

The Olist Commercial Dataset

A geographical analysis of customers

by Alin Cristian Preda

Teachers

Lester Allan Lasrado & Benjamin Flesch



Copenhagen Business School

Business Administration and Information Systems,
Data Science

2018



olist
empowering commerce

Table of contents

Annex

Introduction.....01

Conceptual Framework and Methodology.....02

Results.....05

Facts, Insights and Outcomes.....14

Conclusion and Recommendations.....15

References and Links

Appendix

ABSTRACT

WHAT WAS THE TOPIC?

Is prompt delivery of purchased products a staple of customer satisfaction in e-Commerce?

WHAT WAS THE DATASET?

Olist provided a dataset containing an almost two years period of information about its sellers, buyers, orders, products, customer reviews and location data.

WHAT ARE THE MAIN TOOLS AND METHODS?

The main tools are Alteryx, for cleaning the data and preparing it, and Tableau, which was used to develop dashboards. The method used for analysis was the observation of the data visualizations.

WHAT WAS THE PROBLEM FORMULATION?

Is Olist doing a good enough job at shipping its seller's products to the customers?

WHAT WAS THE RESEARCH QUESTION?

Why do customers give low reviews? Are there differences between regions, states, cities?

WHAT WERE THE CONCEPTS?

e-commerce, customer satisfaction, delivery times

WHAT ARE THE MOST IMPORTANT RESULTS?

Customers are affected by long delivery times and delays. More incoming orders correlate with slower delivery times. Statistically speaking, the customers in each region act almost identically.

WHAT ARE THE CONCLUSIONS AND RECOMMENDATIONS?

Olist could make more shipping free of charge; it could contract more couriers to take the load off, especially on busy periods; it could offer more shipping solutions.

LINK TO DASHBOARD

<https://eu-west-1a.online.tableau.com/#/site/cbsvisualanalytics/views/AlinCristianPreda/Dashboard?:iid=1>

KEYWORDS

Brazil; e-Commerce; Transportation; Shipping; Logistics; Customers; Customer satisfaction;

INTRODUCTION

This research takes a closer look at an online marketplace for local retailers in Brazil. The company Olist has provided the public with a dataset containing commercial information on customers, sellers, orders, products from which we can understand how they interacted with each other. I will attempt to sketch the overall face of Olist and its customers, products and sellers.

This study is most significant to the executives of Olist. It best suits the interests of the marketing and customer relations departments. The conclusions drawn from this analysis could enable the management to better understand their business. So much so that it can strengthen Olist's bonds with its partners and customers. The study can also be of interest to other companies (possible competitors) in Brazil because it tells an interesting story about the demographical and geographical influences the country can have over a business operating in on-line marketplaces.

I decided to use this dataset because it's from a real company, with a novelty business model, who has been operating for a short period. I was attracted by the complexity of having almost a dozen datasets connected to each other. Finding the right connections between them and then ultimately uniting them in an all-encompassing database proved to be a satisfying learning experience.

Olist is Brazil's largest department store within the marketplaces. It offers an online platform where small and large businesses from all over the country can sell their products, with a single contract. The sellers can market their products directly through Olist Store and use the company's logistic partners to ship them to the customers. For business owners, looking to sell, Olist offers the opportunity to benefit from great reputation, qualified traffic, great visibility and higher ranking in search engines. Using their services, small or new businesses can profit off of the reputation and visibility of a large community of well-established sellers. Regarding shipping, Olist offers an exclusive partnership with the post offices, boasting more attractive freight values. The company promises shopkeepers a solution to enable them to sell more and better and attract new customers.

There are, as always, unsatisfied customers or areas where the company could do a better job. The main research questions aim to uncover the strengths, weaknesses and opportunities available to Olist. What is behind the bad reviews that some customers gave? And is there any difference between the commercial activity in one region and the others?

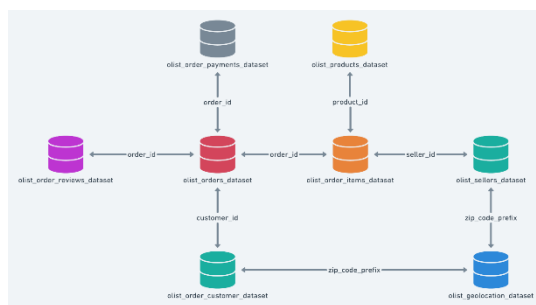
METHODOLOGY

DATA COLLECTION: METHODS AND TOOLS

The data was found on Kaggle, where it was freely publicized by Olist. The datasets are in .csv format and ready for download. Additionally, I enforced the dataset with pieces of information found in other sources. I attached the regions to the geo-spatial data and I introduced the super-categories, made of my own design, in order to restrict the spectrum of observations.

DATASET DESCRIPTION

The research was done on the dataset “Brazilian E-Commerce public dataset by Olist” available on the website Kaggle (<https://www.kaggle.com/olistbr/brazilian-ecommerce/home>). The data has been updated or modified multiple times during its history. The version of the dataset is Version 7, updated on the 1st of December. It has actual anonymized commercial data ranging from 12/30/2016 2016 to 9/24/2018. It is organized in multiple specialized datasets that can be merged together in different ways, according to the research topic: orders; order items; order payments; order reviews; customers; sellers; products; geolocation and product category translation. The customer dataset contains information about the buyers and their location. The geolocation dataset contains Brazilian zip codes and coordinates. The orders file stores information about what was purchased.

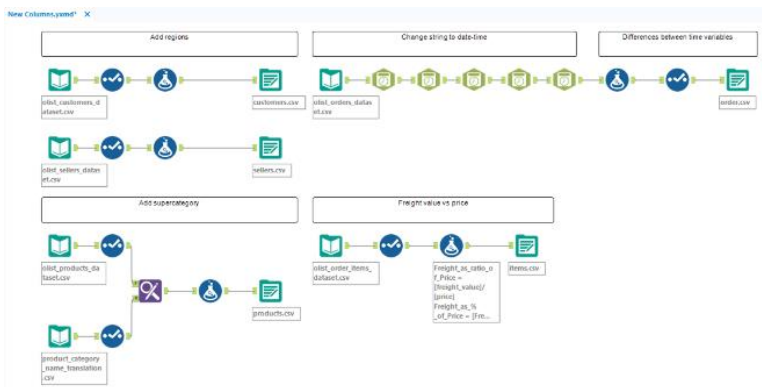


The following schema was provided by Olist. It illustrates all the files available and the relationships are based in common keys or ID's through which they can be merged. I have opted for creating an ultimate dataset which connects all of these into one pool.

DATASET LIMITATIONS

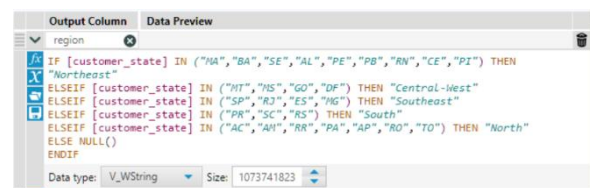
Olist's business is relatively new to Brazil and this is why the dataset is very limited from a temporal point of view. Attempting to identify trends in time series or predicting future KPI is not recommended with having such limited information. But there is still a lot of relevant data to be studied. We can interpret this information such that we can understand the company's current situation and assess the progress that's been going on so far.

DATA PRE-PROCESSING: METHODS, TOOLS AND TECHNIQUES

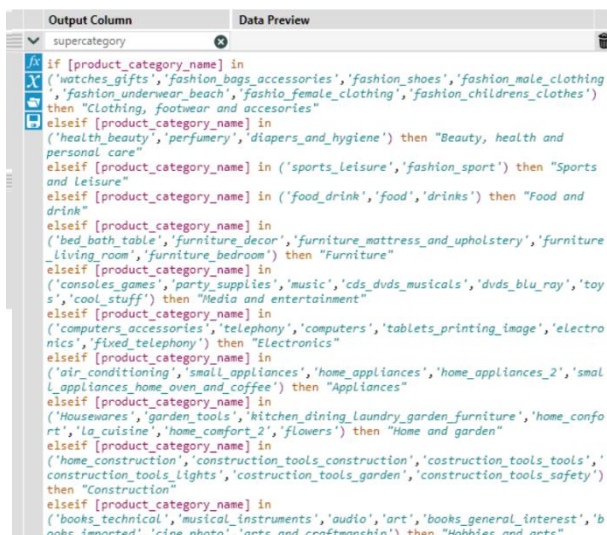


The processes of cleaning, filtering, transforming and combining the pieces of data was done with the help of the amazing Alteryx Designer. The tool came in especially handy in the process of loading and transforming.

Firstly, after loading the .csv files, I used the select function to extract those variables/columns which seemed most useful. Using the select, I also changed the data type of most of the database's members. Both the customer and the seller datasets contained fields about the state, city and ZIP code prefixes. I added the regions using a simple IF statement.



Next, using the Find and Replace tool, I was able to make use of the translation file in order to change the Portuguese product category names with English ones. The next step was to construct a "Supercategory" column, useful for a broader study of the products.

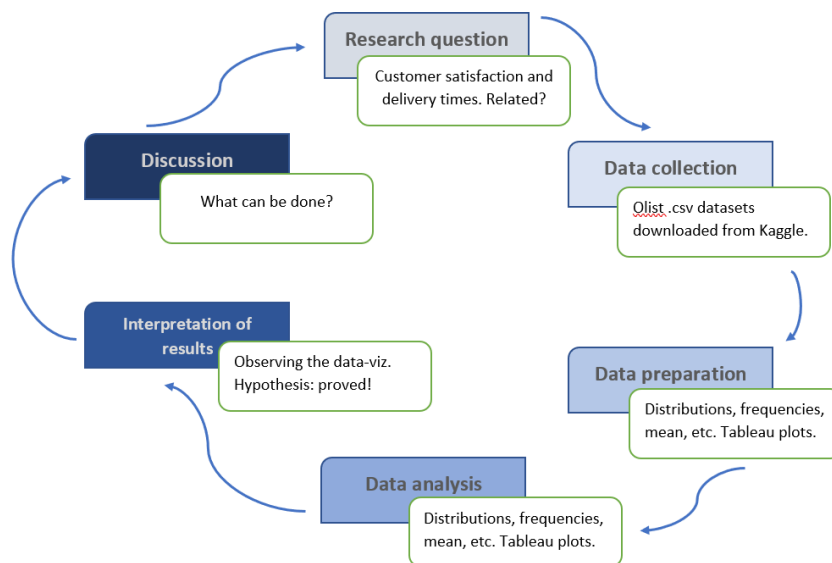


Suspecting the value of the freight price might play a part in the review of the product and also because I believed that the North region, being more isolated would have higher freight values, I used another formula. This one calculates the % of the freight in the price of the product. The data representing timestamps was stored as strings in the .csv. This is why I used the DateTime function to convert the fields into the appropriate format.

```
DateTimeDiff([estimated_delivery_time],[delivered_customer_time],"days")
```

DATA ANALYSIS PROCESS

DATA ANALYSIS PROCESS DIAGRAM



DATA ANALYTICS: MODELLING, METHODS AND TOOLS

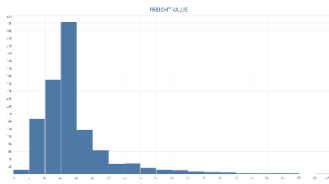
Alteryx and Tableau are the tools of choice in this project. Alteryx was used both in the Data Preparation stage but in the early Data Analysis as well, for getting a good feel of the metadata and of the raw data through summary statistics such as frequencies, mean, median, totals, min, max, etc. After preparing the data and forming a better understanding of it, visualizations were created. Some charts directly attempt to answer the research questions. Other were made to explore which lead to new questions and new answers, some of them uninteresting, some irrelevant, and some that proved surprising.

Ultimately, an interactive Tableau dashboard was created which could spark a discussion about the problem and how the company can choose to act based on this research. By making use of filters and highlighters, the dashboard can enable management to pin-point customers, geographical areas or moments in time based on different combinations of variables that describe the ‘target’.

When analyzing complex datasets of considerable size, it is often recommended to make use of data visualizations in order to better understand the subject matter. Patterns can emerge and subsequent eureka moments can give rise to better decision-making.

RESULTS

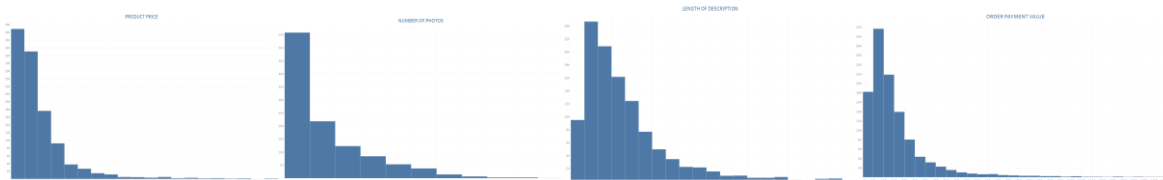
EXPLORATORY ANALYSIS



The freight value follows a right-skewed, positive distribution. That is, most values present themselves towards the lower end. According to the bins, most freights hold values between 10-20 R\$. As a percentage of price, most values situate are between 10%-20%.



Price also follows a positive distribution. Most products cost less than 100 R\$. Very few of them cost more than 400 R\$. Regarding the number of photos in the product's description, we can see a right-skewed histogram, with most of them having just one photo. Most descriptions are short, between 200 and 400 words but some of them can reach over four thousand words.



All of the histograms continued with bins to the extreme right, depicting outliers, which I chose to exclude from the visualization and discard as irrelevant. The seams to be clean. The distributions seem to be normal, considering the nature of the products sold on Olist.

There are 96,096 unique customers. There are 99,441 customer ID's in the 98,666 orders, which means that 3,345 of the buyers engaged in Olist's services more than once. That is just over 3%. The customers come from 4,119 different cities. There are 14,994 different ZIP codes. There are 3,095 sellers associated with Olist. They are spread out in 23 of the 27 states and are based in 611 cities, over 2,246 different ZIP codes. Olist's sellers boast 32,951 different products, viewed in 70,794 photographs, spread over 73 product categories. From these 73 product categories, I managed to synthesize a much shorter list of just 11 product super-categories. When people request an order on Olist, they order between 1 and 21 items at a time. The product's average price is 120.7 Brazilian reals, R\$. The median of prices is 74.99 R\$. The lowest valued product is just 0.85 while the most expensive one costs 6,735 R\$. There are 5,968 different prices. In total, Olist managed to sell products worth a total sum of 13,951,644. In total, the customers paid 2,251,910 in freight value. The average freight is 19.99 R\$, with the median being 16.26 R\$. There are 6,999 different values for freight. The least a freight can

cost is 0 while the most expensive one is 409.7 R\$. There are five types of payment available on Olist. A maximum of 24 payment instalments is accepted. Sequential payment is also possible, which means that a customer can choose to pay for a purchase using multiple payment methods. In total, Olist managed to sell products worth a total sum of 16,008,872. The biggest order managed to attract 13,664 R\$. The orders averaged 154.1 R\$, with a median value of 100 R\$.

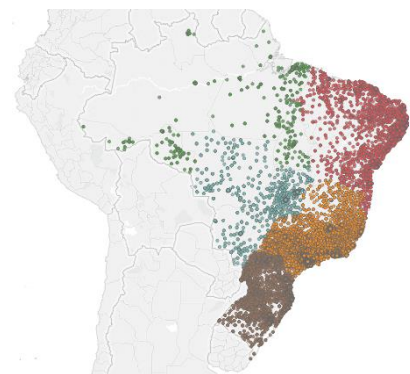
In order to help me simplify the data, I added information about Brazilian administrative regions and I created super-categories for the products. The final merged dataset contains information about 95k customers from 15k zip codes, 4.1k cities, 27 states, 5 regions. There are 33k products divided into 73 categories and 11 product super-categories described in over 70k photos. There are records of almost 100k orders with 100k reviews. Payments were possible through 5 types of payment and a maximum of 24 installments.

REGIONAL ANALYSIS



According to Wikipedia: **NORTH** Transportation is mainly done via rivers, of which the region has many. The very few highways are mostly in the east. The cities are spread very far apart. Sometimes airplanes are needed to access remote areas of the region. **NORTHEAST** Transportation is done mostly via the many highways or by sea, along the coast. **CENTRAL WEST** Transport is done by some highways mainly in the east

and centre, and the rest of the area is taken care of by nautical transport and sometimes airplanes. **SOUTHEAST** It is covered by an extensive network of highways and, to some extent, railways which make transportation fairly easy and fast. This region is both the most populous and the richest, being responsible for almost half the country's GDP. **SOUTH** This region also has good highways and railways with the possibility of transportation via rivers, when necessary. The second map shows the distribution of customers over different Regions, marked with a different colour.



Almost 7 out of 10 of Olist's customers come from the Southeast. Most of them, from Sao Paulo, Rio de Janeiro and MG. This result was to be expected, given the population of the

The figure consists of four bar charts arranged in a 2x2 grid, illustrating the regional distribution of customers and sellers based on their payment method.

Top Left Chart: Region by No. of Customers

Region	No. of Customers	Percentage
Southeast	~650	66.50%
South	~150	15.14%
Northeast	~100	9.52%
Central-East	~70	6.83%

Top Right Chart: States by No. of Customers

State	No. of Customers	Percentage
UT	~400	3.87%
NJ	~140	1.36%
SD	~130	1.26%
RI	~80	0.77%
PA	~70	0.68%
NC	~60	0.58%
VT	~40	0.39%
IL	~30	0.29%
NY	~30	0.29%
DE	~20	0.19%
PR	~10	0.10%
GA	~10	0.10%
MO	~10	0.10%
WY	~10	0.10%
CO	~10	0.10%
AK	~10	0.10%
HI	~10	0.10%
MT	~10	0.10%
ND	~10	0.10%
NE	~10	0.10%
KS	~10	0.10%
OK	~10	0.10%
TX	~10	0.10%
LA	~10	0.10%
AR	~10	0.10%
MS	~10	0.10%
AL	~10	0.10%
SC	~10	0.10%
GA	~10	0.10%
FL	~10	0.10%
CA	~10	0.10%
OR	~10	0.10%
WA	~10	0.10%
MT	~10	0.10%
WY	~10	0.10%
CO	~10	0.10%
NE	~10	0.10%
KS	~10	0.10%
OK	~10	0.10%
TX	~10	0.10%
LA	~10	0.10%
AR	~10	0.10%
MS	~10	0.10%
AL	~10	0.10%
SC	~10	0.10%
GA	~10	0.10%
FL	~10	0.10%
CA	~10	0.10%
OR	~10	0.10%
WA	~10	0.10%

Bottom Left Chart: Region by Payment Value

Region	Payment Value	Percentage
Southeast	~1250	64.57%
South	~350	18.76%
Northeast	~250	11.89%
Central-East	~150	6.75%
Other	~50	2.63%

Bottom Right Chart: Region by No. of Sellers

Region	No. of Sellers	Percentage
Southeast	~2200	72.9%
South	~650	21.6%
Northeast	~50	0.8%
Central-East	~20	0.1%

WHICH STATES MADE LESS THAN IN 2017?

State	Percentage
SF	38000
RI	12000
ME	11000
ES	10000
KS	9000
PW	8000
SC	8000
BA	7000
PT	7000
CT	6000
MA	6000
PS	5000
HI	4000
AK	4000
SD	3000
CO	3000
ER	3000
MT	2000
MS	2000
PA	2000
TN	2000
ND	2000
AR	2000
AC	2000
AP	2000
OR	2000

PAYMENT METHODS ANALYSIS

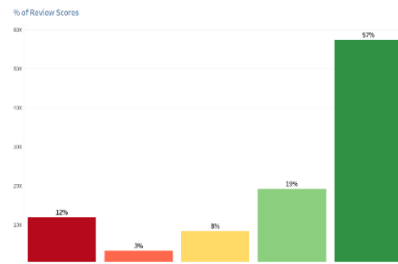
Payment Type

Payment Type	Percentage
boleto	20%
credit_card	76%
debit_card	3%
voucher	2%

7

having an international credit card, I think Olist could reach more of them if it decided to expand and include international marketplaces to its platform. Boleto Bancario payers are 1 in 5 customers already. I would suggest that every business should integrate the newest, most popular payment methods, for convenience's sake. There were no significant results between the region groups, all of them recording similar % for each type of payment method.

REVIEW RATINGS ANALYSIS

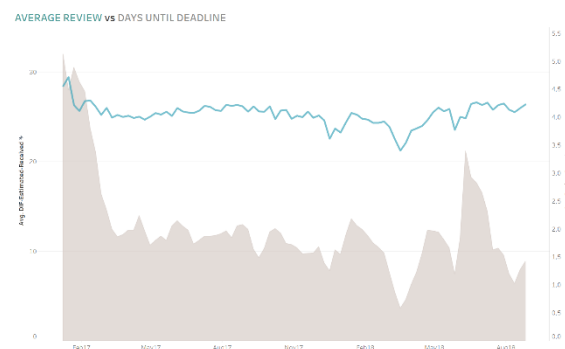


Most customers seem to be satisfied. Indeed, 6 out of 10 orders were rewarded with a review of five stars, 2 out of 10 with four stars. That means 80% of the orders were perceived positively by the buyers. If we take into consideration the fact that the review rating scale has five steps, then a review of three stars can be considered neutral and a two or a one is a negative review, hence the colour choice.

There are not many neutral reviews, very few of two stars and 12% of orders received the lowest possible score. Overall, 15% of the orders were perceived as a negative experience. What can we learn from these low reviews? Is it something that has to do with the products or the transport? There are no significant variations of review preferences between the regions.

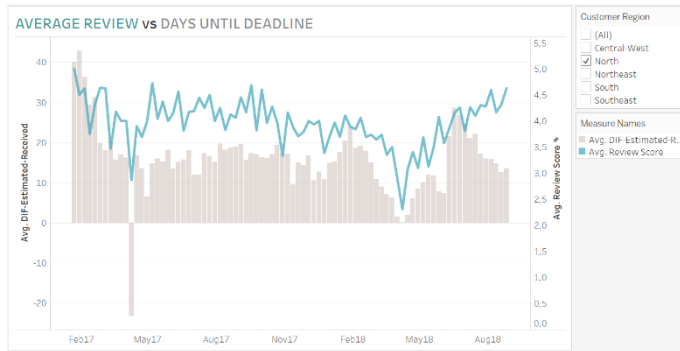
DELIVERY ANALYSIS

The dataset made available various timestamps: from when the order was recorded, to when it was approved, when the product was handed in to the courier, to the actual delivery and including an estimate of the maximum date for the arrival time.



This chart uses an unsynchronized dual axis to analyze the relationship in time between the average review scores and the days remaining until delivery deadline (Estimated delivery time – Actual delivery time, measured in days). Indeed, it seems that there is a somewhat common trend. Although there is strong

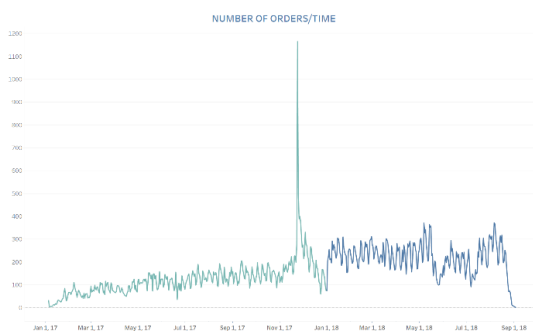
correlation the performance of the couriers cannot fully explain the reasons for all the ratings. A future study using sentiment analysis on the comments written by people who gave low ratings might reveal some patterns and shed light on what might be the main causes of concern.



This becomes more evident in certain periods or when we isolate certain areas. For example, here, filtering in the North, we see a pattern emerging. Coincidentally, the review scores suddenly got much lower while at the same time, the delays skyrocketed. The

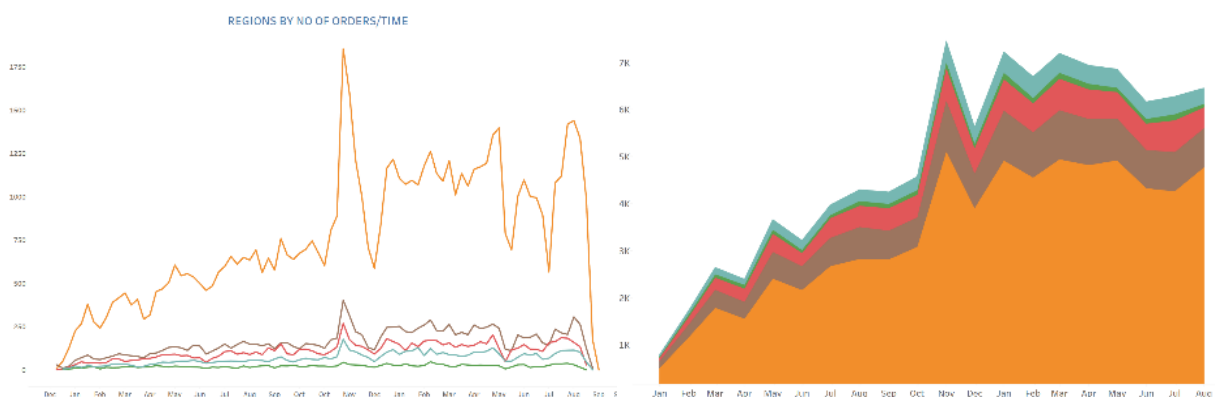
dates are April 17, 2017 and April 18, 2018. Here, the reviews got as low as 2.8 and 2.3 respectively, for average days until deadline of -23 and 0.11 respectively. Although we see an improvement in delivery times, the feedback was worse than last year. Perhaps people remember. This strange coincidence has a low probability of appearance of 1 in 365 or 0.27%. It asks for a more thorough analysis of what happens in the spring quarter. The phenomenon should be expected in 2019. Similar trends appear for all the other regions, reinforcing the fact.

ORDERS ANALYSIS



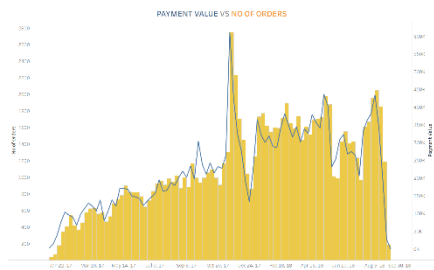
The peak reached by orders/week was reached on Friday, Nov 24, 2017, which was Black Friday in Brazil. 2016 was excluded from the purposes of this study dew to the limited amount of time series data, we can see no seasonality developing but we can expect the same volume of sales in November

2018. Color, light and dark blue, was a filter, was used to highlight the passing of the years.

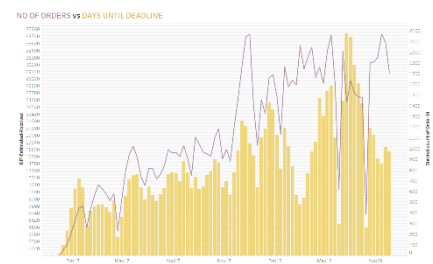


The line chart studies the evolution of orders/week. At a first glance, all regions have overall the same consumer patterns. Lines are best for showing time series. Seasonality would affect each region almost the same, although to very different degrees. The stacked area chart shows

the orders per region from month to month. The areas help the viewer understand the volume of orders, putting the relative differences between regions into a better perspective.



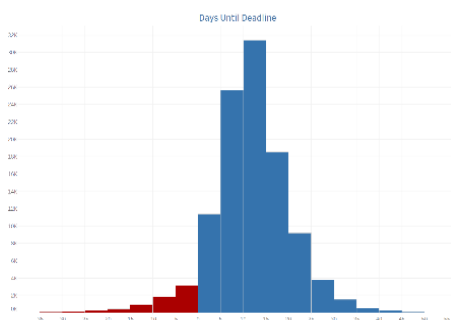
As expected, according to the dual axis plot, the sales amounts are closely tied to the number of orders. But there is also something



contradictory in the data. From this second dual axis chart we can see that, quite expected, when there are many active orders there is also great delay in delivery times. This is generally true except for, most notably, again, spring 2018, when there were both many orders and a period of fast shipment. This means that the couriers sometimes become overwhelmed.



This map explores the cities with the greatest delays. Blue arrive very early, gray is average and the warm colors represent high delay times. We can pick out the problematic cities for further inspection. The method can be further expanded to the level of each individual customer, based on their location.

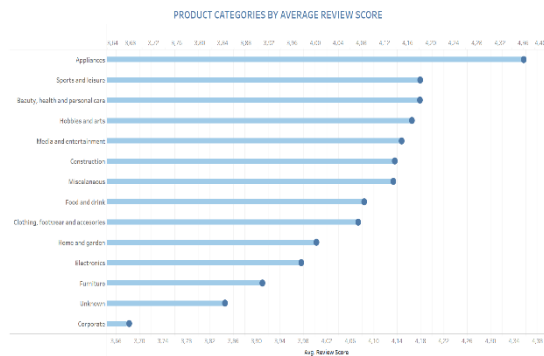


This histogram shows that most orders are actually delivered very early. It would be right to assume that Brazilians value their time dearly and would like to get a hold of their products as fast as possible. In order to appeal to them, courier services must be improved. Faster deliveries means faster money-making.

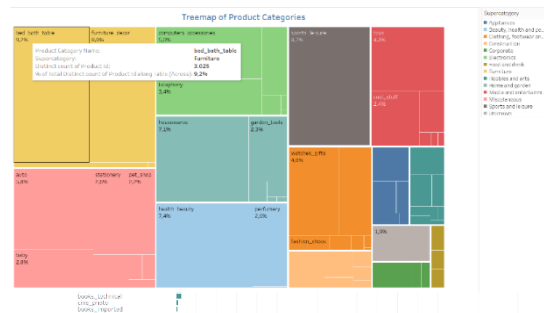
PRODUCT ANALYSIS

The most unpopular product super-category is Corporate. This is an issue, as it means that our corporate clients are not satisfied with the quality of our products and/or services.

Corporate clients are important, as they have greater purchasing power and can place greater orders. Losing the interest of this client segment could cause mayhem to our overall sales.

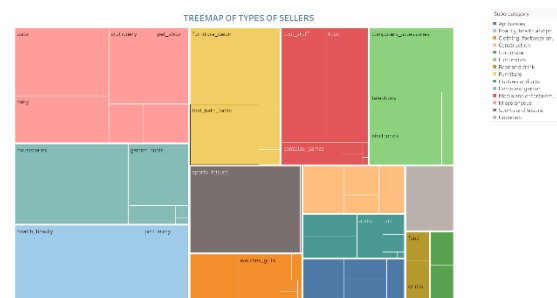


I suggest that a separate study should be made. Interestingly, furniture and electronics, some of the most sold products, merely make the bottom of the review score list with average scores of not even 4. The dotted bar chart with visible dual axis excluding zero, was chosen in order to help the viewer find the exact score more quickly.



Both the nested bar chart and the tree-map reveal the structure of the products. The bar chart deals with sales/category and the tree-map showcases the number of purchases from each category (the product popularity). Thus, the company's marketing department can know which products

sell the best and which ones need an extra push in order to get those areas selling more. A future analysis of the shopping cart would reveal if certain types of products are often-times bought together. Information such as this could help with targeted ads on the website. This could lead to at least an increased level of customer engagement, or, why not, to an increase in sales of complementary products.

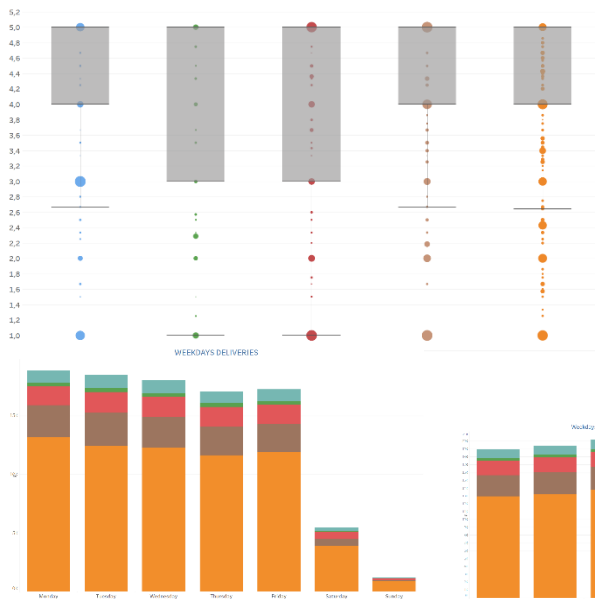


The second treemap showcases the product super categories and category names according to the number the sellers that have them available. As we can observe, there are visible differences between the strongest and weakest products from each category. We can also study how each product

super category and category sold over time. (ANNEX)

CUSTOMER ANALYSIS

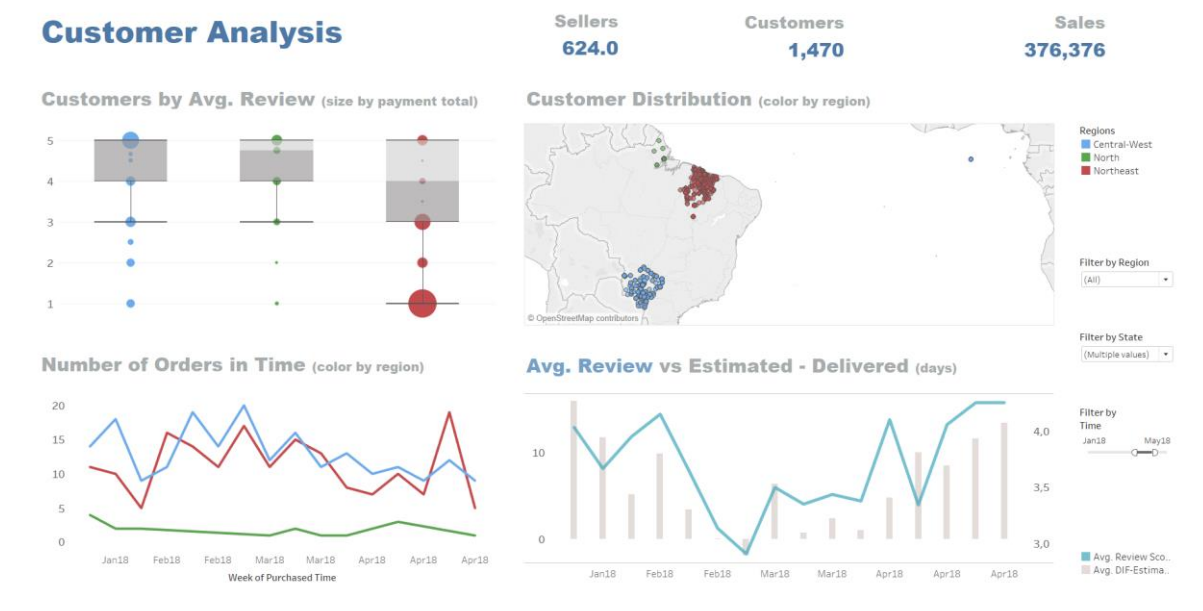
As can be seen from the box plot, some of the biggest clients to date (the size of the circle represents the sum of the payment values) have low average review ratings. These clients with payments in tenths of thousands and extremely negative reviews should be inspected more thoroughly. Our goal should be to attract as many clients as possible. The clients in the North show a wider array of review scores. The south has the most outliers, with extremely low scores compared to the majority.



Big clients should be respected as such and I suggest that their satisfaction be prioritized. One way to address this issue is to give them special attention in terms of customer service, targeted marketing and especially delivery services.

People tend to issue the most purchases on Mondays and decreasing slowly towards the weekend, with more people issuing orders on Saturday than on Sunday. Regarding the delivery times, again the trend is similar. Most products arrive on Monday and decrease until Thursday, with a slight increase on Friday. Very few products reach their buyers on a Sunday. On the other hand, the estimated delivery times never include weekends.

DASHBOARD



The dashboard is the quintessence of the research work, as it those visualizations that can help the viewer decipher the answers to the most important research questions, while at the same

time offering sufficient freedom to explore the dataset in new and innovative ways. I thought of the audience as the decision makers involved in Olist's marketing and customer relationship. What I am trying to achieve with it is transmit the message that Olist could do a better job with its transportation services. It's not doing horrible, by any means, but it could do better and sometimes it seems to choke under the pressure of its own sales success. Customers with problems should be pinpointed using the dashboard and their case should take further interest.

The color scheme is relatively simple. I used the same palette for the regions, for consistency across the report. Dark blue is used highlight important features. Grey is used for writing secondary information. Font and size are also used in this way. In the bottom-left chart, I also used blue in the title in order to help users identify the line chart more easily. The worksheets and other objects were added in tiled version, vertically or horizontally. Blank spaces were used around charts. To the right of the title, information about the number of sellers, number of customers and the actual sales values is provided for any given filtering or selection.

Four worksheets were selected, not too few and not too many, to confuse the viewer, who can focus on only a handful of elements at once. I have chosen to introduce the box plot because it enables the viewer to easily identify persons of interest based on their value as a customer and on their overall experience with Olist. The chart can be used for highlighting. For example, we can select only those customers with negative reviews. The dual axis line/bar chart was chosen because it showcases how reviews are influenced by long shipment times. The map is relevant because Olist's management is familiarized with the landscape of their business. It shows us that the Southern customers are all tightly grouped in a high-density area. Northwest shares similar characteristics on its coastline but many customer locations become scattered to the west and center, much like the Nordic regions and the center. This is consistent with the information found on Wikipedia. Selecting customers from the map highlights them. The map is fixed, so that it zooms in detail.

All the charts are dynamic. They can change dramatically through use of filters and highlighters. First, there are the geolocation filters: regions, states, allowing the user to drill down wherever interest rises. The filters allow for multiple choice of selection, which encourages comparisons. The regions legend can be used for highlighting. Secondly, there is a range filter for time which works on the time series charts. It can be used to study certain moments in more detail.

FACTS, INSIGHTS AND OUTCOMES

MEANINGFUL FACTS

Business is doing good, there is an overall increase in sales and customers. Each region is doing as well as its economies do. The size of the business in each one seems to perfectly relate to the geo-demographical characteristics and the infrastructure of Brazil itself. The average review score is positively tied to the delay time of the couriers. Spring was the weakest quarter for the past two years and the two events seem to be cyclical in nature. Many of Olist's biggest clients to date have had a bad experience with its services for some reason.

ACTIONABLE INSIGHTS

The Southeast makes up the bulk of the business while the North pays out the least, as expected. The company should try to watch out for spring 2019 by coordinating more gracefully with the courier services and communicating them the problems that have been experienced in the past.

Olist would be wise to try to connect the markets with the customers more efficiently. It should establish additional contracts with different courier service providers that are specialized locally, in order to deliver its products faster and offer them more shipping options.

Olist should try to stop adding a freight value to each and every product. A cost vs benefit analysis could possibly establish if such a step would be beneficial to the company because it would certainly suit the customers.

According to a paper published by AlixPartners from March 2018, "Consumers are increasingly seeking the convenience of fast, free shipping. Another study by UPS finds similar results. Shoppers are interested in free/cheap fast shipping. It is one of the main reasons why they choose or abandon certain marketplaces.

VALUABLE OUTCOMES

By using the dashboard, connected to a live stream of data, and updated in real-time, the decision-makers could watch out over the business and act with a quick response time.

By constructing the visualizations used in this research, we have found out many things about the progress of the company and established some pattern of expectations from our customers.

CONCLUSIONS AND RECOMMENDATIONS

The remote North has the lowest number of customers and suppliers. It is also the least populated, with a low density, low GDP/capita and is lacking in adequate transportation systems. Unlike the other regions, which benefit from high urbanization levels, with great purchasing power and are traversed by highways, the North is isolated. This is the region where Olist had the greatest delays in delivery times while the average reviews dropped accordingly.

According to the time series studying both the average review rating and the average difference between the estimated delivery time and the actual delivery time, there seems to be a problem in spring. Both in 2017 and in 2018, in the beginning of April. The cause for these delays needs to be further studied and could benefit from additional data. Nevertheless, they seem correlated and special precautions should be taken in this period in the future to avoid further delays and their subsequent complaints from customers. The business has grown considerably since 2016 and sometimes Olist's courier partners seem to barely reach their deadlines. This is why, along with the increase in sales, we should also consider an increase in courier workers or taking up more options for transporting the goods.

The coastline brings in the most customers and the most revenue and is also the most accessible due to infrastructure. Reaching out to the more remote regions of Brazil does present multiple challenges. On the other hand, these remote regions bear little significance in the overall size of the business. Thus, before jumping to any actions, a cost / benefit analysis should be done in order to ensure that the effort to connect to the customers does not outweigh the benefits.

Studying the time series, I reached the conclusion that there are some periods during the year which seem to repeat themselves in their effects. Spring brings about a period of decadence, for some reason, while we can also expect during 2018 the same surge in sales as was witnessed during last year's Black Friday. As more time passes and more data is gathered, the picture will become clearer and the company can become equipped

All in all, very few orders didn't actually meet the deadline. The business is overall growing but there is room for improvement and vigilance.

REFERENCES

[1]

Your Payment Gateway to Brazil | PagBrasil. (2018). *Boleto Bancário - a leading Brazilian payment method*. [online] Available at: <https://www.pagbrasil.com/payment-methods/boleto-bancario/> [Accessed 30 Dec. 2018].

[2]

Emarketing.alixpartners.com. (2018). [online] Available at: https://emarketing.alixpartners.com/rs/emimages/2018/pubs/EI/AP_Top_Trend_in_Shipping_Mar_2018.pdf [Accessed 30 Dec. 2018].

[3]

UPS Pressroom. (2018). *2016 UPS Pulse of the Online Shopper White Paper*. [online] Available at: <https://pressroom.ups.com/pressroom/ContentDetailsViewer.page?ConceptType=FactSheets&id=1465325363052-863> [Accessed 30 Dec. 2018].

LINKS

[1]

Olist. (2018). *Olist - ANUNCIAR E VENDER em MARKETPLACES sem complicação*. [online] Available at: <https://olist.com/> [Accessed 28 Dec. 2018].

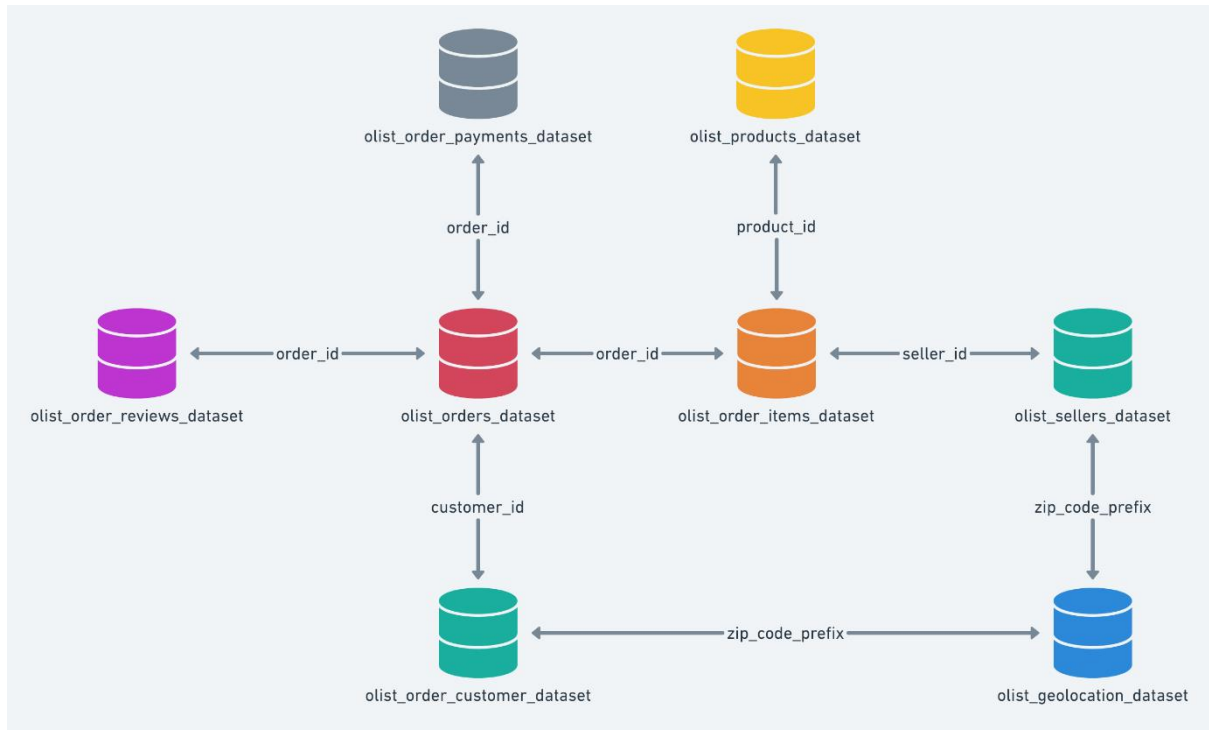
[2]

En.wikipedia.org. (2018). *Regions of Brazil*. [online] Available at: https://en.wikipedia.org/wiki/Regions_of_Brazil [Accessed 30 Dec. 2018].

APPENDIX

The appendix section covers, in order, the data schema and the data visualizations used across the results chapter.

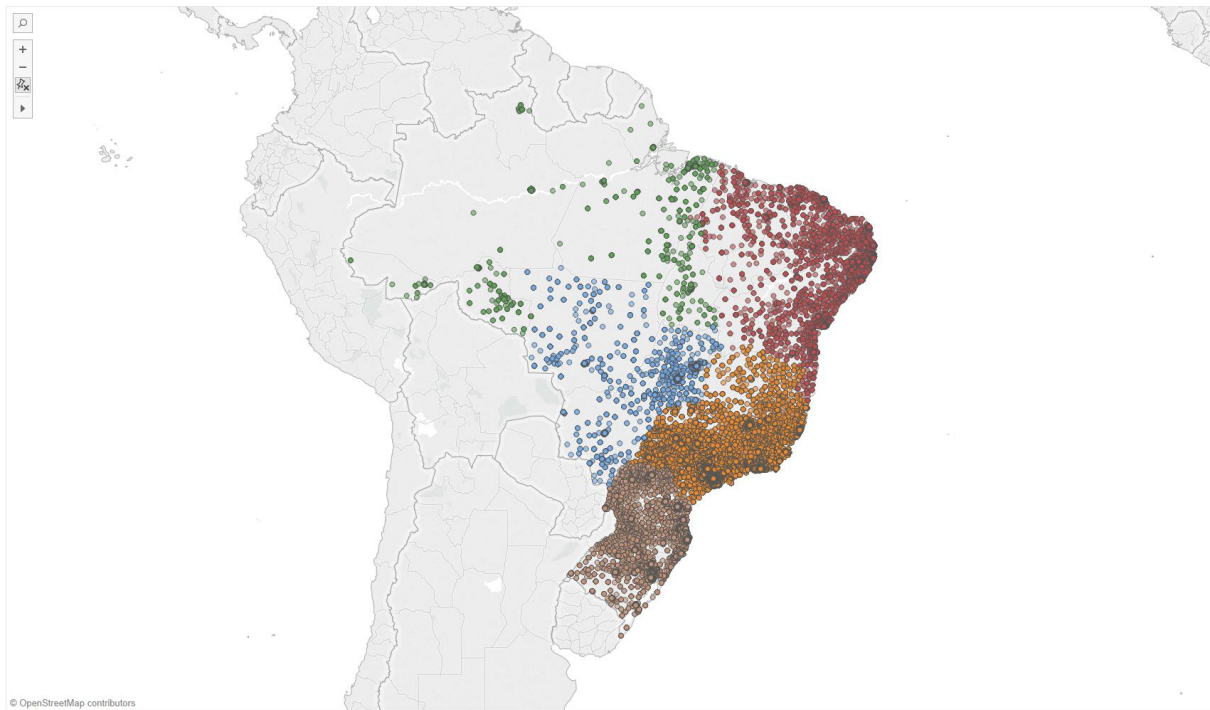
1)



2)

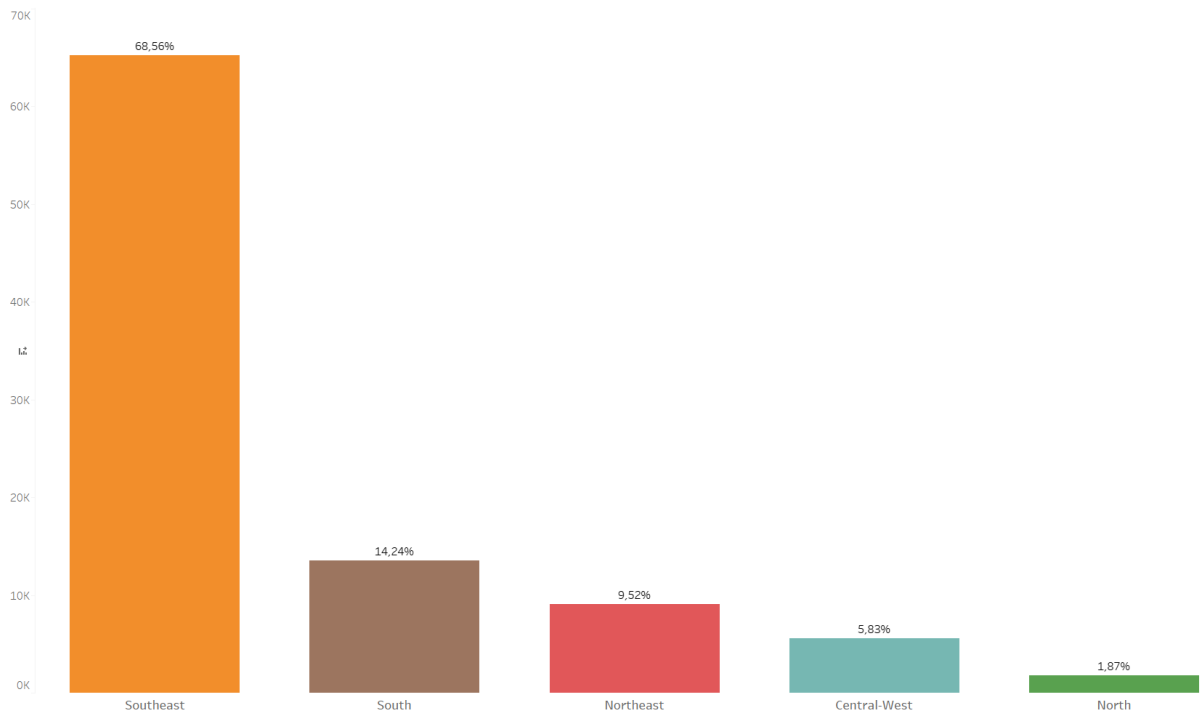


3)



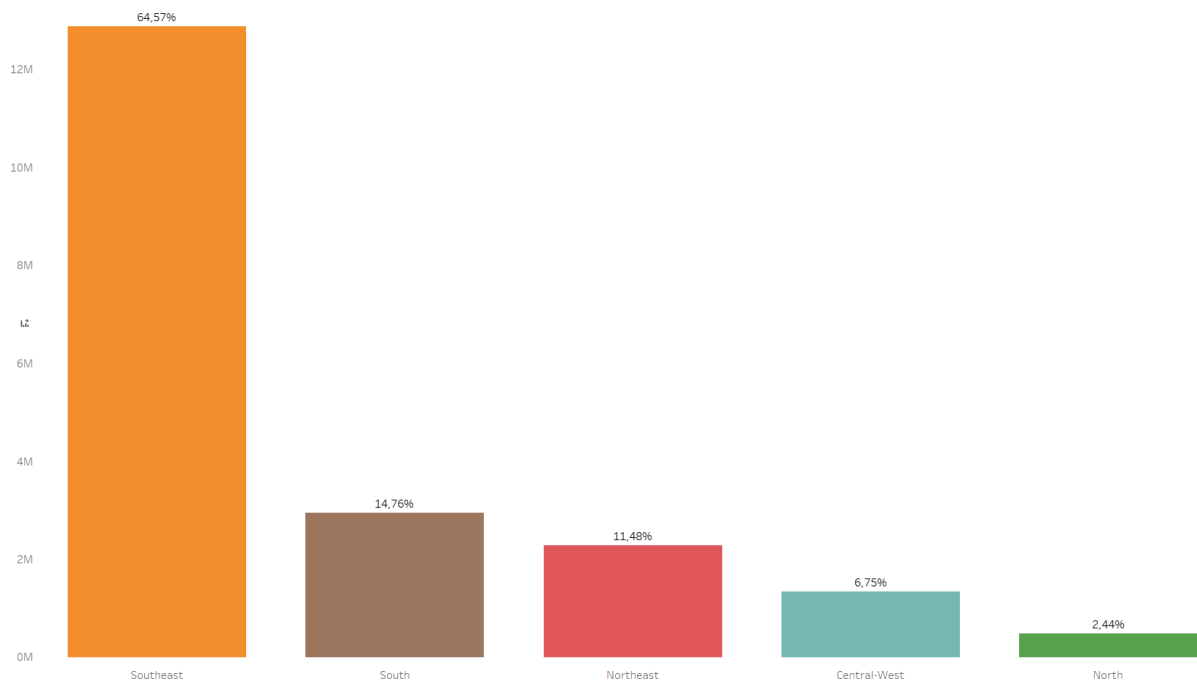
4)

Region by No. of Customers



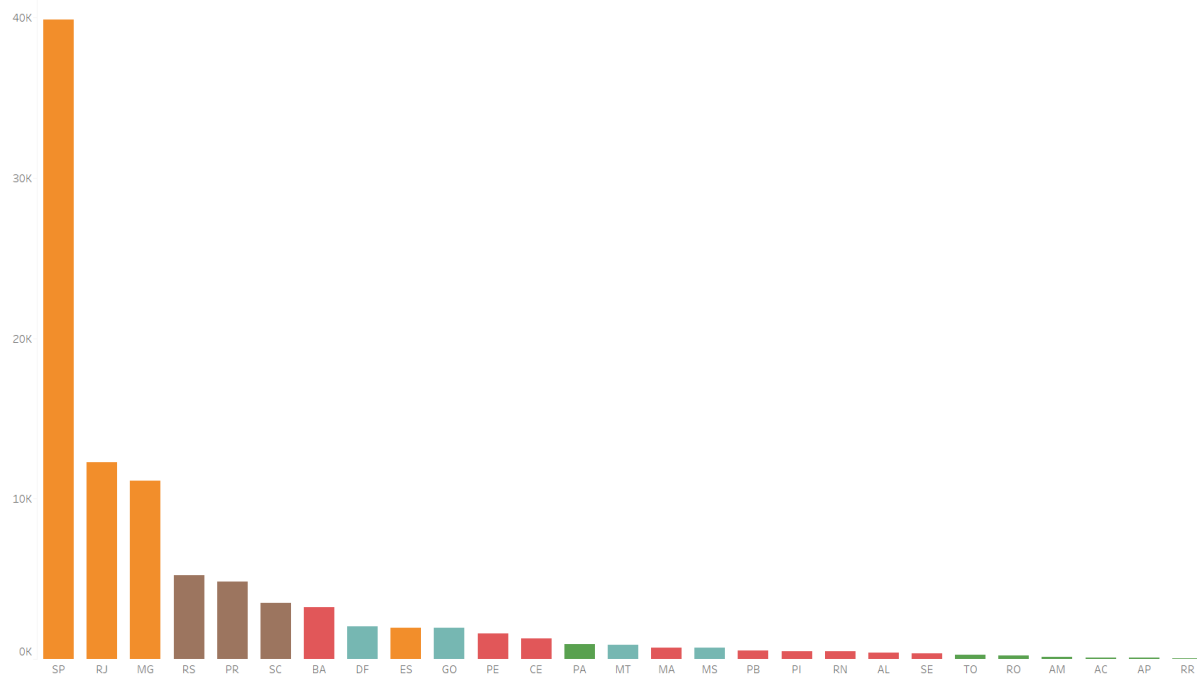
5)

Region by Payment Value

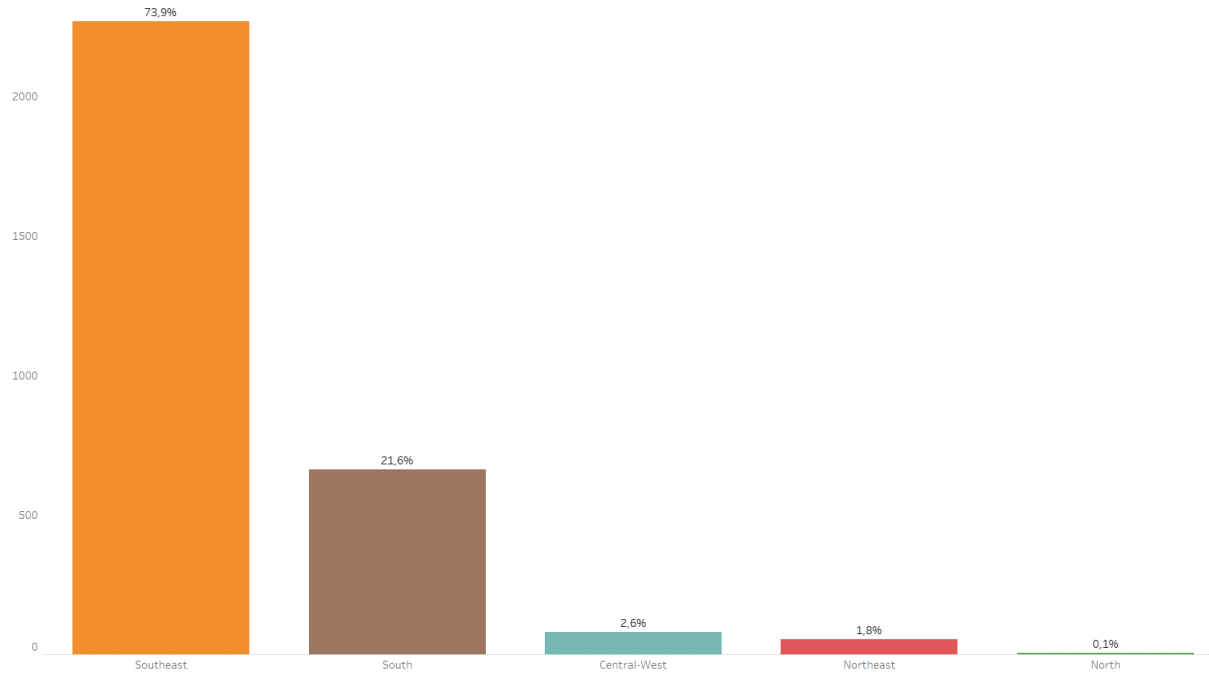


6)

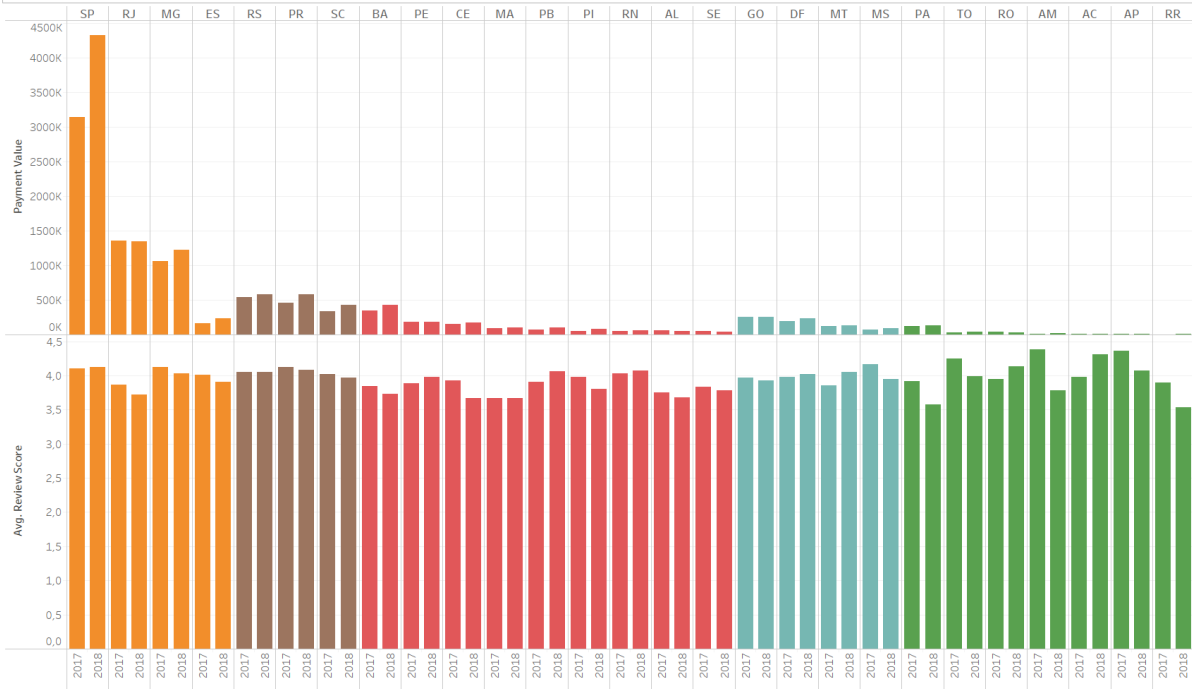
States by No. of Customers



Region by No. of Sellers

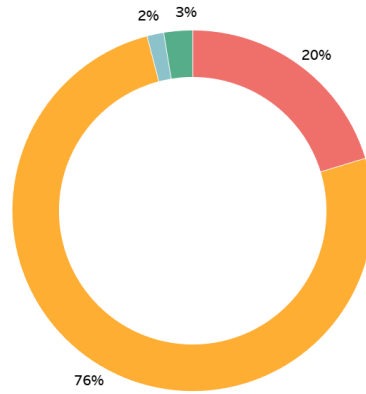


WHICH STATES MADE LESS THAN IN 2017?

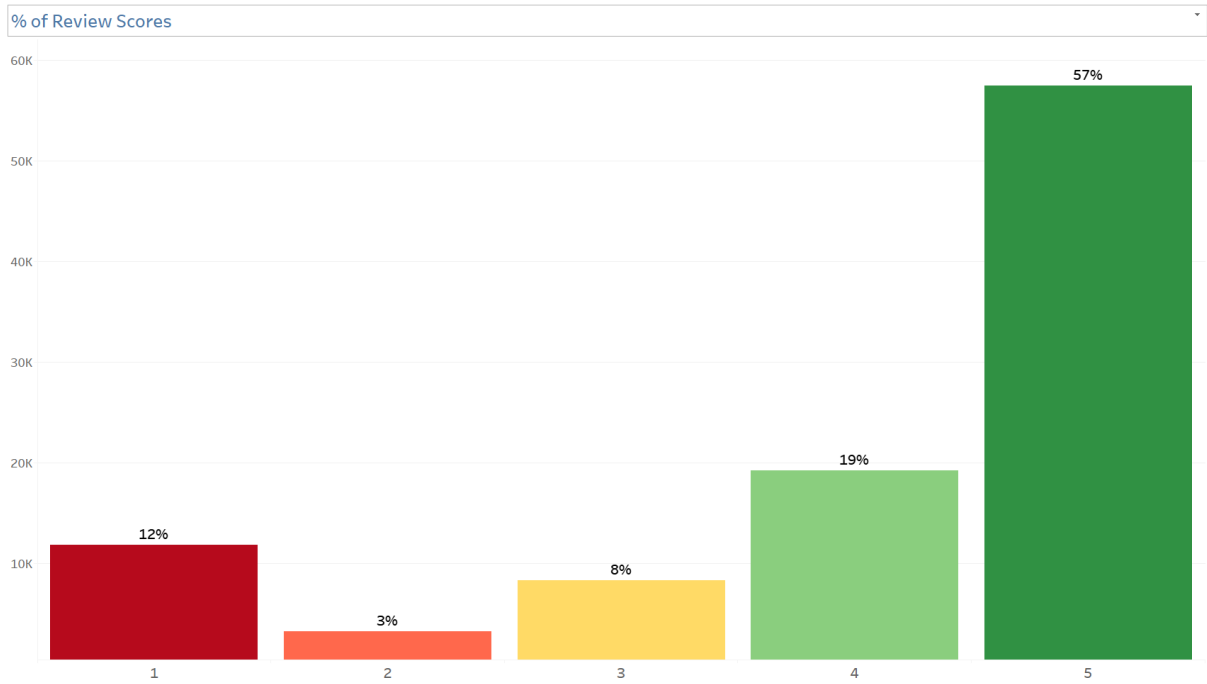


9)

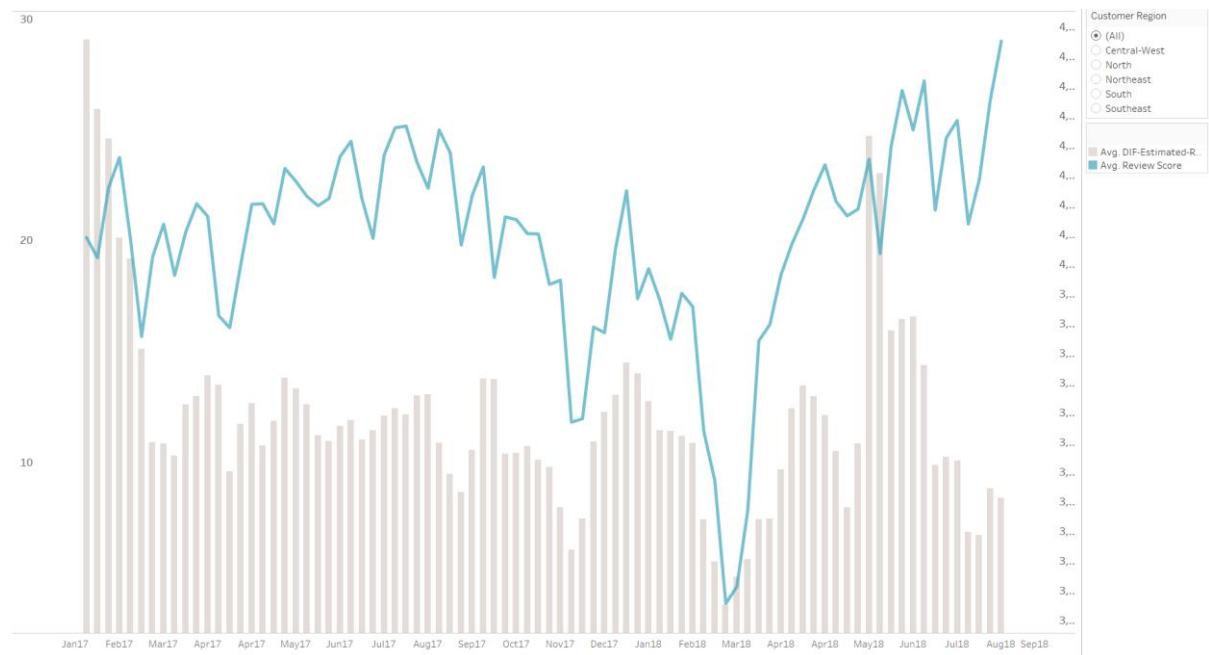
Payment Type



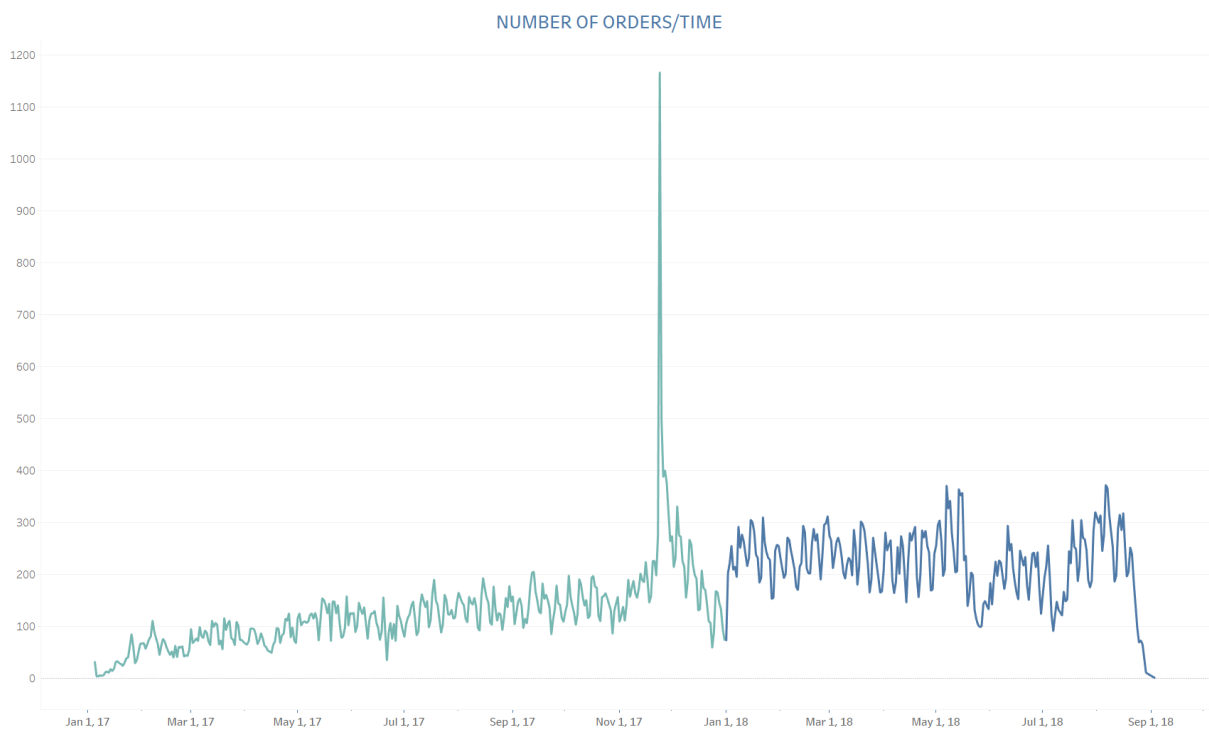
10)



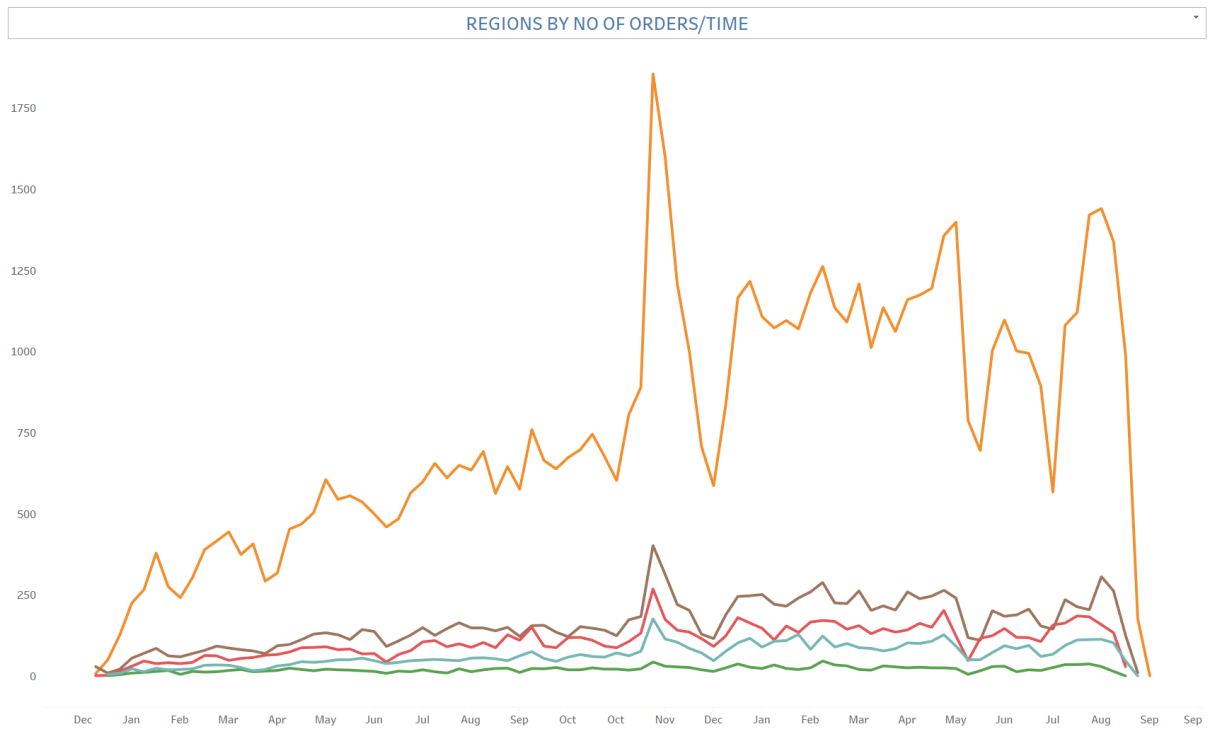
11)



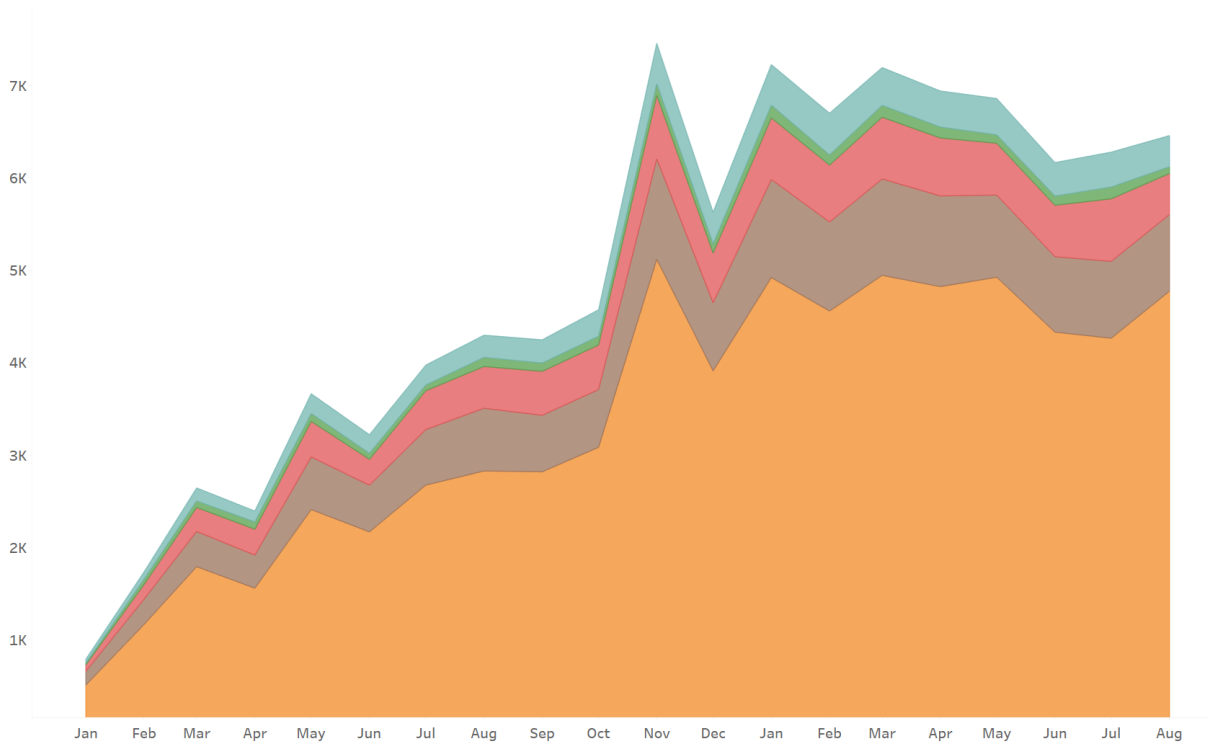
12)



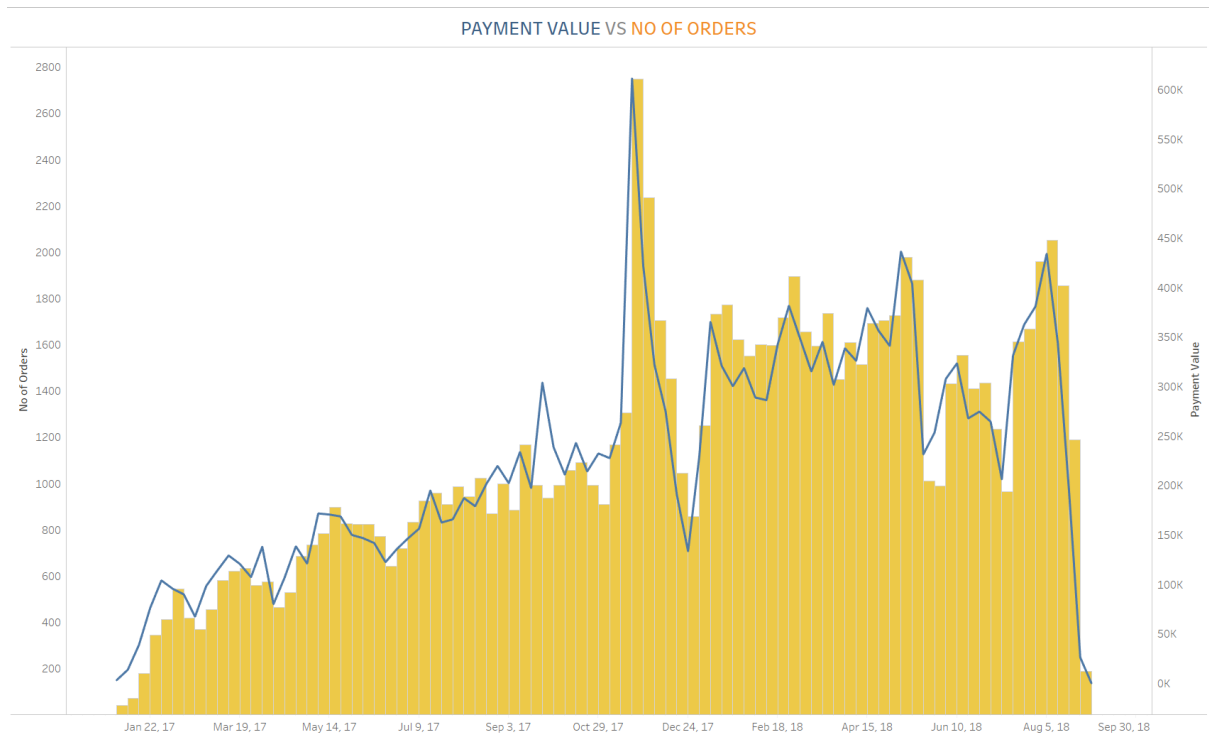
13)



14)

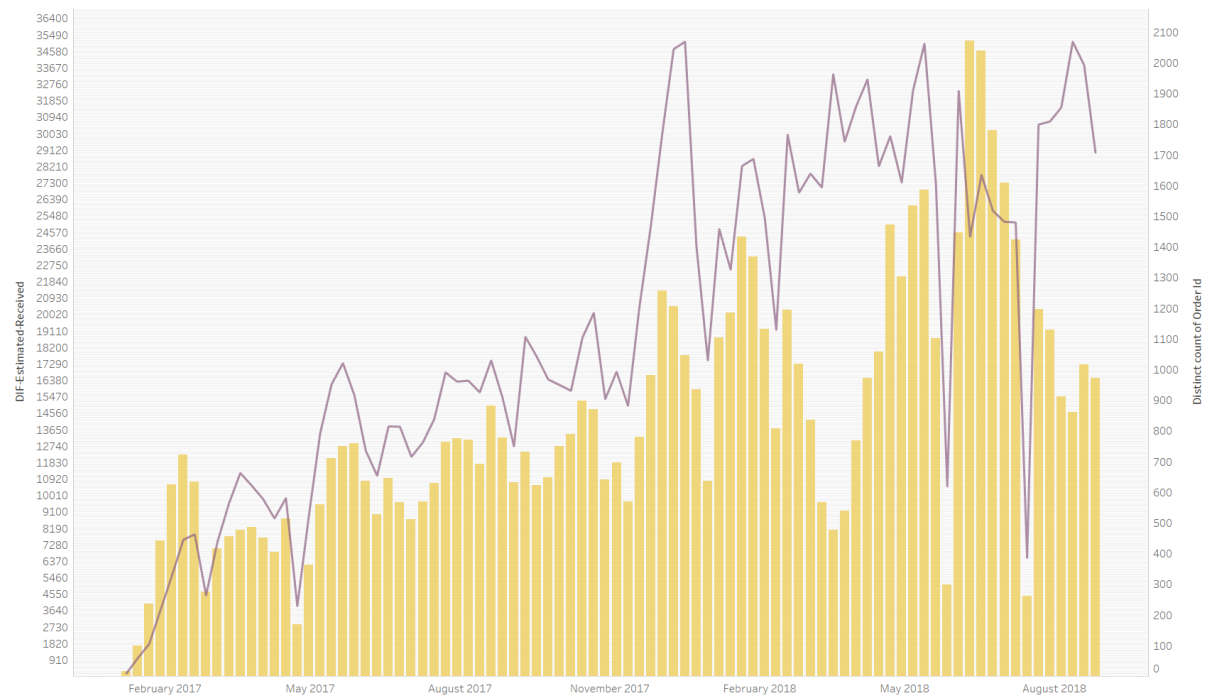


15)

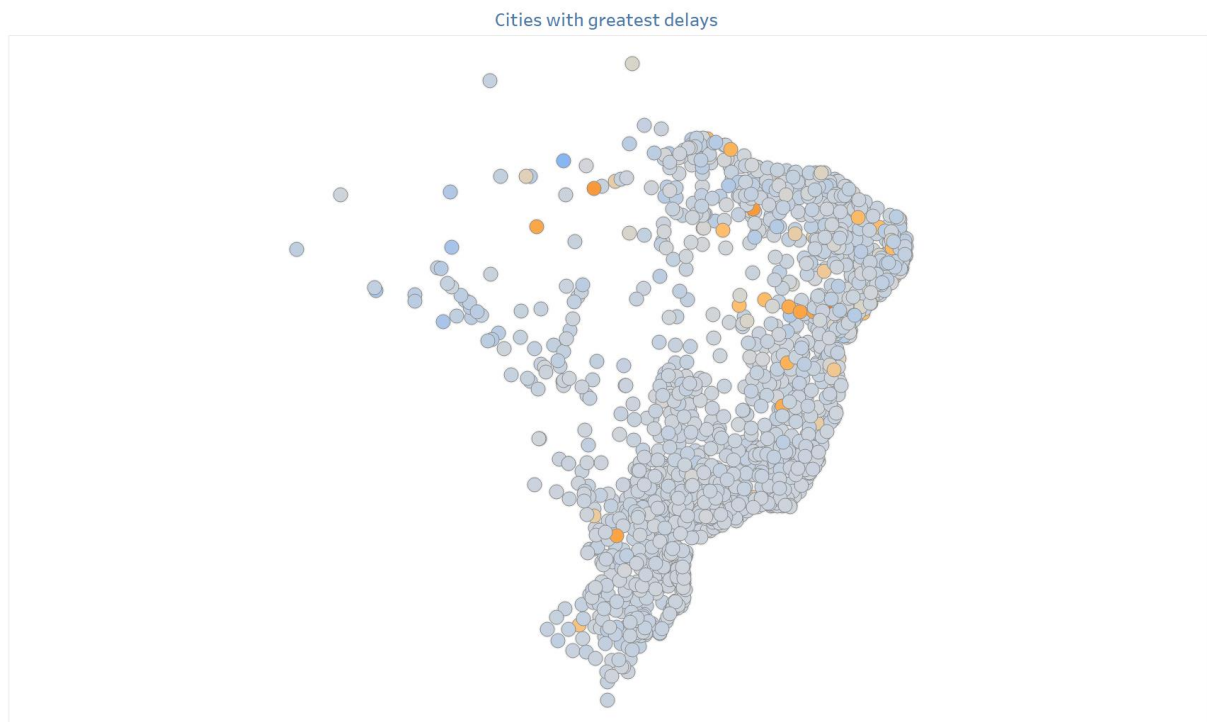


16)

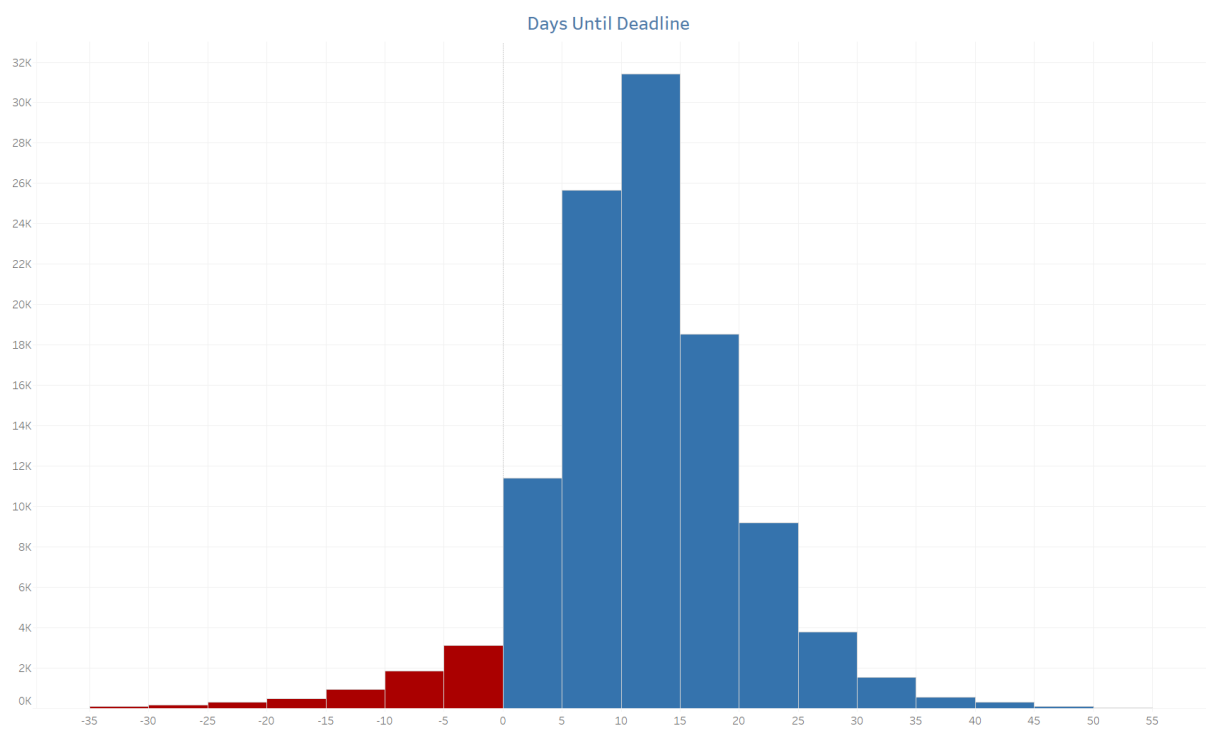
NO OF ORDERS vs DAYS UNTIL DEADLINE



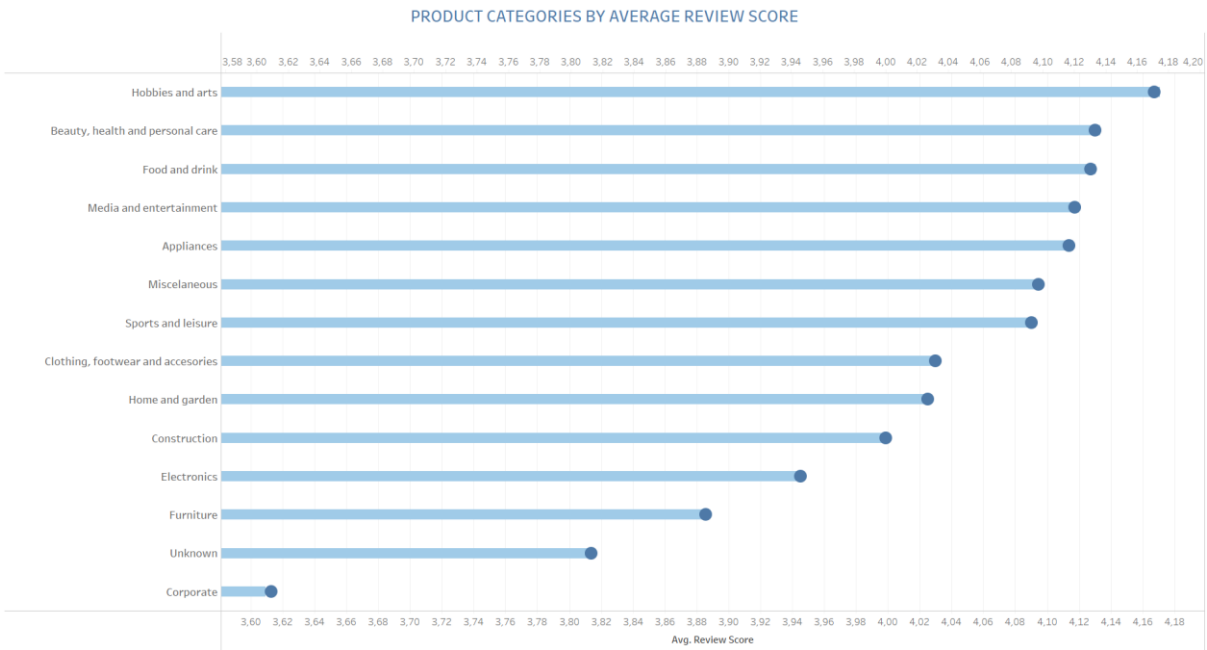
17)



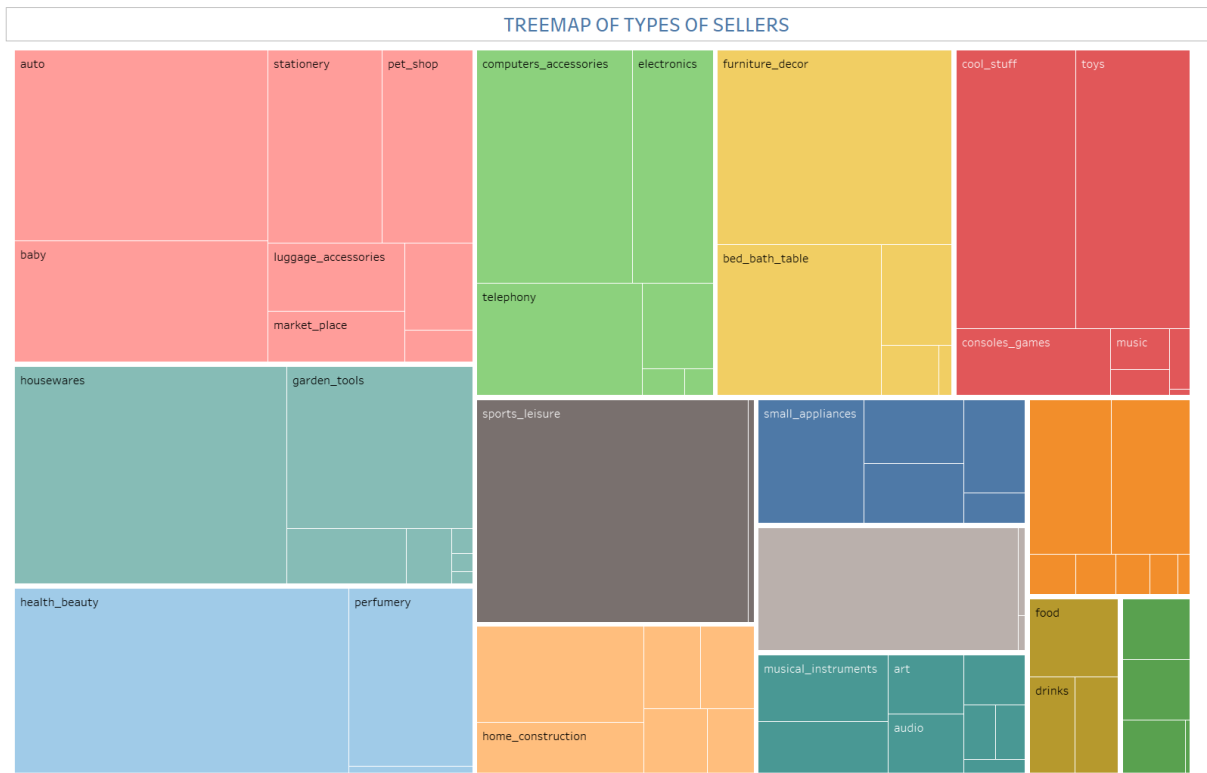
18)



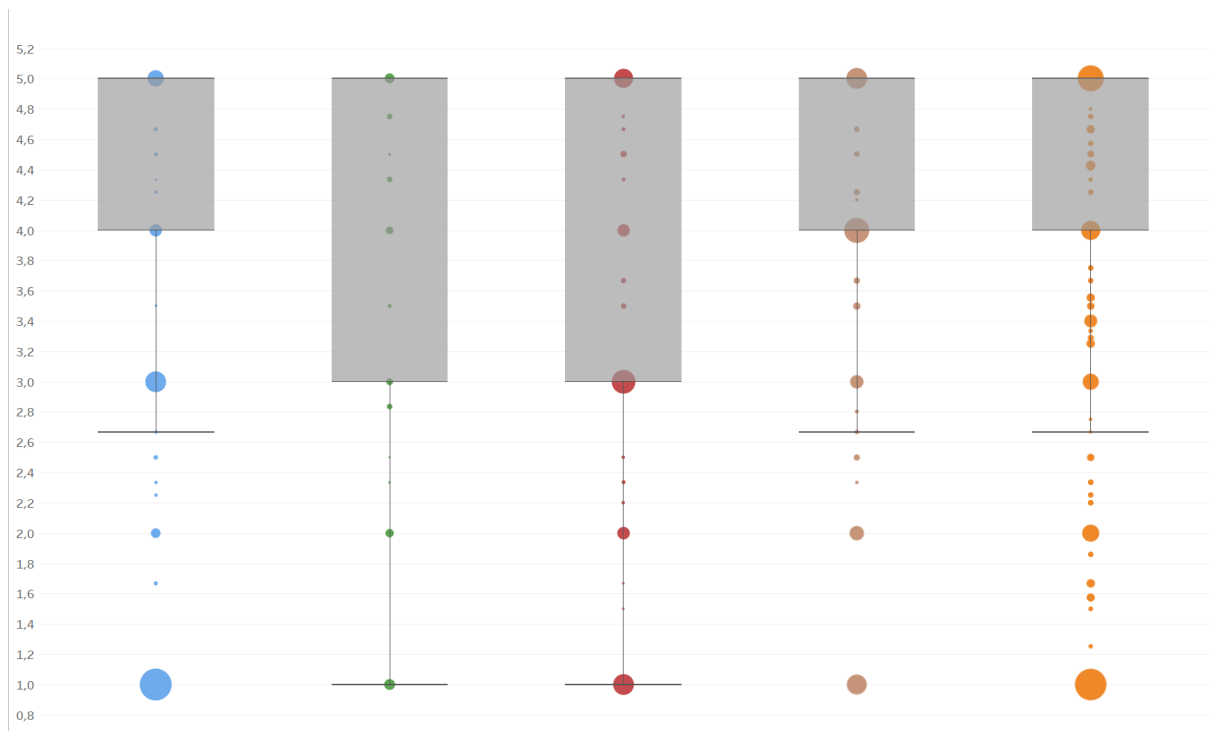
19)



20)

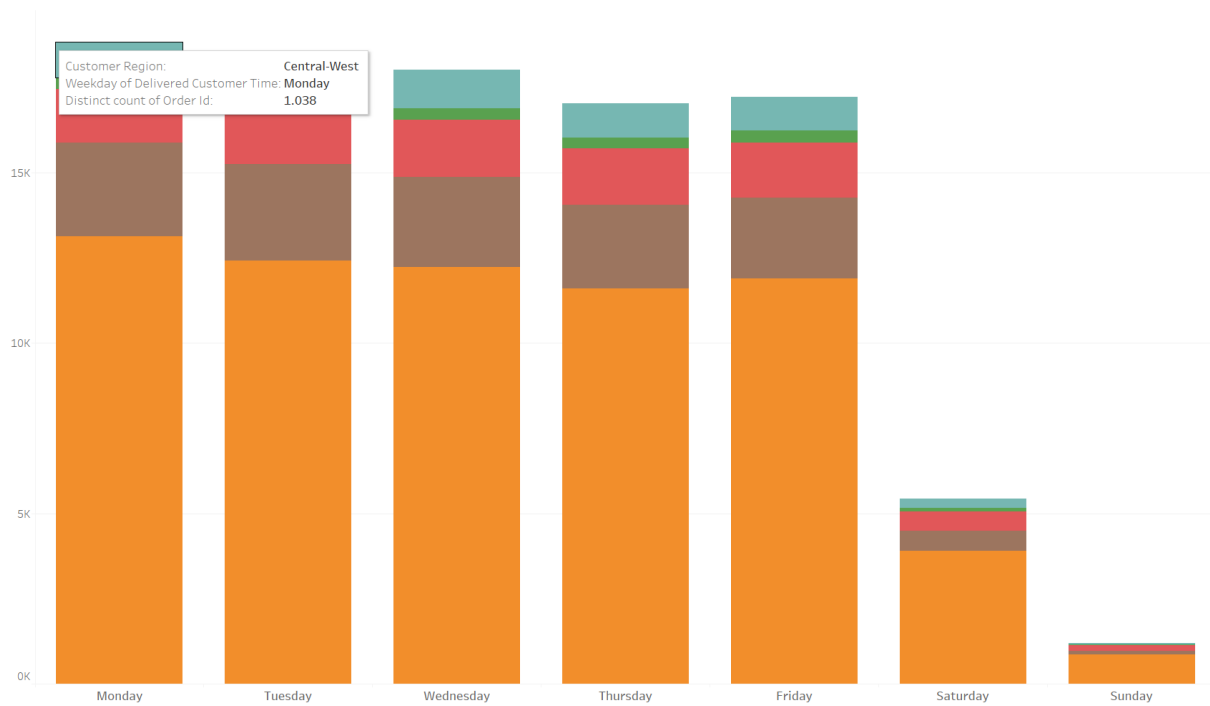


21)

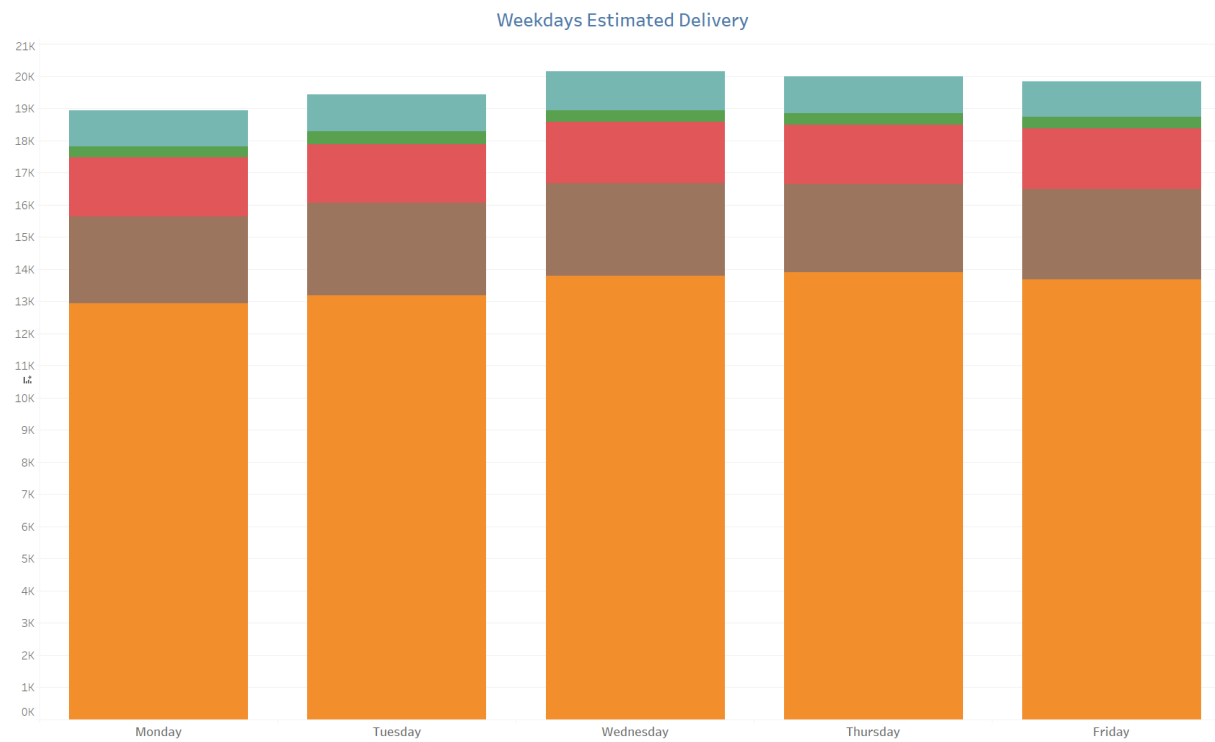


22)

WEEKDAYS DELIVERIES



23)

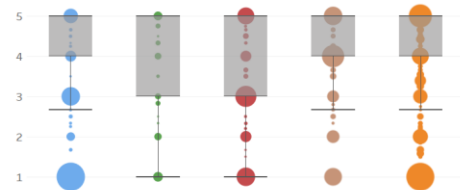


24)

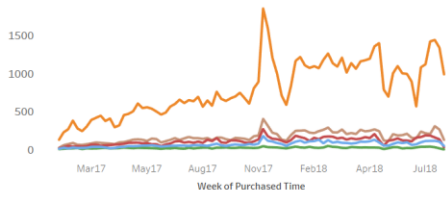


Customer Analysis

Customers by Avg. Review (size by payment total)



Number of Orders in Time (color by region)



Sellers
3,068

Customers
95,121

Sales
19,927,117

Customer Distribution (color by region)



Avg. Review vs Estimated - Delivered (days)

