



Optimal Sales Forecast Model Assessment - eCommerce Startup

By : Kulwinder Kaur

Statistics about sales forecast

- 80% of sales orgs DO NOT have a forecast accuracy of greater than 75%.
 - *As per CSO Insights and Gartner*
- 97% of companies that implemented best-in-class forecasting processes achieved quotas, compared to 55% that did not.
 - *As per Aberdeen research*



“My Goal is to identify the preferred sales forecasting model suitable for eCommerce startup with limited historical data.”



Data collection

Public dataset published by
Olist on Kaggle.

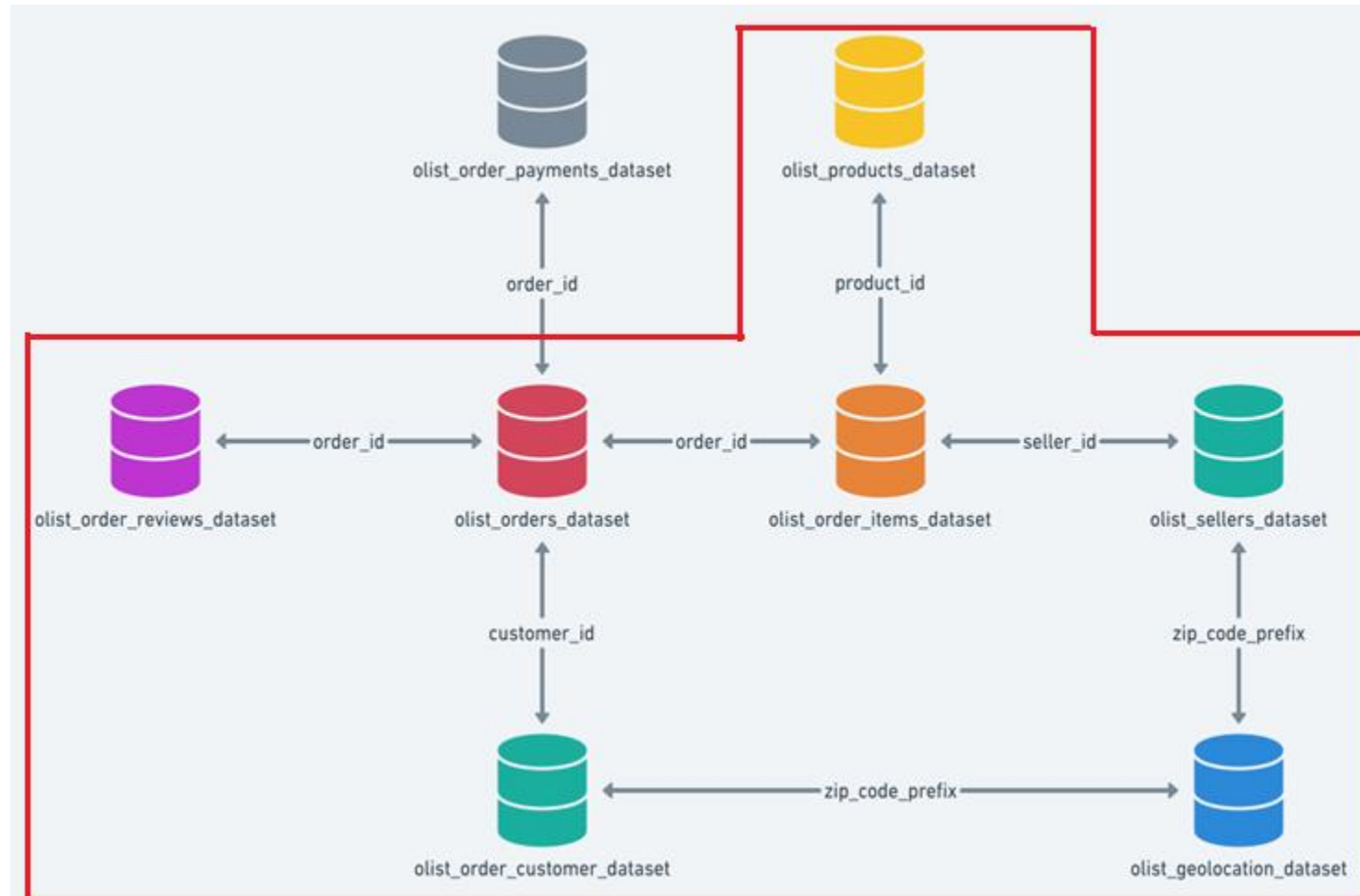
Brazilian Ecommerce platform
Founded in 2015.
Data from 2016 - 2018.

Olist has primarily shared Order
transaction data, product data
and seller data.



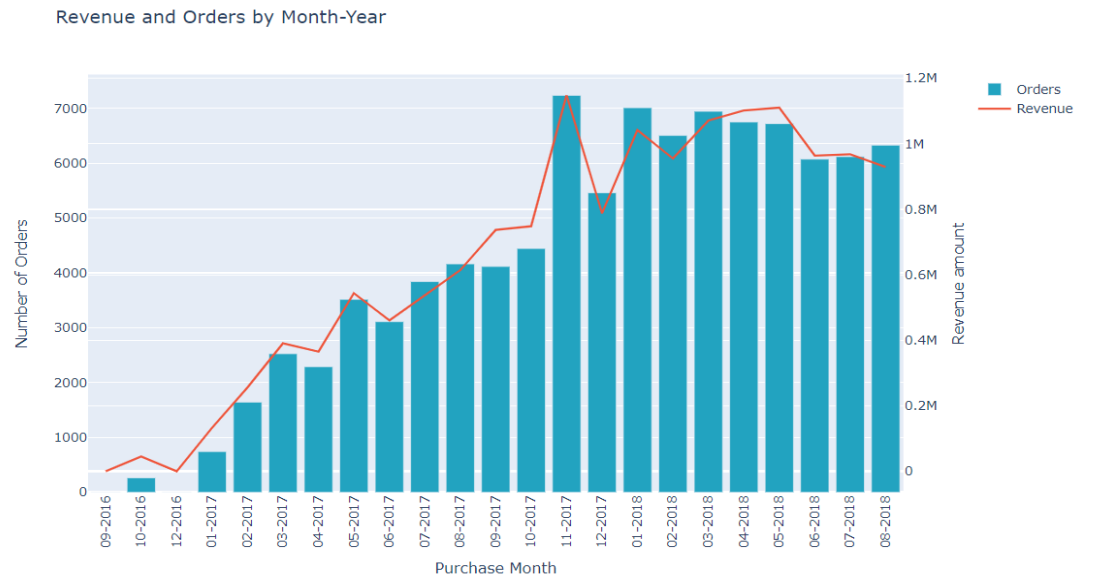
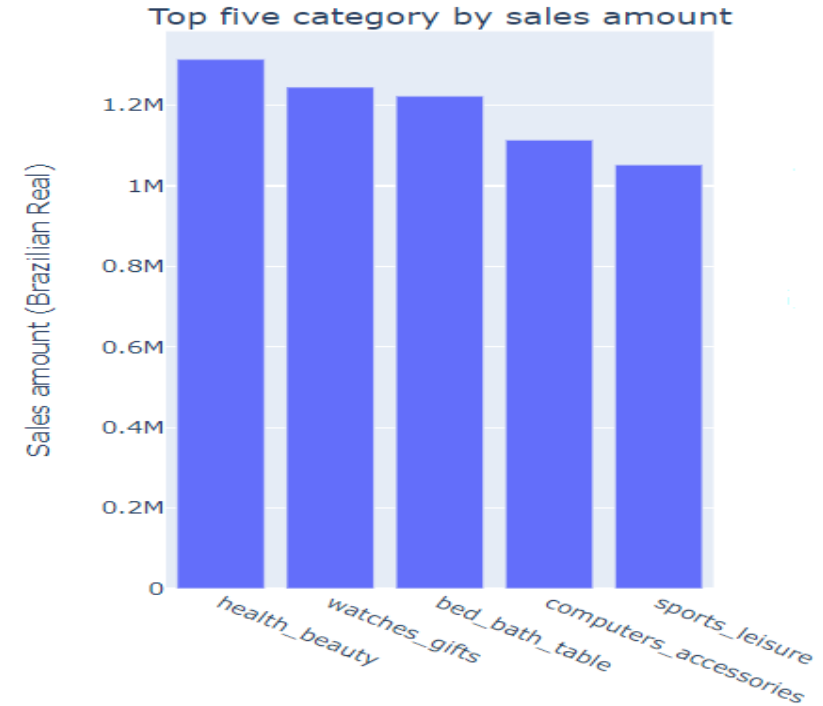
Data Schema

- Joined these csv files to make a master table with 110013 observations with 29 features.
- Removed duplicates and imputed null values.
- After cleaning final data set has all the orders with delivered status.
- Added extra feature 'holiday' after scarping it from Brazilian national holiday website.



Data Insights

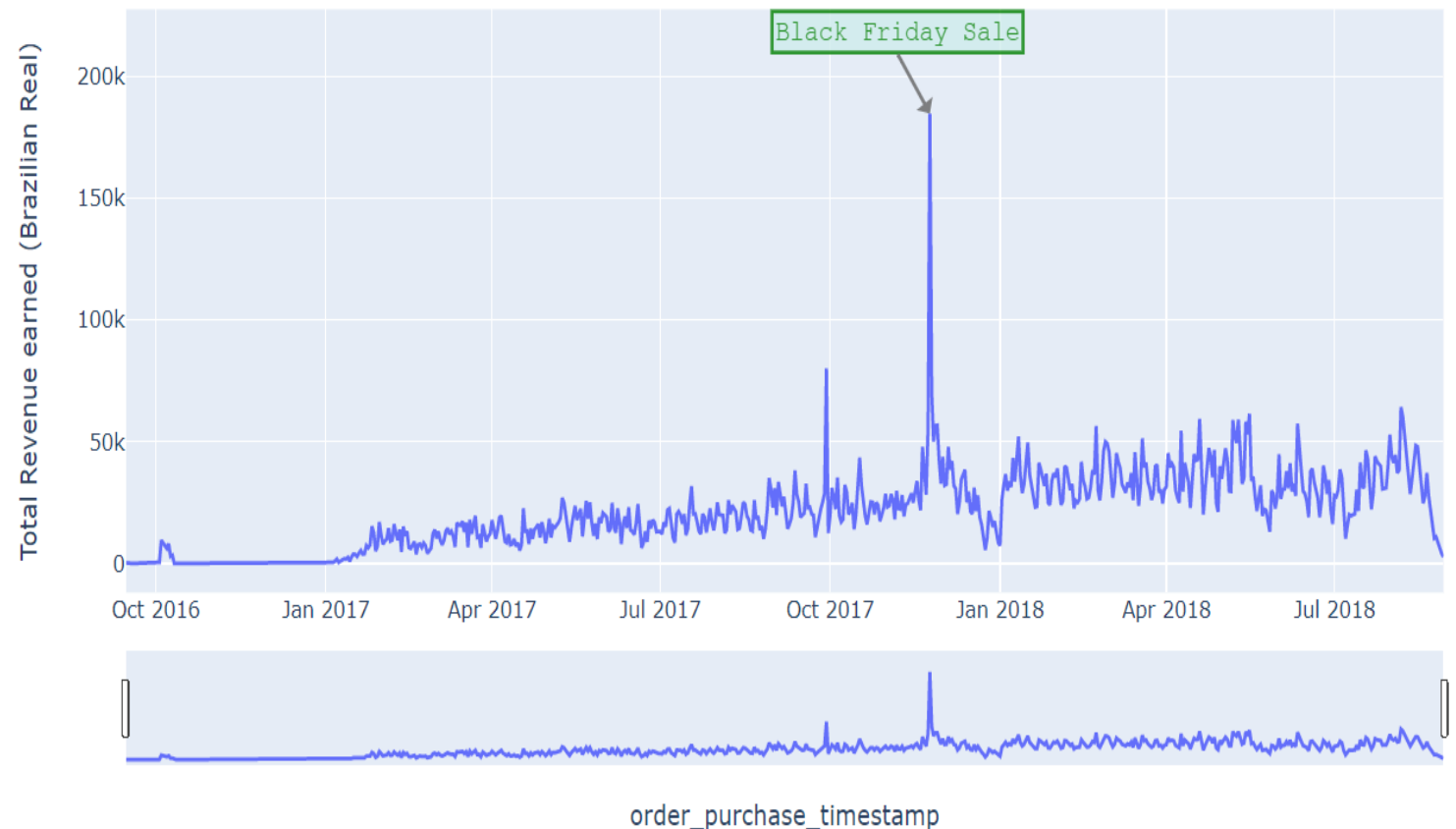
- A total of 96K unique orders.
- Olist platform has 96.79 % as new customers and 3.21% have made repeat purchase..
- A total of 32K different products belonging to 74 categories are sold.
- The overall revenue earned as of Aug 2018 is 14.9 million Brazilian Real (R\$).
- There was a highest sale of 184K R\$ that was recorded on Black Friday 2017.
- The top five categories are shown here in the bar chart.
- The monthly orders and revenue earned show a growth.



Time Series Data Pre-processing

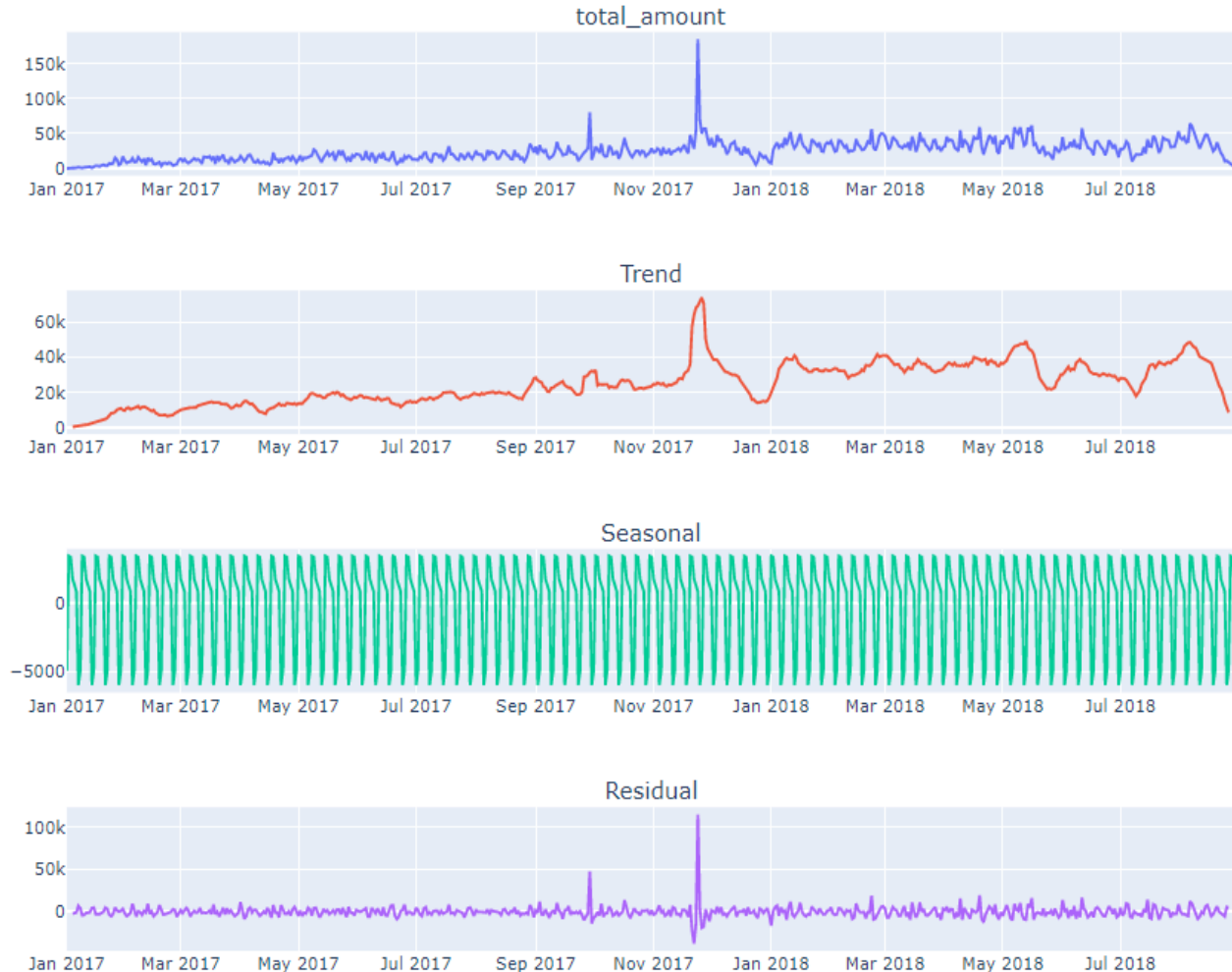
- Aggregated the Sales amount by date and it reduced the datapoints to two years of data with 606 observations.
- **Target variable: Sales**
- Note: Freight charges were not included as Olist outsourced logistics to third party till 2020.

Daily Revenue from Sept 2016 to Aug 2018



Time Series Analysis

- Observed weekly seasonality.
- Positive Trend.
- There was almost 99% increase in revenue from 2016 to 2017 and 17.4% increase in 2018.
- Huge impact of holidays like Black Friday sale.



Models Applied

Applied four different approaches:

- SARIMAX model including holiday impact.
- Tuned Facebook Prophet including holiday impact.
- Tuned XG boost including holiday impact.
- LSTM one step ahead.
- Using grid search did hyper parameter tuning for each model.

Daily Sales amount and forecast using different models



Results [demo](#)

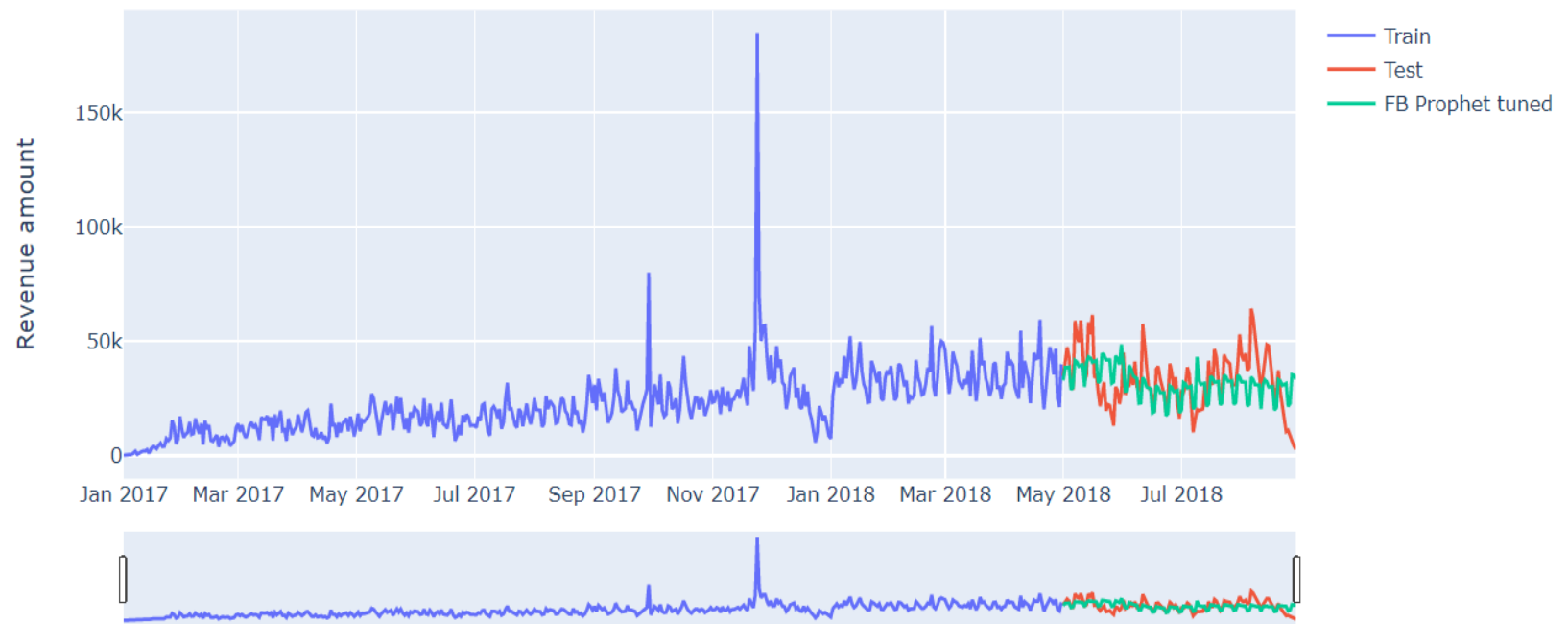
Model Evaluation

- LSTM one step ahead univariate neural network has least MAPE and was trained on limited data. It being a neural network should be fed more data and could be tested further to see actual performance.
- XG Boost next best performance, but it is not able to pick up trend and variations between weeks.
- Choosing tuned FB prophet as final model.

Model	RMSE	MAPE
SARIMA(1,1,1)(0,1,1)(7)	13810.59	68.99
SARIMAX(1,1,2)(0,1,1)(7) Including impact of holidays	13312.72	65.66
Baseline Prophet	27437.77	71.78
Baseline Prophet with holiday	15393.64	77.88
Tuned Prophet with holiday	12142.69	51.45
XGBoost Regression including Holiday	11493.59	52.18
Tuned XGBoost Regression including Holiday	12349.61	47.58
LSTM (one step Prediction)	2803.14	9.37

Facebook Prophet Results

Daily Sales amount and forecast using Tuned FB Prophet



- Going Ahead with FB Prophet.
- Able to pick seasonality, trend and variation within week.
- Gives a MAPE of 51.45% considerable performance.

Parameters of tuned Facebook prophet model

Growth	Changepoint_prior_scale	Holidays_prior_scale	Seasonality_mode	Seasonality	Seasonality_prior_scale
Linear	0.4	0.5	Multiplicative	True for yearly, weekly and daily seasonality	20

Summary

Achieved so far..



Analysed the patterns in time series.
Applied and tested machine learning
and neural network forecast models.

Built a baseline forecast model to
give business an idea about the
future.

Future Scope



External factors can be
added like inflation rate,
customer satisfaction
etc to further improve
this model.



Model can be applied to
forecast sales for each
product category.

Tools used

