Sales Demand Prediction in retail using Time Series forecasting
Walmart Case Study

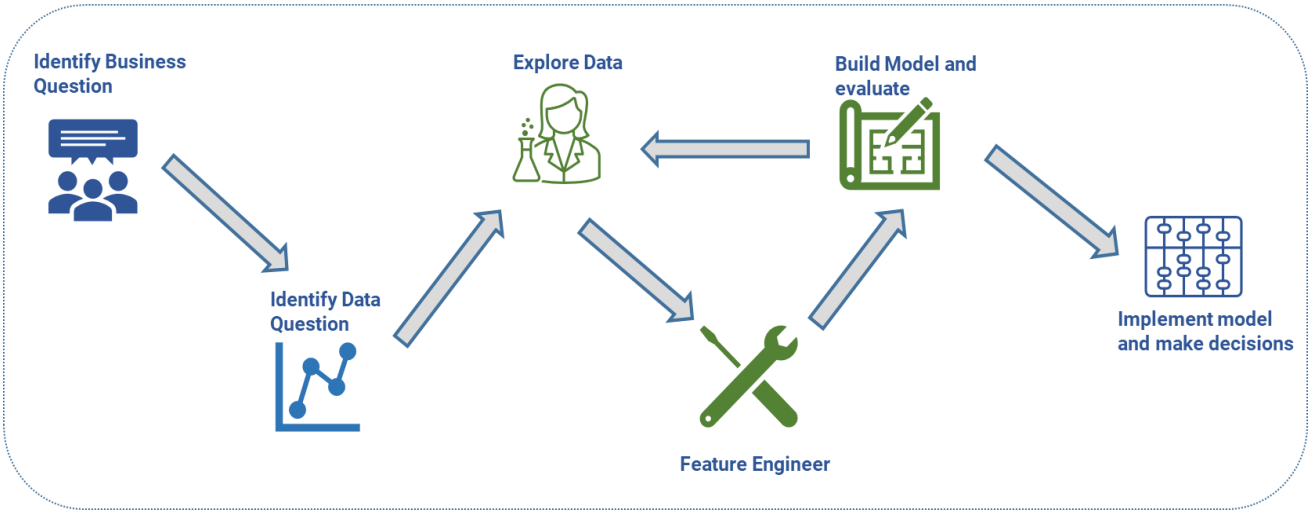
Monica Nino Joya

11th March 2021

Table of Contents

Process overview

The following diagram shows the overall end-to-end process for defining, designing and delivering the Capstone project.



Background

In retail, margins are thin, and competition is fierce. Every decision affects current and feature performance.

Demand forecasting plays a key role in driving decision making to improve customer service, productivity and help the company stay ahead of competitors.

Machine Learning Improves Demand Forecast Accuracy and brings many benefits. A few of these include the ability to automatically predict the behaviour of large portfolios, removing human bias, and using statistics to predict something that is inherently statistical in nature.

A mathematical model is able to accurately and quickly extract information from data that would not be readily evident to a human. Artificial Intelligence has become prevalent in the demand forecasting world, allowing companies to develop deep insights around what is driving their sales and what their sales will look like in the

future, by doing things like finding leading sales indicators, factoring in mergers, major market events like Covid-19, and helping determine how well a new product will perform in various markets.

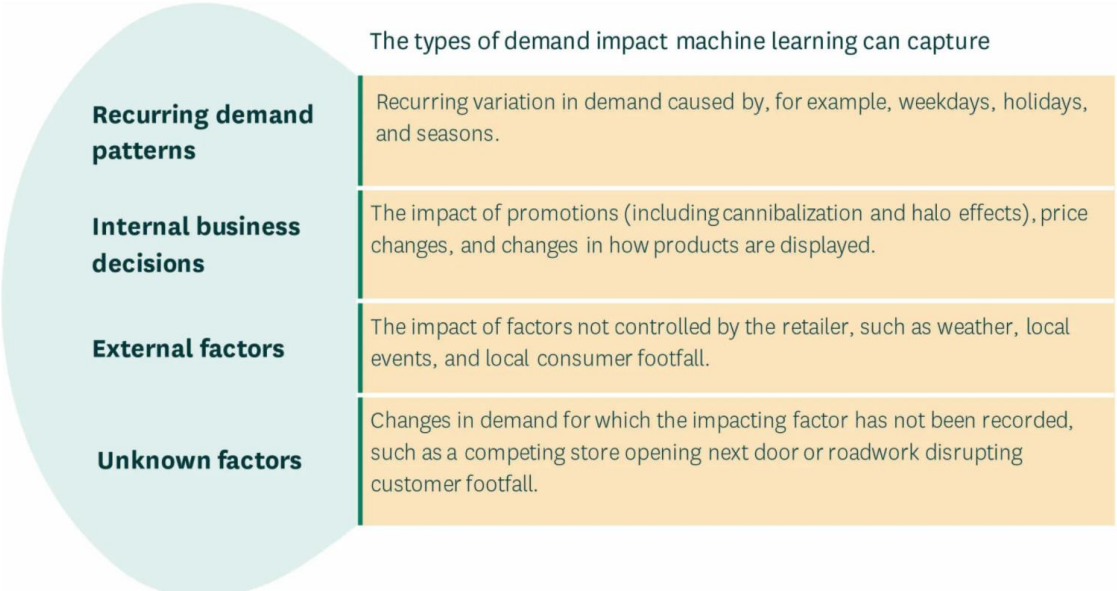


Figure 1: Types of demand machine learning can capture

Problem Statement

This project has identified three problem areas that can be solved with demand forecasting

Inventory Management: by setting Ideal levels of product on shelves will have a positive impact in customer satisfaction. In addition by accurate levels of inventory will have less stock on hand which will be reflected in lower holding cost.

Labour Productivity: optimal level of labour according to sales will improve cost of labour.

Marketing: adequate marketing and advertaising campaigns will increase sales and consequently revenue for the company

Stakeholders

From the three problem areas the following stakeholders has been identified

- Replenishment Team
- Marketing Team
- Customer Service Team
- Clients

Business question

Can we create a machine leaning model to predict sales demand for the next three months that can help to solve the three problem areas?

Data question

By using historical data 5 previous year of data, can we predict sales demand for the next three months?

Time Series

Time series Properties:

- Seasonality: Does the data display a clear periodic pattern. Shows repeated pattern over time. Periodic sign wave behaviour.
- Seasonal variation are short-term fluctuations in a time series which occur periodically in a day/week/month/year
- Trend: does the data follow a consistent upward or downward slope? Increasing or decreasing over time
- Noise/residual: are there any outlier points or missing values that are not consisting with the rest of the data?. Unexplained variance and volatility of the time series

Time series can of course changes because of external factors that are not detectable from the time series itself. Such as economic factors, political events, Covid-19, etc.

Data

The data set was acquired from Kaggle competition challenge it is base on Walmart retail case study. The data contains five years of daily data of 5 different items references in 10 different stores. A sample of the data is presented in Figure 2.

The behaviour of demand of the various items might be different in each store as they are located across different states and are impacted by different variables, therefore a model can perform different for each store. For this Project the focus is on finding the best model that will predict the demand for Store 1.

	date	store	item	sales
0	2013-01-01	1	1	13
1	2013-01-02	1	1	11
2	2013-01-03	1	1	14
3	2013-01-04	1	1	13
4	2013-01-05	1	1	10

Figure 2: Data for demand forecasting

Data science process

Data analysis

Data exploratory was done using Pandas and NumPy libraries, data types change was conducted to convert the date column from object to datetime format and this date column was set as the index for the Data Frame. As the objective is to predict the next three months of data the last three months of the current dataset were taken for testing.

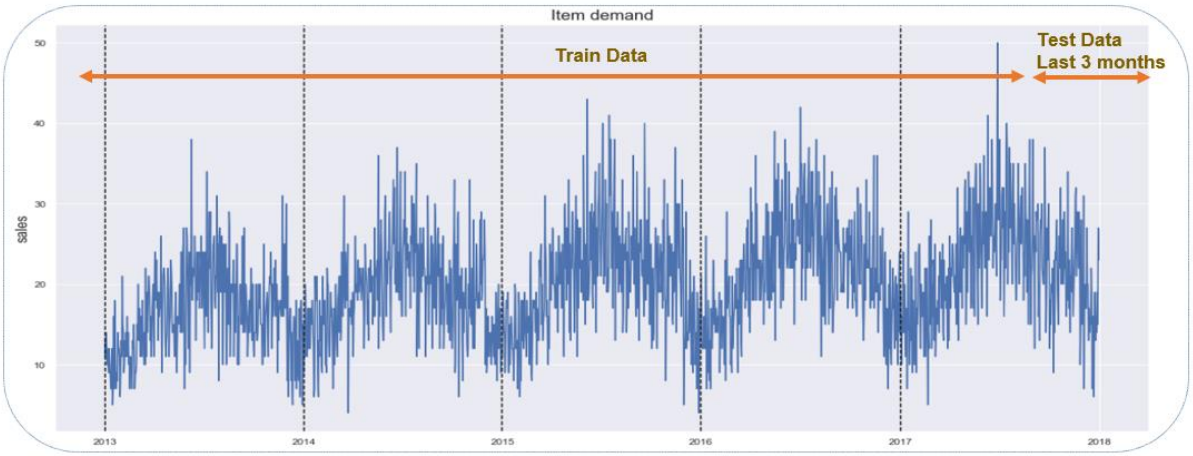


Figure 3: Time series for item 1.

A decomposition of the time series was performed (See Fig 4), as seen in the graph below there are yearly, weekly and monthly seasonal components in the time Serie that repeat periodically.

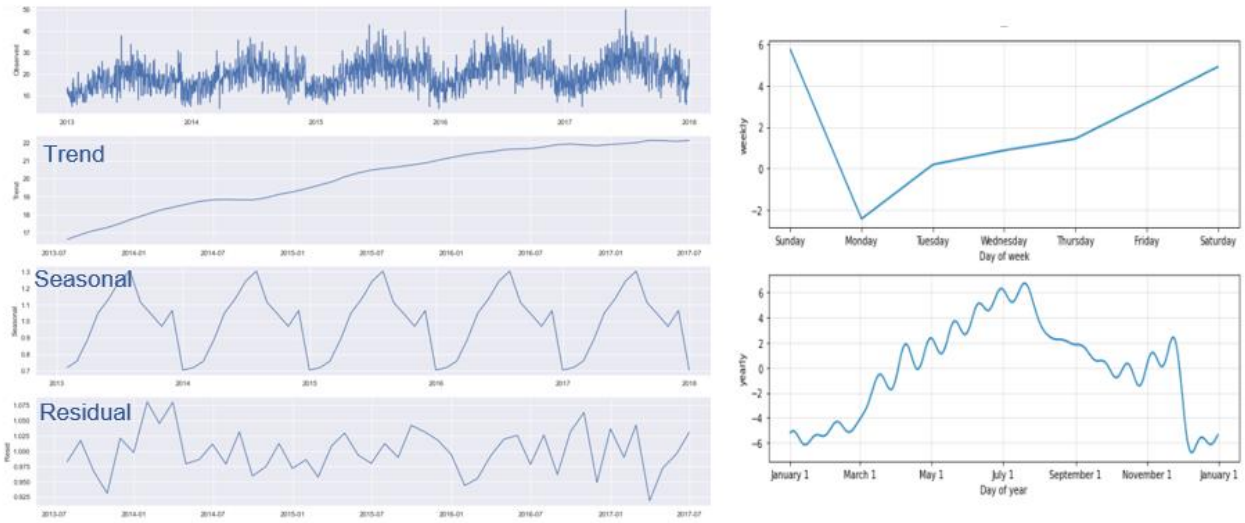


Figure 4: Time Serie Decomposition

The following graph (see Fig 5) shows the correlation for the time series observations with observations with previous time steps which are called lags. From the Autocorrelation plots we can see how seasonal component repeats every 7-time steps.

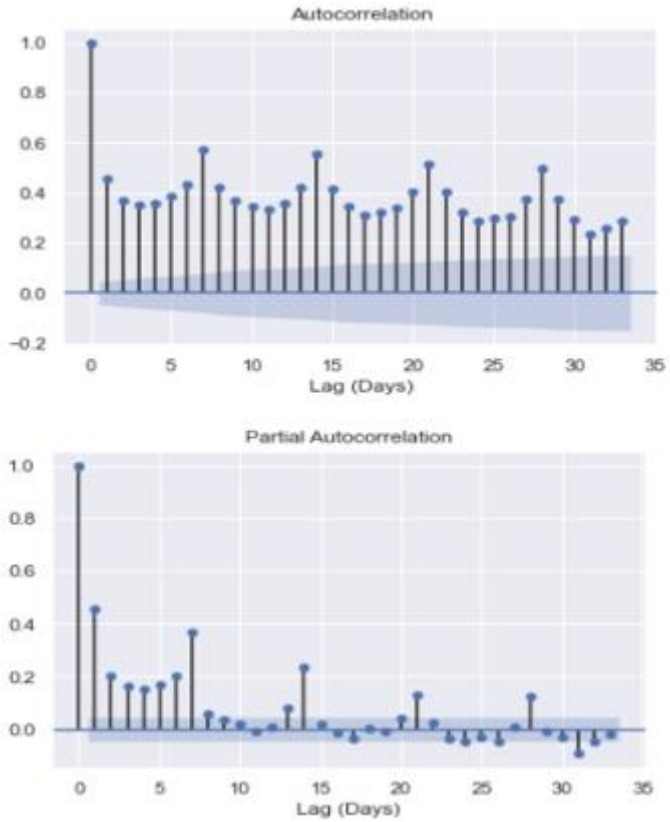


Figure 5: Autocorrelation and Partial autocorrelation

Modelling

The models that were used in this project are shown in the table below with their corresponding python packages.

MODEL	PACKAGE
ARIMA	Statsmodel
SARIMA	Statsmodel
SARIMAX	Statsmodel
PROPHET	fbprophet

TABLE 1: Models used in this project

ARIMA

ARIMA is an Auto Regressive Integrated Moving Average. This model has three parameters

p: AR(Auto Regressive). Autoregressive means that we regress the target variable on its own past values. That is, we use lagged values of the target variable as our X variables

d: Differencing: to remove trend and seasonality. Subtracting current term from the previous term. the future change in Y is a linear function of the past changes in Y.

q: MA(Moving Average).

SARIMA

Seasonal Autoregressive Integrated Moving Average, SARIMA, is an extension of ARIMA that explicitly supports univariate time series data with a seasonal component. For this project two separate models were built, one included weekly seasonality and another for yearly seasonality.

SARIMAX

Seasonal Auto Regressive Integrated Moving Averages with eXogenous regressors. Exogenous variables are parallel time series not modelled directly but used as a weighted input to the model (E.g., Additional seasonality, holydays, Events, etc.).

For the exogenous component I added yearly seasonality using furrier terms, holidays and special events. I performed a variation of this model using Walk Forward.

Prophet

Prophet is a time series forecasting model released by Facebook. This model allows to fit various seasonality patterns. This model require to input the columns in a specific way in order to work so some feature engineering needed to be done.

The model was fit with Exogenous variables such as holidays and special events effects, and added seasonality with Fourier order: yearly, weekly, quarterly, biweekly, 3weeks identifies from the Partial autocorrelation Plot.

Model Evaluation

In the figure below we can see the results from the models compared with the actual values.

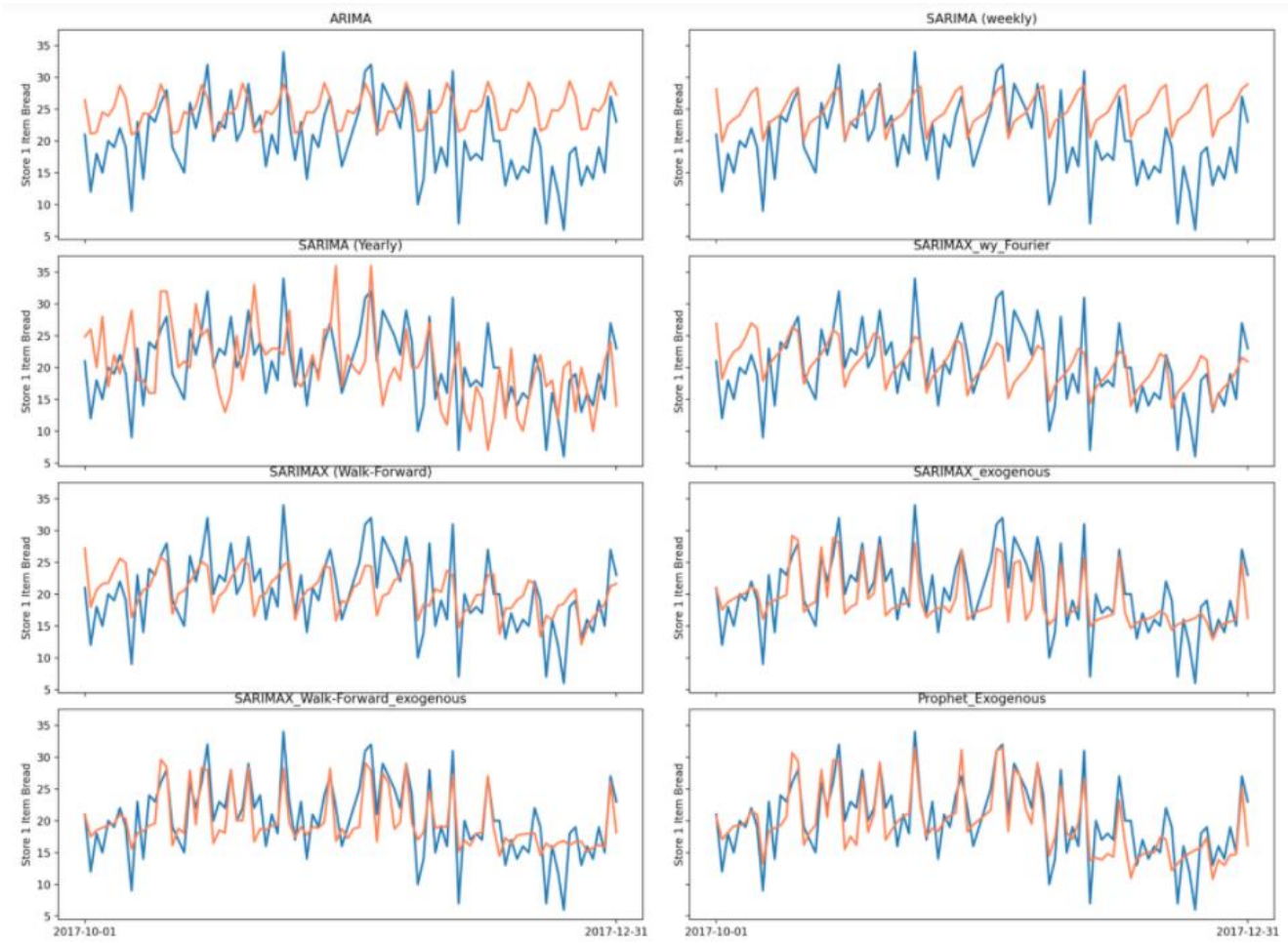


Figure 7: comparison of model outputs and actual values

In addition the performance of the model was assessed by two metrics the MAPE and RMSE.

MAPE: Mean Absolute Percentage Error express accuracy as a percentage of the error

RMSE: Root Mean Square Error is a standard deviation of the residuals which are the prediction errors.

According to these metrics the best performance model was Prophet which is showing the best result with MAPE 16.6% and RMSE 3.52 (See figure 8).

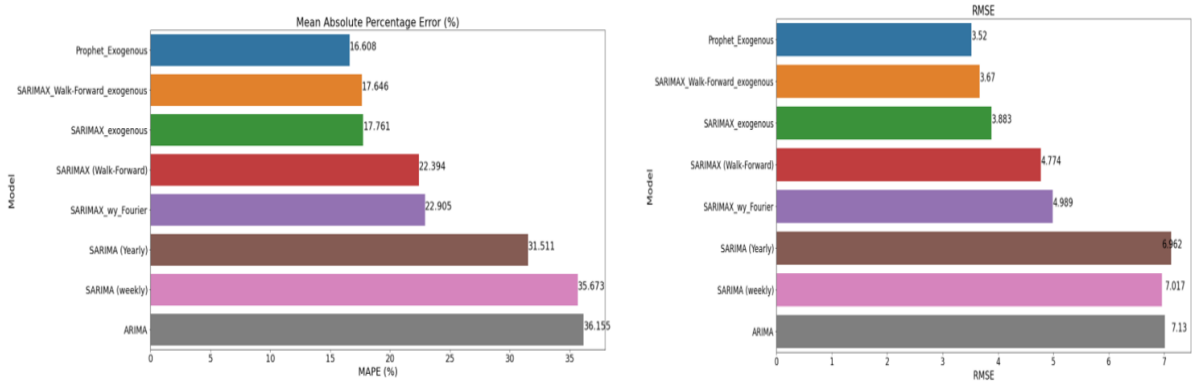


Figure 8. Metrics evaluated in models

Business answer

We can answer the business question saying Yes, by implementing a time series model using historical data, we were able to predict future demand which plays a key role in planning and optimization and helps to guide business decisions.

Conclusions and Next Steps

The Prophet model was found to perform the best forecasting for the sales in the next three months. It showed the best results with MAPE 16.6% and RMSE 3.52.

It was shown that by implementing a time series model using historical data, we were able to predict future demand which plays a key role in planning and optimization and helps to guide business decisions.

The Next steps for the project will be:

- Analyse other external factors that can influence and improve the performance of the model.
- Retrain the model with new data and automate this process to generate a weekly sales demand forecast report that will be distribute to Stakeholders identified for decision making

References

- Jupyter Notebooks
    - Capstone_EDA
    - Capstone_Models_Evaluation
    - Capstone_Arima
    - Capston_Sarima
    - Capston_Sarimax
    - Capstone_Prophet
    - Capstone_models_evaluation
- Rob J Hyndman, “Forecasting Principles and Practice” https://otexts.com/fpp2/
- ARIMA Model – Complete Guide to Time Series Forecasting in Python” https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/
- Jason Brownlee. “What Is Time Series Forecasting?” Machine Learning Mastery. Aug 15, 2020. https://machinelearningmastery.com/time-series-forecasting/
- Andrej Baranovskij, “Forecast Model Tuning with Additional Regressors in Prophet” Jul 16, 2020. https://towardsdatascience.com/forecast-model-tuning-with-additional-regressors-in-prophet-ffcbf1777dda