TIME SERIES FORECASTING
STORE SALES

Emi Chua

Background Story

Basic Information

- Problem Statement
- Datasets
  - Train
  - Oil

● Problem Statement

- To build a time-series model that can accurately predicts the unit sales for items with highest sales and sold at one of the stores with highest sales.
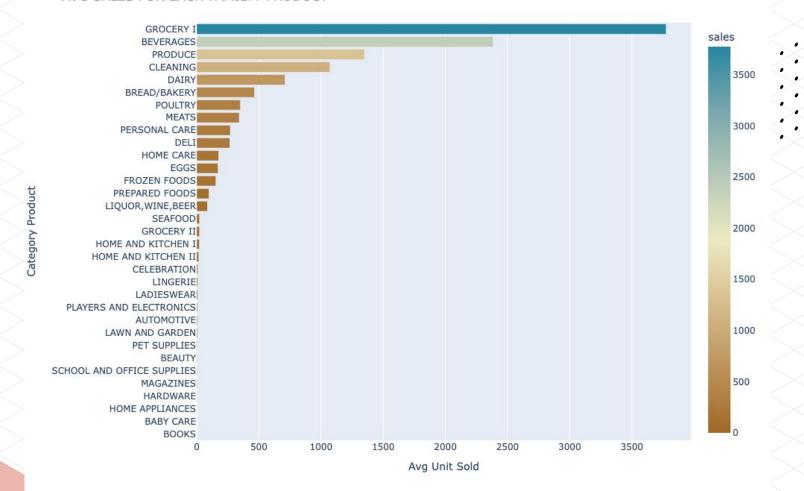
Basic information

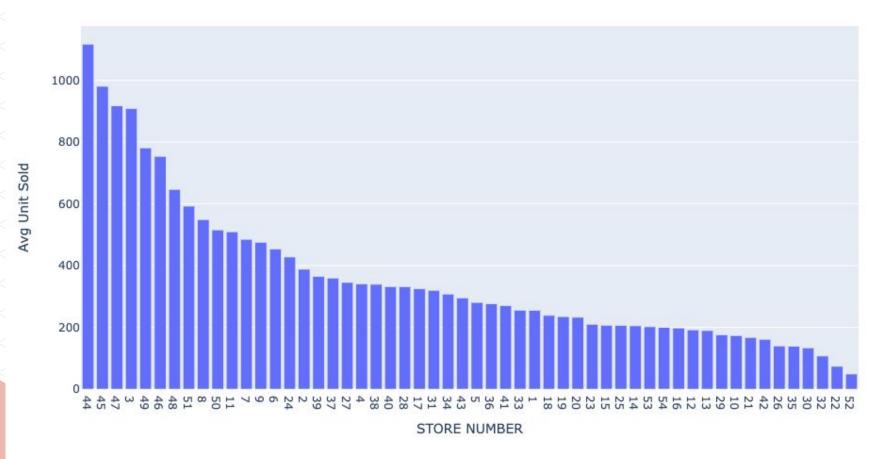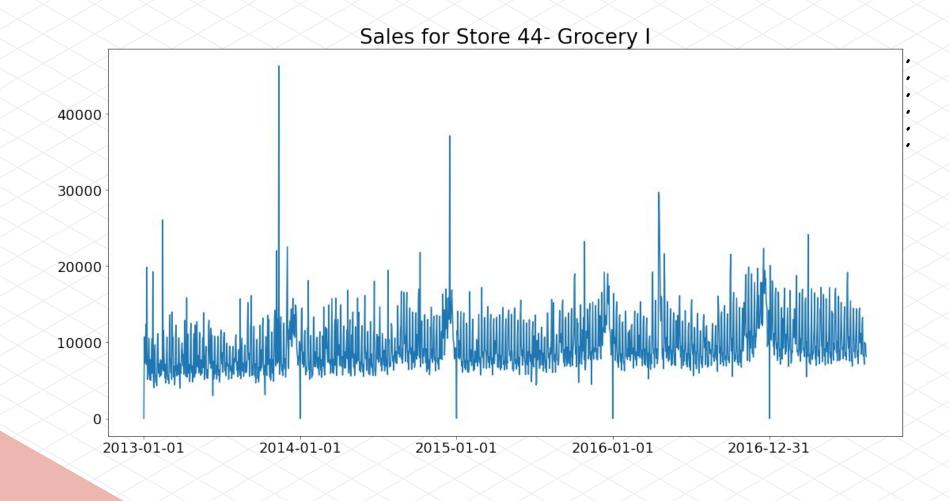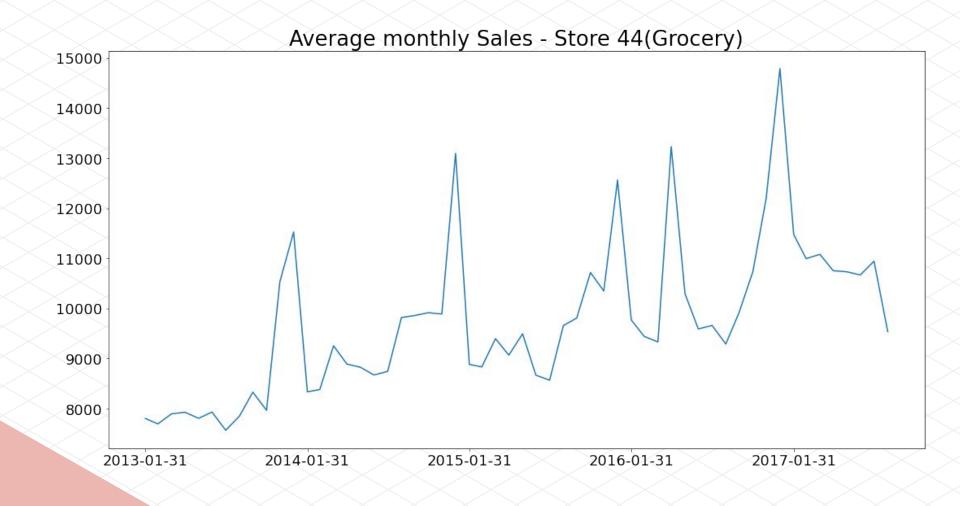| KPI	Value
Number of Stores	54
Number of Different Products	33
Window Start Date	2013-01-01
Window End Date	2017-08-15
Rows in training set	3000888
Date Points in Train Dataset	1684

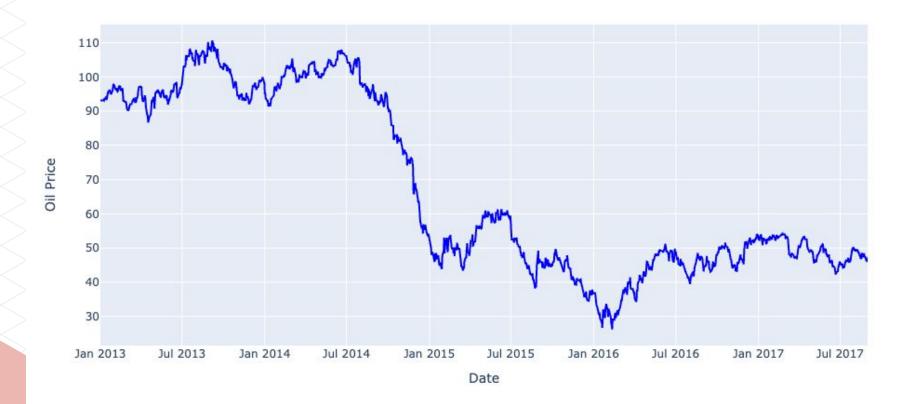# Exploratory Data Analysis

AVG SALES FOR EACH FAMILTY PRODUCT

AVG SALES FOR EACH STORE NUMBER
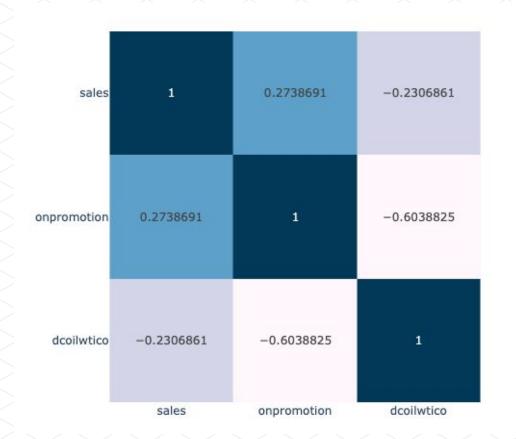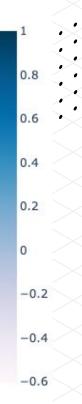
Sales for Store 44- Grocery I

Average monthly Sales - Store 44(Grocery)

Oil Prices Chart
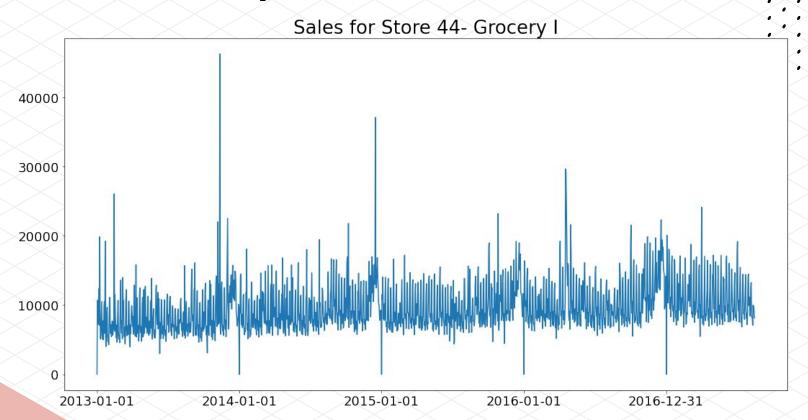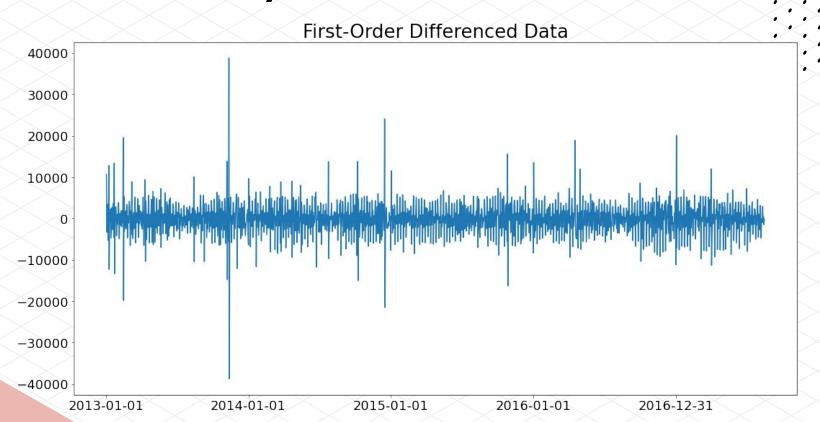
Hyperparameter Tuning

Stationary

Sales for Store 44- Grocery I

Stationary

First-Order Differenced Data
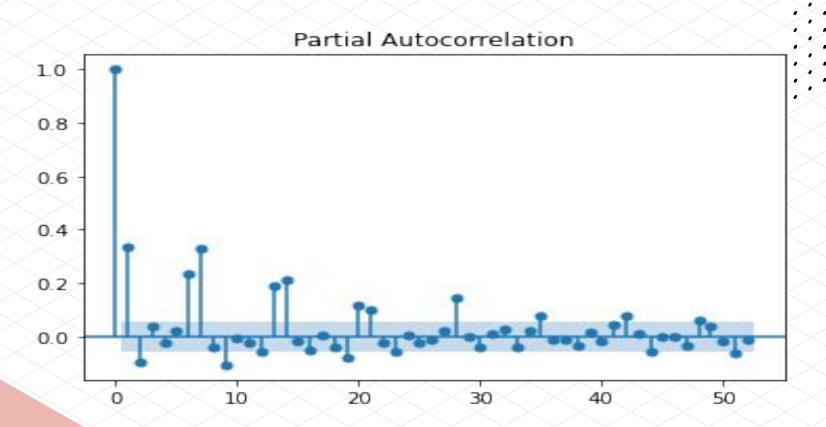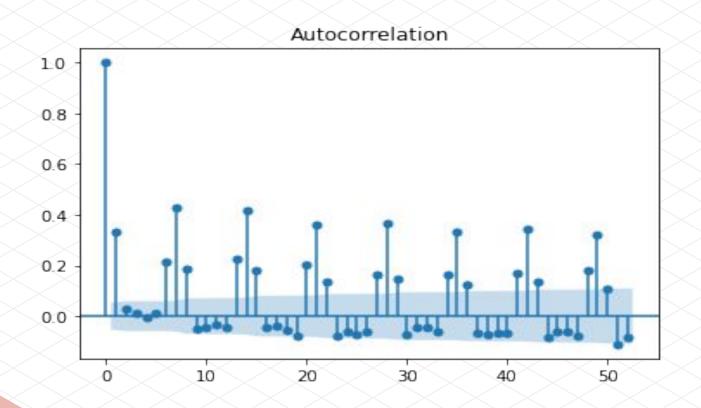
Augmented Dicky-Fuller Test

	Test Statistic	p-value
Original Data	-4.924952	0.000031
Once-Differenced data	-15.160692	0.000000

Partial Autocorrelation Function (PACF)

Partial Autocorrelation

Autocorrelation Function (ACF)

Autocorrelation

Modelling
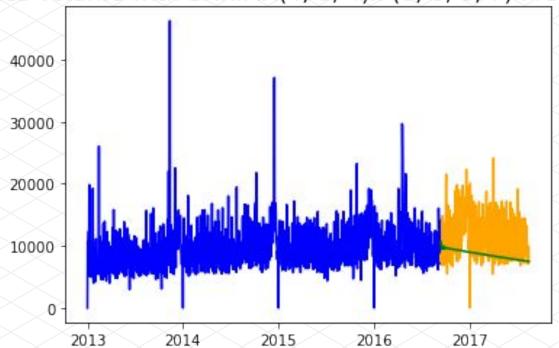
ARIMA

Sales with ARIMA(4,0,4) Predictions
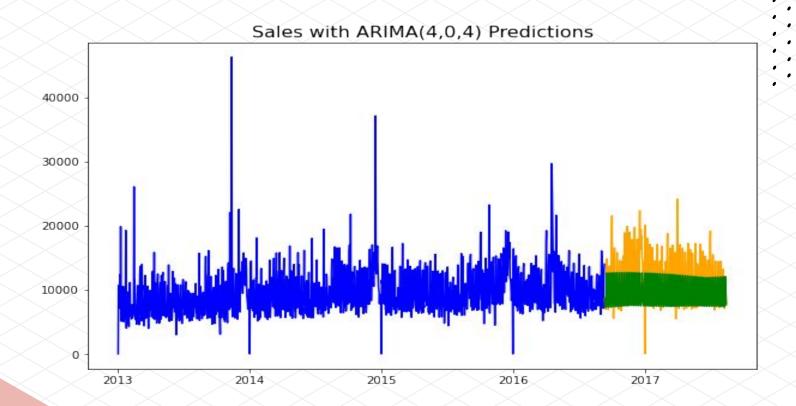
SARIMA

Sales Volume with SARIMA(4, 0, 4)x (0, 0, 0, 7) Predictions

Recommendation

Using qty instead of $

Instead of sales figure, sales qty can be included as promotional item can be in bundle/or cheaper price

Salesforce effectiveness

Besides having promotion, promoter in supermarket will also increase sales for certain items by having food tasting, freebies.

Include population/household income

In order to forecast more accurately, population & household income can be one of the few factors

Conclusion

The time series models I have used ie ARIMA and SARIMA can help to predict the store sales. However, more related feature should be included for higher accuracy.

THANKS