****

**ABHINAV VIJAYAKUMAR**

**19BCE1311**

**CSE3506 – ESSENTIALS OF DATA ANALYTICS LAB-3**

**DR. LAKSHMI PATHI JAKKAMPUTI (L21 + L22)**

**--------------------------------------------------------------------------------------------**

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

**Perform multi-regression and forecasting on weather related dataset “weatherHistory2016.csv”**

**Aim**: To forecast the dependent variable temperature, based multiple independent variables.

**Algorithm:**

**1.** Attach library forecast, dplyr, corrplot, tseries.

**2.** Set working directory and read data.

**3.** Check correlation of variable.

**4.** Make multiple linear regression models.

**5.** Choose the best fit model.

**6.** Make a new dataset using the correlated variables only.

**7.** Formulate time series data.

**8.** Plot the time series data.

**9.** Plot the acf and pacf graph.

**10.** Perform the adf test, to determine the p value.

**11.** Check for stationary values.

**12.** Use auto ARIMA function to get the best fit model.

**13.** Perform forecasting with 95% level confidence.

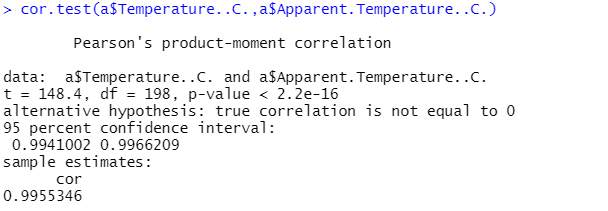
**14.** Plot the forecasted data.

**Statistics:**

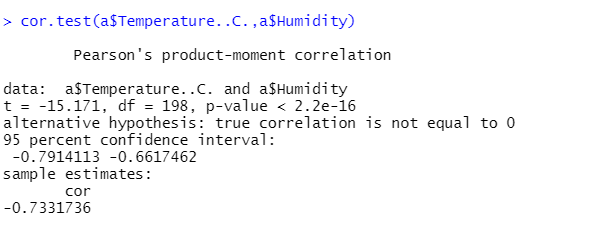
* **Multivariate Regression:**

**Correlation Test:**

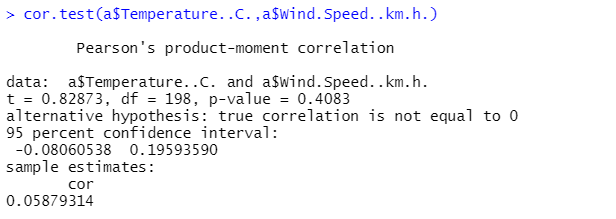
**a. Apparent Temperature**

****

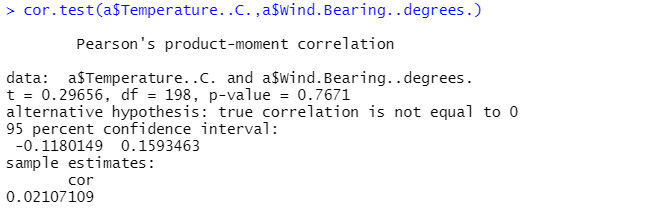
**b. Humidity**

****

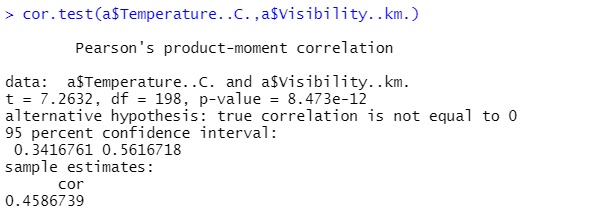
**c. Wind Speed**

****

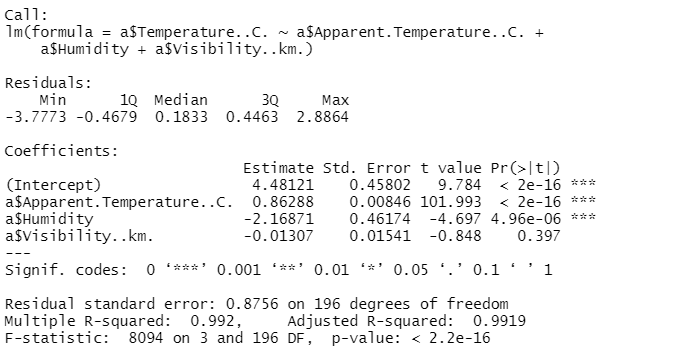
**d. Wind Bearing**

****

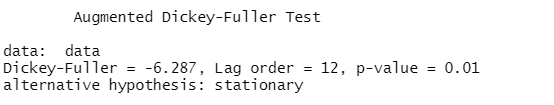
**e. Visibility**

****

**Model Statistics:**



* **Forecasting:**

****

**Best ARIMA Model:**

Fitting models using approximations to speed things up...

ARIMA(2,0,2)(1,1,1)[24] with drift : Inf

ARIMA(0,0,0)(0,1,0)[24] with drift : 11466.33

ARIMA(1,0,0)(1,1,0)[24] with drift : 5676.337

ARIMA(0,0,1)(0,1,1)[24] with drift : 8977.075

ARIMA(0,0,0)(0,1,0)[24] : 11473.89

ARIMA(1,0,0)(0,1,0)[24] with drift : 6252.305

ARIMA(1,0,0)(2,1,0)[24] with drift : 5438.054

ARIMA(1,0,0)(2,1,1)[24] with drift : Inf

ARIMA(1,0,0)(1,1,1)[24] with drift : Inf

ARIMA(0,0,0)(2,1,0)[24] with drift : 11281.55

ARIMA(2,0,0)(2,1,0)[24] with drift : 5374.887

ARIMA(2,0,0)(1,1,0)[24] with drift : 5600.859

ARIMA(2,0,0)(2,1,1)[24] with drift : Inf

ARIMA(2,0,0)(1,1,1)[24] with drift : Inf

ARIMA(3,0,0)(2,1,0)[24] with drift : 5331.394

ARIMA(3,0,0)(1,1,0)[24] with drift : 5559.53

ARIMA(3,0,0)(2,1,1)[24] with drift : Inf

ARIMA(3,0,0)(1,1,1)[24] with drift : Inf

ARIMA(4,0,0)(2,1,0)[24] with drift : 5332.032

ARIMA(3,0,1)(2,1,0)[24] with drift : 5331.313

ARIMA(3,0,1)(1,1,0)[24] with drift : 5558.243

ARIMA(3,0,1)(2,1,1)[24] with drift : Inf

ARIMA(3,0,1)(1,1,1)[24] with drift : Inf

ARIMA(2,0,1)(2,1,0)[24] with drift : 5340.401

ARIMA(4,0,1)(2,1,0)[24] with drift : 5334.033

ARIMA(3,0,2)(2,1,0)[24] with drift : 5332.077

ARIMA(2,0,2)(2,1,0)[24] with drift : 5330.361

ARIMA(2,0,2)(1,1,0)[24] with drift : 5556.545

ARIMA(2,0,2)(2,1,1)[24] with drift : Inf

ARIMA(1,0,2)(2,1,0)[24] with drift : 5343.612

ARIMA(2,0,3)(2,1,0)[24] with drift : 5331.938

ARIMA(1,0,1)(2,1,0)[24] with drift : 5390.12

ARIMA(1,0,3)(2,1,0)[24] with drift : 5332.634

ARIMA(3,0,3)(2,1,0)[24] with drift : 5334.228

ARIMA(2,0,2)(2,1,0)[24] : 5329.467

ARIMA(2,0,2)(1,1,0)[24] : 5555.177

ARIMA(2,0,2)(2,1,1)[24] : Inf

ARIMA(2,0,2)(1,1,1)[24] : Inf

ARIMA(1,0,2)(2,1,0)[24] : 5342.563

ARIMA(2,0,1)(2,1,0)[24] : 5339.546

ARIMA(3,0,2)(2,1,0)[24] : 5331.22

ARIMA(2,0,3)(2,1,0)[24] : 5331.029

ARIMA(1,0,1)(2,1,0)[24] : 5388.923

ARIMA(1,0,3)(2,1,0)[24] : 5331.69

ARIMA(3,0,1)(2,1,0)[24] : 5330.489

ARIMA(3,0,3)(2,1,0)[24] : Inf

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

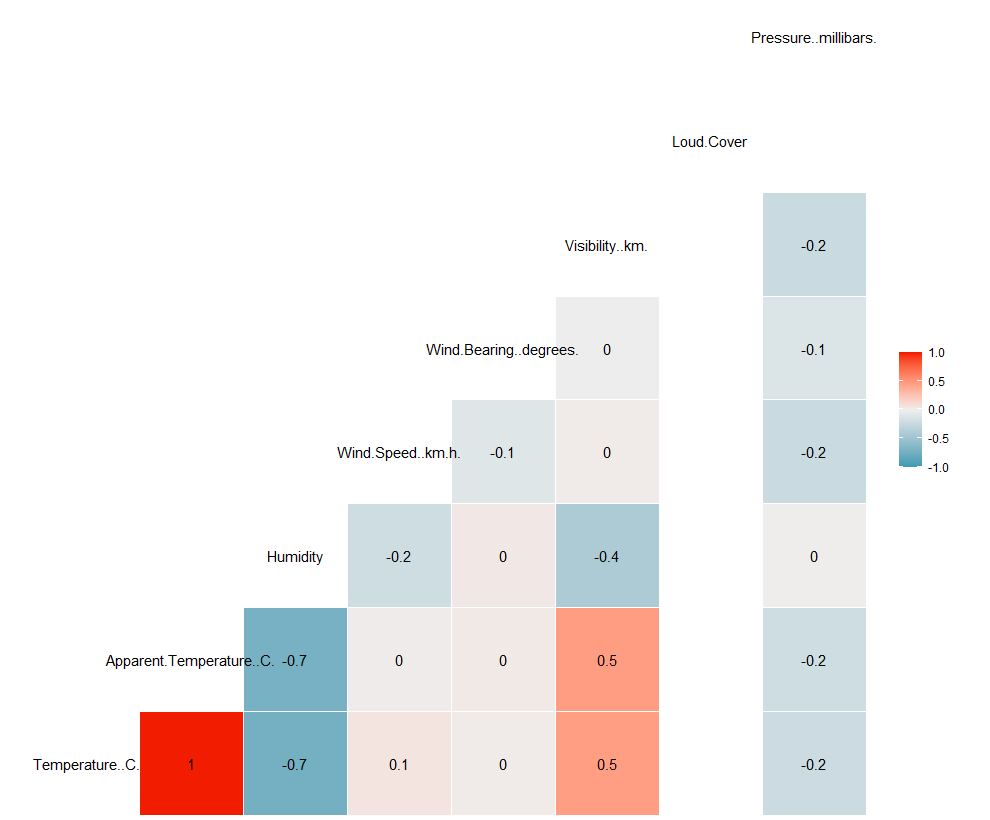
ARIMA(2,0,2)(2,1,0)[24] : 5384.374

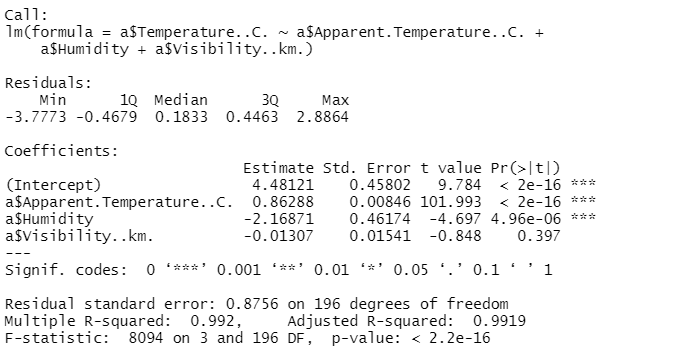
Best model: ARIMA(2,0,2)(2,1,0)[24]

**Inference:**

* **Multivariate Regression:**

The best model was made after considering 3 variable which were highly correlated to the dependent variable and those variables were Apparent.Temperature (0.9955), Humidity (-0.733) and Visibility(0.458).





* **Forecasting:**

**Best ARIMA Model:**

ARIMA(2,0,2)(2,1,0)[24] : 5384.374

Best model: ARIMA(2,0,2)(2,1,0)[24]

**Accuracy of the Model:**

****

**Forecast for 1 day:**

**Point Forecast Lo 95 Hi 95**

17166.04 11.248803 9.60890098 12.88871

17166.08 10.731641 8.28507274 13.17821

17166.12 10.136301 6.97240528 13.30020

17166.17 9.559795 5.80095929 13.31863

17166.21 8.281384 4.03549680 12.52727

17166.25 8.287412 3.64185058 12.93297

17166.29 10.885030 5.90918055 15.86088

17166.33 13.721924 8.47093588 18.97291

17166.38 15.721102 10.23917779 21.20303

17166.42 17.608383 11.93132980 23.28544

17166.46 19.327787 13.48492337 25.17065

17166.50 19.884300 13.89986870 25.86873

17166.54 20.491771 14.38599322 26.59755

17166.58 20.578257 14.36812422 26.78839

17166.62 19.558943 13.25882055 25.85906

17166.67 19.521777 13.14387780 25.89968

17166.71 18.510491 12.06524140 24.95574

17166.75 15.073873 8.57020882 21.57754

17166.79 13.978204 7.42380488 20.53260

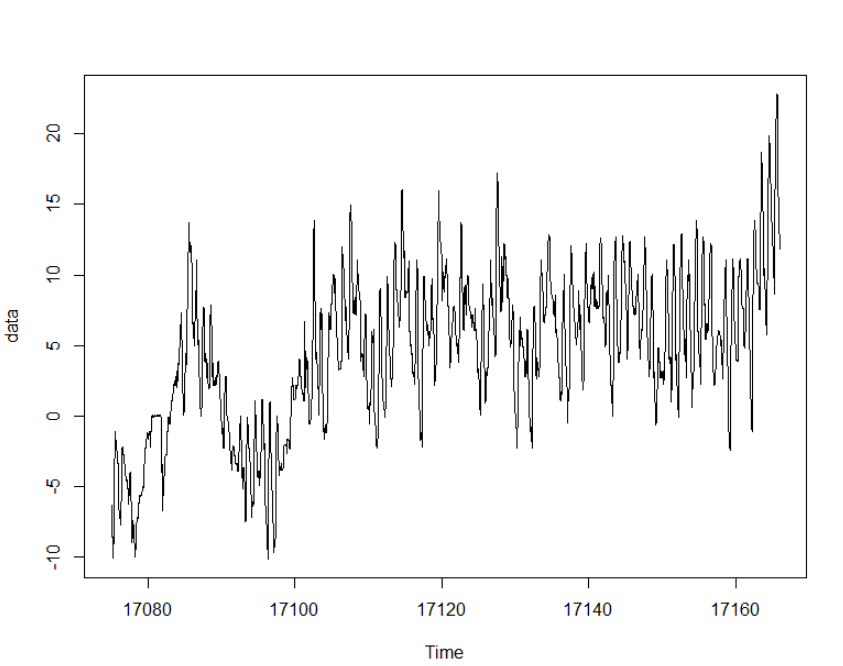
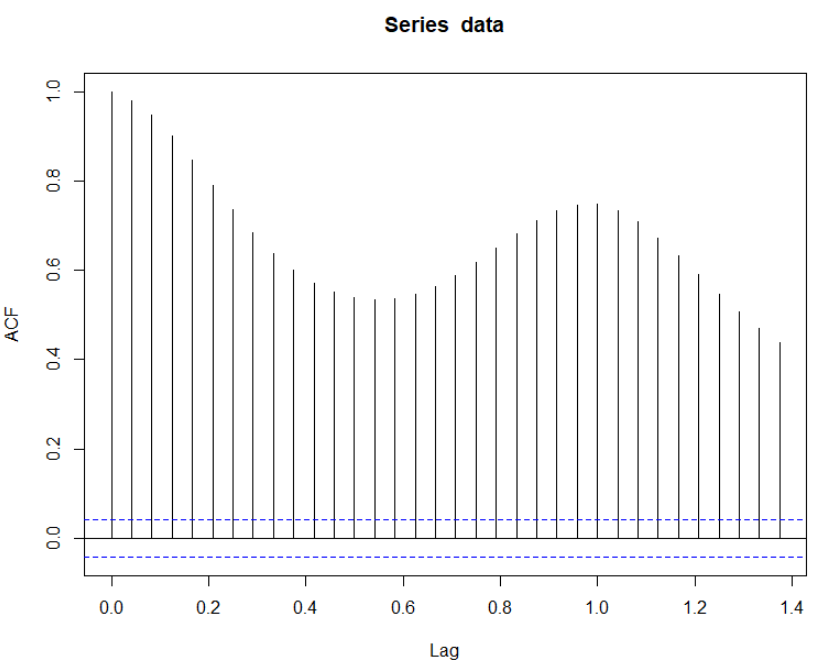
17166.83 13.325428 6.72691417 19.92394

17166.88 13.118927 6.48201677 19.75584

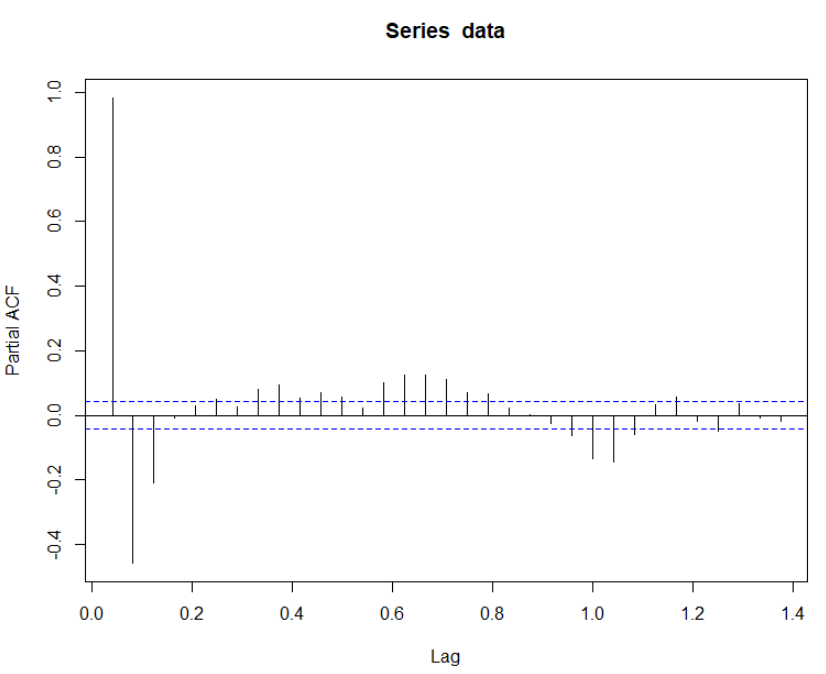
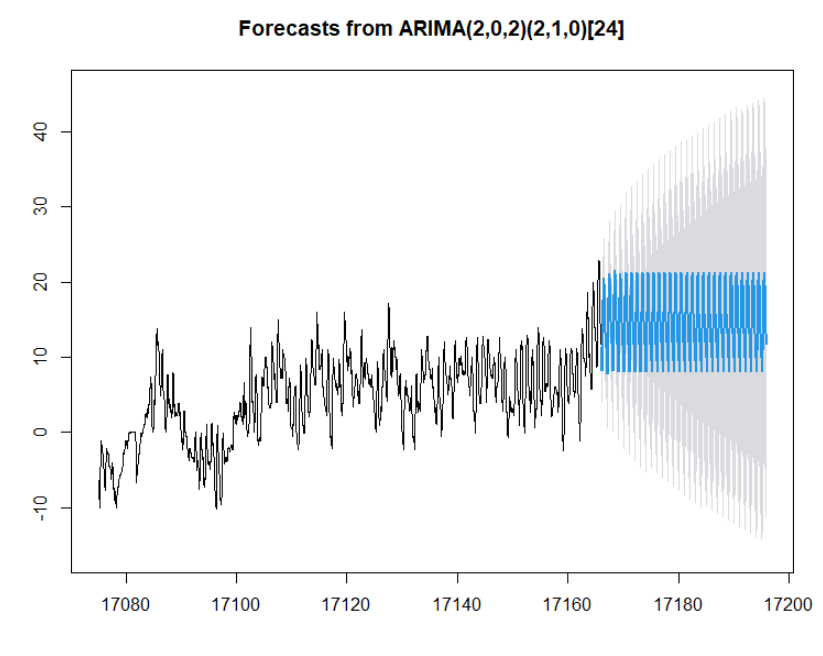
17166.92 12.649515 5.97915833 19.31987

17166.96 11.722233 5.02272037 18.42175

17167.00 11.328230 4.60328548 18.05318

**Plotting the data time series Autocorrelation(acf)**

** **

**Partial acf Plotting the forecast**

**Program:**

**i) Multivariate Regression:**

setwd("C:/Users/Abhinav Vijayakumar/Desktop/VIT Academics/Sem 6/Essentials of Data Analytics/LAB/LAB 3")

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

head(dff)

library(dplyr)

library(GGally)

a=sample\_n(dff,200)

head(a)

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.)

cor.test(a$Temperature..C.,a$Visibility..km.)

cor.test(a$Temperature..C.,a$Loud.Cover)

ggcorr(a %>% mutate\_if(is.factor, as.numeric), label = TRUE)

lmodel=lm(a$Temperature..C.~a$Apparent.Temperature..C.+a$Humidity+a$Visibility..km.)

summary(lmodel)

plot(lmodel)

**ii) Forecast:**

setwd("C:/Users/Abhinav Vijayakumar/Desktop/VIT Academics/Sem 6/Essentials of Data Analytics/LAB/LAB 3")

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

library(forecast)

library(tseries)

data<-ts(dff$Temperature..C.,start = as.Date("2016-10-01"),end =

as.Date("2016-12-31"),frequency = 24)

plot(data)

acf(data)

pacf(data)

adf.test(data)

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

f=forecast(model,level=c(95),h=720)

f

plot(f)

accuracy(model)