D213 Time Series

The purpose of this Time Series model is to be able to predict revenue from the first years of operation. The model used is an ARIMA Time Series model. The model is ideal since the data provided is labeled contains dates and revenue in millions. One research question that is relevant to a real-world situation is to predict the revenue for a set number of dates. The object of the analysis is to determine whether future dates would see an increase of decrease in revenue. This is important as organizations would need this information to determine decisions for the predicted dates. Real life scenarios show that forecast is used to determine whether organizations should increase or decrease employees in order to keep stakeholders happy with revenue reports.

ARIMA model work on the assumption of stationarity which mean they must have a constant variance and [mean](https://www.statisticshowto.com/mean/) (Glen, 2016). If your model is non-stationary, you’ll need to transform it before you can use ARIMA. Autocorrelation refers to how correlated a time series is with its past values whereas the ACF is the plot used to see the correlation between the points, up to and including the lag unit. In ACF, the correlation coefficient is in the x-axis whereas the number of lags is shown in the y-axis (shanan, 2018).

The data preparation phase consisted of transforming the data and converting columns to specific types. Attached is a line graph visualization of the time series:

Chart, scatter chart

Description automatically generated

The formatting steps required for the model consisted of converting the column of days into an actual date. The ARIMA model required that change as it would only work with date formats. Revenue also needed to be converted to a float type object. There were no gaps in the length of the sequence. The adfuller test was performed to determine if our data had stationarity in it. Once performed, the p-value was -2.00757342e which means that our data was not stationary. Since our data was not stationary, we performed the diff() function on our data to calculate the difference and make our data stationary (DataCamp, 2020). Once the data was transformed then the train\_test\_split function was called and 75% of the data was chosen for training purposes and 25% was used for testing purposes.

Analyzing the time series dataset shows us that there is a presence of seasonality in the data. The trend appears to be positive throughout the years. Attached is a screenshot for reference:

Graphical user interface, application, Excel

Description automatically generated

The auto correlation function shows us how each observation is positively associated with its recent past. Attached is a screenshot of the autocorrelation plot: Chart, bar chart, histogram

Description automatically generated

The data was decomposed and the diff() function was applied to our original dataset attached is a screenshot of the dataset looked like once the function was applied: Chart

Description automatically generated

The dataset was then checked for seasonality and the screenshot attached now shows there is no seasonality in the dataset: Timeline

Description automatically generated

A model was created and ran to fit the dataset. The dataset was also forecasted to identify whether there would be a positive or negative trend. Attached is the output of the calculation provided by the model.

Table

Description automatically generated

The model selected was an ARIMA model. The best parameters chosen for the p,d,q order was 1,0,0. The prediction interval chosen was for 30 days. This was to see if there was an increase or decrease in the forecast. Attached is a screenshot where the model shows the actual values which are in blue compared to the forecasted values which are in red. Chart, scatter chart

Description automatically generated.

The best course of action is to continue the process that is being performed by leadership. The forecasted values show a trend upward.

# Bibliography

DataCamp (Director). (2020). *Python Tutorial: Making time series stationary* [Motion Picture].

Glen, S. (2016, February 25). *ARIMA (Box-Jenkins Models): Autoregressive Integrated Moving Average*. Retrieved from Statistics How To: https://www.statisticshowto.com/arima/#:~:text=with%20computer%20software.-,Assumptions,before%20you%20can%20use%20ARIMA.

shanan, S. (2018, October 3). *Time series Forecasting — ARIMA models*. Retrieved from Towards Data Science: https://towardsdatascience.com/time-series-forecasting-arima-models-7f221e9eee06