Name: Vikram Surya R

Reg No: 19BCE1490

Essentials of Data Analytics

Tasks for Week-3: Regression and Forecasting on Weather Data

Perform multi-regression and forecasting on weather related dataset "weatherHistory2016.csv"

Aim:

(i) To perform multi-regression model on the given data using R programming.
(ii)To understand time-series operations/functions and forecast the next two months of data.

Algorithm:

(a)

1. Load the dataset and dplyr library.

2. Take a sample of the dataset

3. Create independent and dependent variables

4. Check the correlation between x and y

5. Plot a scatter plot of the dataset and develop linear model

6. Plot a regression line

7. See the summary of the model

(b)

1. Load the dataset, forecast library and tseries library.

2. Convert this data into time series and plot time series data.

3. Perform adf test

4. Develop an auto arima model.

5. Perform forecast using the model.

6. Plot the forecast and check accuracy of model.

CODE:

```
setwd("C:/Users/VIKRAM SURYA/Desktop/EDA_LAB")

a=read.csv("weatherHistory2016.csv")

library(dplyr)

a=sample_n(a,200)


cor.test(a$Temperature..C.,a$Apparent.Temperature..C.)

cor.test(a$Temperature..C.,a$Humidity)

cor.test(a$Temperature..C.,a$Wind.Speed..km.h.)

cor.test(a$Temperature..C.,a$Wind.Bearing..degrees.)

cor.test(a$Temperature..C.,a$Pressure..millibars.)

x=a$Temperature..C.

y1=a$Humidity

y2=a$Apparent.Temperature..C.

model = lm(x~y1+y2,data=a)

summary(model)

plot(model)


library(forecast)

library(tseries)

data <- ts(a$Temperature..C., start=as.Date("2016-10-01"), end=as.Date("2016-12-31"),frequency = 24)
```

```
plot(data)

adf.test(data)

model1=auto.arima(data,ic="aic",trace=TRUE)

modelf=forecast(model1,level=c(95),h=60)

modelf

plot(modelf)
```

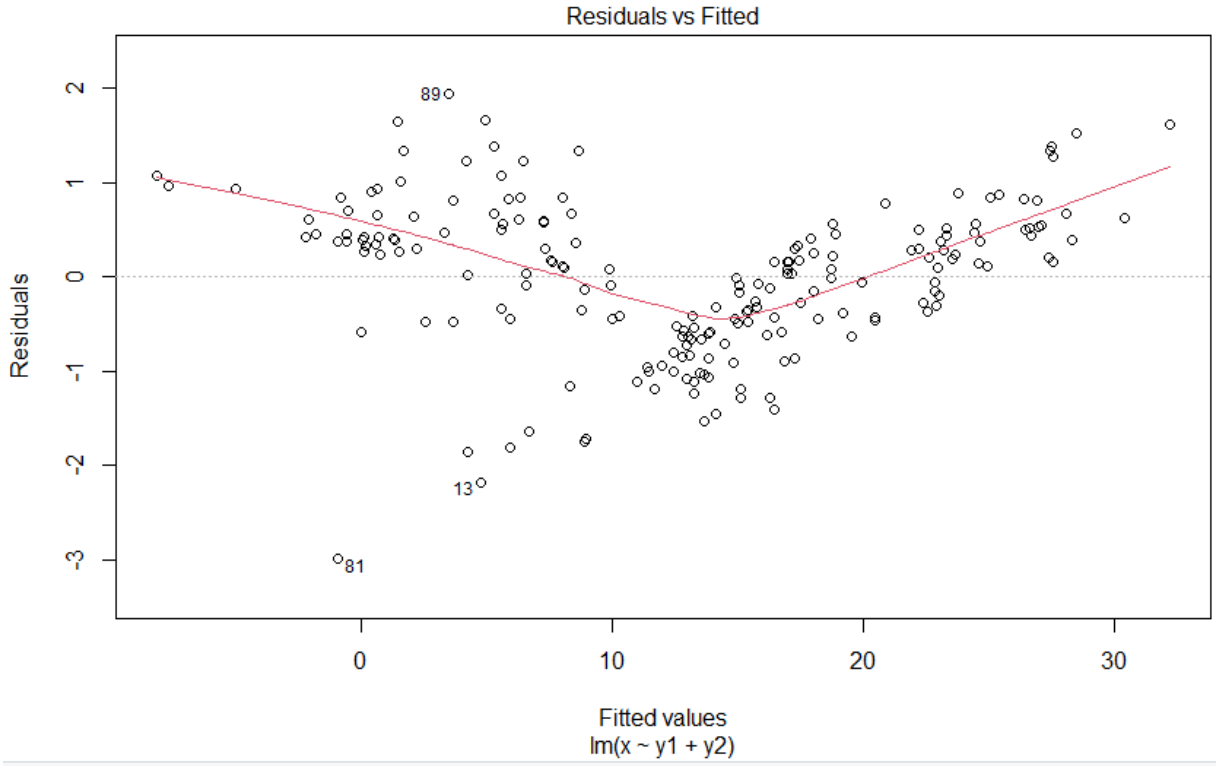
INFERENCE:

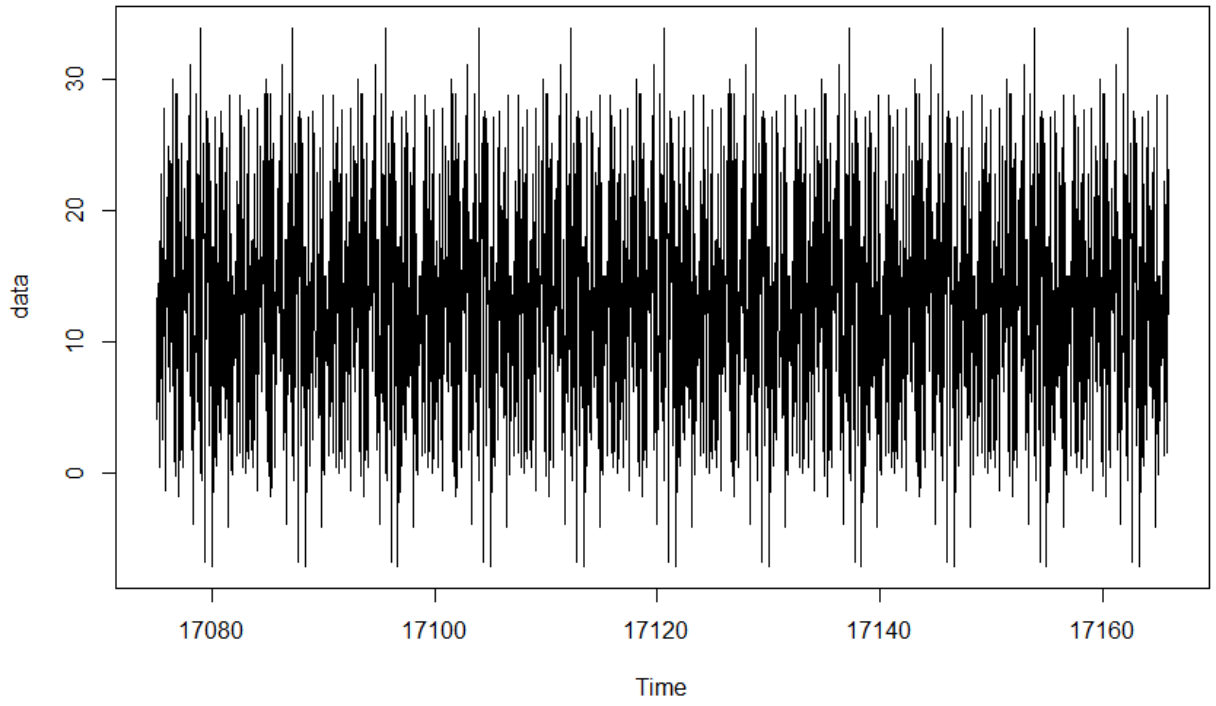
The correlation values for humidity and apparent temperature is greater than 0.5. So these columns are selected.

```
> cor.test(a$Temperature..C.,a$Apparent.Temperature..C.)

        Pearson's product-moment correlation

data:  a$Temperature..C. and a$Apparent.Temperature..C.
t = 146.04, df = 198, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.9939095 0.9965115
sample estimates:
      cor
0.9953902

> cor.test(a$Temperature..C.,a$Humidity)

        Pearson's product-moment correlation

data:  a$Temperature..C. and a$Humidity
t = -13.526, df = 198, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.7587973 -0.6132356
sample estimates:
       cor
-0.6930143

> cor.test(a$Temperature..C.,a$Wind.Speed..km.h.)

        Pearson's product-moment correlation

data:  a$Temperature..C. and a$Wind.Speed..km.h.
t = 0.15829, df = 198, p-value = 0.8744
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.1276921  0.1497555
sample estimates:
       cor
0.01124817
```

```
> cor.test(a$Temperature..C.,a$Wind.Bearing..degrees.)

        Pearson's product-moment correlation

data:  a$Temperature..C. and a$Wind.Bearing..degrees.
t = -0.46827, df = 198, p-value = 0.6401
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.171211  0.105970
sample estimates:
        cor
-0.03326005

> cor.test(a$Temperature..C.,a$Pressure..millibars.)

        Pearson's product-moment correlation

data:  a$Temperature..C. and a$Pressure..millibars.
t = -2.5775, df = 198, p-value = 0.01068
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.31113970 -0.04249786
sample estimates:
       cor
-0.1801765
```

```
> summary(model)

Call:
lm(formula = x ~ y1 + y2, data = a)

Residuals:
    Min      1Q  Median      3Q     Max
-2.9933 -0.4820  0.1325  0.5035  1.9373

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.264437   0.369799  11.532  < 2e-16
y1          -2.089570   0.399888  -5.225 4.41e-07
y2           0.855475   0.007659 111.700  < 2e-16

(Intercept) ***
y1          ***
y2          ***
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8037 on 197 degrees of freedom
Multiple R-squared:  0.9919,    Adjusted R-squared:  0.9918
F-statistic: 1.209e+04 on 2 and 197 DF,  p-value: < 2.2e-16
```


Residuals vs Fitted
Residuals
Fitted values
lm(x ~ y1 + y2)
89
13
81
2
1
0
-1
-2
-3
0
10
20
30


data
Time
30
20
10
0
17080
17100
17120
17140
17160

```
> adf.test(data)

        Augmented Dickey-Fuller Test

data:  data
Dickey-Fuller = -13.189, Lag order = 12,
p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(data) : p-value smaller than printed p-valu
e


 Fitting models using approximations to speed things u
p...

 ARIMA(2,0,2)(1,0,1)[24] with non-zero mean : Inf
 ARIMA(0,0,0)            with non-zero mean : 15740.94
 ARIMA(1,0,0)(1,0,0)[24] with non-zero mean : 15731.93
 ARIMA(0,0,1)(0,0,1)[24] with non-zero mean : 15721.3
 ARIMA(0,0,0)            with zero mean     : 18255.93
 ARIMA(0,0,1)            with non-zero mean : 15720.14
 ARIMA(0,0,1)(1,0,0)[24] with non-zero mean : 15730.98
 ARIMA(0,0,1)(1,0,1)[24] with non-zero mean : 15731.75
 ARIMA(1,0,1)            with non-zero mean : 15698.57
 ARIMA(1,0,1)(1,0,0)[24] with non-zero mean : 15715.31
 ARIMA(1,0,1)(0,0,1)[24] with non-zero mean : 15698.05
 ARIMA(1,0,1)(1,0,1)[24] with non-zero mean : 15717.12
 ARIMA(1,0,1)(0,0,2)[24] with non-zero mean : 15698.55
 ARIMA(1,0,1)(1,0,2)[24] with non-zero mean : 15717.39
 ARIMA(1,0,0)(0,0,1)[24] with non-zero mean : 15723.06
 ARIMA(2,0,1)(0,0,1)[24] with non-zero mean : 15706.89
 ARIMA(1,0,2)(0,0,1)[24] with non-zero mean : 15705.84
 ARIMA(0,0,0)(0,0,1)[24] with non-zero mean : 15740.99
 ARIMA(0,0,2)(0,0,1)[24] with non-zero mean : 15717.25
 ARIMA(2,0,0)(0,0,1)[24] with non-zero mean : 15723.75
 ARIMA(2,0,2)(0,0,1)[24] with non-zero mean : Inf
 ARIMA(1,0,1)(0,0,1)[24] with zero mean     : Inf

 Now re-fitting the best model(s) without approximation
s...

 ARIMA(1,0,1)(0,0,1)[24] with non-zero mean : Inf
```

```
> modelf
         Point Forecast     Lo 95    Hi 95
17166.04       11.34524 -5.881534 28.57202
17166.08       12.61993 -4.707899 29.94775
17166.12       12.75615 -4.615678 30.12797
17166.17       12.91982 -4.464463 30.30411
17166.21       12.81445 -4.573373 30.20226
17166.25       13.00968 -4.379138 30.39850
17166.29       13.21407 -4.175032 30.60318
17166.33       13.05886 -4.330324 30.44805
17166.38       13.16733 -4.221878 30.55654
17166.42       12.99318 -4.396038 30.38239
17166.46       12.89428 -4.494935 30.28350
17166.50       13.53408 -3.855140 30.92330
17166.54       13.30201 -4.087205 30.69123
17166.58       12.67236 -4.716856 30.06158
17166.62       12.79252 -4.596699 30.18174
17166.67       12.73860 -4.650622 30.12781
17166.71       13.37311 -4.016111 30.76233
17166.75       13.29657 -4.092653 30.68578
17166.79       12.42317 -4.966044 29.81239
17166.83       13.52506 -3.864158 30.91428
17166.88       12.43057 -4.958651 29.81979
17166.92       13.07584 -4.313381 30.46506
17166.96       12.61913 -4.770090 30.00835
17167.00       12.89093 -4.498291 30.28015
17167.04       13.11085 -4.293060 30.51476
17167.08       13.08224 -4.321841 30.48633
17167.12       13.06391 -4.340253 30.46807
17167.17       13.05414 -4.350041 30.45832
17167.21       13.04894 -4.355249 30.45313
17167.25       13.04617 -4.358021 30.45036
17167.29       13.04469 -4.359498 30.44888
17167.33       13.04391 -4.360284 30.44810
17167.38       13.04349 -4.360702 30.44768
17167.42       13.04326 -4.360925 30.44745
```