Capstone Project - Walmart

Table of Contents

Problem Statement :

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. To come up with useful insights using the data and make prediction models to forecast the sales for X number of months/years.

Project Objective

1. Come up with useful insights that can be used by each of the stores to improve in various areas.
2. Forecast the sales for each store for the next 12 weeks.

Data Description

Data consists of Store number, Date, Weekly_Sales, Holiday_Flag, Temperature, Fuel_Price, CPI & Unemployment of 45 stores individually to predict the next 12 weeks sales

Data Pre-processing Steps and Inspiration

1. Convert the Date column to datetime object type
2. Changing Date column to index for easier Time Series computation
3. Remove the Date column
4. Plot the time series to check the seasonality & trend
5. Before building the model, it is required to check the time series is Stationary/Non-Stationary using AD-Fuller Test

6. If the data is Stationary proceed with building the model
7. Used Pyramid ARIMA i.e PMDARIMA to calculate the order for the ARIMA model (auto_arima)
8. Splitting the train data & test data as 70% and 30% respectively

Choosing the Algorithm for the Project

SARIMAX (Seasonal AutoRegressive Integrated Moving Average with eXogenous variables)

Motivation and Reasons For Choosing the Algorithm

SARIMAX model that can handle both seasonal and non-seasonal to forecast a wide variety of time series data

Assumptions

1. As most of the forecast data falls of the Festival months of November, December & January the sales are expected to be higher than usual.
2. Holidays are expected on the above months might also have a positive impact on the Sales

Model Evaluation and Techniques

Checking the Accuracy by using MAPE (Mean Absolute Percentage Error) i.e. 100% - MAPE%

Inferences from the Same

1. Unemployment data has a negative correlation with Weekly sales, which shows the when there is low unemployment the weekly sales.
2. Store 20 had the highest sales and Store 33 had the lowest sales
3. CPI (Consumer Price Index) data also has negative effect with Weekly Sales, indicating that as CPI data increases, weekly sales tend to decrease
4. Holiday_Flag data has a (+) positive effect in Weekly Sales
5. Temperature data has a (-) negative effect in Weekly Sales, people tend to shop during colder climates
6. The sales peaks during the Winter season (December) as it is a month of celebration
7. There average sales in holidays are higher than average sales in working days

Stores may need to adjust the Pricing Statergy or provide Festive coupons which will attract more customer and result in increase the sales volume as the forecast weeks fall in festive months.

By marketing the Seasonal products on Holidays will have positive impact on the Weekly Sales

The pricing stratergies can take inference from CPI & Unemployment data

Future Possibilities of the Project

Analysis of Sales pattern can be performed for each store with higher accuracy rate and inferences can be drawn for each stores

Conclusion

By analyzing correlations between unemployment rates, CPI, holiday seasons, and temperature, businesses can take decisions about pricing, promotions, and inventory management to increase sales.

In [ ]: