Forecasting Furniture Sales using SARIMA Model

Introduction

The objective of this project is to forecast the sales of furniture using time series analysis. The data used for this project is monthly furniture sales data from January 2014 to December 2017. The project aims to predict future furniture sales using the SARIMA model.

Methodology

The SARIMA model is a popular time series forecasting technique that incorporates seasonality and autocorrelation. The model is an extension of the ARIMA model and involves three main components: seasonality (S), autoregression (AR), and moving average (MA). The model is denoted as SARIMA (p, d, q) x (P, D, Q) s, where p, d, and q are the order of the autoregressive, differencing, and moving average terms, respectively, and P, D, and Q are the seasonal components. The s denotes the seasonal period.

The first step is to go through our dataset and determine which columns are required. The dataset has been acquired from the Tableau community source. It contains the csv file for superstore sales data and has various columns about quantities, prices and various products. Our aim is to extract the sales of furniture category and the specific dates of sales.

We read the data in Python as a Pandas data-frame and keep the sales for furniture as our target variable and the respective date of sales as the index. We resample the data to make the data monthly by taking monthly average sales. Average is used for resampling as it gives a good measure of central tendency for that month and it provides a summary of the data that is representative of the entire month's samples, and it tends to smooth out any noise or outliers that might be present in the data.

We plot the resultant time series data and also plot the decomposition of the time series. The decomposition shows different components of our time series such as trend, seasonality and residuals. We observe that furniture sales increase towards the end of the year and decrease drastically after the start of the next year. This shows that our data has strong yearly seasonality and hence we require SARIMA model to capture our data's seasonality.

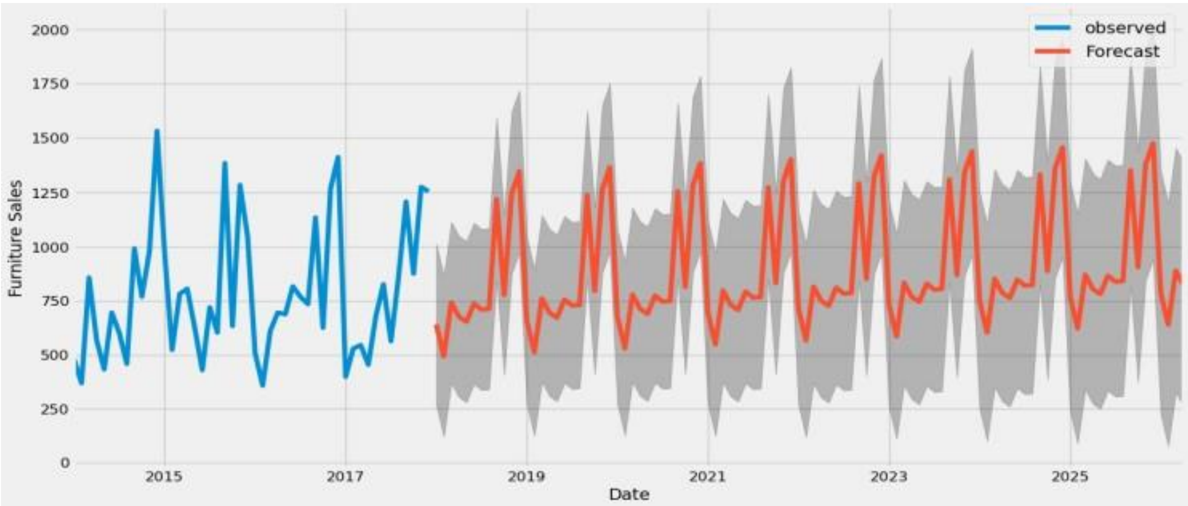
We check the stationarity of our time series before modelling using ADF (Augmented Dickey Fuller) test. We find that our data is stationary and hence, we proceed and plot the ACF (Auto-Correlation Function) and PACF (Partial Auto-Correlation Function) plots. The ACF and PACF plots show a sinusoidal pattern and significant value at lag 12. This further shows that our time series has seasonality over period of every 12 months.

The next step is to find appropriate parameters for our SARIMA model. We know that s = 12, due to the yearly seasonality present in our data. For the rest of the parameter values p, d, q and P, D, Q, we do a grid search over values 0 and 1 for each parameter. We create a SARIMA model for each iteration of the grid search and store the respective parameter values and AIC value of the model in a data-frame.

AIC (Akaike Information Criterion) value is a measure of the relative quality of a statistical model. It estimates the quality of a model by measuring the trade-off between the model's goodness of fit and its complexity. We pick the model with the lowest AIC value and use it for forecasting furniture sales. The resulting best model was SARIMA (0,1,1) x (0,1,1)12 with AIC value of 482. We plot the diagnostics of this model to check if the errors are normally distributed. The errors are almost normally distributed which means our model is decent.

We use this SARIMA model to give us predictions for 2017 and we validate the predicted values with the actual values to evaluate the accuracy of our model. The evaluation metric used is RMSE (Root Mean Squared Error) which comes out to be 185.37

We finally plot the forecasts for the next 100 months and it is clearly observed that the SARIMA model has captured the seasonality of our time series data.



Conclusion

The SARIMA model was used to forecast furniture sales data from January 2021 to December 2017. The best-fit model was found to be SARIMA (0,1,1) x (0,1,1)12. The model's performance was evaluated using the RMSE metric, and the results indicated that the model's performance was good. The model can be used to forecast future furniture sales, which can help furniture retailers and manufacturers plan their operations and inventory management more efficiently.