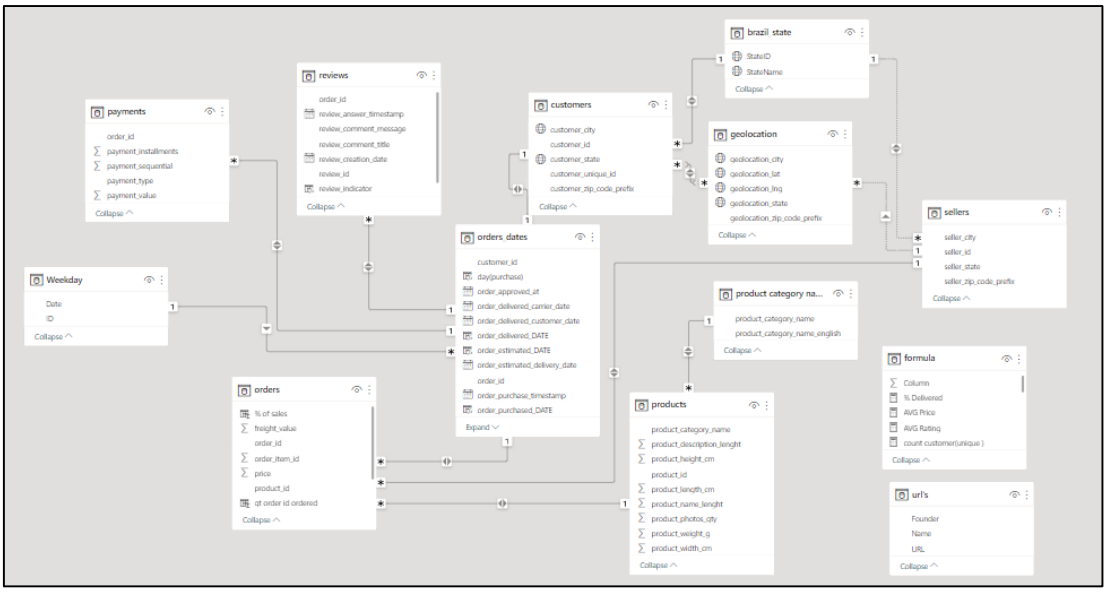

Figure 3: Model After

3.2 Table analysis

I will dive into some of these tables that were added, I have renamed them to make it easier catch and this will obtain as the final data model that allows us to make analytics.

I. Orders Table is at the center of the data model. It represents the most important part of the dataset describing best what can be obtained from the data. Thus, in this case, it about the orders made by a customer buying a specific product. This table contains only primary keys that are necessary to infer knowledge about the data. I have created two columns in this table namely "% of sales" to divide the price of the total price and shows a percentage as a profit ratio, and "qt ordered" where this counts the unique orders.

II. The geolocation table is the Space table represented by geolocation that is now connected to the data model in order to use the information it contains for the two other tables: olist_customer and olist_seller. Indeed, with the olist_geolocation table, we are able to create secondary keys about customers and sellers to obtain knowledge about their postal

code, city or state that were previously impossible to understand. Moreover, we created a hierarchy in order to go from the country and drill down to the state then the city.

III. Orders date have a precise temporality about the data time. Thanks to this table it is possible to make knowledge about temporal data by classifying it chronologically. Thus, in the orders_date table, we are able to tell the difference in days between the orders that allow to group them by day date and obtain indicators such as the deliveries or the purchases. With this, it is again possible to drill down from the year to the quarter, the month and the day. I have created “delivery_days” column which calculate the difference between estimated date and delivered date in days, and delivery indicator to indicates whether it was delivered before estimated date or after. I’ve also created time_day and time_hour to differentiate the time between approved, delivered and orders in days and hours respectively for special visual I have created.

IV. In review table the idea was to create relevant indicators that could be analyzed to obtain insights about the products. Thus, I have created the review_indicator that tells if a comment has been made after a purchase or not. It will give us knowledge on how the customers liked the product or not and if they want to share it to others. Then, the review_sorting indicator sorts the reviews according to the recurrent keywords. It will be possible to know what can describe best the products when customers have similar opinions.

V. Brazil state table shows the state ID and name and modeled in a relationship one-to-many to both sellers and customers tables.

3.3 Dashboard

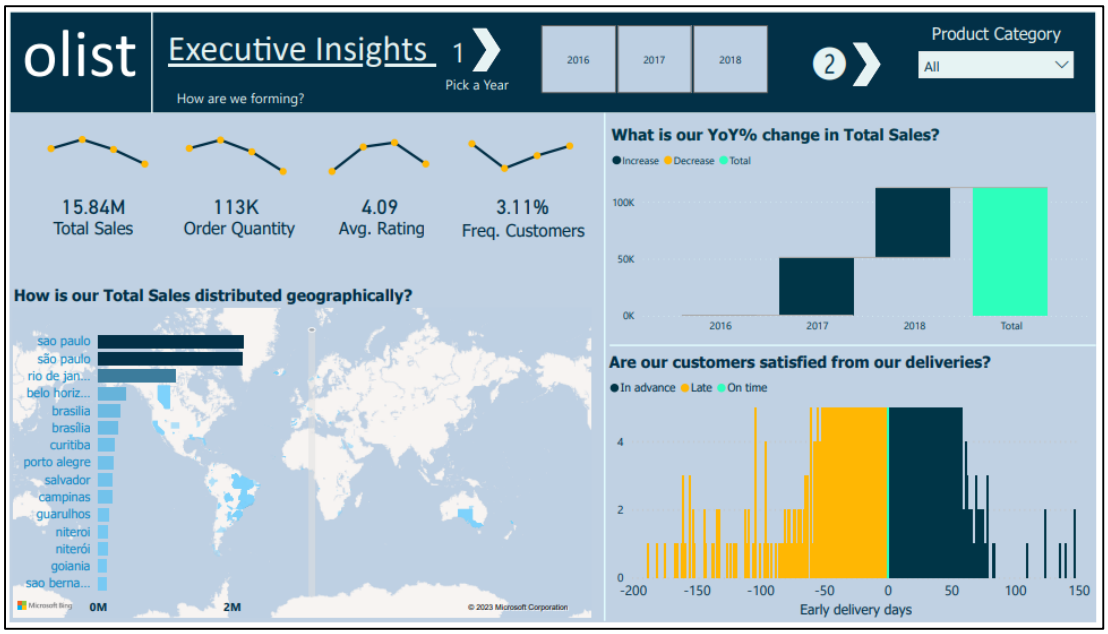

Figure 4: Executive Insights by Decisive Data

Often asking, “How are we performing?” can be a question that cascades into a series of further questions, spinoffs and investigative research. This is especially true for globally minded companies. I wanted to create a report that pre-emptively addressed this kind of exploration. This report is meant to provide data-driven decision making, while emphasizing user-flexibility and visual analysis. Thus, this dashboard can scale as the needs of the global business changes.

The executive Insights page shows that this dashboard is heavily focused on sales and customers, as to fulfil the objective and to achieve customer’s satisfaction and increase sales by discovering insights from the dashboard.

By looking at Figure again. We can conclude that Sao Paulo has the highest sales in all time. Although sales have increased rapidly last three years, the waterfall graph doesn’t show a decrease values.

We can also see that there are a lot of late deliveries as well as earlier deliveries.

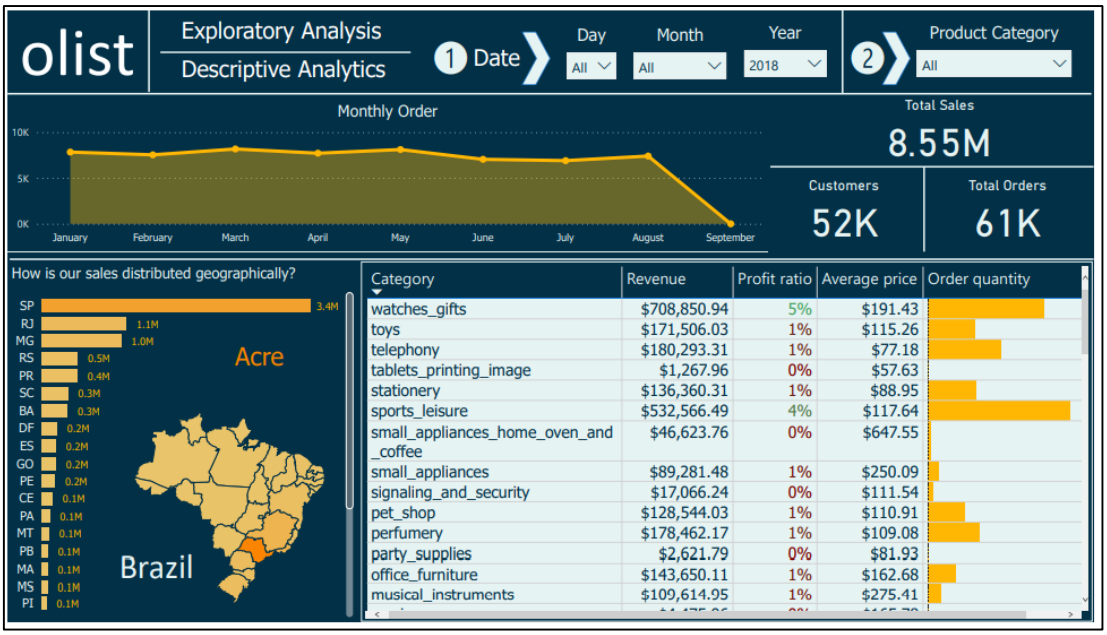

Figure 5: Descriptive Analytics

Next, lets dive deeper into descriptive analytics, we have multiple options we can make use of, for instance to choose a specific day, month or year from the slicer at the top of the page.

The table graph is distributing the product category by the features of their average price, sum of price which is revenue and its ratio, and lastly the number of quantities customer's ordered the specific category. The wining category is health beauty which has been ordered the most, no wonder as females known to spend more on fashion. It shows $772,238 revenue with a profit ratio 6%

The descriptive analysis page shows that we have approximately 16M sales, 94k customers and more than 100k orders.

Figure 6: Customer Investigation

Ever wondered how many customers join our business recently. As customers are our power, with them we can increase revenue and without them we can't maintain growth. This is why customers inveistigation is quite important. And this page a illustrate that among nearly 100k customers.

By looking at the Figure again, we can notice that most of new customers join between May and August throught the three years. The preffered payments methods are by credit-card and boleto paymen.

The top three customers spent more than 21 quantities with approximetly four thousands dollars.

Monday's has the highest order quantities while Sunday's has the highest average price spent. It's quite logical because people tend to spend on weekends.

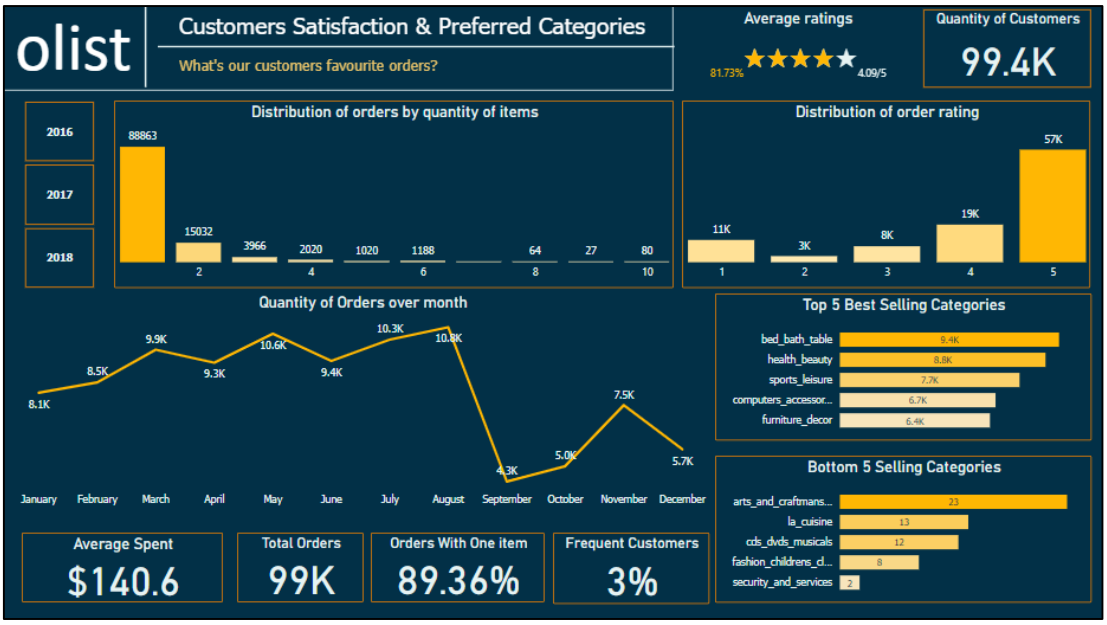

Figure 7: Customer Satisfaction

By looking at the charts we can simplify that the overall average rating looks pretty good showing 4 star from approximately 99.5k customers. While the number of quantity has decreased in December after a slight increase three-months before.

The top five selling categories has approximately 40k orders out of 99k, while the bottom 5 categories didn't exceed 60 orders.

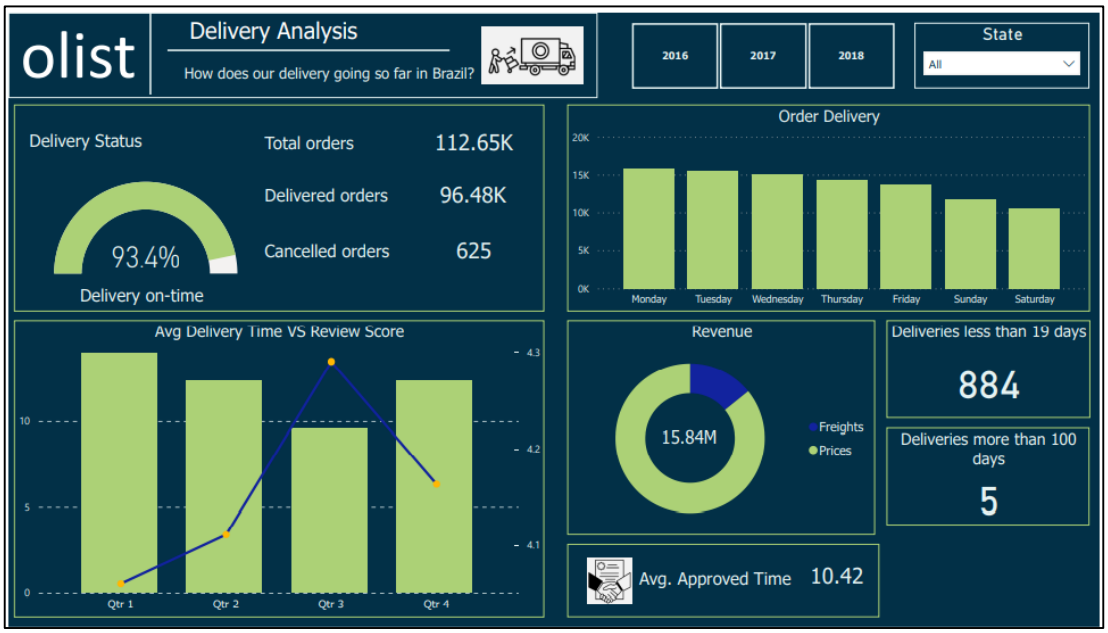

Figure 8: Delivery Days

The delivery page illustrates the processing of delivers from seller's location to customer's desire location, the freights value looks acceptable, there were 96.4 orders delivered out of cancelled orders.

We can notice that average review score is not high at the first quarter, because it took more days to deliver, as opposite in the third quarter higher review for few days. And order delivery days are slightly higher in weekends

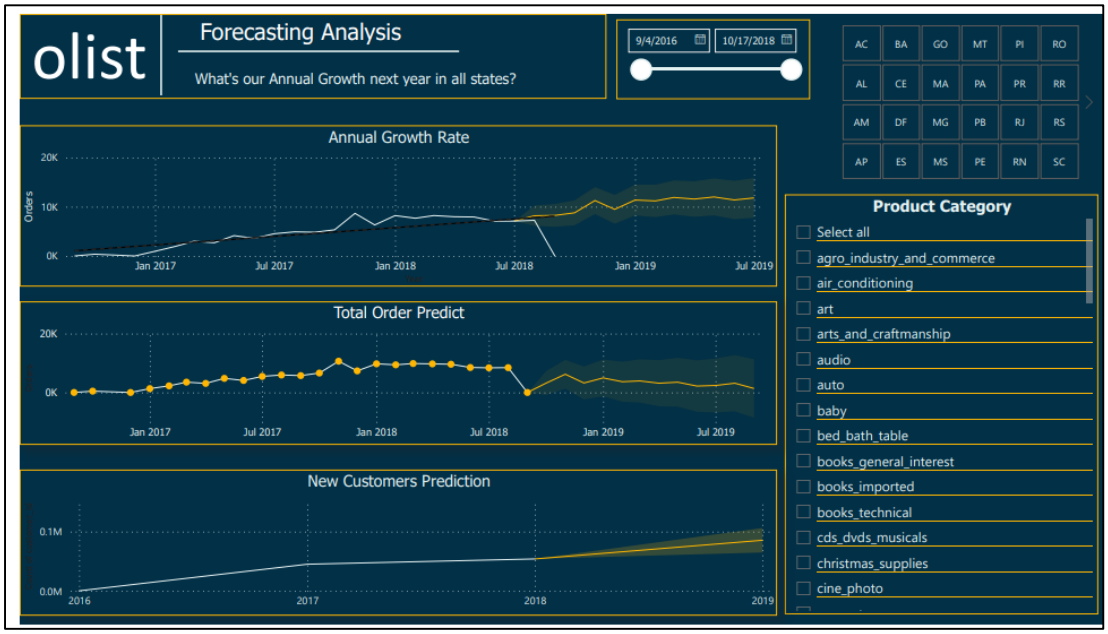

Figure 9: Forecast

What's our annual growth the upcoming years I the question you are looking for, anyway, the annual growth looks great as the orders quantities showed the highest orders quantities score in all three years, as well as new customer prediction showed incensements.

Note that this forecast as in the overall states and time, but it may change regularly.

4.0 INSIGHTS

Olist has a delivery success rate of approximately 85%. This may indicate that the company is facing some challenges with its delivery process, and it may be worthwhile to investigate the causes of the undelivered orders and take steps to improve the delivery success rate. Some potential areas to look into could include the efficiency of the company's logistics and fulfillment processes, the reliability of its transportation partners, and any potential issues with the quality or accuracy of the orders being placed.

Olist company has a high level of customer satisfaction overall, with a significant number of positive reviews and scores. However, the fact that the lowest-rated product category is "Security and Services" suggests that this type of product may need improvement.

5.0 Recommendations

- Monitor and analyze customer reviews regularly to identify trends and areas for improvement. This could involve using data analysis tools to identify common themes in customer feedback and using this information to make changes and improve the customer experience.

- Investigate the causes of undelivered orders. This could involve analyzing the undelivered orders to identify common themes or factors that may be contributing to the problem. For example, are certain regions or customer demographics more likely to have undelivered orders? Are there particular products or types of orders that are more likely to be undelivered?

- Communicate with customers about the delivery process. Olist should be transparent with customers about the delivery process and provide them with regular updates on the status of their orders. This will help

to build trust and create a positive customer experience. It will also give customers the opportunity to provide feedback on their experiences with the delivery process, which can be used to identify areas for improvement.

- Overall, the key is to continue providing high-quality products and services, while also being responsive to customer feedback and working to improve areas that may need attention. Also regularly monitor and review the delivery success rate and communicate with customers about the process.

6.0 Suggestions:

- Special offerings to boost overall sales on low sales period
- Improve low sales category
- Outsourcing drivers for delivery during Sales or Festival period
- Investigate and Review the partner company with low review score and
- analysis customer's comment's provided in the dataset with NLP or any kind of language processing models