Brazilian E-Commerce Data Analysis

olist
empowering commerce

Presented by-
Alpesh Gorde

Outline

Summary

Welcome! The analysis which is presented in upcoming slides is done in Jupyter notebook.

- I have used Brazilian e-commerce public dataset of orders made at Olist Store.

- The dataset has information of 100k orders from 2016 to 2018 made at multiple marketplaces in Brazil.

- Its features allows viewing an order from multiple dimensions: from order status, price, payment and freight performance to customer location and product attributes.

Problem Statement

In this EDA, I have tried to find-out some insights from the e-commerce dataset:
1. How e-commerce has evolved over time in Brazil?
2. Highest Monthly sales over 3 years?
3. On what days of week Brazilians customers tend to do online shopping?
4. Which are top 10 categories?
5. What is average price of top 10 categories?
6. How many payment methods are available? and which one is the most popular?
7. Payment methods vs number of instalments?
8. How customers are distributed over the states?
9. Which state has longest and shortest delivery time?
10. Highest sales for the year 2016, 2017, 2018?
11. Which customer city has the highest and lowest shipping Values?
12. Product category with the highest number of payment instalments?

Data Understanding
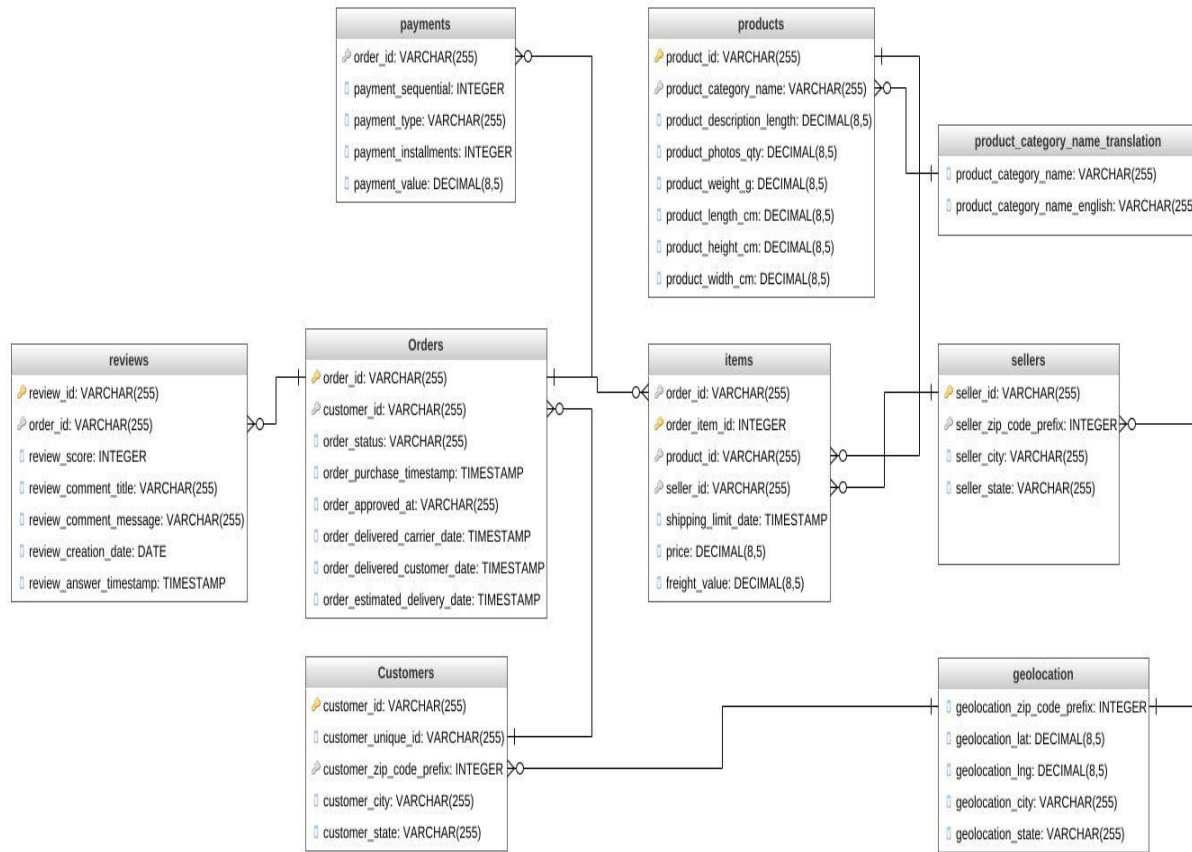
There are 8 datasets(in CSV format) with different attributes which are helpful for drawing some useful insights.

1. Customers Dataset
2. Order Dataset
3. Order Payment Dataset
4. Seller's Dataset
5. Product Category Dataset
6. Geo-location Dataset
7. Order items dataset
8. Products Dataset

Data Merging and Cleaning

These Datasets are merged and null values has been identified as follows:

	dataset	cols	cols_no	null_no	null_cols_no	null_cols
0	customers	customer_id, customer_unique_id, customer_zip_code_prefix, customer_city, customer_state	5	0	0	
1	items	order_id, order_item_id, product_id, seller_id, shipping_limit_date, price, freight_value	7	0	0	
2	payments	order_id, payment_sequential, payment_type, payment_installments, payment_value	5	0	0	
3	orders	order_id, customer_id, order_status, order_purchase_timestamp, order_approved_at, order_delivered_carrier_date, order_delivered_customer_date, order_estimated_delivery_date	8	4908	3	order_approved_at, order_delivered_carrier_date, order_delivered_customer_date
4	products	product_id, product_category_name, product_name_lenght, product_description_lenght, product_photos_qty, product_weight_g, product_length_cm, product_height_cm, product_width_cm	9	2448	8	product_category_name, product_name_lenght, product_description_lenght, product_photos_qty, product_weight_g, product_length_cm, product_height_cm, product_width_cm
5	category_translation	product_category_name, product_category_name_english	2	0	0	
6	geolocation	geolocation_zip_code_prefix, geolocation_lat, geolocation_lng, geolocation_city, geolocation_state	5	0	0	
7	sellers	seller_id, seller_zip_code_prefix, seller_city, seller_state	4	0	0	

Exploratory Data Analysis

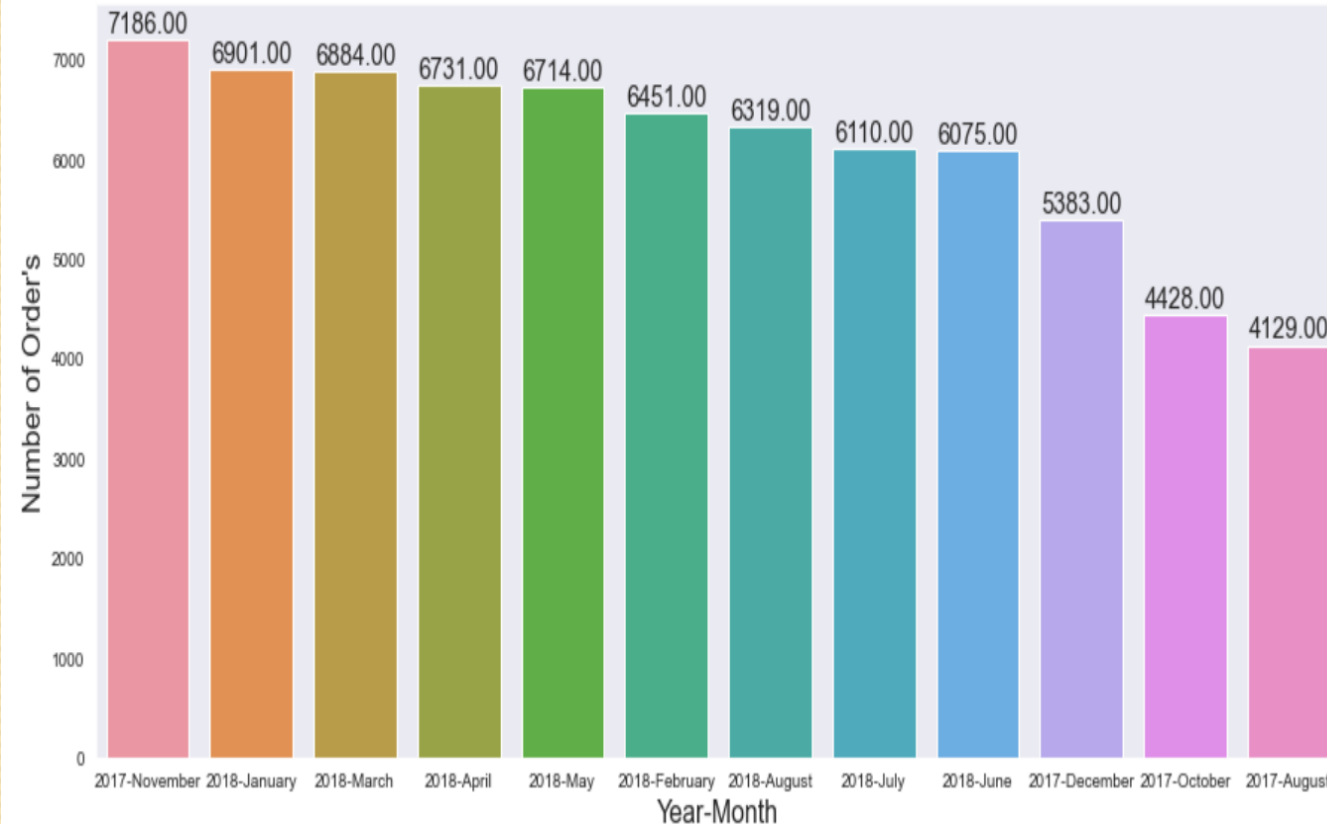**1. How e-commerce has evolved over time in Brazil?**
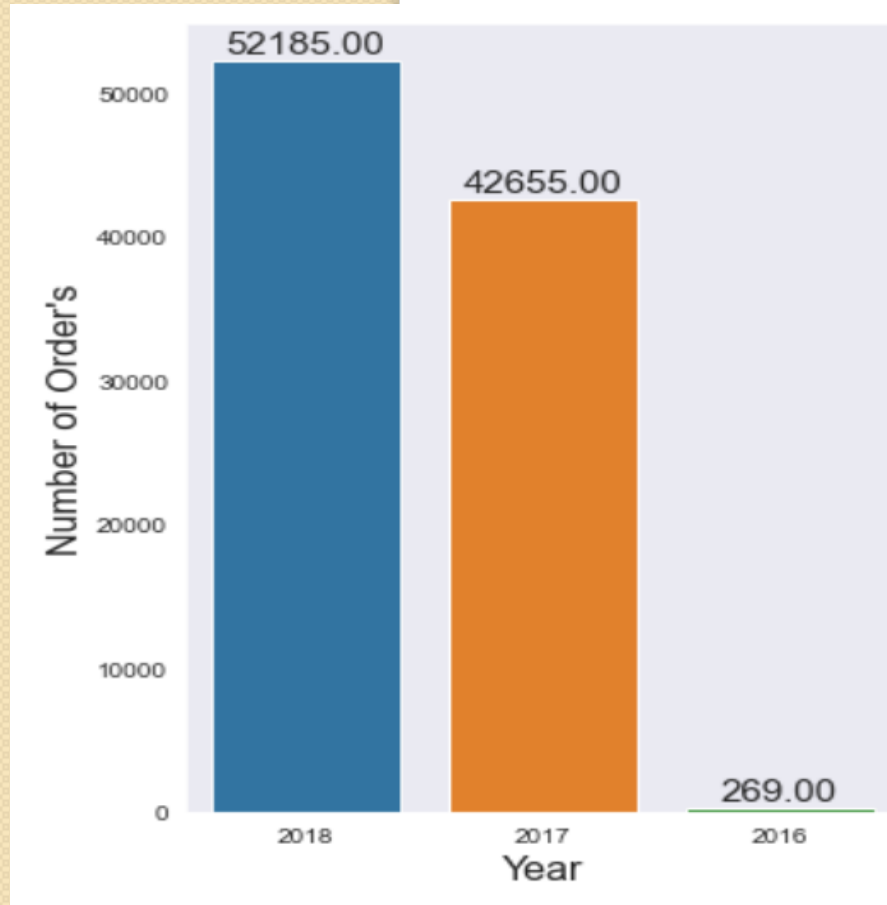
It can be seen that after 2016 e-commerce in Brazil increased rapidly.

Exploratory Data Analysis
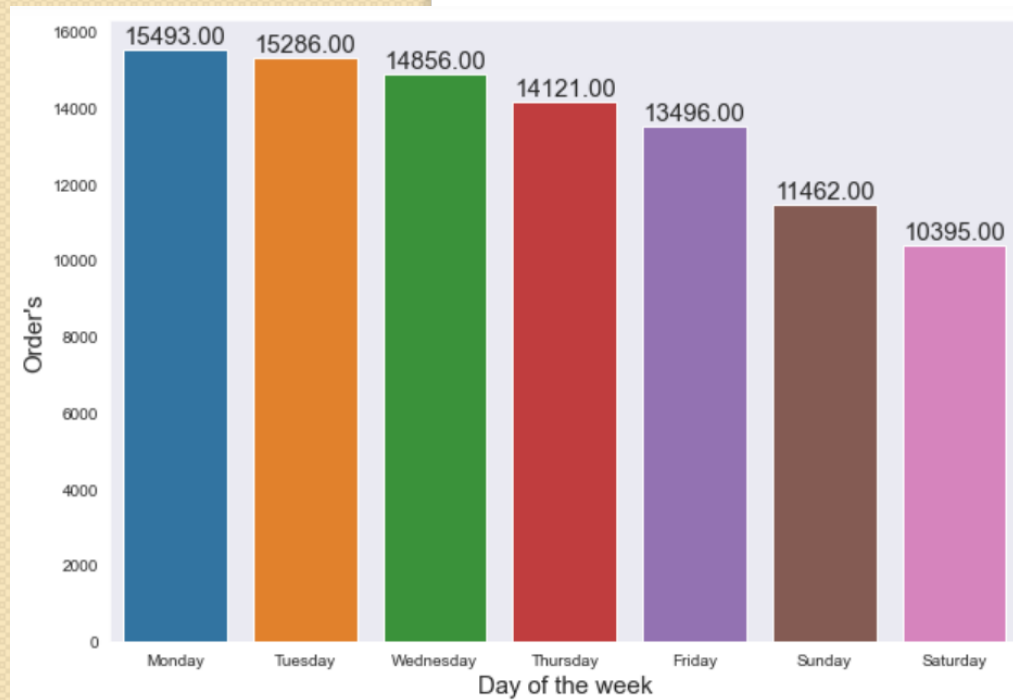
2. Highest Monthly sales over 3 years?

It can be clearly seen that Nov 2017 is the highest sales months over the period of 3 years.
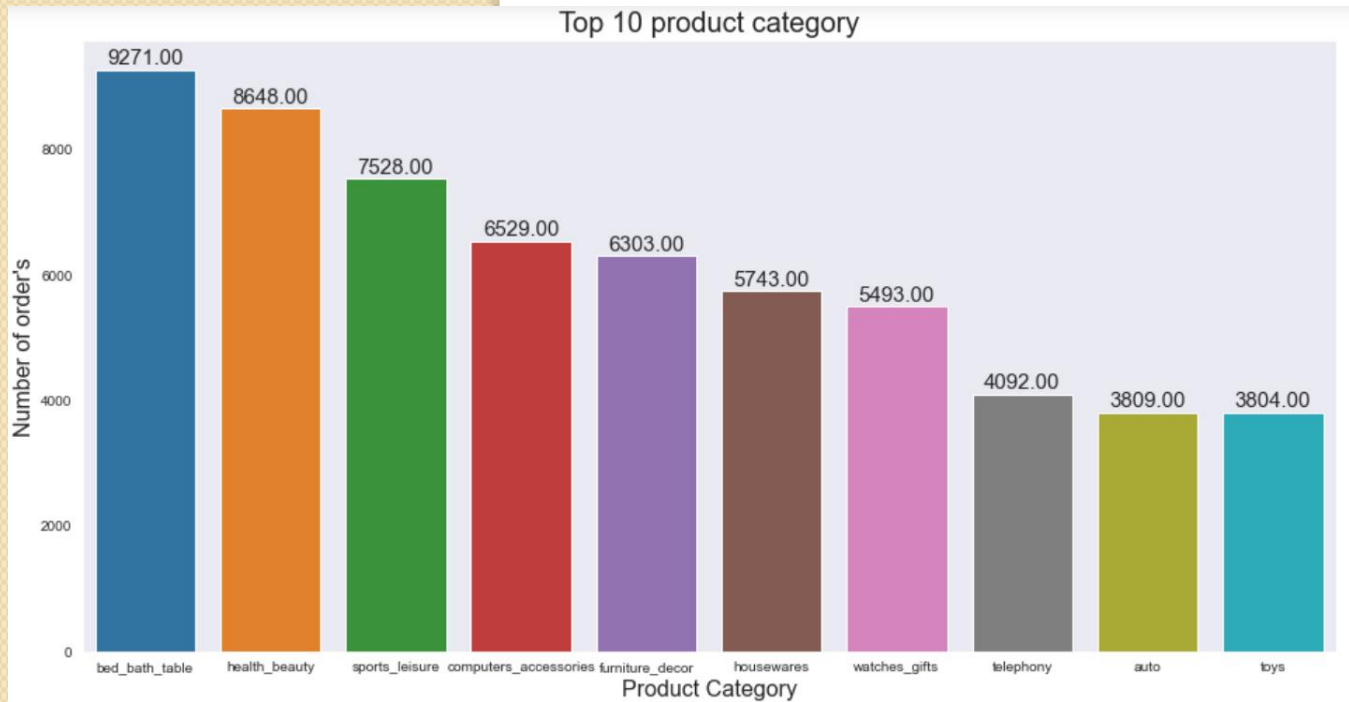
Exploratory Data Analysis

3. On what days of week Brazilians customers tend to do online shopping?

We can see more online shopping happen on weekdays and less on weekends.
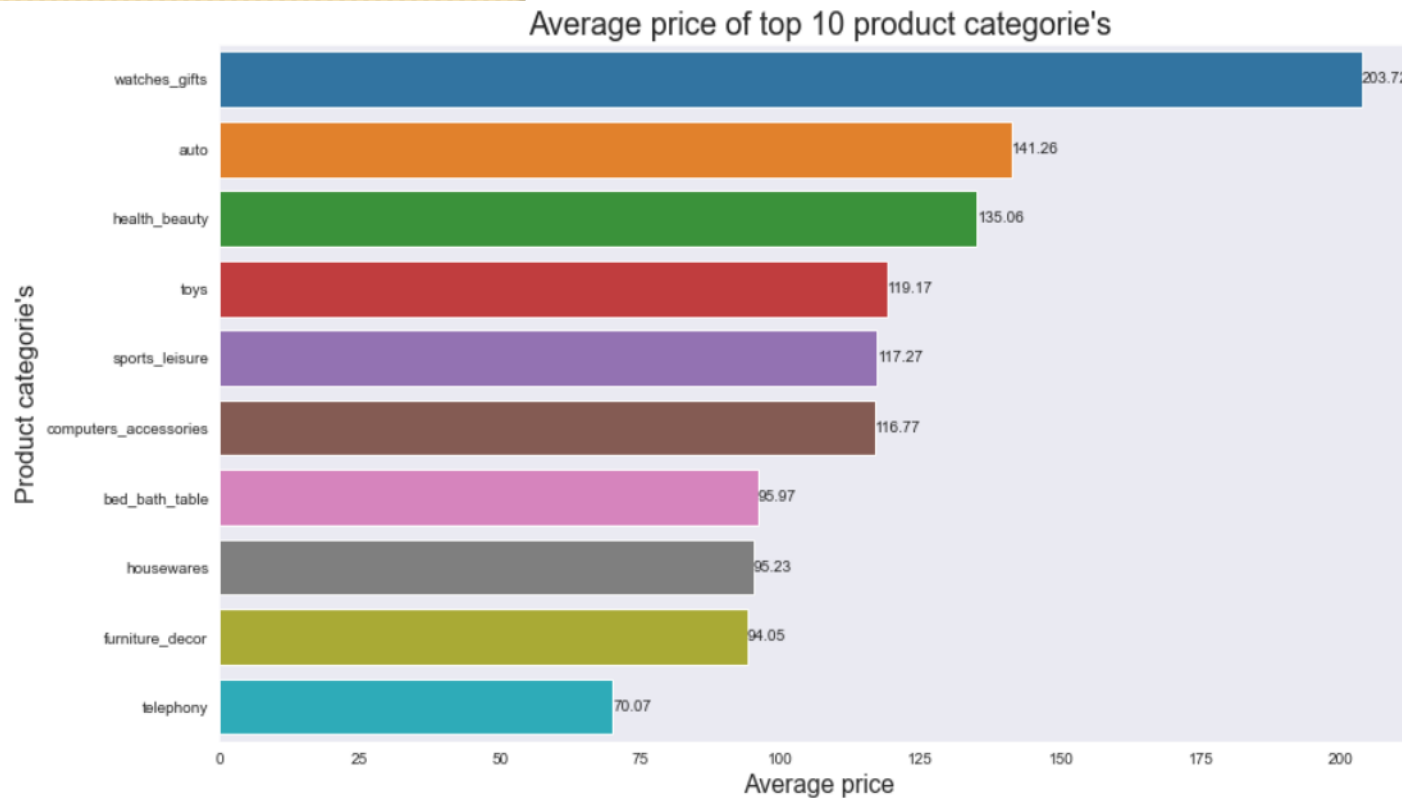
Exploratory Data Analysis

4. Which are top 10 categories?

Bed_bath_table category is the top selling category.

Exploratory Data Analysis

5. What is average price of top 10 categories?
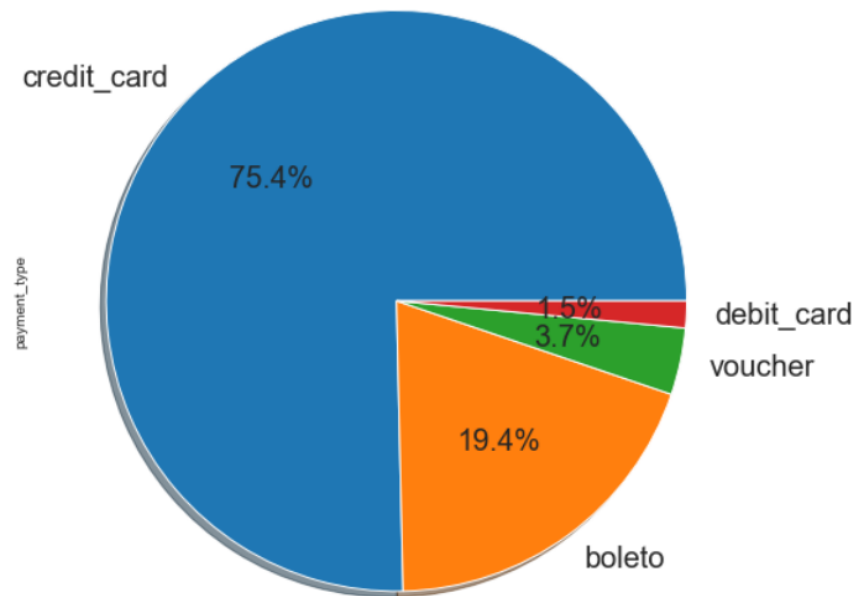

Average price of top 10 product categorie's

Watches and gifts is the top product with highest average price

Exploratory Data Analysis

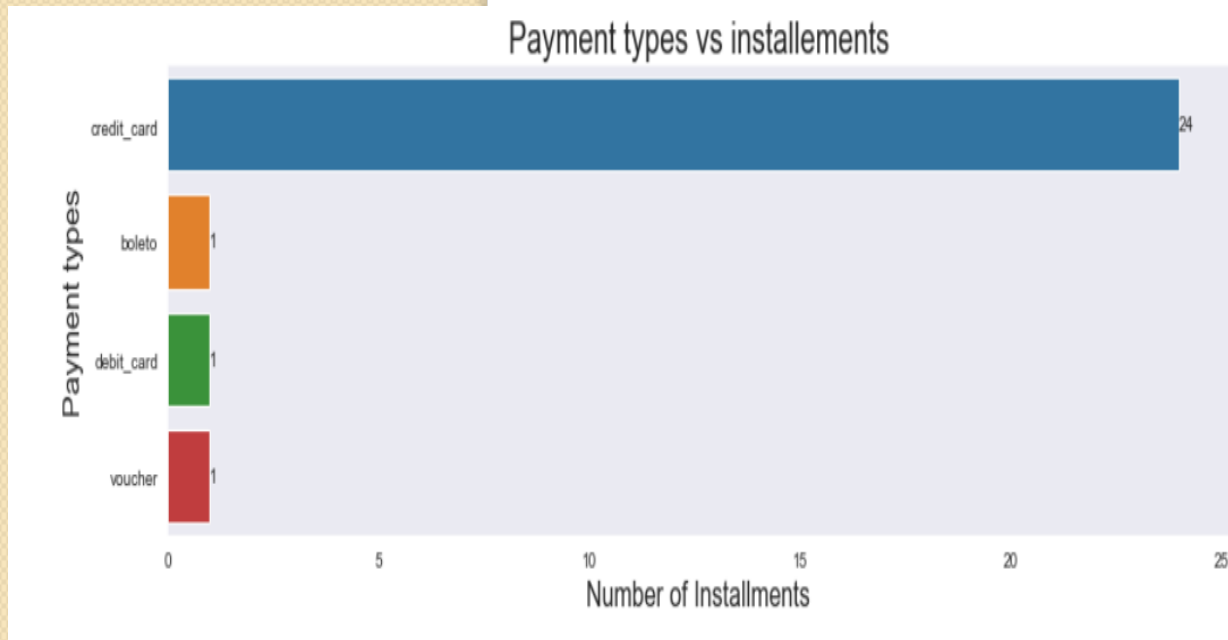**6. How many payment methods are available? and which one is the most popular?**

The Most Frequent Payment Type

There are 4 payment types in Brazil with Credit Card solely account of 75 % and it is the most common payment method.
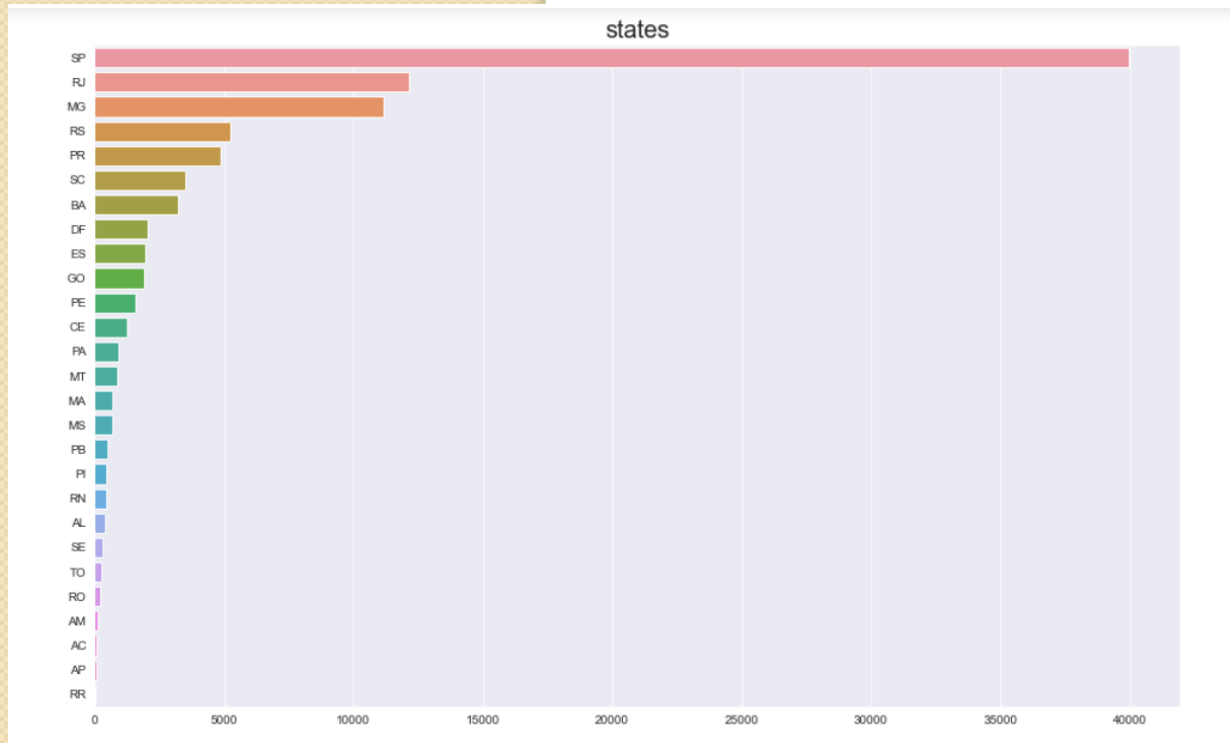
Exploratory Data Analysis

7. Payment methods vs number of installments

Credit Cards having highest number of Instalments
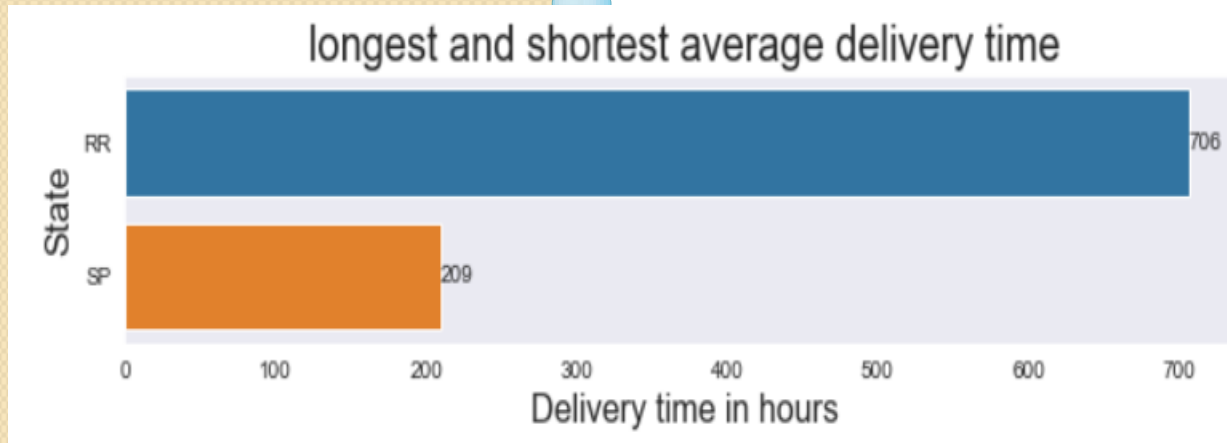
Exploratory Data Analysis

8. How customers are distributed over the states?

Most customers are from Sao Paulo, followed by Rio de Janeiro.
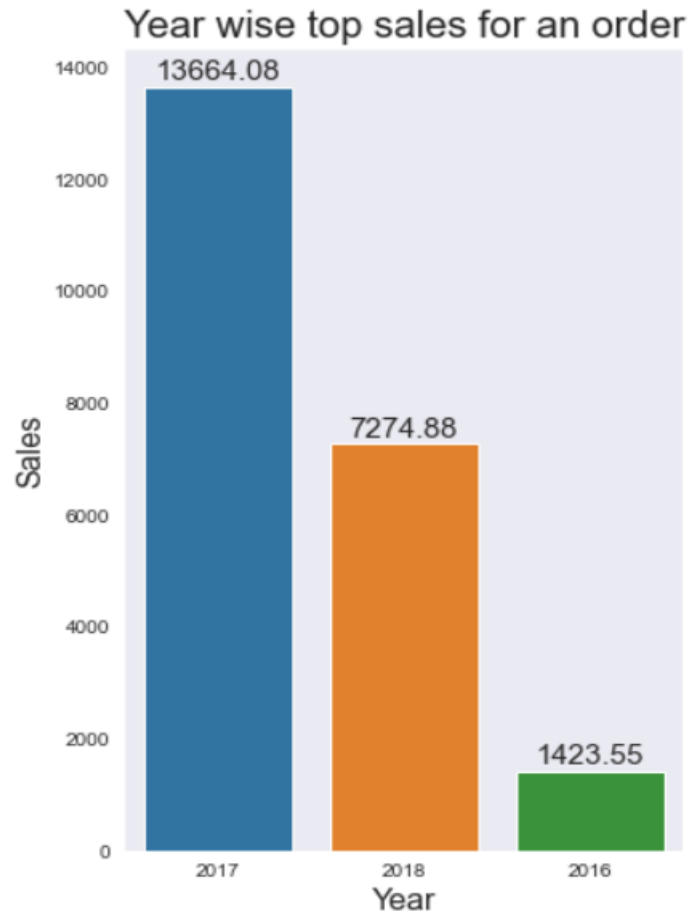
Exploratory Data Analysis

9. Which state has longest & shortest average delivery time?

longest and shortest average delivery time

The longest and shortest delivery time belongs to the State of Roraima(RR) and Sao Paulo(SP).

Exploratory Data Analysis
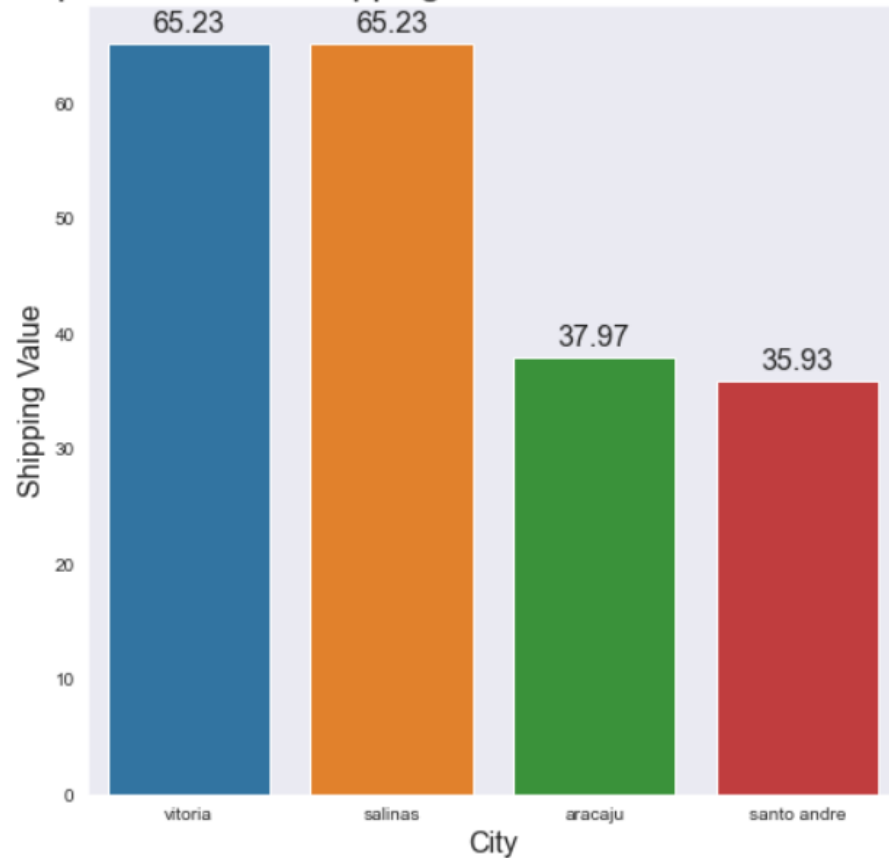
10. Highest sales for the year 2016, 2017, 2018

2017 year has the highest sales with almost double the amount compared to 2018

Exploratory Data Analysis

11. Customer City vs Freight Value

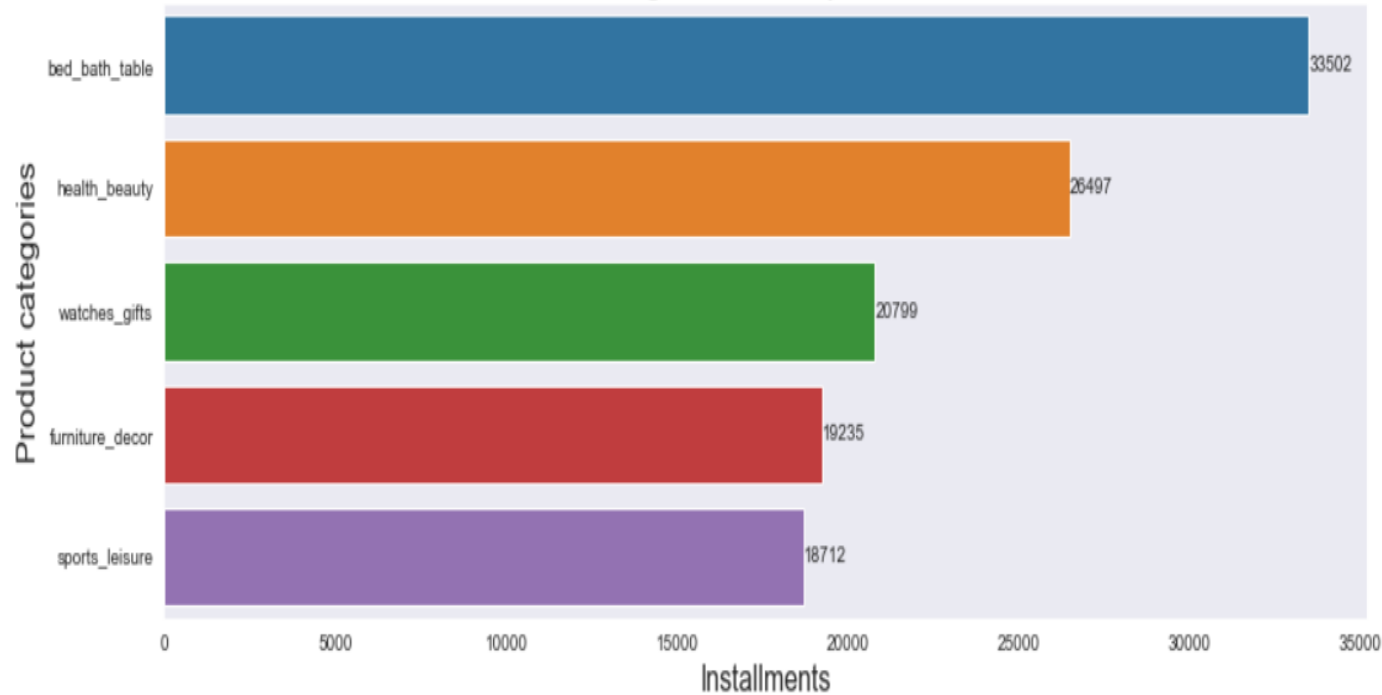Top 2 and least 2 shipping values across the cities in Brazil

Vitoria and Salinas has the highest shipping value and Santo Andre has the lowest shipping value

Exploratory Data Analysis

12. Product category with the highest number of payment installments

Product categories with top 5 installements

Bed bath table has the higesht number of payment installment

Conclusion

With these 12 most important and insightful findings, further action can be taken to work on the weaknesses and improve the services.

- We can improve the delivery time for the city the State of Roraima(RR), by considering factors affecting the delivery.
- We can improve the spread of customers to the shown city by segmented advertising.
- We can concentrate on top selling products by offering variety.
- By analysing factors affecting highest monthly sales we can apply same methodology for rest of the year.