**Assignment 2**

**Uchenna Ilodigwe**

**Big Data Solutions Architecture, Conestoga College**

**Prog8430-Data Analysis, Mathematics ,Algorithms**

**Dr. David Marsh**

**October 12,2021**

**Assignment 2 - Time Series Analysis**

**Author: Uchenna Ilodigwe**

**Course:** PROG8430

**Background**

Given two datasets, Woodstock\_21F.Rdata and Ayr\_21F.Rdata which contain average temperature over different periods for the cities of Woodstock and Ayr respectively, we are to use time series analysis to describe and forecast their temperature

**Section 1.1 Data Transformation and Cleaning (Description)**

We create a dataframe called TempStudy\_UI , identify it as a ts(timeseries) data, identify the dataframe we are taking it from which is the dataframe which was loaded from the Woodstock data file(Temperature\_UI) .The frequency is 12 i.e. it cycles through 12 since there are 12 months in a year. We identify when the data starts i.e. what time period is associated with the first piece of data.

Temperature\_UI <-get(load("Woodstock\_21F.Rdata"))

TempStudy\_UI <-ts(Temperature\_UI, frequency =12 , start=c(1988,1))

**Section 1.2 Descriptive Data Analysis**

**1.2.1 The table below represents the various temperature information of the Woodstock data from 1988 to 2018**

|  |  |
| --- | --- |
| Mean Temperature | 5.6543634 |
| Minimum Temperature | -14.3180129 |
| Maximum Temperature | 23.1180589 |
| Median | 5.3108081 |
| Variance | 78.7126550 |
| Standard Deviation | 8.8720153 |

**1.2.2 Time series representing the average temperature for Woodstock from 1988-2018**Chart

Description automatically generated

This is the plot of average temperature in Woodstock over time. It has a repeated pattern, so it is Seasonal and from observation, the average temperature seems to be going up overtime.

**1.2.3 Time Series decomposition**A picture containing application

Description automatically generated

Seasonal: It has got a repeated pattern

Trend: The average temperature seems to be going up overtime

Error: The pseudo random fluctuation overtime seems to be constant overtime

1**.2.4** The p-value is 0.01 which is less than 0.05, hence the data is likely stationary i.e. we reject the null hypothesis which says the data is non stationary

**1.2.5** The chart below represents the Deseasonalized Average Temperature for Woodstock from 1988-2018

Chart, line chart

Description automatically generated

**1.2.6 Add any comments about what you observe:**

From the Deseasonalized chart above, it is quite clear to spot a trend as average temperature increases over the period.

The sum of the seasonal, trend and random temperature values for a particular period is equal to the temperature value.

**Section 2.1 Data Transformation**

We create a dataframe called TempStudy2\_UI , identify it as a ts(timeseries) data, identify the dataframe we are taking it from which is the dataframe which was loaded from the Ayr data file(Temperature2\_UI) .The frequency is 1 since it is annual data. We identify when the data starts i.e., what time period is associated with the first piece of data.

Temperature2\_UI <-get(load("Ayr\_21F.Rdata"))

#Convert to a timeseries datatype

TempStudy2\_UI <-ts(Temperature2\_UI, frequency =1 , start=c(1968))

**Section 2.2 Descriptive Data Analysis**

**2.2.1 The table below represents the various temperature information of the Ayr data from 1968-2003**

|  |  |
| --- | --- |
| Mean Temperature | 12.57479309 |
| Minimum Temperature | 10.98669590 |
| Maximum Temperature | 13.98229284 |
| Variance | 0.48895305 |
| Standard deviation | 0.69925178 |
| Median Temperature | 12.72217789 |

**2.2.2 This represents the timeseries of Ayr from 1968-2003**Chart, line chart

Description automatically generated

This is the plot of average temperature in Ayr over time. It has a repeated pattern so it is Seasonal and from observation, the average temperature seems to be going up overtime.

**Section2.2.3 Moving Average Charts**Chart, line chart

Description automatically generated

Section2.2.3 Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

The temperature chart with the moving average 12 is the best because the nuances are just about enough to make meaningful interpretations.We have been able to get rid of most of the noise but not loose the details.Also the details to interpret the trend are not lost.

**2.2.4**. The p-value is 0.09458 which is greater than 0.05 and so the timeseries is probably non stationary

**2.2.5 The auto correlation chart for the Ayr Data**

Chart, box and whisker chart

Description automatically generated

Any lag below the bottom blue line and above the top blue line is significant, therefore, the lags that are significant are the lags paired 3-6 ,6-9,9-12 because they are all above the top blue line.

**3.1. The chart represents the simple moving Average Temperature forecast for Ayr for 5 years with a 75% prediction interval**

Chart, line chart

Description automatically generated

3.2 **The chart represents the exponentially smoothed Average Temperature forecast for Ayr for 5 years with 75% prediction interval**

Chart, line chart

Description automatically generated

The grey window gives the intervals of 75% that means that we are 75% certain that the average temperature for the next 5 years will be between the lower bound of 12.25 and the upper bound of 13.75

**3.3** The exponential forecast puts more emphasis on more recent events.it suggests that more recent data is better at predicting the future than older data in general. Comparing both forecasts, the trend in the exponential forecast is quite concise and it clearly has an upward trajectory. Therefore, it is superior for forecasting the average temperature for Ayr over the next 5 years with a 75% prediction