Team members: Jason Sanchez
Group name: Alpha
Email: jasonfs2@illinois.edu
College: Harold Washington College
Specialization: Data Science
Problem Description:
The large company which 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 a forecast of each of the products at item level every week in weekly buckets.

Data Intake Report

Name: Cross selling recommendation
Report date: September 19, 2023
Internship Batch: 09454990337545509761
Version: N/A
Data Intake by: Jason Sanchez
Data Intake reviewer: Data Glacier Reviewer
Data Storage location: https://github.com/Kuzma12/VC/tree/Project-week-1

Tabular data details:

Test.csv:

Total numbers of observations	929,615
Total number of files	2
Total number of features	24
Base format of the file	csv
Size of the data	105.18 MB

Train.csv:

Total numbers of observations	13647309
Total number of files	2
Total number of features	48
Base format of the file	csv
Size of the data	2.14 GB

Proposed Approach:

Cleaning data is an essential first step in any data analysis process. Before diving into the analysis, it's crucial to ensure that the dataset is free of errors, inconsistencies, and missing values. Data cleaning involves tasks such as handling duplicates, filling in missing data points,

and correcting erroneous entries. This process not only ensures the accuracy of the analysis but also enhances the reliability of any patterns or insights derived from the data.
Once the data is cleaned, the next step is to identify common patterns within it. This involves conducting exploratory data analysis (EDA) to understand the distribution of variables, relationships between variables, and potential trends. Visualization tools like histograms, scatter plots, and heat maps can be used to visualize data patterns effectively. Common patterns might include correlations between variables, seasonality trends, or outliers that could significantly impact the analysis. Sorting the data is another crucial step, especially when dealing with time-series or sequential data. Arranging data in a logical order, such as by date or numerical values, can reveal temporal or ordinal patterns. Sorting can make it easier to identify trends, anomalies, or seasonality within the data. After cleaning, identifying common patterns, and sorting the data, it's time to report a data analysis strategy based on observations. This strategy may include defining research questions, hypotheses, or goals for the analysis. It should also outline the analytical techniques, statistical methods, or machine learning algorithms to be used to extract meaningful insights. Additionally, the strategy should address how to handle and interpret the identified patterns, whether they are used for predictive modeling, optimization, or decision-making.

In conclusion, data analysis is a multi-step process that begins with data cleaning to ensure data integrity. Identifying common patterns, sorting data, and reporting a well-defined analysis strategy are subsequent steps that enable meaningful insights to be derived from the data. A thorough and systematic approach to data analysis is essential for making informed decisions and solving complex problems in various domains, from business analytics to scientific research.

Github Repo Link:
https://github.com/Kuzma12/VC/tree/Project-week-1