



Sales Forecast for an
ecommerce start-up to help
business understand future
demands.

Statistics about sales forecast

- 93% of sales leaders are unable to forecast revenue within 5 percent, even with two weeks left in the quarter.
- 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 aim is to

- analyse current business performance of E-Commerce startup,
- try different sales forecast models, compare them and find out the best model suitable for limited historical data.

Data collection

This is a public dataset published by Olist and is available at Kaggle

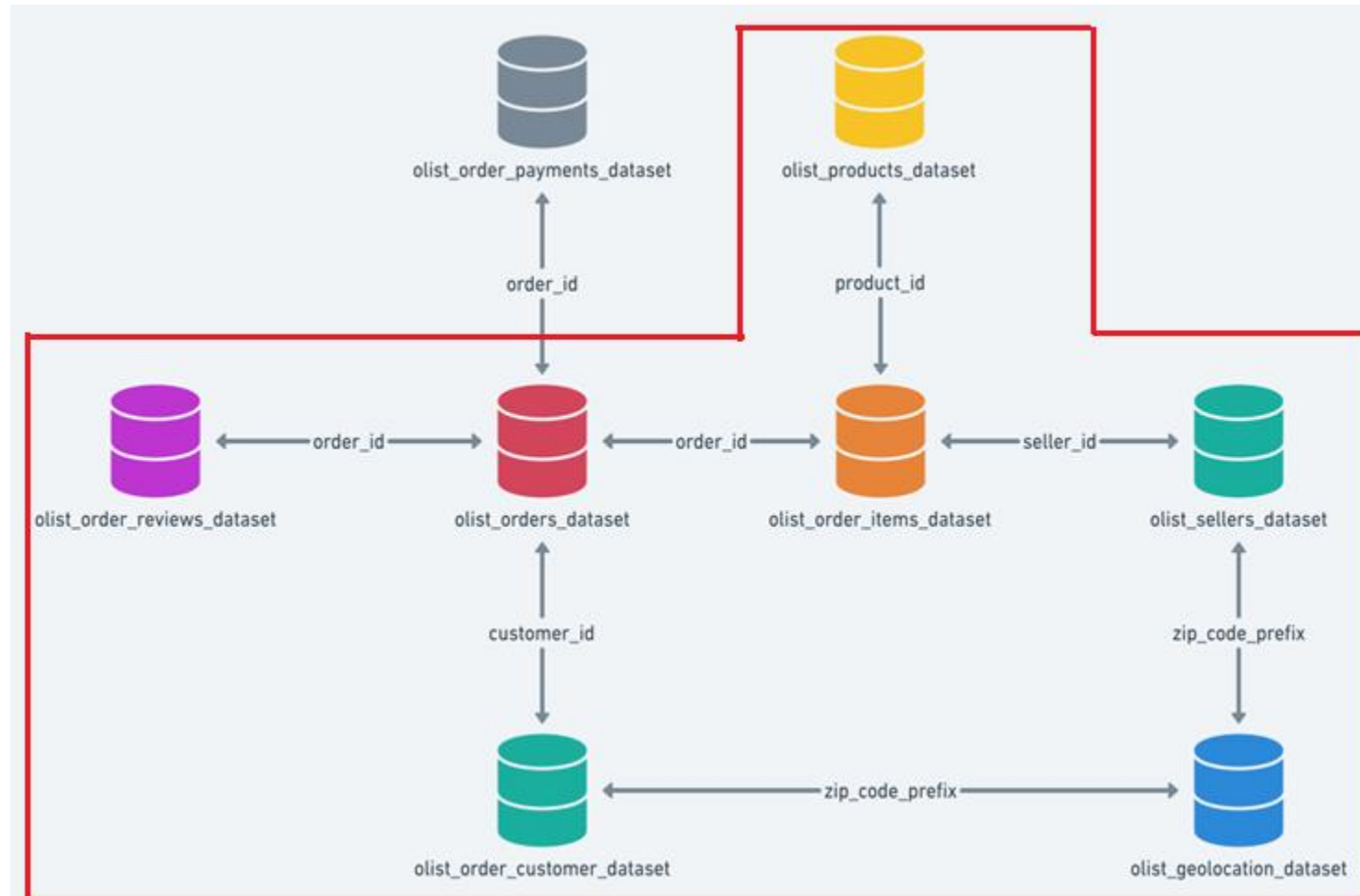
Data points of 100k orders from 2016 to 2018 made at multiple marketplaces in Brazil.

Scraped Brazilian National Holiday information from OfficeHolidays.com



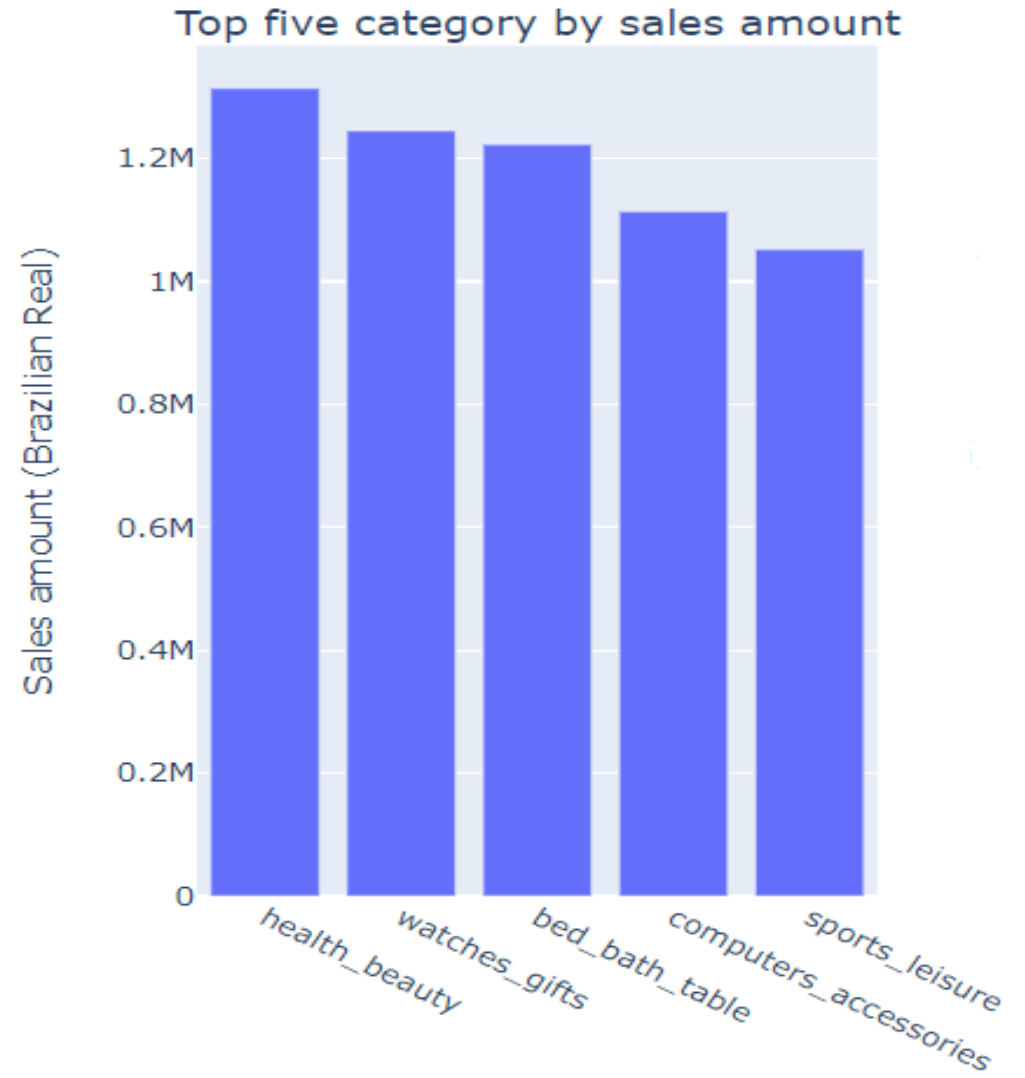
Data Schema

- Joined seven tables to make a master table with 110013 orders with 29 features.
- Removed duplicates and imputed null values.
- All the orders are delivered.



Some Insights

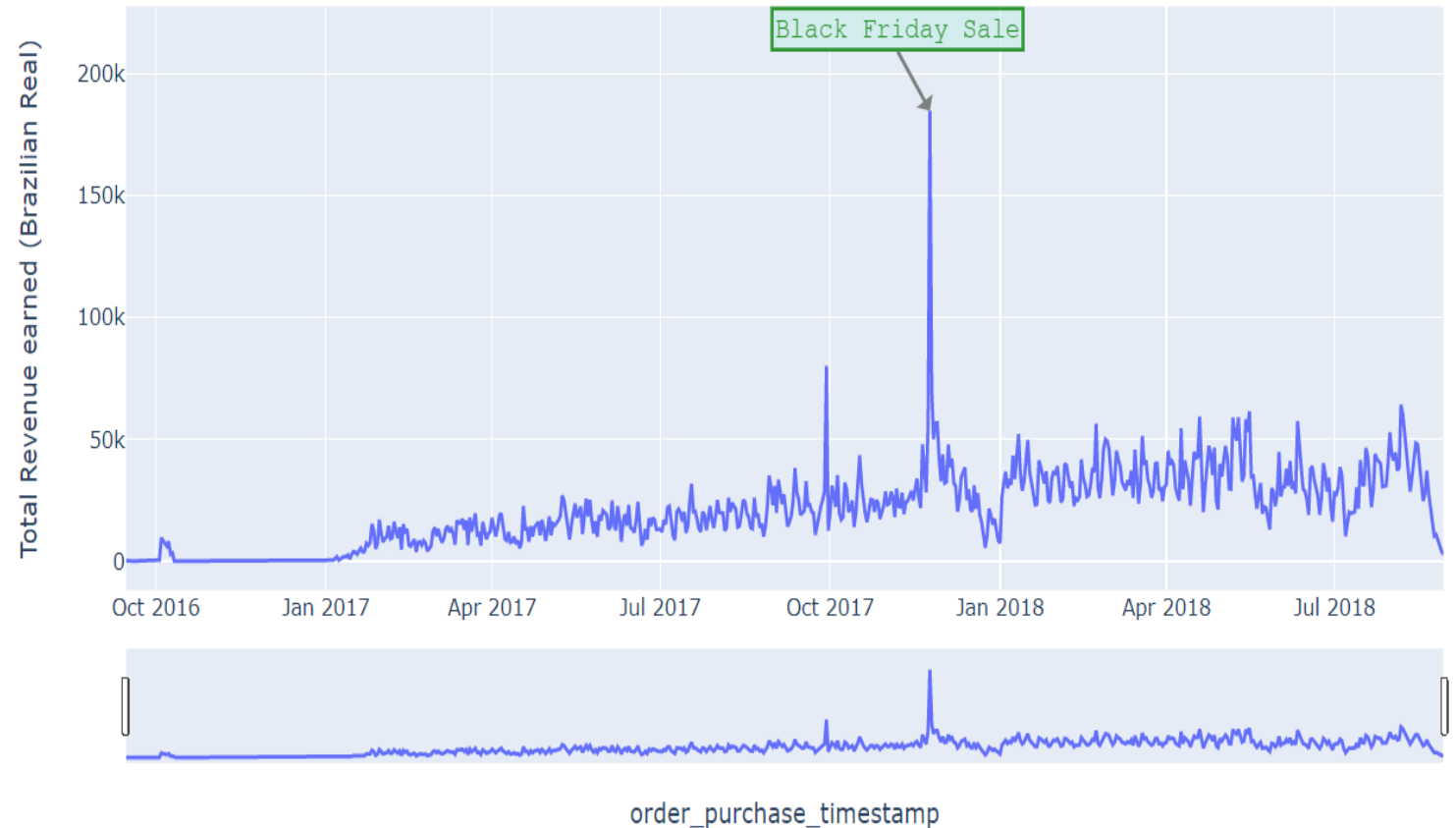
- A total of 96K unique orders.
- 96.79 % are new customers with 3.21% repeat customers.
- A total of 32K different products belonging to 74 categories are sold.
- The overall revenue earned as of Aug 2018 is 14.9 million Brazilian Real (R\$).
- There was a highest sale of 184K R\$ that was recorded on Black Friday event.



Time Series Modelling

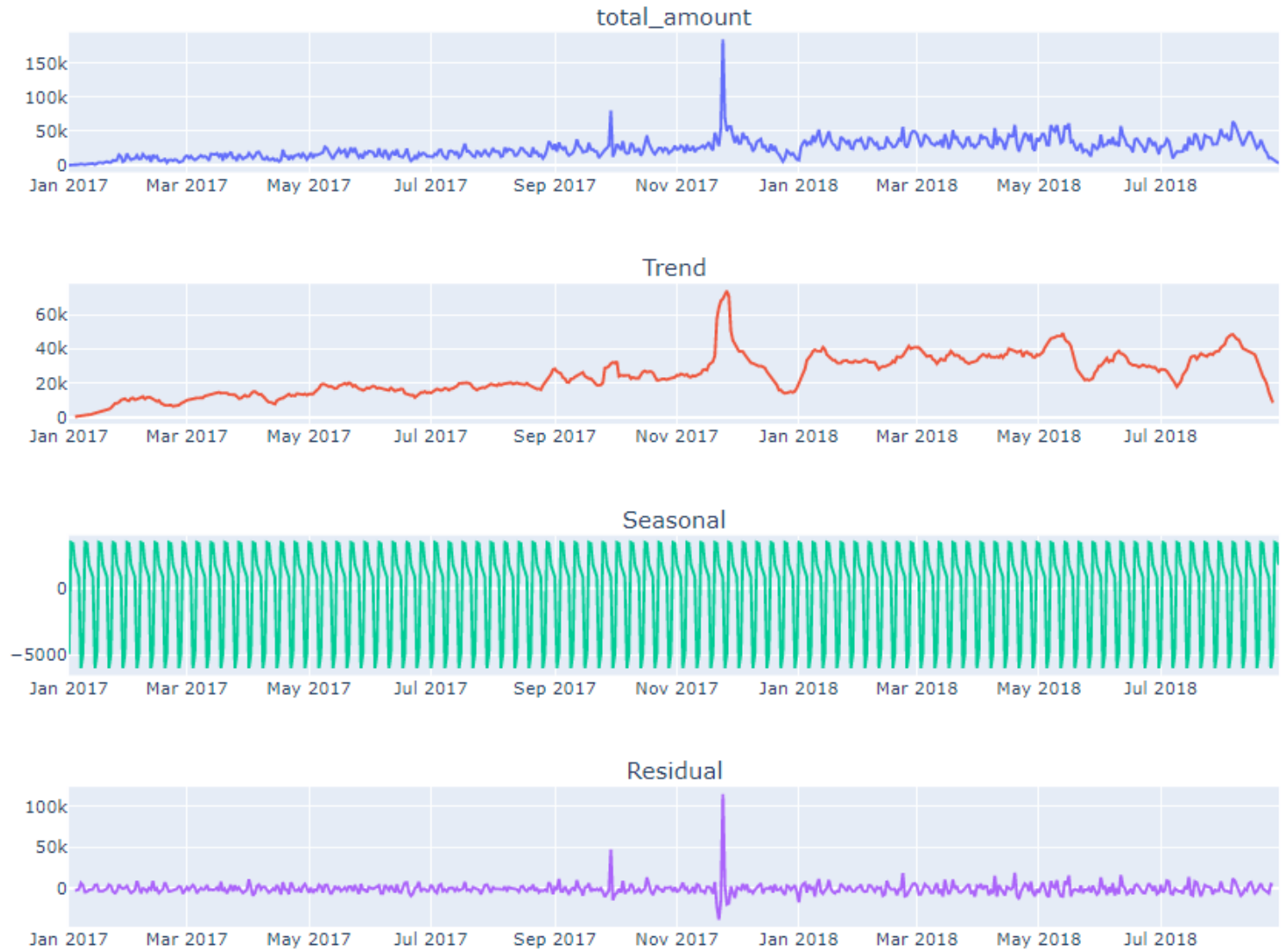
- Aggregated the Sales amount by date and it reduced the datapoints to two years of data with 606 observations.
- **Target: Sales**
- Note: Not considered the freight charges as Olist outsourced logistics to third party till 2020.

Daily Revenue from Sept 2016 to Aug 2018



Time Series Modelling

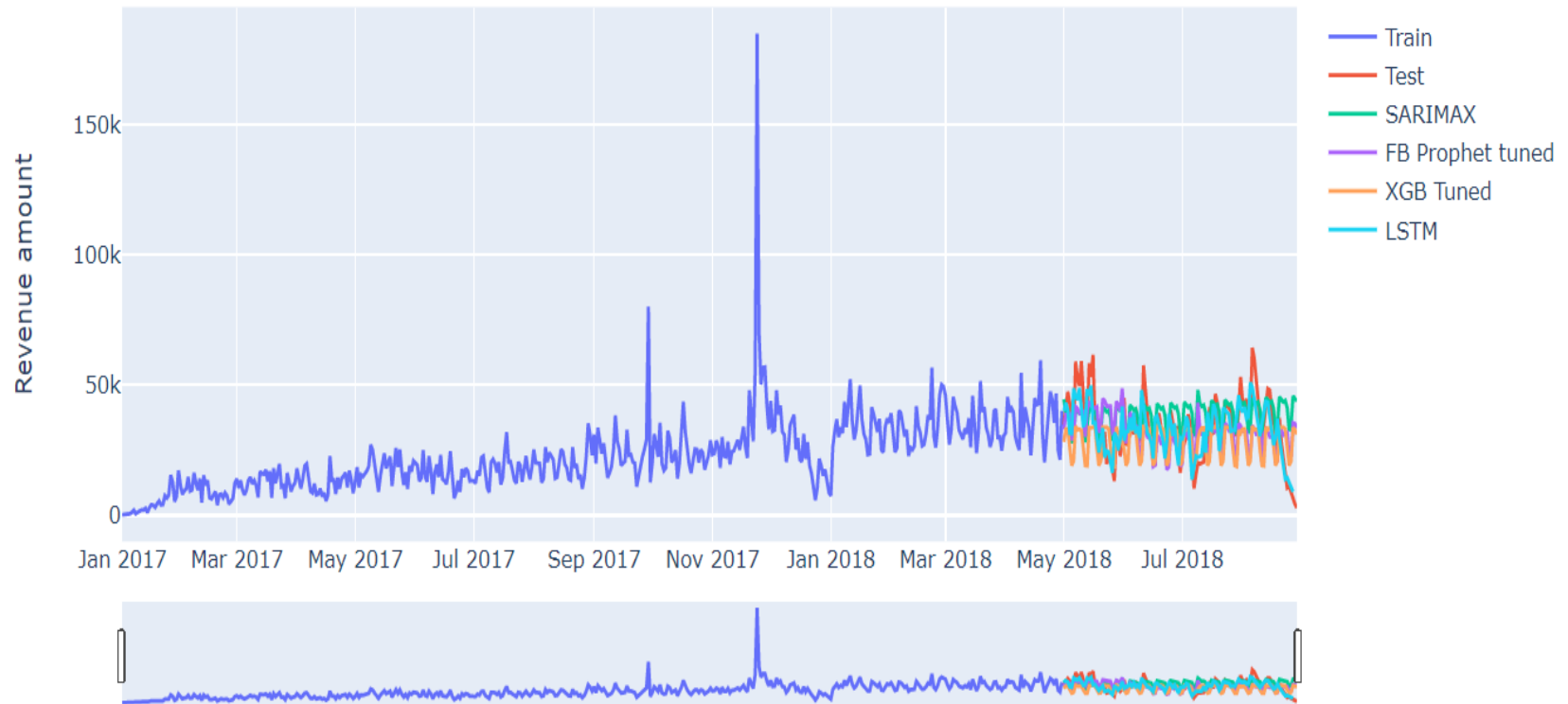
- Observed weekly seasonality.
- Positive Trend.
- Huge impact of holidays like Black Friday sale.



Time Series Modelling

- SARIMAX model including holiday impact.
- FB Prophet.
- Tuned Fb Prophet including holiday impact.
- XG boost
- Tuned XG boost
- LSTM Model

Daily Sales amount and forecast using different models



Result comparision

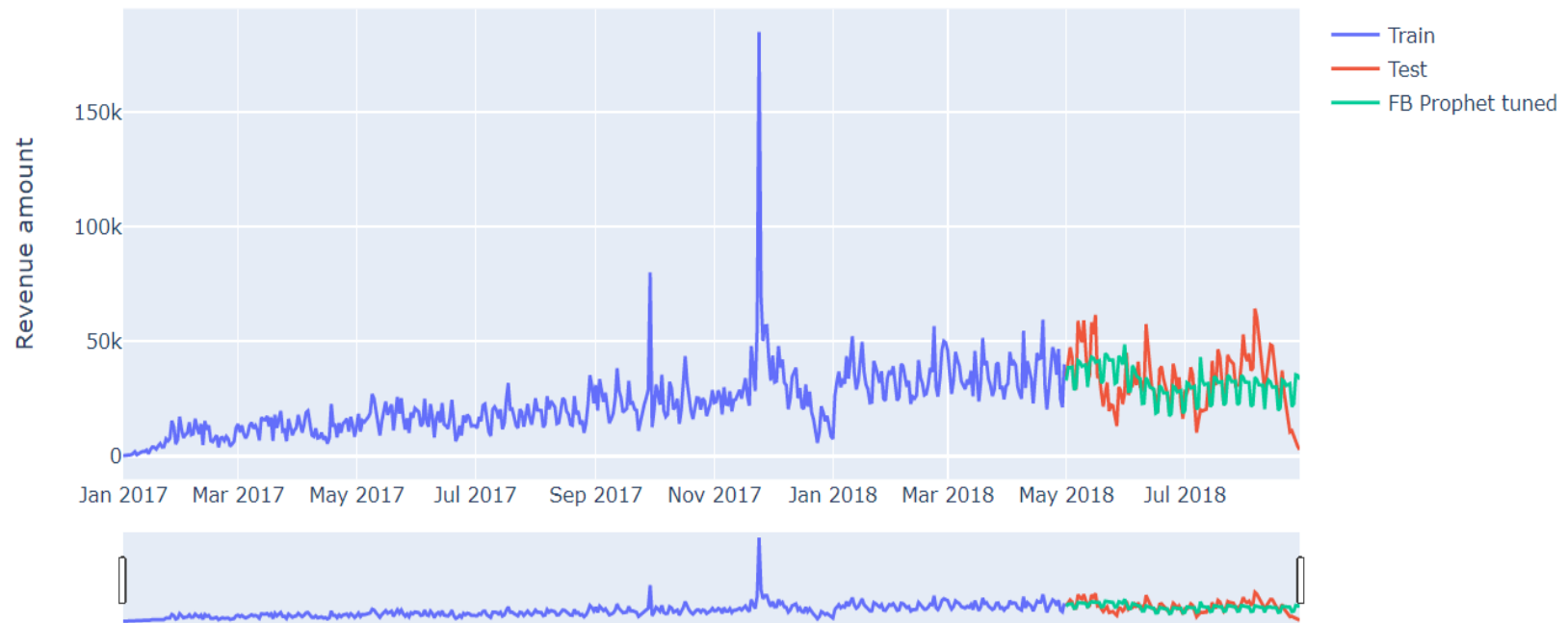
- LSTM has least MAPE but I am reluctant to proceed with this since we have limited historical data.
- XG Boost has MAPE of 47.58% but it is not able to pick up trend and variations between weeks.
- FB prophet next best with MAPE 51.45%

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

FB Prophet Results

- Going Ahead with FB Prophet.
- Able to pick seasonality, trend and variation within week.
- Gives a MAPE of 51.45% not very impressive but considerable.

Daily Sales amount and forecast using Tuned FB Prophet



Summary

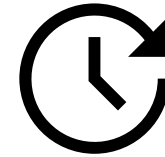
Achieved so far..



We analysed the current performance of business.

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.