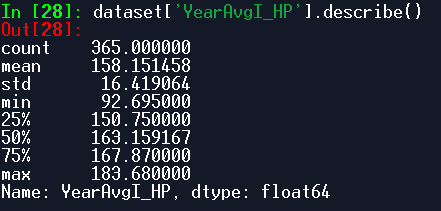
Methodology

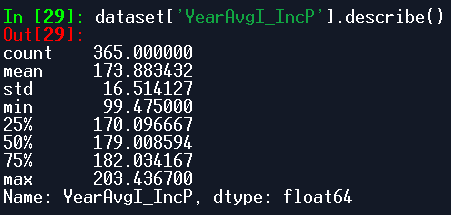
The Satellite Data collected is stored in the format as seen in the ’ Y\_Solar\_data ‘ file .

After some statistical studies on the dataset the descriptions of the values to be predicted are as shown below.

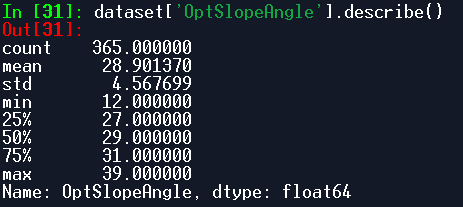
* YearAvgI\_HP (Yearly average of Irradiation on a horizontal Plane in KW/m^2)



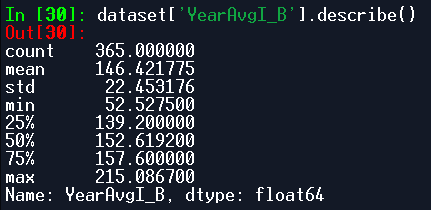
* YearAvgI\_IncP (Yearly average of Irradiation on optimally inclined Plane in KW/m^2)



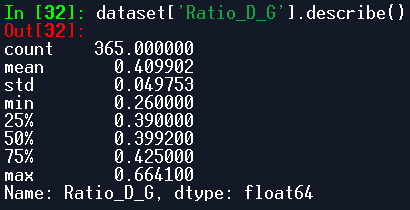
* OptSlopeAngle (Optimal Slope Angle in degrees)



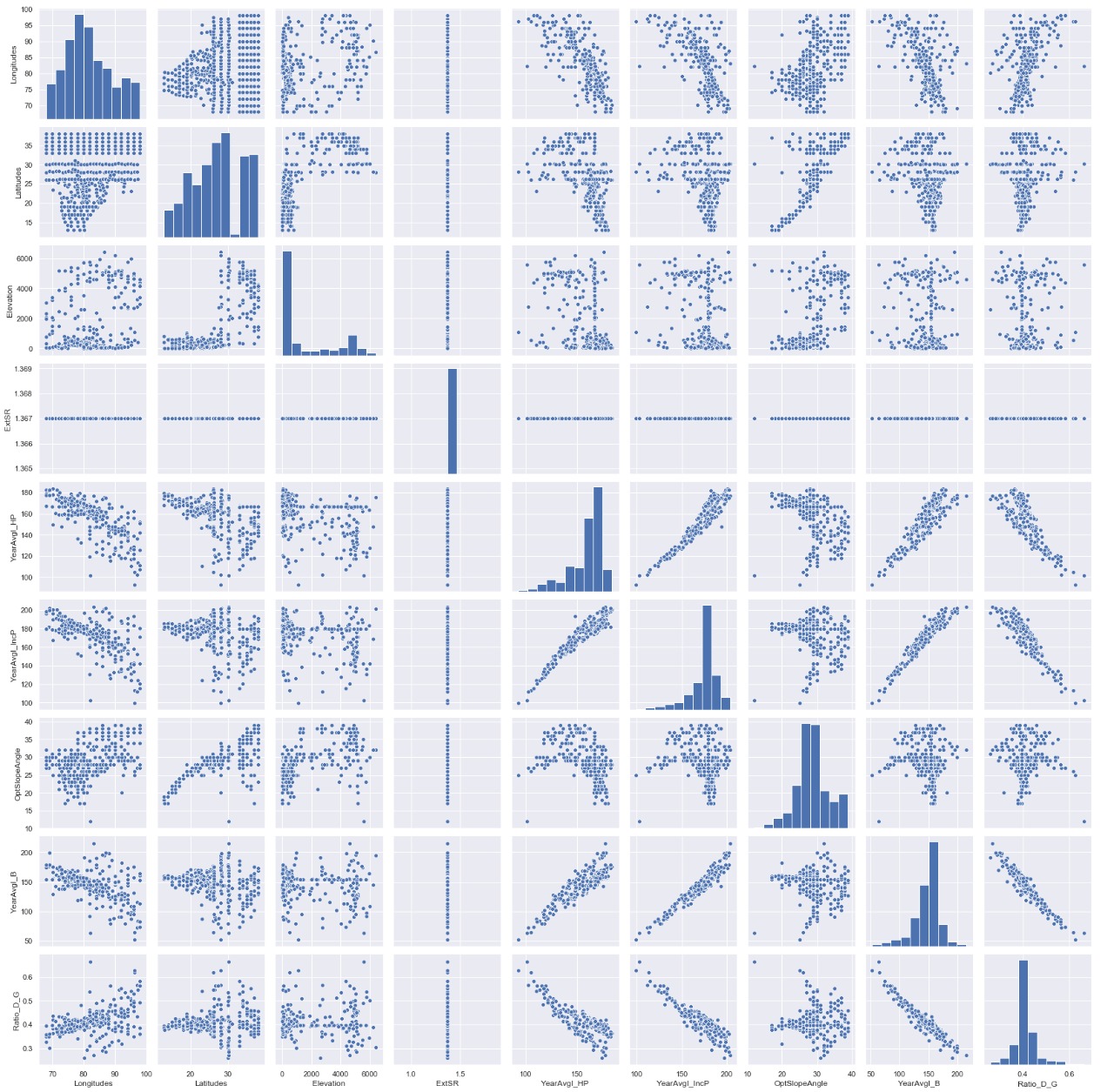
* YearAvgI\_B ( Yearly average of Beam Radiation in KW/m^2)



* Ratio\_D\_G (Diffuse Ratio)



The Pair Wise Plots of the Features is as shown below:



From the above plot the key information obtained :

1: All most all of the values of YearAvgI\_HP, YearAvgI\_IncP,YearAvgI\_B lie above a certain fix line and is seen to decrease with increasing Longitude.

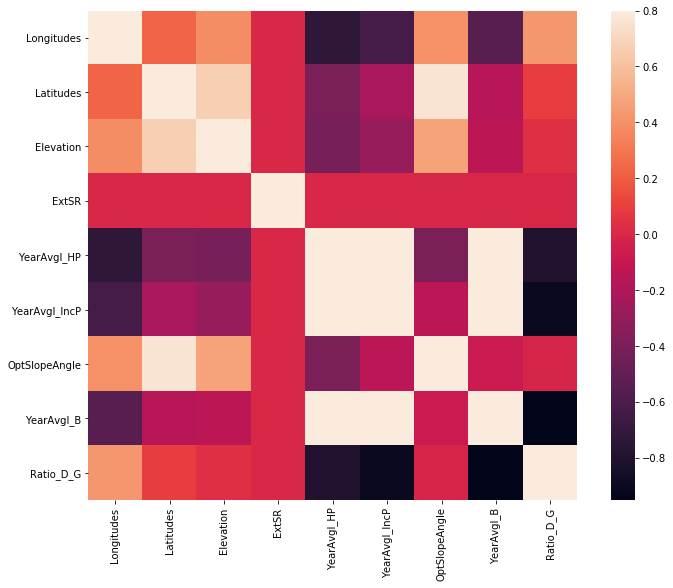
2: The Value of Ratio\_D\_G has all of its value below a certain fix line and also shows linear dependency on longitude where the value of Ratio\_D\_G seems to increase with increasing Longitude.

3: The OptSlopeAngle seems to have a linear Dependency on Latitude where the value of OptSlopeAngle seems to increase with increasing latitude.

4: The YearAvgI\_B and the YearAvgI\_IncP have a linear dependency on YearAvgI\_HP .

5:The value of Ratio\_D\_G decreases with increasing values of YearAvgI\_HP, YearAvgI\_B and YearAvgI\_IncP.

Studying the Correlation Matrix :



The above figure is the correlation matrix where white colour has the value 1. Gives the information about the correlation among the feature variables.

This Shows that the correlation of OptSlopeAngle has high correlation with Longitude,latitude,anf elevation.

The correlation value for YearAvgI\_HP, YearAvgI\_IncP and YearAvgI\_B is also high.

There is a significant correlation value for Ratio\_D\_G and Longitude.

The Code file for these plot :p1.py

Model Building:

Depending upon the above observations –

A linear model was choosed for prediction of the Output Parameters (YearAngI\_HP,YearAvgI\_IncP,YearAvgI\_B,OptSlopeAngle and Ratio\_D\_G).

A linear Regression model (regressor1 as used in model code) was used to predict the value of YearAvgI\_HP.

The model was provides with the values of longitude,latitude , elevation, and ExtSR (exterior solar radiation) as the input parameters and the YearAvgI\_HP as the target value to learn a relation between them.

A Linear Regression Model Uses the input parameters and predicts the output parameters according to the following equation:

y = a1\*x1+a2\*x2+a3\*x3+a4\*x4………

where y stands for the output parameter and xi ( i =1,2,3…) stands or input features.

ai (i = 1,2,3….) are called the weights.

The model begins with very small value of weights and during the training the weights a accordingly updated to give the best possible values for predicting the output.

Similarly the OptSlopeAngle was predicted using the longitude, latitude, elevation and ExtSR as the input features using Linear Regression (regressor3 as used in the model code)

Now due to significant dependency of YearAvgI\_IncP , YearAvgI\_B on the values on YearAvgI\_HP the predicted value of YearAvgI\_HP was also included to predict the remaining values.

While predicting the values of YearAvgI\_IncP the Input features were Latitude, longitude , elevation , ExtSR and the YearAvgI\_HP . A similar Linear Regression model (regressor2 as used in the model code)

For the prediction of YearAvg\_B, predicted valus of both, YearAvgI\_HP and YearAvgI\_IncP were used as input features along with longitude, latitude, elevation and ExtSR. Linear Regression model was used (regressor4 as used in the model code)

For the prediction of Ratio\_D\_G , Linear Regression model(regressor5 as used in the model code) was used. And the corresponding input features all the predicted values of YearAvgI\_HP, YearAvgI\_IncP, YearAvgI\_B were also used.

Code File : YearlyAvg.py

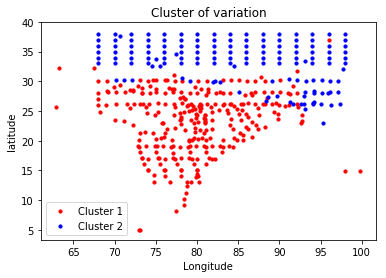
Studying the Month Wise data:

The trends of variation the values of MonthAvgI\_HP (Monthly Irradiation received on a horizontal Plane) were observed.

A K-means Clustering Algorithm was used to find if some sets of points showed similar trends in variation in the value of MonthAvg\_HP.

Code file: Clusters.py

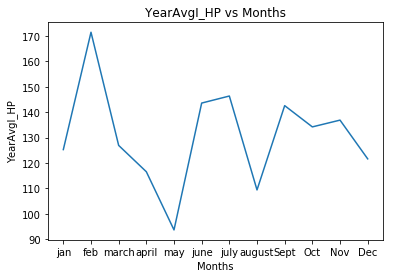
Based on the results the complete dataset could be divided into two clusters as shown:



And the complete dataset was divided into two parts to obtain the pattern of variation.The variation of MonthAvgI\_HP for different months in cluster1 was more or less observed to be same as

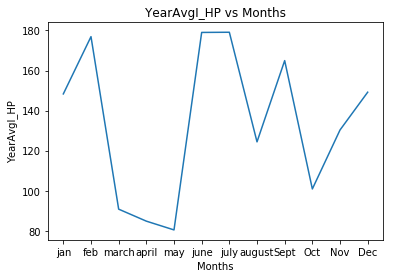
The variation plots are as shown below:

**MonthAvgI\_HP vs Months**



For Cluster2 the patterns observed was

**MonthAvgI\_HP vs Months**



As there is a linear correlation between MonthAvgI\_HP, MonthAvgI\_B, MonthAvgI\_IncP the similar pattern was observed.

Model Code:

First a different dataset was created unlike Y\_Solar\_data the values of Month wise parameter values were kept intact and the dataset was named as M\_Solar\_data.

Another dataset was created named clusters which had the information of locations and their MonthAvgI\_HP value for all the months. The columns of this dataset were: Longitude, Latitude, Elevation, ExtSR, Jan, Feb, March, April, May, June, July, August, September, October, November, December

Using the K-means Clustering on array of Columns [Jan, Feb, March, April, May, June, July, August, September, October, November, December] we got the visual of the dataset having two clusters.

Then the datapoints were divided into two parts referred as dataframe1 and dataframe2 in model code.

Code File: Clusters.py

Then using months.py model, dataframes being its input dataset the dataframes were further divided into specific months wise dataset as jan\_data, feb\_data,march\_data, april\_data, may\_data,j une\_data, july\_data, aug\_data, sept\_data,oct\_data,nov\_data, dec\_data.

Then using mean function, the mean of the MonthAvgI\_HP for every month was calculated for both the clusters separately and the trends plot is shown above.

When a new unseen location is input then the corresponding cluster is calculated and then the corresponding dataframe(dataframe1/dataframe2) is used to study the month wise trend of variation of parameters on that location.

Code file preforming the task: newpoint.py

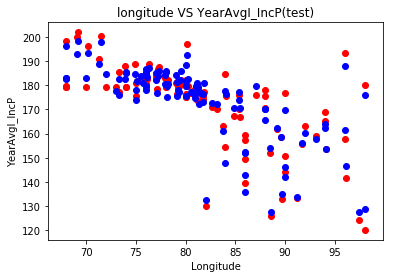
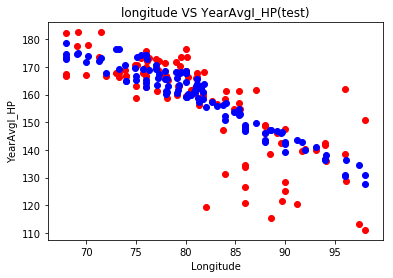
Using the above plots and the YearAvgI\_i(i=HP,IncP,B) , OptslopeAngle and Ratio\_D\_G the approx. values of monthly parameters can be estimated.

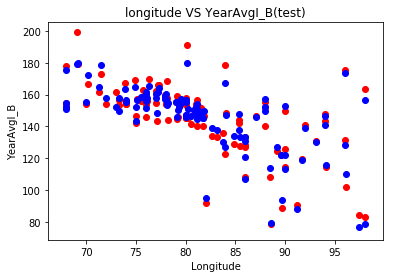
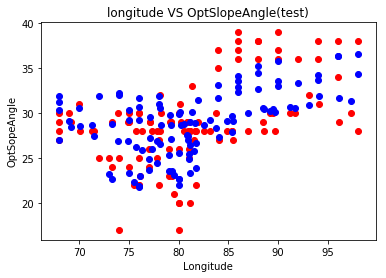
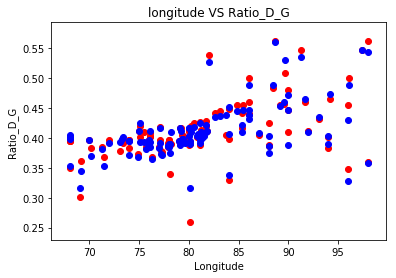
Running the predictions.py file will ask the user to input any new location coordinates (Longitude, Latitude, Elevation respectively in the order) and the output values of predicted parameters is stored in YearAvg variable.

Yearly value prediction plots:

Code File:YearlyAvg.py

The Output plots of the actual values and the predicted values are listed below.



Actual Values –Red Colour

Predicted Values – Blue Colour

The above plots display the results of prediction of the model for the Interested Yearly average output parameters showing variation along the longitude.

\*The Parameters can be changed from the code file to view any other pair of prediction

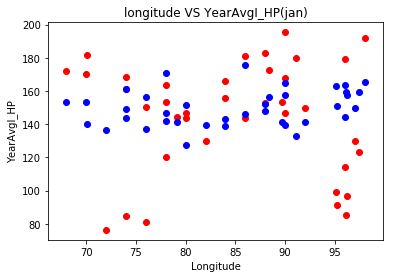
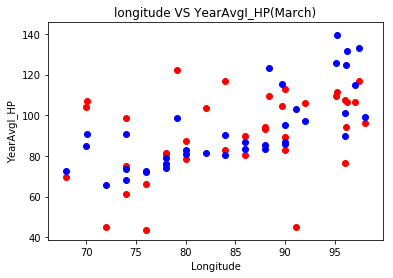
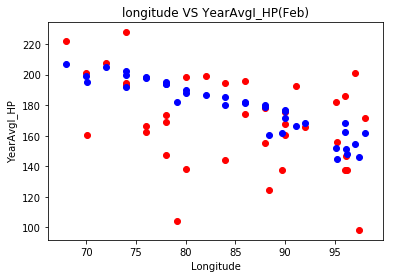
Monthly Value Prediction:

Code Files: jan\_model.py, feb\_model.py, march\_model.py, april\_model.py, may\_model.py, june\_model.py, july\_model.py, aug\_model.py, sept\_model.py, oct\_model.py, nov\_model.py, dec\_model.py

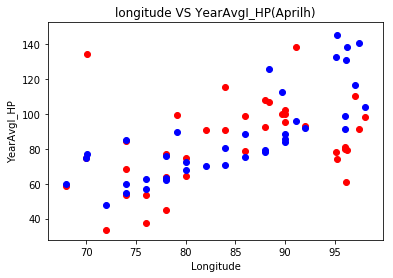
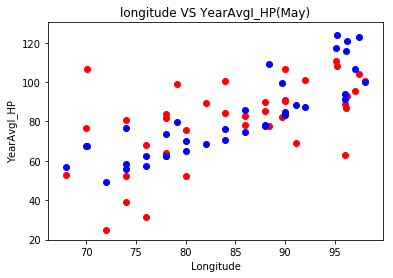
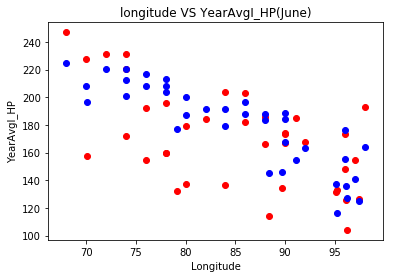
The Prediction of values for different months along the range of Longitudes for the two clusters is shown below

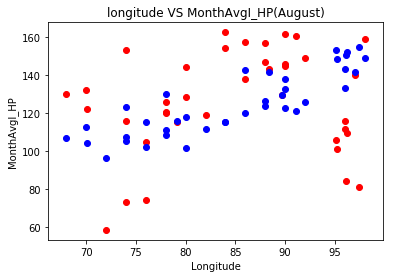
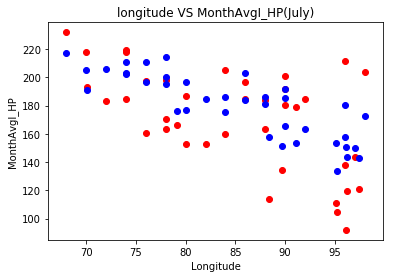
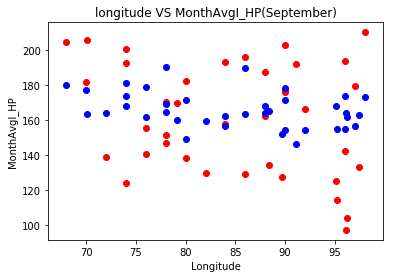
Region 2 plots displaying the actual and predicted values of MonthAvgI\_HP along longitude:

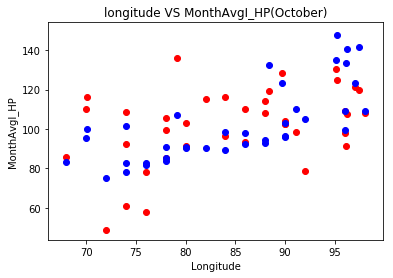
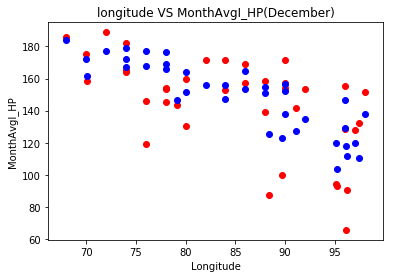
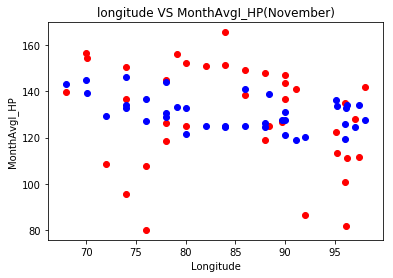
MonthAvgI\_HP vs Longitude (jan) MonthAvgI\_HP vs Longitude(feb) MonthAvgI\_HP vs Longitude(march)

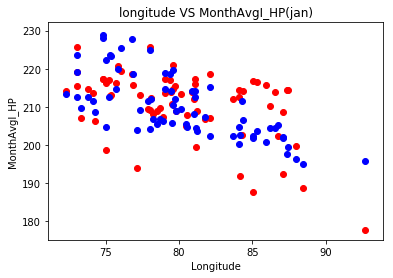
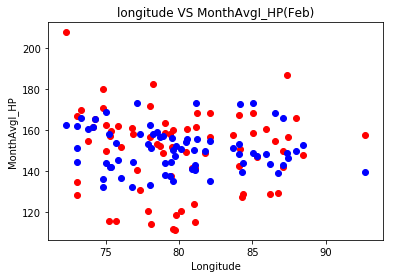
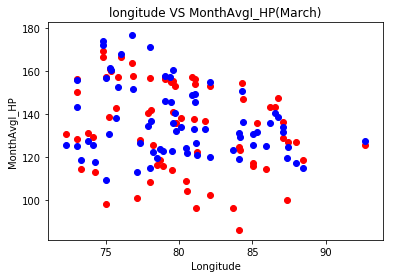
MonthAvgI\_HP vs Longitude(April) MonthAvgI\_HP vs Longitude(May) MonthAvgI\_HP vs Longitude(June)

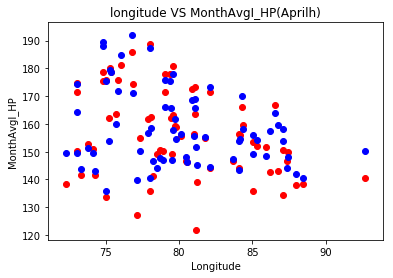
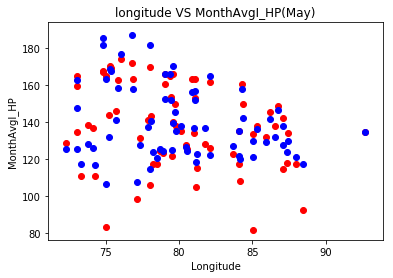
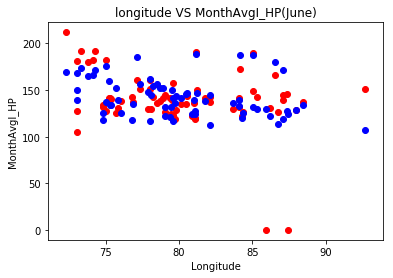
  

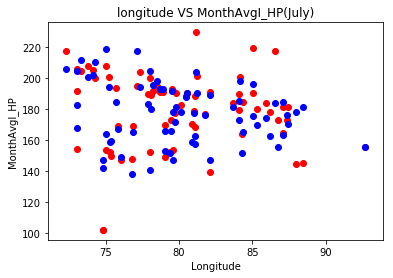
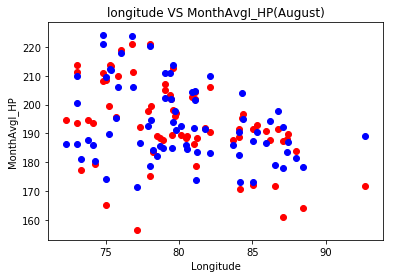
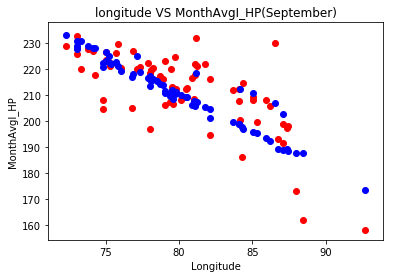
 

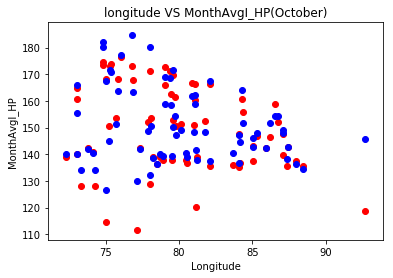
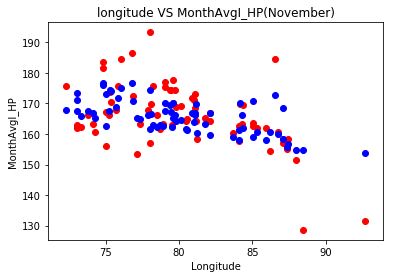
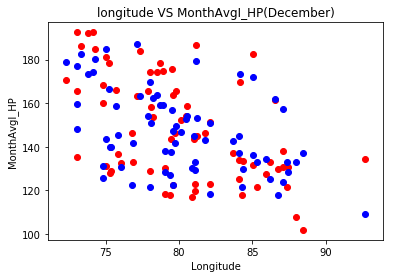
 

Similarly for Region 1:

In the above plots Blue points represent the predicted values and the red points represent the atual value.

For testing purpose a unseen data point (Longitude,Latitude,Elevation: 91.693,26.189,67) is a coordinate of Library Roof top IIT G . The model was used to predict the Yearly average values the result obtained is as shown:

|  |  |  |
| --- | --- | --- |
| Parameters | Actual value | Predicted value |
| YearAvgI\_HP | 139.6192 | 142.963 |
| YearAvgI\_IncP | 155.5275 | 159.844 |
| OptSlopeAngle | 30 | 30.634 |
| YearAvgI\_B | 118.7192 | 124.218 |
| Ratio\_D\_G | 0.463333 | 0.454 |

Both the files are Uploaded with this file.

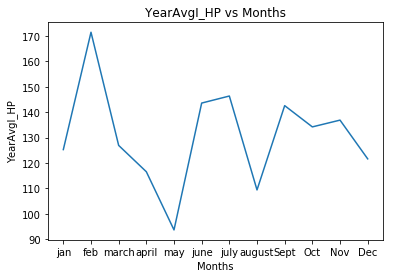
Conclusion:

Using the Model , user can predict Avg value of solar Irradiation parameters for a year of any specific location in India. Only for India, as the datset consisted of data points from different locations in India.

Also after Finding out in which region(cluster) the location belongs to , one can also get a approximate information of the trend of varaion in the data along the months.

For eg. The location used above (Longitude,Latitude,Elevation: 91.693,26.189,67) , this location belongs to region 1. So using the MonthAvgI\_HP plot for region 1 as shown below and using the predicted value of YearAvgI\_HP one can obtain the information that The Solar irradiation received on a horizontal surface will have a value less than the average in the month of January and the value will go on to increase more than the average in the month of February and then the value will continue to decrease till the month of May and again increase to approx. value of the average in the month of June and so on.

This information can help the engineer do plan accordingly the use of power generated throughout the year.



\*The scale in the figure can be adjusted according to the Average value obtained at that location.

The horizontal line in the figure represent the YearAvgI\_HP value predicted using the model and the variation can be studied.

The Prediction value of the individual month can also be achieved, the amount of data required will be huge. As depending upon the data of 450 points across India and some neighbouring region the variation pattern was divided into two regions (cluster1 and cluster2) and the complete data was divided into two parts reducing the datapoints belonging to each clusters. Hence the learning algorithm was not able to learn the exact relation but could only give some tentative value.