Capstone Project

support@intellipaat.com

+91-7022374614

US: 1-800-216-8930 (Toll-Free)

Problem Statement 1:

A retail store that has multiple outlets across the country are facing issues in managing the inventory - to match the demand with respect to supply.

Dataset Information:
The walmart.csv contains 6435 rows and 8 columns.

Feature Name	Description
Store	Store number
Date	Week of Sales
Weekly_Sales	Sales for the given store in that week
Holiday_Flag	If it is a holiday week
Temperature	Temperature on the day of the sale
Fuel_Price	Cost of the fuel in the region
CPI	Consumer Price Index
Unemployment	Unemployment Rate

1. You are provided with the weekly sales data for their various outlets. Use statistical analysis, EDA, outlier analysis, and handle the missing values to come up with various insights that can give them a clear perspective on the following:
    a. If the weekly sales are affected by the unemployment rate, if yes - which stores are suffering the most?
    b. If the weekly sales show a seasonal trend, when and what could be the reason?
    c. Does temperature affect the weekly sales in any manner?
    d. How is the Consumer Price index affecting the weekly sales of various stores?
    e. Top performing stores according to the historical data.
    f. The worst performing store, and how significant is the difference between the highest and lowest performing stores.
2. Use predictive modeling techniques to forecast the sales for each store for the next 12 weeks.

Problem Statement 2:

You are working in an e-commerce company, and your company has put forward a task to analyze the customer reviews for various products. You are supposed to create a report that classifies the products based on the customer reviews.

Dataset Information:
The Reviews.csv dataset contains 60145 rows and 10 columns.

Feature Name	Description
Id	Record ID
ProductId	Product ID
UserId	User ID who posted the review
ProfileName	Profile name of the User
HelpfullnessNumerator	Numerator of the helpfulness of the review
HelpfullnessDenominator	Denominator of the helpfulness of the review
Score	Product Rating
Time	Review time in timestamp
Summary	Summary of the review
Text	Actual text of the review

1. Analyze the customer reviews data, perform EDA and statistical tests to gather insights about the products.
    a. Highest and lowest rating for the products. Percentage wise product ratings for the entire data.
    b. Total number of reviews by unique profiles. How many customers or profiles have reviewed more than one product?
2. Perform sentiment analysis on the reviews data, and classify the reviews based on the sentiment associated with the same.