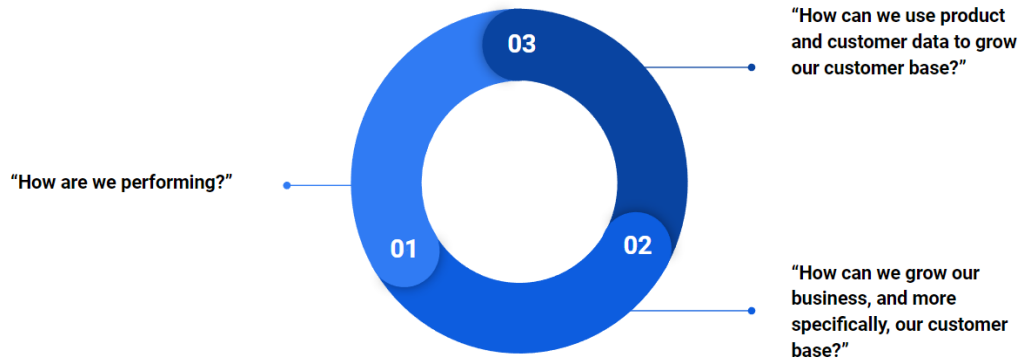
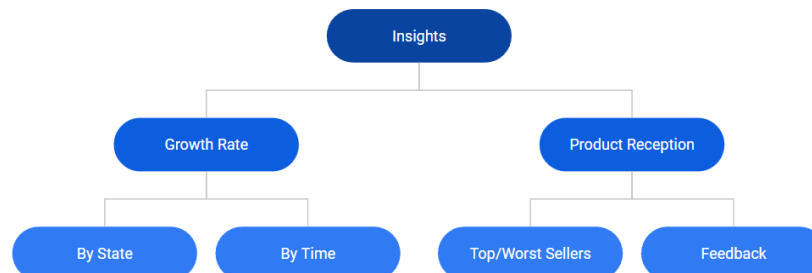


Report on E-Shop.com's Performance

Osamwonyi, David, Nene



Our task was to find out “How are we performing?”, yet there is more to meet this eye in that question itself. The ultimate goal of this analysis is to conjure results that aim to improve this company’s customer base. So, the better question would be: “How can we grow our business, and more specifically, our customer base?” The process to get to this question requires understanding what we have available to us and what is the end-goal. What we have available at our disposal is several data frames that encompass 2 main categories- product and customer data- that are essential to the analytical process of business performance. Product and customer data can help us come up with insights on opportunities that we can use as a company to grow. The end goal is to grow our customer base so when you acknowledge these circumstances, we can then reframe the question even further to “How can we use product and customer data to grow our customer base?”

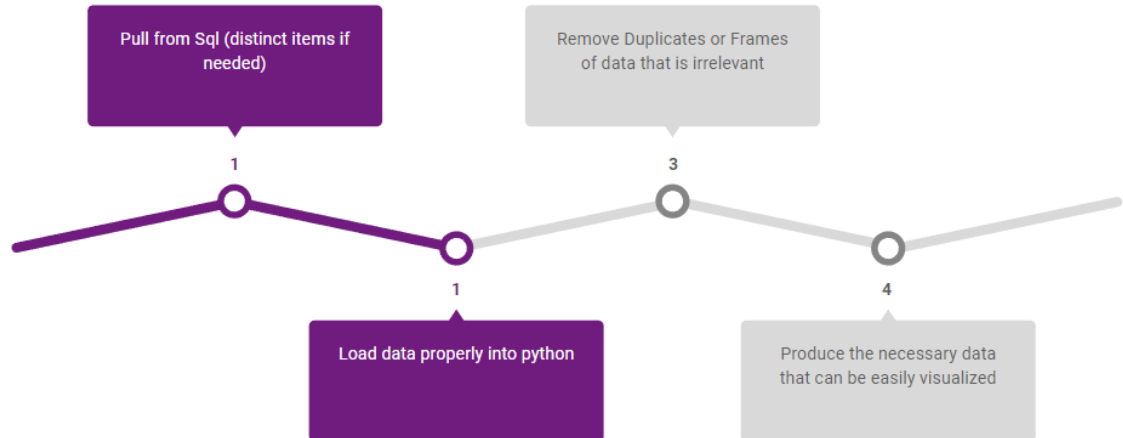


As alluded to before, the reframing process involved understanding what was the end goal. It is stated above that insights are needed to find opportunities to grow. But what are those insights? One of them is the growth rate, and in this case, the growth rate of the customer base in

Brazil. That can be broken down into the growth rate by state as there are 27 federal states of Brazil (including the capital district). It can also be viewed as the growth rate of the overall customer base annually, biannually, and quarterly. Furthermore, one could combine both methods and search for time-based growth rates in each state. Another insight is the products and their relation to the region of Brazil. This can be evaluated by looking for top sellers in the nation and its states, and as well as the reviews of those products themselves. Knowing where your products are going and the quantity of them is part of the battle, not the whole. The other part is the feedback. We need to know which products aren't liked just as much as which products aren't selling, as correlation does not mean causation. Meaning, we cannot assume that the worst selling products are the worst liked products. It might just be a case of low activity in the economy or as well as the population distribution across the country. External factors make it such that we cannot assume in such a manner.

The information that we have available is as follows:

- Customer Based Data
 - This type of data refers to the general details of the average customer that buys our products. Ranging, from state and city residency, to the unique id that the customer has been given within the system. This is important to note as when it comes time to organize the data, it must be done with the unique ID and not the general customer ID.
- Order Based Data
 - This type of data refers to the general details of orders and along with it shipping details. This consists of processing and shipping dates, identification of orders, and as well as identification of items.
- Product Based data
 - Similar to the order based data category, the product based data we have here is by its namesake, more item prioritized. It entails item ID, dimensions, respective pricing, Seller information, and review information.

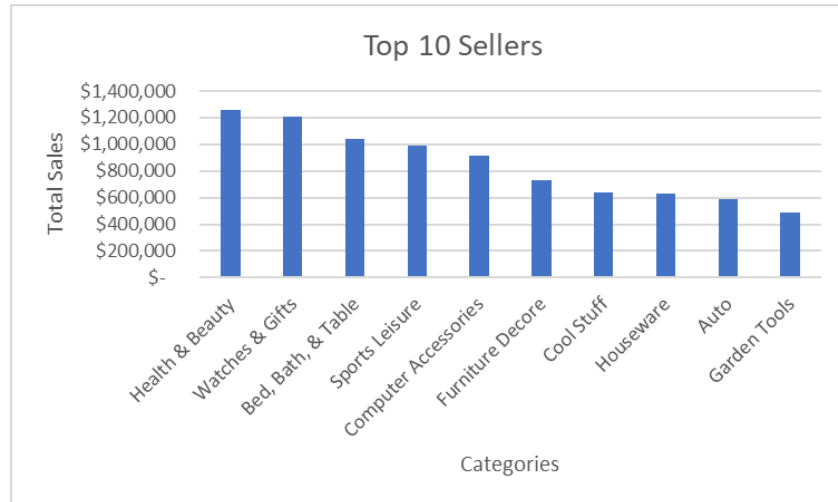


The data cleaning process overseas is an integral step in getting the results we want as without assorting and mending the dataframe(s) we have at our grasp, we will get inaccurate results which could skew the truth that we are searching for. To avoid this it is important to always have two things in mind: Awareness of duplicates and what we are looking for. Duplicate rows/columns/cells can appear and it is important to be aware of them. Our team dealt with this issue by:

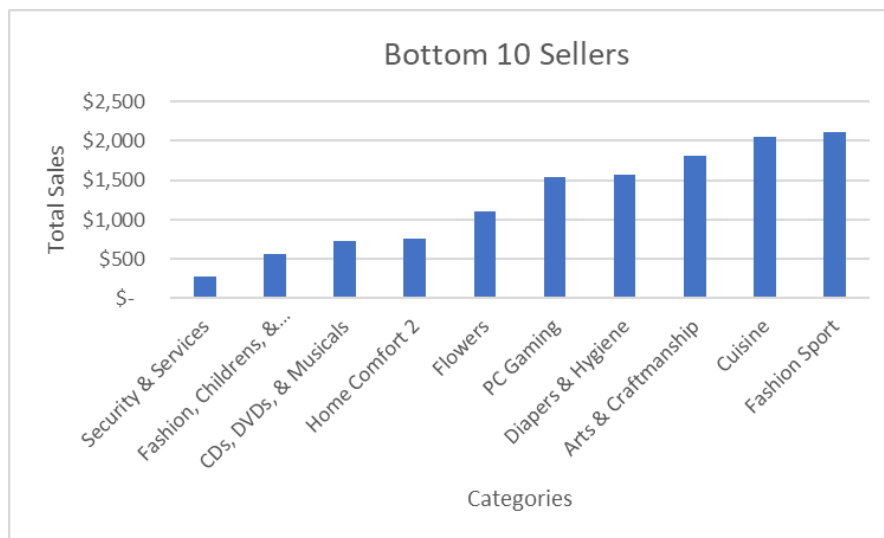
- 1) Uploading every data frame that we are receiving from SQL into python
- 2) Confirming the dimensions of the dataframe (its shape and its elements)
- 3) Using the duplicated() function to get back rows of data (if they exist)
- 4) removing such data with the remove_duplicates() command.
- 5) Using dropna() or fillna() to remove or replace rows of data that are filled with null values.

To further ensure data integrity was followed through this was also done when we began to look at individual dataframes (at a lower level) just to confirm that all duplicates and nulls were removed. Now we can focus on our key insights.

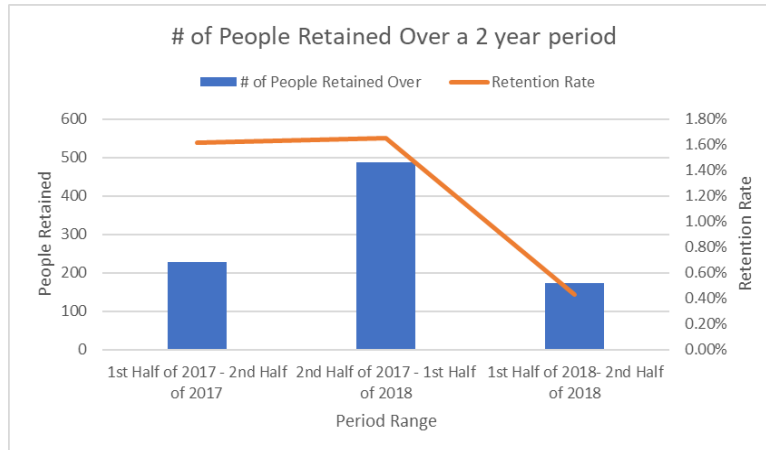
This is what we found after analyzing the data..



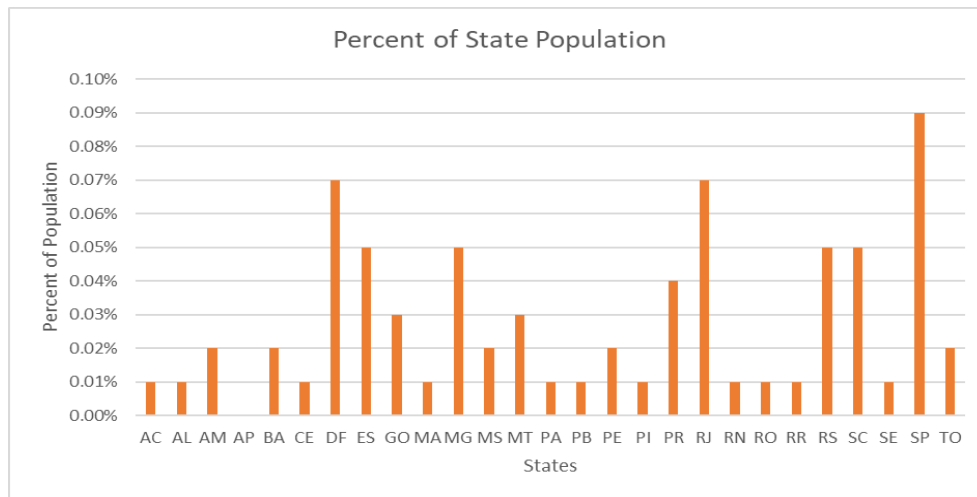
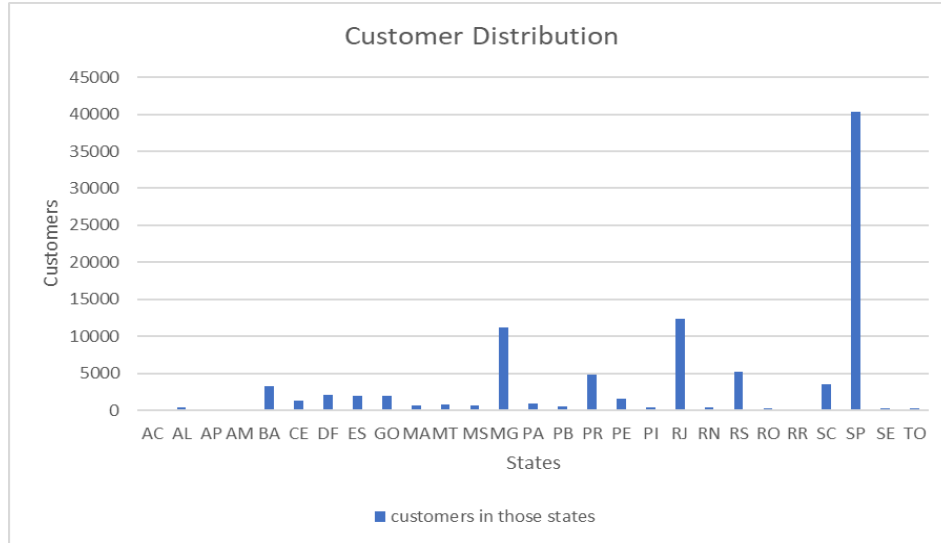
Here we have made a list of products that sell the most. The graph shows the top 10 highest earning products, products that generate the highest amount of revenue. This is a clear indication of the products that the customers want and a good indication of products we should be pushing to sell more. The top 10 highest earning products show great demand and hence, should be the products we push to states with few customers in order to increase our sales.



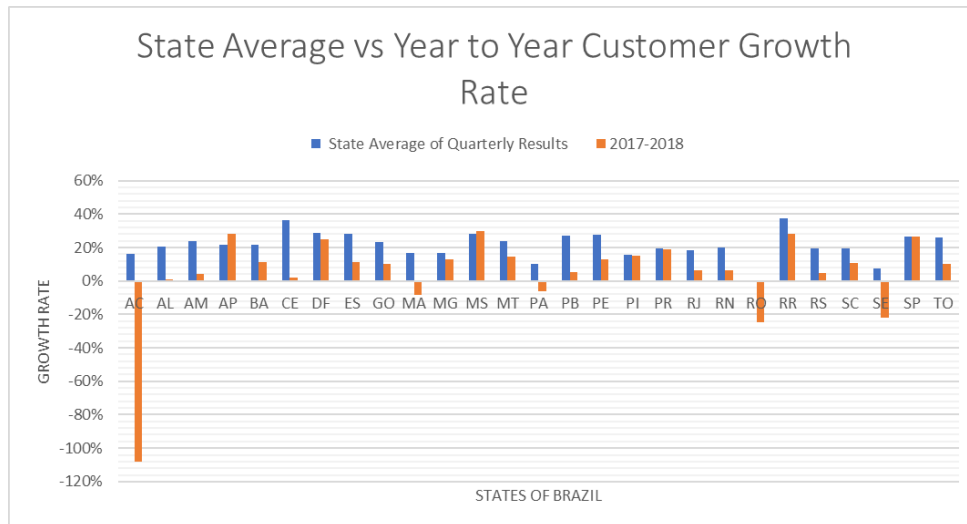
The bottom 10 products that sell the least gives us an indication on which product we should be focused on expanding. Our fashion section creates the least amount of revenue. This is something that can be improved upon as Fashion is seen as a very important aspect of a person's life nowadays. The bottom 10 products are the products we need to give more time in order to try to increase more revenue through them, starting off in states like Sao Paulo with an already established customer base.



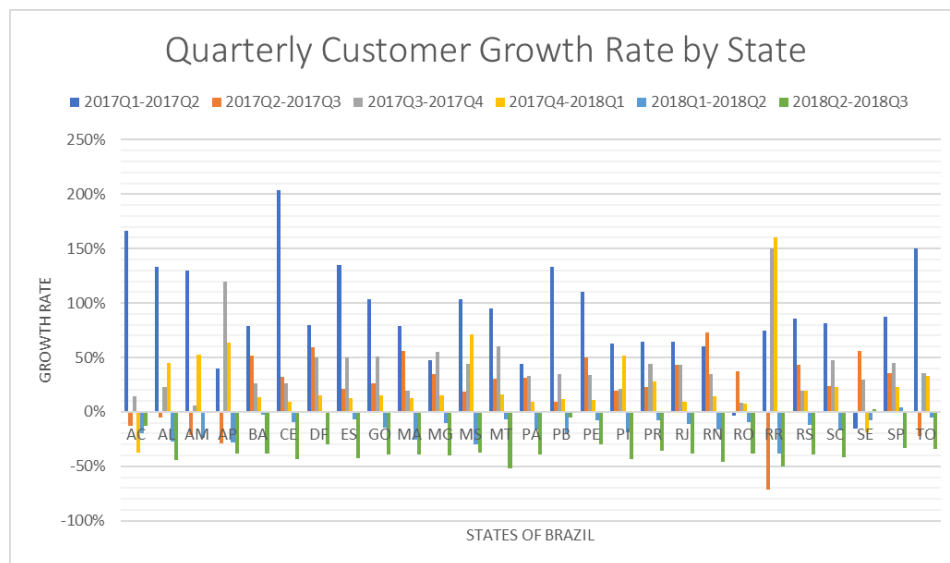
This graph shows the amount of customers the company has retained over the years. The time period ranges from the 1st half of 2017 to the second half of 2018. To understand this graph better, a second axis of retention rate was added. How these numbers were achieved was by collecting all the unique customer ids of the first half of 2017 and then the second half of 2017. The dataset of the second half is then compared to the first half to see how many people in the second half were also from the first half. The number displayed is the number retained over those two periods and the percentage displayed is the fraction of those retained over the initial period. It can be seen that the retention rate of the customer base relative to the total number of customers of those periods is significantly small. Any retention rate within the single digits is a pause for concern as it indicates that the company is retaining a small percentage of its customers.



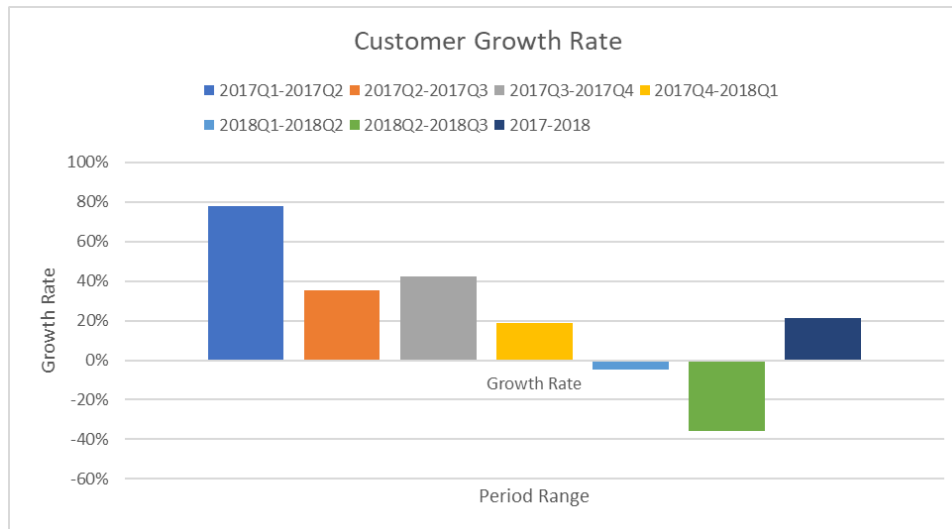
The bar graphs above go hand in hand in showcasing the general distribution of our customers across the country of Brazil. An important thing to note is that the bulk of our customers reside in Sao Paulo, the most populous state of Brazil. The data indicates that customers from Sao Paulo resonate with our products the most, with sports, leisure, health & beauty, and bed/bath/table being the top 3 categories in that state in terms of items sold. However, there is space for opportunity in all 27 districts as there are millions of people we have not reached out to. Even in Sao Paulo in which 0.9% of the population is our customer base represents the total population of that state.



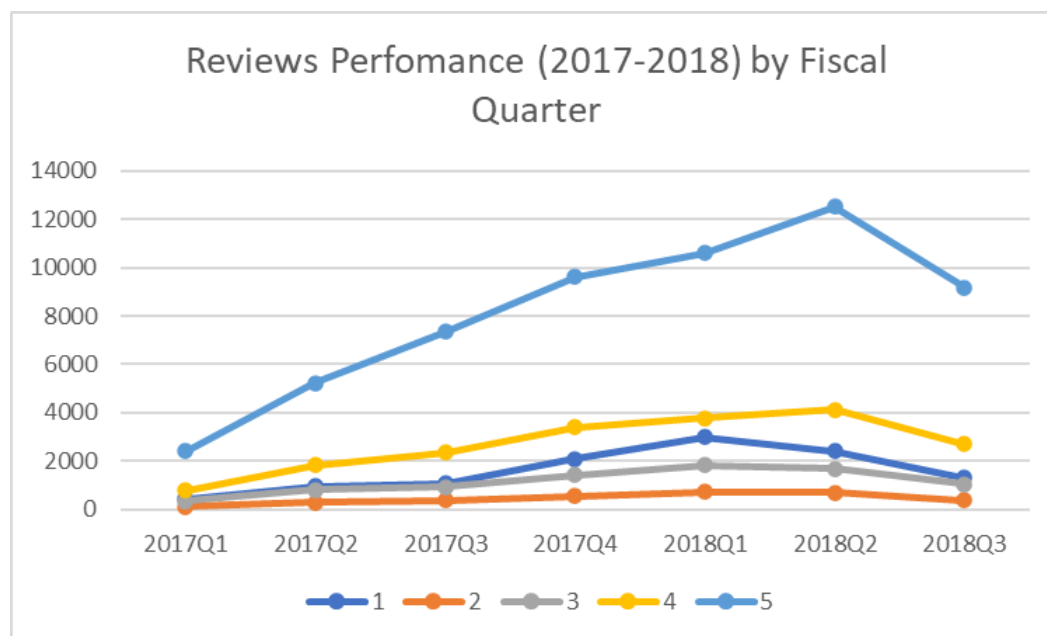
This graph shows the yearly customer growth rate in all states compared to the state average. There is a clear indication of decrease in the yearly customer growth rate in most states. States like Amapá, Mato Grosso do Sul, and Sao Paulo have a very positive yearly customer growth rate whereas states like Acre, Maranhão, Pará, Rondônia, and Sergipe show a negative customer growth rate which is going to require a lot more attention. Looking at all states, we can conclude that most states do not match the state average, indicating lack of state growth in most of the states. However we must consider that certain states having smaller populations are more prone to erratic changes in the growth rate since there is such a low number like Acre (AC).



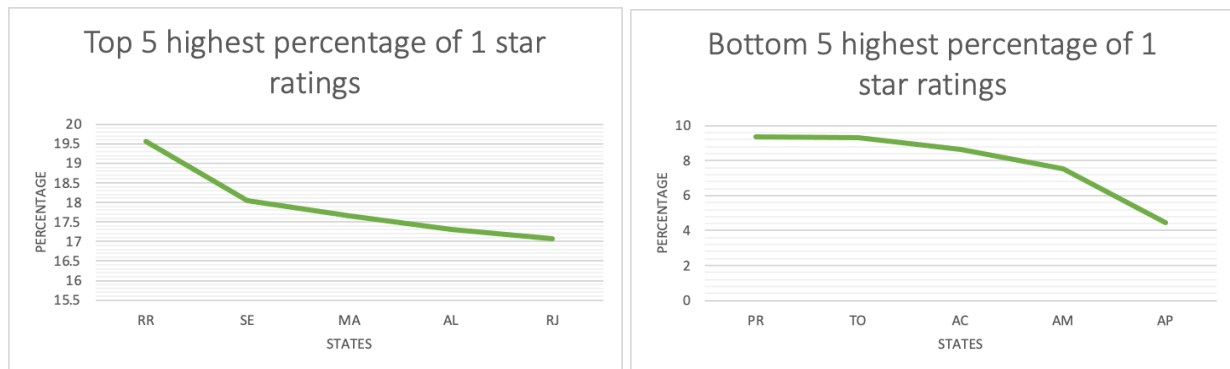
This graph shows the quarterly customer growth rates in all states. There are many ups and downs but all states see a decrease in customer growth rates in the Q1-Q3 for 2018 while most customer growth rate is negative for all states for 2018 Q2- 2018 Q3. This period of 2018Q2-Q3 is notably becoming a trend to which we will acknowledge later.



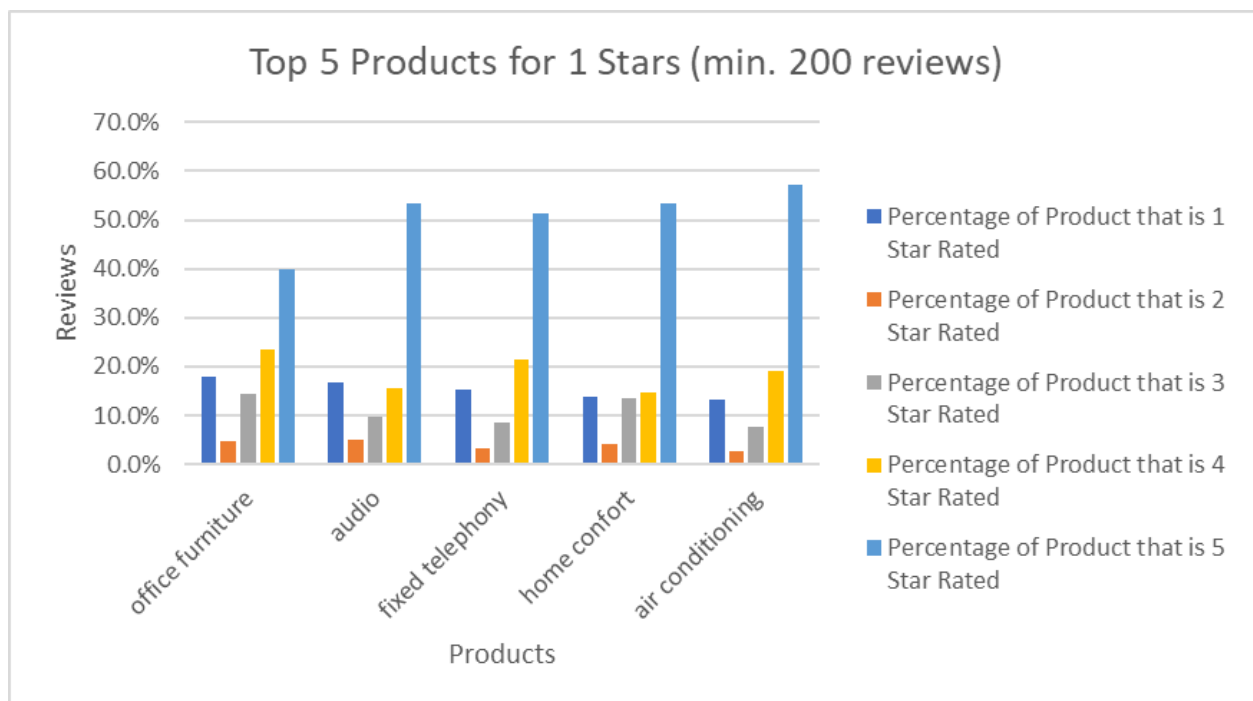
This graph shows the customer growth rate for the entire country of Brazil. Although the customer growth rate is up and close to 20% from 2017-2018, it shows that the growth rate has decreased over each quarter and if the trend continues, the overall growth rate will decrease a lot more. We will be gaining less customers each quarter if a change is not made.



This graph shows the number of reviews that we have received in different quarters from 2017-2018. We have received 5 star ratings the most, followed by 4 star rating but all forms of reviews see a decline in 2018. The immediate decrease of 5 star ratings from 2018Q2 to 2018Q3 is the most concerning as this shows a decrease in overall customer satisfaction. But a large growth in 5 star reviews until 2018Q2 is a great sign and shows what we did had worked but we need to work a little harder to help increase our reviews.



These graphs are from the total percentage of 1 star ratings we have received. The Top 5 states with the highest percentage of 1 star ratings amongst all the reviews and the 5 states with the lowest percentage of 1 star ratings are shown here. In states like Roraima, Sergipe, Maranhão, Alagoas, and Rio de Janeiro, the very highest percentage of customers have shown their dissatisfaction with our products and this shows a need to do better in all states of Brazil but especially in those states.



This graph shows the products with the most number of 1 star reviews which is a clear indication of low customer satisfaction from these products. The 5 star reviews are still a lot more than 1 star ratings but 1 star ratings here should be decreased as we need to maximize customer satisfaction as much as we can. Office Furniture has the most number of 1 star ratings which is close to 18%.



On the contrary, this graph shows the top 5 products with the most 5 stars reviews. What can be taken away from this is that on a percentage basis, these products are doing relatively well in comparison to a product that is selling higher volumes but the percentage of the products being bought is seen at a lower quality. For example, Health & Beauty have a total number of 5377 5-star rated scores in comparison to Books and General Interest which only had 370. However 73% of the Latter was rated at 5 stars. A 10+% difference in comparison to H&B's 61.8%. This analysis was necessary in order to find items that could be selected as target products that can be pushed to increase sales. Their volumes are low but people love them when they buy them.

A trend that was noticed between the retention, customer, and review growth rate was the stark drop in growth from the 2nd Quarter of 2018 to the 3rd Quarter. This is concerning as it was seen throughout 3 different metrics assessing growth rate. Furthermore it ultimately was a factor in the growth rate from 2017-18 being in the 20% range rather than in the 30-40% range. People are buying less, reviewing less, and just overall interacting less with our products towards the end of 2018 and it is putting a dent into the gradual progress that was made since the start of 2017. 2 key suggestions can be made in order to combat this issue. One suggestion is to consider researching more into the top 5 products that scored high on review percentage. These products could help rejuvenate sales if more attention was brought to said items given their high scoring rate on percentage. Another suggestion is to research more onto what made Sao Paulo resonate with our customer base so much. It is indeed obvious that it is the most populous city, however. Given that almost 42% of our products were either purchased or requested from there we should consider putting more time into the state as well as develop a model based on that state that we

can apply to others. Overall, gradual gains were made in creating a base of operations in Brazil in 2017 but losses in 2018 may prove to undo those efforts. Changes must be made in order to turn the tide.

References

<https://www.ibge.gov.br/en/statistics/social/population/18448-estimates-of-resident-population-for-municipalities-and-federation-units.html?=&t=resultados>

<https://www.worldatlas.com/articles/brazilian-states-by-population.html>

<https://www.worldatlas.com/articles/the-richest-and-poorest-states-of-brazil.html>