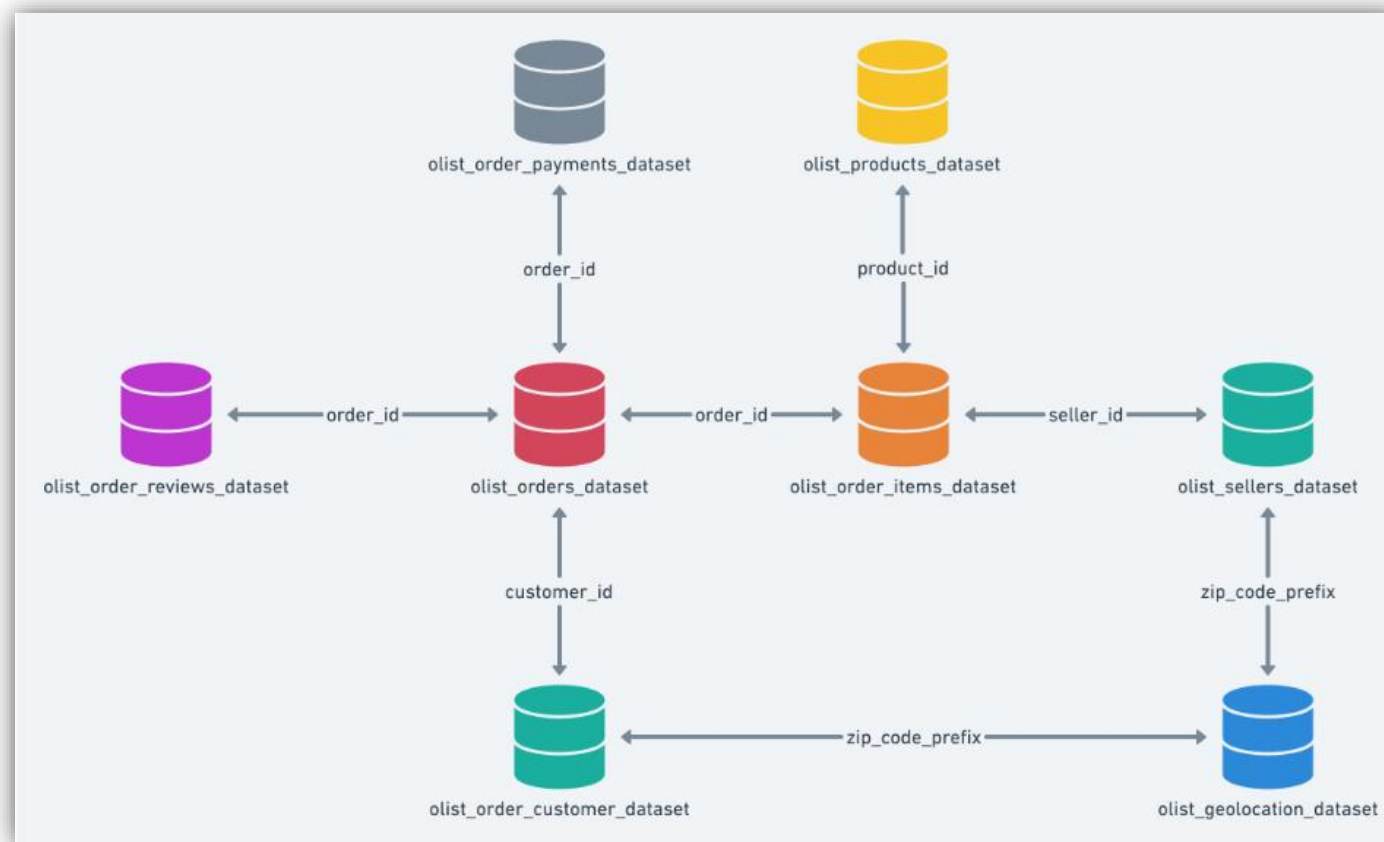


# Customer Segmentation

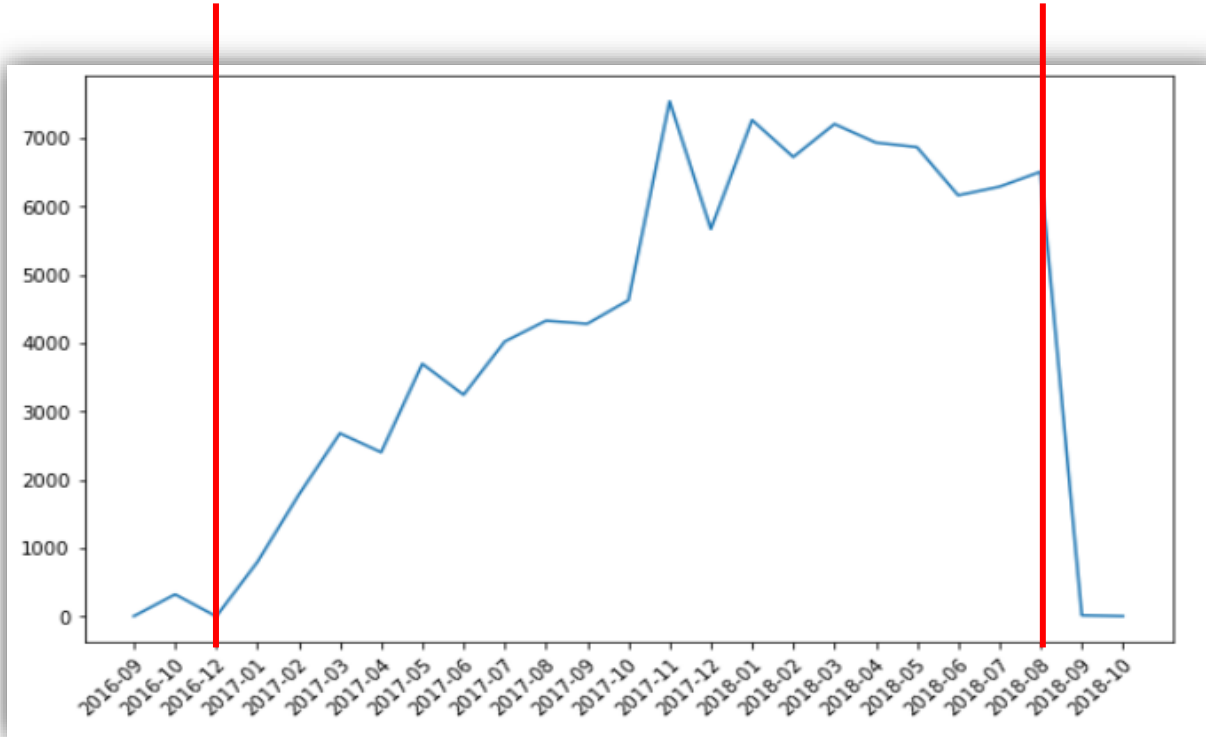


# Datasets

- **Customers:** This dataset has information about the customer and their location.
- **Geolocation:** This dataset has information about the Brazilian zip codes and their latitude/longitude coordinates.
- **Order Items:** This dataset has information about the items purchased within each order.
- **Order Payments:** This dataset has information about the order payment options.
- **Order Reviews:** This dataset has information about the reviews made by the customers.
- **Orders:** This dataset has information about all customer orders.
- **Products:** This dataset has information about all the products sold by Olist.
- **Sellers:** This dataset has information about the sellers that fulfilled the orders made at Olist.
- **Category Name Translation:** This dataset has English/Portuguese translations for all products sold at Olist.



Orders between Jan 2017 – Aug 2018

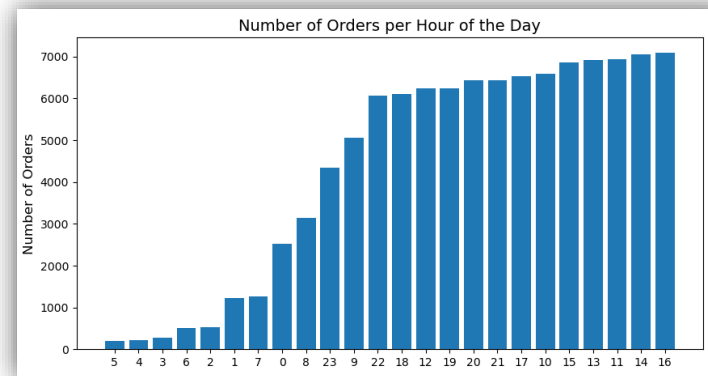
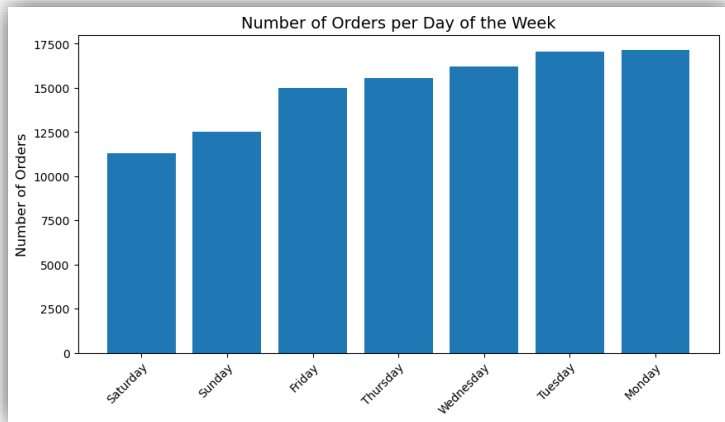
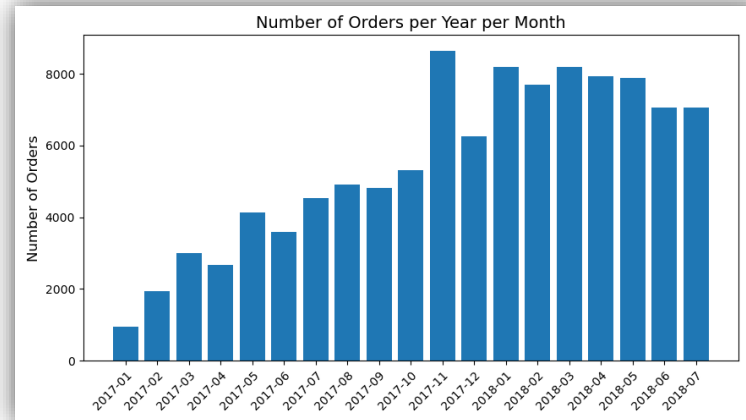
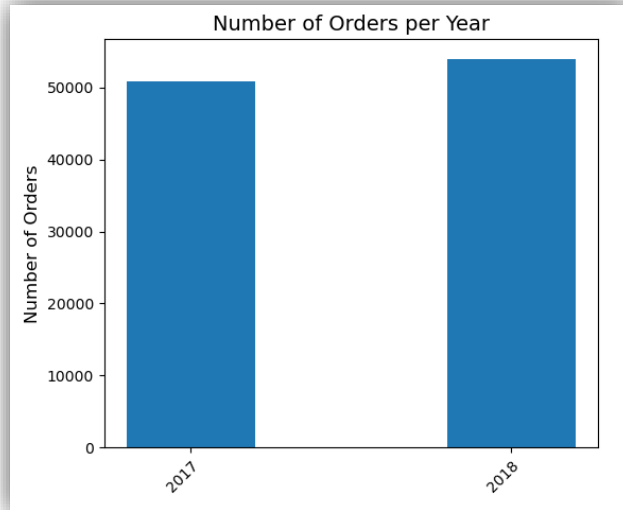


## Data Cleaning: Orders

- Dropped orders from 2016 and after August of 2018 as there were very small numbers
- Empty Values:
  - Order Approved At: Used the Order Purchase Timestamp for those orders that were delivered
  - Order Delivered Carrier Date: Used the Order Approved Date for those orders that were delivered
  - Order Delivered Customer Date: Used the difference between the Order Delivered Customer Date and the Order Delivered Carrier Date for all orders and got the median number of days. Added the days to the Order Delivered Carrier Date

# Exploratory Data Analysis



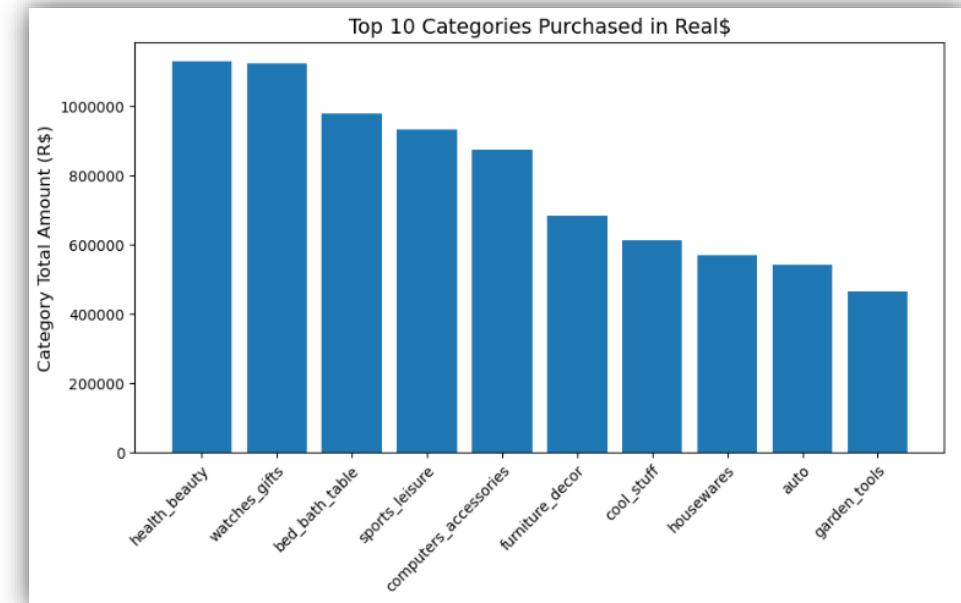
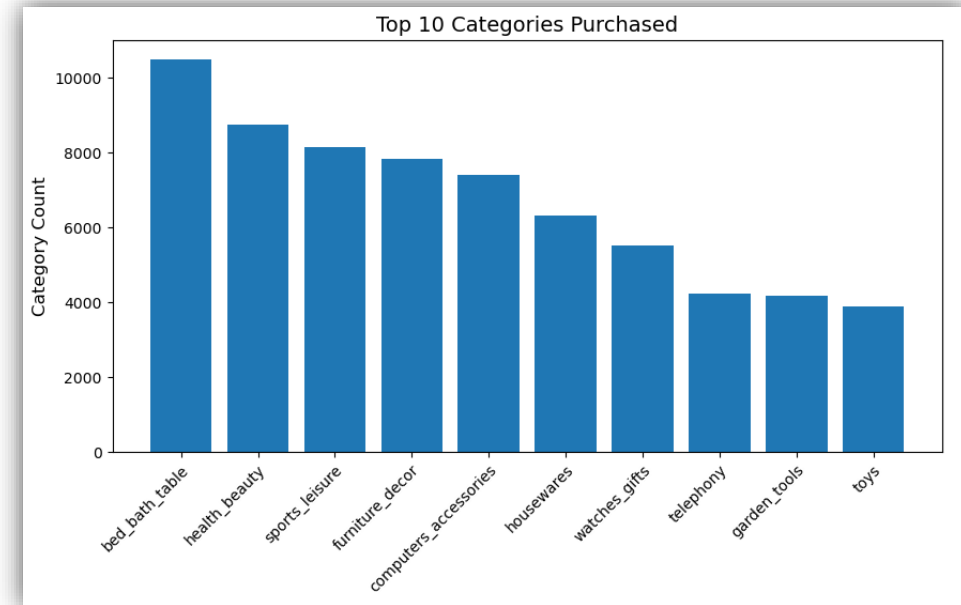


## Order Purchases Timeframes

- The number of orders grew between 2017 to 2018.
- November and January had the highest number of orders probably due to the holidays.
- March, April, and May were also good purchasing months.
- Weekdays were also more popular purchasing days as the beginning of the week was best probably due to items needed from the weekend and early evenings.

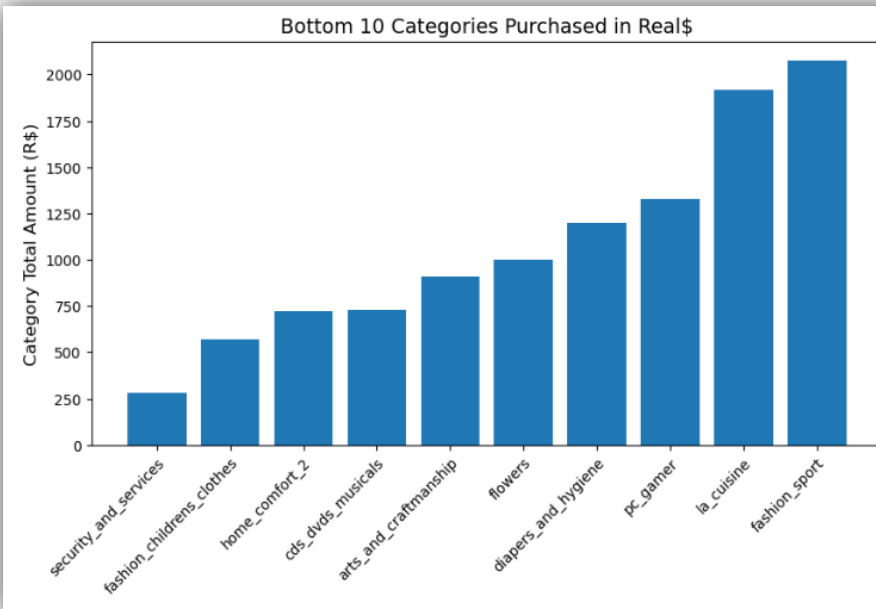
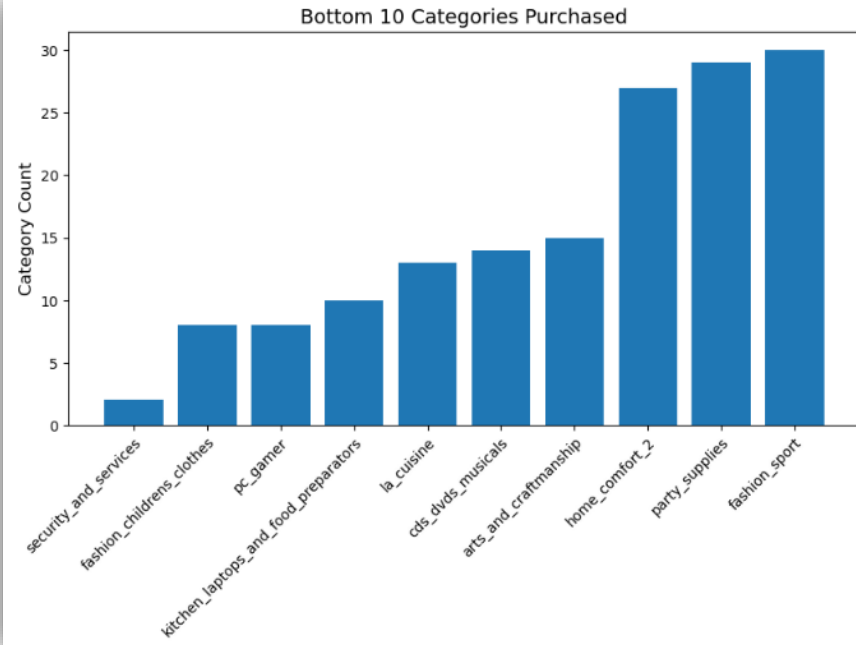
# Categories

- Most items that were purchased were from the Bed, Bath & Table category but the top category in the amount purchased was Health & Beauty.
- Most customers purchased items in the bedroom/bathroom area or that were used in the bedroom/bathroom area.



## Categories (con't)

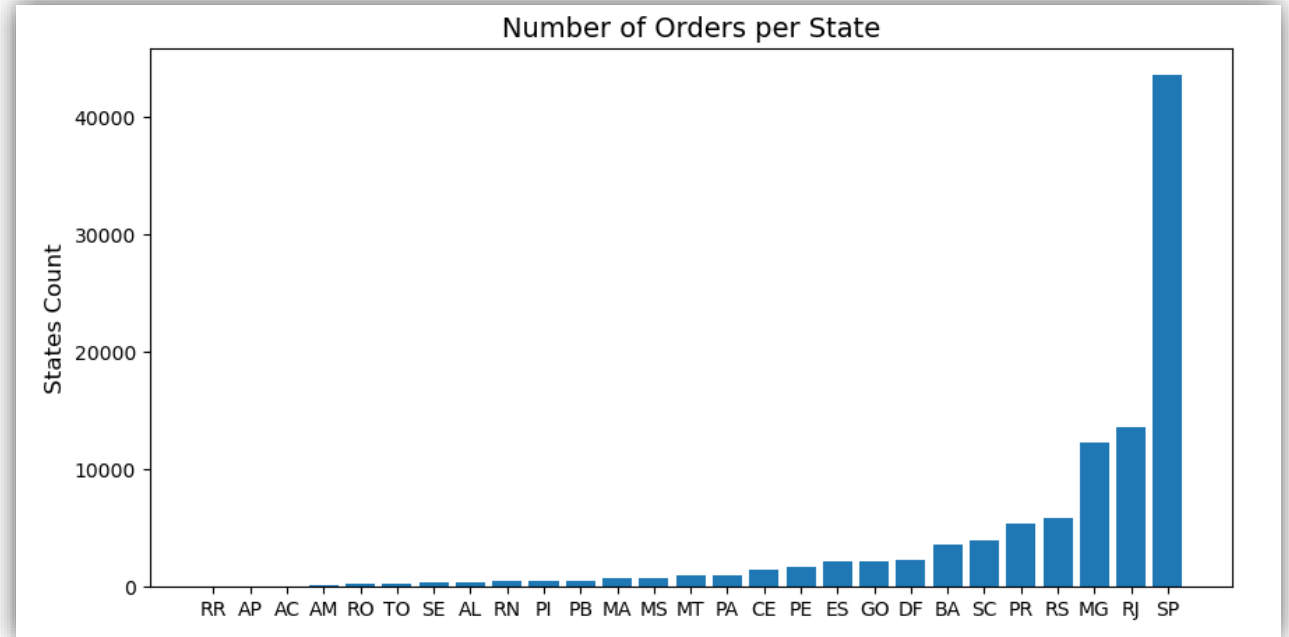
Looking at both the number of items purchased, and the amount spent, the Security and Services and Children's clothes categories are at the bottom. It could be that there are not many sellers of these items or more marketing needs to be allotted for these bottom categories.



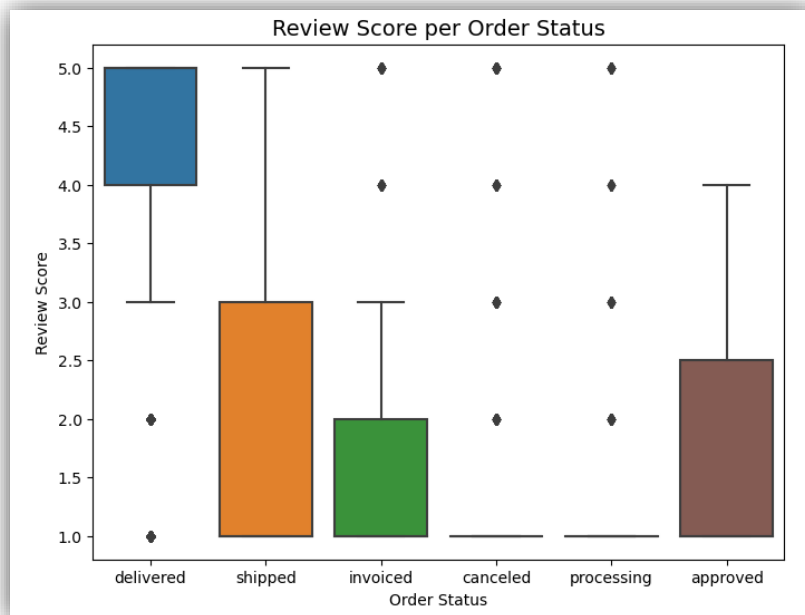
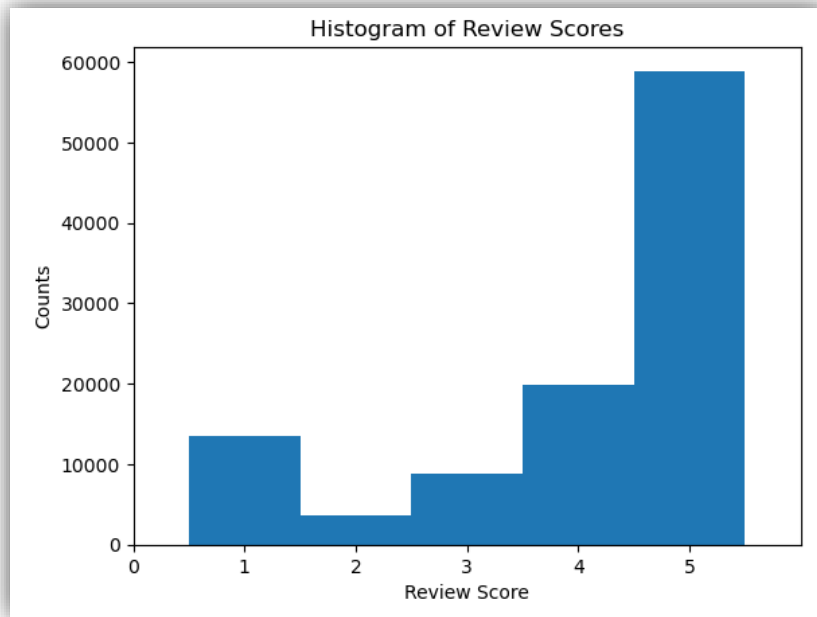
# States

---

The top State was São Paulo by far compared to the other States in Brazil. Marketing may investigate advertising more in States other than São Paulo.







## Review Score

- The review score of 5 is by far the most used by Olist customers which is a good sign
- Looking at the review score based on the order status does show purchases that were delivered by far getting a review score of 5 and the other statuses, not surprisingly, were lower.

Review Score = 5

|       | customer_id                      | review_score | order_status | price  | freight_value | product_category_name | cust_delivery_diff | est_delivery_diff |
|-------|----------------------------------|--------------|--------------|--------|---------------|-----------------------|--------------------|-------------------|
| 4201  | d6646ea91d8cd9fc7e6882a7068779d4 | 5            | delivered    | 81.99  | 14.51         | computers_accessories | 7                  | 20                |
| 4350  | 679f84ceb2ee4ca5bca0c3ea34647746 | 5            | delivered    | 59.90  | 17.67         | garden_tools          | 20                 | 4                 |
| 13293 | b4afeb58ac51bc903c5362286c6a5cfe | 5            | delivered    | 19.30  | 11.73         | drinks                | 11                 | 5                 |
| 49414 | 10de381f8a8d23fff822753305f71cae | 5            | delivered    | 65.49  | 16.22         | furniture_decor       | 19                 | 5                 |
| 54499 | b7770073b02ed1d626a027ce86a4ff82 | 5            | delivered    | 66.90  | 31.65         | sports_leisure        | 10                 | 44                |
| 67152 | 0d93f21f3e8543a9d0d8ece01561f5b2 | 5            | delivered    | 20.70  | 16.11         | housewares            | 8                  | 8                 |
| 67939 | 1ff773612ab8934db89fd5afa8afe506 | 5            | delivered    | 284.99 | 16.87         | drinks                | 14                 | 18                |
| 77657 | 20c93357daf05d1c3a092be59aea2c2b | 5            | delivered    | 20.50  | 16.91         | drinks                | 10                 | 14                |
| 90355 | 0e772d9e02b17408e716f35cd1dcc222 | 5            | delivered    | 36.99  | 11.85         | bed_bath_table        | 10                 | 13                |
| 96279 | adb32467ecc74b53576d9d13a5a55891 | 5            | delivered    | 51.00  | 1.20          | garden_tools          | 14                 | 20                |

## Customers

Looking at the two charts, you can't distinguish between those that gave a high review score to those that gave a low review score.

|        | customer_id                      | review_score | order_status | price  | freight_value | product_category_name | cust_delivery_diff | est_delivery_diff |
|--------|----------------------------------|--------------|--------------|--------|---------------|-----------------------|--------------------|-------------------|
| 6639   | 91f92cfee46b79581b05aa974dd57ce5 | 1            | delivered    | 108.00 | 15.52         | watches_gifts         | 11                 | 13                |
| 19725  | d5f2b3f597c7ccafbb5cac0bcc3d6024 | 1            | delivered    | 59.00  | 13.43         | garden_tools          | 14                 | 10                |
| 24699  | 4a60b2ce1ee8c7b828e4bbcca5b86b41 | 1            | delivered    | 137.90 | 38.81         | computers_accessories | 14                 | 1                 |
| 25887  | be1c4e52bb71e0c54b11a26b8e8d59f2 | 1            | delivered    | 49.99  | 7.10          | bed_bath_table        | 5                  | 11                |
| 37089  | 78fc46047c4a639e81ff65f0396e02fe | 1            | delivered    | 109.97 | 34.04         | furniture_living_room | 5                  | 13                |
| 46421  | be1b70680b9f9694d8c70f41fa3dc92b | 1            | delivered    | 100.00 | 10.12         | computers_accessories | 10                 | 2                 |
| 68674  | cb87122c4871e202777cf243fba2d12  | 1            | delivered    | 149.91 | 0.14          | computers_accessories | 11                 | 23                |
| 77621  | a7693fba2ff9583c78751f2b66ecab9d | 1            | delivered    | 29.99  | 7.78          | telephony             | 8                  | 5                 |
| 102470 | fc3d1daec319d62d49bfb5e1f83123e9 | 1            | delivered    | 1.20   | 7.89          | health_beauty         | 14                 | -4                |
| 102724 | 9eb3d566e87289dcb0acf28e1407c839 | 1            | delivered    | 5.31   | 15.23         | housewares            | 10                 | 9                 |

Review Score = 1

# Final Dataset for Modeling

|        | customer_id                       | order_purchase_timestamp | order_id                         | price  |
|--------|-----------------------------------|--------------------------|----------------------------------|--------|
| 0      | 9ef432eb6251297304e76186b10a928d  | 2017-10-02 10:56:33      | e481f51cbdc54678b7cc49136f2d6af7 | 29.99  |
| 1      | a20e8105f23924cd00833fd87daa0831  | 2017-08-15 18:29:31      | 128e10d95713541c87cd1a2e48201934 | 29.99  |
| 2      | 26c7ac168e1433912a51b924fbd34d34  | 2017-08-02 18:24:47      | 0e7e841ddf8f8f2de2bad69267ecfbcf | 29.99  |
| 3      | 53904ddbea91e1e92b2b3f1d09a7af86  | 2017-10-23 23:26:46      | bfc39df4f36c3693ff3b63fcbea9e90a | 29.99  |
| 4      | b0830fb4747a6c6d20dea0b8c802d7ef  | 2018-07-24 20:41:37      | 53cdb2fc8bc7dce0b6741e2150273451 | 118.70 |
| ...    | ...                               | ...                      | ...                              | ...    |
| 104777 | 609b9fb8cad4fe0c7b376f77c8ab76ad  | 2017-08-10 21:21:07      | e8fd20068b9f7e6ec07068bb7537f781 | 356.00 |
| 104778 | 609b9fb8cad4fe0c7b376f77c8ab76ad  | 2017-08-10 21:21:07      | e8fd20068b9f7e6ec07068bb7537f781 | 356.00 |
| 104779 | a2f7428f0cafbcb8e59f20e1444b67315 | 2017-12-20 09:52:41      | cfa78b997e329a5295b4ee6972c02979 | 55.90  |
| 104780 | 39bd1228ee8140590ac3aca26f2dfe00  | 2017-03-09 09:54:05      | 9c5dedf39a927c1b2549525ed64a053c | 72.00  |
| 104781 | edb027a75a1449115f6b43211ae02a24  | 2018-03-08 20:57:30      | 66dea50a8b16d9b4dee7af250b4be1a5 | 68.50  |

104782 rows × 4 columns



Modeling

## RFM Analysis + K-Means Clustering

### Recency, Frequency, and Monetary Analysis (RFM)

Looks at historical customer behavior to predict how might a new customer act in the future

Recency:

How recently has a customer purchased?

Frequently:

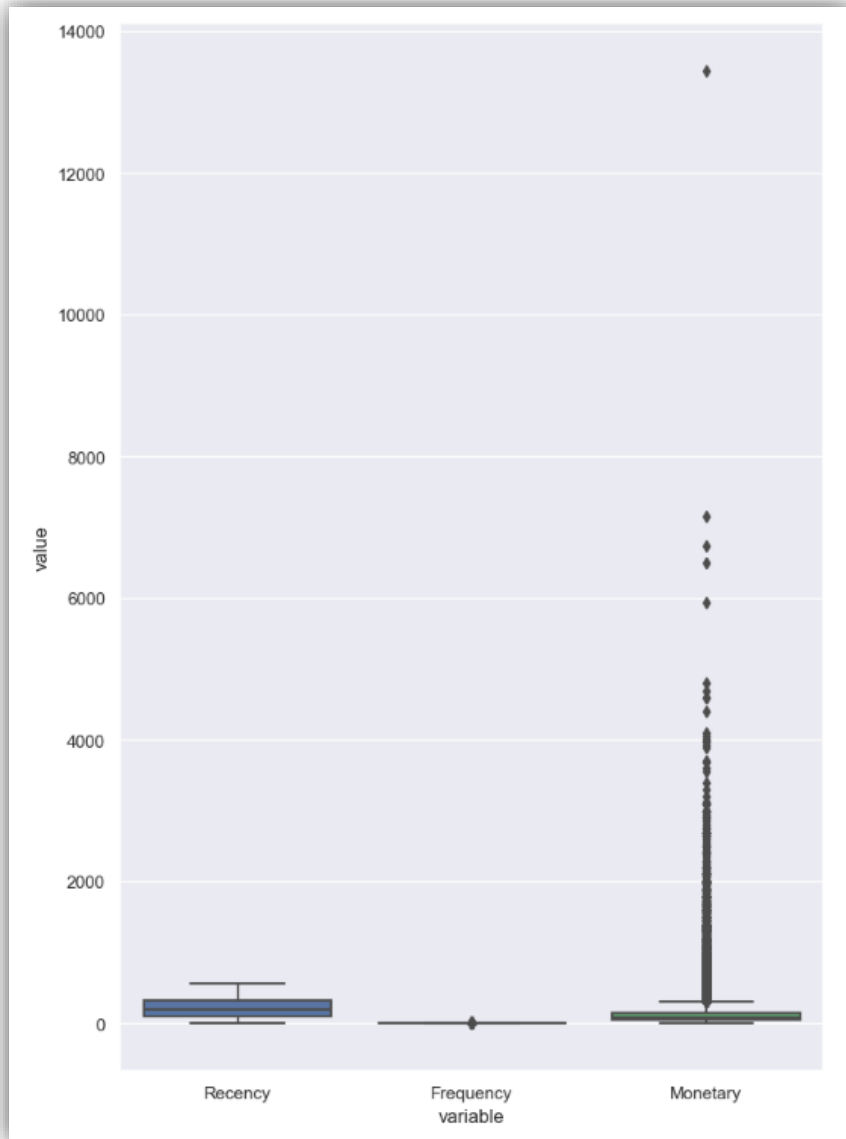
How many orders did a customer purchase?

Monetary:

How much money has a customer spent?

### K-Means Clustering

Grouping customers based on historical data according to how recently they purchased, how frequently they purchased and how much money they spent on purchases.



## RFM Analysis

### Recency

Mean: 256 days

Min: 0 days

Max: 572 days

### Frequency

Mean: 1.12 times

Min: 1 time

Max: 22 times

### Monetary Value

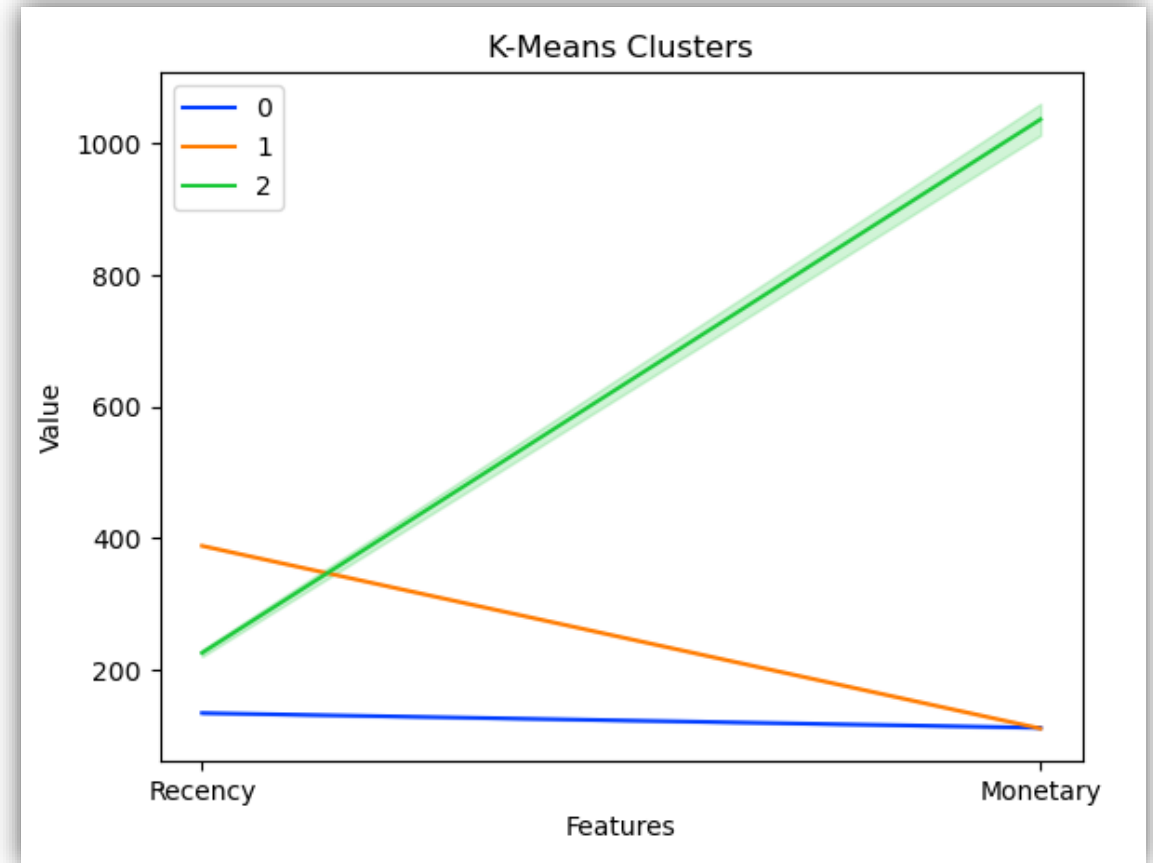
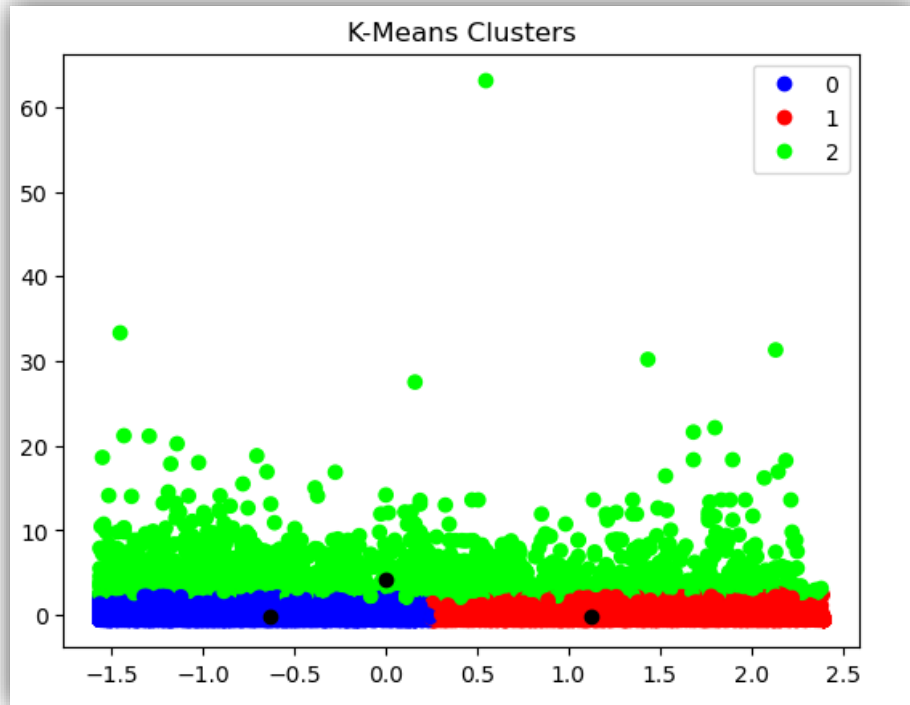
Mean: R\$138

Min: R\$0.85

Max: R\$13,440

# K-Means Analysis

Number of clusters: 3



Looking at the two graphs, we can see the differences between the different clusters.

- **Cluster 0:** Customers that haven't purchased items in a long time and their purchases amount to a very low dollar amount.
- **Cluster 1:** Customers that have purchased recently but their purchases amount to a very low dollar amount
- **Cluster 2:** These are customers that have purchased items somewhat recently but these customers' purchases amount to a very high dollar amount.