Data Glacier Internship – Project

Batch: LISUM41 (30 December, 2024 – 30 March, 2025)

Team Member Details:

Name: Mahnoor Farhat

Email: mahnoor.farhat@gmail.com

Country: United Kingdom

College: University of Hertfordshire (Sep 2023 – Oct 2024)

Specialization: Data Science

GitHub: https://github.com/mahnoor-farhat/data-glacier-project

Problem Description:

The time series data showed a range of patterns, some with trends, some seasonal, and some with neither. At the time, they were using their own software, written in-house, but it often produced forecasts that did not seem sensible. The beverage company wanted to explore power of AI/ML based forecasting to replace their in-house local solution.

Exploratory Data Analysis (EDA) Summary

1. Sales Trend Analysis:

- Created a **line plot** to visualize the trend of Sales over time, categorized by Product.
- Observed patterns in sales fluctuations across different time periods.

2. Distribution Analysis:

- Generated a **histogram with KDE (Kernel Density Estimation)** to examine the distribution of Sales.
- Identified potential skewness or irregularities in the sales data.

3. Impact of Discounts on Sales:

- Created a **scatter plot** of Price Discount (%) against Sales.
- Identified potential correlations between discount rates and sales volume.

4. Promotional Effect on Sales:

- Created a **box plot** to analyze the effect of In-Store Promo on Sales.
- Evaluated whether promotional activities significantly impact sales performance.

5. Correlation Heatmap:

- Plotted a **heatmap** to visualize correlation coefficients between numerical features.
- Identified relationships between sales and promotional activities, mobility trends, etc.

6. Seasonal Analysis:

- Extracted the **month** from the date column.
- Created a **box plot** to analyze the distribution of sales across different months.
- Investigated seasonal variations in sales trends.

7. Moving Average Analysis:

- Calculated a **4-week moving average** of Sales to smooth short-term fluctuations.
- Overlaid actual sales and moving average trends in a line plot for better trend identification.

8. Outlier Detection:

- Created a **box plot** to detect and visualize outliers in Sales.
- Analyzed data distribution to assess potential anomalies.