**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Internet of Things Analytics**

Advances in Data Science and Architecture

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Under the guidance

Professor Srikanth Krishnamurthy

**Team 8**

* Ankita Suresh Shanbhag
* Vignesh Karthikeyan
* Sriniketan G.S.

Index

Abstract 4

Visualization and Analysis 5

2.1 Plot the time series dataset 6

2.2 Detect the seasonality using Fourier transform 6

2.3 DE trend the time series 9

2.4 Average Seasonality 9

2.5 Random Noise extraction 10

2.6 Decompose 11

ARIMA Forecast 11

3.1 Estimation and Forecasting using Box-Jenkins (B-J) methodology 12

3.1.1 Identification 12

Azure Experiment 14

4.1 Data collection 14

4.2 Data Cleaning 15

4.3 Compute the Events 22

4.4 Predict the ‘Events’ using dependent variable mentioned above. 25

4.5 Create a Web Service for the above model 26

POWER BI 27

Web Application 30

Reference 32

# Abstract

We have the weather station data from Sensors (the sensor we have chosen is <https://www.wunderground.com/>), for this analysis we have chosen Boston, MA and weather data from the year 2008 to present. We are going to forecast the weather for the next few days. Our Web application will display the forecast for the future period and will get updated every day.

Most of the cities have a weather station at the Airport but many places like Universities and Government properties have their weather stations too, if they can forecast with their own stations they can feed a more accurate data into their building (services) controls.

We are forecasting the future weather using the past data so we have used AR + I + MA modeling (ARIMA) for the forecast as the future data depends only on the past trend. And we have tried to observe the trend and seasonal components out of our data and the trend represents the gradual increase in temperature over the year, which might be due to global warming.

Our data set has the following components.

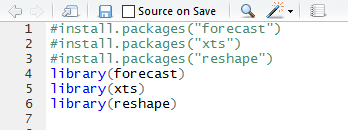
|  |  |
| --- | --- |
| **Column** | **Type** |
| Timestamp | Date |
| EST | character |
| Max.TemperatureF | numeric |
| Mean.TemperatureF | numeric |
| Min.TemperatureF | numeric |
| Max.Dew.PointF | numeric |
| MeanDew.PointF | numeric |
| Min.DewpointF | numeric |
| Max.Humidity | numeric |
| Mean.Humidity | numeric |
| Min.Humidity | numeric |
| Max.Sea.Level.PressureIn | numeric |
| Mean.Sea.Level.PressureIn | numeric |
| Min.Sea.Level.PressureIn | numeric |
| Max.VisibilityMiles | numeric |
| Mean.VisibilityMiles | numeric |
| Min.VisibilityMiles | numeric |
| Max.Wind.SpeedMPH | numeric |
| Mean.Wind.SpeedMPH | numeric |
| Max.Gust.SpeedMPH | numeric |
| PrecipitationIn | numeric |
| CloudCover | numeric |
| Events | character |
| WindDirDegrees.br... | character |

# Visualization and Analysis

To forecast the future weather data and understand how it behaves, so beginning with the temperature data over 9 years lets plot it against time.

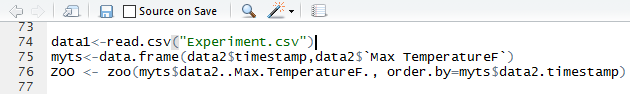
We will do this in R

Packages you might need:



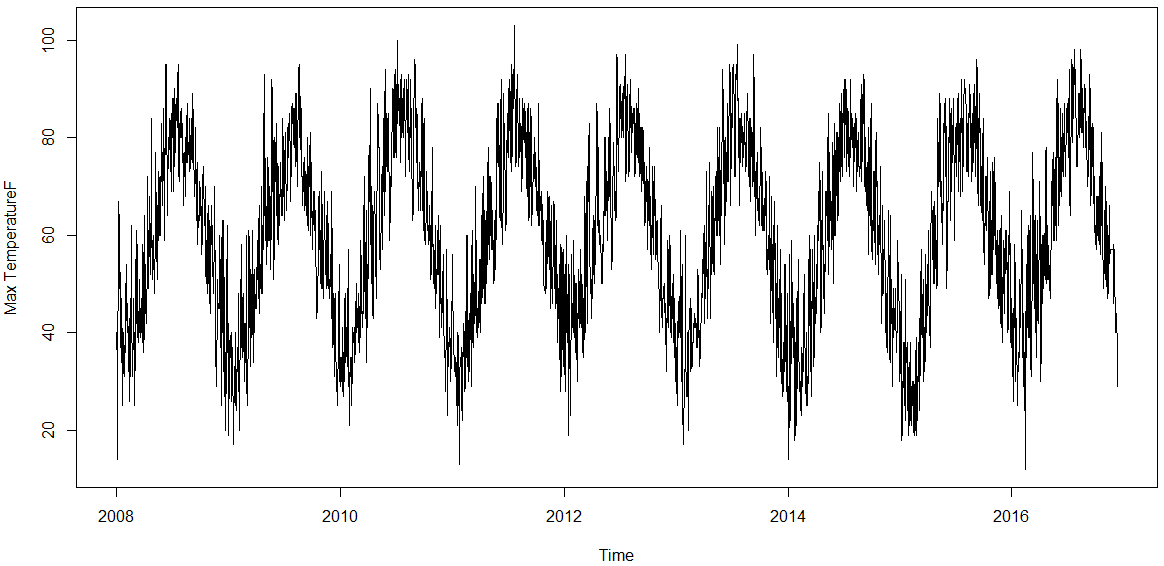
We have the weather data till date from Azure ML(Experiment.csv)

Choose the Max Temperature and create a Time series data set(ZOO)



## 2.1 Plot the time series dataset

We get the following plot

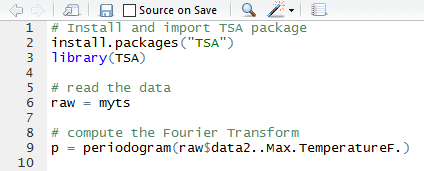


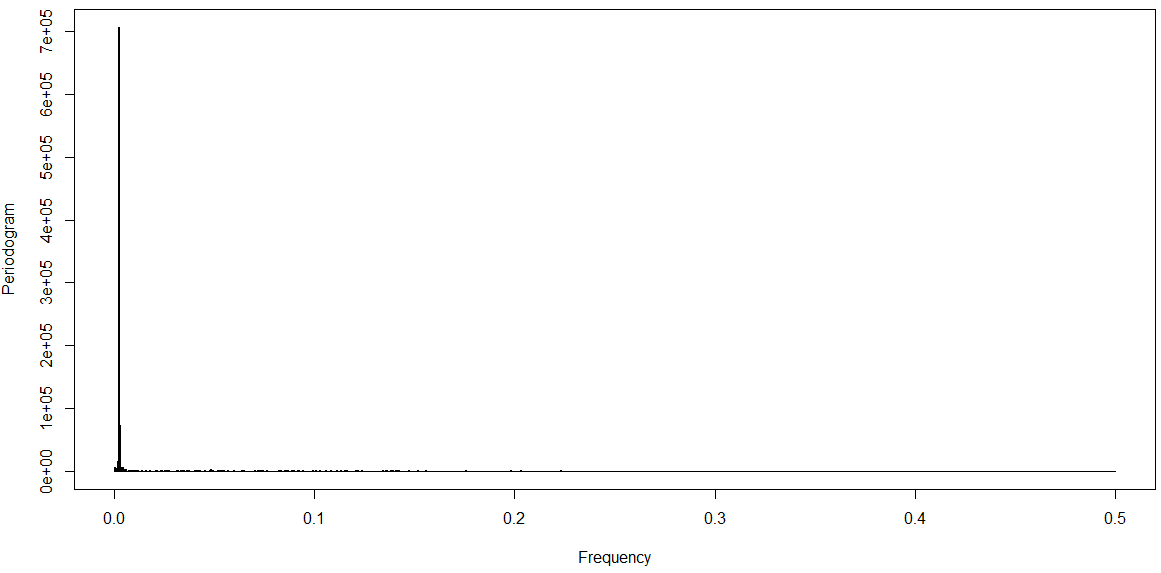
It can be observed that there is a seasonality component to data but the trend is not clearly visible so on decomposing we get.

## 2.2 Detect the seasonality using Fourier transform

For this data, you can observe that the seasonality is 365/366 days as the temperature varies with respect to the earth’s rotation which is a cycle of 365/366 days.

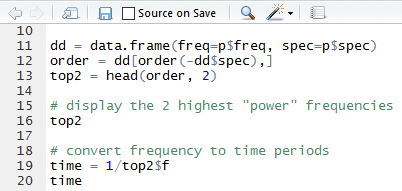
But for rest of the data where it is not clear we can use Fourier transform to get the frequency.

****

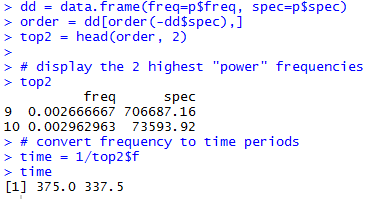
****

It can be observed that there is some trend around0.01 frequency

Get the seasonality

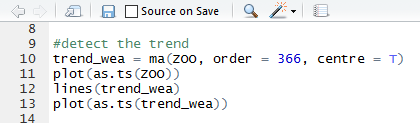


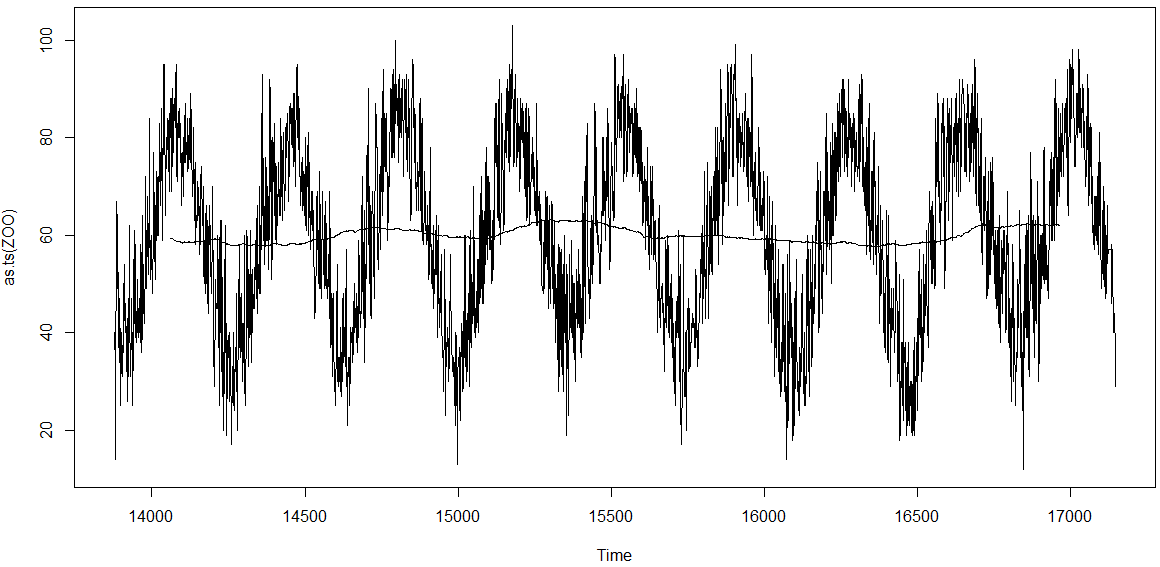
The seasonality is

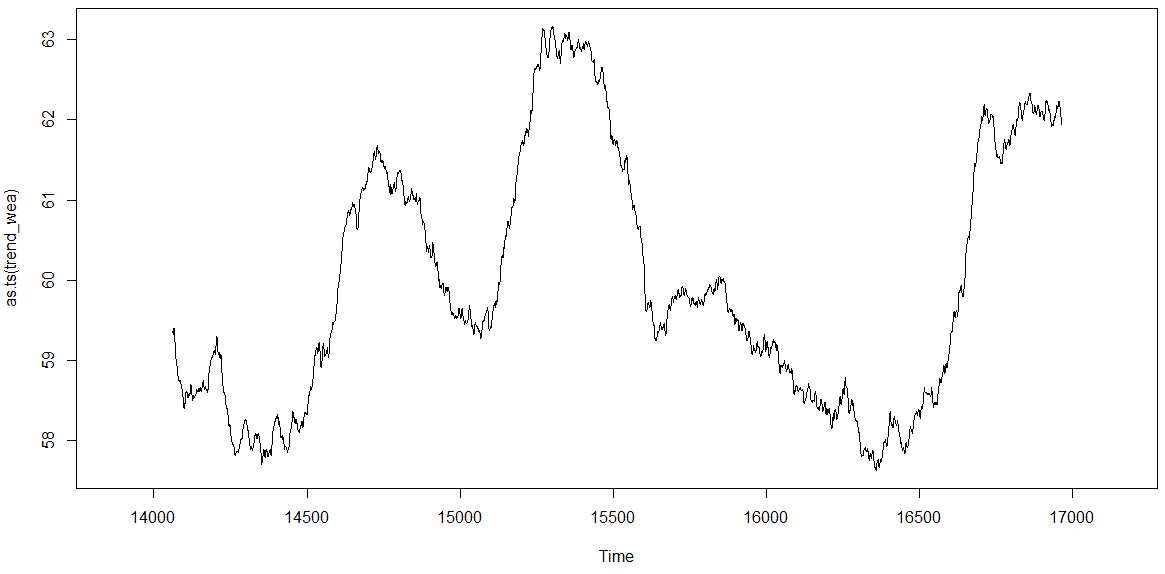


The average of which is 356.25, but we will use 365 as we know it for this data (which is reconfirmed)

*Use moving average windows of 365 as the data is collected for every day for 9 years*





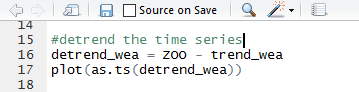


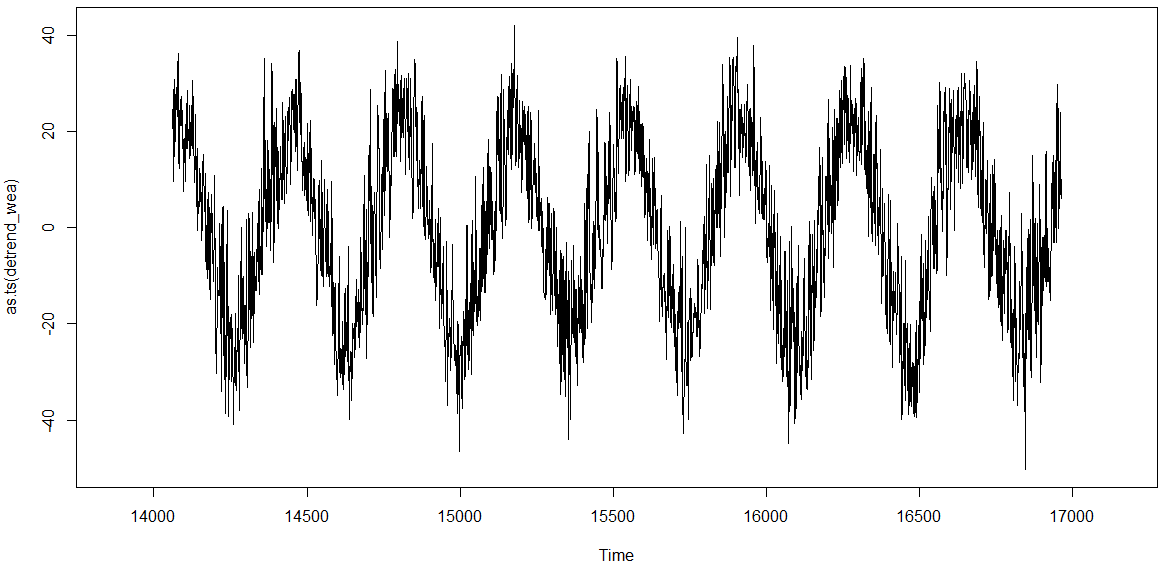
The trend component seems random, but on close inspection you can see the temperature it is varying between is 58F and 61F and only for a period of 9 years, which is not enough the observe a trend of increasing temperature.

We will still remove it from the seasonal data.

From the graph we can observe that the trend and seasonal components are additive.

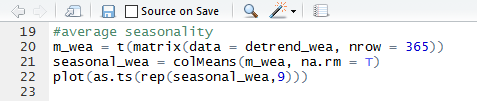
## 2.3 DE trend the time series

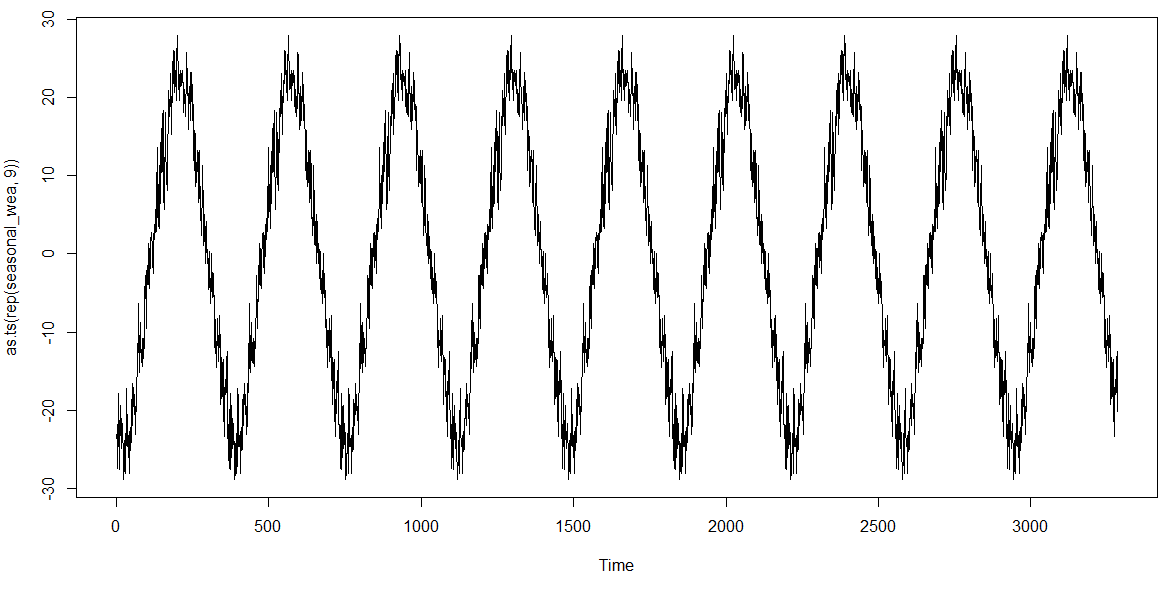




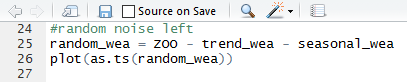
The trend component has been removed from the time series

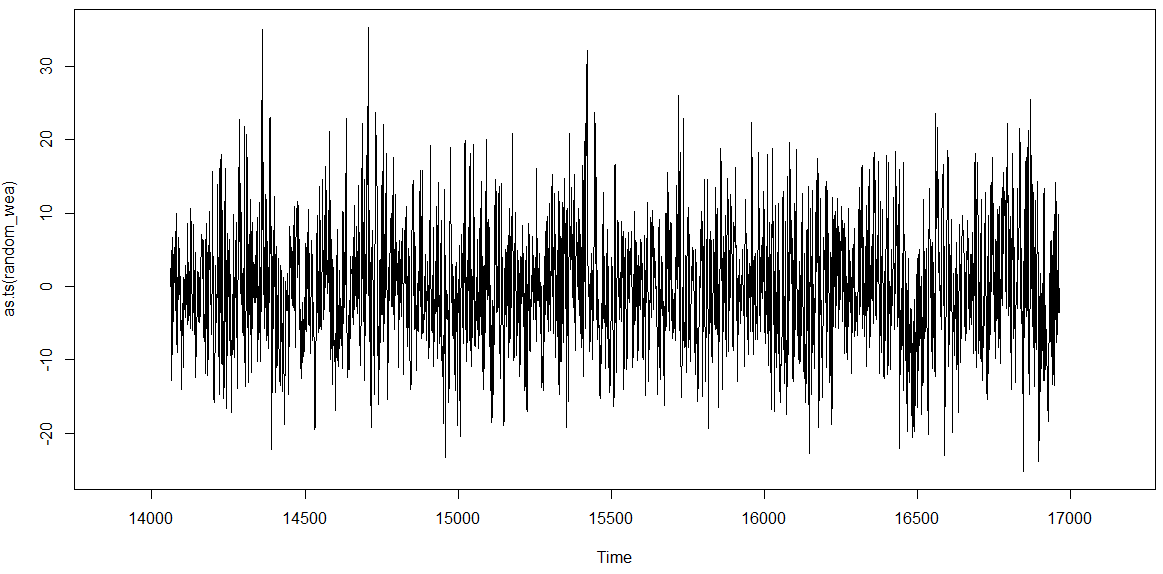
## 2.4 Average Seasonality



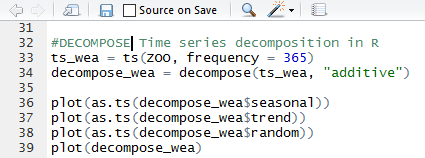


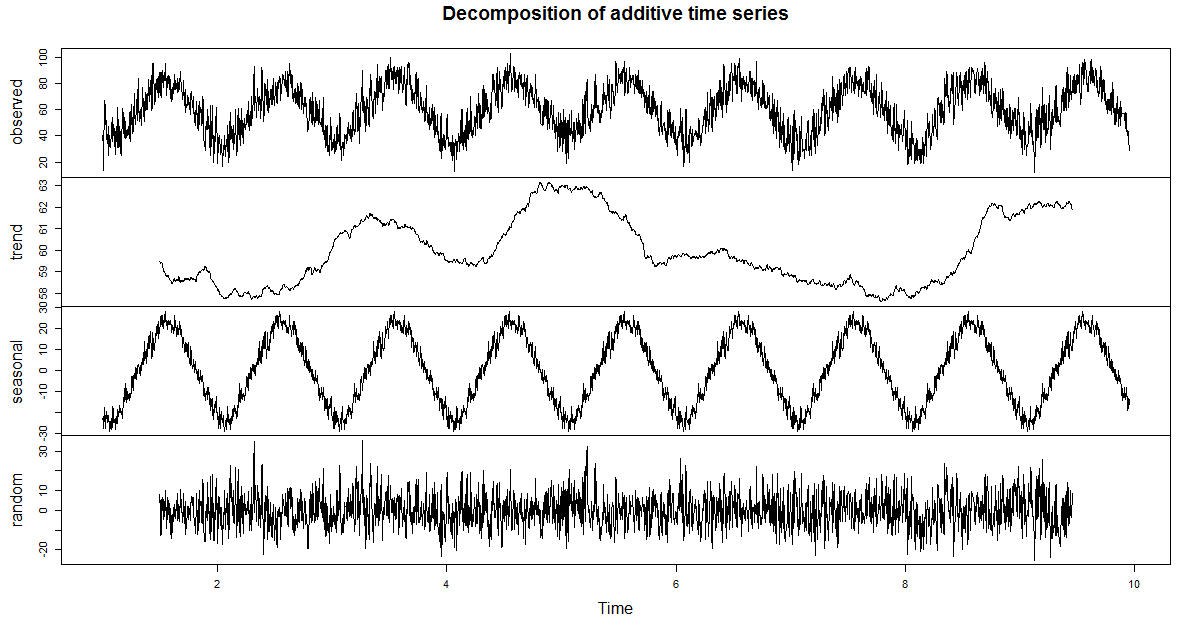
## 2.5 Random Noise extraction





## 2.6 Decompose

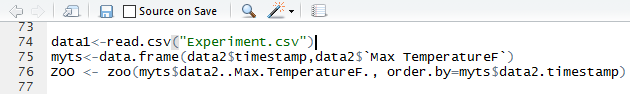




ARIMA Forecast

Our data is Multivariate Time Series as we have multiple time series dependent variables.

For each Model we will have to convert it to Univariate Time Series for modeling.



Now this data just has Max Temperature.

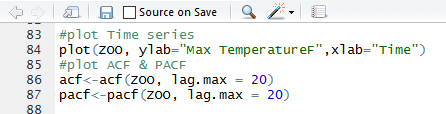
For our Data we have observed that there is some trend observed and hence we will be using ARIMA model instead of ARMA model as it is not stationary Time Series.

## 3.1 Estimation and Forecasting using Box-Jenkins (B-J) methodology

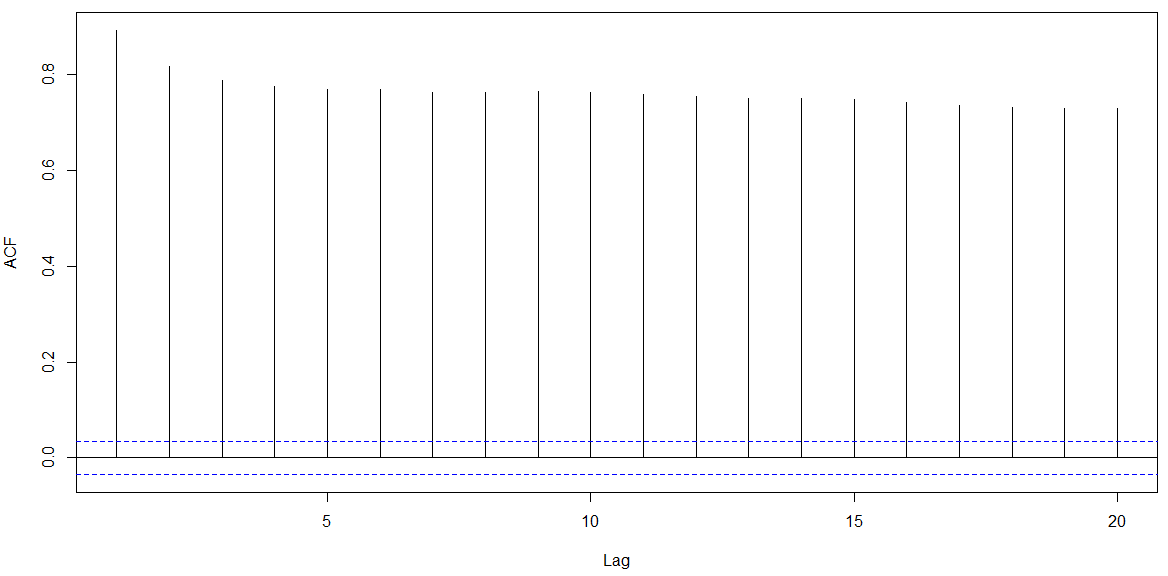
### 3.1.1 Identification

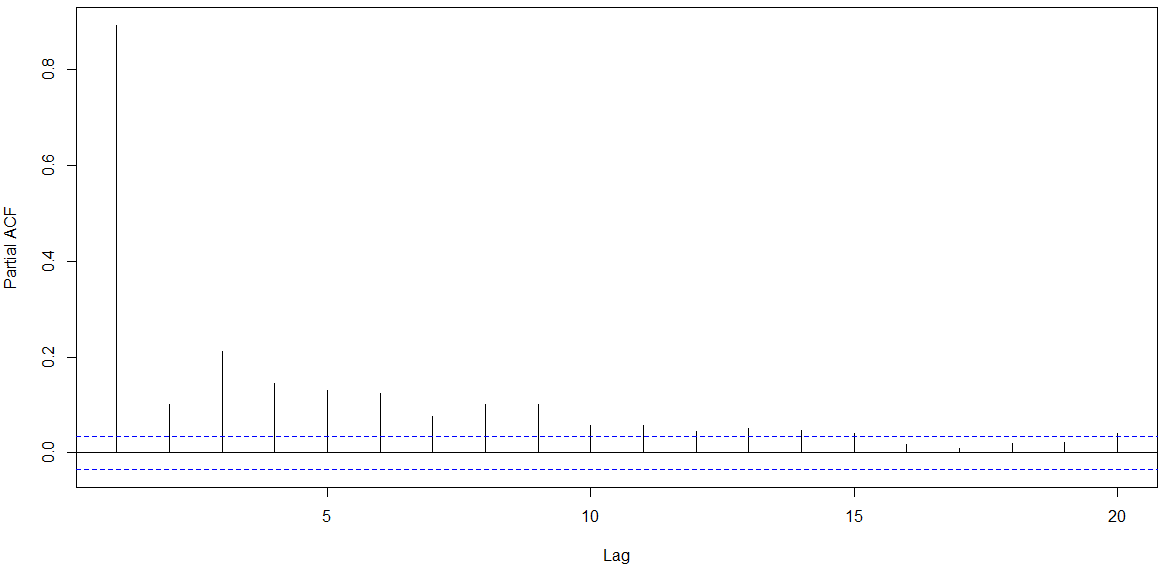
Autocorrelation Function(ACF)

Partial Autocorrelation Function(PACF)



We have seen the Time Series before so now we will see the ACF $ PACF Plots

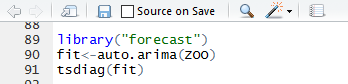




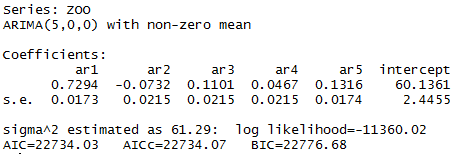
To see how the present values are related to past value and how many past value.

And depending on that you pick ARIMA(p, q)

In R you can get it Automatically

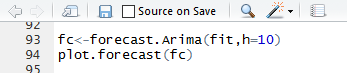


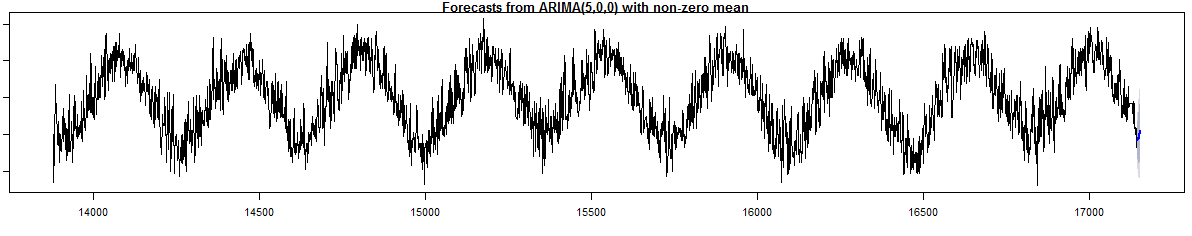
Which gives us the best result



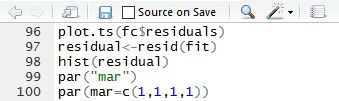
The best Model is ARIMA(5,0,0)

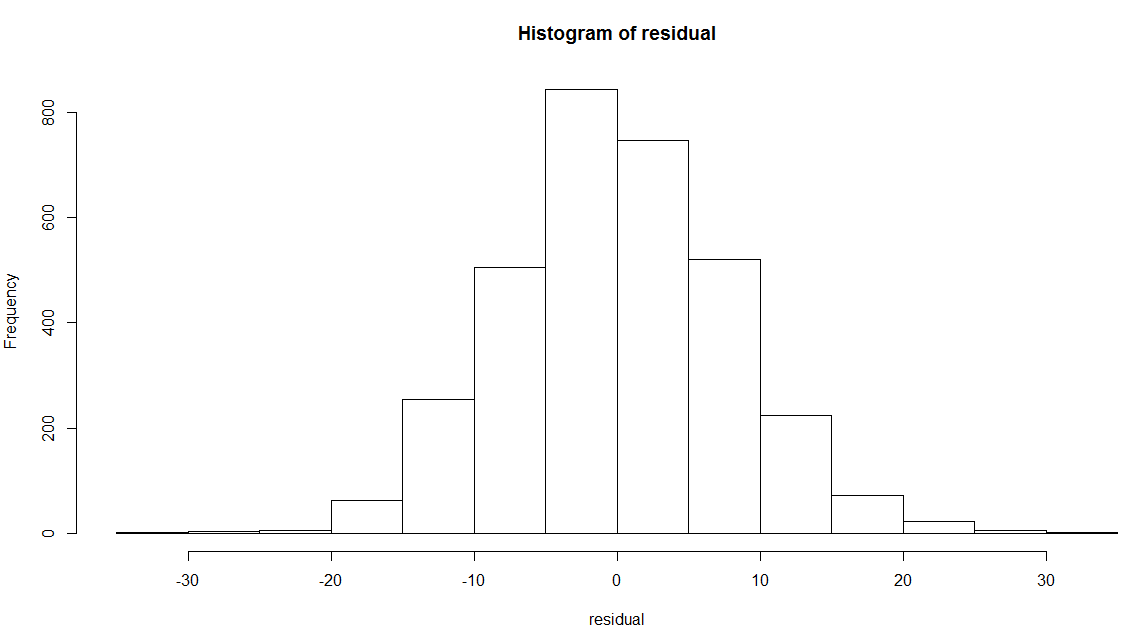
Forecast for next 10 days





The forecast for next 10 days is near the red star



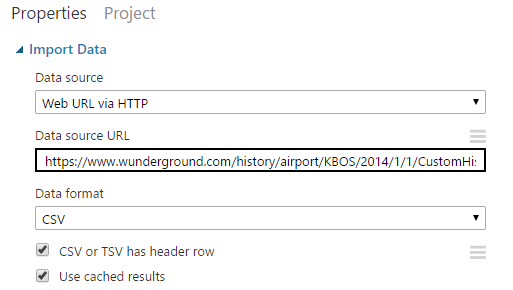


# Azure Experiment

Now that we Know how we can forecast the data for next 10 days we will create an azure experiment which gets the forecast.

## 4.1 Data collection

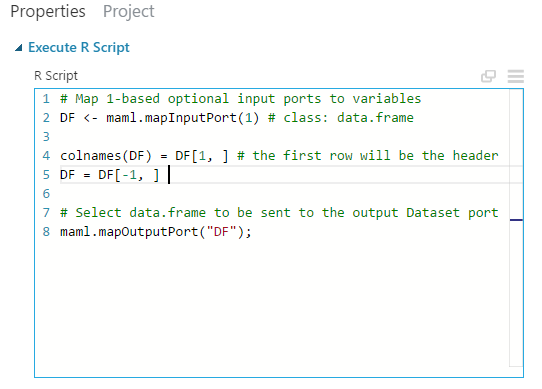
The data is collected in Azure from the URL, the past year data is stored in cache and the new data is not stored in cache and hence every time we call the model we get the latest data.



## 4.2 Data Cleaning

**4.2.1 Add all 9 years of data using ADD ROW module**

**4.2.2 Make the first column the header of data**

****

**4.2.3 Run Rscript to complete the time series & remove NA**

# Map 1-based optional input ports to variables

df1 <- maml.mapInputPort(1) # class: data.frame

library(dplyr)

library(stringi)

library(zoo)

df1$EST<-as.Date(df1$EST)

a<-Sys.Date()

ts <- seq.POSIXt(as.POSIXct("2008-01-01",'%m/%d/%y'), as.POSIXct(a,'%m/%d/%y'), by="day")

ts <- seq.POSIXt(as.POSIXlt("2008-01-01"), as.POSIXlt(a),by="d")

ts <- format.POSIXct(ts)#,'%y-%m-%d')

df <- data.frame(timestamp=ts)

df$timestamp<-as.Date(df$timestamp)

df1$timestamp<-df1$EST

data<- merge(df,df1,by="timestamp",all.x = T)

data1<-na.locf(data)

# extract the numbers:

nums <- stri\_extract\_all\_regex(data1$WindDirDegrees.br..., "[0-9]+")

# Make vector and get unique numbers:

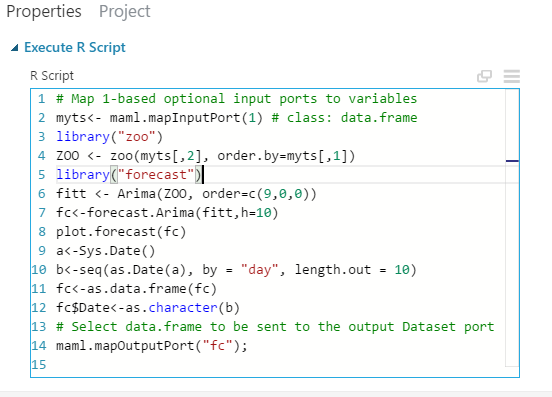
nums <- unlist(nums)

data1$WindDirDegrees.br...<-nums

# Select data.frame to be sent to the output Dataset port

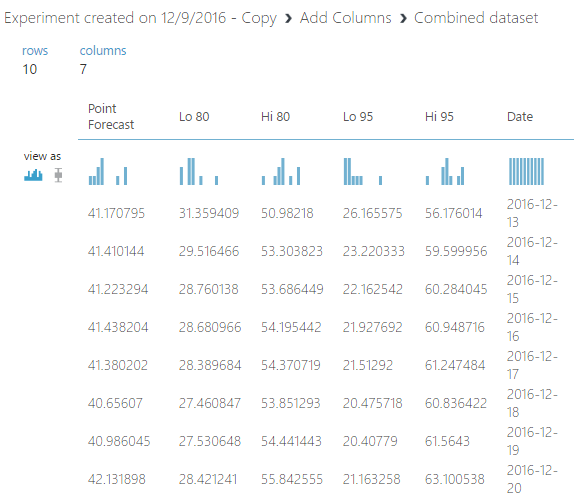
maml.mapOutputPort("data1");

**4.3 Forecast the Temperature for next 10 days**

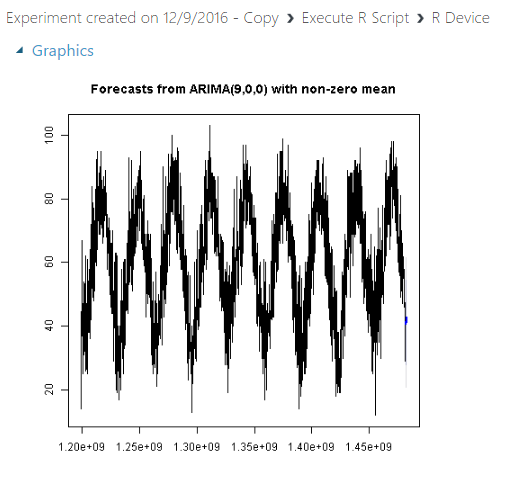
****

When checked for AIC ARIMA(9,0,0) has smaller AIC than ARIMA(5,0,0) so we have modelled it for ARIMA(9,0,0)

The forecast Data we get is:- (snap shot taken on 12 Dec 2016)

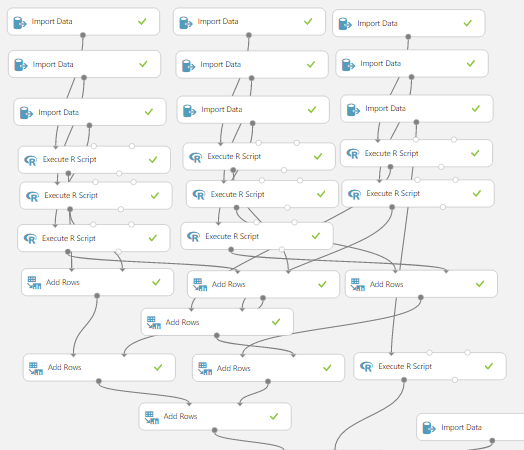


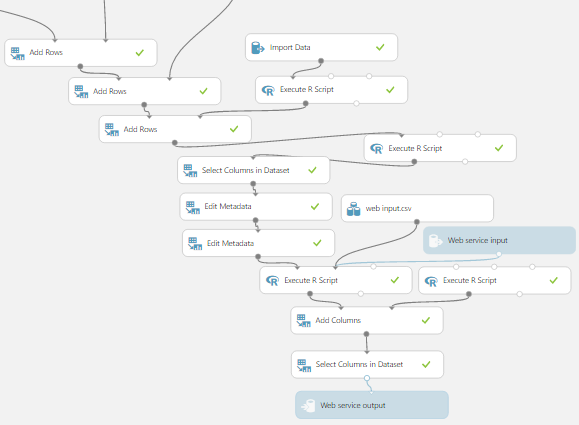
And the forecast graph on Azure is



The total Azure Experiment is below

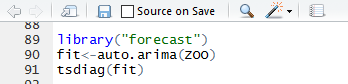
**Part 1**



**Part 2**

**4.3 Add other weather components**

Calculate the ARIMA(p,d,q) for rest of the weather components



|  |  |
| --- | --- |
| **Weather component** | **(p,d,q)** |
| Mean.TemperatureF | (9,0,0) |
| MeanDew.PointF | (1,0,1) |
| Mean.Humidity | (2,0,2) |
| Mean.VisibilityMiles | (0,1,2) |
| PrecipitationIn | (0,0,1) |
| CloudCover | (0,0,1) |
| Events | Cannot be computed |

Except Events add the above to the experiment

And like before with the Max. TemperatureF we forecast the weather for next 10 days.

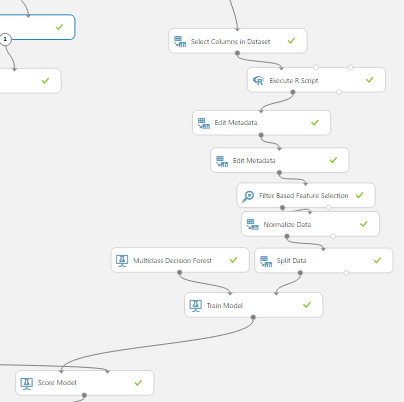


## 4.3 Compute the Events

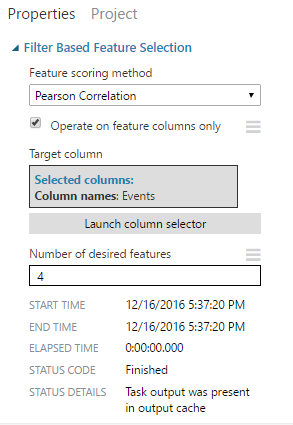
Events is the parameter of the weather which cannot be computed by using past year data. So we are going to perform multi-class classification for Events using dependent variables mentioned above

|  |  |  |
| --- | --- | --- |
| **Weather component** | **(p,d,q)** | **Variable Type** |
| Mean.TemperatureF | (9,0,0) | Independent |
| MeanDew.PointF | (1,0,1) | Independent |
| Mean.Humidity | (2,0,2) | Independent |
| Mean.VisibilityMiles | (0,1,2) | Independent |
| PrecipitationIn | (0,0,1) | Independent |
| CloudCover | (0,0,1) | Independent |
| Events | Perform Regression | Dependent |

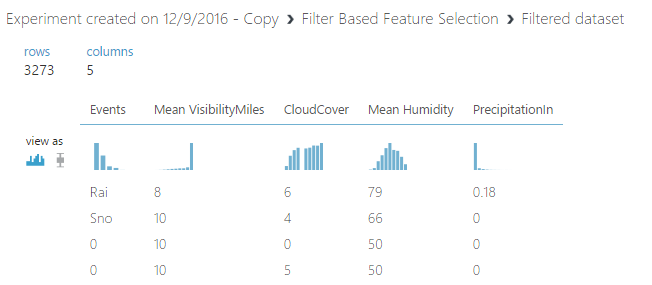
Now get the data from the Rscript and filter out only the above components. And do a multi-class regression.



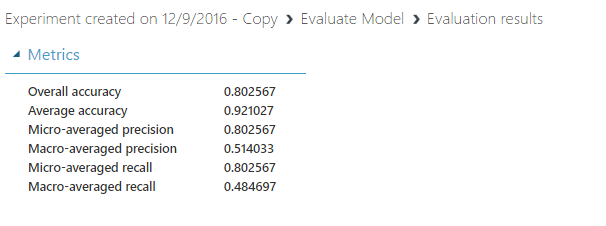
Perform Filter Based Feature Selection.



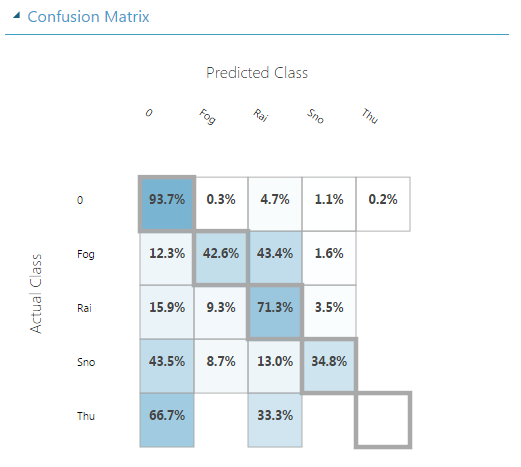
We get the following dependent variables which impact the Events



The Evaluated model has the following result.

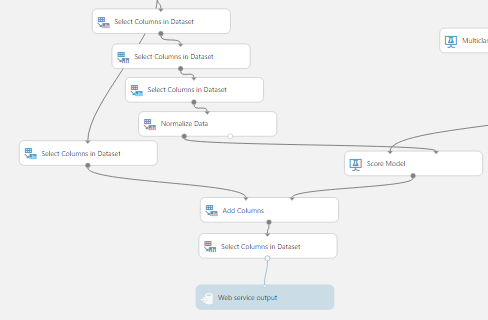


And a confusion matrix



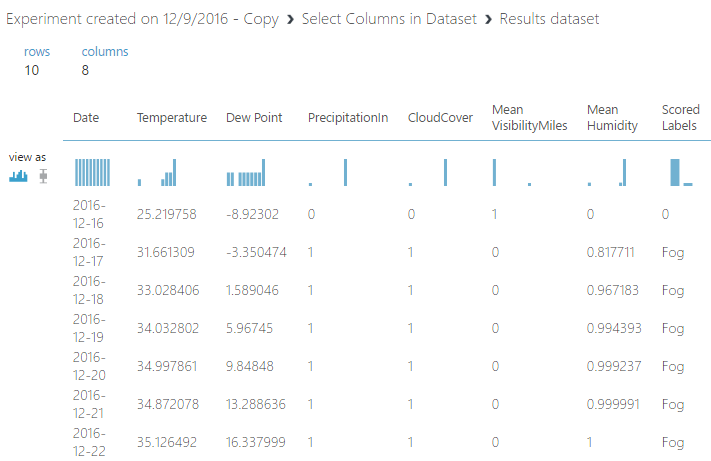
Once the confusion matrix is ready score the model.

## 4.4 Predict the ‘Events’ using dependent variable mentioned above.

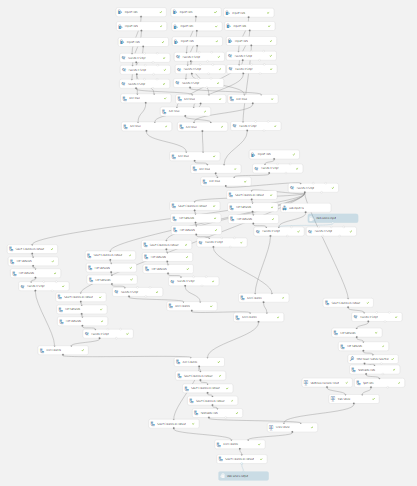


To the score model we will add the forecasted data and predict the Events.

We get the following results.



## 4.5 Create a Web Service for the above model

****

For reference login into azure as the pic is not clear

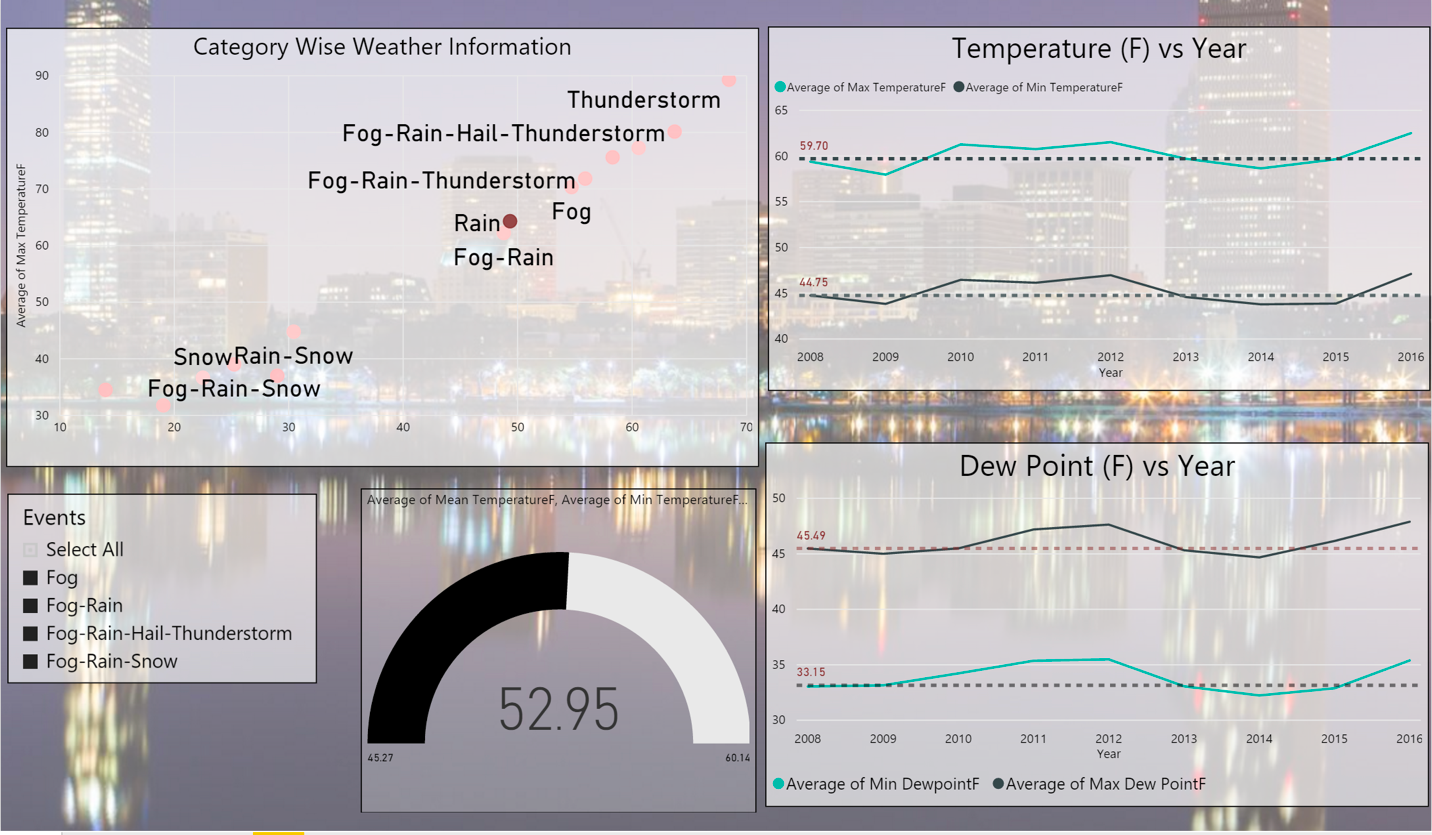
**Azure login**

**Login Id :** [**teameight2017@gmail.com**](mailto:teameight2017@gmail.com)

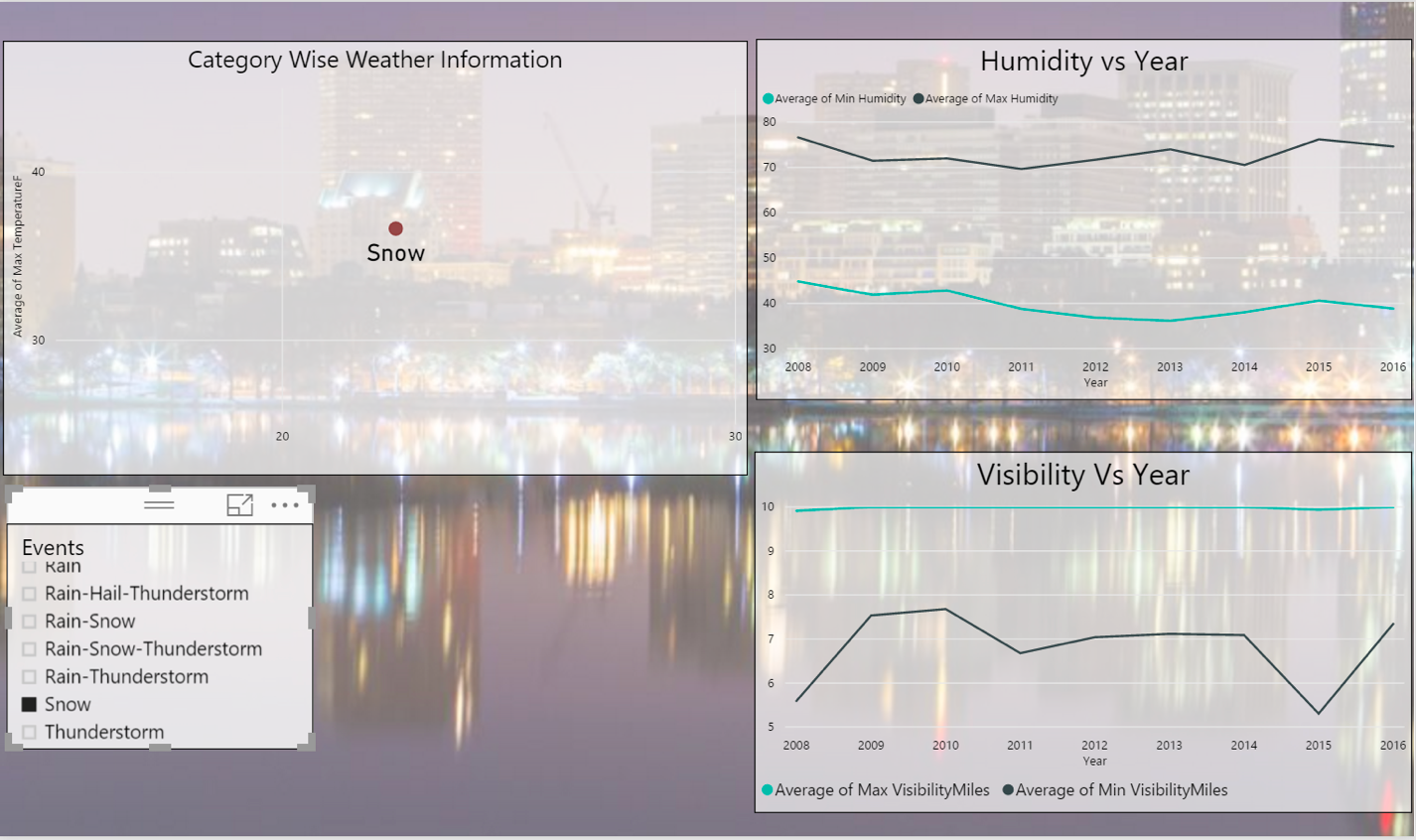
**Password : team8ads**

# 

# POWER BI:

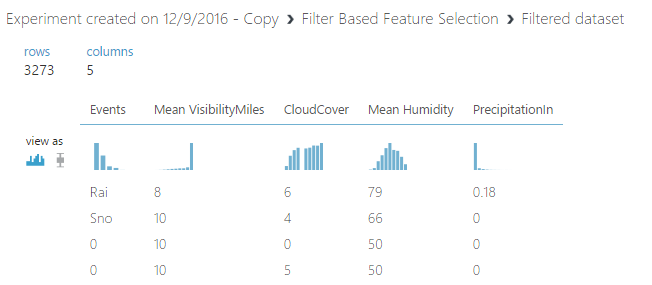


* The Above dashboard provides information about weather events occurring over span of 8 years and, trend of max. and min temperature, dew point, visibility recorded over the years.



* We can see that the visibility had dropped drastically in the year 2015 during the snow in Boston.

Using which we have decided that we can compute the “Events” using the other weather data as they had correlation between them.

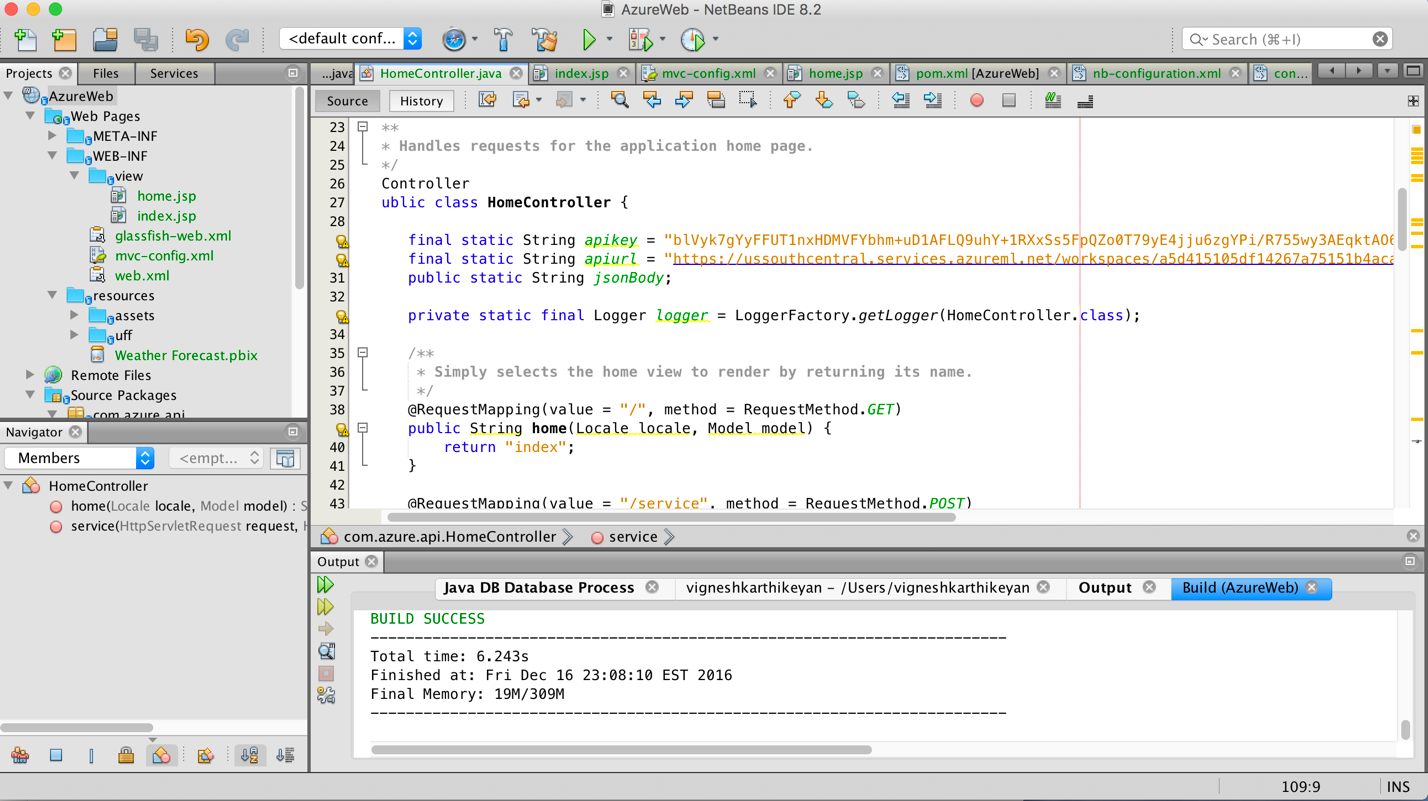


# Web Application:

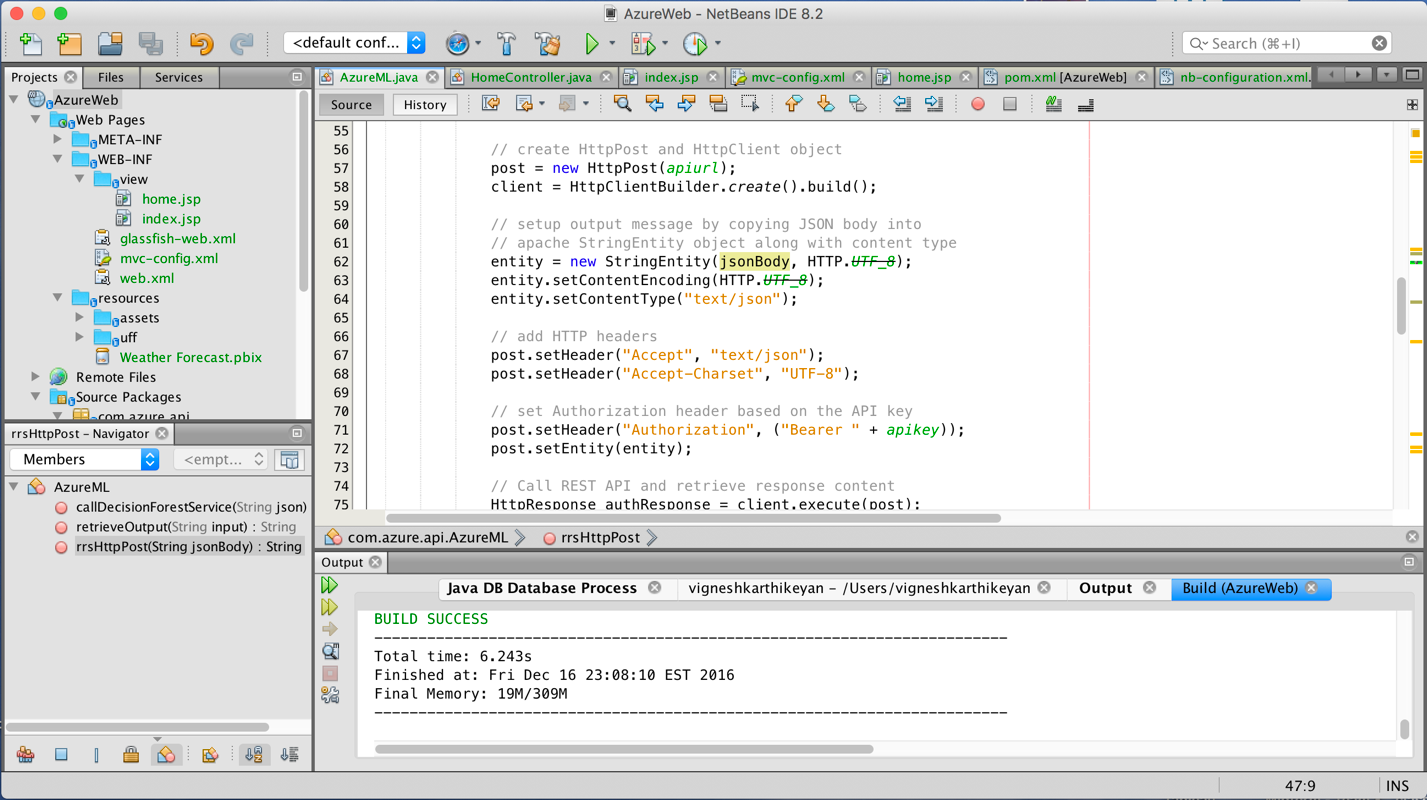
AWS Link: http://sample-env-1.qmntzmwe52.us-west-2.elasticbeanstalk.com

We created a MVC Java application on NetBeans and deployed the application on Amazon AWS.

Sample Code:



The Home Controller - where we make the API call



# Reference

* Extract Seasonal & Trend: using decomposition in R: <https://anomaly.io/seasonal-trend-decomposition-in-r/>
* Time Series Forecasting Theory | AR, MA, ARMA, ARIMA: <https://www.youtube.com/watch?v=Aw77aMLj9uM>
* Using R for Time Series Analysis: <http://a-little-book-of-r-for-time-series.readthedocs.io/en/latest/src/timeseries.html>
* https://www.otexts.org/fpp/8/8
* Detect Seasonality using Fourier Transform in R: <https://anomaly.io/detect-seasonality-using-fourier-transform-r/>
* Data set : <https://www.wunderground.com/>