DEMAND FORECASTING

GROUP MEMBERS -
ISHANA VIKRAM SHINDE -
DEEKSHA GANGADHARAN SRINIVAS -

INTRODUCTION -

Demand forecasting is the process of using predictive analysis of historical data to estimate and predict customers' future demand for a product or service. Demand forecasting helps the business make better-informed supply decisions that estimate the total sales and revenue for a future period of time. Demand forecasting allows businesses to optimize inventory by predicting future sales. By analyzing historical sales data, demand managers can make informed business decisions about everything from inventory planning and warehousing needs to running flash sales and meeting customer expectations.

GOAL: To predict the unit sales for thousands of items sold at different Favorita stores in Ecuador.

DATA: Our Dataset was taken from Kaggle.

Link for the Dataset -
The training data includes dates, store and item information, whether that item was being promoted, as well as unit sales. Additional files include supplementary information that may be useful in building your models.

File Descriptions and Data Field Information -

train.csv

- Training data, which includes the target unit_sales by date, store_nbr, and item_nbr and a unique id to label rows.
- The target unit_sales can be an integer (e.g., a bag of chips) or float (e.g., 1.5 kg of cheese).
- Negative values of unit_sales represent returns of that particular item.
- The onpromotion column tells whether that item_nbr was on promotion for a specified date and store_nbr.
- Approximately 16% of the onpromotion values in this file are NaN.

test.csv

- Test data, with the date, store_nbr, and item_nbr combinations that are to be predicted, along with the promotion information.

Additional files include supplementary information

stores.csv

DEMAND FORECASTING

GROUP MEMBERS -
ISHANA VIKRAM SHINDE -
DEEKSHA GANGADHARAN SRINIVAS -

- Store metadata, including city, state, type, and cluster.
- cluster is a grouping of similar stores.

items.csv

- Item metadata, including family, class, and perishable.

transactions.csv

- The count of sales transactions for each date, store_nbr combination. Only included for the training data timeframe.

oil.csv

- Daily oil price. Includes values during both the train *and* test data timeframe. (Ecuador is an oil-dependent country and it's economical health is highly vulnerable to shocks in oil prices.)

holidays_events.csv

- Holidays and Events, with metadata

APPROACH -

1. FACEBOOK PROPHET -

Prophet provides us with two models - one is the logistic growth model and the other one is the piece-wise linear model. By default, the prophet uses the piece wise linear model but it can be changed by specifying the model. In our case, we are using the prophet model to forecast the sales of the Favorita stores in Ecuador. The seasonality of our model is multiplicative and the reason we give it as multiplicative is that we have only 5 years of data so the product follows the same pattern. The trend shows the tendency of the data to increase or decrease over a long period of time and it filters out the seasonal variations. Seasonality is the variations that occur over a short period of time and is not prominent enough to be called a “trend”.

Understanding the Prophet Model

The general idea of the model is similar to a generalized additive model. The “Prophet Equation” fits, as mentioned above, trends, seasonality, and holidays. This is given by,

$$y(t) = g(t) + s(t) + h(t) + e(t)$$

Where, g(t) refers to trend (changes over a long period of time), s(t) refers to seasonality (periodic or short term changes), h(t) refers to effects of holidays to the

DEMAND FORECASTING

GROUP MEMBERS -
ISHANA VIKRAM SHINDE -
DEEKSHA GANGADHARAN SRINIVAS -

forecast, e(t) refers to the unconditional changes that are specific to a business or a person or a circumstance. It is also called the error term, and y(t) is the forecast.

2. REGRESSION MODELS -

We implemented various regression and statistical models

i) Linear Regression -

Linear regression is widely used in practice and adapts naturally to even complex forecasting tasks. The linear regression algorithm learns how to make a weighted sum from its input features. For two features, we would have:

target = weight_1 * feature_1 + weight_2 * feature_2 + bias

During training, the regression algorithm learns values for the parameters weight_1, weight_2, and bias that best fit the target

ii) Moving/Rolling Average -

The moving average is a statistical method used for forecasting long-term trends. The technique represents taking an average of a set of numbers in a given range while moving the range. A simple moving average takes the sliding window over a given time period. It can be termed as an equally weighted mean of n records.

Why Moving Average?

The moving average method is used with time-series data to smooth out short-term fluctuations and long-term trends.

We computed the rolling average of unit sales over a window of 7 days.

iii) Decision Tree Regressor -

Decision Tree predictions are averages of subsets of the training dataset. These subsets are formed by splitting the space of input data into axis-parallel hyperrectangles. Then, for each hyper rectangle, we take the average of all observation outputs inside those rectangles as a prediction. For regression against time, those hyperrectangles are simply splits of time intervals. Predictions are then the arithmetic means of the time-series observations inside those

DEMAND FORECASTING

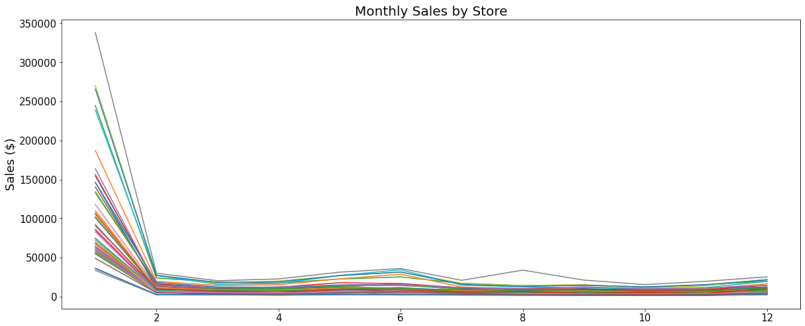
GROUP MEMBERS -
ISHANA VIKRAM SHINDE -
DEEKSHA GANGADHARAN SRINIVAS -

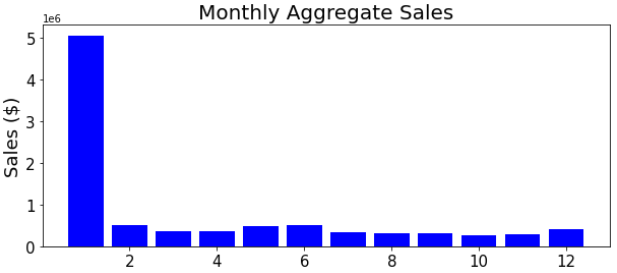
iv) Random Forest -

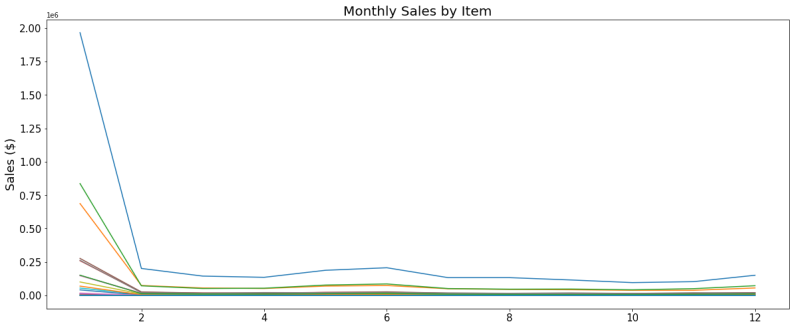
A prediction on a regression problem is the average of the prediction across the trees in the ensemble. We used Random Forest Regressor and tuned our model with various hyper-parameters like (max-depth, number of trees,max_bins, etc.)

RESULTS -

Analysis of the time series data -

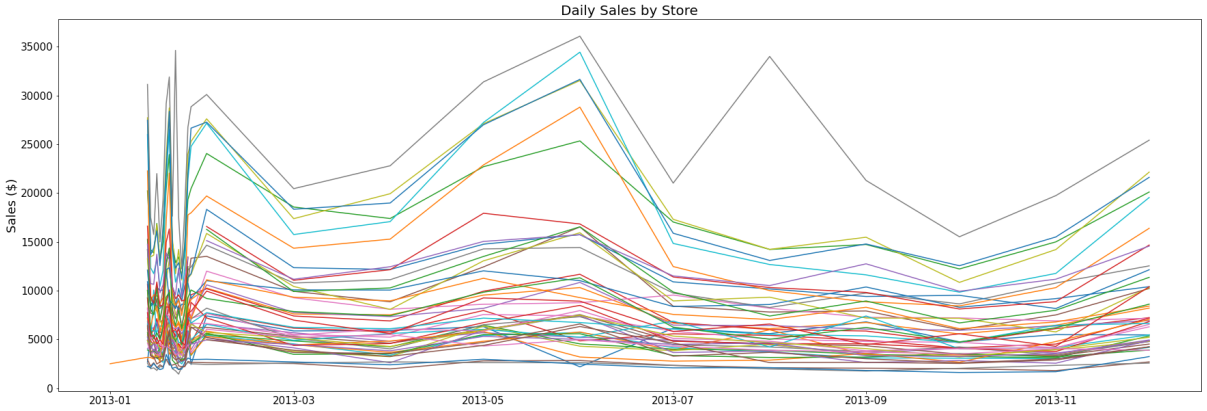




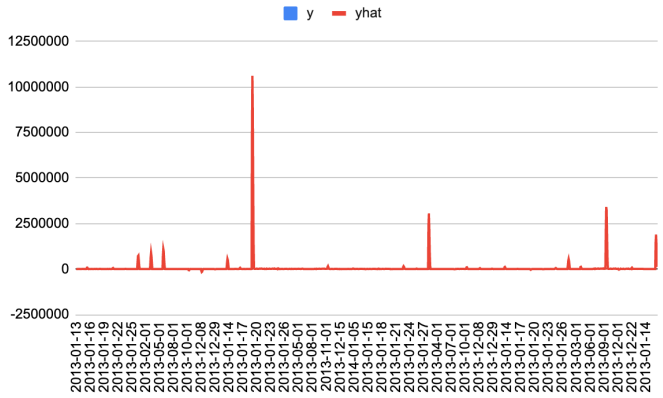


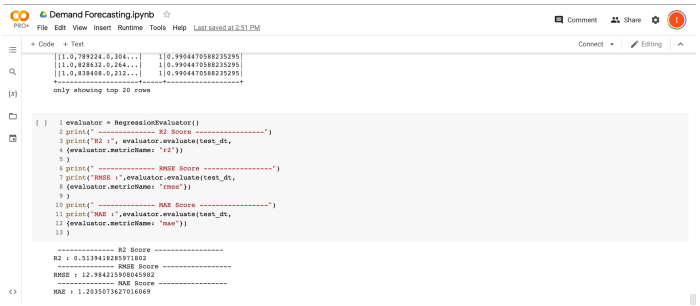
DEMAND FORECASTING

GROUP MEMBERS -
ISHANA VIKRAM SHINDE -
DEEKSHA GANGADHARAN SRINIVAS -



Results of the Regression Models and Time series model -





CLUSTER: We used the Prophet model by Facebook for Demand Forecasting we ran our code on Databricks Cluster. We have included instructions for setting up the cluster in our readme file.

Conclusion -

In conclusion, we can say that demand forecasting is important for businesses to decide what items they can sell during which period. We explored statistical and time series models for the task and reported our results.