ITS ML Challenge

RIGOROUS REGRESSION

BY - SLT VIGHNESH TIWARI

COURSE : 15 'X' IT

SCHOOL : INFORMATION TECHNOLOGY SCHOOL

Github Repository Link : https://github.com/halfbloodprince16/ITS-ML-Challenge

ITS ML Challenge

Problem Statement – I

The dataset consist of two column (X,Y), this is a univariate regression problem statement where model will train itself on X while predicting Y.

Since we have only one independent variable and one dependent variable, I will begin with my analysis.

ITS ML Challenge

Problem Statement I - (Data Description)

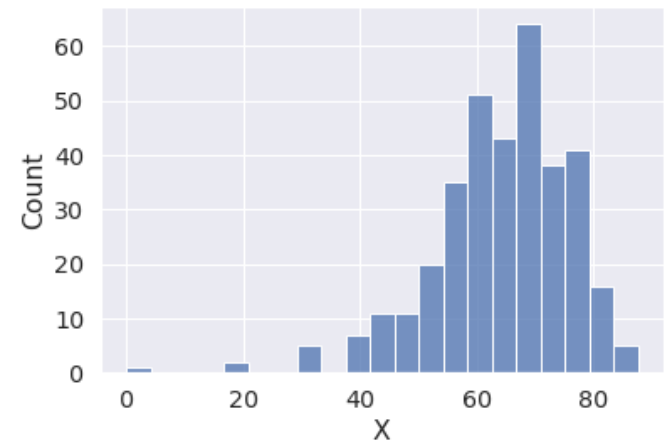
```
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   X       350 non-null    float64
 1   Y       350 non-null    float64
dtypes: float64(2)
memory usage: 5.6 KB
```

	X	Y
count	350.000000	350.000000
mean	63.912880	69.001553
std	11.855688	7.494282
min	0.000000	35.000000
25%	57.500000	65.511811
50%	65.039370	69.921260
75%	72.500000	73.385827
max	87.500000	90.000000

- Train set consist of 350 records.
- Test set consist 100 records.

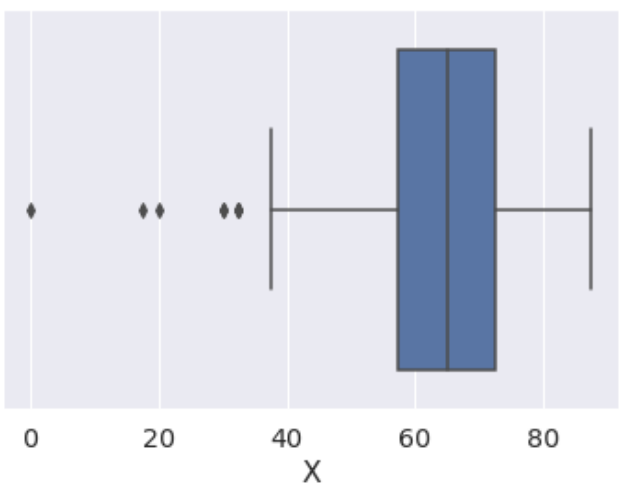
ITS ML Challenge

Problem Statement I - (Exploratory Data Analysis)



- Our dependent variable is clustered around values between 60-80.
- Very few values are around for X < 40.

- Another very smart representation of Box Plot in seaborn gives exact representation that values are clustered around 60-80.

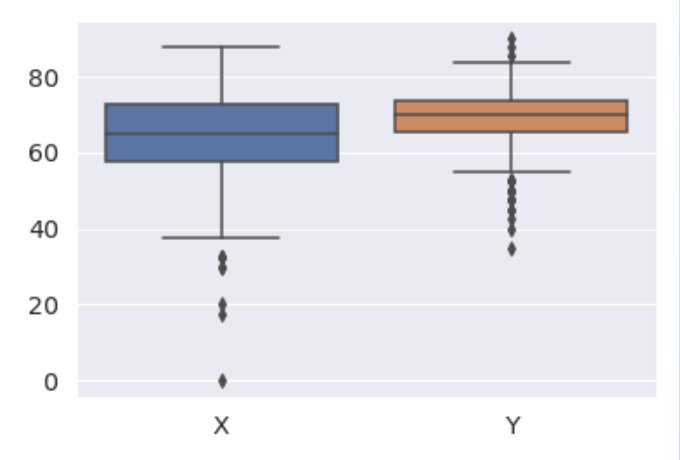


ITS ML Challenge

Problem Statement I - (Exploratory Data Analysis)

- The correlation between X and Y is around 0.058 which tells us that there correlation is quite useful.

- The Box Plot for X, Y also tells same that the values for X are correlated with Y.



ITS ML Challenge

Problem Statement I - (Model Validation Report)

Model Name	RMSE Score	Ranking	Code Notebook and Submission File
Linear Regression	8.086785	3	itsc-ml-P.ipynb Code Notebook P1.zip Submission File
Random Forest	8.024946	2	
LightGBM	8.217616	4	
GradientBoosting	9.897876	5	
RANSAC	7.824255	1	

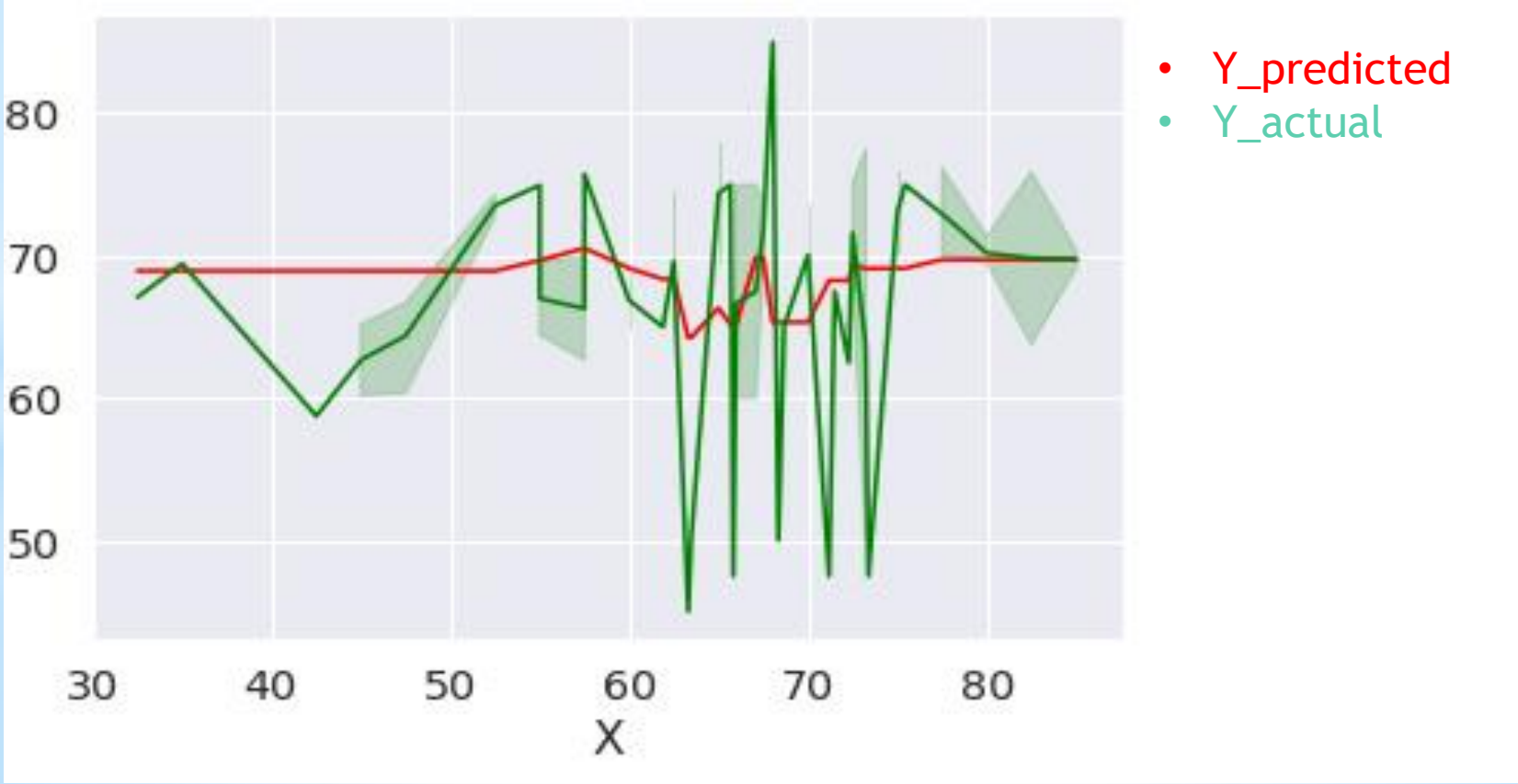
ITS ML Challenge

Problem Statement I - (Model Performance Report)

Model Name	RMSE Score	Ranking	Code Notebook and Submission File
Linear Regression	**7.695102**	3	itsc-ml-P.ipynb Code Notebook P1.zip Submission File
Random Forest	**7.697825**	4	
LightGBM	**7.523122**	1	
GradientBoosting	**8.191966**	5	
RANSAC	**7.635156**	2	

ITS ML Challenge

Problem Statement I – (Model Performance Report)



ITS ML Challenge

Problem Statement – II

The dataset consist of two column (X,Y), this is a univariate regression problem statement where model will train itself on X while predicting Y.

Since we have only one independent variable and one dependent variable, I will begin with my analysis.

ITS ML Challenge

Problem Statement II - (Data Description)

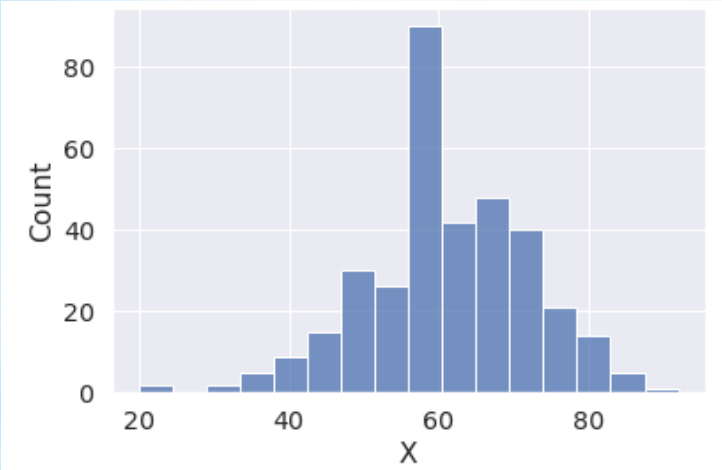
```
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   X       350 non-null    float64
 1   Y       350 non-null    float64
dtypes: float64(2)
memory usage: 5.6 KB
```

	X	Y
count	350.000000	350.000000
mean	61.080000	70.677355
std	11.408005	5.447000
min	20.000000	57.586207
25%	52.000000	67.825093
50%	60.000000	70.000000
75%	68.000000	73.793103
max	92.000000	87.878788

- Train set consist of 350 records.
- Test set consist 100 records.

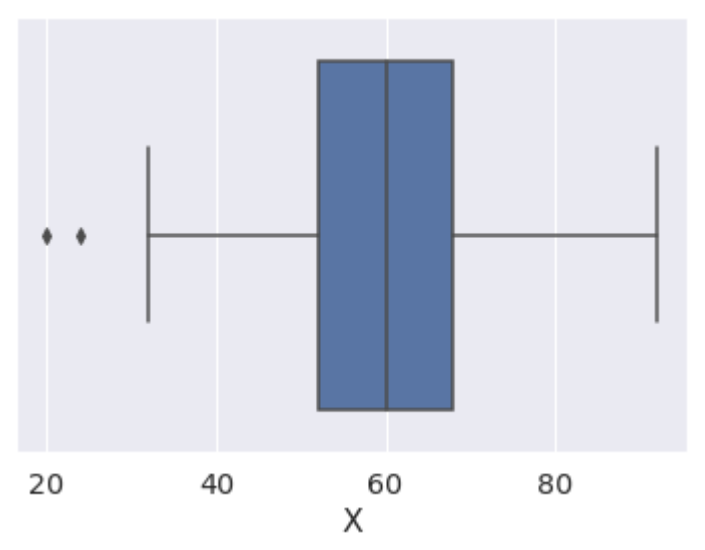
ITS ML Challenge

Problem Statement II - (Exploratory Data Analysis)



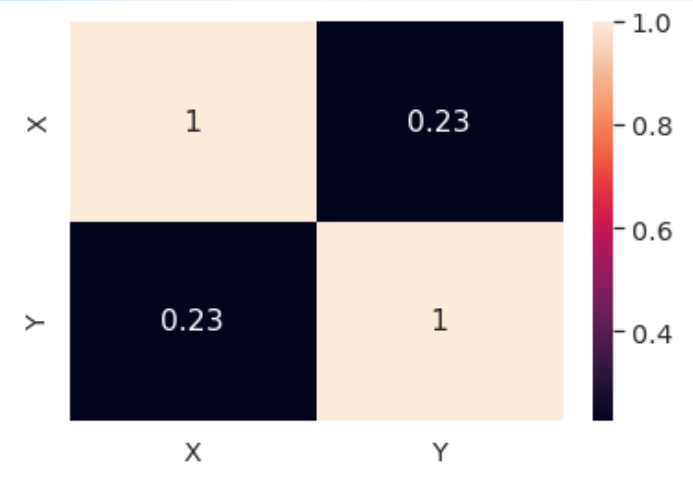
- Our dependent variable is clustered around values between 50-70.
- Very few values are around for X < 50.

- Another very smart representation of Box Plot in seaborn gives exact representation that values are clustered around 50-70.



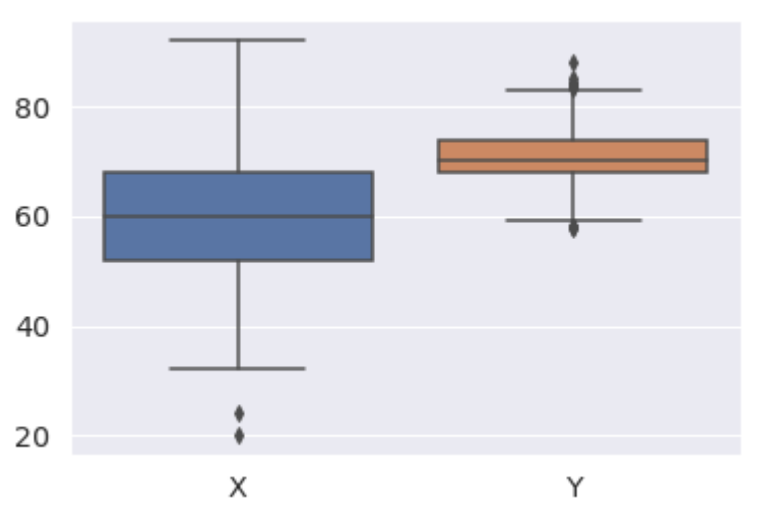
ITS ML Challenge

Problem Statement II - (Exploratory Data Analysis)



- The correlation between X and Y is around 0.23 which tells us that there correlation is quite useful.

- The Box Plot for X, Y also tells same that the values for X are near to correlated with Y.



ITS ML Challenge

Problem Statement II - (Model Validation Report)

Model Name	RMSE Score	Ranking	Code Notebook and Submission File
Linear Regression	5.265927	1	itsc-ml-R.ipynb Code Notebook P2.zip Submission File
Random Forest	5.348536	2	
LightGBM	5.492377	4	
GradientBoosting	5.610987	5	
RANSAC	5.374011	3	

ITS ML Challenge

Problem Statement II - (Model Performance Report)

Model Name	RMSE Score	Ranking	Code Notebook and Submission File
Linear Regression	**7.695102**	5	itsc-ml-R.ipynb Code Notebook P2.zip Submission File
Random Forest	**5.319863**	2	
LightGBM	**5.505367**	3	
GradientBoosting	**6.287577**	4	
RANSAC	**5.296517**	1	

ITS ML Challenge

Problem Statement II – (Model Performance Report)