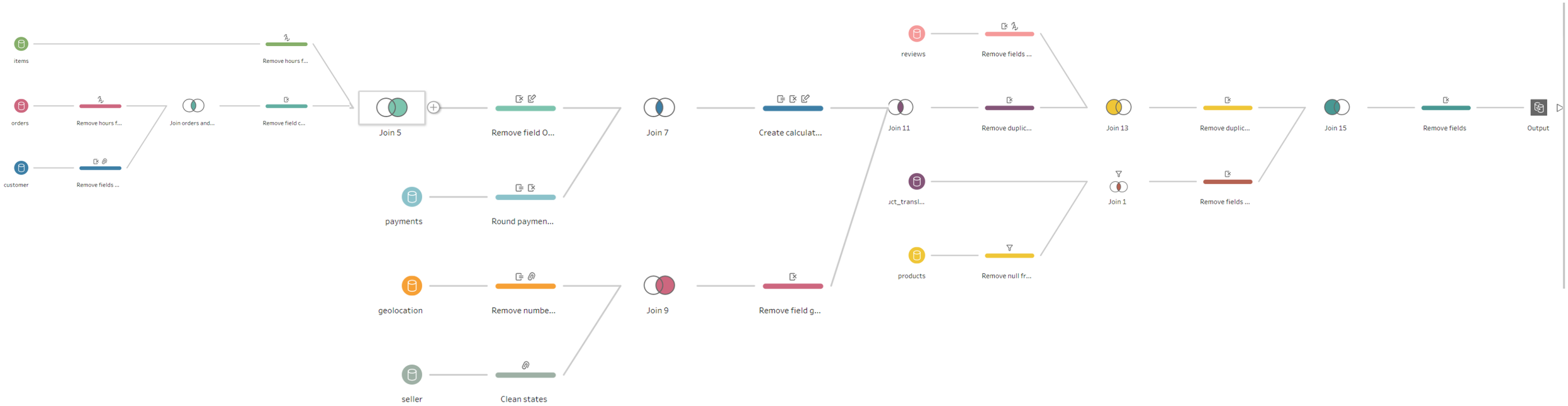
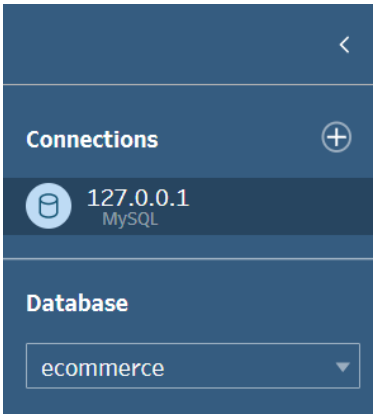
Lab 2 – Group 1

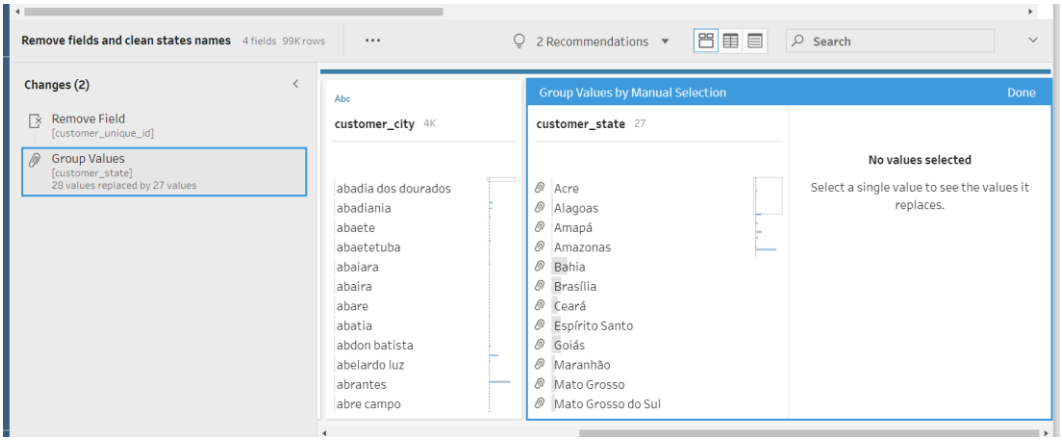
Cleaning our data with Tableau Prep



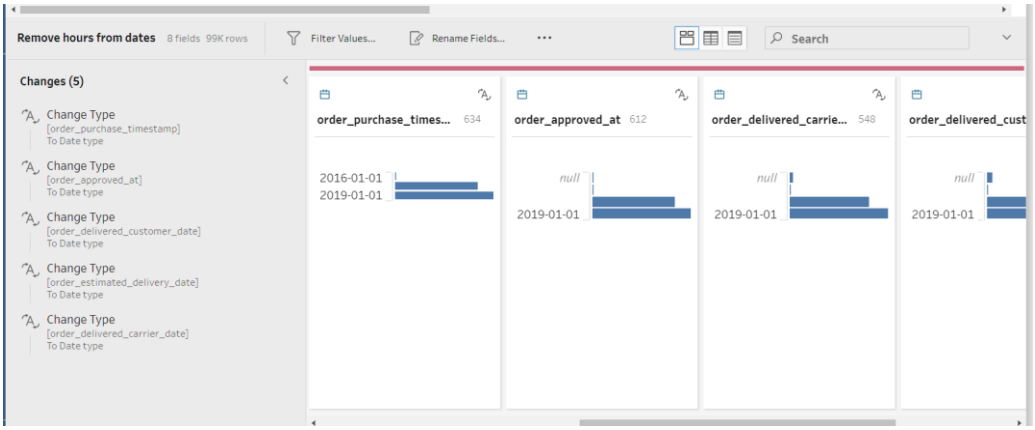
1. Add a connection. For this step, we made a connection in MySql.



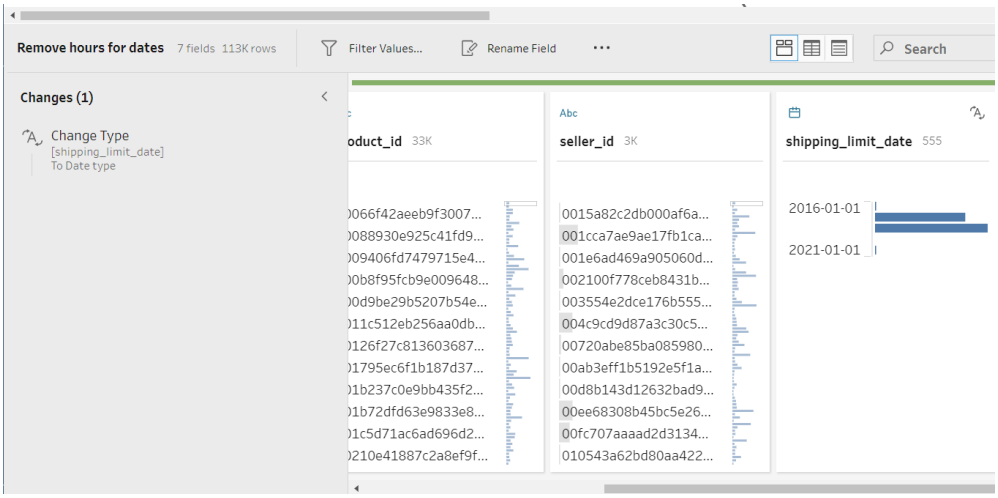
2. After we loaded all tables, we found some issues in our data. For example, states were written like AC, AP, RJ, and so on. Another example was prices were with more than 10 decimals. We also found that all dates columns had hours included, and for our analysis we decided to not consider it. Finally, we removed fields that were not relevant to our project.
    a. Customer: Group values by Manual Selection. – we replaced the names of the states AC for Acre, and so on. Also, we eliminated column customer_unique_id because this column wasn't show in any table.



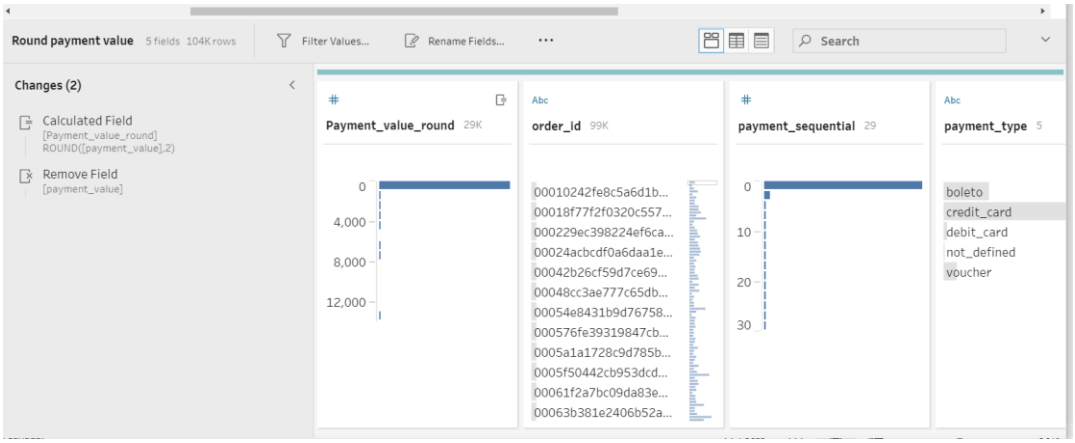
    b. Orders: We changed the datatype to date for all date columns in order to eliminate hours, minutes and seconds from our data.



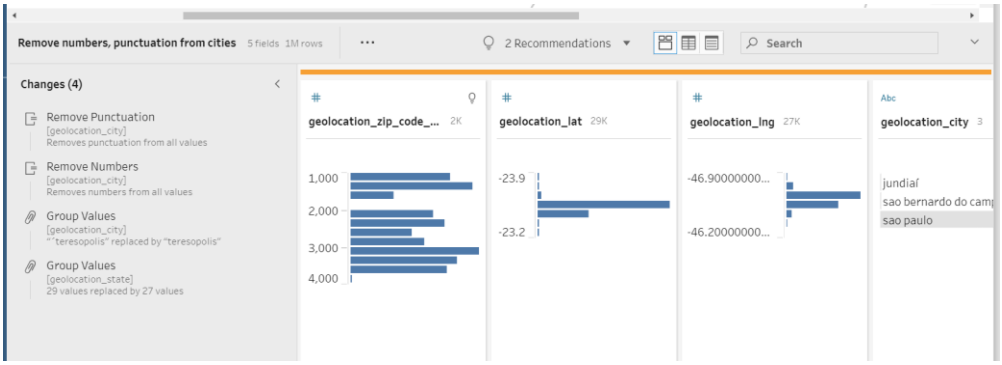
c. Items: We changed the datatype to date for all date columns in order to eliminate hours, minutes and seconds from our data



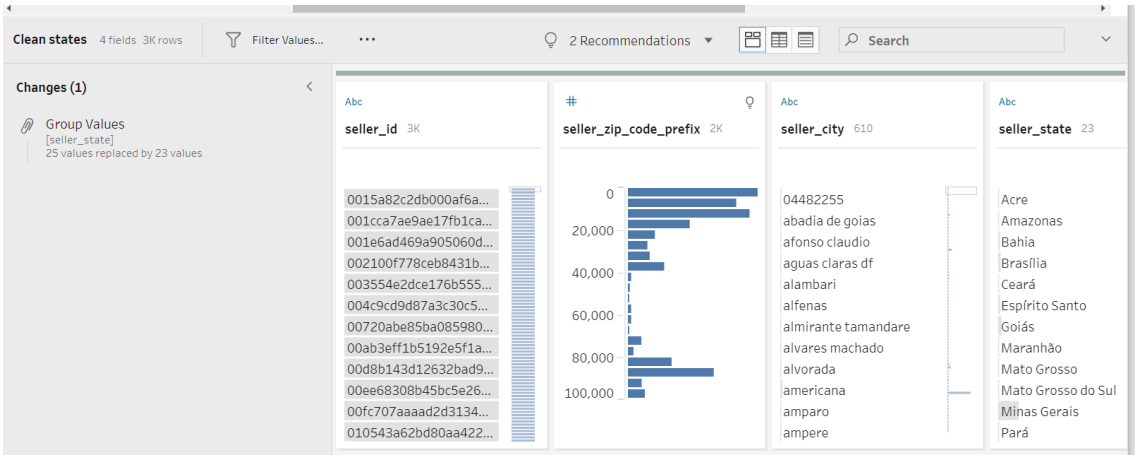
d. Payment: We created a calculated field(“payment_value_round”) in order to round the price in 2 decimals and remove the field “payment_value”.



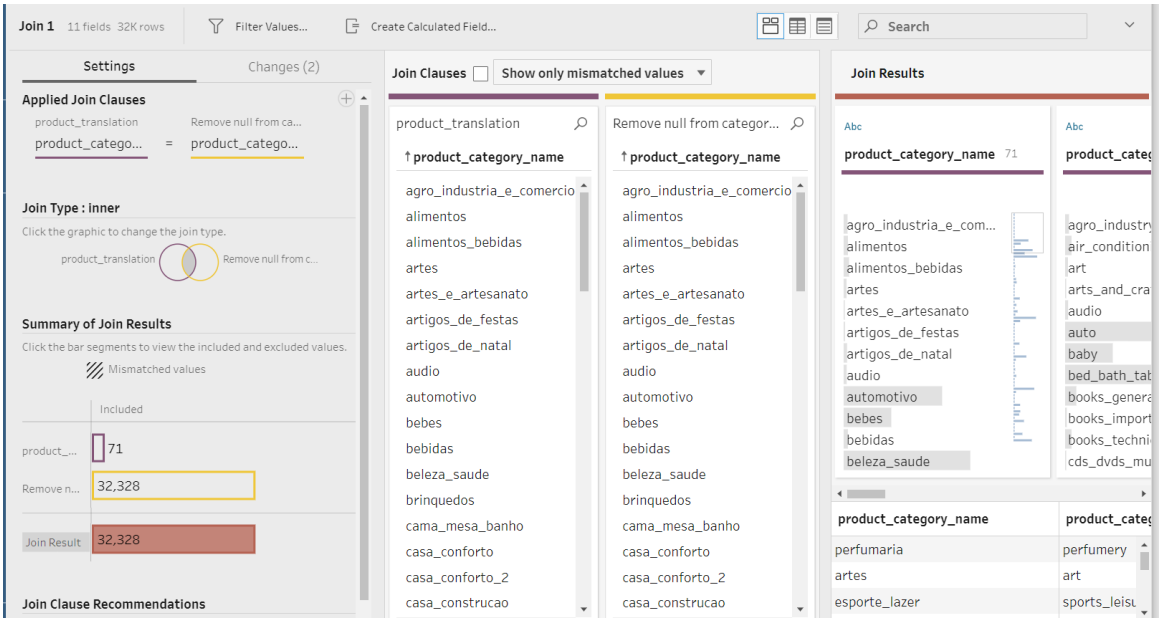
e. Geolocation: Group values by Manual Selection. – we replaced the names of the states AC for Acre, and so on. Also, we eliminated punctuation in geolocation city.

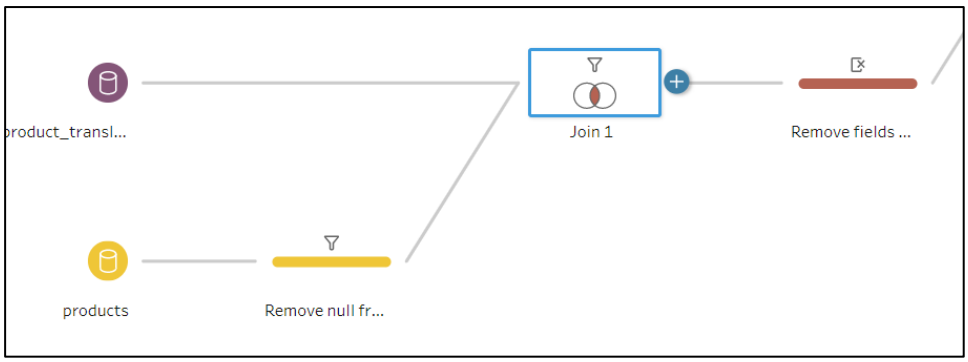


f. Seller: Group values by Manual Selection. – we replaced the names of the states AC for Acre, and so on.

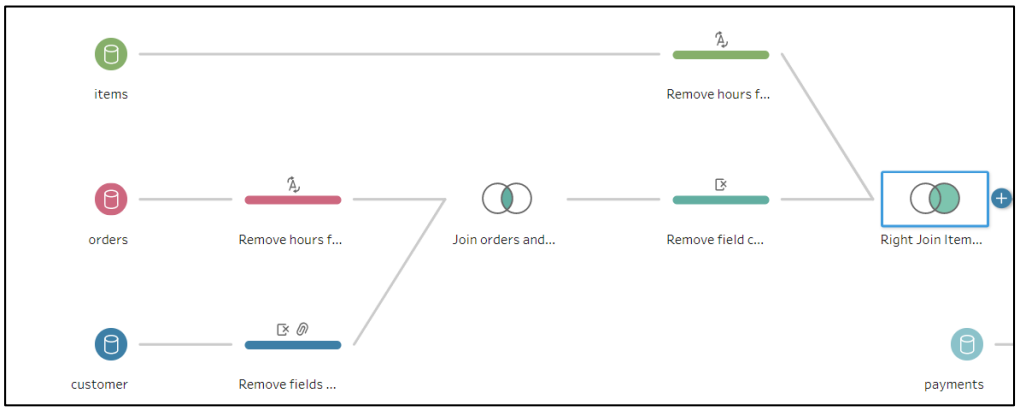


g. Products and product_translation: In order to find the category English translation, we decided to do an inner join between Products and Produc_translation. After we did the inner join between those tables, we proceed to eliminate product_category_name and keep the English version.

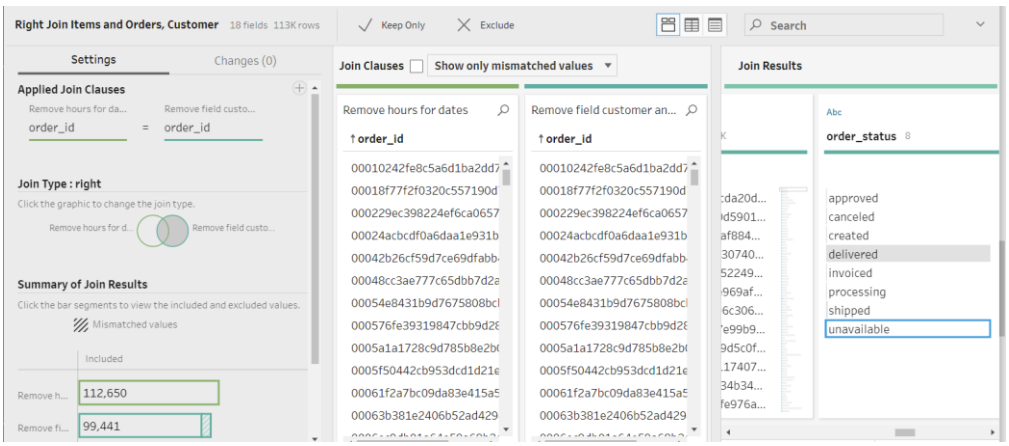




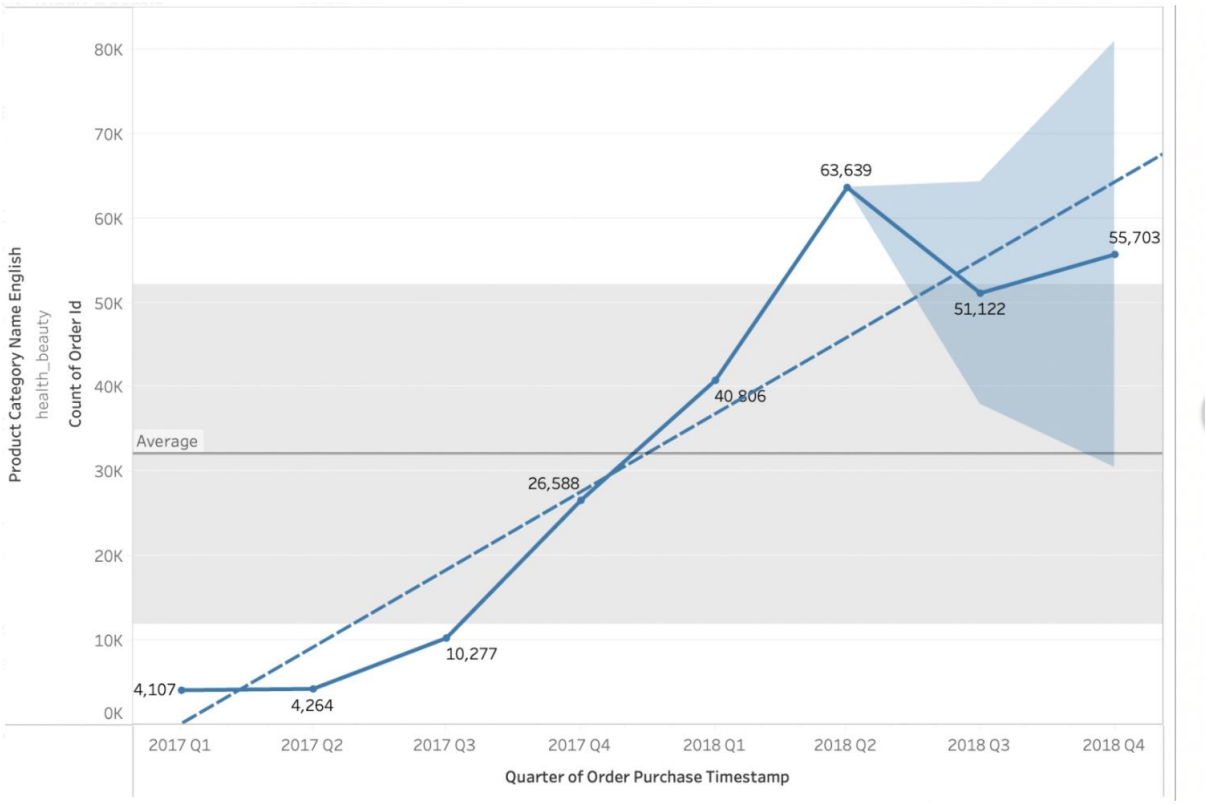
After we clean our data, we proceed to use left join and inner join in order to have a Fact table with all the columns that we will need for our analysis. It is important to mention that after any join we eliminated fields that we were duplicated.



For example, we decided to use right inner join with Items and orders, because we noticed that were some orders that weren't in Items. Also, we noticed that those orders had as an order status cancelled/unavailable that we could use as a future KPIs.



Analysis 1 – Quarterly revenue generated for health and beauty product for 2017-2018



2017-2018 Orders for "Health and Beauty" category

In the above figure we have shown the count of orders for the "health and beauty" category starting from 2017Q1 till 2018Q3. Using the historical data here the graph also shows predictions for the number of orders for 2018Q4. The graph also plots a confidence interval around the prediction.

Issue

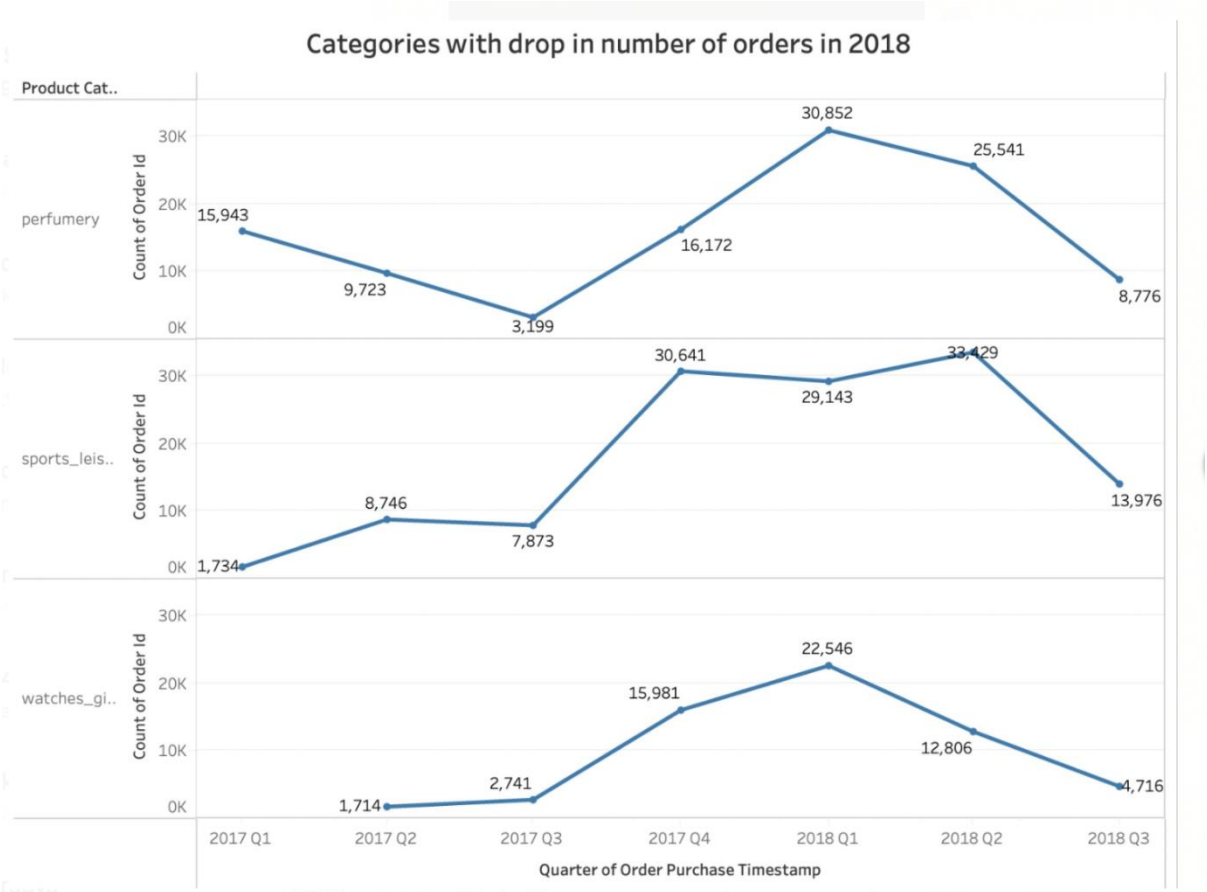
In the figure we observe that the number of orders for the "health and beauty" category significantly decreased from 63,369 in 2018Q2 to 51,122 in 2018Q3. This was in sharp contrast to the consistent 50% quarter over quarter growths from the past quarters.

Starting from 2017Q2 this segment had seen consistent growth in the number of orders and was growing at close to a 50% growth rate per quarter.

Event Correlation

The drop in number of orders was not consistent across all the categories in our data. A similar amount of drop was also observed in categories like "sport leisure", "perfumery", and "watches". The categories in which we observed this drop usually had >10,000 orders in 2018Q1-Q2. We did not

observe this drop in categories which had <10,000 orders. This observation can also be seen in the below figure.

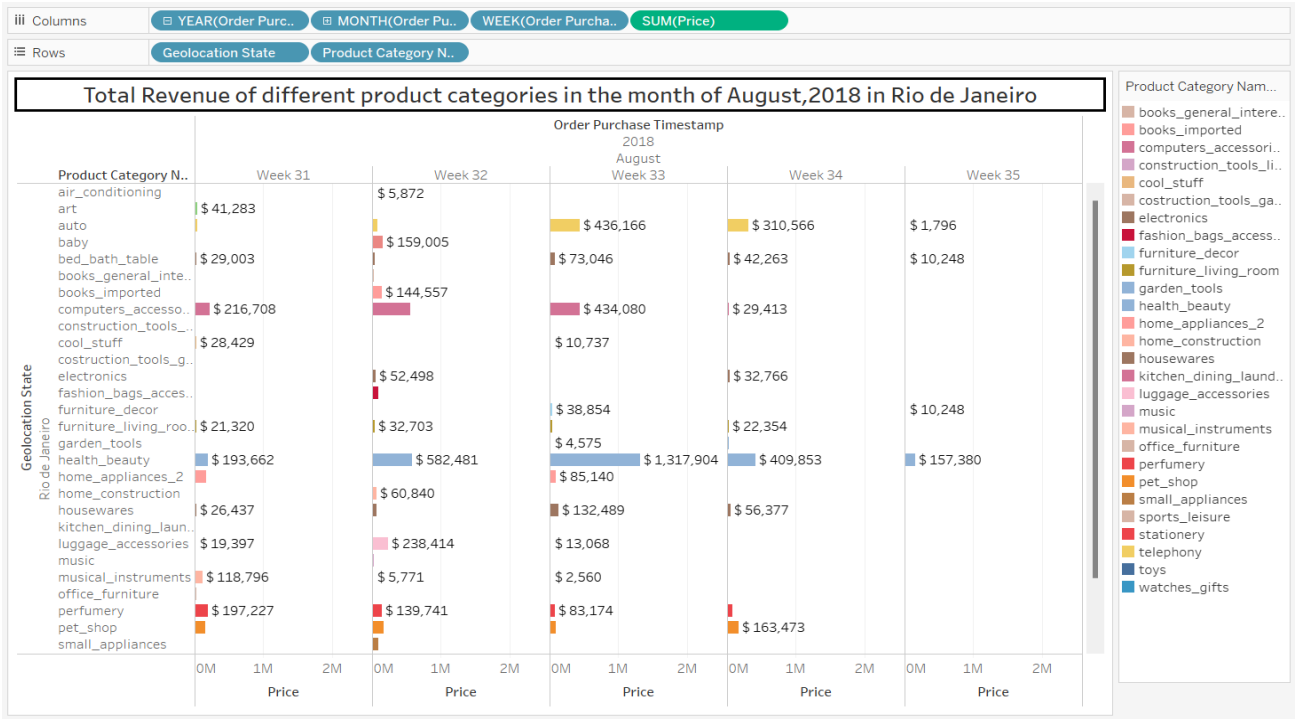


Analysis

Looking at our observations, the drop in number of orders occurred in cosmetic and fashion related categories. The drop also occurred in categories which had a significantly high order count compared to other categories. Brazil's GDP also dropped in 2018 from 2T USD to 1.7T USD. This drop was significant compared to the growth in GDP from 2016 to 2017.

From these observations we can hypothesize that the shrinking of Brazil's economy in 2018 had a significant hit on the sales of cosmetics category products. One possible way to account for this reduction for the coming years could be to reduce the amount of inventory of products from cosmetic categories or to focus on providing discounts and promotions for products from these categories to potentially increase their sales.

Analysis 2 - The weekly sales of different product categories in the month of August, 2018 in Rio De Janeiro.



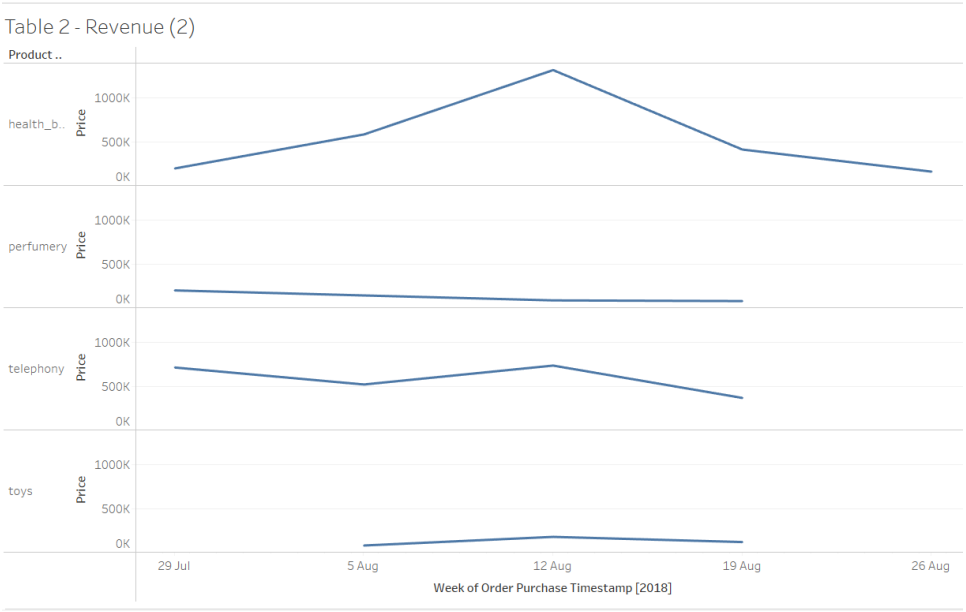
In the above image, we have shown the weekly sales of different product categories in the month of August, 2018 in Rio De Janeiro.

Issue

By looking over the data, we can see that the revenue generated by all the different product categories in Rio De Janeiro dropped significantly in the 4th week, which is week 35, suggesting that people usually don't spend much during the later end of the month.

Event Correlation -

If we just take the sample of some of the top revenue generating product categories, and compare them using the line charts, we can see that almost each one of them almost negligible sales during the last week of August.

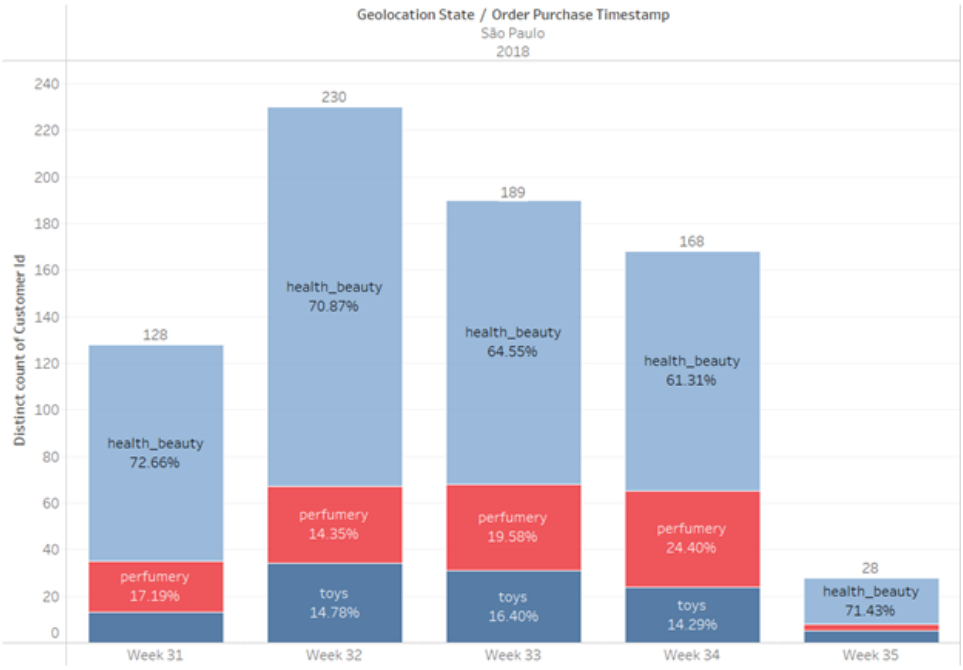


Analysis:

Since the Brazil's economy already started shrinking from 2 trillion to 1.7 trillion, the significant drop is evident through the purchasing power of people. Most of the people didn't spend much during the last week of the month because usually it's the time to pay for their credit bills and other loans.

It's important to understand the need of adjusting the inventory size for different products based on the demand of the market. From the vertical bar graph, we can see that the demand is mostly for the product categories like health beauty, furniture, auto, furniture decor.

Analysis 3: % Comparison of different products based on unique customer visits in Sao Paulo



The stacked bar chart above shows the percentage of comparing three product categories that are health_beauty, perfumery, and toys, based on unique customer visits in Sao Paulo for August 2018.

We can observe that Week 32 represents the highest number of unique customer visits by 230, followed by Week 33 - 189, then Week 34 - 168. In August 2018, the least unique customer visits were Week 35 by 28 on the right.

The graph shows that the preferred product category for unique customers is health_beauty, the most significant percentage compared to perfumery and toys.

Although there was a significant improvement in unique customer visits between Week 31 and Week 32, from 128 to 230, the bar chart indicates the decreasing number of those through Week 33 to Week 35.

It is noticeable that there is a problem in Week 35, so a brief problem statement can state as "Unique customer visits in Week 35 was not able to meet our expectation".

"5-whys" root-cause analysis

Problem: Unique customer visits in Week 35 was not able to meet our expectation

Asking why this problem is occurring to analyze as below:

1. Why?

Answer 1: The significant decrease in new customer visits is aligned with the rise of the national unemployment rate that started simultaneously.

2. Why?

Answer 2: Our store did not proactively respond to curate marketing efforts to accommodate the low consumer purchasing power due to the declining job market.

3. Why?

Answer 3: The store did not foresee the impact of the shrinking consumer market and did not have a plan B for the deteriorating market.

4. Why?

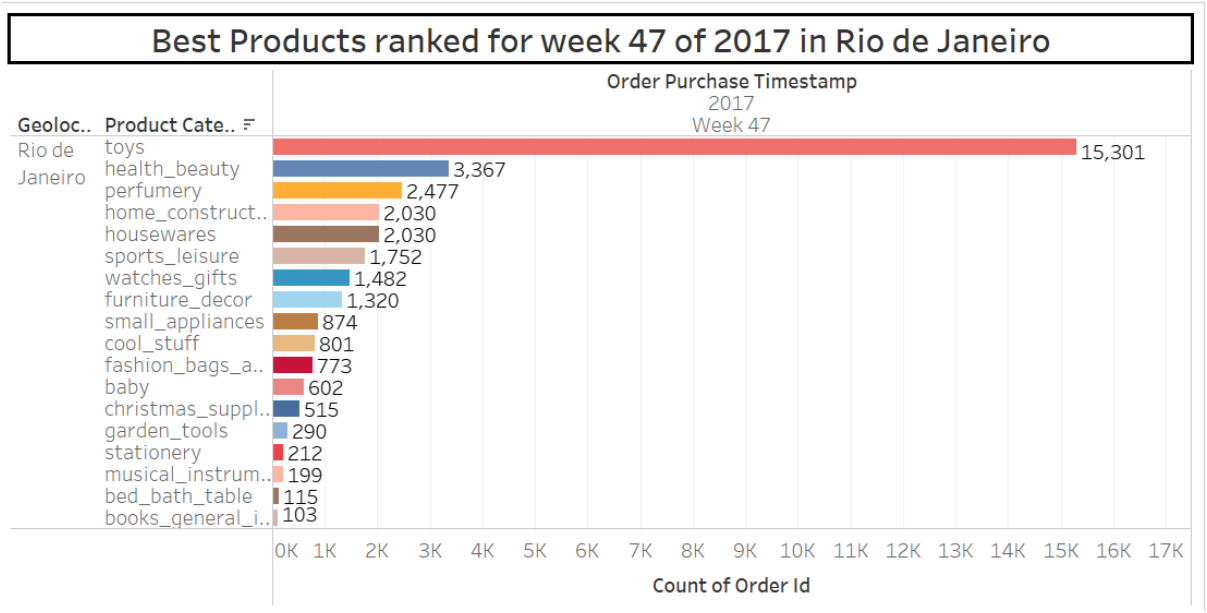
Answer 4: The current strategic marketing practices did not consider national economic predictions, and plan B was not readily available for deteriorating market confronted.

5. Why?

Answer 5: The current marketing strategic capacity had no experience forecasting unseen situations and could not respond promptly and effectively.

Overall, it is vital to take immediate actions to improve performance to cope with the current job market downturn. The efforts should thoroughly examine all the marketing options to overcome consumer resistance due to low purchase power. The traditional marketing strategies are not likely to work in the same way. It requires very innovative measures that will be effective for an extended period if the economic downturn should last longer than expected.

Analysis 4: Best Products ranked for week 47 of 2017 in Rio de Janeiro



In the above figure we have shown the count of orders for each category in the week 47 in 2017 in Rio de Janeiro. Using this data, we can know which products were sold more and less in that week.

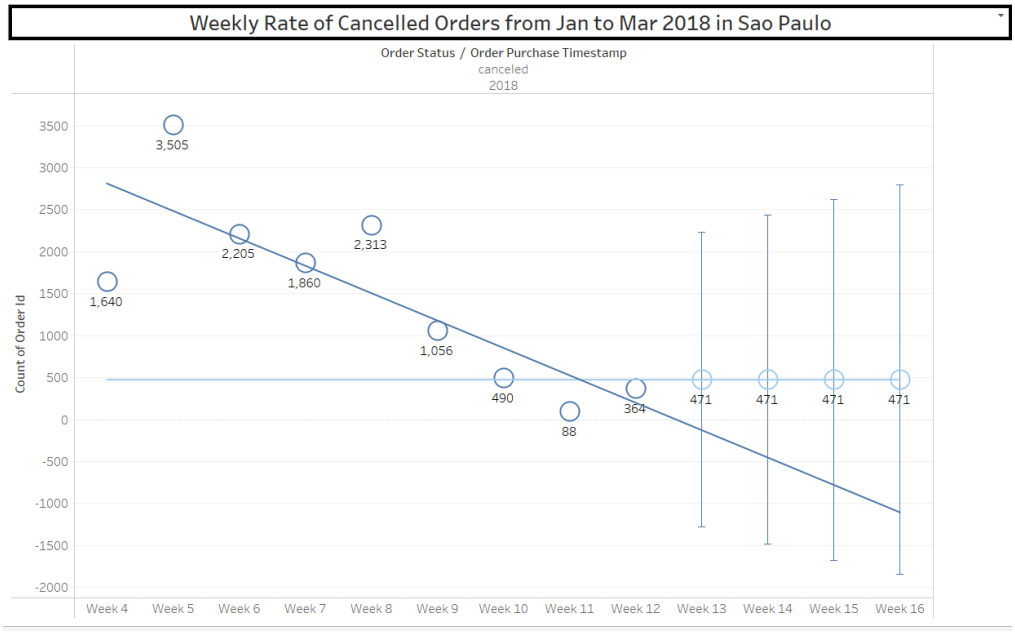
Issue

In the figure we observe that the number of orders for toys were sold more than 400% than other categories. Also, we know that for the last week of November, in many countries the holiday Black Friday is celebrated. So, as we known this information, we need to create a marketing campaign to sell more products in the categories which has lowest sales.

Analysis

Looking at our observations, we can notice that the categories which has lowest sales are Christmas supplies, garden tools, musical instruments, bed & bath, and books. Furthermore, none of those categories rise above the average which make us take the decision to create a marketing campaign to increase those sales in quantity and sales.

Analysis 5: Weekly Rate of Cancelled Orders

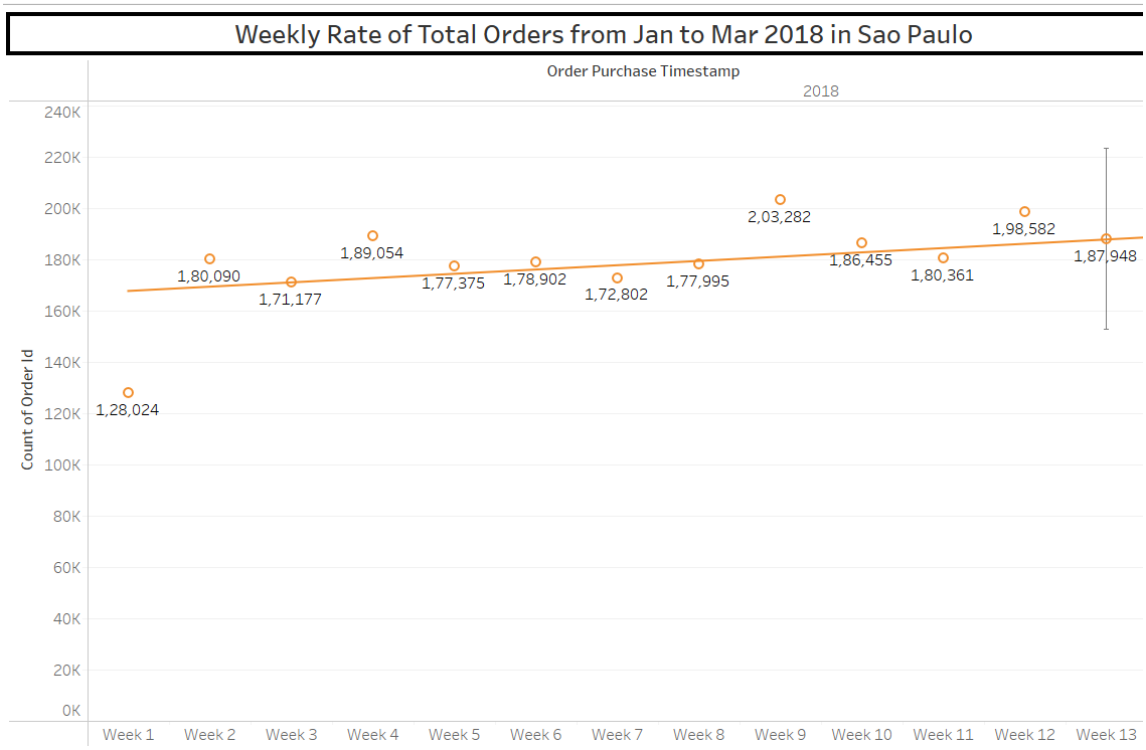


Problem: The above graph illustrates the weekly rate of cancelled orders from January to March 2018 in Sao Paulo. This could give us a general understanding about how satisfied customers are with our services and products.

Analysis: The above graph illustrates the weekly rate of cancelled orders from January to March 2018 in Sao Paulo. From week 3, the number of orders cancelled saw a sharp increase from 1640 to 3505 in week 4. However, following this there is a steady decline all the way to week 11 hitting a minimum of just 88 cancelled orders. Finally, for week 12 the number of cancelled order sightly increased again to 364.

There are some fluctuations in between as seen in week 8 and week 12 but overall, there is a constant negative trend through the weeks. This trend emphasizes the facts that the quality of our services is increasing, and customers are facing less issues as the weeks follow, Furthermore, the graph also depicts a forecasted rate of cancelled orders of 471 for the following 4 weeks with increasing range of confidence intervals. The forecasted value denotes that there is going to be a slight but constant increase in our number of cancelled orders. This information could help us take necessary measures to prevent this.

Event Correlation: The above graph illustrates that there is a negative trend in the number of cancelled orders which shows that people are overall satisfied with out services and product. However, this decline could also be attributed to the fact that the overall sales itself has declined for the period. So, to find out if people are happy with our products, we have plotted the weekly rate of total orders placed between the same period.



The above graph depicts a positive trend in the rate of the overall orders placed by customer. This enforces the fact that the quality of our services and products has improved.