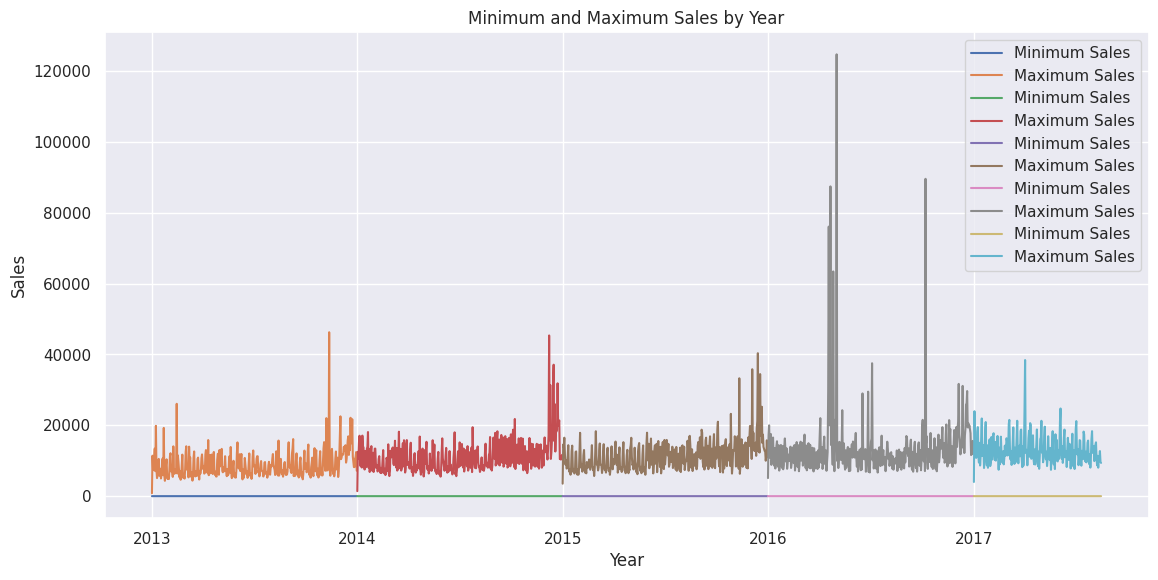
Sales is one of the most important units of any business enterprise. Its efficiency brings more revenue into the business accounts for the business continuity. The cumulative record of sales shows interesting patterns which may be beneficial to the data analyst and other interested parties for informed decisions. Sales could be affected by a lot of other factors which may include, seasonal constraints, bad economic conditions, oil prices, product promotions or simply the incidence of a natural disaster, sales records could provide a basis for employee promotion, appraisals or downsizing in other unforeseen situations. The above directive could be informed by a time series analysis of sales record.

This article is intended to take the reader through a data science project involving real world data set from an Ecuadorian grocery retailer, Corporation Favorita. In the project, the main goal was to perform a time series analysis of the sales at the various stores of Favorita scattered across Ecuador. With these patterns identified, a time series forecasting model was to be built to aid predictions of sales of the corporation. CRISP-DM approach of the data mining process was followed throughout the process.

In total, there were five datasets to be analyzed. The fields and dates were variable throughout all the datasets. Therefore, it was very necessary to understand the data fields to help the data concatenation into one to aid analysis. Exploratory analysis on the merged datasets was performed, after the business and data understanding phase. Major steps included: data cleaning, data overview, univariate, bivariate and multivariate analysis. Seven main questions were asked and explored to better understand the sales trends over the years (between 2013 – 2017). The question and hypothesis are made available at:

Below are the major insights from the sales trend:

1. There were missing dates for Christmas Days (between, 2013 to 2016). This could be attributed to the no sales information on Christmas days.
2. We had the highest sales in the year 2016, the trend showing the Second and Third Quarter of that particular year with a maximum of 124,717 sales. 
3. It was reported that, an earthquake struck Ecuador on April 16, 2016, and from the analysis the impact didn’t have a long-term negative effect on sales but rather an increase in yearly sales from the visual below.

Chart, histogram

Description automatically generated

After the analysis, time series forecast models were built to help in sales predictions. They are AR, ARIMA, SARIMA, Prophet Models. Both ARIMA and SARIMA models performed better with the lowest RMSLE of (0.27). Overall, SARIMA model was chosen after Hyperparameter Tuning of both models.

All analysis and models were developed in python. Follow the link below to access it: