I

PREDICTIVE ANALYSIS ON WEATHER IN SZEGED (2006-2016)

CASE STUDY BY ALI AHMAD

CONTENTS

CASE STUDY

This dataset comprises the scrapped information about the factors which have an impact on the Temperature in Hungary. It includes an hourly/daily summary for Szeged, Hungary area, between 2006 and 2016. It consists of 12 variables with 96453 rows..

- Formatted
- Date Summary
- Precip Type
- Temperature (C)
- Apparent Temperature (C)
- Humidity
- Wind Speed (km/h)
- Wind Bearing (degrees)
- Visibility (km)
- Loud Cover
- Pressure (millibars)
- Daily Summary

IV

Objective

The objective here is to perform a Linear regression analysis to arrive at a model that can be used to predict the Temperature in Szeged (2006-2016) and find out the factors having an impact on the Temperature.

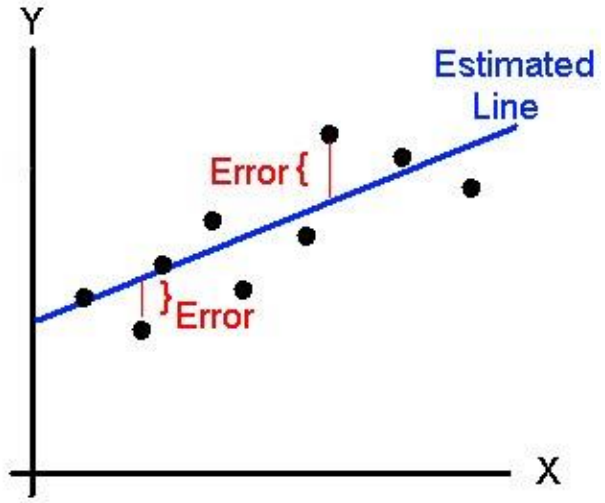
Estimated (or predicted) Y value for observation i

Estimate of the regression intercept

Estimate of the regression slope

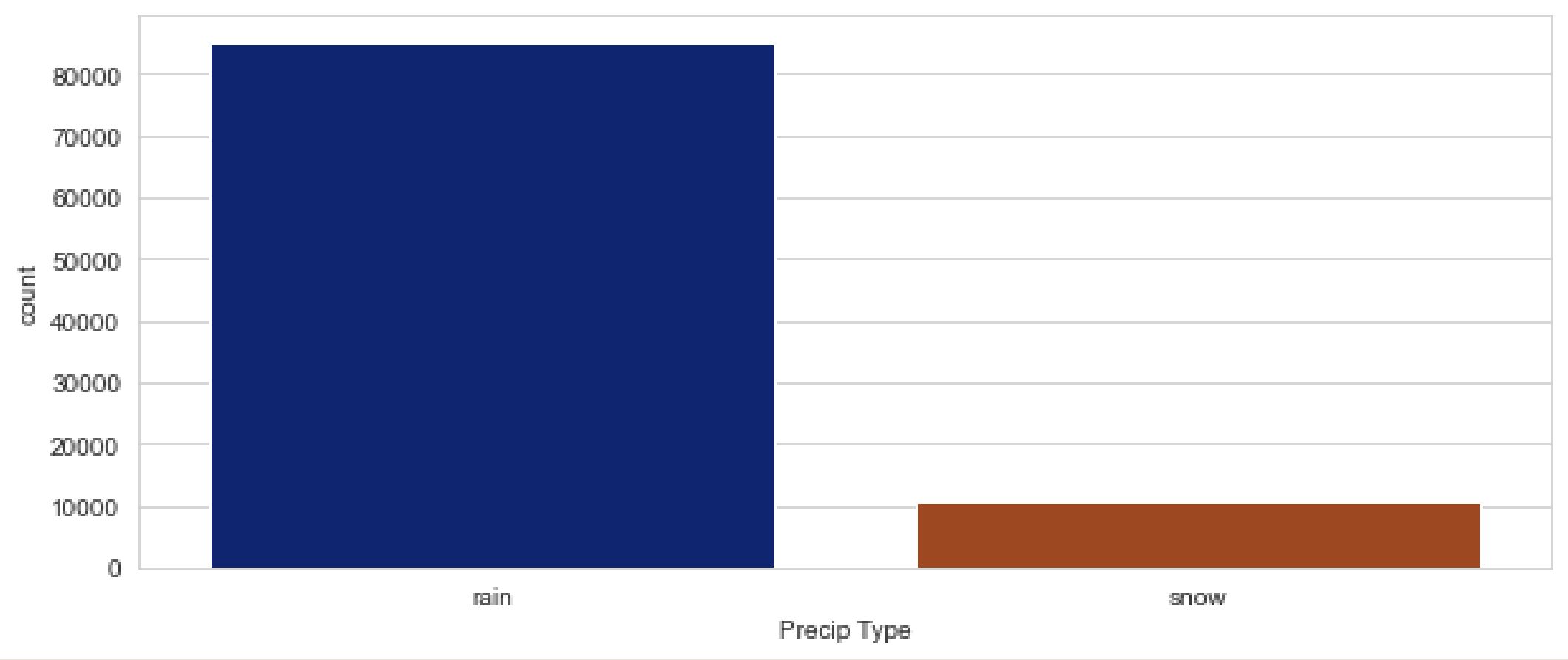
Value of X for observation i

$$\hat{Y}_i = b_0 + b_1 X_i$$

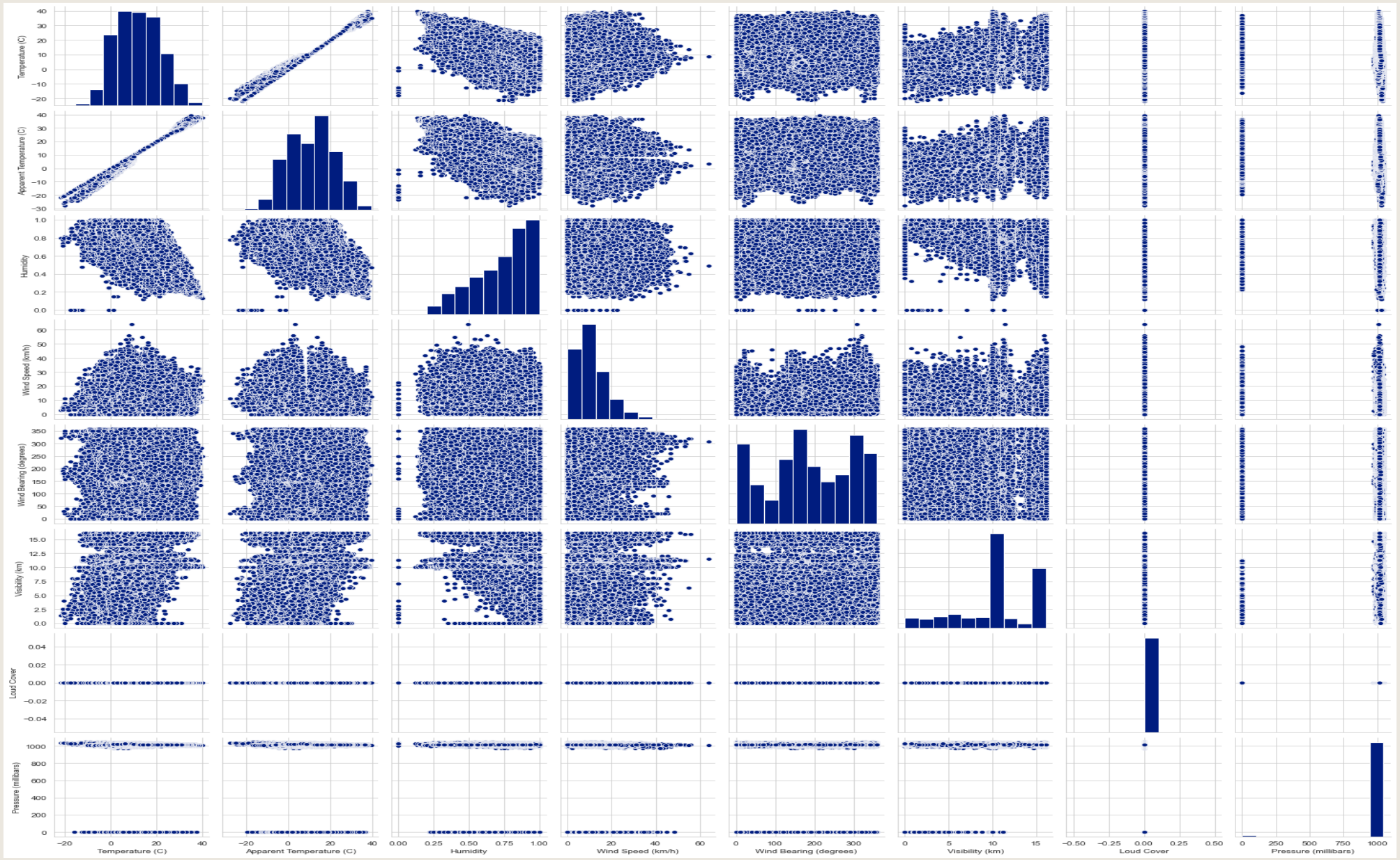


V

UNIVARIATE ANALYSIS

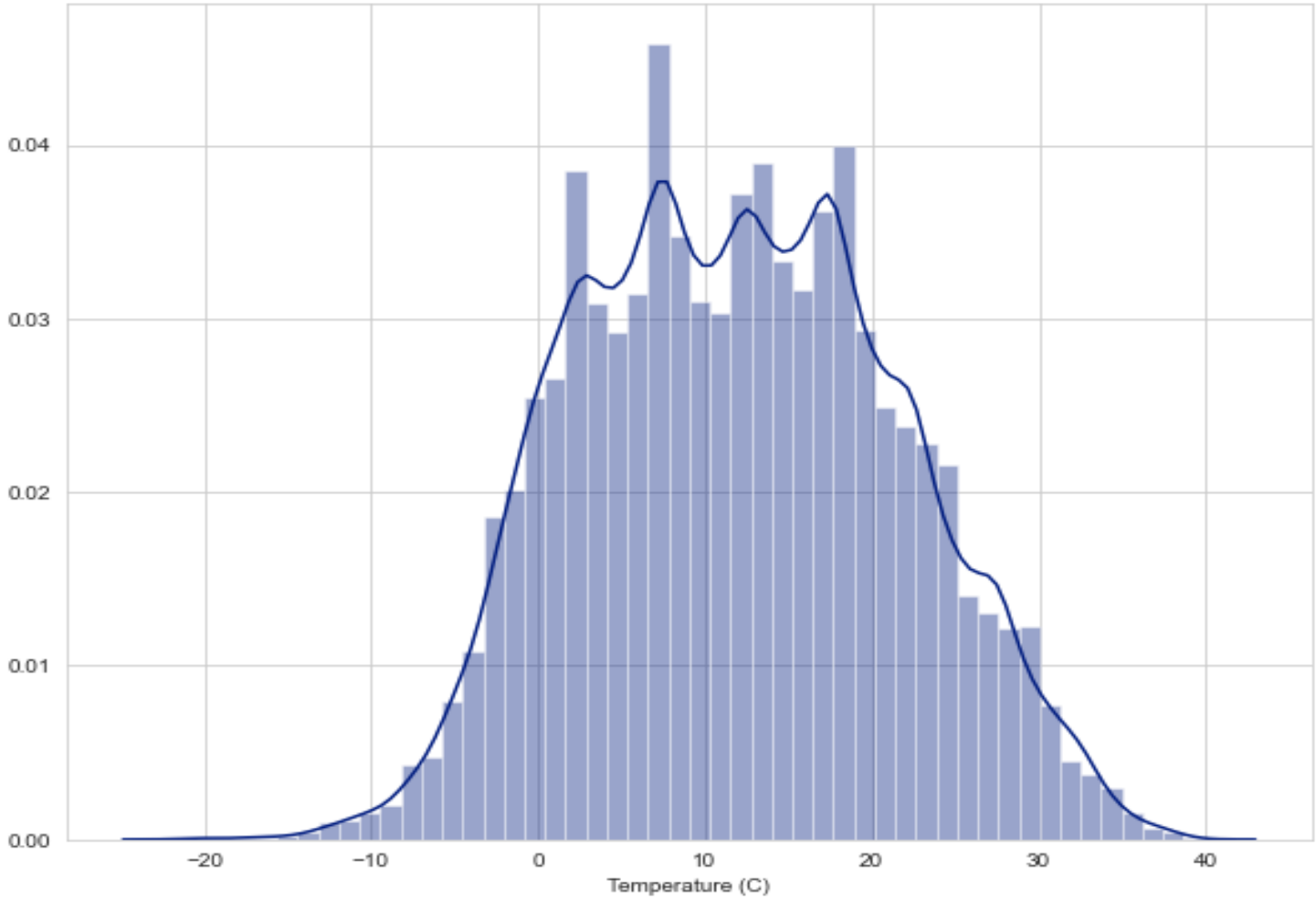


MULTIVARIATE ANALYSIS : Pairplot



VII

DENSITY DISTRIBUTION FOR TARGET VARIABLE

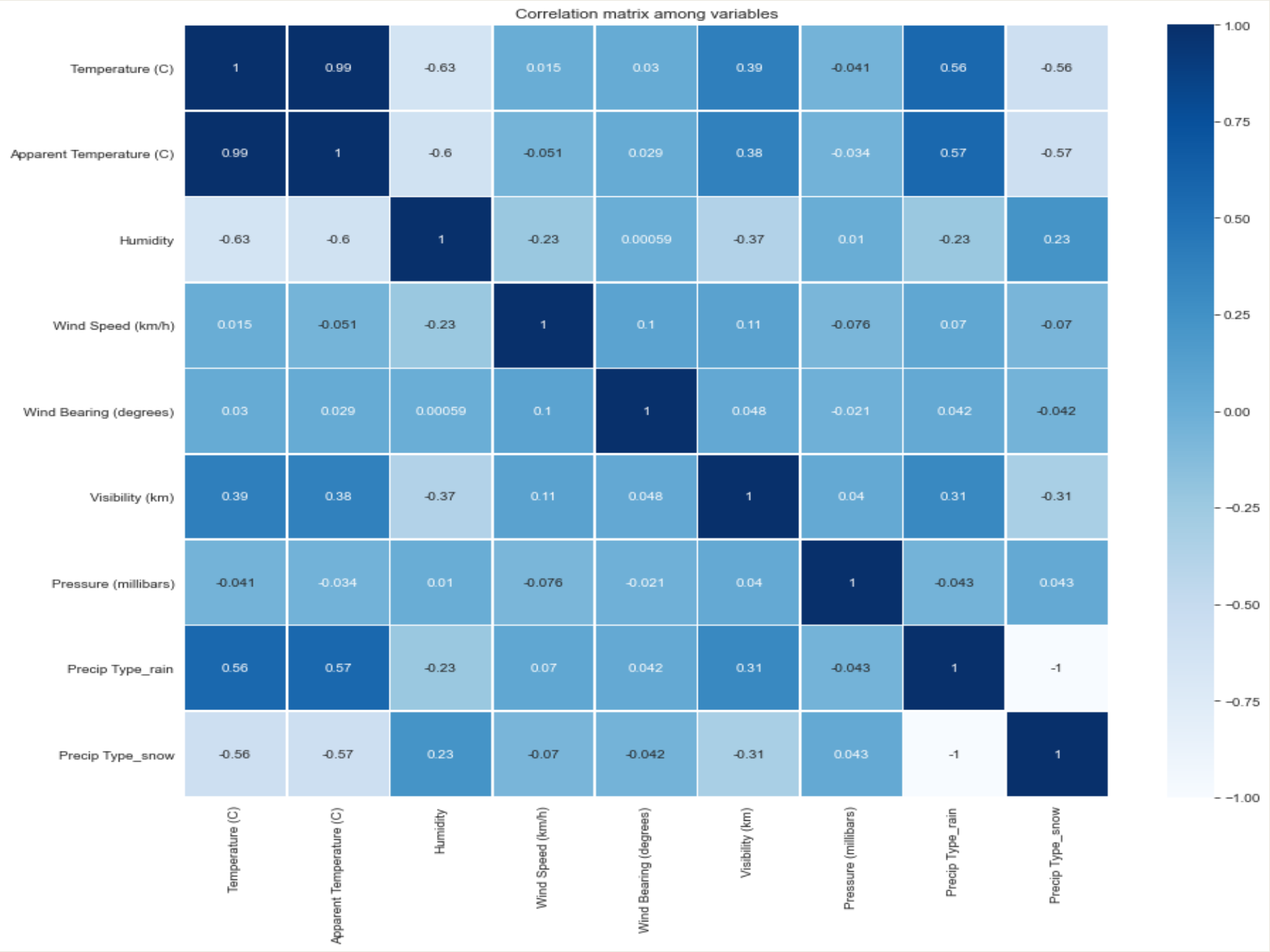


Correlation Matrix

	Temperature (C)	Apparent Temperature (C)	Humidity	Wind Speed (km/h)	Wind Bearing (degrees)	Visibility (km)	Pressure (millibars)	Precip Type_rain	Precip Type_snow
Temperature (C)	1.000000	0.992652	-0.633350	0.015442	0.029995	0.392771	-0.041252	0.562523	-0.562523
Apparent Temperature (C)	0.992652	1.000000	-0.603616	-0.050812	0.029005	0.381696	-0.033546	0.565130	-0.565130
Humidity	-0.633350	-0.603616	1.000000	-0.230628	0.000592	-0.369596	0.010005	-0.232782	0.232782
Wind Speed (km/h)	0.015442	-0.050812	-0.230628	1.000000	0.099943	0.105185	-0.075718	0.070080	-0.070080
Wind Bearing (degrees)	0.029995	0.029005	0.000592	0.099943	1.000000	0.047594	-0.020593	0.041984	-0.041984
Visibility (km)	0.392771	0.381696	-0.369596	0.105185	0.047594	1.000000	0.039725	0.312875	-0.312875
Pressure (millibars)	-0.041252	-0.033546	0.010005	-0.075718	-0.020593	0.039725	1.000000	-0.043474	0.043474
Precip Type_rain	0.562523	0.565130	-0.232782	0.070080	0.041984	0.312875	-0.043474	1.000000	-1.000000
Precip Type_snow	-0.562523	-0.565130	0.232782	-0.070080	-0.041984	-0.312875	0.043474	-1.000000	1.000000

IX

HEATMAP


Correlation matrix among variables
Temperature (C)
Apparent Temperature (C)
Humidity
Wind Speed (km/h)
Wind Bearing (degrees)
Visibility (km)
Pressure (millibars)
Precip Type_rain
Precip Type_snow
1 0.99 -0.63 0.015 0.03 0.39 -0.041 0.56 -0.56
0.99 1 -0.6 -0.051 0.029 0.38 -0.034 0.57 -0.57
-0.63 -0.6 1 -0.23 0.00059 -0.37 0.01 -0.23 0.23
0.015 -0.051 -0.23 1 0.1 0.11 -0.076 0.07 -0.07
0.03 0.029 0.00059 0.1 1 0.048 -0.021 0.042 -0.042
0.39 0.38 -0.37 0.11 0.048 1 0.04 0.31 -0.31
-0.041 -0.034 0.01 -0.076 -0.021 0.04 1 -0.043 0.043
0.56 0.57 -0.23 0.07 0.042 0.31 -0.043 1 -1
-0.56 -0.57 0.23 -0.07 -0.042 -0.31 0.043 -1 1
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

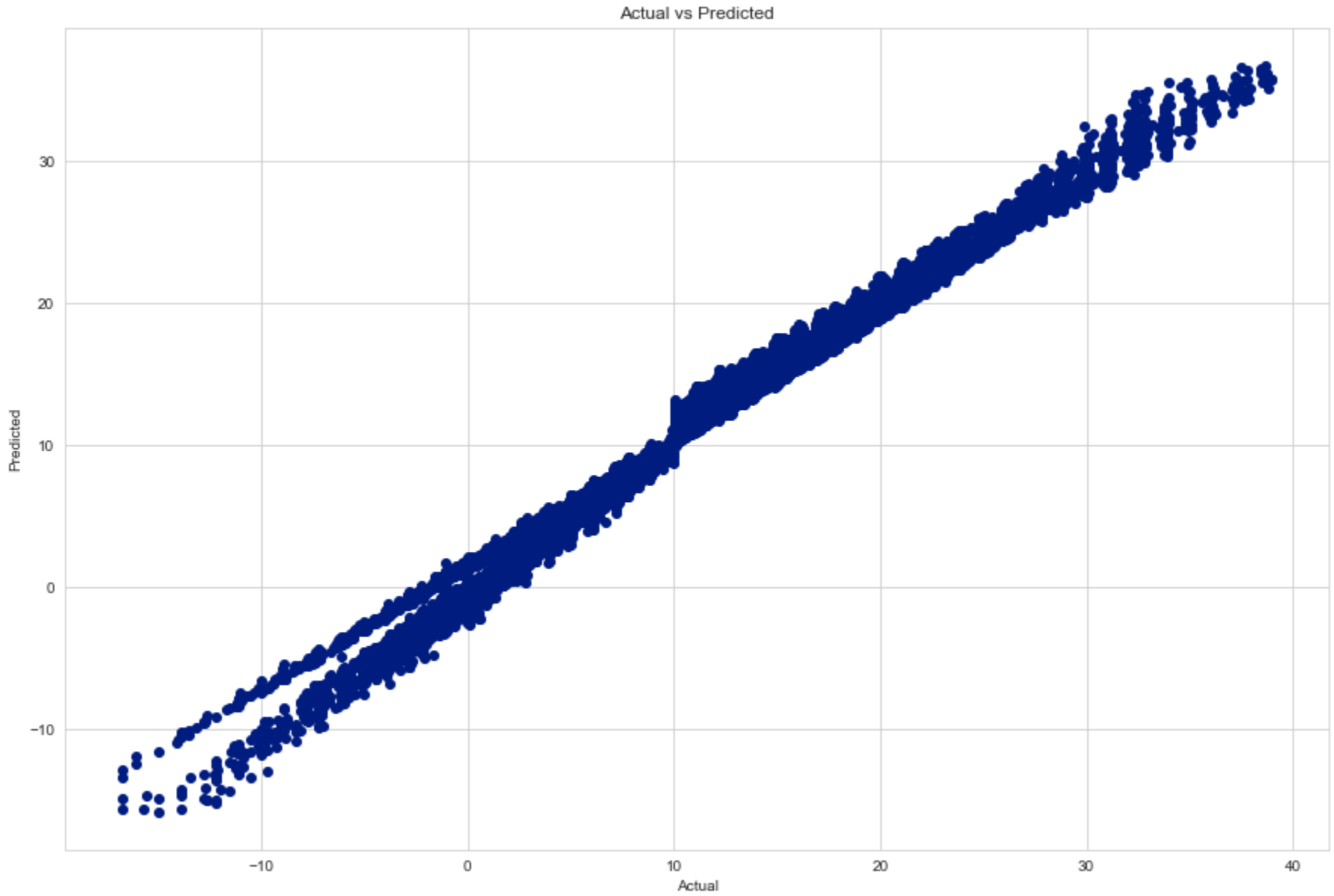
Model Prediction

Interpretation of actual value and the predicted value

	Actual value	Predicted value	Difference
6119	14.422222	14.172452	0.249771
51443	12.155556	13.496475	-1.340920
18754	23.961111	23.338621	0.622490
34070	15.000000	15.113622	-0.113622
26082	12.800000	13.138845	-0.338845
75095	11.088889	11.607233	-0.518345
29781	0.555556	-0.078981	0.634537
68559	12.733333	13.523960	-0.790626
71611	27.222222	26.636426	0.585796
62588	20.211111	19.428311	0.782800

XI

Plot Result : Actual Vs Predicted Data



XII

OBSERVATIONS

- R-square score on train data of Linear Regression : 0.9902
- R-square score on test data of Linear Regression : 0.9901
- Mean Absolute Error (MAE) of Linear Regression : 0.736
- Mean Squared Error (MSE) of Linear Regression : 0.889
- Root Mean Squared Error (RMSE) : 0.943

RECOMMENDATIONS

- When apparent temperature increases, it causes an increase of temperature. This is because apparent temperature has a high positive coefficient value given by the model.

- When humidity increases, it also causes an increase in the temperature. This is because humidity has a positive but low coefficient value given by the model

- When wind speed increases, it causes a decrease of the temperature. This is because wind speed has a negative coefficient value given by the model.

THANK YOU