TABLEAU PROJECT REPORT

Analysis of E-Commerce Dataset

Submitted by

Rahul Kumar

Registration No. 11704509

Course Code: INT233

Under the Guidance of

Assistant Professor Shaina Gupta

UID: **24968**

Lovely School of Computer Science and Engineering

Lovely Professional University, Phagwara

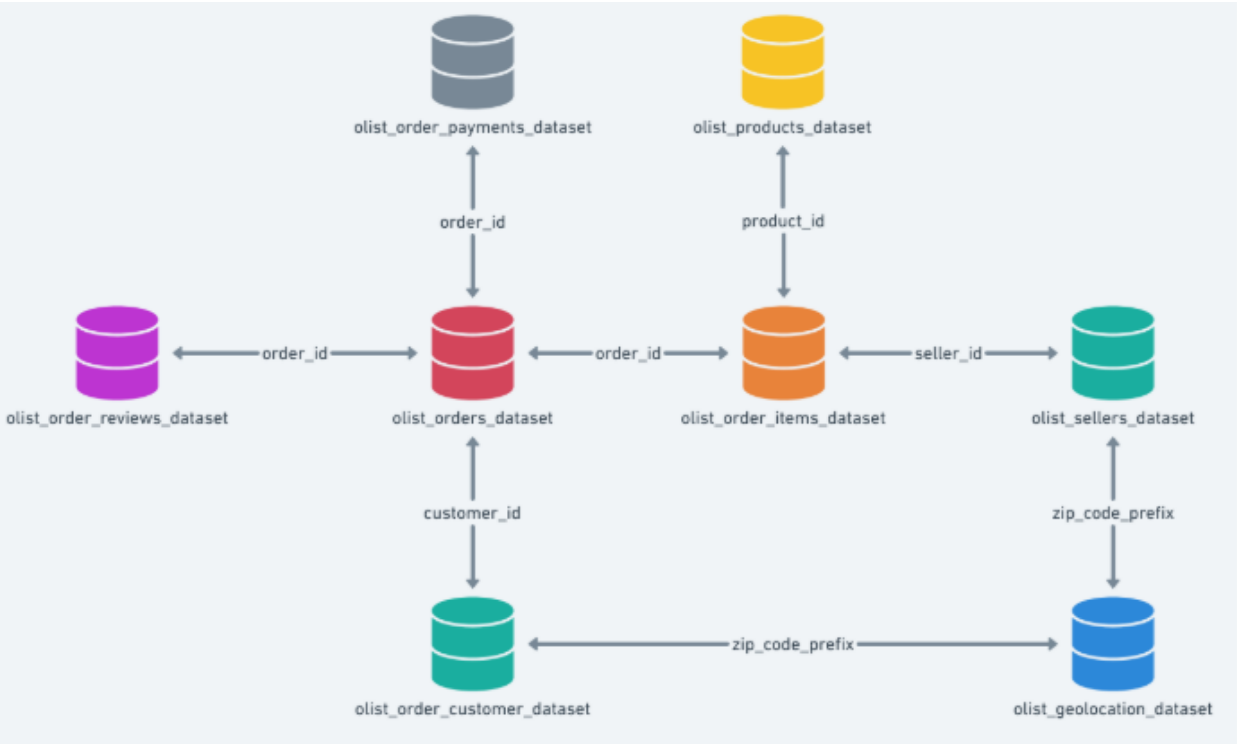
Introduction:

In this Analysis, I used a dataset of E-Commerce Company named “Olist Store”. The dataset has information of more 100k orders from 2016 to 2018. Its features allow viewing an order from multiple dimensions: from order status, price, payment and freight performance to customer location, product attributes.

This Dataset contains 8 CSV files named as follows:

1. Dataset of Orders.
2. Dataset of Order Reviews
3. Dataset of Order Items
4. Dataset of Products and its categories
5. Dataset of Order Payments
6. Dataset of Customers
7. Dataset of Sellers
8. Dataset of Geolocation of Customers and Sellers

Data Schema: - for Better understanding of Relationship between Datasets



Walkthrough of Datasets

Order Dataset

This is the core dataset. From each order you might find all other information.

order_id: -unique identifier of the order.
customer_id: -key to the customer dataset. Each order has a unique customer_id.
order_status: -Reference to the order status (delivered, shipped, etc).
order_purchase_timestamp: -Shows the purchase timestamp.
order_approved_at: -Shows the payment approval timestamp.
order_delivered_carrier_date: -Shows the order posting timestamp. When it was handled to the logistic partner.
order_delivered_customer_date: -Shows the actual order delivery date to the customer.
order_estimated_delivery_date: -Shows the estimated delivery date that was informed to customer at the purchase moment.

Order Items Dataset

This dataset includes data about the items purchased within each order.

order_id: -order unique identifier
order_item_id: -sequential number identifying number of items included in the same order.
product_id: -product unique identifier
seller_id: -seller unique identifier
shipping_limit_date: -Shows the seller shipping limit date for handling the order over to the logistic partner.
Price: -item price
freight_value: -item freight value item.

Order Reviews Dataset

This dataset includes data about the reviews made by the customers.

review_id: -unique review identifier
order_id: -unique order identifier
review_score: -Note ranging from 1 to 5 given by the customer on a satisfaction survey.
review_comment_title: -Comment title from the review left by the customer, in Portuguese.
review_comment_message: -Comment message from the review left by the customer, in Portuguese.
review_creation_date: -Shows the date in which the satisfaction survey was sent to the customer.
review_answer_timestamp: -Shows satisfaction survey answer timestamp.

Products Dataset

This dataset includes data about the products sold by Olist.

product_id: -unique product identifier
product_category_name: -root category of product, in Portuguese.
product_name_length: -number of characters extracted from the product name.
product_description_length: -number of characters extracted from the product description.
product_photos_qty: -number of product published photos
product_weight_g: -product weight measured in grams.
product_length_cm: -product length measured in centimeters.
product_height_cm: -product height measured in centimeters.
product_width_cm: -product width measured in centimeters.

Customers Dataset

This dataset has information about the customer and its location.

customer_id: -key to the orders dataset. Each order has a unique customer_id.
customer_unique_id: -unique identifier of a customer.
customer_zip_code_prefix: -first five digits of customer zip code
customer_city: -customer city name
customer_state: -customer state

Sellers Dataset

This dataset includes data about the sellers that fulfilled orders made at Olist.

seller_id: -seller unique identifier
seller_zip_code_prefix: -first 5 digits of seller zip code
seller_city: -seller city name
seller_state: -seller state

Payments Dataset

This Dataset Contains data about the orders payment options.

order_id: - unique identifier of an order.
payment_type: -method of payment chosen by the customer.
payment_installments: -number of instalments chosen by the customer.
payment_value: - transaction value.

Geolocation Dataset

This dataset has information Brazilian zip codes and its latitude /longitude coordinates.

geolocation_zip_code_prefix: -first 5 digits of zip code
geolocation_lat: -latitude
geolocation_lng: -longitude
geolocation_city: -city name
geolocation_state: -state

Pre-processing of Data: For Pre-processing of the data, I used TABLEAU PREP. Because of its flow chart feature it is quite easy to understand Dataset Joins and it has much powerful data cleansing tools than Tableau Desktop

1. First I joined all 3 order relates dataset using key “order_id”.

2. Join Product with Product Description using key “product_name”

3. Remove Duplicate Columns after joins

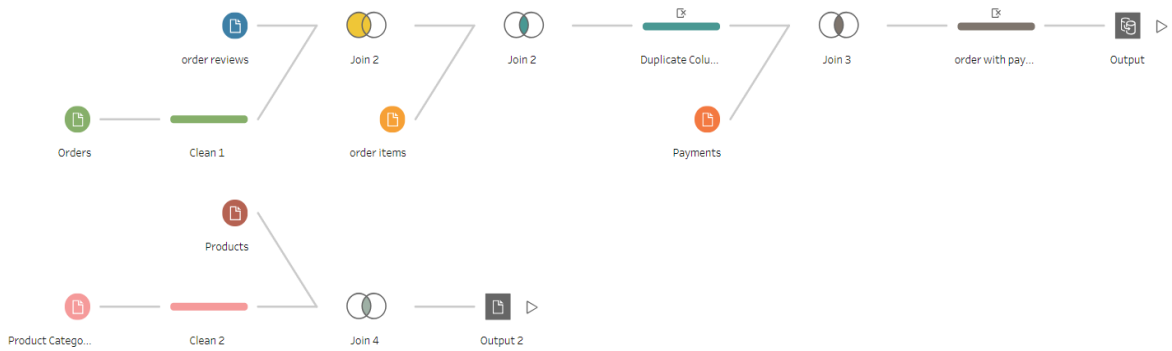


Fig 2. Tableau Prep Flow Chart

Project Objectives: - The dataset provides a great opportunity to let us explore in different angles like:

1. Most Bought Product Categories

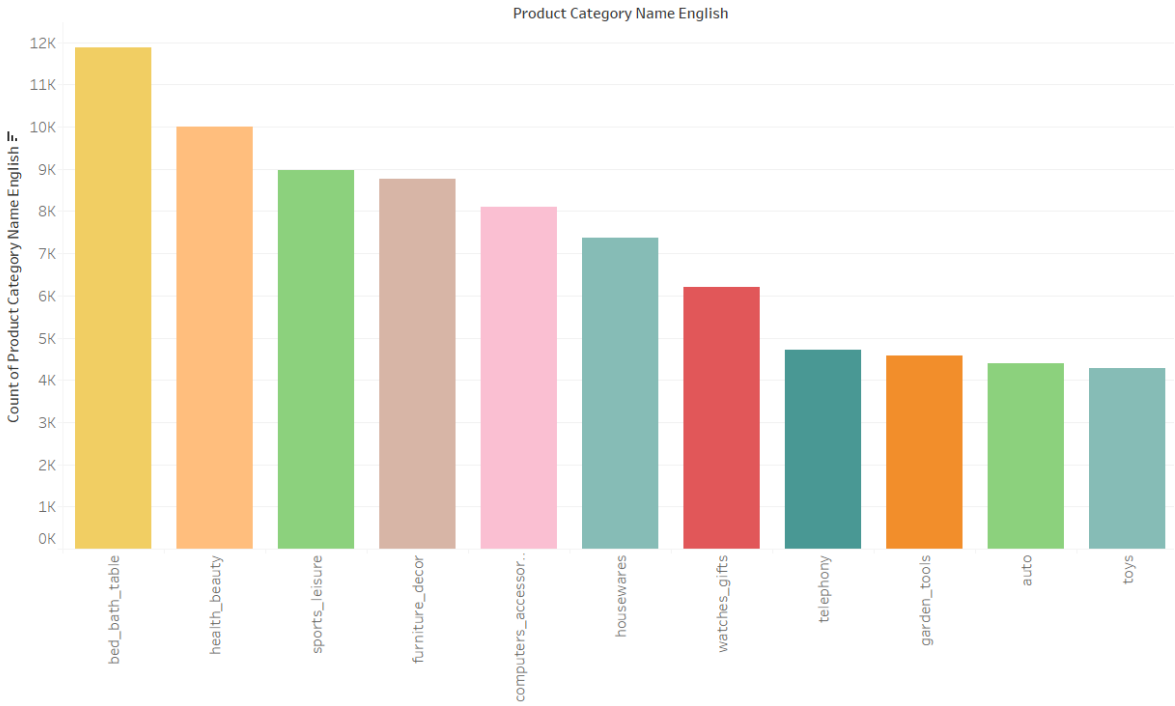
2. Finding trends or insights in orders and transactions

3. Cities or States with Highest Transaction Values

4. Average number of days between order and delivery

5. Mode of Transactions used by customers and some analysis on sellers like top sellers, no of products etc.

OBJ 1: - Most Bought Product Categories:



Bed Bath table, health beauty, sports are some categories that are bought most often by the customers. The difference in order count between bed bath table and health beauty is high.

OBJ 2: - Trends or Insights in Order and Transactions:

a. Let's check the order status

A majority of the orders in the database are delivered products, while some are in transition.

Order Status	
approved	3
canceled	551
delivered	1,13,617
invoiced	359
processing	362
shipped	1,219
unavailable	7

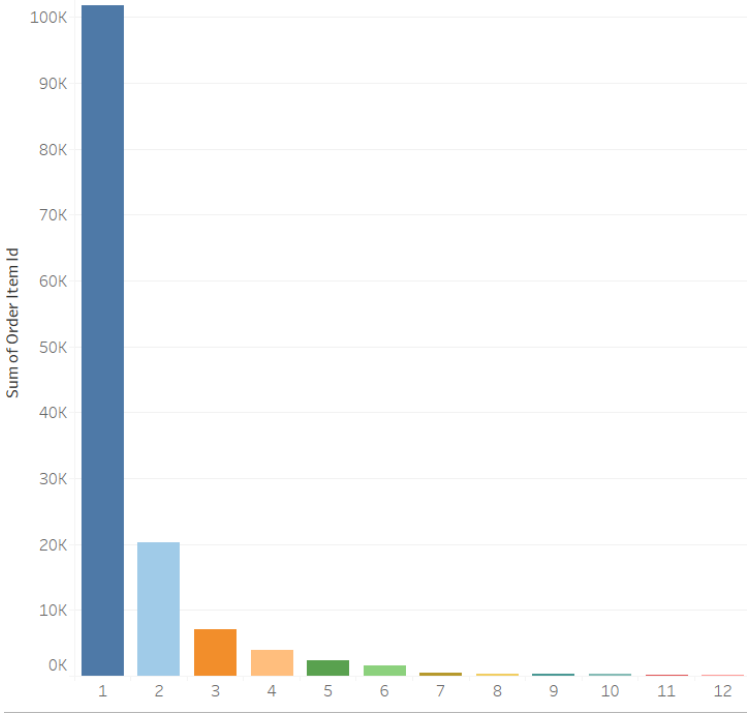
Let's take a look at the order value column. We see that there is no currency value specified and we assume that it is in local currency unit -BRL (Brazil's Currency).

Max order amount: - 13440

Min order amount: - 2

Average order value: - 129.2

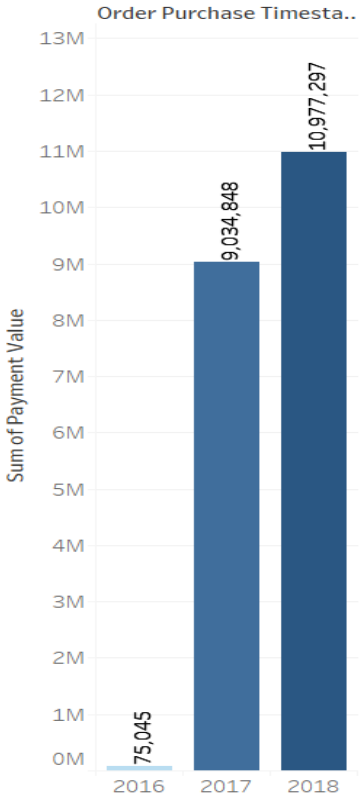
b. Number of products people usually order:



Most of them have ordered only 1 product. The number of people ordering more than 2 items is very less.

2c. Transaction by Years:

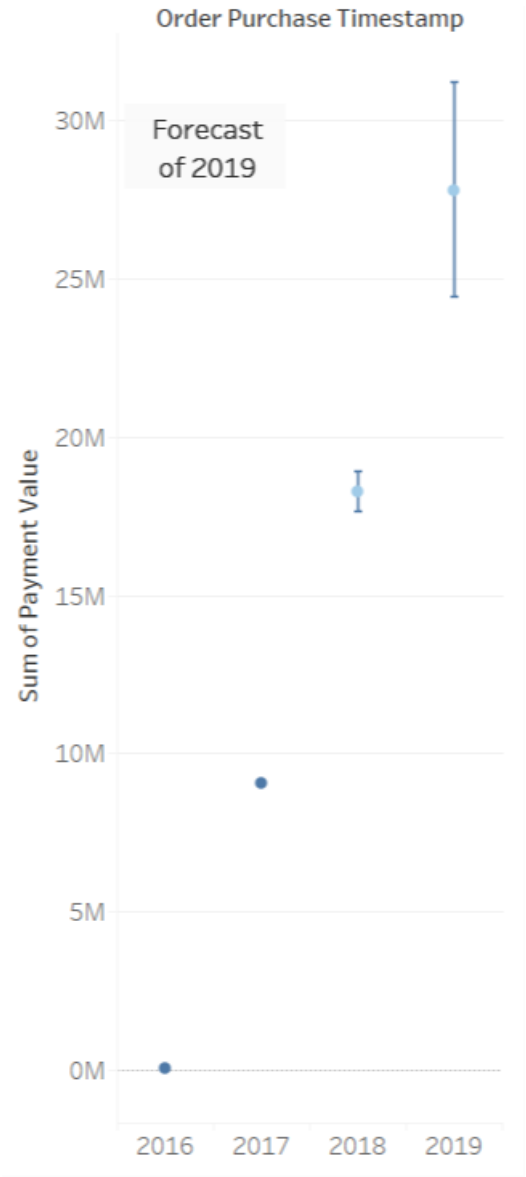
The data is available from Sept 2016 to Sept 2018. Therefore, we see a very low sales for the year 2016. Entire year data is available for 2017 and hence that year is on the higher side whereas sales till Sept for 2018 is plotted. For lack of entire data, we are unable to conclude any significant findings here.



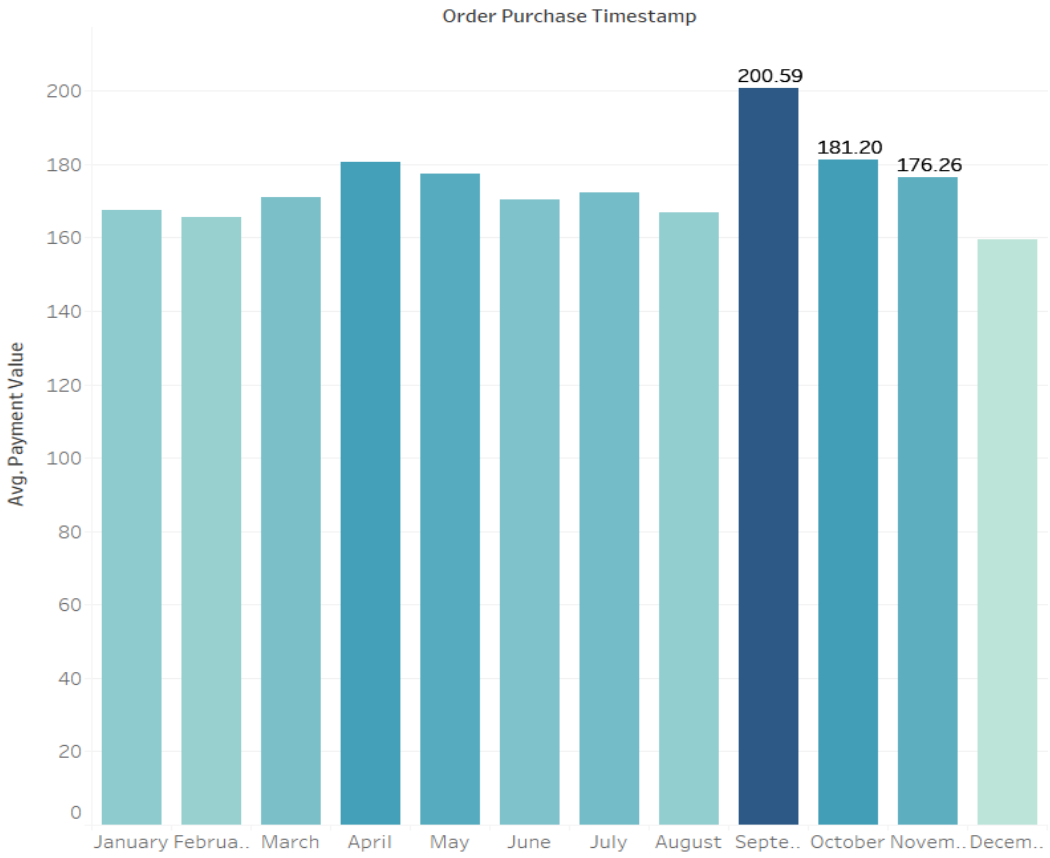
2d. Forecast of Transaction for 2019 :

From Box Plot: We find that most of the transaction fall below BRL 2000. There are more outliers for the year 2017 (since there is complete one-year data for this). The maximum transaction value has been 14000.

I also **forecast** the transaction value for 2019 on the basis of last three years

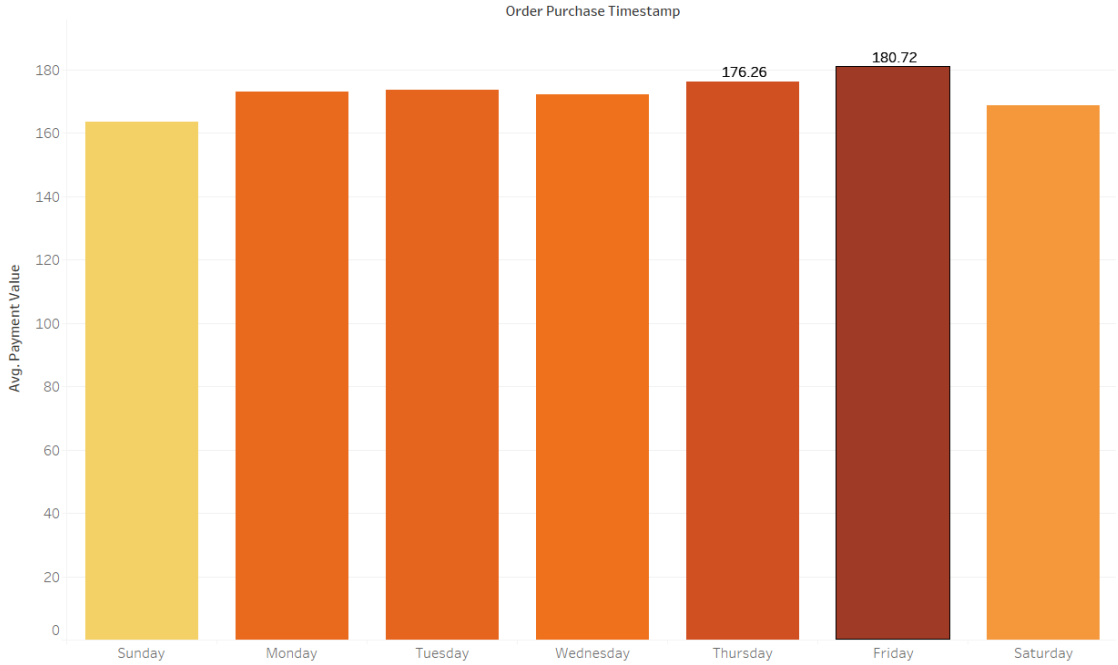


2e. **Average value of transactions per month**



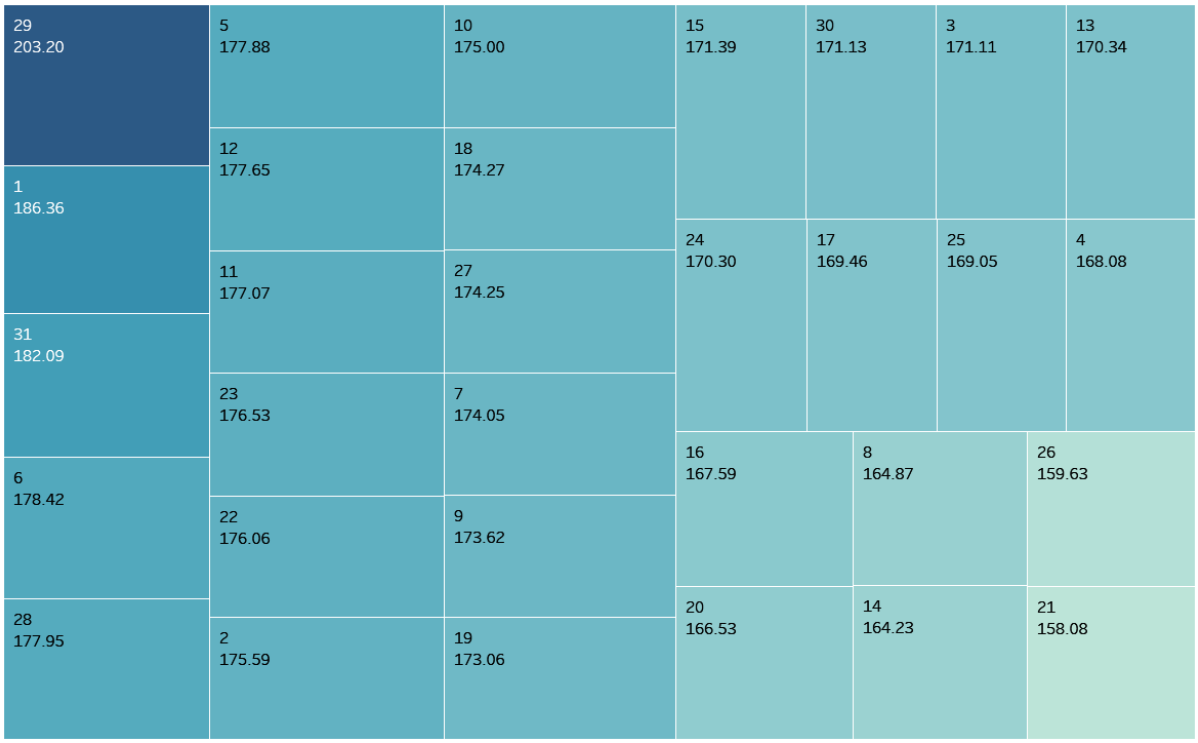
The average value of transaction is high during the month of October and September.

2f. **Average value of transactions by days of Week**



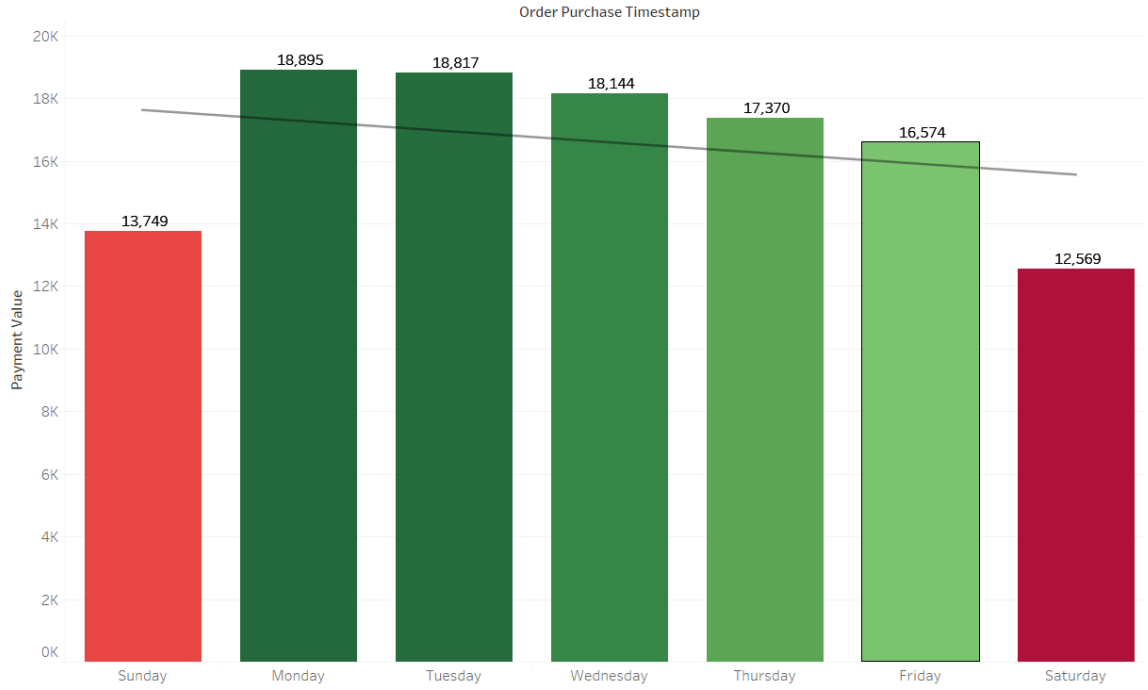
There seems to be not much trend observed during the day of the transaction . Lets check the frequency of the orders.

2g. **Average value of transactions by Days in a Month**



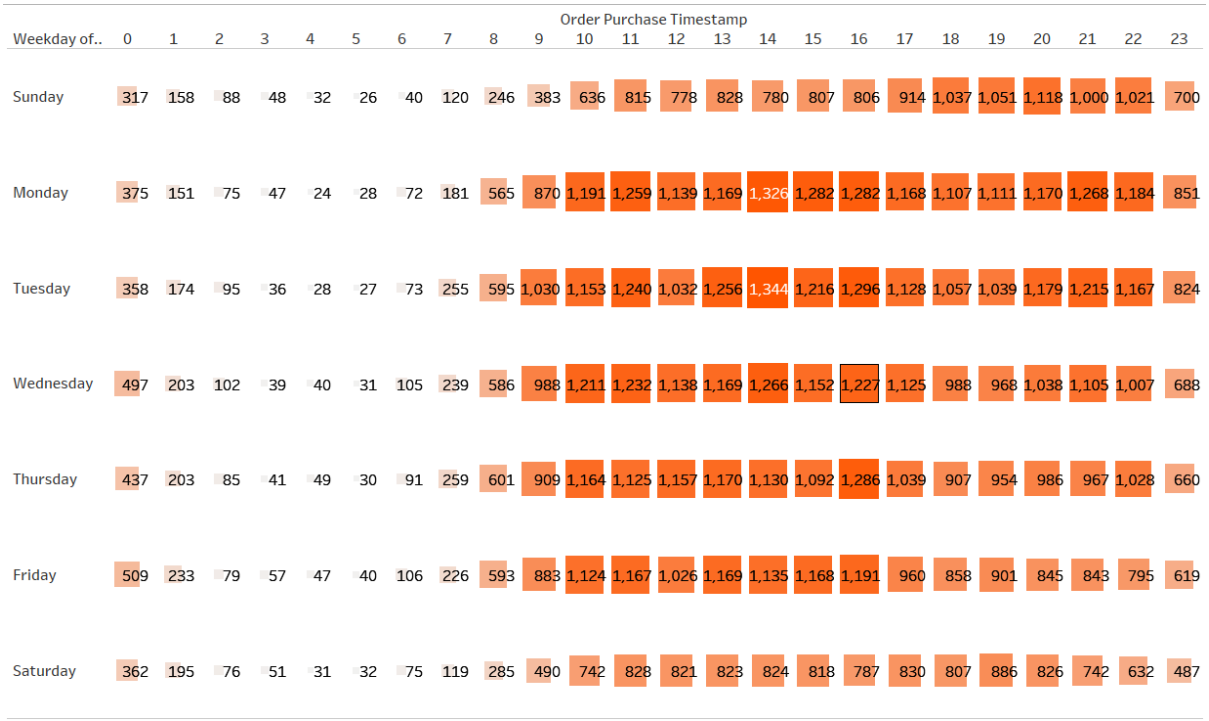
There seems to be not much trend observed during the day of the transaction. But the transactions are little high in end of month.

2h. **frequency of transactions by days of Week**



Transactions decreases as the week goes to end.

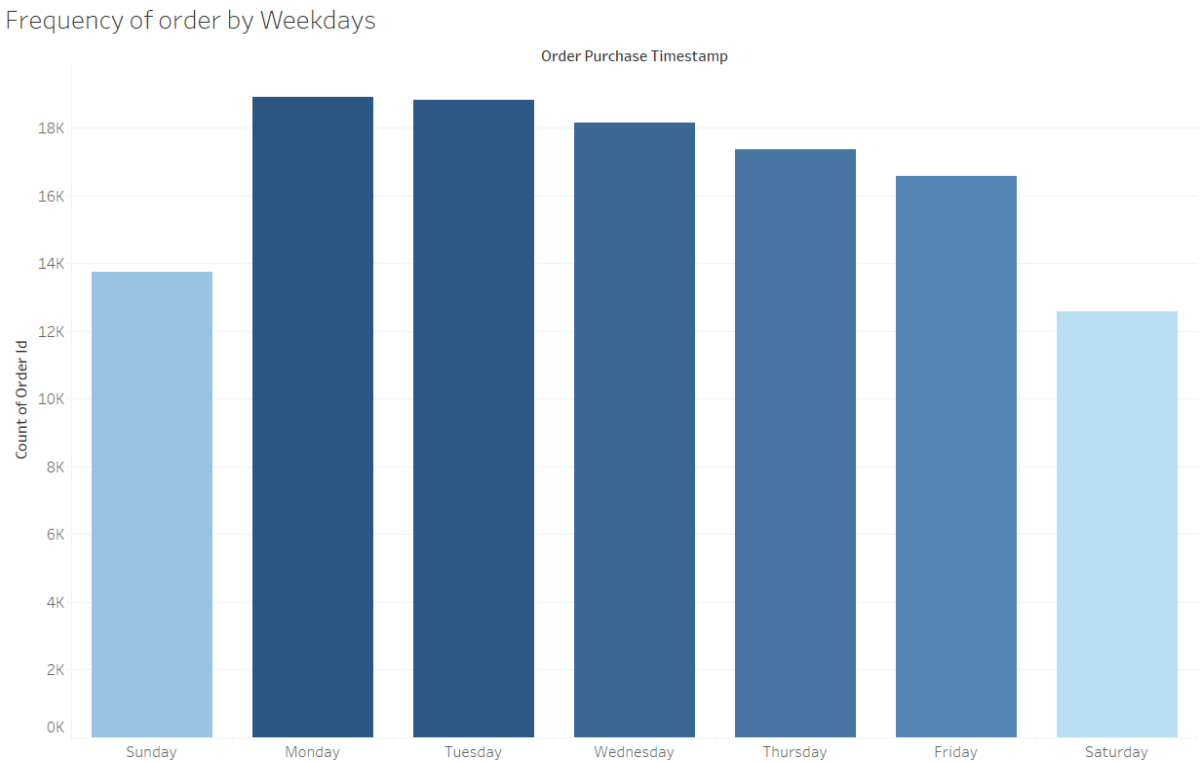
2i. **Average value of transactions by Weekdays and Hours**



Weekday of..	Order Purchase Timestamp																							
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Sunday	317	158	88	48	32	26	40	120	246	383	636	815	778	828	780	807	806	914	1,037	1,051	1,118	1,000	1,021	700
Monday	375	151	75	47	24	28	72	181	565	870	1,191	1,259	1,139	1,169	1,326	1,282	1,282	1,168	1,107	1,111	1,170	1,268	1,184	851
Tuesday	358	174	95	36	28	27	73	255	595	1,030	1,153	1,240	1,032	1,256	1,344	1,216	1,296	1,128	1,057	1,039	1,179	1,215	1,167	824
Wednesday	497	203	102	39	40	31	105	239	586	988	1,211	1,232	1,138	1,169	1,266	1,152	1,227	1,125	988	968	1,038	1,105	1,007	688
Thursday	437	203	85	41	49	30	91	259	601	909	1,164	1,125	1,157	1,170	1,130	1,092	1,286	1,039	907	954	986	967	1,028	660
Friday	509	233	79	57	47	40	106	226	593	883	1,124	1,167	1,026	1,169	1,135	1,168	1,191	960	858	901	845	843	795	619
Saturday	362	195	76	51	31	32	75	119	285	490	742	828	821	823	824	818	787	830	807	886	826	742	632	487

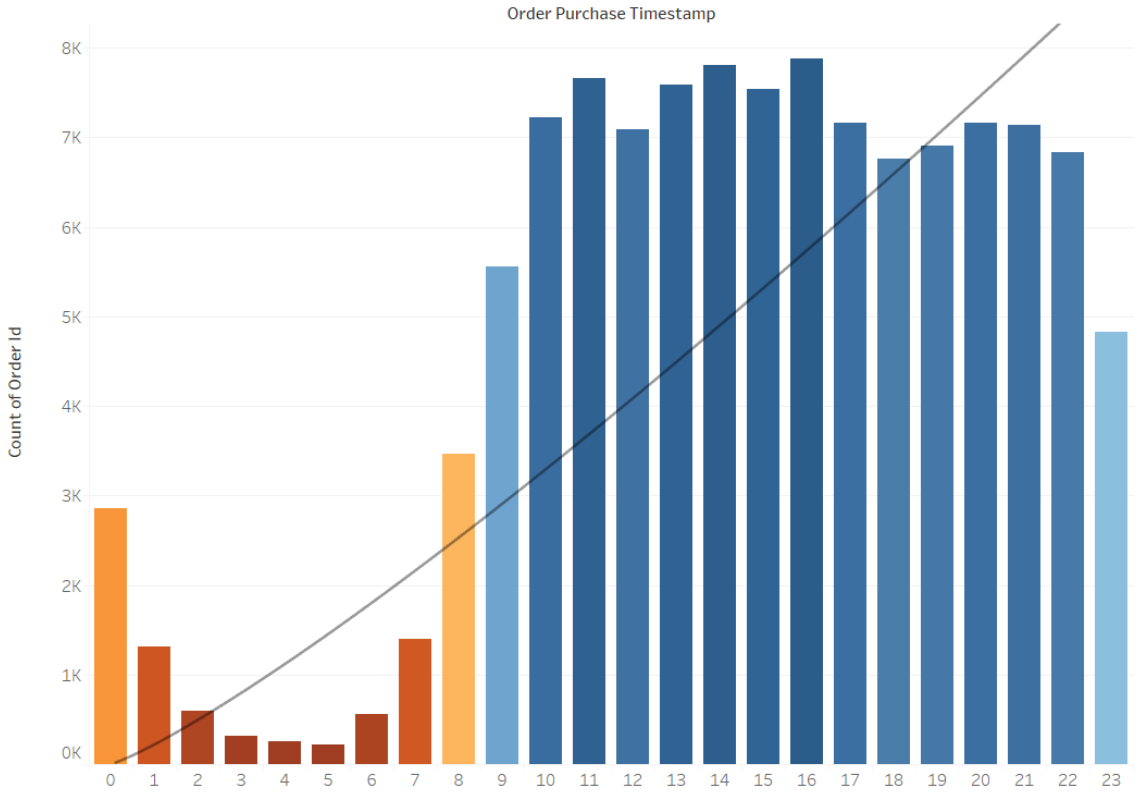
- As the day progresses, the number of orders placed increases.
- There is clearly a difference in the order frequency between weekdays and weekends.
- While during weekdays, the order frequency increases steady after 9 AM , the order frequency picks up only after 15:00 hrs during Sundays .

2j. **No of Orders in Weekdays**



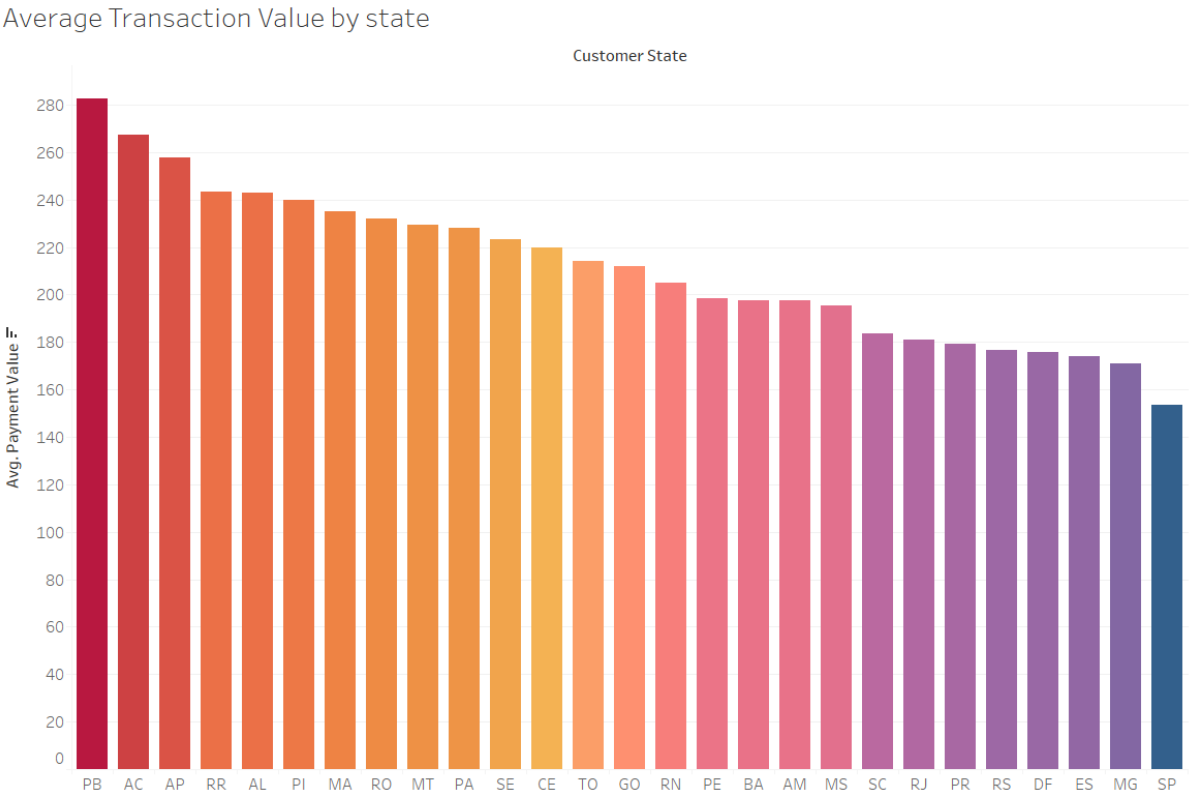
The frequency of the orders has been higher on Mon, Tue whereas the freq of orders is low during Saturday and Sundays. This means that during weekend people are not interested in online shopping going only by the frequency of the orders but combining this with the average value of transactions during the day there is a relatively high average value of transaction happening during Saturdays compared to other days.

2k. **Frequency of Orders in Hour of the Day**



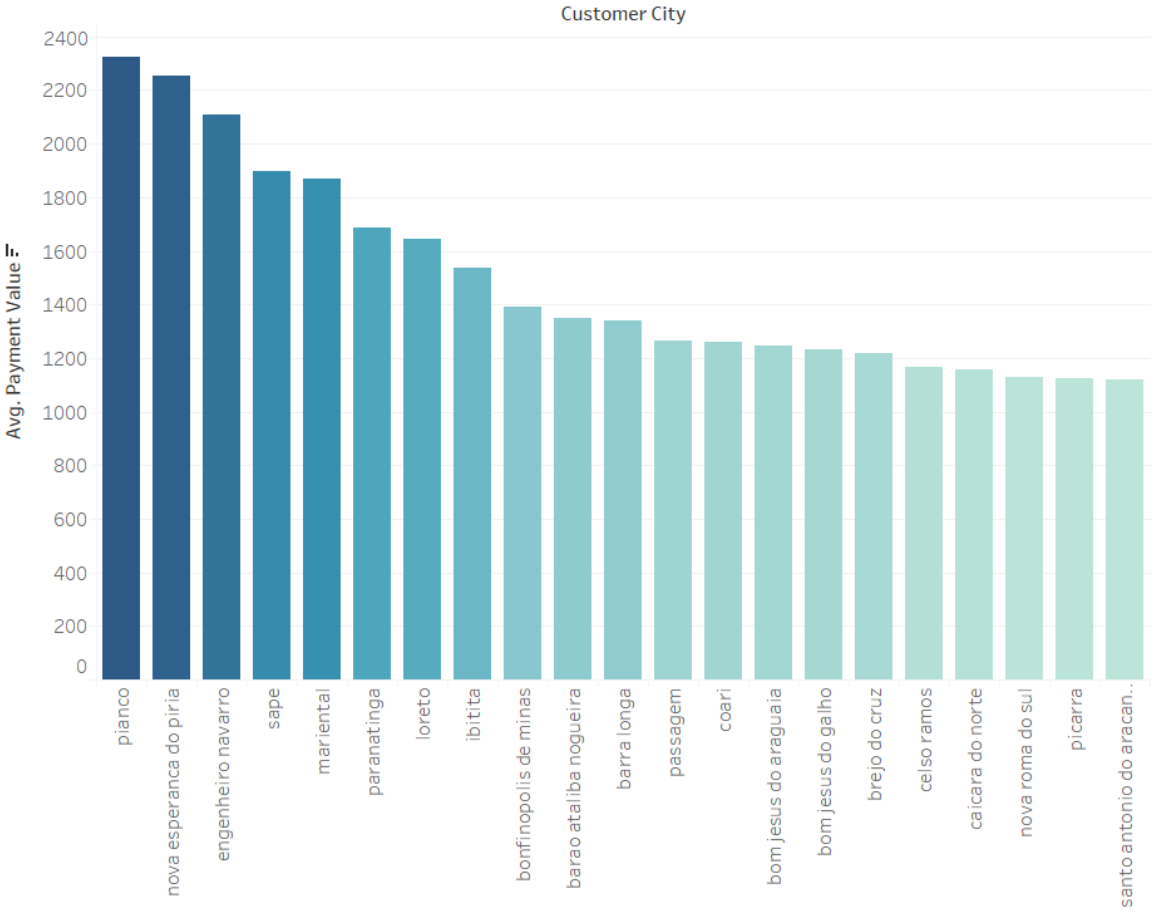
From the plot we see that the frequency of the order steadly rises as the day progresses and reaches the peak after noon and continues till 22 hrs . There is a dip in the transaction during evening time between 18-19 hrs and it sees a rise after that.

Objective 3. States, Cities with Highest Transaction Values



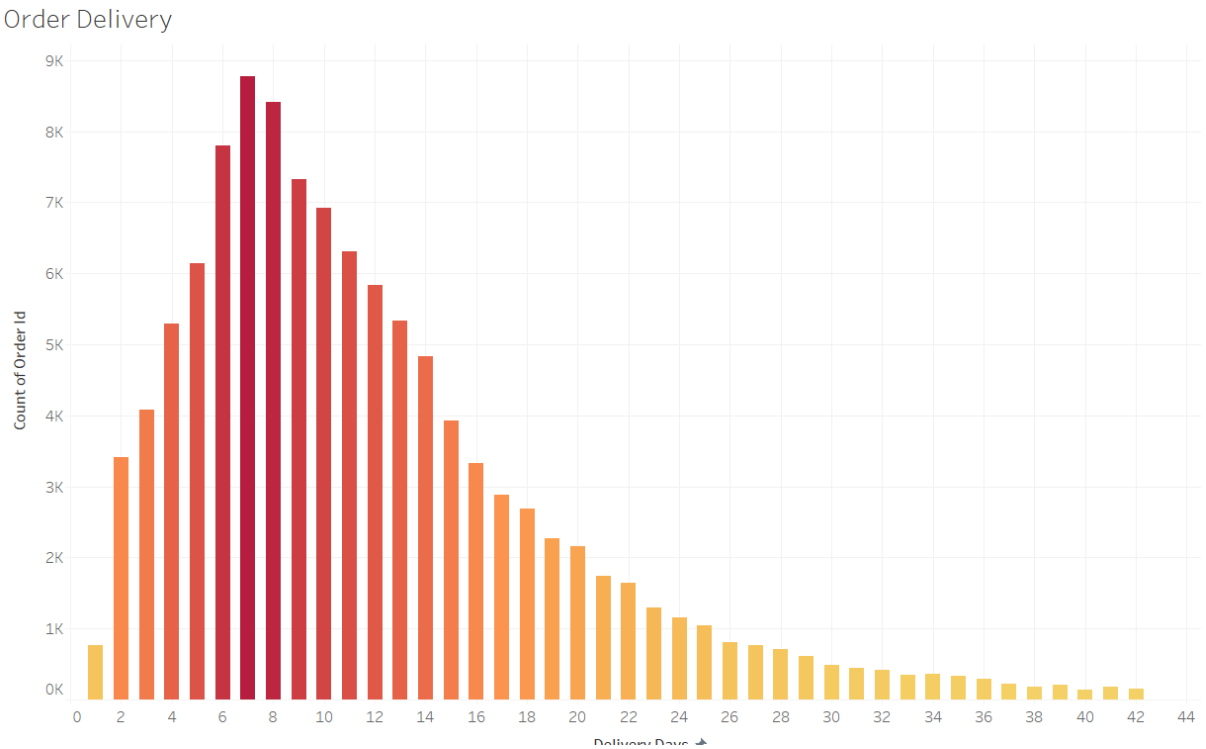
States Acre, Rondônia, Alagoas s have a higher transaction value whereas Minas Gerais, Sao Paulo have the lowest average transaction.

Average Transaction Value by City



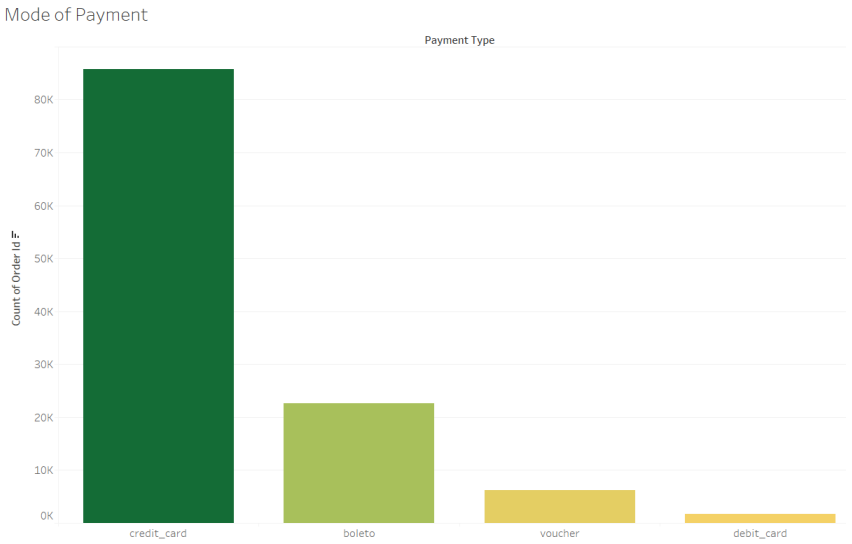
Cities like pianco, nova have a higher transaction value whereas Picarra and santo have the lowest average transaction.

Objective 4. Average number of days between order and delivery



A majority of the orders are getting delivered within a week whereas there were few orders that is taking over 1.5 months too.

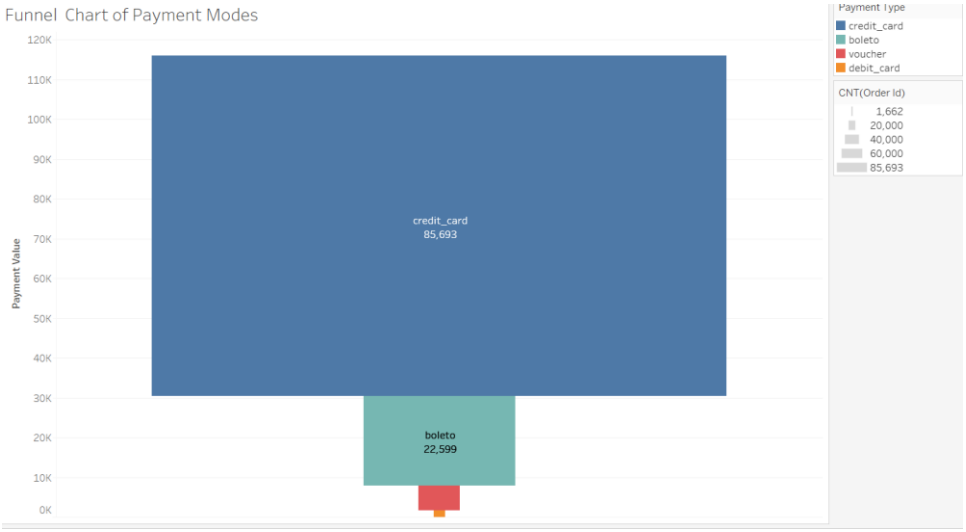
Objective 5. Mode of Transactions used by customers and some analysis on sellers like top sellers, no of products etc.



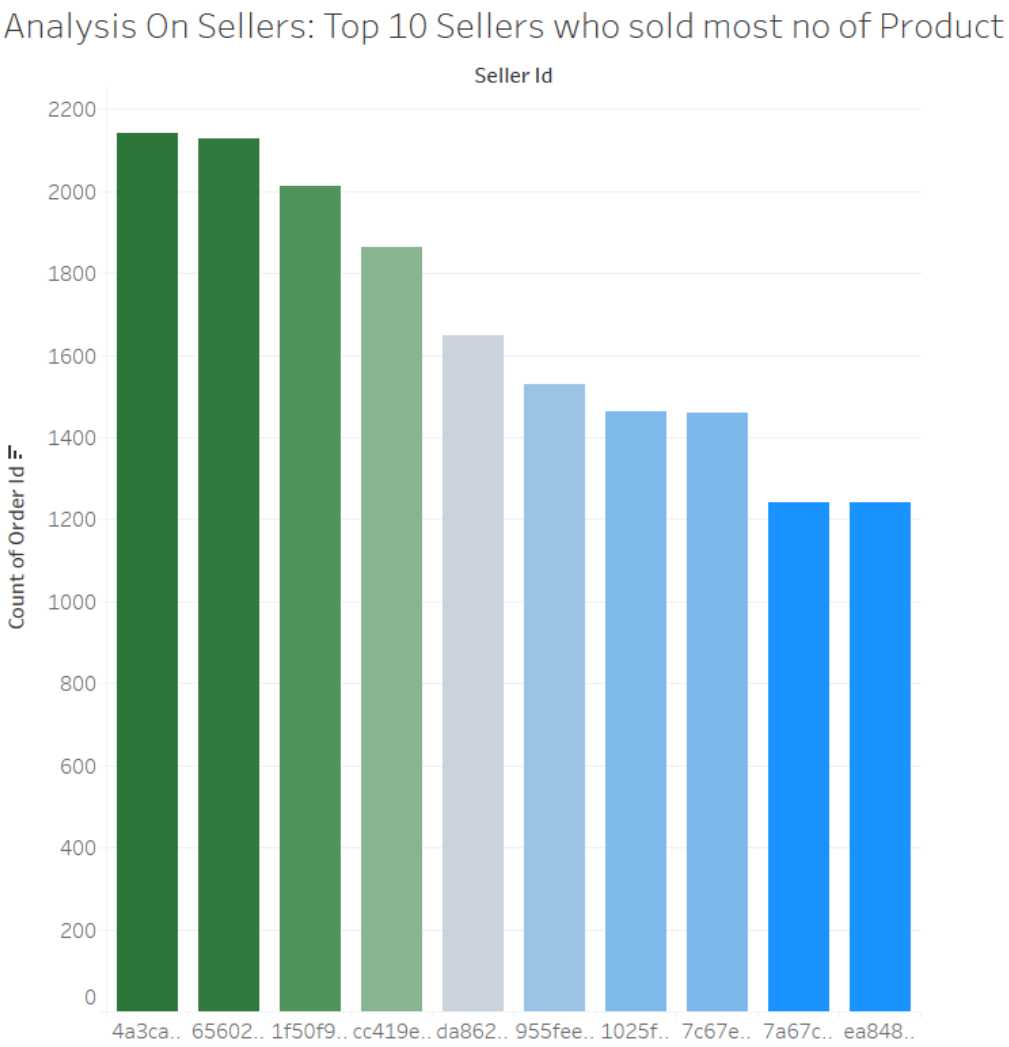
A large number of online buyers use credit card their preferred mode of payment followed by boleto .According to wiki , boleto is a payment method in Brazil regulated by FEBRABAN, short for Brazilian Federation of Banks. A boleto can be paid at ATMs, branch facilities and internet

banking of any Bank, Post Office, Lottery Agent and some supermarkets until its due date. After the due date it can only be paid at the issuer bank facilities.

Lets check the average value of transaction used for each type of payment.



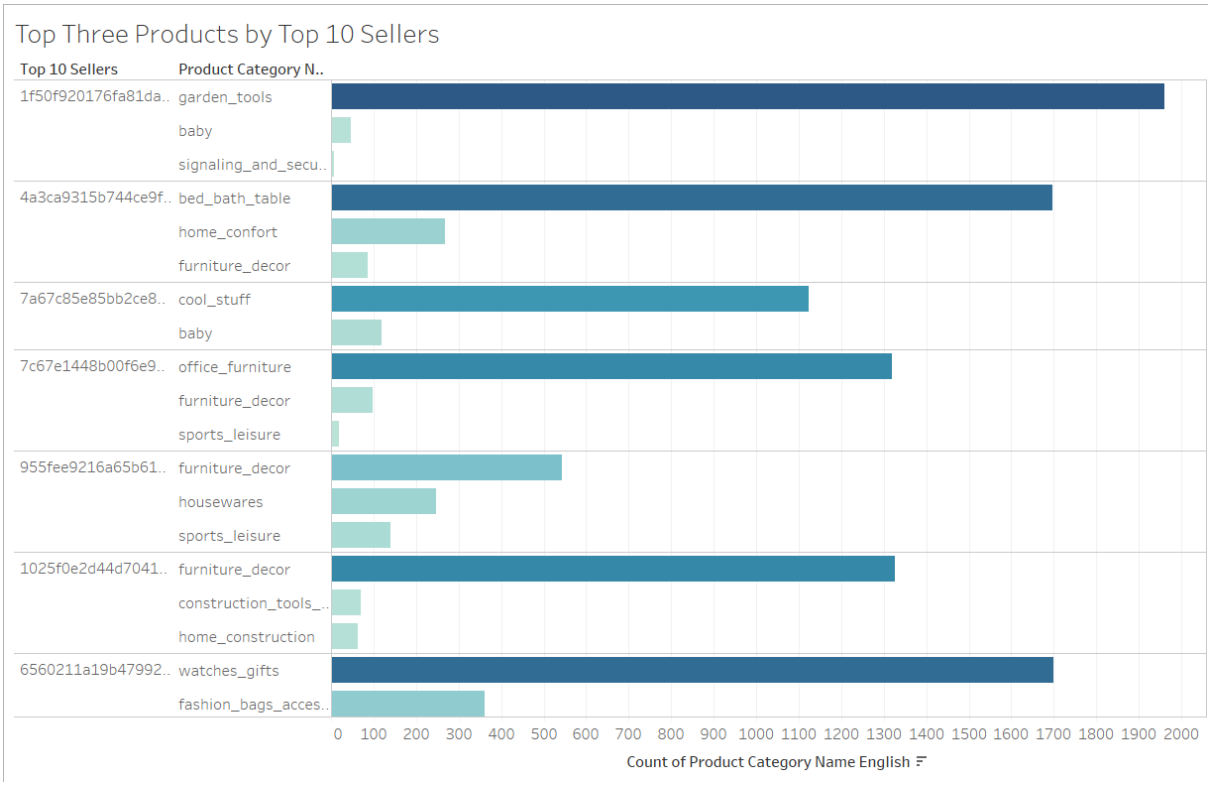
Sellers who have sold most number of products:



There are 23 products sold by one seller. Overall the top 10 sellers by product count have 2 to 23 products in their portfolio.

Portfolio of Top Sellers

Top 10 Sellers	
1f50f920176fa81da..	3
4a3ca9315b744ce9f..	7
7a67c85e85bb2ce8..	2
7c67e1448b00f6e9..	6
955fee9216a65b61..	23
1025f0e2d44d7041..	4
6560211a19b47992..	7
cc419e0650a3c5ba7..	4
da8622b14eb17ae2..	4
ea8482cd71df3c196..	2



Summary :

The following are some of the key points noted from the Olist E commerce analysis.

- Maximum order amount is BRL 13440 and Minumum order amount is BRL 2.
- Most of the time,the number of products ordered has always been < 3 .Bed Bath table,health beauty,sports are some categories that are bought most often by the customers.
- The frequency of the orders has been higher on Mon,Tue whereas the frequency of orders is low during Saturday and sundays.
- In a day,the number of transactions happening rises after 11 and continues till 22:00 Hrs.
- Average time taken for delivery is a week and maximum has gone up to 1.5 months.
- Credit card,boleto have been used for transactions of high value whereas people prefer using vouchers for low transaction values.
- The anonymised seller data tells that the top 15 sellers with maximum portfolio of products have 2 to 23 products in their category .