# **Forecasting Unit Sales (Task 1)**

# This is the sales data for a brand that sells on Amazon.

**KAGGLE LINK:** [**https://www.kaggle.com/t/ef77ef06a5524357b4fca7b347ed56b6**](https://www.kaggle.com/t/ef77ef06a5524357b4fca7b347ed56b6)

**Close By: 02/08/24 6PM IST**

**Contact:** [thiruvikraman@napqueens.com](mailto:thiruvikraman@napqueens.com) for any queries

**Note:**

Please also include the R&D you conducted on time series analysis in a PDF file. It should clearly explain why and how you arrived at a particular model based on the research you made.

**Please Enable This Option:** “Share your email address with the host” - It will be under team column

**Task:**

In this task, you will work with dummy sales data of a well-known brand on Amazon. Your objective is to build a time series forecasting model that predicts the number of units sold for each item ID.

**Description:**

**This data pertains to Amazon (USA), so feature engineering should be conducted accordingly.**

The dataset contains dummy sales data for various items sold by a well-known brand on Amazon. Each row in the dataset represents the sales data for a particular item on a specific date. The columns included in the dataset are:

**date:** The date on which the sales data was recorded. This column is essential for time series analysis and forecasting. The format is typically YYYY-MM-DD.

**Item Id:** A unique identifier for each item. This column helps in distinguishing between different products.

**Item Name:** The name of the item. While this column is not directly used in numerical analysis, it provides a human-readable reference to the items.

**anarix\_id:** An internal identifier that might be used for tracking or categorizing items within the brand's system.

**ad\_spend:** The amount of money spent on advertising for the item on the given date. This column can be useful in understanding the relationship between advertising spend and sales performance.

**units:** The number of units sold for the item on the given date. This is the target variable for forecasting in this task.

**orderedrevenueamount:** The total revenue generated from the units sold on the given date. This can be used to calculate metrics like average unit price and return on ad spend.

**unit\_price:** The price per unit of the item. This column helps in understanding the pricing strategy and its impact on sales.

### 

### **Evaluation**

Your model will be evaluated based on its ability to accurately forecast the number of units sold for each item ID. The primary evaluation metric will be the Mean Squared Error (MSE). This metric measures the average magnitude of the errors between the predicted and actual values, with lower MSE values indicating better model performance.

**Submission Requirements:**

Ensure that you submit your work on Kaggle and also create a Google Drive folder containing the following:

* The .ipynb notebook file
* A PDF version of the notebook
* The file with the predicted results

**Expected Content:**

Your submission should include the following sections:

* Exploratory Data Analysis (EDA)
* Feature Engineering
* Model Selection
* Hyperparameter Tuning

**Include the link to your Google Drive folder in the provided sheet:**

Give access to the following mail alone: [thiruvikraman@napqueens.com](mailto:thiruvikraman@napqueens.com)

Sheet: [Submission](https://docs.google.com/spreadsheets/d/1XpBSkWIL0zhilhYSKbsXvDpdsRw1kt0F8Cuu3OKMU2w/edit?usp=sharing)