Table of Contents

I. Problem Description

Sales forecasting of a pharmaceutical retail chain

KioMed, a huge pharmaceutical retailer, operates in in 10+ cities across India. The company has one distribution warehouse in each of the cities it operates in. Unfortunately, the warehouses are not able to consistently meet the demand of the stores in their respective cities. Kio, the parent company, being a data driven corporation wants to solve the stocking / inventory management problem using their in-house data science team.

The retailer has provided you with historical sales data and is looking to forecast the sales for the period of one month after the end of the data. These forecasts will be used to ensure that the company is able to stock its supplies of medicines in a warehouse accordingly in each city for a period of one month.

The company will also provide you with the footfall data for all the stores across each of its cities. You can use this data, but as in the case of the real world, the footfall data is only available at train time and not at test time

II. Dataset Description

You will be given the following datasets:

- **train_data.csv**: *year*, *month*, *day*, *city*, *medicine*, *sales*
- **test_data.csv**: *id*, *year*, *month*, *day*, *city*, *medicine*
- **foot_fall.csv:** *footfall at all the stores in each city on a given day*
- **city_dict.json**: *a dictionary mapping the integer value present in the train and test datasets to the actual city*
- **sample_submission.csv**: *id*, *sales*

III. Hackathon Tasks

As part of this hackathon, you are expected to complete three tasks: i) Exploratory Analysis ii) ML Modelling iii) Final Recommendations

Exploratory analysis

Exploratory Data Analysis using visualizations, numerical analysis, and describing the findings.

- List down the insights/patterns observed from the visualizations
- Explain the impact of the most important attributes on the sales

ML Modelling

You are expected to create a robust fraud detection framework by engineering new features, tuning, and improving the baseline ML model performance.

Recommendations to the business

- What are your final recommendations for the problem?
- Any final visualizations you would use to convey your recommendations?
- Can you explain your ML model using non-technical terms?

IV. Evaluation Metric

- The evaluation metric for this hackathon is the **RMSE Score**

V. Submission Timelines

Submission No	File	Submission Format	Start Date	End Date
Submission - I	Exploratory Data Analysis and Initial Report	A zip file containing: • R Notebook or Jupiter notebook • HTML notebook	03-Dec 9:00 (Fri)	09-Dec 20:00 (Thu)
Submission - II	Predictions of test.csv (Target attribute: sales)	samplesubmission.csv	10-Dec 9:00 (Fri)	11-Dec 20:00 (Sat)
Submission - III	Improved version of predictions	samplesubmission.csv	12-Dec 9:00 (Sun)	12-Dec 20:00 (Sun)
Submission – IV	Improved version of predictions	samplesubmission.csv	14-Dec 9:00 (Tue)	14-Dec 20:00 (Tue)
Submission – V	Improved version of predictions	samplesubmission.csv	16-Dec 9:00 (Thu)	16-Dec 20:00 (Thu)
Submission – VI	A final report including answers to all the hackathon tasks.	A zip file containing: • R Notebook or Jupiter notebook • HTML notebook	17-Oct 9:00 (Fri)	18-Oct 20:00 (Sat)