
Olist

Overview

Olist is an ecommerce company from Brazil. It provides an online platform for business-to-customer.

While the transaction to customer is done online. Olist has a sales team where they look for a potential seller to sell in their platform.

Total Seller
3,090

Total Customer
96,096

Total Order
96,561

State Served
29

12 Days
Avg Delivery Time

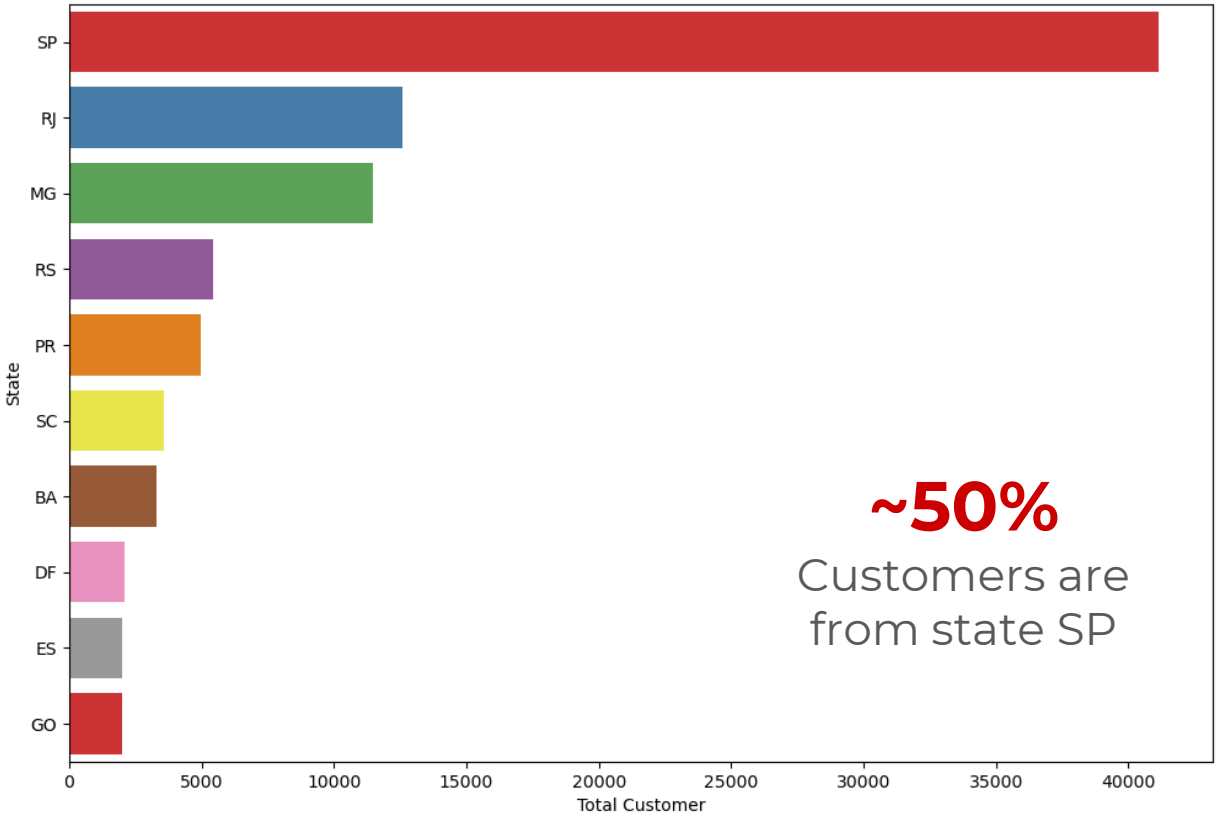
72
Total Product Category

Data Recorded From
Sept 2016

Objective

- Customer segmentation.
- Predicting future sales & total order (given that all the features, e.g: sellers' delivery duration stay the same).

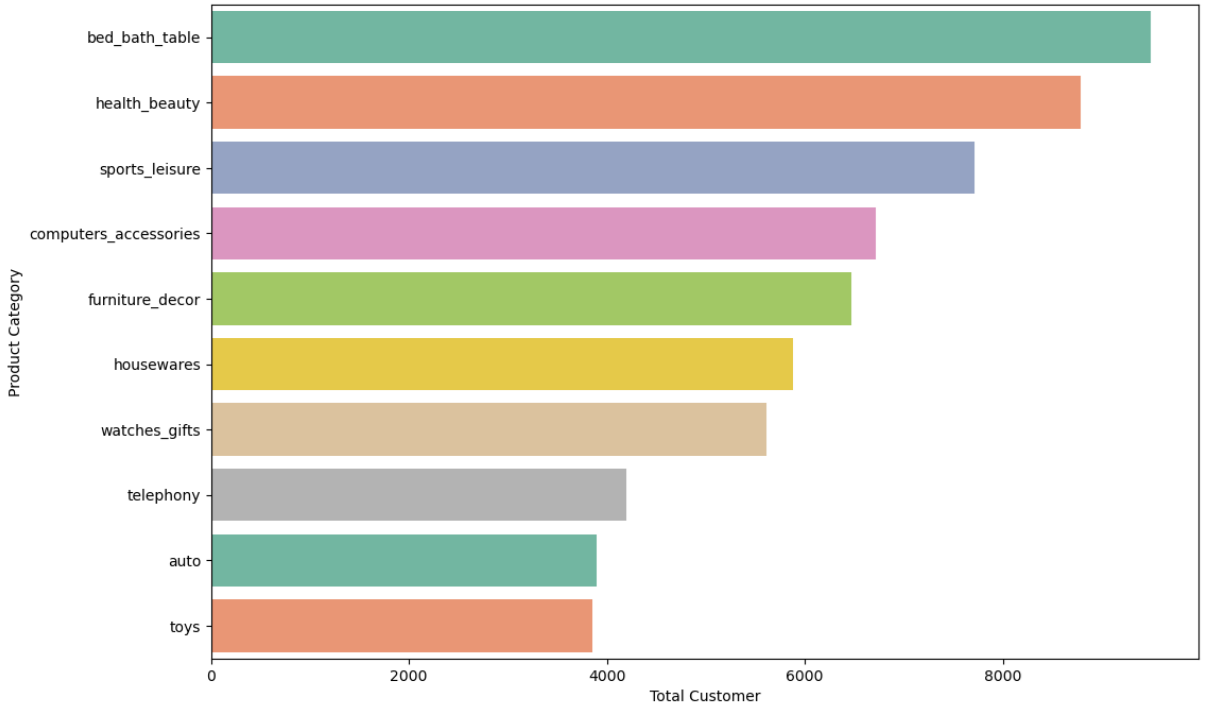
Characteristic of Customer



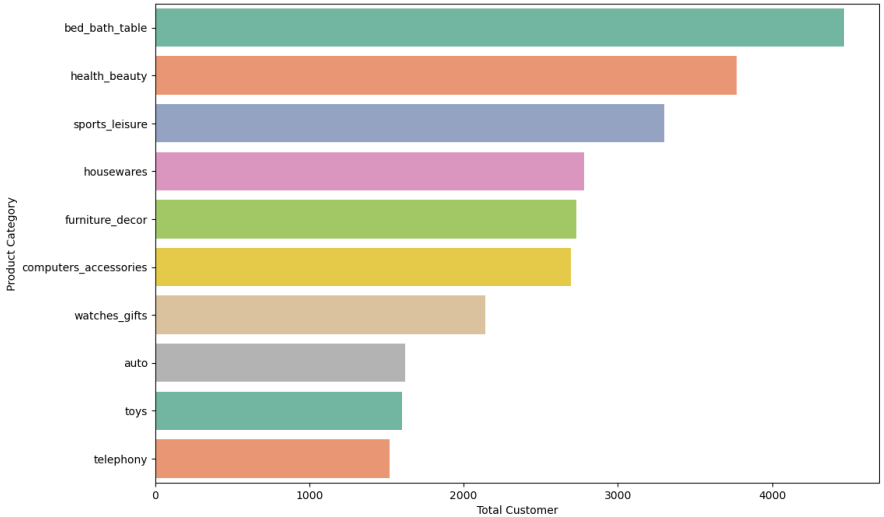
~50%

Customers are from state SP

Product



Top-ordered product by its total order is **bed_bath_table**.

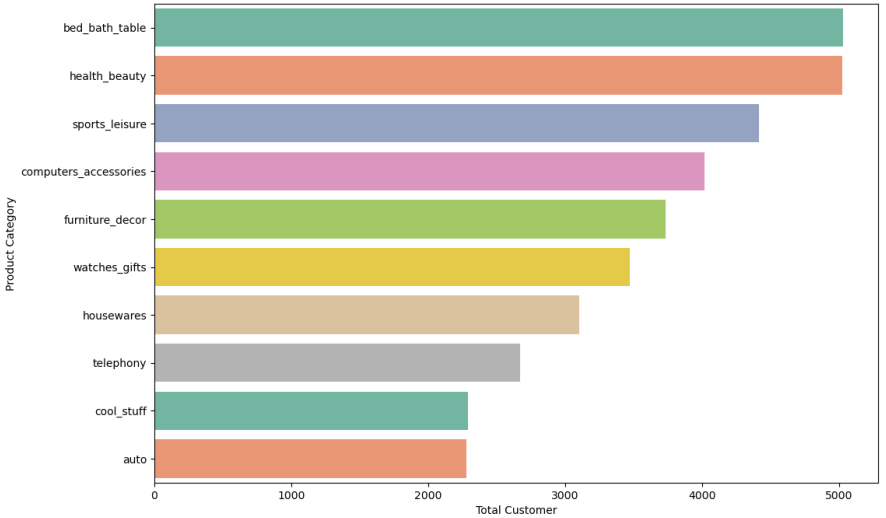


Total Order by Product in State SP

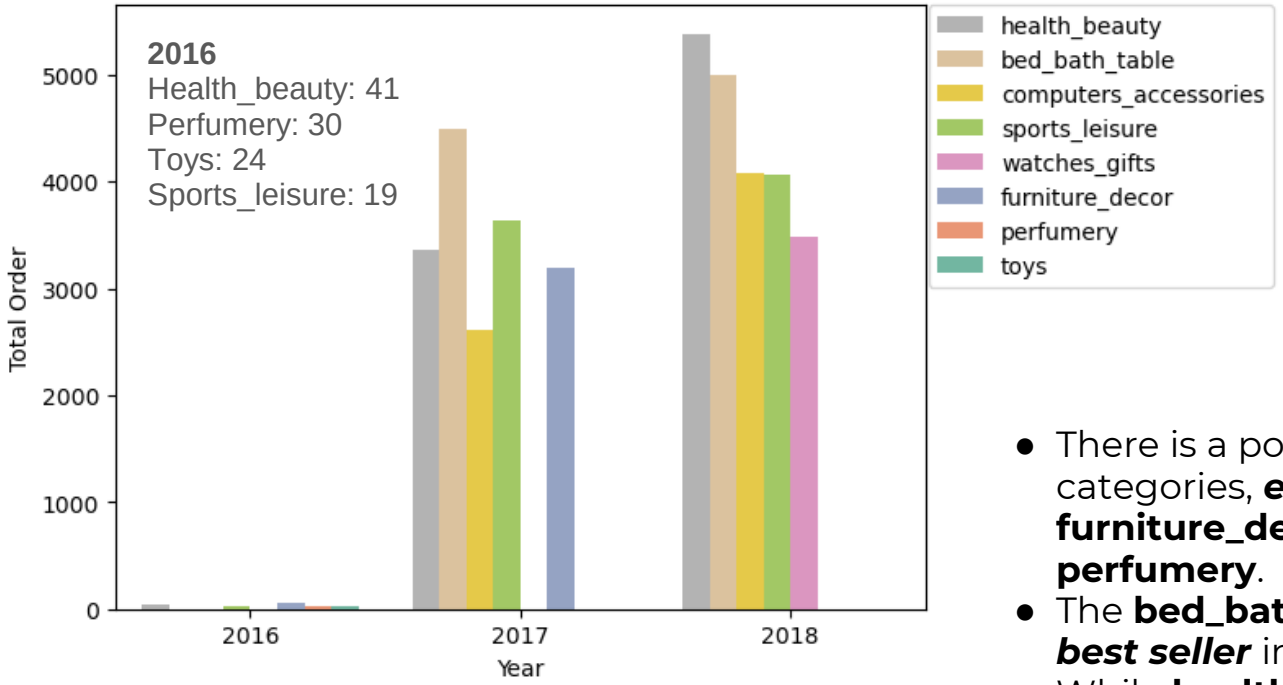
There are **50% orders** of **bed_bath_table** is from state SP

In state non-SP, there are more than **50% orders** forf **bed_bath_table** and **health_beauty**.

Total Order by Product in State non-SP

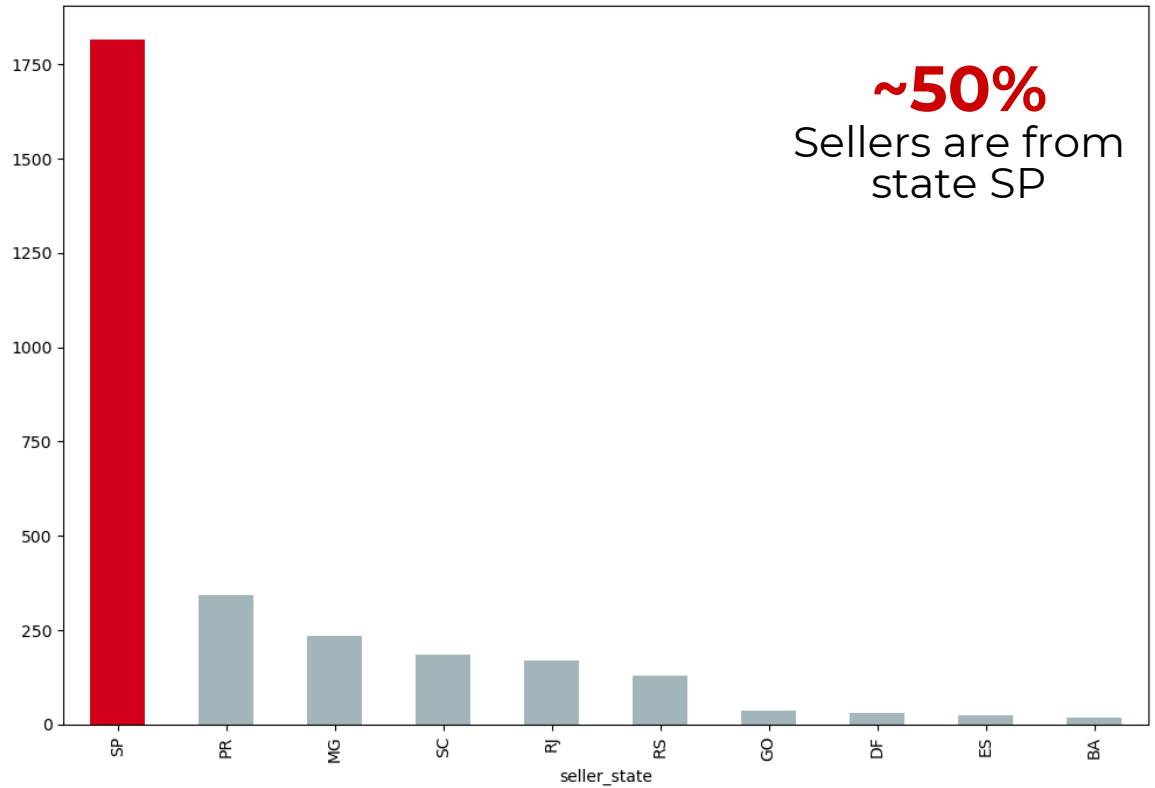


Product Trend by Year

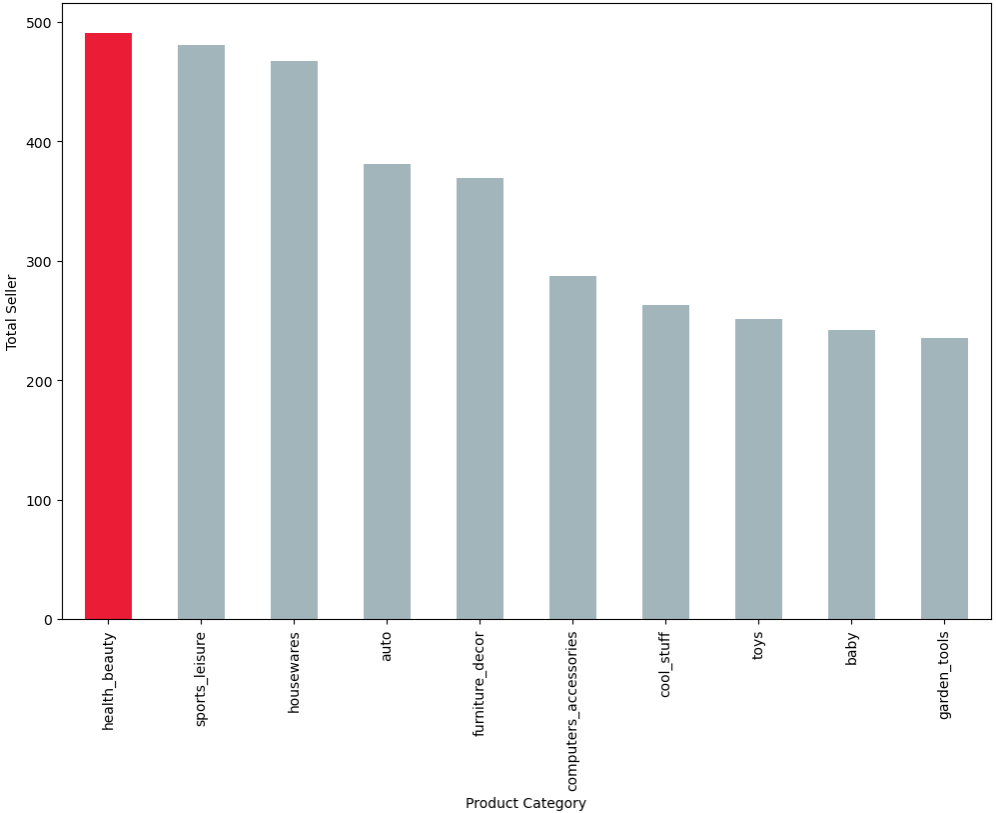


- There is a positive trend from all categories, ***except*** for **furniture_decor, toys,** and **perfumery**.
- The **bed_bath_table** started to be a ***best seller*** in 2017.
- While **health_beauty** is a ***best-seller*** in 2016 & 2018.

Characteristic of Seller



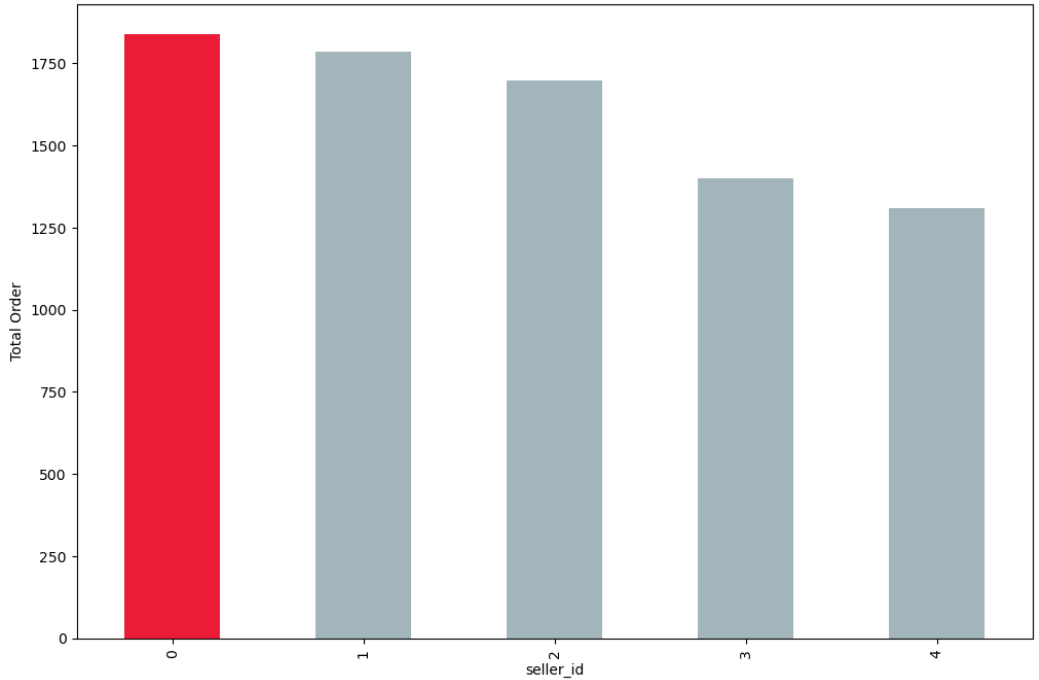
Total Seller by Product Category



There are **15.85% sellers** that are selling **health_beauty product**.

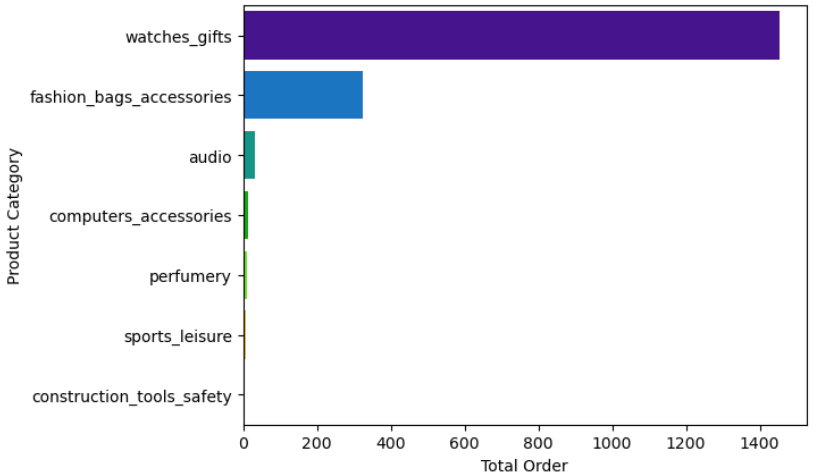
While bed_bath_table has the highest total order, *but* the sellers are not on top 10.

Total Order by Seller



The top-seller has **finished 1,838 orders** with average of **9 days delivery time**.

The **best-selling product** is watches_gifts.



Customer Segmentation

Cluster Delivered

- Customers with all order delivered.
- Average delivery time is 139 days.

Cluster Canceled

- Spent more.
- Customers with cancelled order.
- Average of delivery time 46 days.

Customer Segmentation in State SP

The total order is the same for both cluster, but **2.06% customers** are more likely to have a cancelled order and give bad ratings.

For customer in state SP, the concern is more to the seller/buyer, due to the delivery time is smaller (which means the order got cancelled before it's delivered)..

Customer Segmentation in State non-SP

Cluster Canceled

- Customers with cancelled order.
- The delivery time reach 1800 days on average.

Cluster Delivered

- Customers with all order delivered.
- Average of delivery time 300 days.

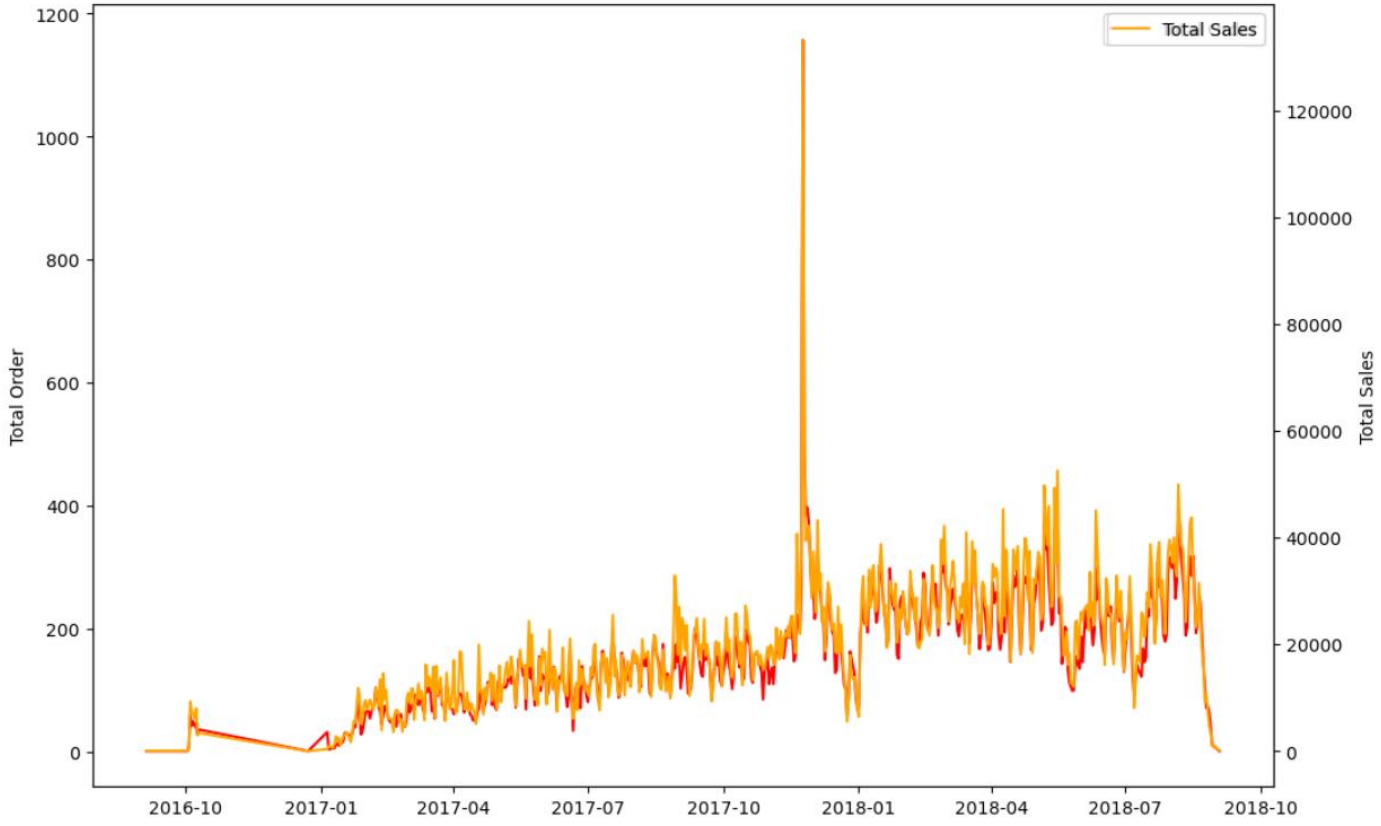
The total order is the same for both clusters, but **2.8% customers** are more likely to cancel their order and give bad ratings.

For customer in state non-SP, the concern is the delivery time.

Customer Segmentation Insight

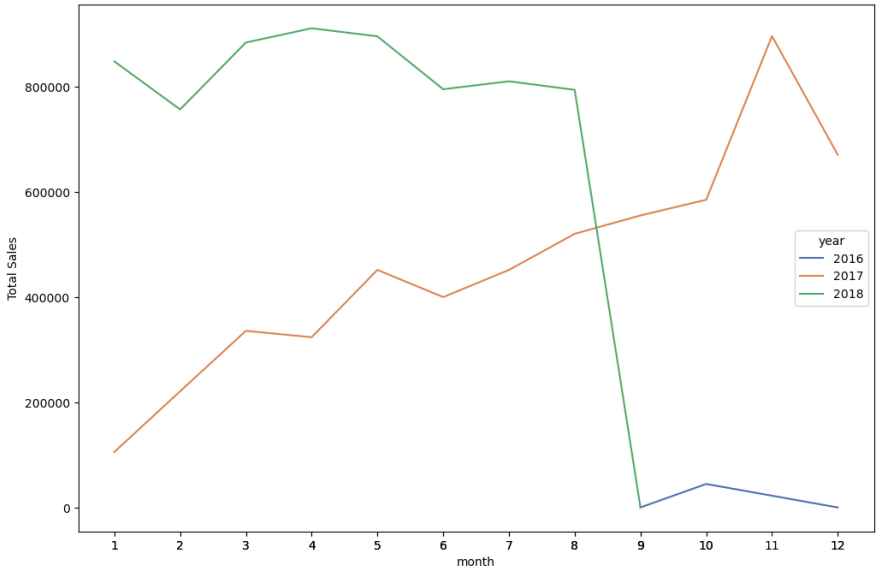
Based on the chart, the total number of sellers and customers at Spare the highest number. It means that the transactions mostly happened inside SP which makes the delivery time is lower than the transaction outside SP.

Order & Sales

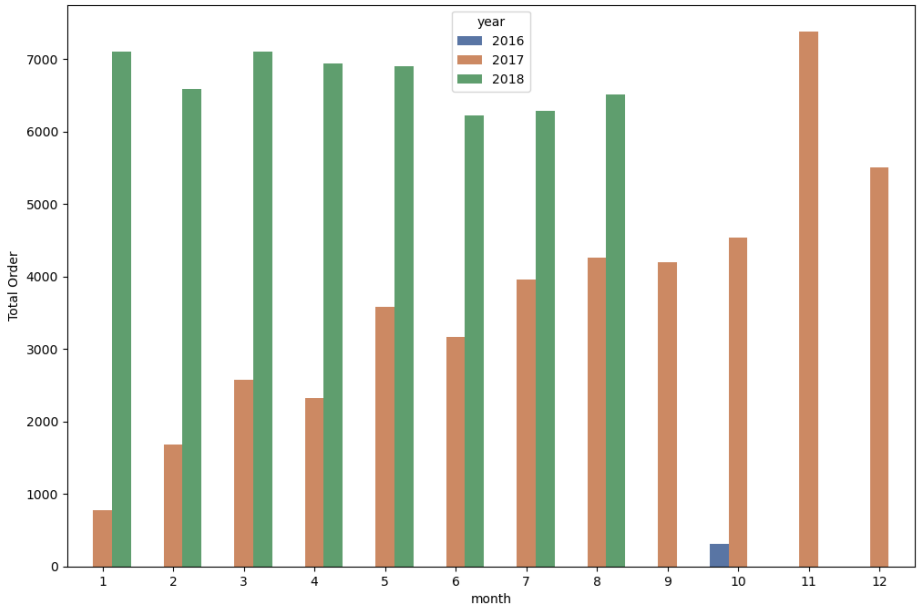


Total Order & Sales daily chart

Total Sales by Month

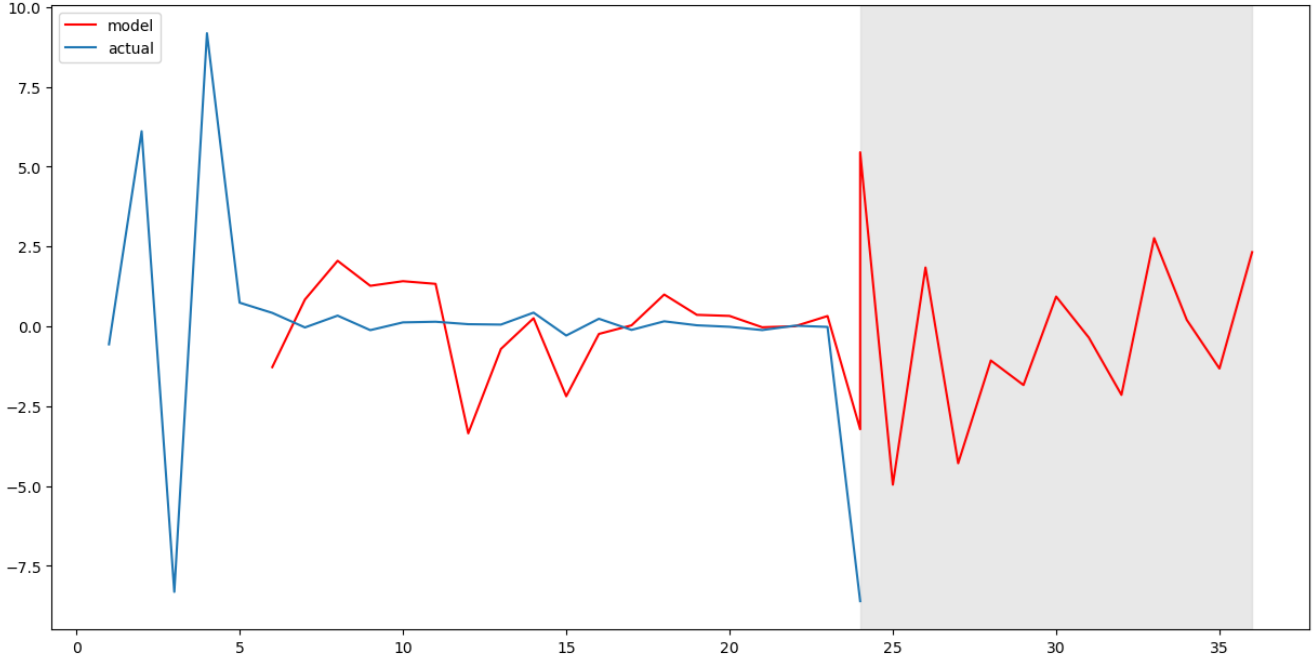


Total Order by Month



The ‘sales’ is achieved through calculating price of goods times the number it is purchased by finished order, and lastly grouped by date.

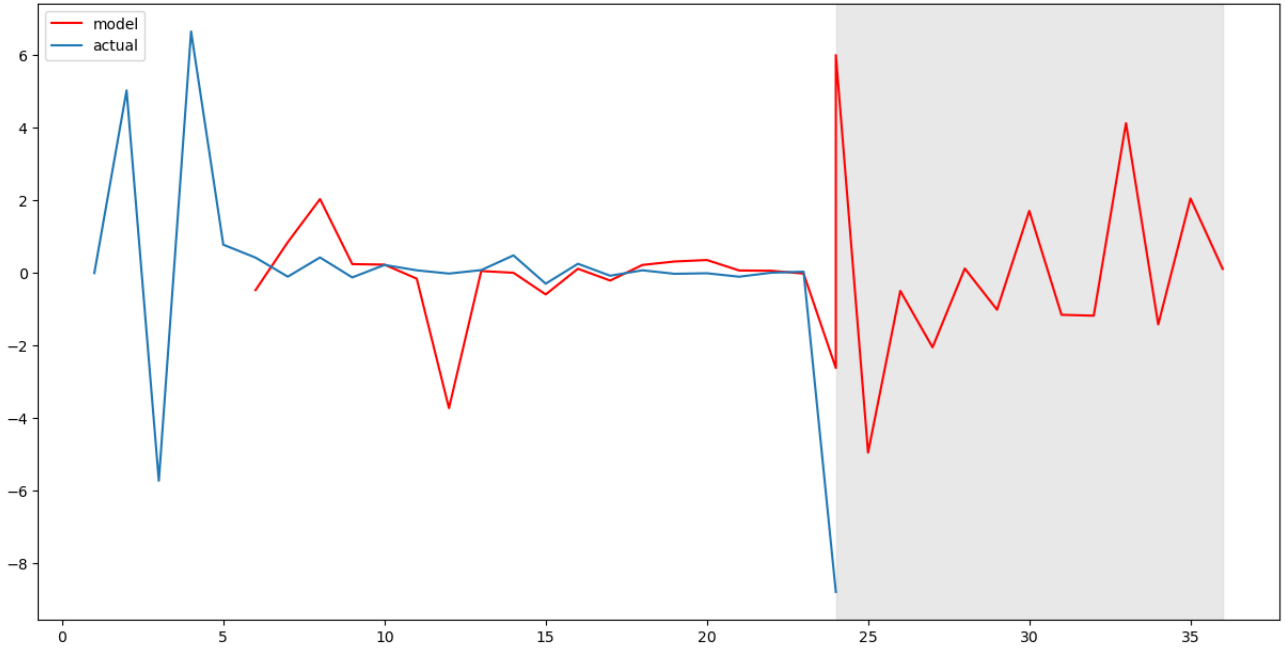
Monthly Sales Prediction



Using Auto-ARIMA

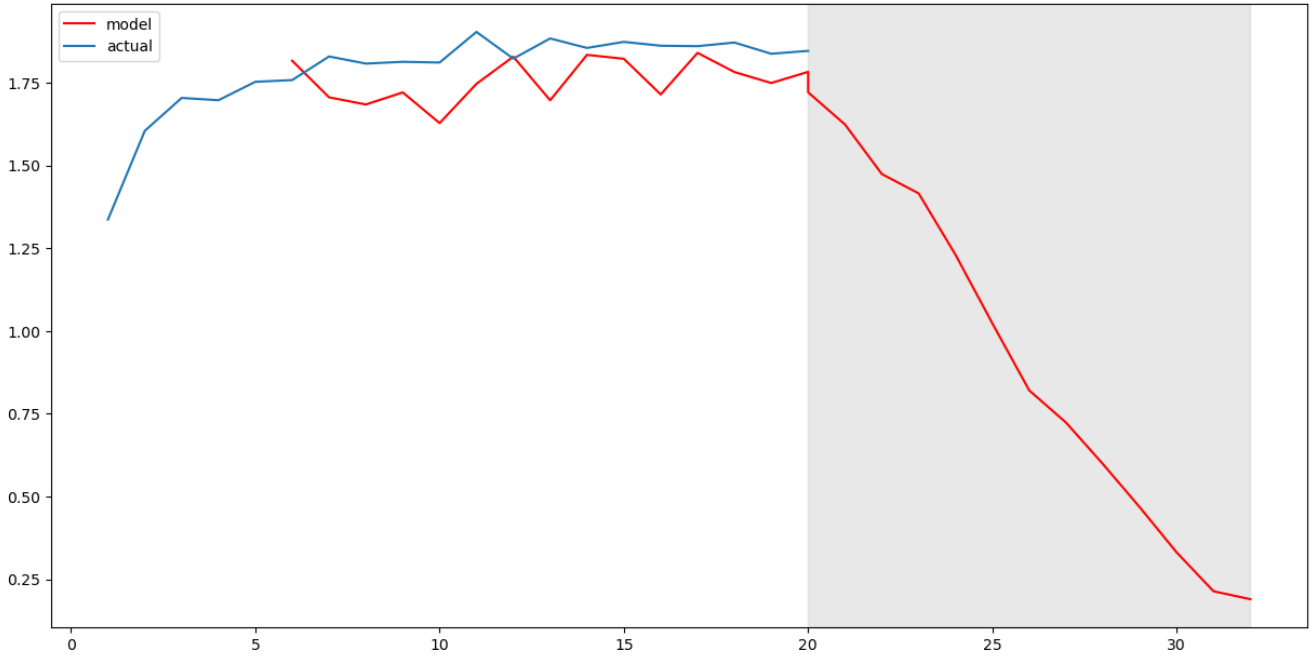
There's around **0.89 points difference** between prediction and the actual forecast value of monthly sales.

Monthly Order Prediction



There's around **0.69 points difference** between prediction and the actual forecast value of monthly sales.

Prediction of Monthly Order for Bed_Bath_Table



There is a **difference around 4.9% (0.09 points)** between the prediction and the actual forecasted values.

The order is most likely to be decreased to almost 0 in 12 months.

Business Recommendations

- Olist could assist seller in giving a shorter delivery time, e.g: building drop-point outside state SP and also near seller.
- Regarding the cancelled order, Olist could have reward system for those who finished order (both sellers & customers).
- Given that Olist has 72 product categories, but the top-4 best sellers have not changed in 2 years. Olist could focus on product in top-4 and specialized in health, beauty, and hobbies.
- Given that the order for bed_bath_table will be decreased in a year, Olist could cooperate with sellers to give promotion.

Technical Recommendation

Even though the MAE score is tolerable, the timeseries models for total order & sales both have **>50% MAPE** and a **minus R2-score** which means the models have a really-really bad performance.

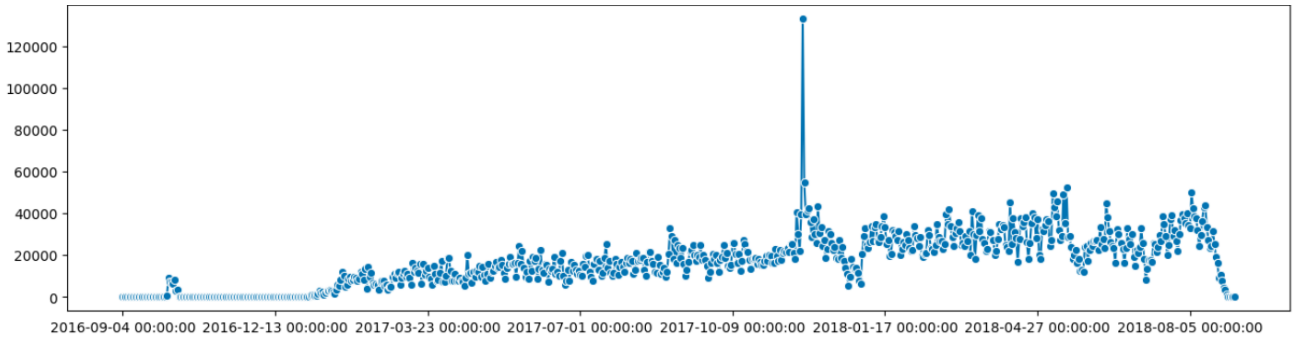
It is better to use other models or trained with different timeframe (weekly, daily) or add more data.

Daily Sales Data

- There is 615 non-null data, but there are some dates with no sales data. Nonetheless, to use timeseries model we need all data (data in all dates should be filled).
- To create the data from missing dates, ‘resample’ is used (based on ‘d’) and the data is filled with 0.
- In result, there is 730 data for daily sales.

	date	price
0	2016-09-04	39.99
1	2016-09-05	59.50
2	2016-10-02	100.00
3	2016-10-03	463.48
4	2016-10-04	9252.87

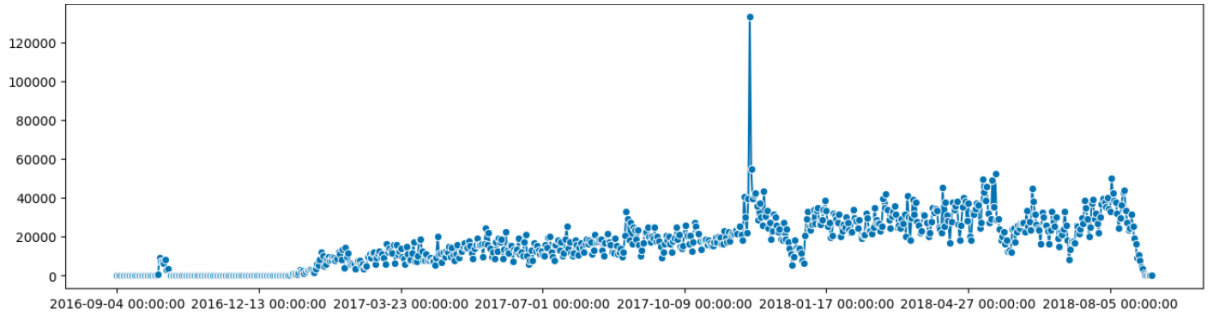
There is no data from ‘2016-09-06’ to ‘2016-10-01’.



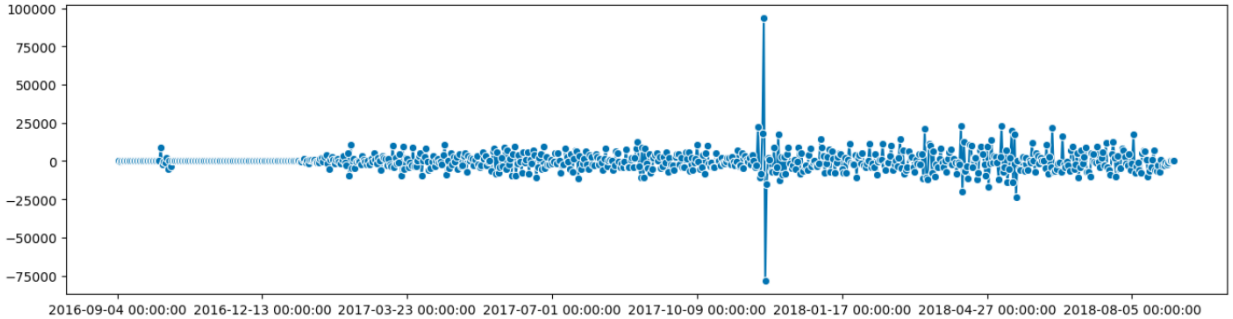
Data Preparation

With the p-value < 5% for KPSS test, the null-hypothesis is rejected, the data is considered non-stationary. So the diff transformation is applied to the dataset for timeseries modelling.

The ‘Daily Sales’ plot

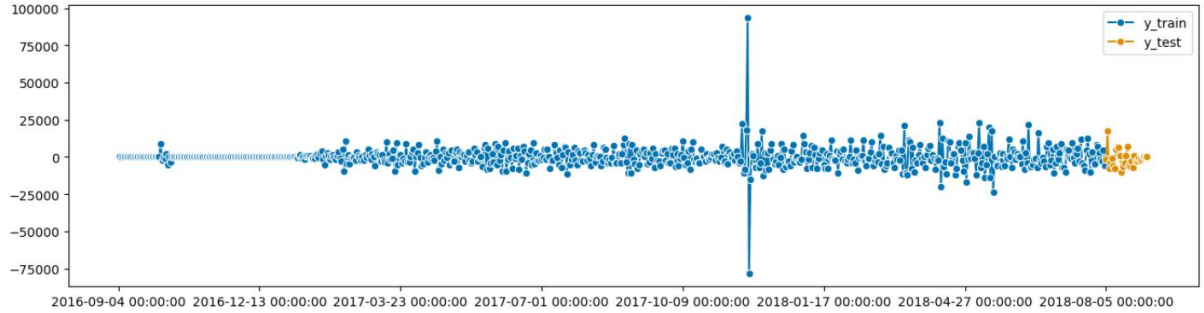


The ‘Daily Sales’ plot after Diff Transformation

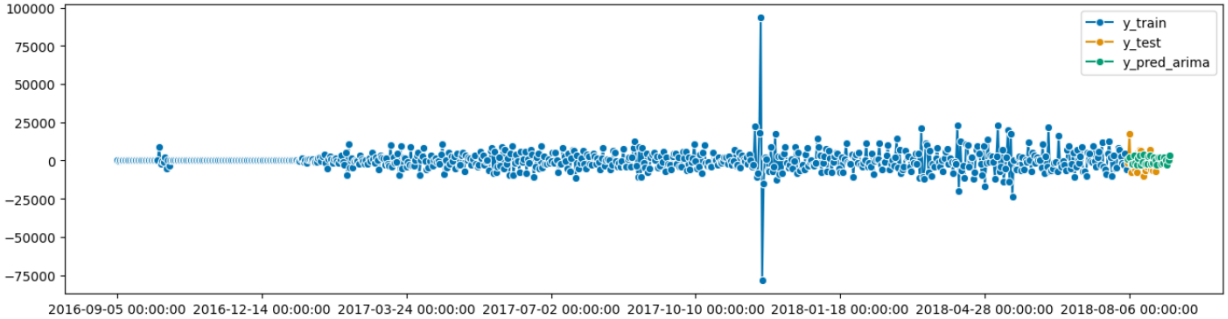


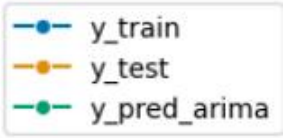
Auto-ARIMA

- The data split into test (30) & train.
- Seasonality period: 30 days (1 month).



The result:



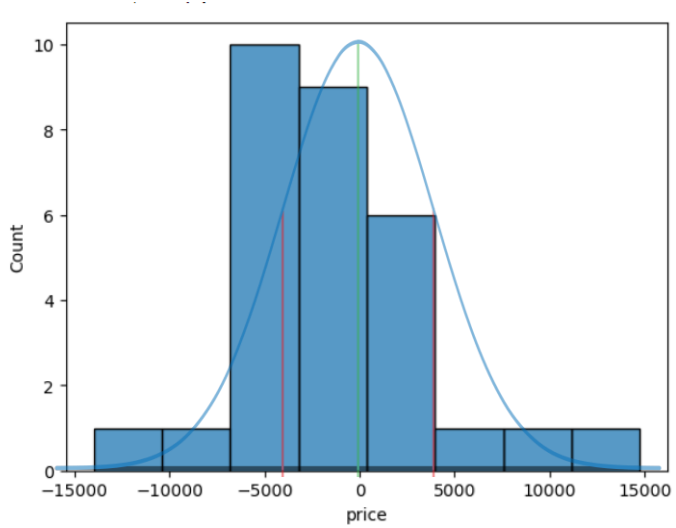


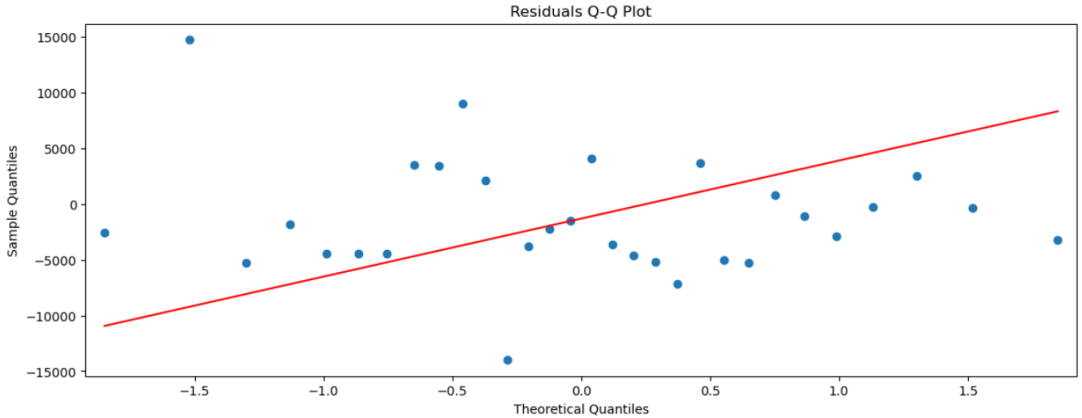
The prediction model seen in the chart is not well-fitted to the actual data. But how is the model performance?

Model Diagnostics

Error occurred in the model: 4.6464733387939 25e+19 (so-so-so bad)

Which means the model can not be used for forecasting.





The residuals are still skewed to the left, even though it is almost normally distributed.

The residuals also spread-out and not aligned with the regression line.

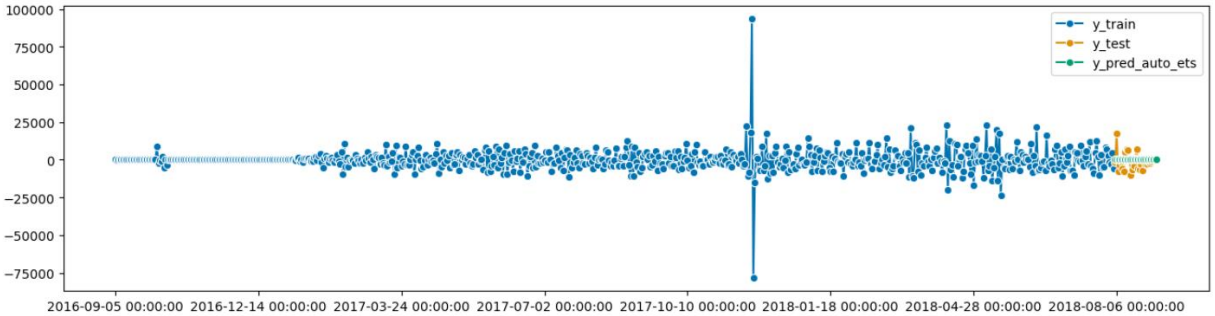
Exponential Smoothing (ETS) Holt-Winters

ETS Holt-Winters cannot be done because the data contains negative values.

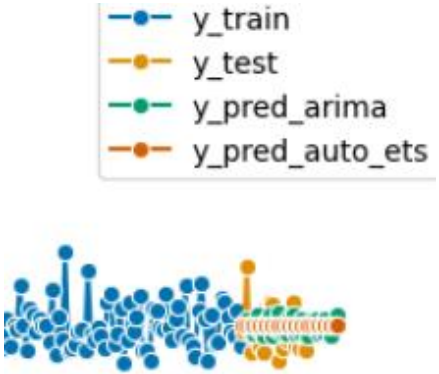
Auto-ETS

- The data split into test (30) & train.
- Seasonality period: 30 days (1 month).

The result:



Comparison with Auto-ARIMA:

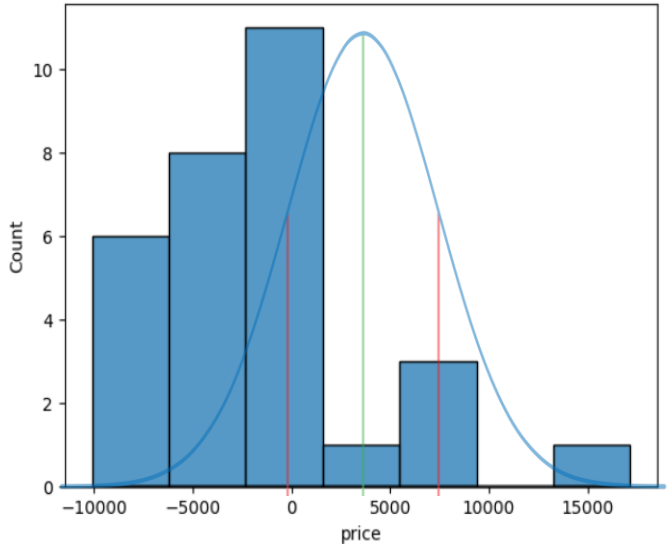


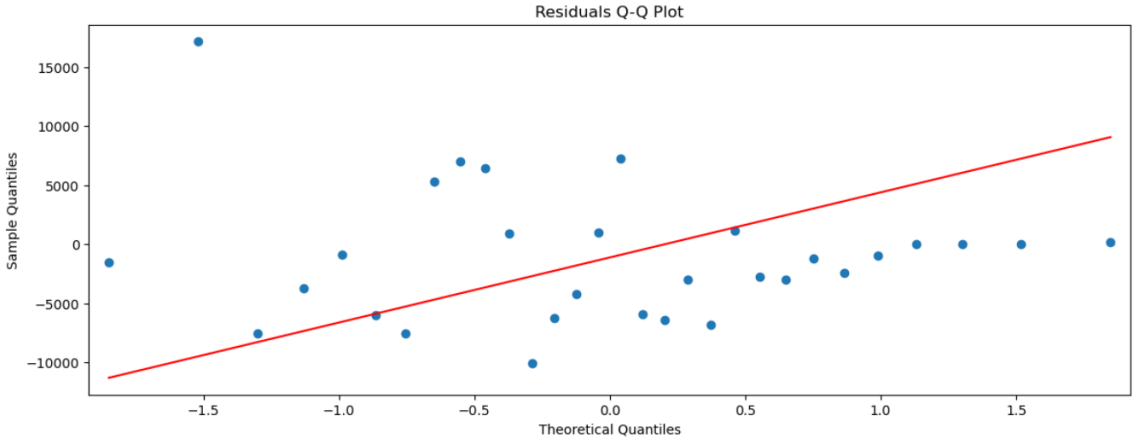
The prediction model with Auto-ETS is not well-fitted to the actual data. Even it has smaller range than the Auto-ARIMA model.

Model Diagnostics

Error occurred in the model:
3.5599435047877l5e+17 (> 100%, so-so-so bad)

Which means the model can not be used for forecasting.





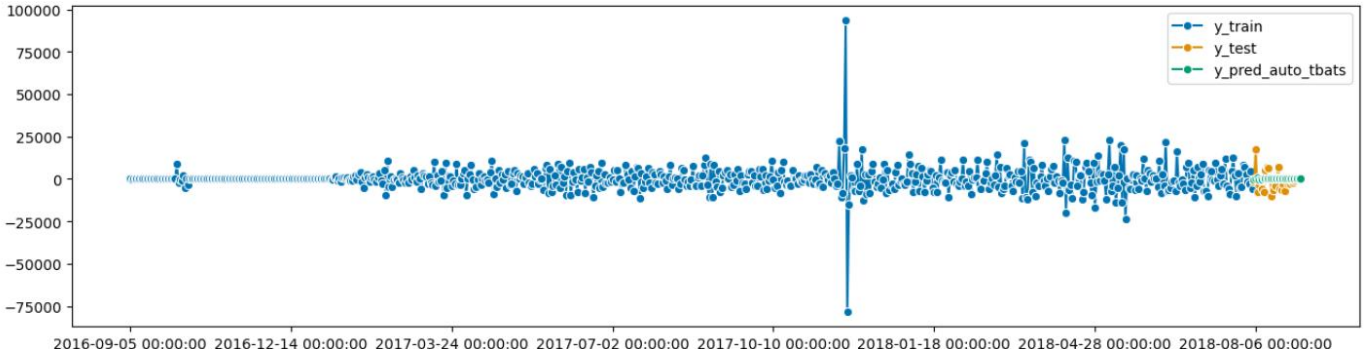
The residuals are obviously right-skewed.

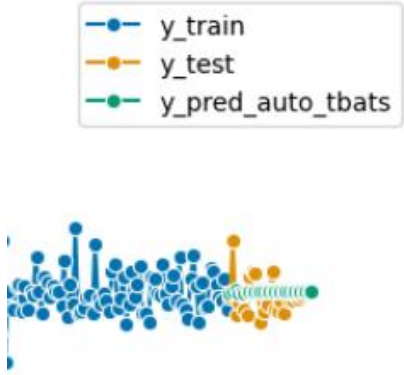
The residuals also spread-out and not aligned with the regression line.

TBATS

- The data split into test (30) & train.
- Seasonality period: 30 days (1 month).

The result:



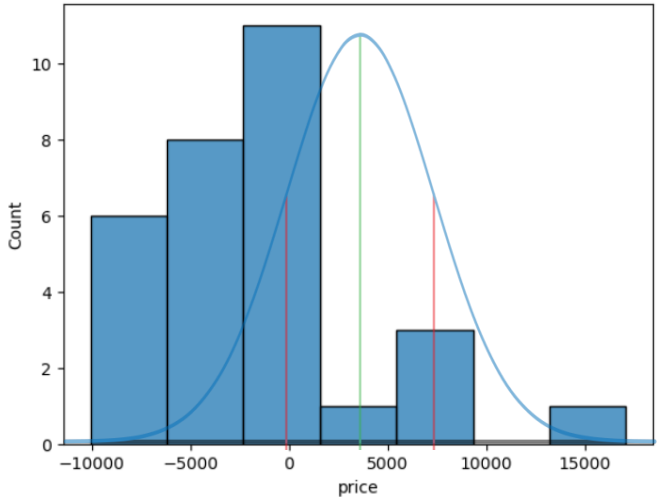


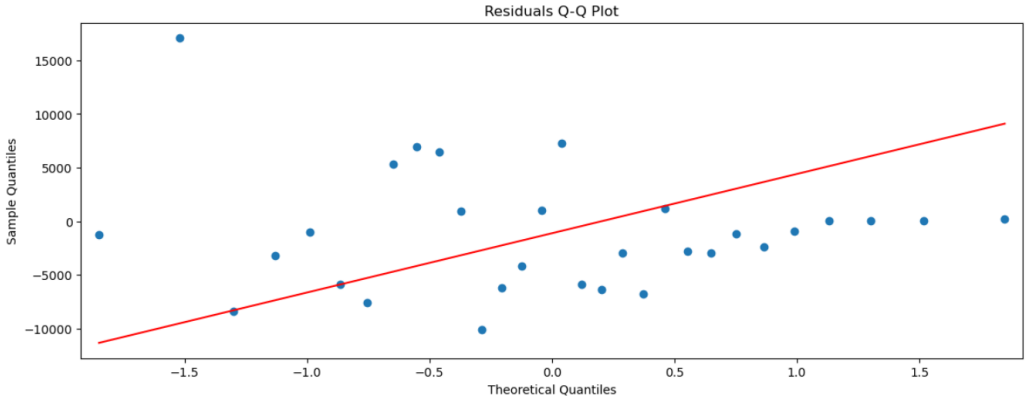
The prediction model seen in the chart is not fitted to the actual data. It is smaller than the first two models.

Model Diagnostics

Error occurred in the model:
1.2435406599272229e+18 (> 100%, so-so-so bad)

Which means the model can not be used for forecasting.





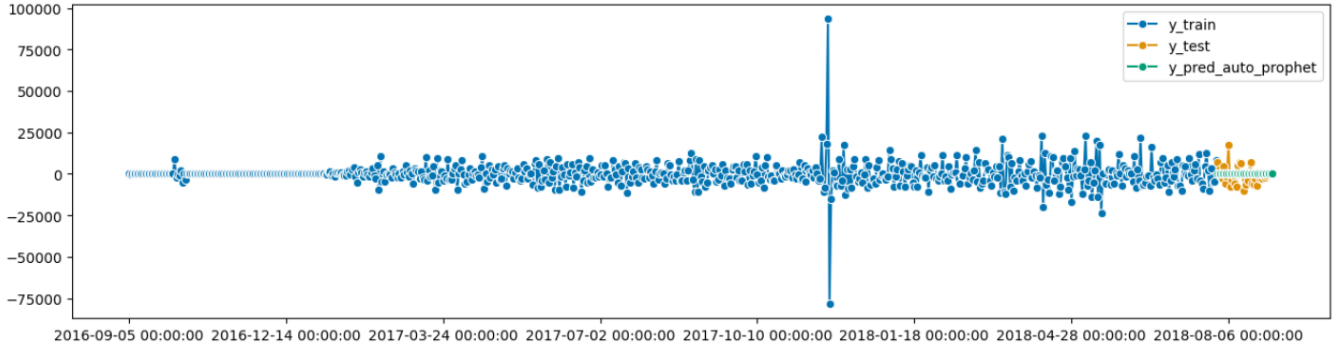
The residuals are obviously right-skewed.

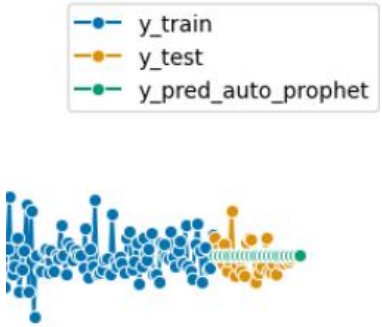
The residuals also spread-out and not aligned with the regression line.

Prophet

- The data split into test (30) & train.
- Seasonality period: 30 days (1 month).

The result:



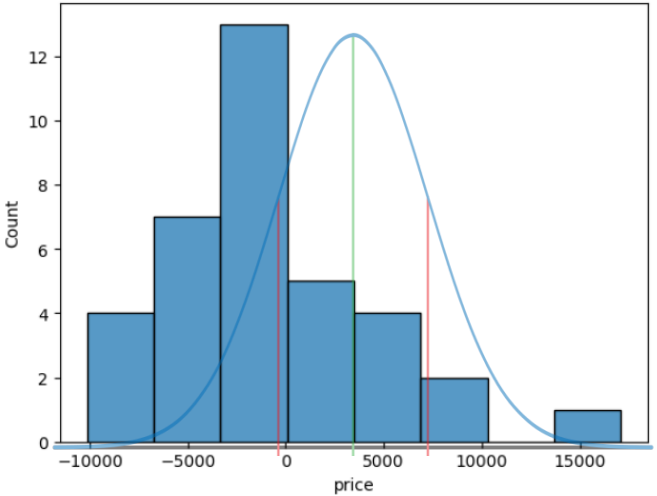


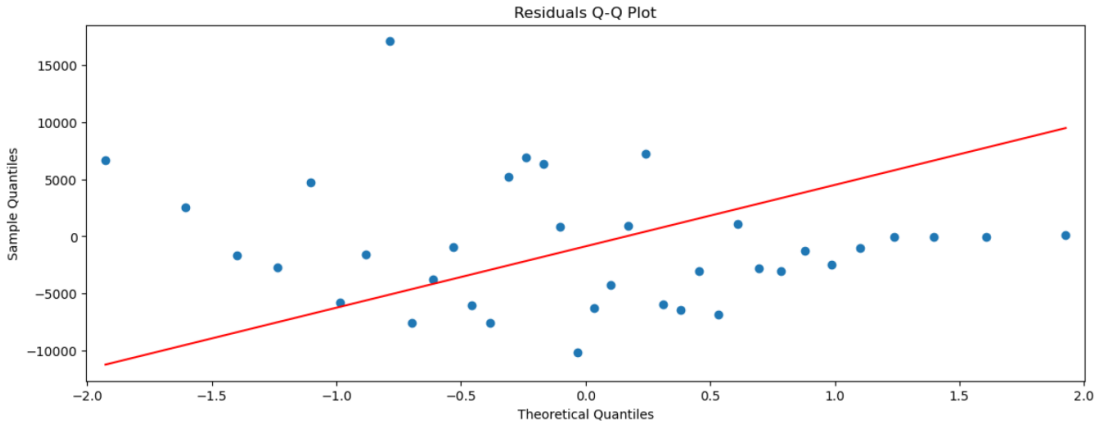
The prediction model seen in the chart is not fitted to the actual data. It has the smallest among all models.

Model Diagnostics

Error occurred in the model:
1.4688978608267612e+18 (> 100%, so-so-so bad)

Which means the model can not be used for forecasting.





The residuals are obviously right-skewed.

The residuals also spread-out and not aligned with the regression line.

Further Works

The models need to be re-evaluate.