# CP2403 - Project – Part 2 – Time Series Analysis

First Name: Kaung Khant

Last Name: Naing

|  |
| --- |
| **1: Time Series Plot** |
| It can be seen from the data above that the sales of pulltab is quite stable with a significant drop around 2017-01, however it start to increase and stabilize later on and has an extreme rise which is the highest sales in around 2018-05 then drop back in around 2018-06. |
| **2: Box Plot** |
| In the box plot for Iowa Weekly lottery sales for pulltab, the first week ending date is higher than the rest of the time only closely followed by the fifth week ending date. Moreover, it can be concluded that the median for week ending 5 and 8 is the highest out of all 12 week ending dates, despite that all week ending date has small difference of median with each other. Furthermore, in the fourth, eleventh, and twelfth week ending date, the data consist of outliers which can affect the whole data. |
| **3: Rolling Mean & Standard Deviation Plot of Time Series** |
| Based on the data it can be concluded that the blue line represents original data whereas the red one represents rolling mean and lastly the black one represents rolling standard deviation |
| **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 |
| **5: Log Plot of Time Series** |
|  |
| **6: Seasonal Decompose plot of Time Series** |
|  |
| **7: Rolling Mean & Standard Deviation Plot of Residual** |
| Based on the data it can be concluded that the blue line represents original data whereas the red one represents rolling mean and lastly the black one represents rolling standard deviation |
| **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  It can be concluded from the Dickey-Fuller test that the test statistic value is -5.6, which is less than the critical value which mean the null hypothesis is rejected therefore the data of this timeseries is stationary. Moreover, it can be seen that after decomposing for one time, the p-value is now below 0.05. |
| **9: ACF and PACF plot** |
|  |
| **10: p, d and q values** |
| **p = 1**  **q = 1**  **d = 1**  **the two dotted lines on either sides of 0 are the confidence intervals.**  **p = is where the PACF chart crosses the upper confidence interval for the first time. in this case p=1.**  **q = is where the ACF chart crosses the upper confidence interval for the first time. in this case q=1.**  **The number of differencing to make the time series is 1, thus the d- value is 1** |
| **11: ARIMA Plot** |
|  |
| **12: Prediction Plot** |
| Based on the data above the orange line represents the prediction whereas the blue represents the actual data |