Brazilian e-commerce company Olist performance analysis

Technical Report

Introduction

As a collective we set out to complete what we had proposed to Olist in the initial proposal, and using multiple tools in order to observe and visualise our findings:

- Azure Data Studio: Making the use of SQL and queries to retrieve data from CSV files
- Microsoft Excel: Used to perform data cleaning
- Power Query: Also used to perform data cleaning and to transform data
- Power BI: Used to produce visualisations to demonstrate statistical analysis

Review Sentiment and Score Analysis

Feedback sentimental analysis was used to demonstrate the polarity with customer feedback. As a large corporation, it is important that customer satisfaction is maintained. Customer feedback can be a cheap and easily attainable form of market research and a very powerful tool in illustrating Olists' perception in the public eye. Using the Olist dataset, after performing the necessary data cleaning by removing null results. We

created a custom calendar table suitable for the data set we are working with i.e. 2016 to 2018.

To translate the reviews from Portuguese to English, the data was exported to Excel google Docs and a column was created with the title translate. Then this formula was used in the new column:

```
=GOOGLETRANSLATE(cell with text ”source of language”, “target language”)
```

These translated reviews were analysed using Valence Aware Dictionary for Sentiment Reasoning (VADER) which is sensitive to both polarity (positive/negative) and intensity (strength) of emotion.

review_id	order_id	review_score	review_comment_message_translate	polarity scores	Sentiment	review_creation_dat	review_answer_timestamp_con
e64fb393e7b32834bb	658677c97b385a9be1	5	I received well before the	0.27	Neutral	21/04/2017	21/04/2017
f7c4243c7fe1938f181b	8e6bfb81e283fa7e4f1	5	Parking lannister stores I loved buying the internet safe and practical parabaval to all	0.89	Positive	01/03/2018	02/03/2018
8670d52e15e00043ae	b9bf720beb4ab37287	4	Efficient apparatus. On the site the brand of the device is printed as 3Desinfector and when it arrives with another name ... Update with the correct brand since it is the same	0.42	Neutral	22/05/2018	23/05/2018
4b49719c8a200003f70	9d6f15f95d01e79bd1	4	But a little, crashing ... for the value is good.	0.79	Positive	16/02/2018	20/02/2018
3948b09f7c818e2d86c	e51478e7e277a83743	5	Trust seller, product OK and delivery before the	0.74	Positive	23/05/2018	24/05/2018
			I would like to know what happened. I				

The sentiment score of a text was obtained by summing up the intensity of each word in the text. The VADER’s Sentiment Intercity Analyser takes in a string and returned a score in each of this categories (Neutral, Positive, or Negative).

After the data was cleaned in Excel, the scores and labels were added to our data-frame.

```
a = 'Very good high quality!'
sid.polarity_scores(a)
OUTPUT-{'neg': -0.65, 'neu': 0.00, 'pos': 0.56}
```

Text mining

Introduction

Text mining is a sophisticated data analytics technique that involves extracting valuable insights and meaningful information from unstructured text data. This process employs natural language processing (NLP) and machine learning algorithms to analyse vast amounts of text data and identify patterns, sentiments, and topics.

In the case of the Olist database in our project, text mining can provide us with several advantages. By leveraging this technique in our project, businesses can extract sentiments and opinions from customer reviews and identify common themes that customers are discussing. This can enable Olist to better understand the needs and preferences of their customers and use data-driven decisions to improve their products and services.

Moreover, text mining can aid businesses in categorizing customer reviews into different topics, such as delivery times, product quality, and customer service in this project. This categorization can help businesses identify the most common issues that customers face and take proactive steps to address them.

Overall, text mining is a powerful tool in data analytics that can help businesses derive valuable insights from unstructured textual data. In the context of the Olist database, text mining can provide businesses with a deeper understanding of their customers' needs and preferences and enable them to make data-driven decisions to improve their products and services.

Polarity score

VADER (Valence Aware Dictionary and Sentiment Reasoner) is a popular open-source sentiment analysis tool that uses lexicon and rule-based approaches to determine the sentiment of a piece of text. VADER assigns a polarity score to each piece of text, indicating whether the sentiment is positive, negative, or neutral.

VADER's sentiment analysis is based on a predefined lexicon of words and phrases, each of which has a score that reflects the degree of sentiment associated with it. These scores are combined to calculate a sentiment score for a given piece of text.

The polarity score calculated by VADER reflects the degree of positive or negative sentiment in a piece of text. The score ranges from -1 to 1, where -1 indicates extremely

negative sentiment, 0 indicates neutral sentiment, and 1 indicates extremely positive sentiment.

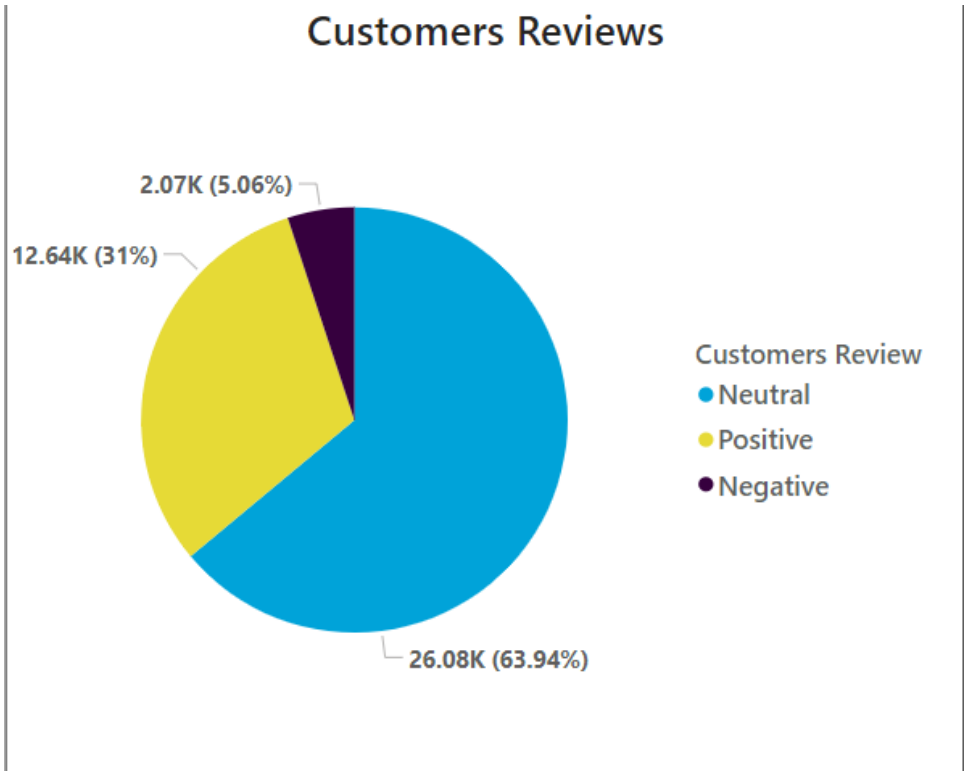
VADER uses a set of predefined rules to handle negation, punctuation, and capitalization, which can affect the polarity score of a text. For example, VADER recognizes that the sentiment of a sentence can be reversed by adding negating words like "not" or "never." VADER can also recognize the difference in sentiment between "I like it" and "I really like it" by considering the intensity of the sentiment.

By using VADER, text mining algorithms can quickly and accurately analyse the sentiment of large volumes of text data. For example, a business may use VADER to analyse customer reviews of their products or services and identify areas for improvement. The polarity score provided by VADER can also be used to monitor the sentiment of a brand or product over time, providing valuable insights into customer satisfaction and sentiment trends.

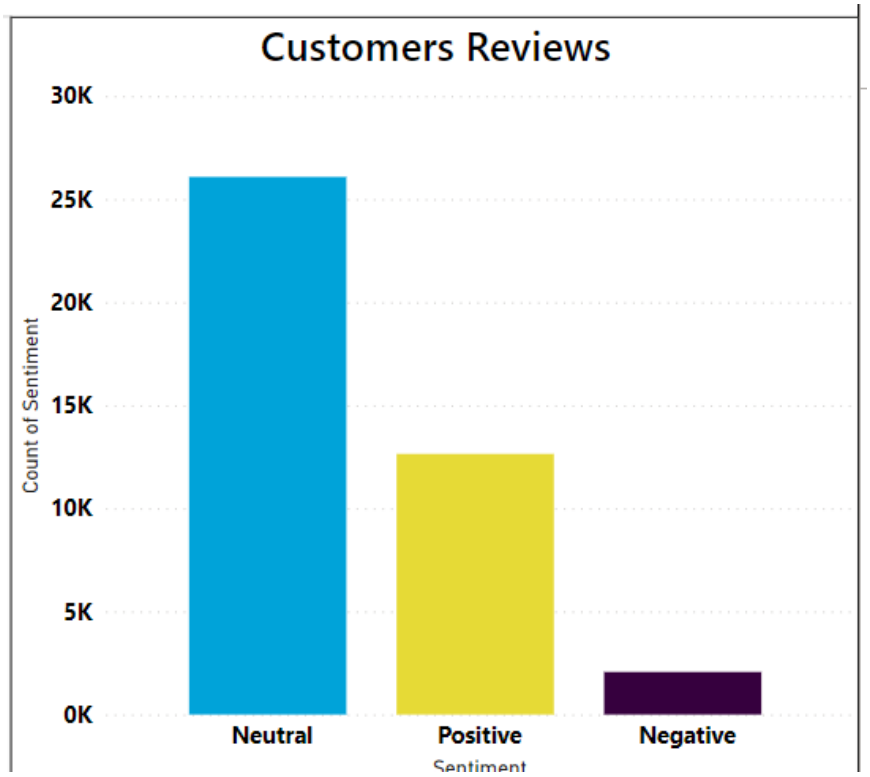
In summary, VADER is a widely used tool in text mining for sentiment analysis. The polarity score provided by VADER reflects the degree of positive or negative sentiment in a piece of text, and can be used to gain valuable insights into customer opinions and brand sentiment in the project of Olist.

Therefore, we use VADER to get the polarity score from the customer comments to analysis to customer satisfaction.

Review score by VADER



Pie chart of customer reviews(Neutral, Positive, Negative)

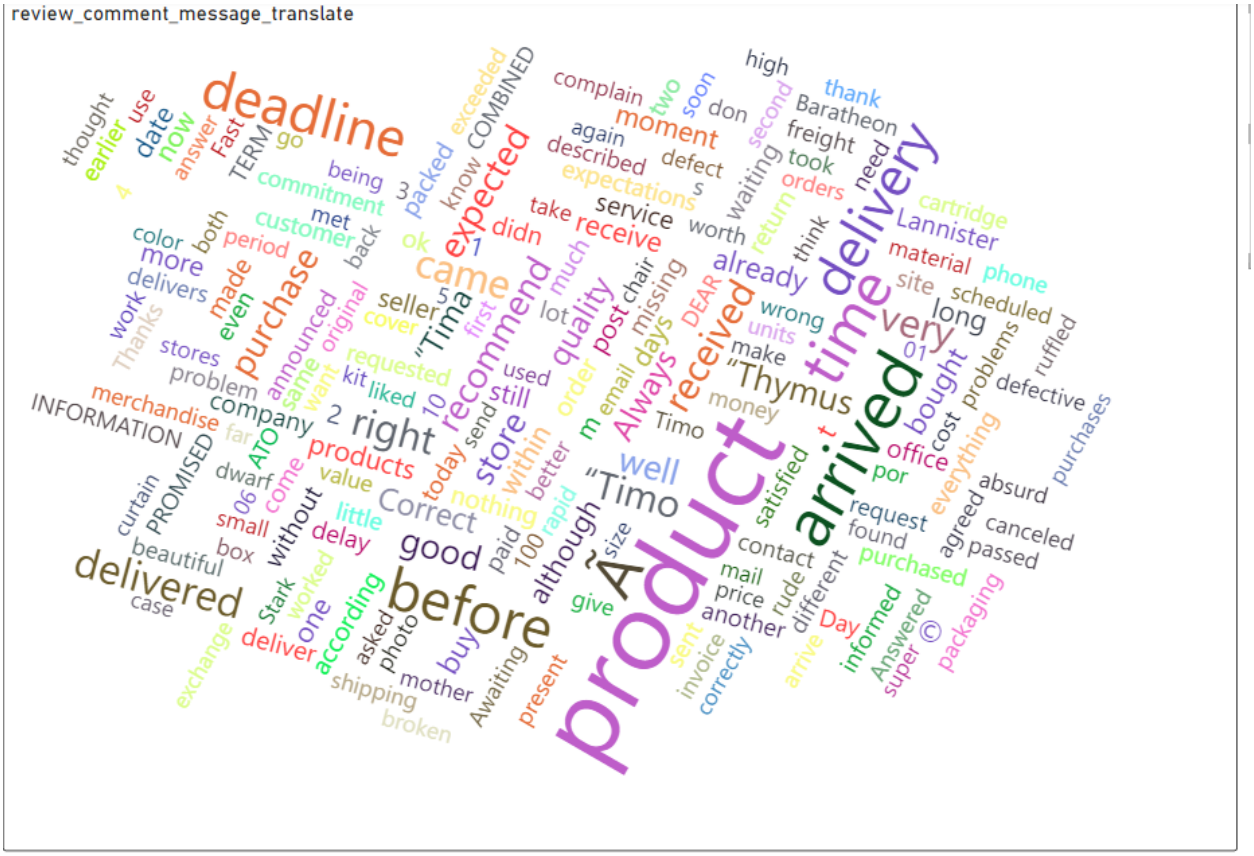


Bar chart with the count of review in category(Neutral, Positive, Negative)

The bar graph above show the customer review analysis using the VADER text mining tool, we can see that there is 25k Neutral review, 12.5k positive while 2.5k of them are negative. This suggests that the majority of customers had a neutral or positive experience with the product or service being reviewed, but there were still some negative reviews that should be taken into consideration. It's important to analyse customer feedback in order to make improvements and address any issues that may arise.

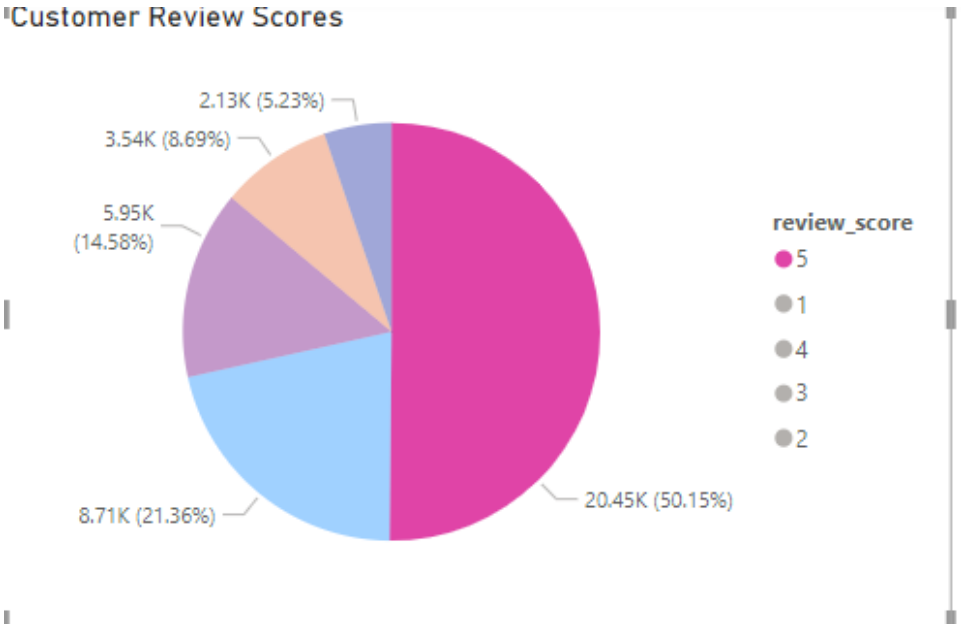
The pie chart represents the distribution of customer comments, and it shows that 64% of comments were classified as neutral, 31% were positive, and 6% were negative. This suggests that the majority of customers had a neutral opinion about the product or service, while a significant proportion of customers had a positive experience.

However, it's still important to address any negative feedback and work to improve the overall customer experience. Therefore, we will look into the segment of why there is negative comments and what it is about, before that, let’s move to analysis the keyword of neutral comments.



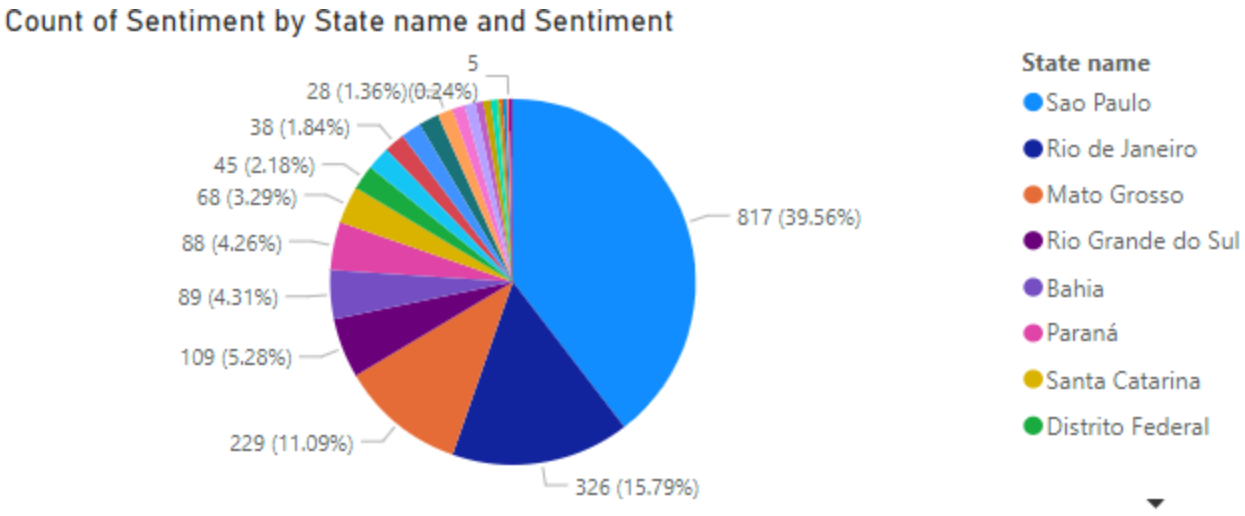
Word Cloud representing words that were most common within Neutral Reviews

The neutrality of customers' reviews can be attributed to the fact that the ordered products were delivered within the stipulated deadline and were found to be correct. Such outcomes indicate that the customers' expectations were fulfilled, and the products were deemed to be of satisfactory value.



Breakdown of Customer Review Scores

While Feedback Sentimental Analysis provided us with an overall view of customer perception of Olist and the pie chart above also illustrating the overall review score breakdown, we decided to delve deeper looking into the geographical breakdown of each state in terms of many different factors; from feedback sentimental analysis to the average price of order per state.



Breakdown of Negative Sentiment by State in Brazil

This graph illustrates the breakdown of the all the negative sentiments about Olist. This visualisation shows that over half the negative sentiments by customers who have left a review can be pinpointed to two main states in Sao Paulo and Rio de Janeiro. These two states are two of the most densely populated states within Brazil, therefore to ensure that the perception of Olist improves, focusing within these two areas would be ideal. To illustrate the reasons why the sentiment were negative, we decided to make use of word clouds which sift through the translated reviews and increasing the size of words that are repeated through the reviews, to pinpoint the areas which can be improved in terms of customer satisfaction as seen below:

Word Cloud illustrating the words which appeared the most within negative reviews

The word that glaringly stands out is product, having been mentioned 1210 within negative reviews as mention of quality, defect, missing also prominent throughout the visualisation. While Olist as an organisation act as a middleman between sellers and customers, a perception of lack of quality of products listed on Olist, can do nothing but harm the perception of the business.

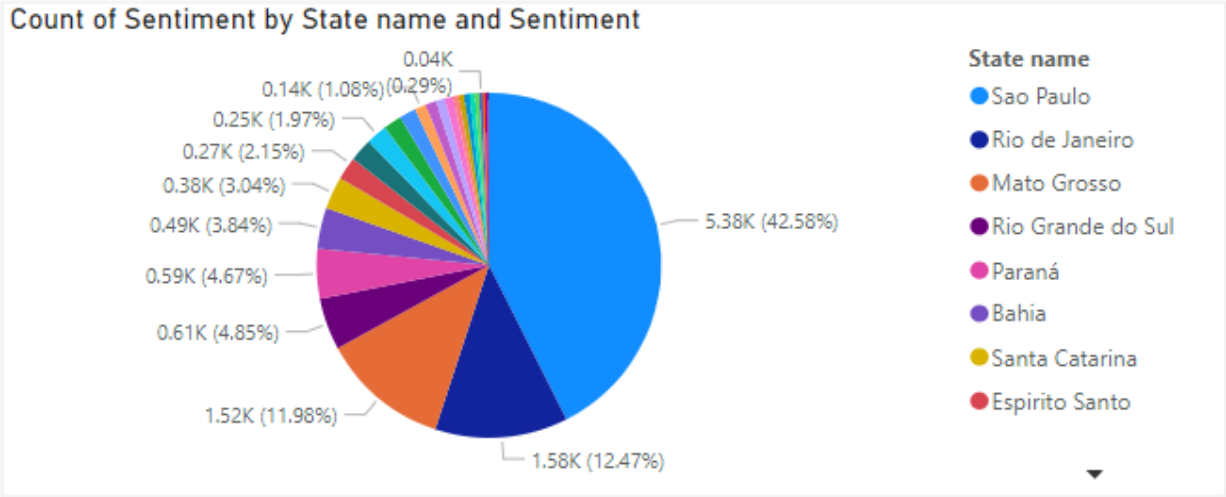
To improve the perception, we would suggest taking considerable more care in the screening of sellers while punishing the sellers who are repeat offenders and are falling below the standards, through imposing a levy or potentially completely removing them off the platform. The nature of the punishment is up to Olist, but we strongly recommend having a disciplinary system being put in place. Using a form of ID during the screening will also discourage sellers with nefarious intentions from abusing the platform and plunging Olist's perception.

A point that has been shown to repeatedly appear within reviews are points of poor customer service relations between customer and seller/Olist employee, this again shall have a negative impact on customer sentiment, in order to improve in this area we

would suggest providing training in customer relations, or providing increased customer relations training to the support team already in place. We suggest the recording of phone calls for training purposes, and Olist taking control of Customer Service in order maintain a high standard.

Another issue which seems to arise throughout customer reviews, is products which appear to be damaged before arriving to the customer, or issues within shipping which shall be analysed later in the report and its suggested improvements.

However it is also important to highlight the positives within the service that Olist provides and how to maintain and even improve on



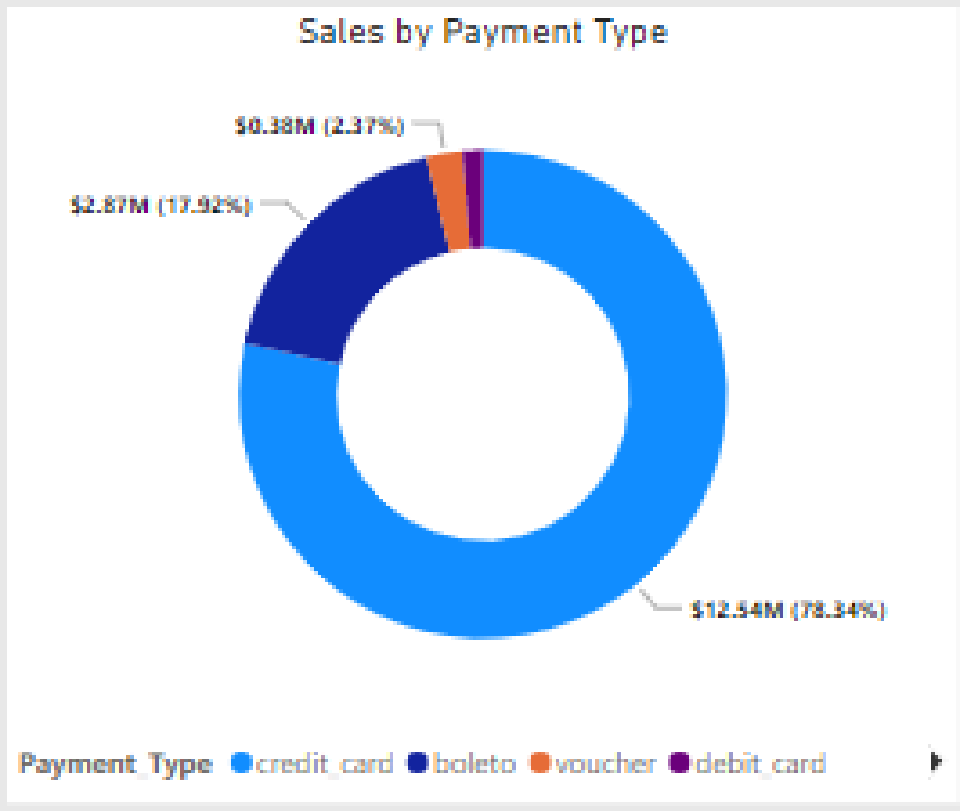
Breakdown of Positive Sentiment by State

As illustrated above, Sao Paulo and Rio de Janeiro are by and large, the states with the highest number of positive reviews with the number of positive reviews greatly outweighing the negative reviews at 6960 positive to 1143 negative reviews, which suggests that the public opinion of Olist is a greatly positive one. Within the word cloud below:

Positive Sentiment Word Cloud

The word that stand out the most are the product, deadline and delivery. This suggests that many of the issues that were negative, were not as wide spread as what it initially seemed with many of the positive reviews actually praising these factors, however to reduce number of the negative reviews, taking what recommendations we have illustrated through this report are greatly recommended over the long term future of the business.

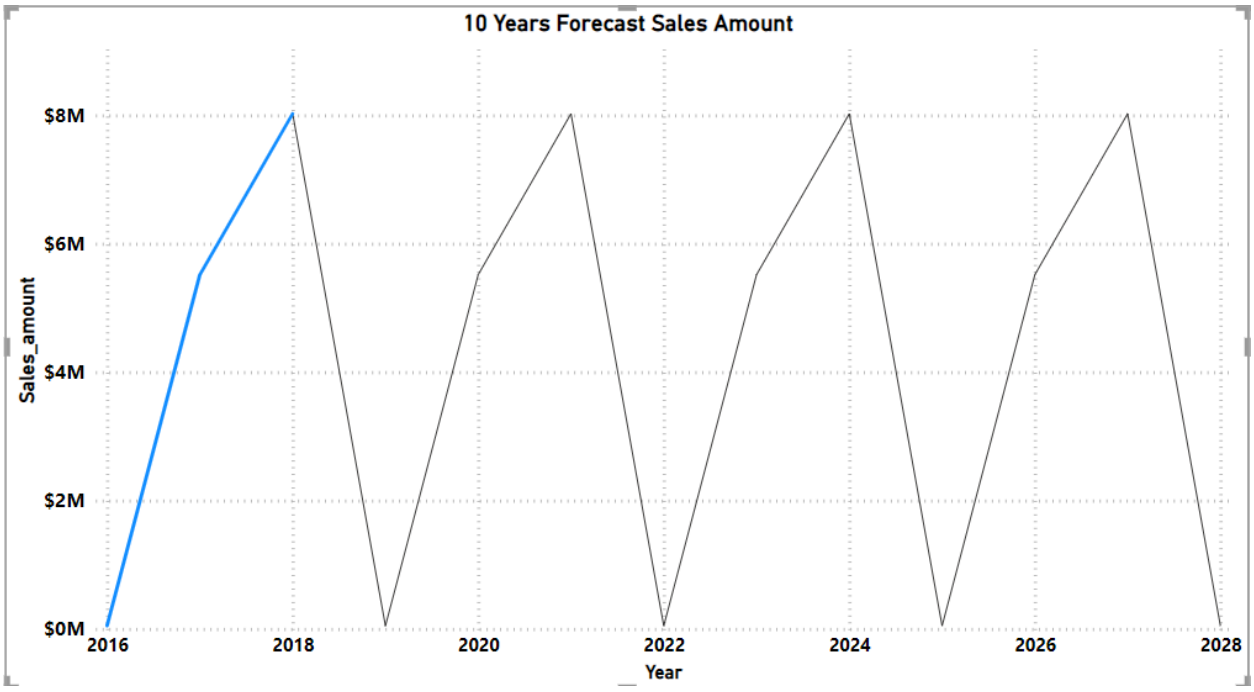
Breakdown of Payment Types



The most preferred payment method is the credit card representing 78.34% of sales payment followed by Boleto with 17.92%. Recent trends show that Boleto is becoming increasingly popular among consumers in Brazil. Therefore Olist should also look to capitalize early on this trend and look to partner with the Central Bank, or partner with banks, in order to promote its use and promote the business also, allowing for a harmonious relationship. Due to the nature of Boleto, reducing the amount of personal information being given out, advertising the safety aspect similar to Whatsapp's advertising methods would be advantageous especially given the current public perception of data sharing.

Sales Prediction

To predict the future sales of Olist eCommerce, we used exponential smoothing which provided different weights at each point to obtain the forecast. To calculate this, the sales amount was tabulated against the three years period of data provided (the weights calculated using Power BI), and set our forecast period to 10 years.



As Olist was only established in 2015, the data provided in terms of sales are from the inception of the company, therefore the forecasting would use this data to extrapolate the predicted revenue, which would not be an accurate or fair prediction, therefore we suggest that a few fiscal years to pass before producing a 10 year sales forecast, this will also account for any global phenomena which may have a drastic effect on the sales figures, and skewing the data and therefore forecasting, such as the Covid19 pandemic or global recessions which had extreme effects on sales figures around the world both in negative and positive ways.

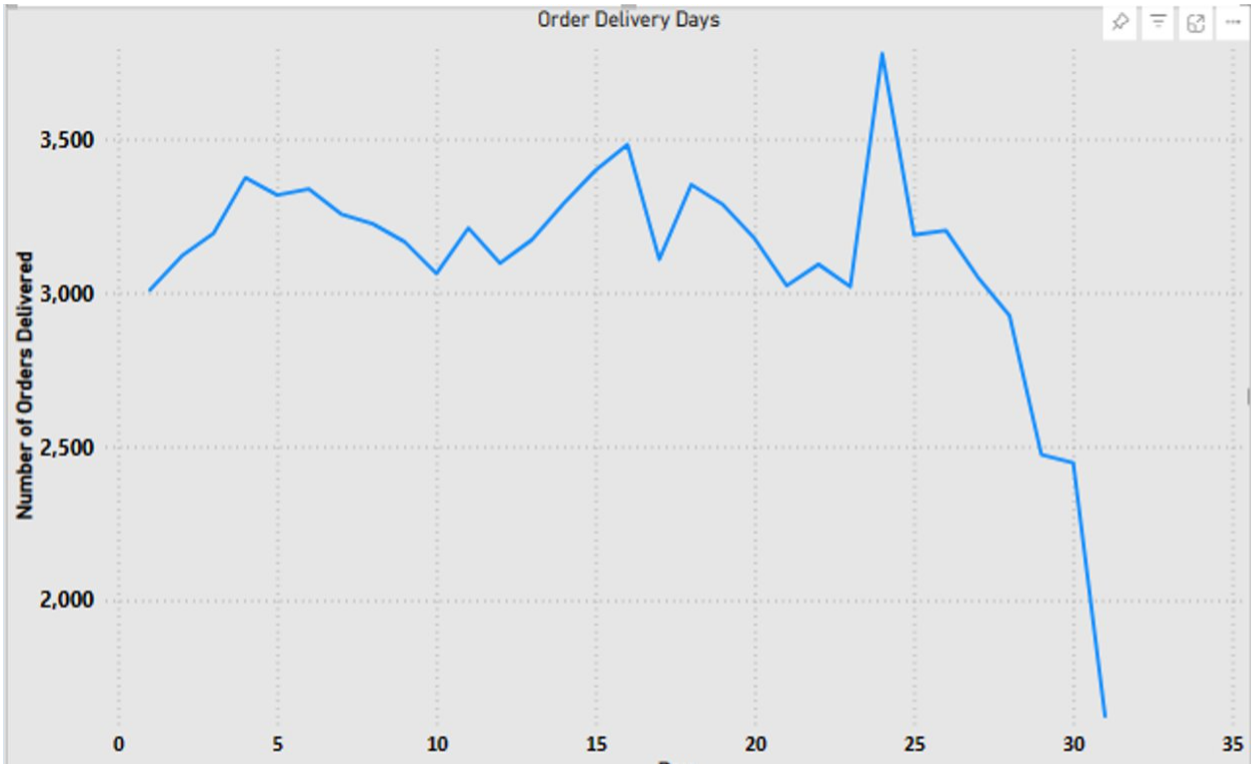
Delivery Performance

When assessing delivery performance, it is important to consider the overall customer experience throughout the order process. By analysing customer reviews, we were able to identify specific areas of concern and negative feedback related to the service provided. This allowed us to gain a deeper understanding of the factors that contributed to poor delivery performance such as logistics, shipping, and inventory management, as well as identifying opportunities to streamline processes and improve overall efficiency.

Certain reviews received a low rating due to non-compliance with the specified delivery date, resulting in a score of 1. As a customer-centric organization, Olist acknowledges the importance of adhering to the agreed delivery timelines, and in such scenarios, the

customer's expectation is considered to be valid. Therefore, addressing such concerns is a priority for Olist in order to enhance customer satisfaction and maintain a high level of service quality.

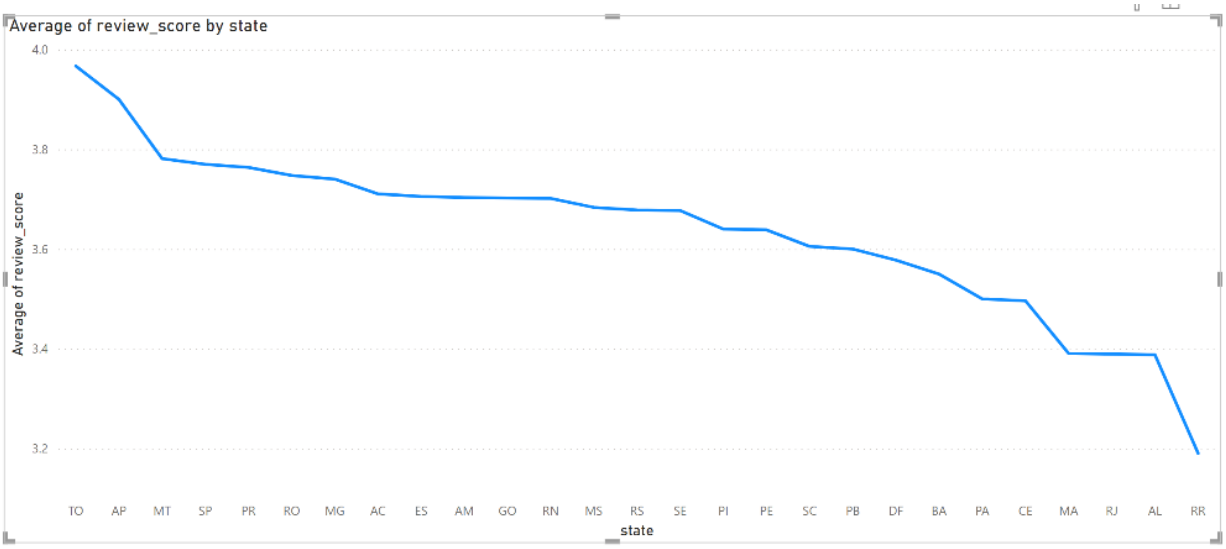
In order to calculate and evaluate the Delivery Performance, we used multiple different metrics such as the average time for the customer to receive the product after placing their order which we discovered to be 11.88 days, while the orders seem to take 2.71 days to be processed and sent to the carrier within 9.28 days, however it can be see in the graph below a high amount of packages were delivered after 24 days which is too large a number to maintain a high customer satisfaction.

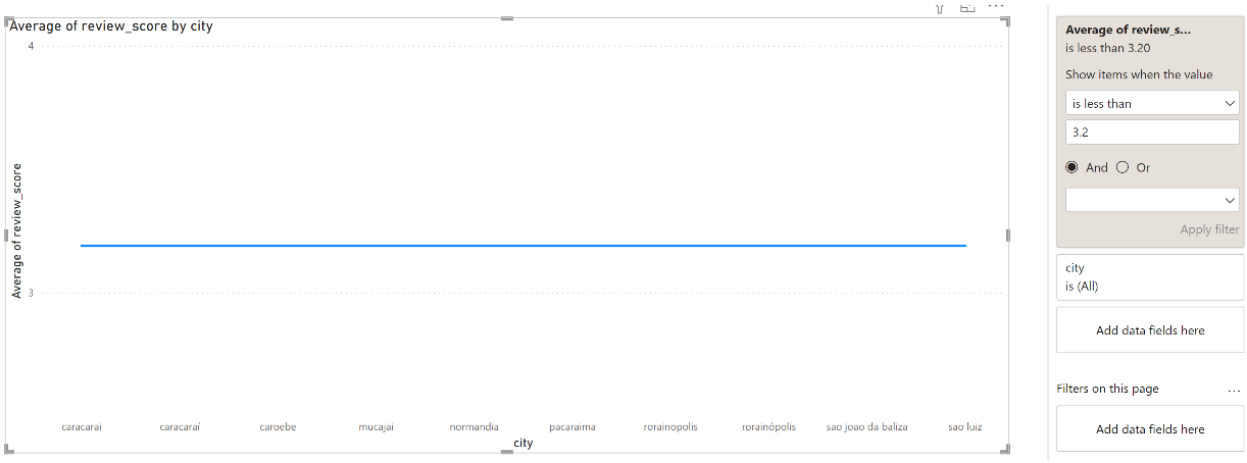


The chart below presented in the report displays the average review score of the Olist data set across different states, offering insights into the relationship between location and review scores. By applying a filter to display only the states with the lowest average review score, we found that these states were located in sparsely populated regions that are farther away from major urban centres. These findings highlight the impact of geographic location on delivery performance and suggest that Olist may need to adjust its delivery strategy to better serve customers in rural or less accessible areas. This

could involve implementing targeted marketing campaigns, optimizing delivery routes, or partnering with local delivery providers to improve delivery times and overall customer satisfaction.

The lowest average review score was 3.19 and the highest by state was 3.97





Conclusion

Olist while a relatively new company having only been established in 2015, it has been extremely successful and how shown large growth year on year. However as a new

company in order to establish themselves in the market and to maintain good customer perception for returning customers and potential new customers.

The analysis of the Olist ecommerce 2016-2018 dataset has provided valuable insights into the performance of the Olist platform and its sellers during the period under review. The dataset contains information on various aspects of the ecommerce business, including customer orders, payments, reviews, and seller metrics.

Through the analysis, we have identified trends and patterns in customer behaviour, such as the most popular products and payment methods, as well as the factors that influence customer satisfaction and loyalty. We have also explored the performance of sellers on the platform, including their sales volume, customer ratings, and shipping times.

Overall, the findings of this report can be used to inform business decisions aimed at improving the performance of the Olist platform and its sellers. For example, insights on customer behaviour can be used to optimize marketing and sales strategies, while seller metrics can be used to identify areas for improvement in product quality and shipping efficiency.

The Olist ecommerce 2016-2018 dataset provides a valuable resource for understanding the dynamics of ecommerce businesses and for developing data-driven solutions to improve their performance. The dataset is however limited and now quite out of date as the data is from 2018. Therefore to improve the analysis in the future, using a larger and more recent dataset will produce more relevant insights to Olist that they can use to produce more effective data driven solutions