**Data Science Hackathon**

**All Star Cohort 10**

**Store Sales Forecasting**

**Problem statement**

You have been given historical sales data for 30 retail stores of a retail brand situated in various regions, where each store consists of multiple departments. Your task is to make predictions for the sales figures of each department in every store.Furthermore,

**Dataset**

**stores.csv**

This file contains anonymized information about the 30 stores, indicating the type and size of store.

**train.csv**

This is the historical training data, which covers 30 stores with 68 unique departments and weekly sales numbers from 05-02-2016 to 27-07-2018. Within this file you will find the following fields:

* Store\_No - the store number
* Department - the department number
* Date - the week
* Weekly\_Sales -  sales for the given department in the given store
* Holiday\_ind - whether the week is a special holiday week

**test.csv**

This file is identical to train.csv without weekly sales from 03-08-2018 to 26-10-2018. You must predict the sales for each triplet of store, department, and date in this file.

**additional\_features.csv**

This file contains additional data related to the store, department, and regional activity for the given dates. It contains the following fields:

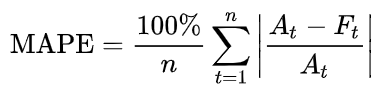
* Store\_No - the store number
* Date - the week
* Temperature - average temperature in the region
* Fuel\_Price - cost of fuel in the region
* CPI - the consumer price index
* Unemployment\_Ind - the unemployment rate
* Holiday\_Ind – holiday week or not

**sample\_submission.csv**

* Store\_No - the store number
* Department - the department number
* Date - the week
* Weekly\_Sales -  sales forecasted for the given department in the given store

**Evaluation**

Evaluation is based on following factors.

* Mean Absolute Percentage Error (MAPE)
* 
  + where *At* is the actual value and *Ft* is the forecast value. Their difference is divided by the actual value *At*. The absolute value of this ratio is summed for every forecasted point in time and divided by the number of fitted points *n*.
* Code Quality

## Submissions

* For each row in the test set, you should predict the weekly sales of that department. Submissions should be in the format sample\_submissions.csv.
* Codes (jupyter notebook or python files)