Sales Demand Prediction in retail using Time Series forecasting

Monica Nino Joya

11th March 2021

Agenda

1. **BIO**
2. **Project Context**
    - Industry / Domain
    - Problem area / statement
    - Stakeholders
3. **Data & Design**
    - Data Source
    - Data exploration, analysis and visualisation
    - Overall flow used
4. **Deliver**
    - Machine models
    - Model Evaluation
5. **Conclusions and next steps**

Biography

Education
- Bachelor of Industrial Engineer
- Diploma of Database Design and Development
- Project Management Certificate Program

Professional experience
- Service Analyst – Woolworths Group
- Technology Services Coordinator – Sydney Airport

Data science learnings and experience
- Data Science & AI Certificate - Current
- Python for Data Analytics
- Visualization software: Power BI, Tableau

Relevance to the project
- Experience in the retail Industry

Project Context

Background

- In retail, margins are thin, and competition is fierce. Every decision affects current and feature performance.
- Demand forecasting to improve customer service, productivity and stay ahead of competitors.

Demand Forecasting: How much of a product customers will want to purchase during time period according to historic information.

Leverage Machine Learning for Accurate Demand Forecasting

The types of demand impact machine learning can capture

Recurring demand patterns	Recurring variation in demand caused by, for example, weekdays, holidays, and seasons.
Internal business decisions	The impact of promotions (including cannibalization and halo effects), price changes, and changes in how products are displayed.
External factors	The impact of factors not controlled by the retailer, such as weather, local events, and local consumer footfall.
Unknown factors	Changes in demand for which the impacting factor has not been recorded, such as a competing store opening next door or roadwork disrupting customer footfall.

Project Context

Problem areas to solve with demand forecasting

Inventory Management
- Ideal levels of product = Happy Customers
- Less stock on hand = lower holding costs

Labour productivity
- Optimal level of staff = improve Cost of labour.

Marketing
- Plan marketing campaigns = increase revenue

Stakeholders
- Replenishment, marketing, and 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?

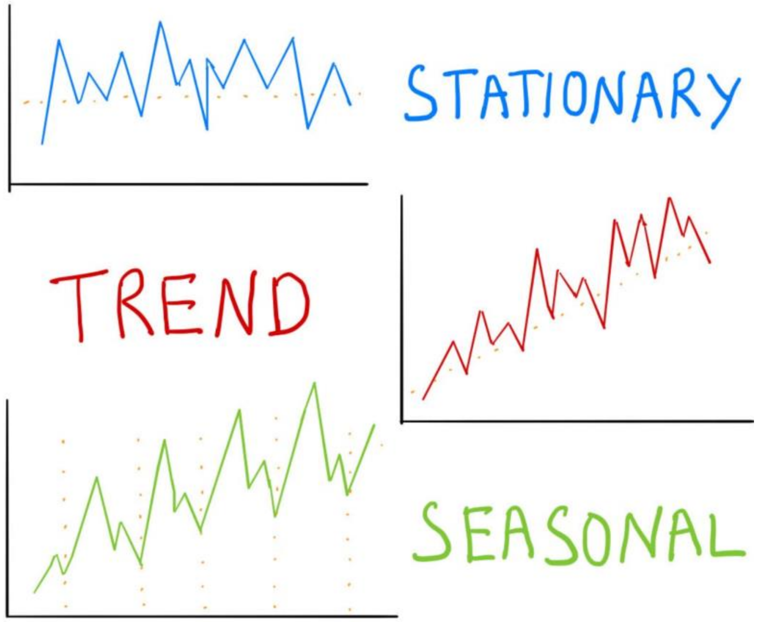
Data Question: By using historical data, can we predict sales demand for the next three months ?

What is Time Series?

A time series is a sequence of observations taken sequentially in time. The goal is to predict future, based on the past observational data.

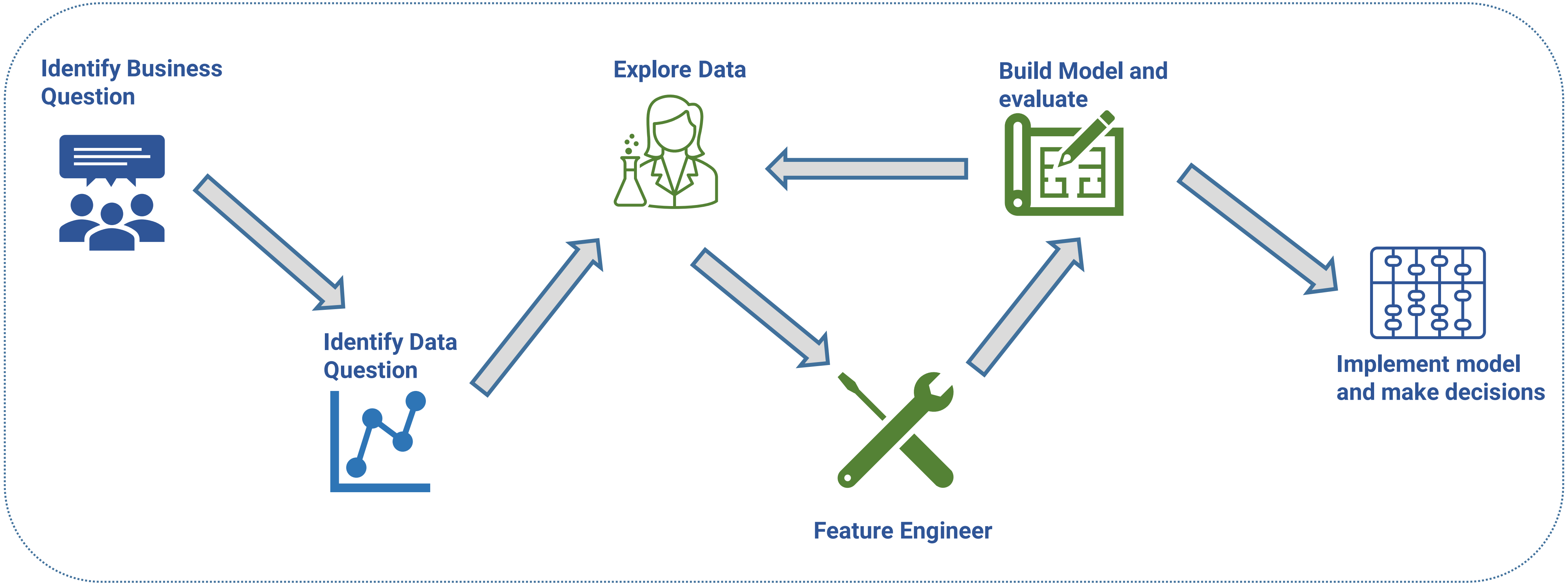
Characteristics of time series:

- **Seasonality:** Periodic fluctuations over time.
- **Trend:** data Increase or decrease over time.
- **Noise/residual:** Unexplained variance/volatility of the time series.
- **Stationarity:** Statistical properties do not change over time.
- **Autocorrelation:** Correlation between a time series and a delayed copy of itself.



Note: Model them well to obtain accurate forecast.

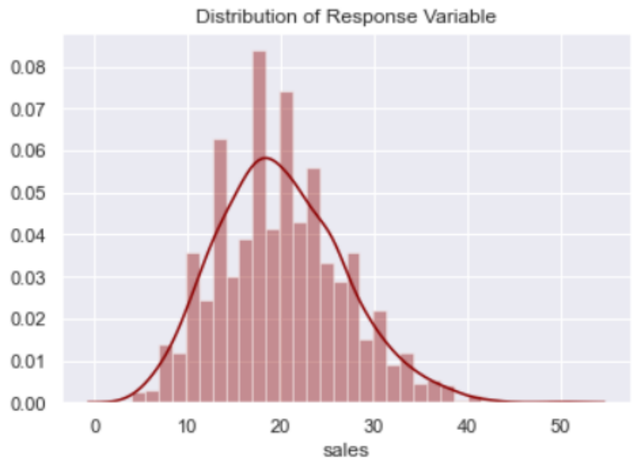
Overall Process Flow Used



DATA

- Walmart Case Study acquired from Kaggle Forecasting Challenge.
- Dataset contains 5 years of daily sales volumes. Predict 3 months.

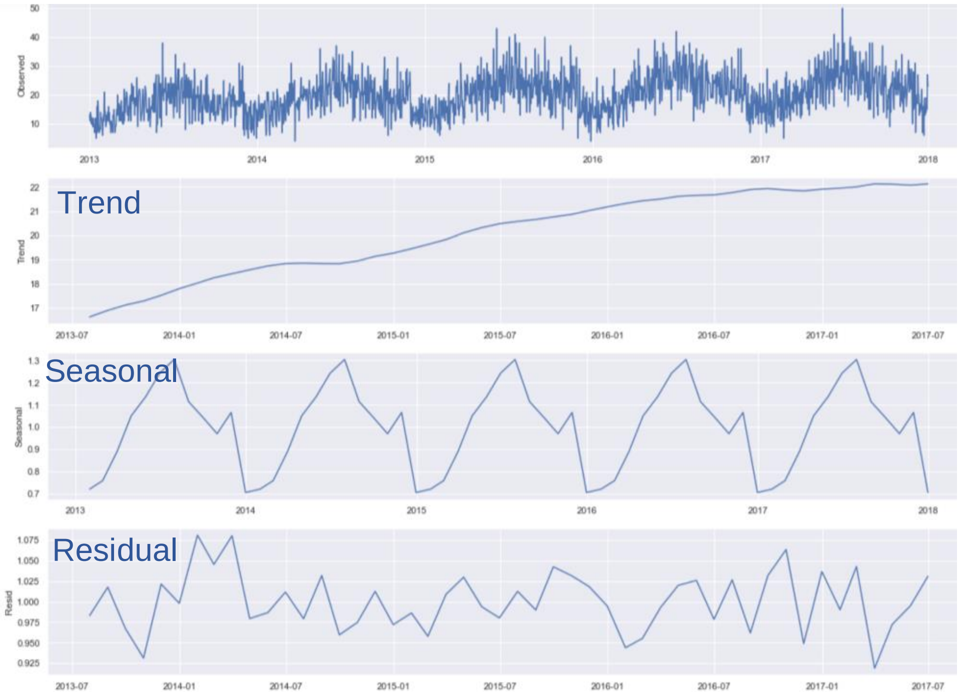
	date	store	item	sales
0	2013-01-01	1	1	13
1	2013-01-02	1	1	11
2	2013-01-03	1	1	14
3	2013-01-04	1	1	13
4	2013-01-05	1	1	10

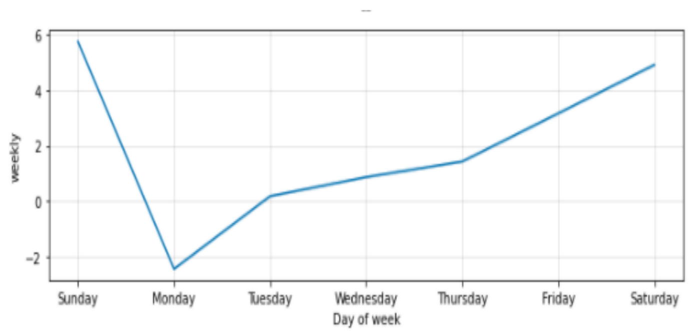


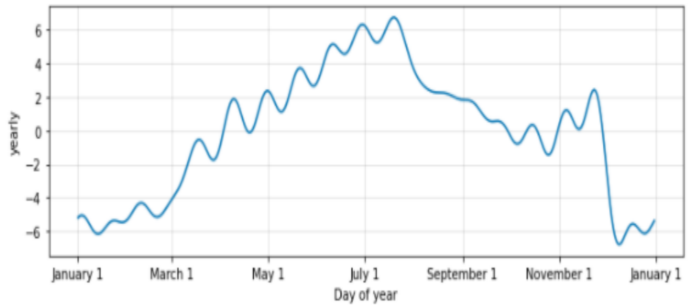
Explore Data

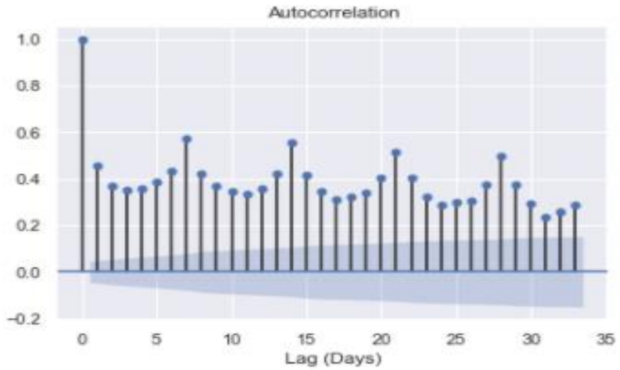
Analysis of Time series Characteristics

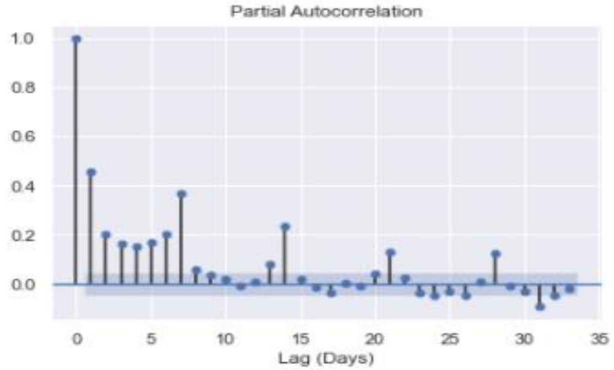
- Decomposed results: Weekly, yearly, monthly for seasonal pattern.
- Sales peak: Jun to July and Dec decrease Sep-November.
- Sales are slightly increasing over the years.
- From ACF seasonal component repeats every 7-time steps.











Models Used for Time Series Forecasting

ARIMA

SARIMA

SARIMAX

Prophet

Models Used

ARIMA

Auto Regressive Integrated Moving Average model

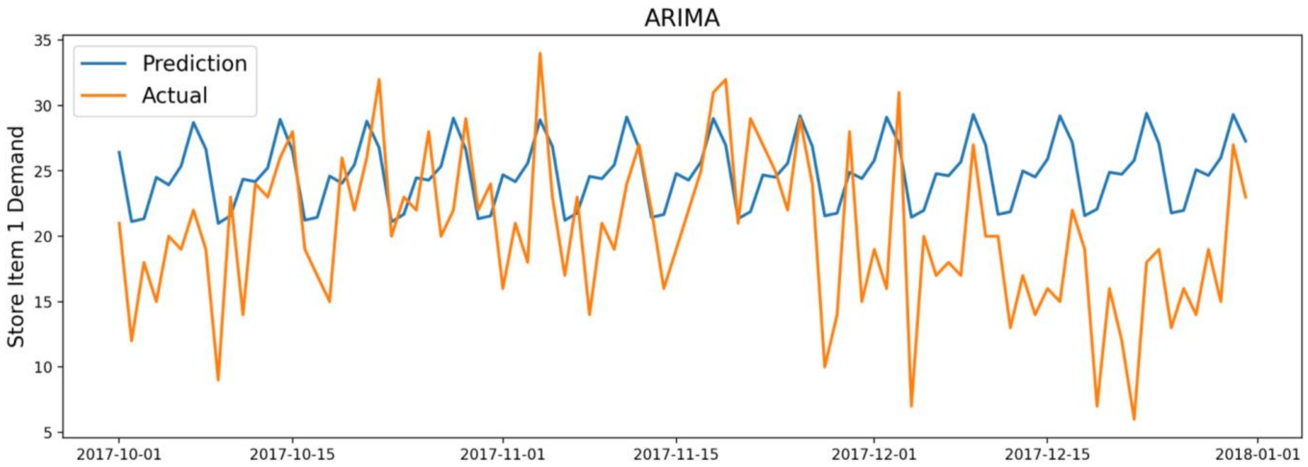
Parameters:

- p= AR(Auto Regressive). use lagged values of the target variable as X variables
- d: Differencing to remove trend and seasonality
- q= MA(Moving Average).
- From autocorrelation 1st guess (1,1,0) and then lower AIC (6,1,1)

AIC: For model selection. Estimator of prediction error

Using ACF and PACF to choose model order

	AR(p)	MA(q)	ARMA(p,q)
ACF	Tails off	Cuts off after lag q	Tails off
PACF	Cuts off after lag p	Tails off	Tails off



Models Used

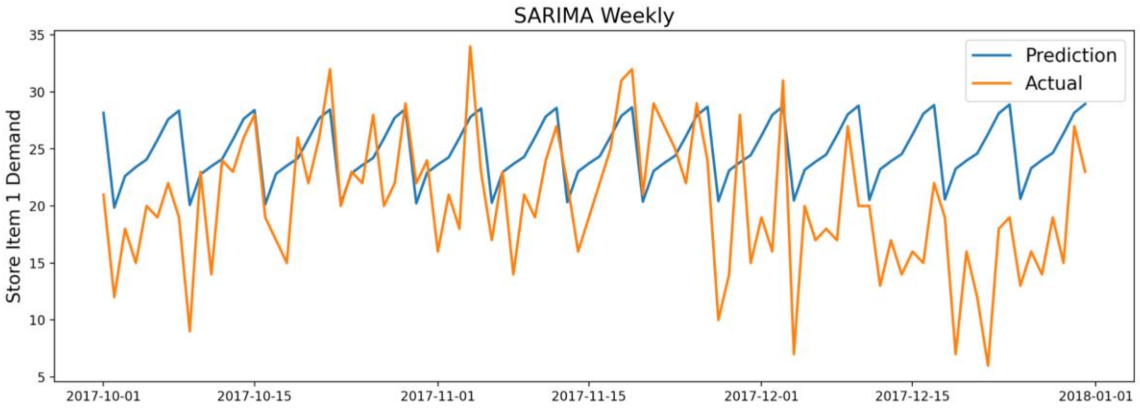
SARIMA

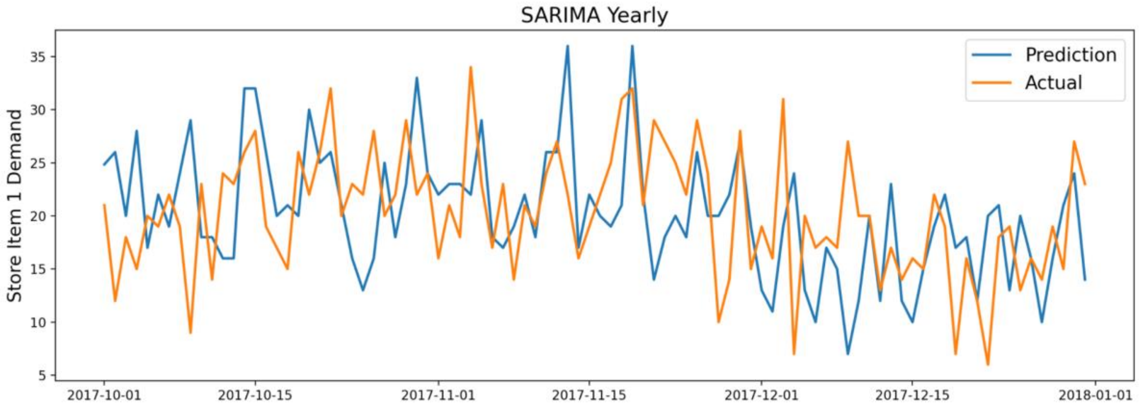
- ARIMA + Seasonal component
- pmdarima for hyperparameter optimization (order = (4,1,1), seasonal order = (1,0,2,7))
- Two variation: Yearly & Weekly seasonality

Dep. Variable:	y	No. Observations:	1734
Model:	SARIMAX(4, 1, 1)x(1, 0, [1, 2], 7)	Log Likelihood	-5168.898
Date:	Wed, 24 Feb 2021	AIC	10357.796
Time:	17:50:39	BIC	10412.372
Sample:	0	HQIC	10377.980
	- 1734		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
intercept	9.856e-05	0.001	0.144	0.886	-0.001	0.001
ar.L1	0.0914	0.027	3.400	0.001	0.039	0.144
ar.L2	0.0191	0.027	0.707	0.480	-0.034	0.072
ar.L3	0.0405	0.026	1.565	0.118	-0.010	0.091
ar.L4	0.0569	0.026	2.209	0.027	0.006	0.107
ma.L1	-0.9561	0.009	-101.765	0.000	-0.975	-0.938
ar.S.L7	0.9850	0.007	134.154	0.000	0.971	0.999
ma.S.L7	-0.8307	0.026	-31.610	0.000	-0.882	-0.779
ma.S.L14	-0.0252	0.025	-1.015	0.310	-0.074	0.023
sigma2	23.2869	0.764	30.471	0.000	21.789	24.785

Ljung-Box (Q):	61.58	Jarque-Bera (JB):	12.37
Prob(Q):	0.02	Prob(JB):	0.00
Heteroskedasticity (H):	1.32	Skew:	0.12
Prob(H) (two-sided):	0.00	Kurtosis:	3.34





Models Used

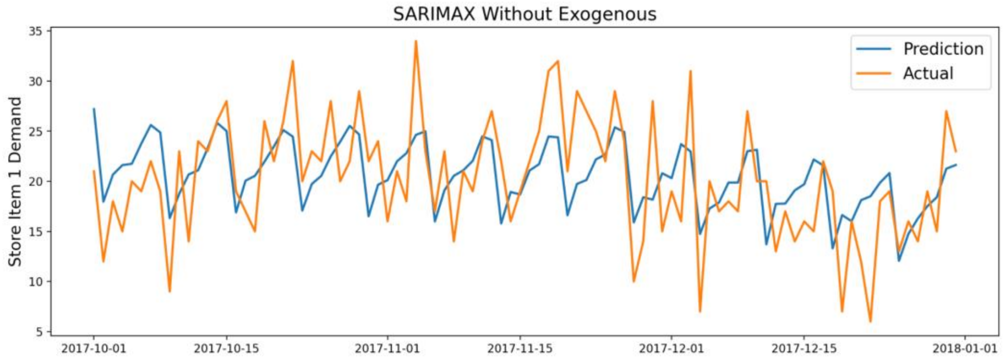
Exogenous variable: Parallel time series not modelled directly but used as a weighted input to the model (E.g., Additional seasonality, holydays, Events..)

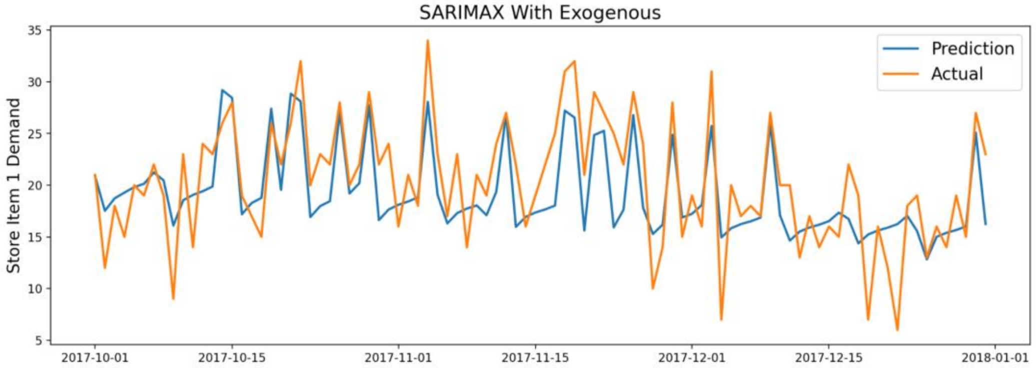
SARIMAX

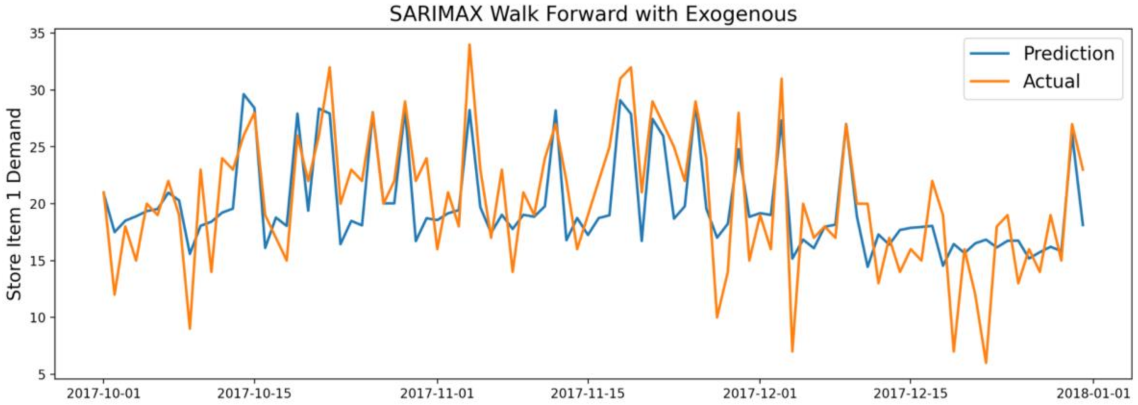
SARIMA with Exogenous Variables:

- **SARIMA +**
- **Weekly seasonal component + exogenous variables (Yearly(Fourier terms), holidays, event)**
- **Variation: Walk forward validation**

Fourier series is a way of representing a periodic function as a sum of sine and cosine functions







Models Used

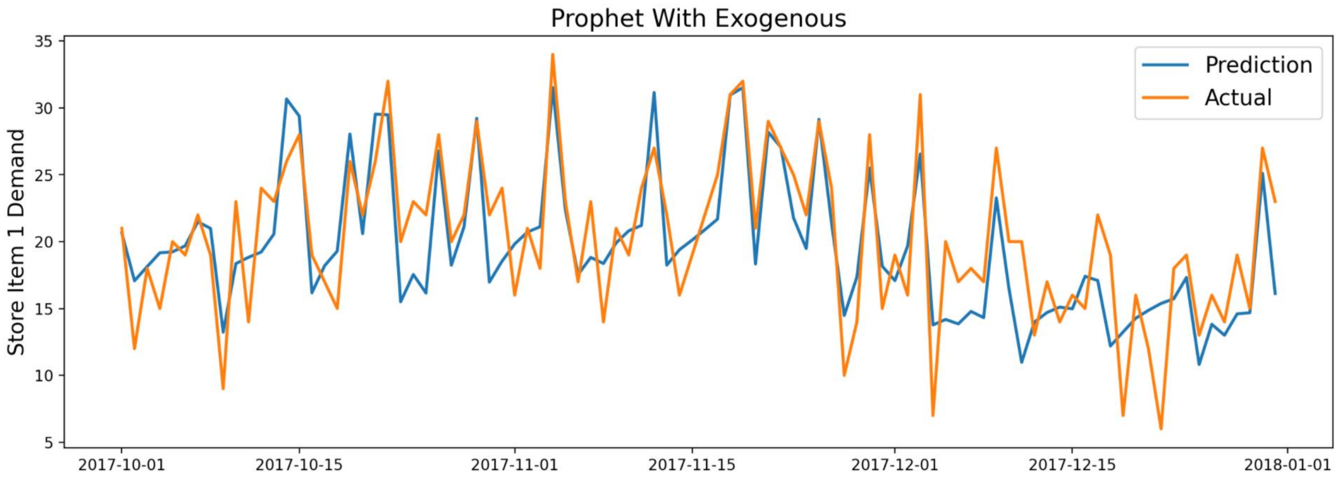
Prophet

Prophet is released by Facebook.

- Built with multiplicative decomposition

- Fit with Exogenous variables (holidays and special events effects).

- Added seasonality with Fourier order: yearly, weekly, quarterly, biweekly, 3weeks(PACF Plot)



Models Evaluation- Metrics

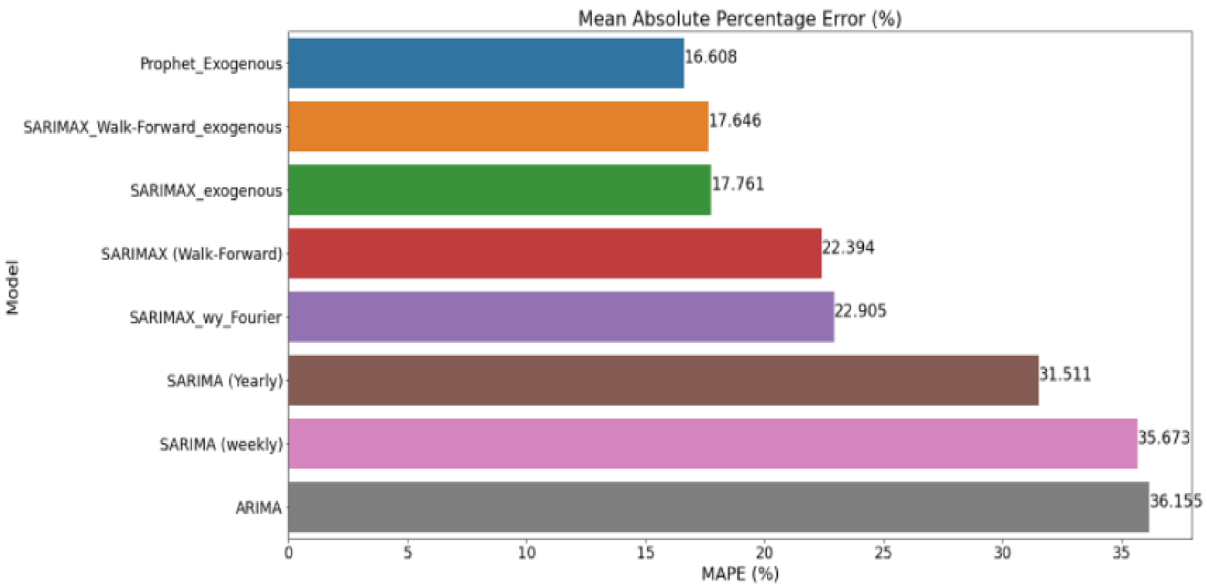
MAPE (Mean absolute percentage error): Express accuracy as a percentage of the error

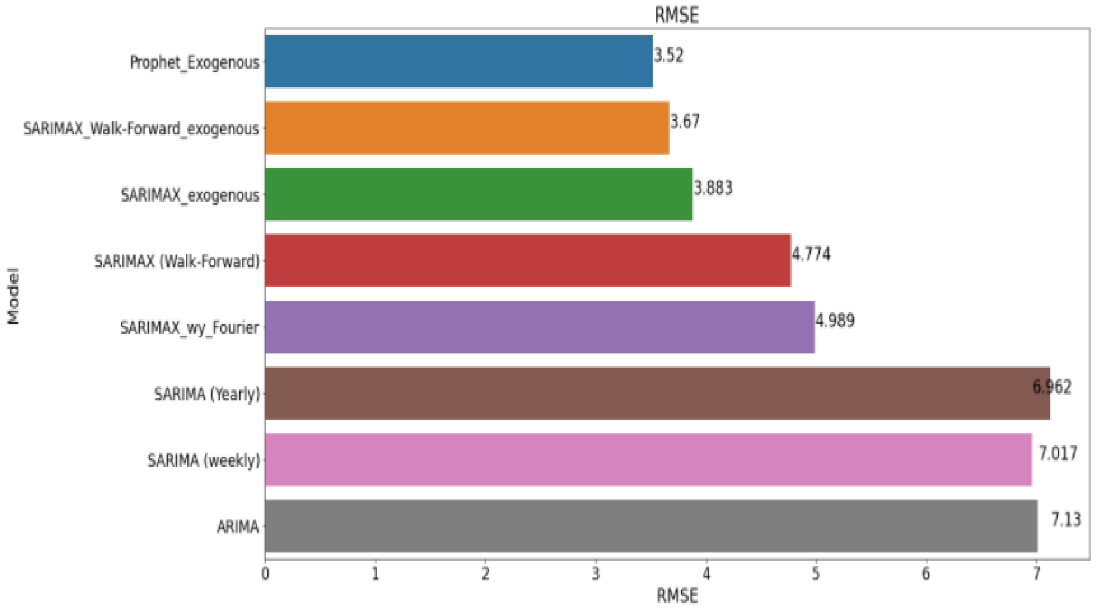
$$\frac{1}{N}\sum_{t=1}^{N}\frac{ABS(Actual_t - Forecast_t)}{Actual_t} * 100\%$$

RMSE (Root Mean Square Error): Standard deviation of the residuals (prediction errors).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(Predicted_i - Actual_i)^2}{N}}$$

Best Model:
Prophet is showing the best results with MAPE 16.6% and RMSE 3.52





Conclusions and Next Steps

Conclusions

- Best performance model for this case study is Prophet with MAPE 16.6% and RMSE 3.52.
- Exogenous variables incorporated in the model increased performance and forecasting accuracy.
- 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.

Next steps

- 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.

Questions

References

- 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

THANK YOU!