Machine Learning Assignment 1

M22MA003

Bhawna Bhoria

Task 1: Simple Linear Regression

Steps Followed:

1. Imported the libraries
2. Import the input dataset

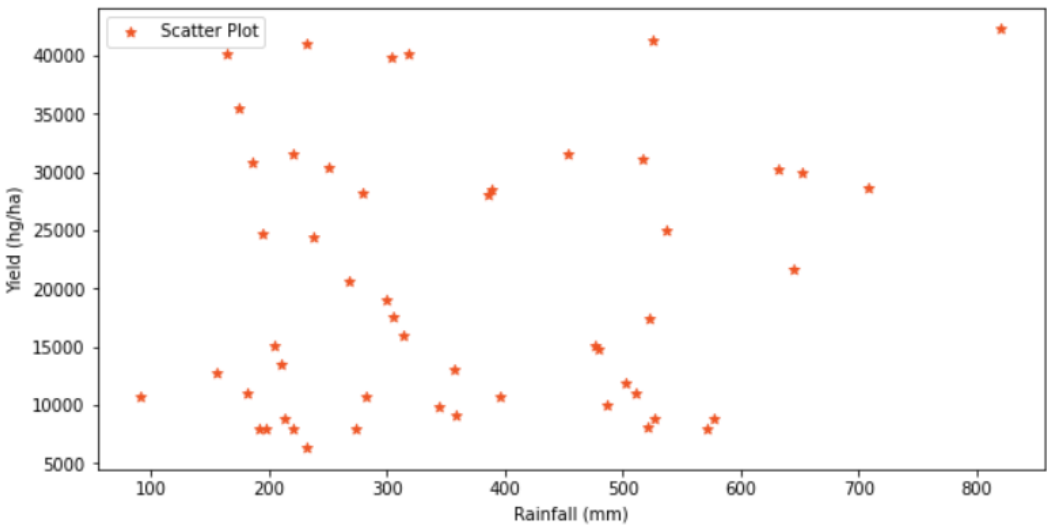
```
Saving Rainfall_dataset.csv to Rainfall_dataset (1).csv
(52, 5)
```

	Unnamed: 0	Year	Rainfall (mm)	Yield (hg/ha)	Remarks
0	0	1970	631.8	30197	Normal
1	1	1971	268.5	20698	Normal
2	2	1972	237.0	24388	Deficit
3	3	1973	651.8	29976	Excess
4	4	1974	194.4	24745	Deficit

---

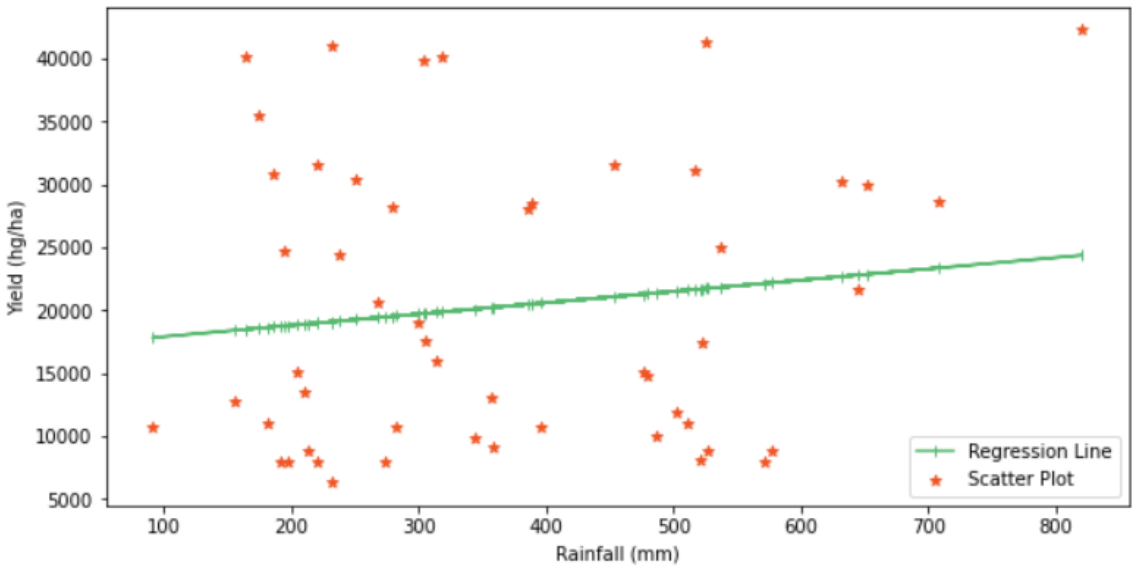
3. Identified independent and dependent variables
4. Obtained the value of mean value of variables
5. Draw the scatter plot

```
Mean value of Rainfall is  367.3807692307692
Mean value of Yield is     20330.51923076923
Value of coefficient of regresion is
8.972274380148036
b0 is 17034.278167240922
```



---

6. Obtained the coefficient of regression
7. Obtain the y_predicted values
8. Draw the regression line



9. Obtained the value of MAE and MSE
10. Predicted the value of crop yield for year 2022

```
MAE is 9764.615081494698
MSE is 122363336.42449355
Predicted crop yield for the year 2022 is  22058.75182012382
```

Colab link : https://colab.research.google.com/drive/1Wx_ROesJlFUUtjeM-XDRQ9kBB9BXyZBs#scrollTo=tuwK0bwI6J99

Reference : https://www.youtube.com/watch?v=E5RjzSK0fvY

Task 2: Multi Variant Regression

Steps Followed:

1. Imported the libraries
2. Cleaning the dataset to handle '?' missing data and outliers
3. Import the input dataset

```
Saving Ques_2.csv to Ques_2 (2).csv
(159, 26)
```

	symboling	normalized_losses	make	fuel_type	aspiration	num_of_doors	body_style	drive_wheels	engine_location	wheel_base	...	engine_size
3	2	164.0	audi	gas	std	four	sedan	fwd	front	99.8	...	109
4	2	164.0	audi	gas	std	four	sedan	4wd	front	99.4	...	136
6	1	158.0	audi	gas	std	four	sedan	fwd	front	105.8	...	136
8	1	158.0	audi	gas	turbo	four	sedan	fwd	front	105.8	...	131
10	2	192.0	bmw	gas	std	two	sedan	rwd	front	101.2	...	108

5 rows × 26 columns

4. Add Index column to the dataset to maximize the cardinality of dataset.

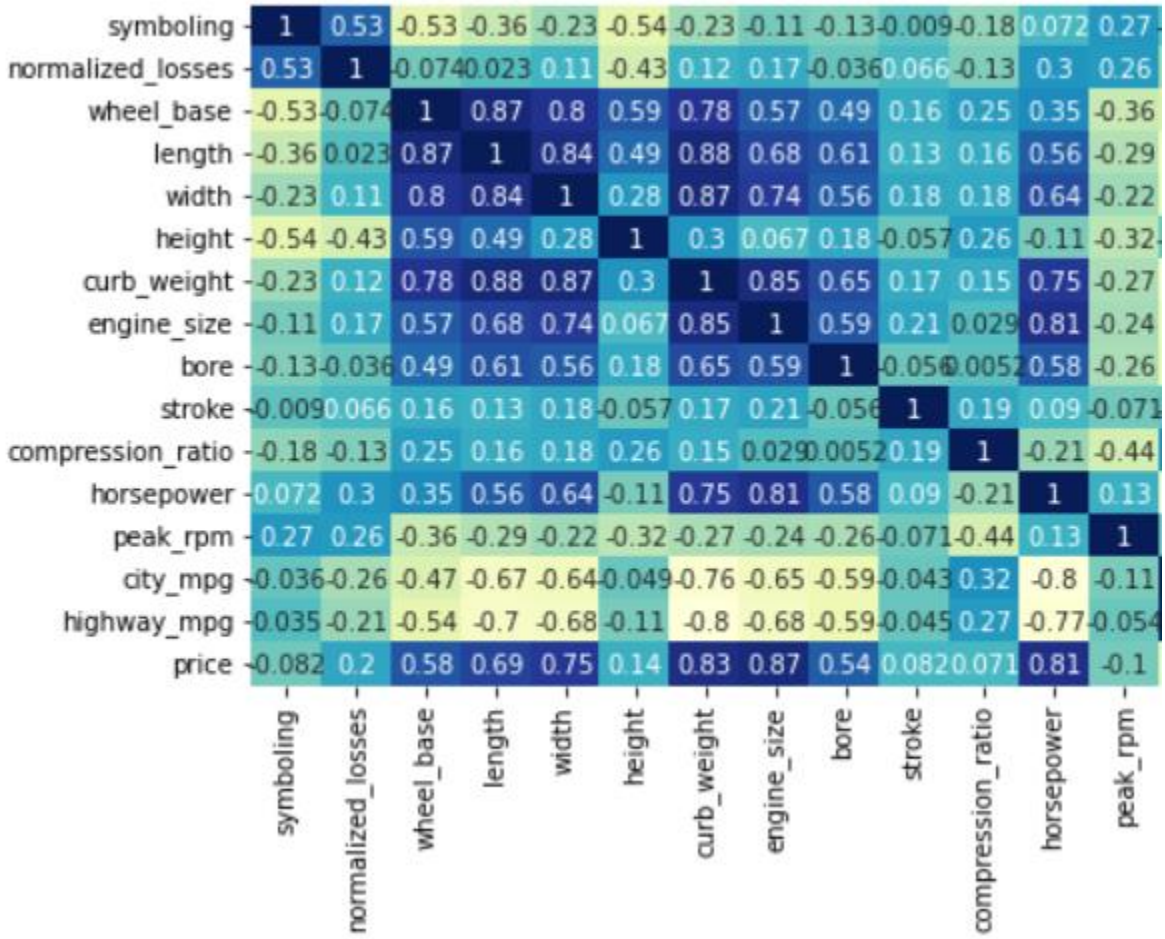
```
data['index_col'] = data.in
print(data.shape)
data.head()
```

```
(205, 27)
```

	symboling	normalized_lo
0	3	
1	3	
2	1	
3	2	
4	2	

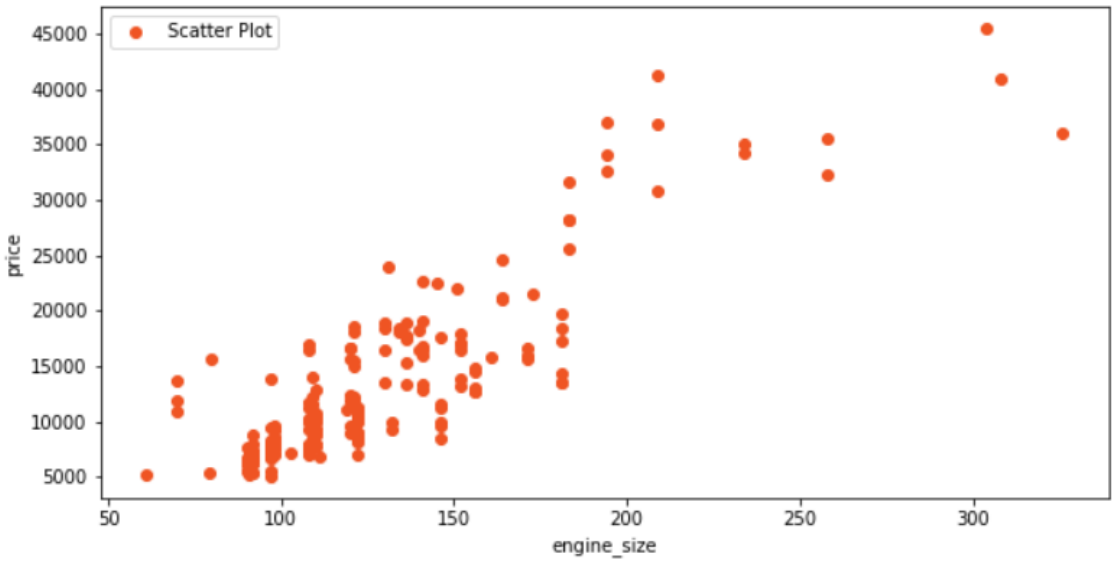
5 rows × 27 columns

5. Find the correlation between the variables to identify dependent and independent variables
6. Using Heatmap for visualization

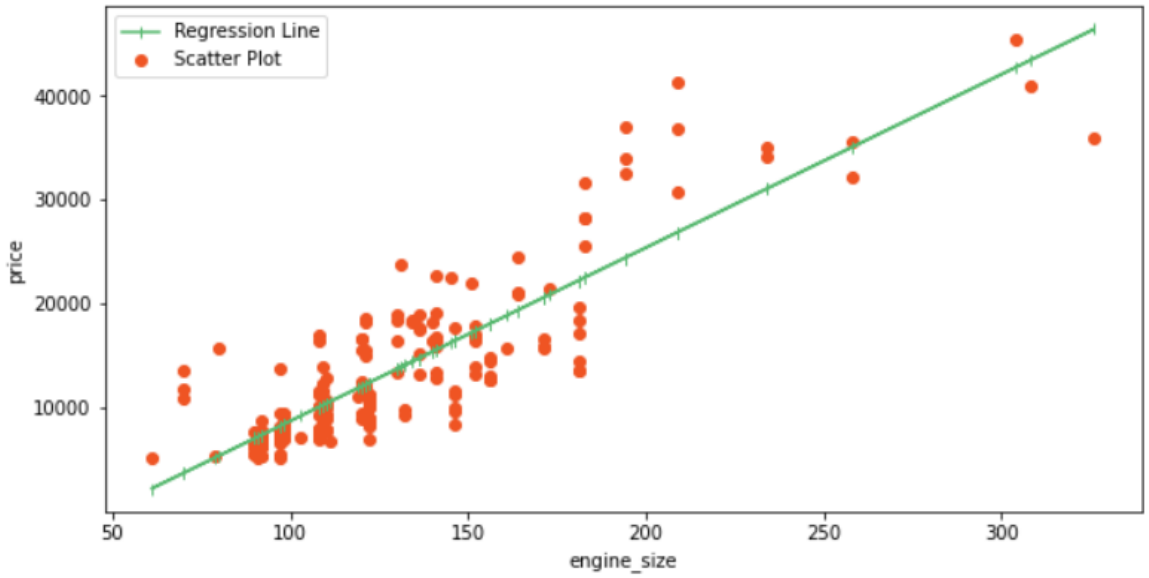


7. Identify the variables with high correlation factor.

8. Draw scatter plot between the dependent and independent variables having high correlation



9. Plot regression line.




```
r2 value is  0.7609686443622008
```

Colab Link:
https://colab.research.google.com/drive/14imVu3hKidMqbyoiFw57Y4aeFMYbtNRJ#scrollTo=OkLAijYz5gS1

Reference: https://towardsdatascience.com/how-to-identify-the-right-independent-variables-for-machine-learning-supervised-algorithms-439986562d32

Reference: https://www.youtube.com/watch?v=VCVhwjbI6h8

Task 3: Polynomial Regression

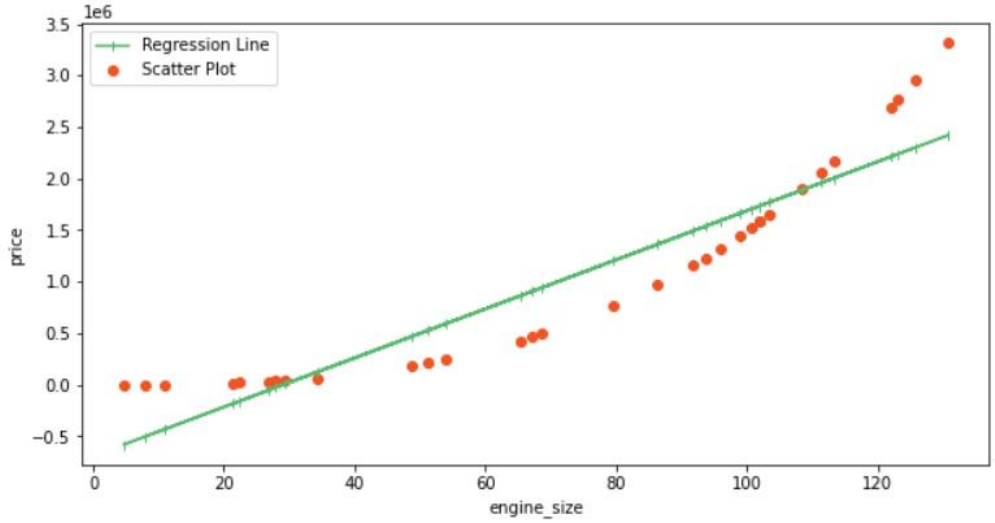
Steps Followed:

1. Imported the libraries
2. Import the input dataset

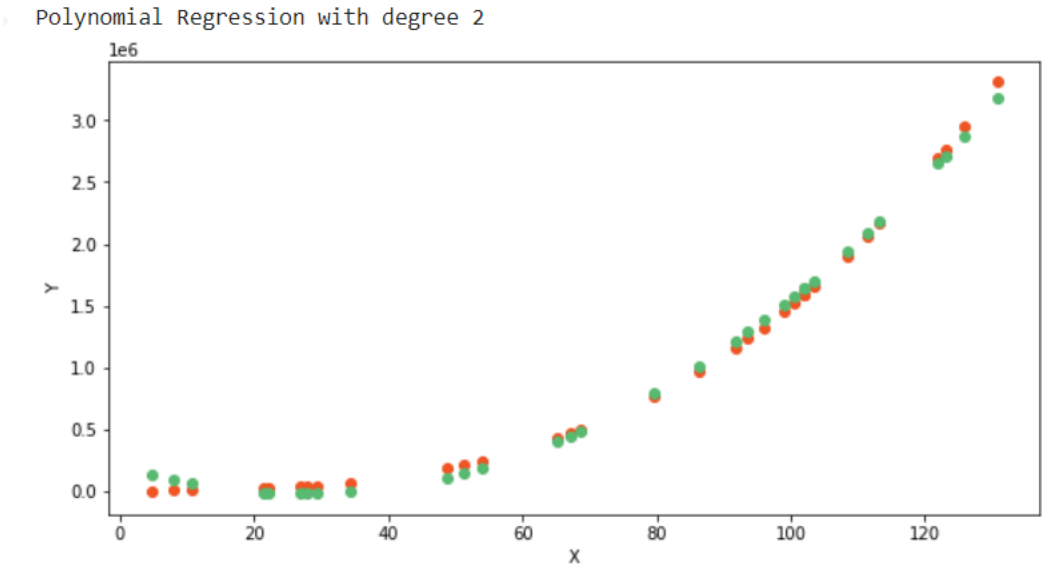
```
Saving Ques_3.csv to Ques_3 (1).csv
(32, 3)
```

	Unnamed: 0	x	y
0	0	34.33159	6.504940e+04
1	1	26.94935	3.238074e+04
2	2	111.47824	2.061468e+06
3	3	51.18001	2.079118e+05
4	4	7.95172	1.083245e+03

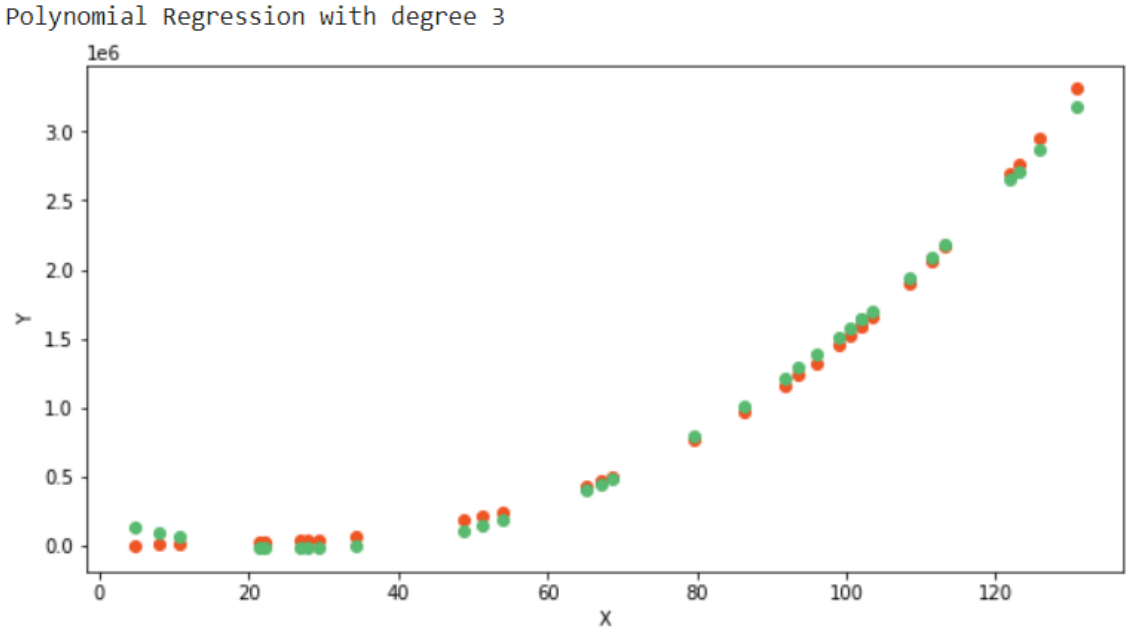
3. Draw Linear Regression Line



4. Draw Polynomial Regression with degree of freedom 2



5. Draw Polynomial Regression with degree 3



Colab Link : https://colab.research.google.com/drive/1YAGcwsdGZFOK9IKX1Bi4aM7CEvkTOtHS#scrollTo=O0YbGFtaRDsU

Reference : https://www.youtube.com/watch?v=SSmpyoIdJTc

Task 5: Logistic Regression

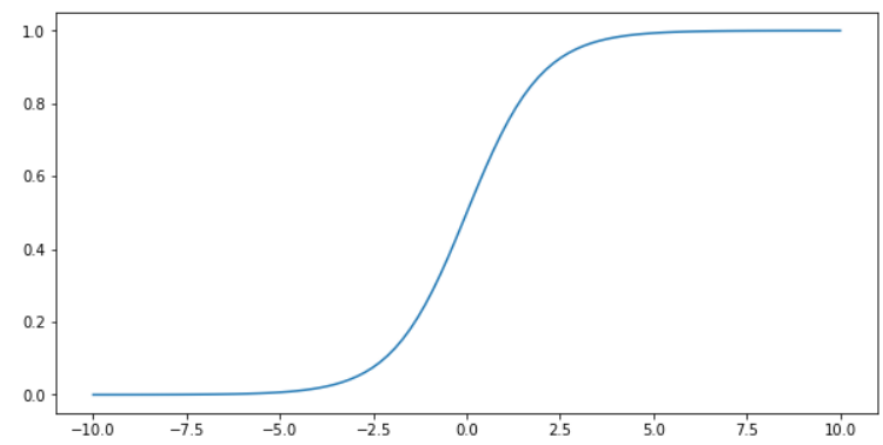
Steps Followed:

1. Imported the libraries
2. Import the input dataset

```
Saving Algerian_forest_fires_dataset_UPDATE.csv to Algerian_forest_fires_dataset_UPDATE
(244, 14)
```

	day	month	year	Temperature	RH	Ws	Rain	FFMC	DMC	DC	ISI	BUI	FWI	Classes
0	1	6	2012	29	57	18	0.0	65.7	3.4	7.6	1.3	3.4	0.5	not fire
1	2	6	2012	29	61	13	1.3	64.4	4.1	7.6	1.0	3.9	0.4	not fire
2	3	6	2012	26	82	22	13.1	47.1	2.5	7.1	0.3	2.7	0.1	not fire
3	4	6	2012	25	89	13	2.5	28.6	1.3	6.9	0.0	1.7	0	not fire
4	5	6	2012	27	77	16	0.0	64.8	3.0	14.2	1.2	3.9	0.5	not fire

3. Plotting sigmoid function in range(-10,+10)



4. Importing Algerian forest fires dataset (Odd Roll Number)
5. Cleaning the dataset.
6. Marking the class of Non fire as 0 and fire as 1

	day	month	year	Temperature	RH	Ws	Rain	FFMC	DMC	DC	ISI	BUI	FWI	Classes
0	1	6	2012	29	57	18	0.0	65.7	3.4	7.6	1.3	3.4	0.5	0
1	2	6	2012	29	61	13	1.3	64.4	4.1	7.6	1.0	3.9	0.4	0
2	3	6	2012	26	82	22	13.1	47.1	2.5	7.1	0.3	2.7	0.1	0
3	4	6	2012	25	89	13	2.5	28.6	1.3	6.9	0.0	1.7	0	0
4	5	6	2012	27	77	16	0.0	64.8	3.0	14.2	1.2	3.9	0.5	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
239	26	9	2012	30	65	14	0.0	85.4	16.0	44.5	4.5	16.9	6.5	1
240	27	9	2012	28	87	15	4.4	41.1	6.5	8.0	0.1	6.2	0	0
241	28	9	2012	27	87	29	0.5	45.9	3.5	7.9	0.4	3.4	0.2	0
242	29	9	2012	24	54	18	0.1	79.7	4.3	15.2	1.7	5.1	0.7	0
243	30	9	2012	24	64	15	0.2	67.3	3.8	16.5	1.2	4.8	0.5	1

244 rows × 14 columns

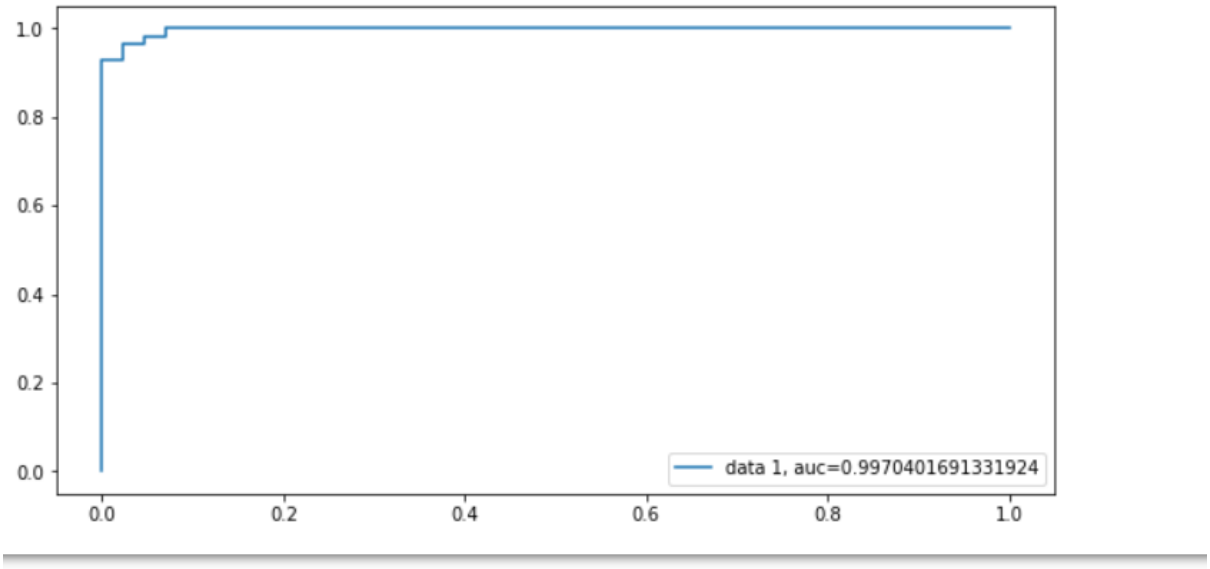
7. Splitting data into test and training set of ratio 60:40.
8. Making model with logistic regression of binary classification
9. Prediction of test data

```
array([1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0,
       0, 1, 1, 1, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 0,
       0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 0, 0, 1, 0, 0, 1, 1,
       0, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 0, 0, 0, 1, 0, 1, 1, 1,
       0, 1, 0, 1, 0, 0, 1, 1, 1, 1])
```

10. Making CNF Matrix for evaluating the performance of a classification model

```
array([[42,  1],
       [ 3, 52]])
```

11. Draw ROC Curve



Colab Link:
https://colab.research.google.com/drive/1pXMVK8WZpAZvc_ug6d0XEYYCDoGTGw8T#scrollTo=E6NihJHGelu1

Reference : https://www.youtube.com/watch?v=VCJdg7YBbAQ
Reference : https://www.youtube.com/watch?v=OCwZyYH14uw
Reference : https://towardsdatascience.com/logistic-regression-detailed-overview-46c4da4303bc

Colab Folder Link :
https://drive.google.com/drive/folders/13Mgi2A4JBoUA5gl0iwNgmeTCu7F-3ErU?usp=sharing