**Time Series Lab 2**

Create Time Based Plot and Adjust for Seasonality

Text

Description automatically generated

As seen above, a time based plot has been created that takes into account the monthly increments (seasonality) that starts in 1940. An augmented dickey fuller test is performed to determine stationarity. In this instance, because the p value is 1% so the null hypothesis is rejected and the data is stationary.

Decompose Model

Graphical user interface

Description automatically generated

Here in the decompose portion of the Rmarkdown, three different observations can be found in the time series; trend, seasonal, and random. AS shown, the observations between 1940 and 2020 show the rain statistics in the City of San Diego. The trend is obtained by observing the moving averages covering one year. Next, seasonality is gathered by averaging the de-trended values for the time period.

Model Output – exponential smoothing

Graphical user interface

Description automatically generated

Graphical user interface

Description automatically generated

As seen above that as the time series prediction model predicts, the accuracy of the prediction shown in the ACF lag improves greatly from the initial prediction. The tapering pattern in the lag’s shows the autocorrelation to be positive. As shown above in the holt-winters exponential smoothing formula takes in historical plot predictions and gauges them to the predicted model. As the model continues, the smoothing increases and the prediction model improves.

Model Output – exponential smoothing - residuals

Chart

Description automatically generated

Graphical user interface, histogram

Description automatically generated

Chart, histogram

Description automatically generated

With initial observation of the ACF line, there are three spikes above the blue line which indicate these three are significant. These spikes are initial spikes close to lag0 and have three spikes that decrease as it approaches lag 0.01. As with the observation in the PACF chart where there is one initial spike and no other significant spikes, then the ACF chart is the better chart to use. As seen below in the forecasts model, there are a couple of huge spikes in the model around the early 1940’s and ~2007 which would have indicated a large weather event that created the prediction model to increase substantially. Overall the model predicts less rain in the future coming years.

Timeline

Description automatically generated

Model Output – auto.Arima and residuals

Chart, histogram

Description automatically generatedAs shown with the ACF and PCF models below, the ACF model shows the better prediction of the arima model versus the PACF.

Text

Description automatically generated

Forecasts

As seen below in the prediction forecast model, the model is predicting low likelihood of rain to occur over the next 500 months. As shown the likelihood that it rains is very low which makes sense as San Diego it hardly ever rains.

Chart

Description automatically generated

Throughout the exercise, I ran a mean absolute error on each analysis and the following MAE’s were given.

|  |  |
| --- | --- |
| residualsHolt1 | .05528824 |
| residualsHolt2 | .05041531 |
| RainForecast1 | .04399371 |
| RainForecast2 | .04442833 |
| RainArima | .04352928 |
| RainArima2 | .04218151 |
| RainArima3 | .04213866 |

With these numbers above, you can see that the RainArima3 model gave the lowest mean absolute error which would be the best model overall to use for predictability. Given that each of the prediction models have a MAE close to 0 this indicates that the prediction model is very accurate and does not need any improvement.