Contents

Sales Forecasting for SmallBasket

Problem Description:

Small Basket is a huge online / mobile application based grocery retailer in India, founded in 2011. Small Basket is trying to manage its supply chain and delivery partners and would like to accurately forecast the sales for the period starting from 1st January 2019 to 15th January 2019 . You are also given a few features that were developed by the Business Intelligence team, that might or might not provide additional insights to your analysis.

In this hackathon, you are tasked with creating a model that can accurately forecast the sales for smallbasket across several locations and product categories.

About Data:

There are 5 datasets along with a sample submission file provided to you in this competition. The datasets are named as follows and have the following attributes:

- ***'train.csv':***
  - **date:** The sales measured during a particular day
  - **location_id:** The location from where the sold units were dispatched
  - **item_id:** The identifier of a product that was sold
  - **unit_sales:** The number of sales of a particular item from a particular location at the given date
  - **onpromotion:** Whether the given product was sold in a promotion or a discount

- ***'train_transactions.csv':***
  - **date:** The day during which the number of transactions are measured (Only given for the duration of the 'train.csv' file)
  - **location_identifier (location_id in train.csv):** The location from where the transactions were handled
  - **transactions:** The number of transactions handled by the particular location

- ***'items.csv':***
  - **item_id:** An identifier of a product
  - **category_of_item:** The category to which the product belongs to
  - **class:** Another way to categorize the product (Provided by the Business Intelligence Team)
  - **perishable:** Whether the item is perishable or not

- ***'locations.csv':***
    - **location_id:** The location of the store / warehouse (unit)
    - **city:** The city where the unit is located
    - **state:** The state in which the city is located
    - **type:** The type of business unit ('A', 'B', 'C', 'D', 'E')
    - **cluster:** The cluster that the unit belongs to (Provided by the Business Intelligence team)

- ***'test.csv' and 'sample_submission.csv' are also provided to you. They have an added column called 'id' which will be used to compute the evaluation metric.***

Objectives:

In this hackathon, you are expected to:

1. Explore the data and engineer new features
2. Predict the number of sales for a given item for the dates given in the 'test.csv' file
3. Answer questions from the operations team using the machine learning models that you have developed

Answering questions from the operations team at SmallBasket

The operations team at small basket has asked you the following questions

- Business units belonging to which cluster will see the highest amount of sales for the first 15 days of 2019?
- What are the top 10 selling items in this cluster?
- What is the **rate of purchase per week** for these items?

Evaluation Metric:

The evaluation metric used for this hackathon would be the **Mean Absolute Percentage Error** (**MAPE**)