Text Mining Group Project Proposal
OPIM-5671- Data Mining & Business Intelligence
Professor: Sudip Bhattacharjee

Walmart Sales in Connecticut

Group 4:

Harsha Alla
Nimish Pastaria
Pratik More
Sri Vinay
Zilon Mao

PROBLEM STATEMENT:

To develop a predictive model that accurately forecasts weekly sales for a Walmart store by taking into account historical sales data and other non-intuitive predictive parameters including temperature, fuel prices, unemployment rates in the state, holidays and Consumer Price Index (CPI) information.

RELEVANCE:

In the current supply chain industry, the cost of storing inventory is very high, given the huge demand and supply. Through accurate forecasting of the sales, we can optimize inventory management and timely delivery of stock to different store locations. The proposed sales forecasting model aims to have the capability to predict sales for a Walmart store based on attributes such as unemployment rate, CPI, fuel prices etc. This initiative also implicitly aims to enhance inventory management, staffing optimization, and strategic planning, ultimately improving operational efficiency and profitability by aligning stock levels with forecasted demand, improving staffing schedules to ensure adequate customer service during peak and off-peak periods, and informing strategic decisions such as promotions and pricing strategies based on anticipated sales trends. This targeted approach can also enhance customer satisfaction, reduce operational costs, and increase profitability by ensuring resources are efficiently allocated to meet consumer needs.

DATA DESCRIPTION:

Parameter	Description
Date	The Week of Sales. It is in the format of dd-mm-yyyy. The date starts from 05-02-2010
Weekly_Sales	The sales of the store in the given week
Holiday_Flag	If the week has a special Holiday or not. 1 - The week has a Holiday 0 - Fully working week
Temperature	Average Temperature of the week in the area
Fuel_Price	Price of the Fuel in the region
CPI	Customer Price Index
Unemployment	Unemployment rate of the region

DATA SOURCE:

The primary dataset for this project is from Kaggle (link mentioned below), featuring historical sales data for a Walmart store. This dataset includes Weekly Sales, Holidays, and other relevant features like Temperature and Fuel Prices, affecting sales performance.
https://www.kaggle.com/datasets/varsharam/walmart-sales-dataset-of-45stores

PLAN OF ACTION:

1. Data Preparation: We'll make sure the data is structured in a time series format, while ensuring data is appropriately indexed and missing values are addressed.
2. Exploratory Data Analysis: Utilize SAS to perform exploratory analysis, identifying underlying patterns, trends, and seasonal variations in sales data using Time Series Forecasting and Model Exploration.
3. Feature Engineering: We'll also try to develop some features that capture seasonal effects, holidays, and other relevant exogenous variables that could impact sales from the list of predictor variables.
4. Model Building in SAS: We'll try to apply different AR and MA models or ARIMAX modeling to integrate external variables (like promotions or economic indicators) with the time series data.
5. Model Evaluation and Selection: Using the results of different models, we'll use SAS to evaluate model performance through appropriate metrics (e.g., AIC, SBC, MAPE etc) and select the best performing model for forecasting future sales.
6. Evaluating the Work: We'll measure how well our model did, using some of the metrics we learned about, to see how close our predictions are to the actual sales.