# CP2403 - Project – Part 2 – Time Series Analysis

First Name: Caleb

Last Name: Webster

|  |
| --- |
| **1: Time Series Plot** |
| Figure 1: Line plot of raw time series data. The lottery values used for this time series analysis were the “Pulltab” column. At first glance, the time series does not appear to be very consistent or stationary. |
| **2: Box Plot** |
| Figure 2: Box plot of time series with values grouped by month. This plot shows that the deviation from the mean is fairly large for most months, which could indicate trends and inconsistencies in the time series. |
| **3: Rolling Mean & Standard Deviation Plot of Time Series** |
| Figure 3: Rolling mean and standard deviation plot. In a stationary time series, the rolling mean should not vary, and the standard deviation should resemble a consistent periodic shape. The shape of the red line reveals that this time series may not be stationary. |
| **4: Results of Dickey-Fuller Test of Time Series** |
| Results of Dickey-Fuller Test:  Test Statistic -2.387211  p-value 0.145379  #Lags Used 4.000000  Number of Observations Used 178.000000  Critical Value (1%) -3.467632  Critical Value (5%) -2.877918  Critical Value (10%) -2.575501  H0: The time series is non-stationary  Ha: The time series is stationary  The Dickey-Fuller stationarity test produces a p-value of 0.145379, which is > 0.05. This indicates that the “Pulltab” time series is non-stationary. |
| **5: Log Plot of Time Series** |
| Figure 4: Log plot of “Pulltab” time series. It appears to closely resemble the raw data, but the variance has been stabilized. |
| **6: Seasonal Decompose plot of Time Series** |
| Figure 5: Decomposition of time series. From these graphs, we can note a decreasing then increasing trend, fairly consistent seasonality, and residuals that resemble the original data points. |
| **7: Rolling Mean & Standard Deviation Plot of Residual** |
| Figure 6: Rolling mean and standard deviation plot of “Residual” part of decomposed time series. The rolling mean and standard deviation fluctuate, but there appears to be less of a significant trend than was present in the raw data. |
| **8: Results of Dickey-Fuller Test of Residual** |
| Results of Dickey-Fuller Test:  Test Statistic -5.602576  p-value 0.000001  #Lags Used 1.000000  Number of Observations Used 129.000000  Critical Value (1%) -3.482088  Critical Value (5%) -2.884219  Critical Value (10%) -2.578864  H0: The time series is non-stationary  Ha: The time series is stationary  The Dickey-Fuller stationarity test produces a p-value of 0.000001, which is < 0.05. This indicates that the Residual of the “Pulltab” time series is stationary. |
| **9: ACF and PACF plot** |
| Figure 7: Autocorrelation and Partial Autocorrelation functions. The p and q values for the ARIMA model were obtained from these graphs. On each of the plots, the value where the line crosses the upper confidence interval is closest to 1, so we set p and q to 1. |
| **10: p, d and q values** |
| P = 1  D = 1  Q = 1 |
| **11: ARIMA Plot** |
| Figure 8: Plot of ARIMA prediction model. The predicted values (red) are extremely close to the actual values (blue), indicating a good fit for the model on the training data. |
| **12: Prediction Plot** |
| Figure 9: Plot of predictions vs actual time series. The predictions appear to reflect the actual values well at the start and end of the time series, but wander in the middle, spiking to an extreme height. This model can be used to predict the future sales of the Iowa lottery to inform business decisions for the institution that owns this data. |