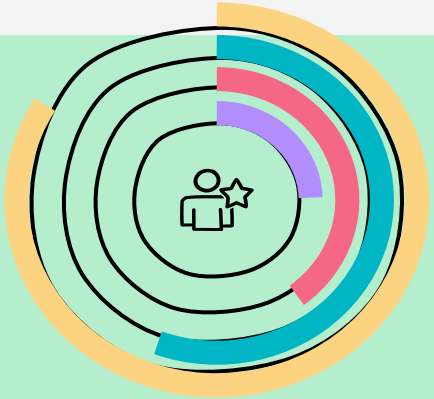
Walmart Sales Analysis

DS207: Time Series Forecasting

Inna Krmoyan, Irena Torosyan, Gor Mkrtchyan

Table of Contents



Walmart

Save money. Live better.

Abstract

Many different American multinational retail corporations that operate a chain of hypermarkets, discount department stores, and grocery stores that sell household goods which are geographically located at various locations. It is not always possible for the retailers to understand the condition of the market at different geographical locations. The retail store corporations need to understand the conditions of their markets to intensify the sales of their products for a large number of customers to get attracted in that particular direction. Time Series Forecasting helps the retailers to understand and visualize the big picture of the sales depending on various time periods. Forecasting also helps the retailer to understand when a certain product is demanded so that they increase their sales at that given period of time. In this research report, we will make an attempt to understand the driving factors that lead to huge sales, try to find trends, and seasonality, and understand how time series forecasting can benefit the chosen company that is Walmart.

Introduction

As curious young enthusiasts, we are very interested in the future and its all possible developments.

The aim of our project is to understand the reasons, relationships between the variables, the main driving factors that affect the sales so that we can predict the possible future pattern of the sales of ‘‘Walmart’’.

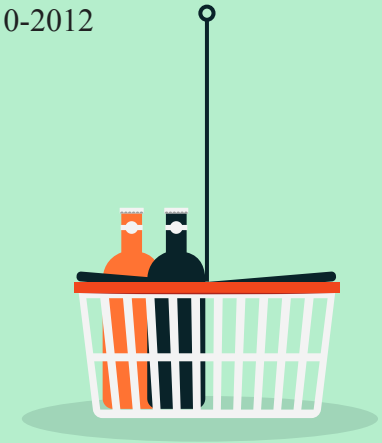
We want to understand our data make any necessary changes to make good visualizations that will show us the fluctuations of the sales

Analysis Question

There are many organizations that find it difficult to forecast sales because of some reasons as as constant introduction of new products, seasonal/weather changes, etc. Retailers have moved to large-scale demand forecasting that can handle a lot of transaction data in an effort to solve these problems. Retailers can mine this data and predict future consumer behavior by gathering them. Retailers have the chance to optimize their revenue system by using the capacity to forecast at such a big scale, which enables them to make better decisions on promotions and price. For our project we will face the challenge of making valid sales predictions that will actually benefit the company itself and other stakeholders who want to make reasonable changes to their business plans.

Data Description

The file has information about the Weekly Sales of 45 stores for the years 2010-2012



Column names

Store, Date Weekly_Sales, Holiday_Flag, Temperature, Fuel_Price, CPI, Unemployment

Number of observations: 6435

The covered period of time / frequency: 2010 - 2012 / annual

Exploratory Analysis

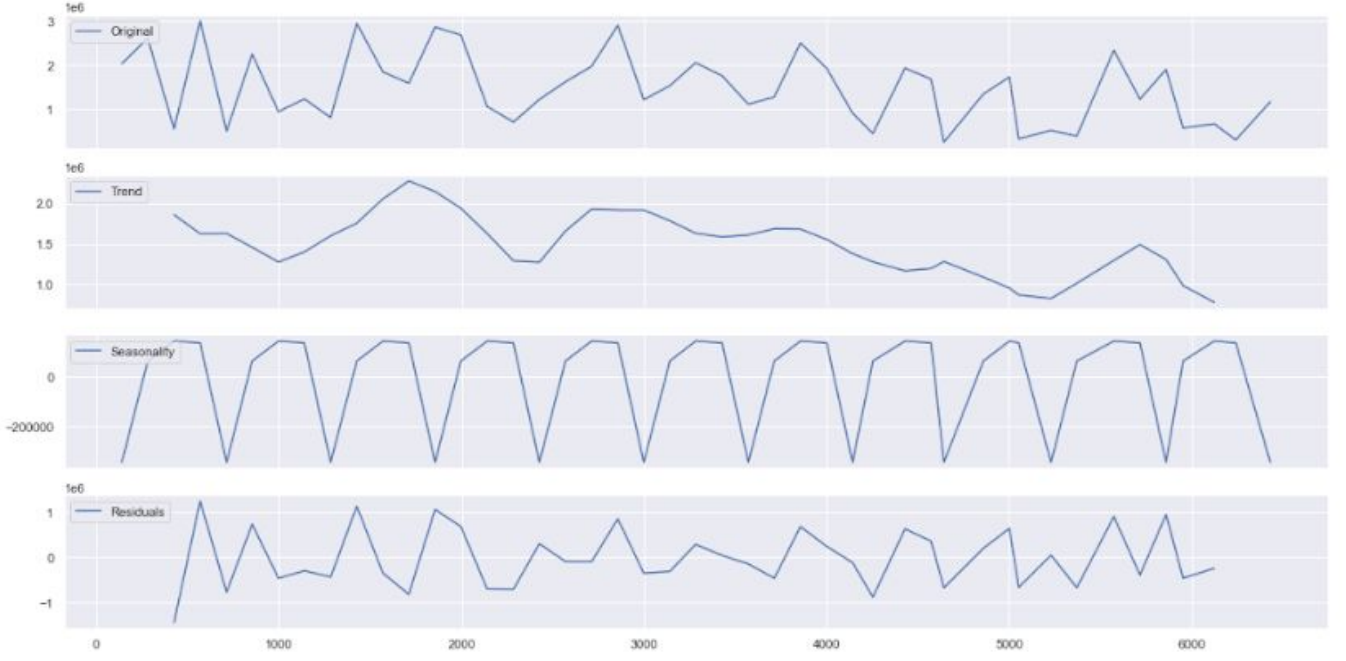
In this part we got to know the dataset we explored it and understood more thoroughly what we are working with.

Methods used in this part

Modified the date, got the top stores, printed the plot of the Weekly_Sales, extracted seasonality and trend etc.

An example from the exploratory part

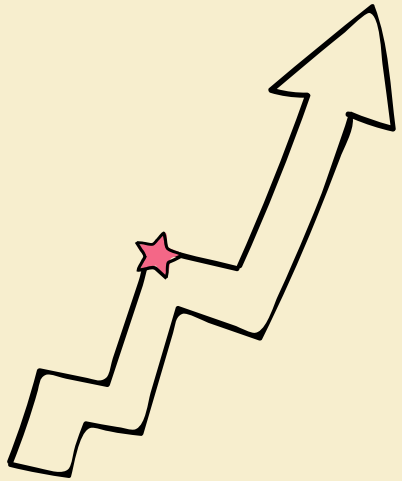
Thanksgiving Sales decomposition



Literature Review

- Wine dataset
- Amazon dataset
- Supermarket dataset

We chose these literatures as they were strongly correlated to our topic as they all are about sales and production in general and were also interesting to read.



Wine Dataset

The first literature that we found is the wine sales analysis for time series forecasting. The main goal for that project was to perform forecasting analysis on the Rose and Sparkling dataset. They tried to analyse the wine dataset by using Linear Regression, Naïve Model, Simple and Moving Average models, Simple, Double and Triple Exponential Smoothing.

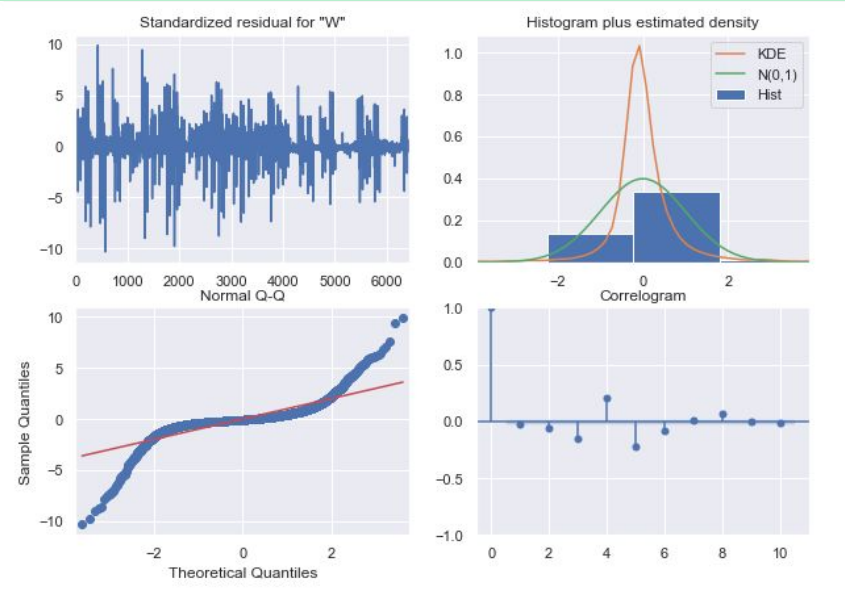
Amazon Dataset

This paper attempts to forecast future sales at Amazon.com, Inc. based on historical sales data. Firstly, it proposes three possible forecasting approaches according to the historical data pattern, that is Holt-Winters exponential smoothing, neural network auto regression model and ARIMA(Autoregressive integrated moving average). Secondly, it specifies certain accuracy measures using which well determine the suitability of the forecast methods on the available sales data. Finally the three methods will be implemented to forecast Amazons quarterly sales in 2019. The results can help Amazon well manage its future operations.

Supermarket Dataset

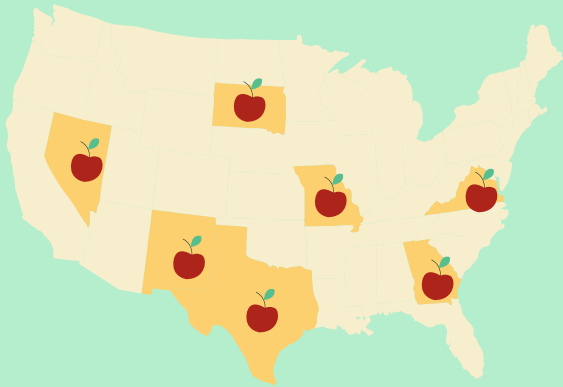
This research work has proposed a FB Prophet tool for the sales prediction of the supermarket data. The proposed research work has examined few forecasting models such as- The additive model, the Autoregressive integrated moving average (ARIMA) model, FB Prophet model. From the proposed research work, it is concluded that, FB Prophet is a better prediction model in terms of low error, better prediction, and better fitting.

Estimation I



An example plot from our paper, the residual diagnosis

During this stage we calculated ADF, KPSS, found the dependence order of our model, did residual diagnosis, Ljung-Box test, Holt Winter's test, Calculated MSE and compared with the best SARIMA model



Estimation II

	AIC	BIC	FPE	HQIC
0	32.31	32.32	1.082e+14	32.32
1	27.91	27.92	1.326e+12	27.92
2	27.77	27.78	1.145e+12	27.77
3	27.74	27.76	1.118e+12	27.75
4	27.68	27.70	1.050e+12	27.69
5	27.58	27.61	9.524e+11	27.59
6	27.55	27.57	9.184e+11	27.56
7	27.51	27.54	8.833e+11	27.52
8	27.47	27.51	8.552e+11	27.49
9	27.45	27.49	8.369e+11	27.47
10	27.45	27.50	8.365e+11	27.47
11	27.43	27.48	8.204e+11	27.45
12	27.43	27.48	8.176e+11	27.45
13	27.42	27.48*	8.104e+11	27.44
14	27.42	27.48	8.083e+11	27.44
15	27.41*	27.48	8.054e+11*	27.44*

The main point was estimating a VAR model. A VAR model is a generalization of the univariate autoregressive model for forecasting a vector of time series.It consists of one equation for each variable in the system. The right side of each equation contains a constant as well as the lags of all the variables in the system.

An example figure from this part, VAR order selection

Conclusion

By using various methods to construct different kinds of models to explore our data, we came to the conclusion that holidays are actually the cause of the trends and seasonality in our time series as during those times the sales have big shocks that is because people shop more during the holidays such as Christmas, Thanksgiving and so on. By forecasting the series we understood that Walmart company uses a great technique of keeping their holiday sales on peak as we noticed seasonality that indicates some repetitions of data over the years. There are also some peaks that are trends that may be caused because of some in-store sales and many other factors.

Thank You!

Walmart

Save money. Live better.

References

[1] Jha, B. K., & Pande, S. (2021, May 6). *Time Series Forecasting Model for Supermarket Sales using FB-Prophet*. https://ieeexplore.ieee.org. https://ieeexplore.ieee.org/abstract/document/9418033

[2] Khan, K. (2022, January 13). *Project - Time Series Forecasting - Wine Sales Analysis*. https://www.scribd.com. https://www.scribd.com/document/552600283/Time-Series-Forecasting-Project-Report

[3] Singh, B., Kumar, P., Sharma, N., & Sharma, K. P. (2020, April 20). *Sales Forecast for Amazon Sales with Time Series Modeling*. https://ieeexplore.ieee.org. https://ieeexplore.ieee.org/abstract/document/9071463