Name: Andrea Perez

Email: apereza1@asu.edu

Country: United States

College/Company: ASU/Department of VA

Specialization: Data Science

GitHub Repo link: https://github.com/aperezace20/Data-Science-Retail-Forecasting/tree/main/Week%209:%20Deliverables

Problem description:

The large company who is into beverages business in Australia. They sell their products through various super-markets and also engage into heavy promotions throughout the year. Their demand is also influenced by various factors like holiday, seasonality. They needed forecast of each of products at item level every week in weekly buckets.

ML Problem:

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. Company wanted to explore power of AI/ML based forecasting to replace their in house local solution.

Steps taken:

1. Imported necessary Python modules for data analysis and visualization.
2. Imported the dataset from a GitHub repository or other source.
3. Applied Data Cleaning Best Practices to remove duplicates, correct data types, and handle missing values.
4. Explored the type of data by reviewing column data types and unique values.
5. Used the **`head()`** and **`tail()`** functions to inspect the first and last few rows of the data.
6. Examined the shape of the data by checking the number of rows and columns.
7. Checked for missing values in the data using functions like **`isnull()`** and **`sum()`**.
8. Determined the size of the data by calculating the memory usage of the data using **`info()`** function.
9. Split the data by month and added a new column 'Month' to the data.
10. Checked the head of the updated dataset to confirm the changes made.

11. Grouped the data by month and calculated the sales for each month.
12. Prepared to do the visuals in the following week using Python visualization libraries.

Deadline:

Deliverables	Date
Week 9: deliverables	2 March 2023
Week 10: deliverables	9 March 2023
Week 11: EDA Presentation and proposed modeling technique	16 March 2023
Week 12: Model Selection and Model Building/Dashboard	23 March 2023
Final Project Report and Code	30 March 2023

Data Intake report

Project Name: Data Science Retail Forecasting

Internship Batch: LISUM17

Data Intake by: Andrea Perez

Data Storage: GitHub link

Meta Data:

Number of Observations	1219
Number of files	1
Number of features	12
Formal of file	.csv
Data Size	1.09 MB