DSO 522 Time Series Analysis Final Project Proposal

Team 8: Freda Lin, Zihan Ling, Pizheng Zhang, Zhiyu Zhang

Overview

The objective of the project is to build an optimal time series model to forecast the store sales of Walmart, which supports the company to make data-driven decisions in finance, marketing, supply chain, etc. The inspiration of conducting the sales forecast on the Walmart dataset comes from the demand for accurate performance prediction of the traditional retailing industry with the rise of e-commerce platforms in recent years.

Dataset Description

The *Walmart Dataset* is from Kaggle and it contains historical sales data from 45 different Walmart department stores between 2010-02-05 and 2012-11-01. The time series data was reported weekly as "weekly_sales" and there is no missing value in the dataset. In addition to date and weekly sales, the dataset also includes some relevant features such as whether it's holiday season, temperature, fuel price, and consumer price index, which really helps us build a more robust model to forecast the store sales by taking these factors into consideration. The data also includes holiday events such as Super Bowl, Labor Day, Thanksgiving, and Christmas.

Methodology

We plan to apply dataset in processes such as data cleaning, predictive modeling, validating, graphing, analyzing, and presenting. In the data preparation stage, we will conduct exploratory data analysis on the data and use the mean of the 45 stores on the same date for later forecasting. We will split the data to training and validation set to forecast the sales for the next 10 months. We will first apply classical decompositions method by evaluating elements such as trend, seasonality, cycle, and noise. In the modeling process, we will use different models to forecast and pick the best model by looking into the accuracy and the robustness of the model as well as the analysis from business insight. For instance, we will evaluate the models based on RMSE, MAPE, and more. Methods including but not limited to naïve forecasting, simple exponential smoothing, linear regression model, multiplicative seasonality forecast, and ARIMA model will be implemented during the modeling stage.