Assignment to build a Machine Learning model

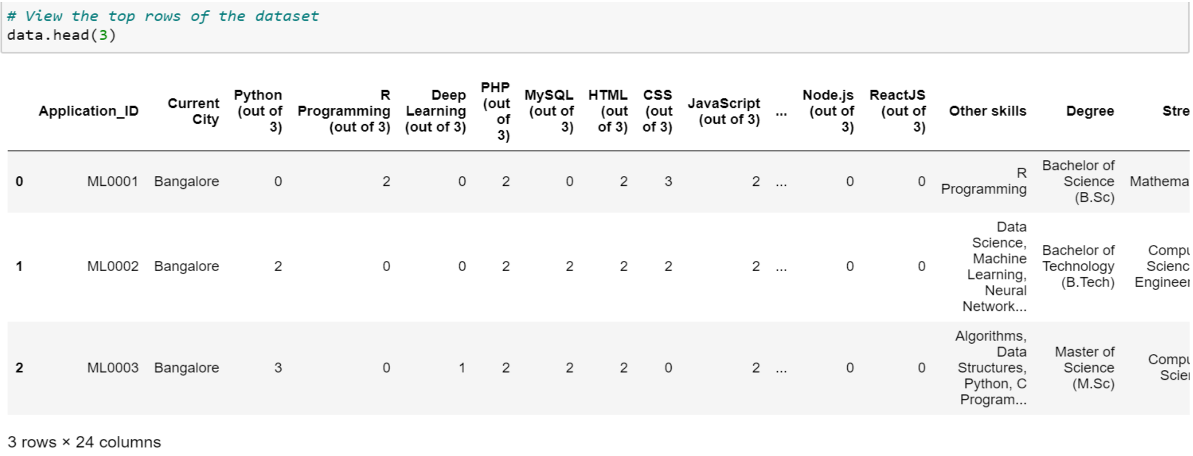
sample-data from the GitHub repository

https://github.com/internbuddy/foster-app.git

The data set contains the following columns.

- Application_ID
- Current City
- Python (out of 3)
- R Programming (out of 3)
- Deep Learning (out of 3)
- PHP (out of 3)
- MySQL (out of 3)
- HTML (out of 3)
- CSS (out of 3)
- JavaScript (out of 3)
- Unnamed: 10
- AJAX (out of 3)
- Bootstrap (out of 3)
- MongoDB (out of 3)
- Node.js (out of 3)
- ReactJS (out of 3)
- Other skills
- Degree
- Stream
- Current Year Of Graduation
- Performance_PG
- Performance_UG
- Performance_12
- Performance_10

The Data set visualization

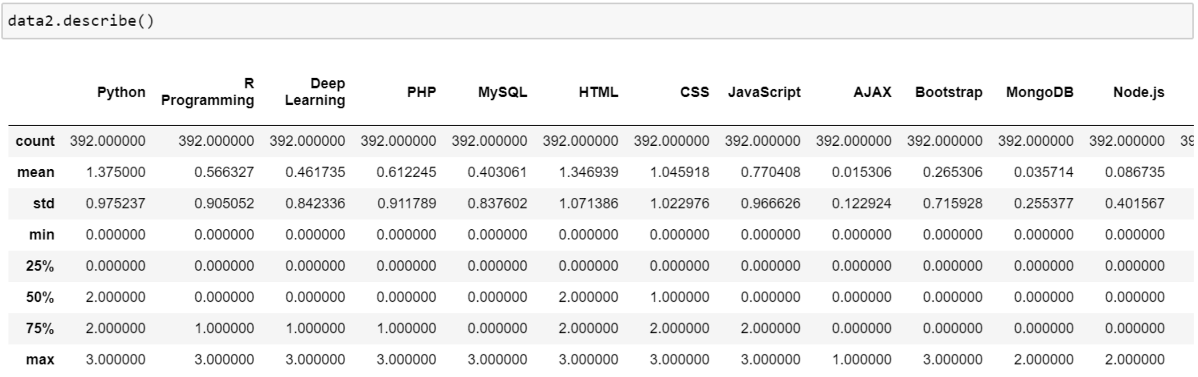


```
# View the top rows of the dataset
data.head(3)
```

	Application_ID	Current City	Python (out of 3)	R Programming (out of 3)	Deep Learning (out of 3)	PHP (out of 3)	MySQL (out of 3)	HTML (out of 3)	CSS (out of 3)	JavaScript (out of 3)	...	Node.js (out of 3)	ReactJS (out of 3)	Other skills	Degree	Stre
0	ML0001	Bangalore	0	2	0	2	0	2	3	2	...	0	0	R Programming	Bachelor of Science (B.Sc)	Mathema
1	ML0002	Bangalore	2	0	0	2	2	2	2	2	...	0	0	Data Science, Machine Learning, Neural Network...	Bachelor of Technology (B.Tech)	Compu Scienc Engineer
2	ML0003	Bangalore	3	0	1	2	2	2	0	2	...	0	0	Algorithms, Data Structures, Python, C Program...	Master of Science (M.Sc)	Compu Scie

3 rows × 24 columns

Statistical description of data set

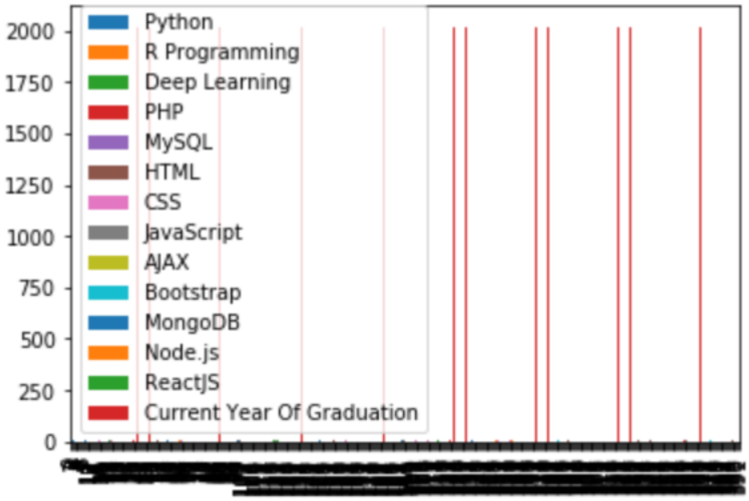


```
data2.describe()
```

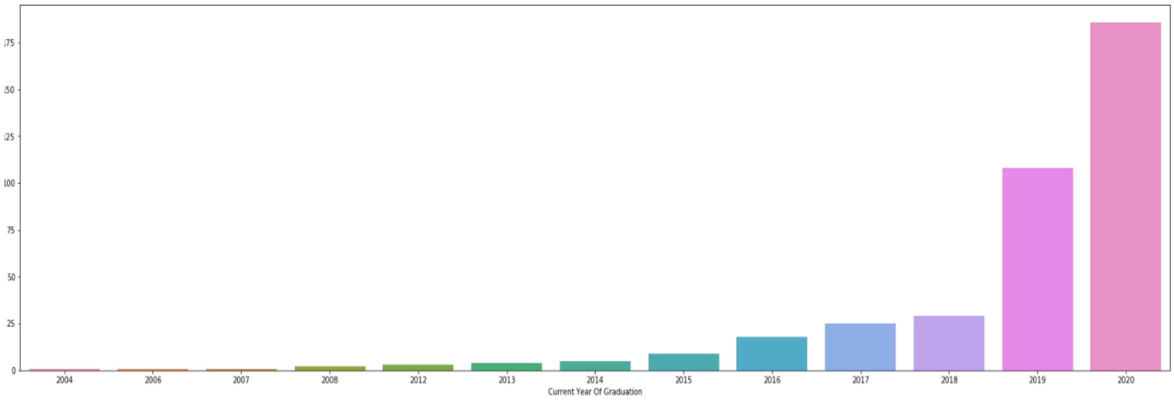
	Python	R Programming	Deep Learning	PHP	MySQL	HTML	CSS	JavaScript	AJAX	Bootstrap	MongoDB	Node.js	
count	392.000000	392.000000	392.000000	392.000000	392.000000	392.000000	392.000000	392.000000	392.000000	392.000000	392.000000	392.000000	39
mean	1.375000	0.566327	0.461735	0.612245	0.403061	1.346939	1.045918	0.770408	0.015306	0.265306	0.035714	0.086735	
std	0.975237	0.905052	0.842336	0.911789	0.837602	1.071386	1.022976	0.966626	0.122924	0.715928	0.255377	0.401567	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
50%	2.000000	0.000000	0.000000	0.000000	0.000000	2.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
75%	2.000000	1.000000	1.000000	1.000000	0.000000	2.000000	2.000000	2.000000	0.000000	0.000000	0.000000	0.000000	
max	3.000000	3.000000	3.000000	3.000000	3.000000	3.000000	3.000000	3.000000	1.000000	3.000000	2.000000	2.000000	

Data visualization

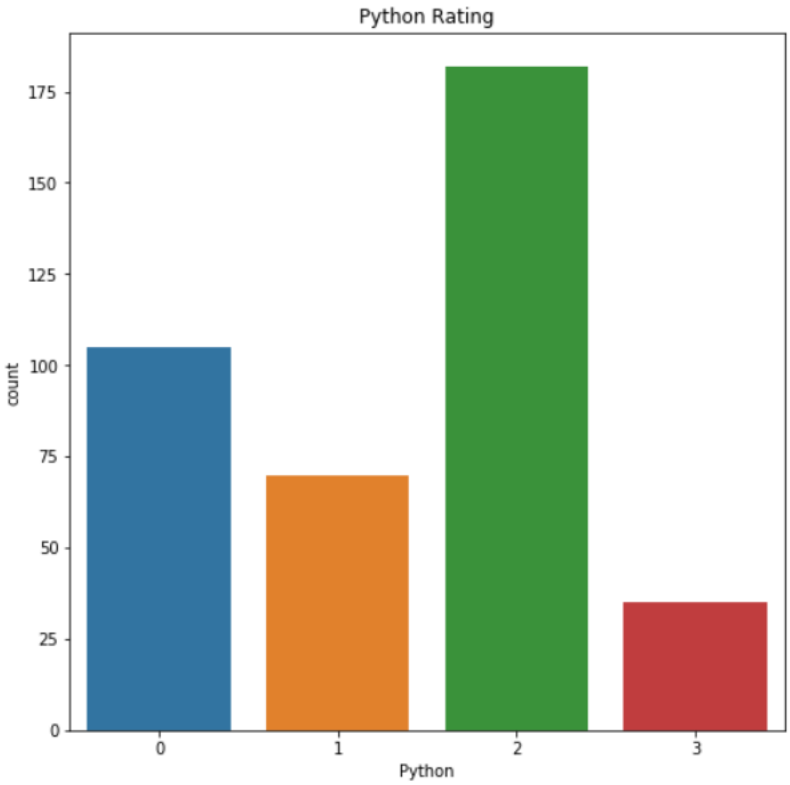
Bar Plot of Data



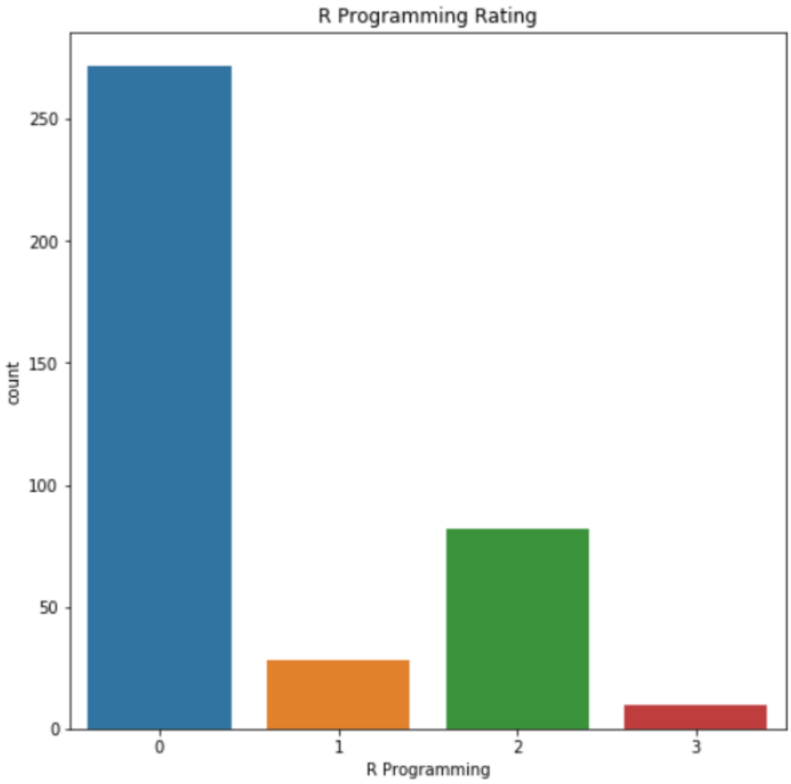
Current Year Of Graduation



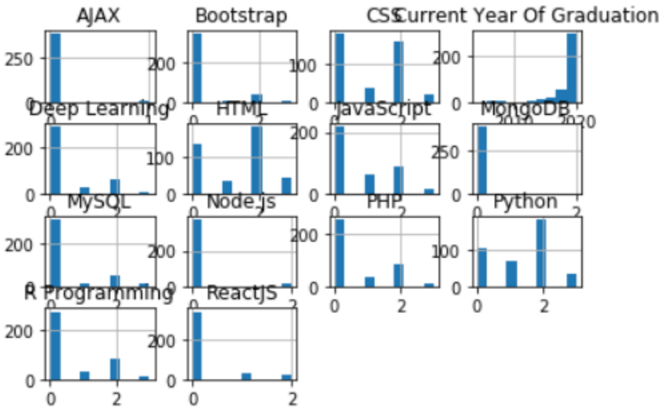
Python Rating



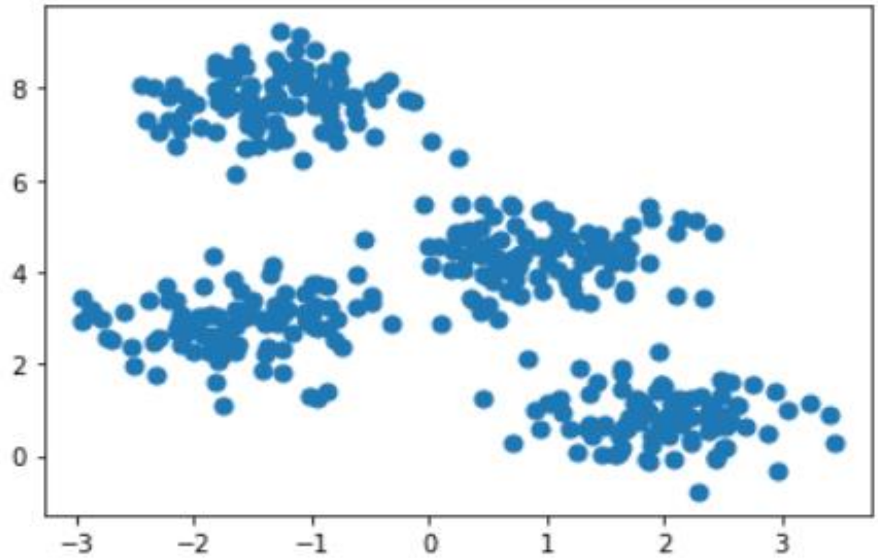
R Programming Rating



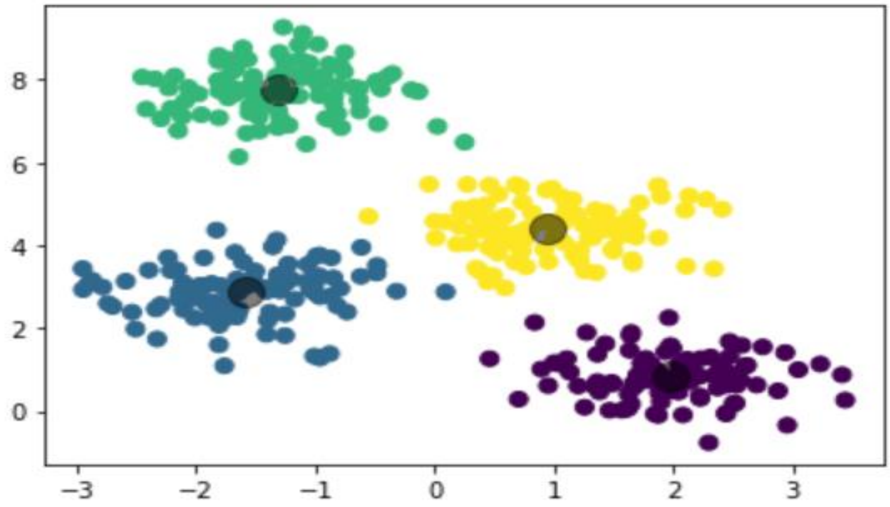
Histograph of different language



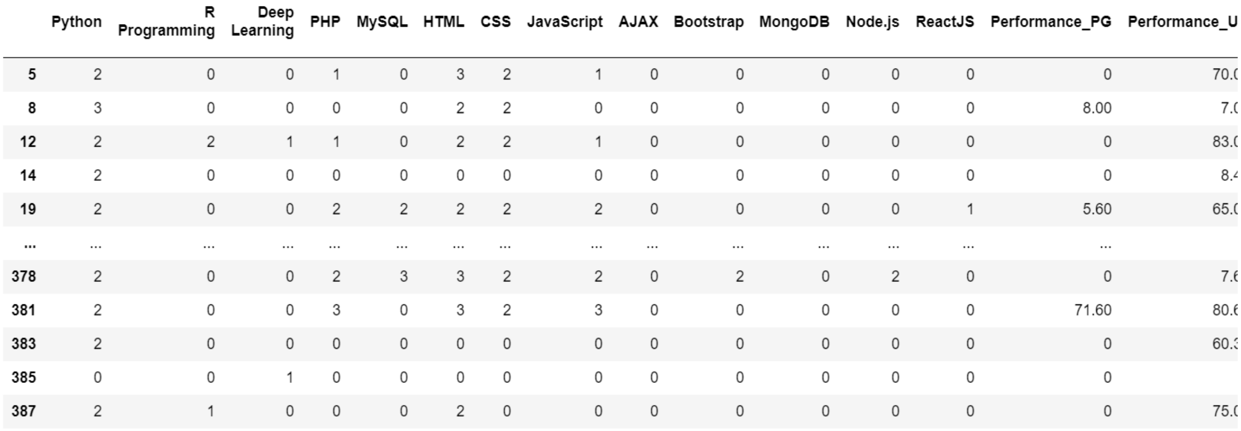
Plot1: Scattered data set visualizing



Plot2: Finally, let's visualize the resulting clusters

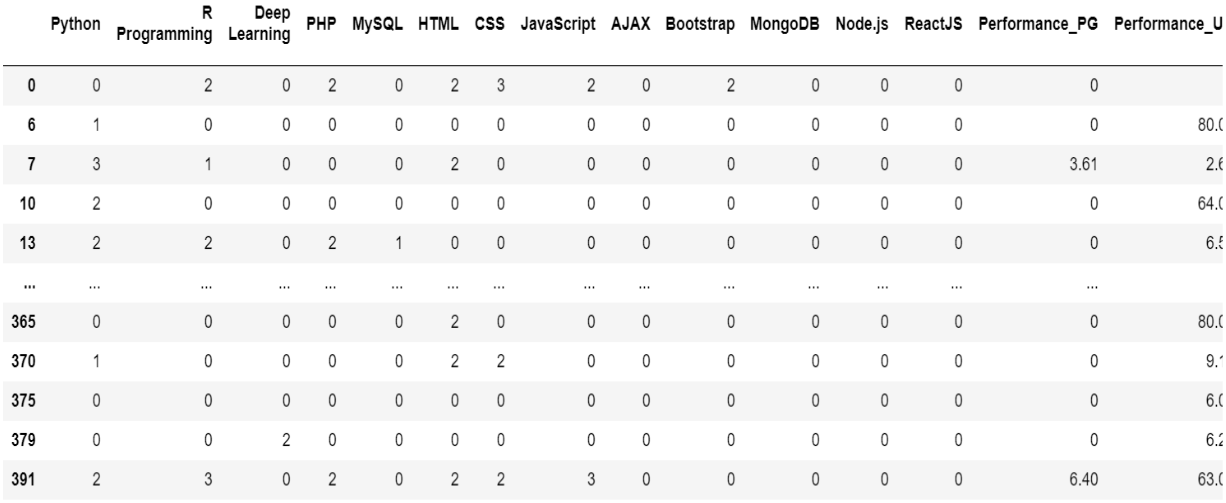


Cluster 1



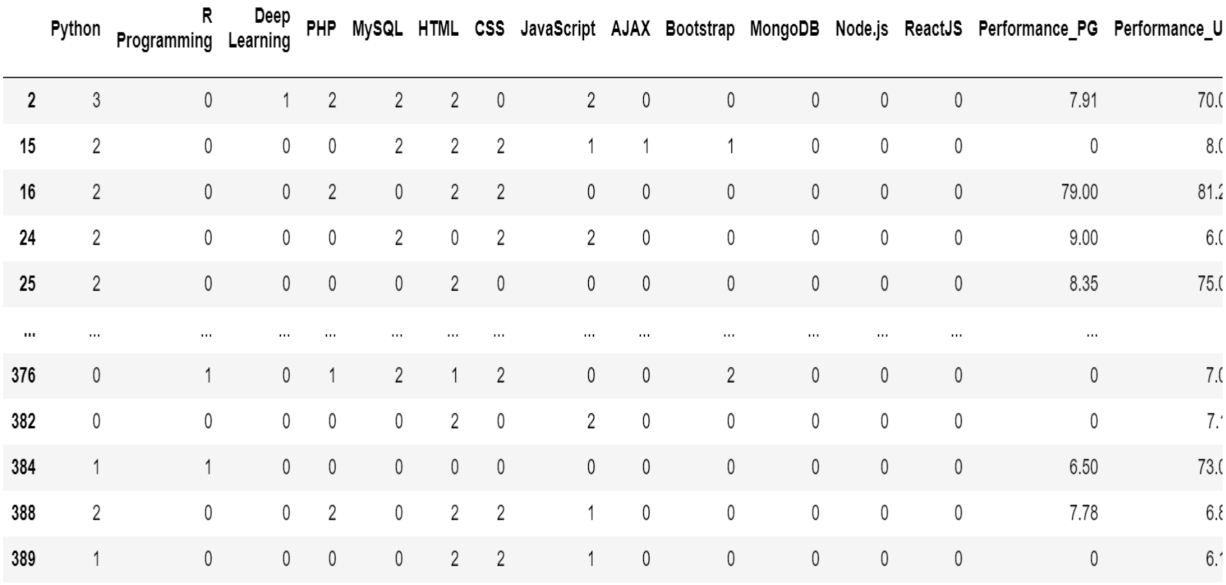
	Python	R Programming	Deep Learning	PHP	MySQL	HTML	CSS	JavaScript	AJAX	Bootstrap	MongoDB	Node.js	ReactJS	Performance_PG	Performance_U
5	2	0	0	1	0	3	2	1	0	0	0	0	0	0	70.
8	3	0	0	0	0	2	2	0	0	0	0	0	0	8.00	7.
12	2	2	1	1	0	2	2	1	0	0	0	0	0	0	83.
14	2	0	0	0	0	0	0	0	0	0	0	0	0	0	8.
19	2	0	0	2	2	2	2	2	0	0	0	0	1	5.60	65.
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
378	2	0	0	2	3	3	2	2	0	2	0	2	0	0	7.
381	2	0	0	3	0	3	2	3	0	0	0	0	0	71.60	80.
383	2	0	0	0	0	0	0	0	0	0	0	0	0	0	60.
385	0	0	1	0	0	0	0	0	0	0	0	0	0	0	
387	2	1	0	0	0	2	0	0	0	0	0	0	0	0	75.

Cluster 2



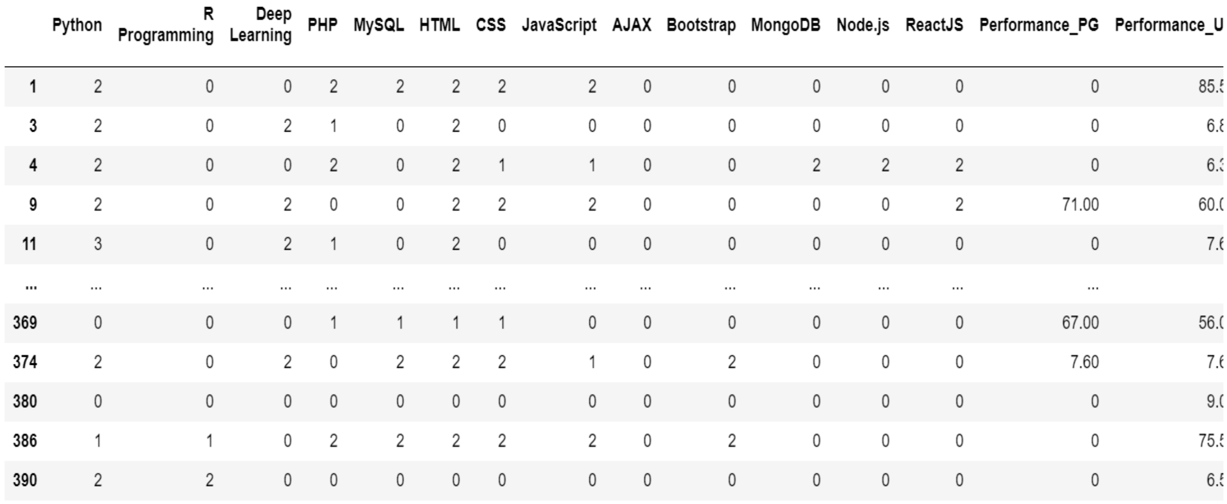
	Python	R Programming	Deep Learning	PHP	MySQL	HTML	CSS	JavaScript	AJAX	Bootstrap	MongoDB	Node.js	ReactJS	Performance_PG	Performance_U
0	0	2	0	2	0	2	3	2	0	2	0	0	0	0	
6	1	0	0	0	0	0	0	0	0	0	0	0	0	0	80.
7	3	1	0	0	0	2	0	0	0	0	0	0	0	3.61	2.
10	2	0	0	0	0	0	0	0	0	0	0	0	0	0	64.
13	2	2	0	2	1	0	0	0	0	0	0	0	0	0	6.
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
365	0	0	0	0	0	2	0	0	0	0	0	0	0	0	80.
370	1	0	0	0	0	2	2	0	0	0	0	0	0	0	9.
375	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6.
379	0	0	2	0	0	0	0	0	0	0	0	0	0	0	6.
391	2	3	0	2	0	2	2	3	0	0	0	0	0	6.40	63.

Cluster 3



	Python	R Programming	Deep Learning	PHP	MySQL	HTML	CSS	JavaScript	AJAX	Bootstrap	MongoDB	Node.js	ReactJS	Performance_PG	Performance_U
2	3	0	1	2	2	2	0	2	0	0	0	0	0	7.91	70.
15	2	0	0	0	2	2	2	1	1	1	0	0	0	0	8.
16	2	0	0	2	0	2	2	0	0	0	0	0	0	79.00	81.
24	2	0	0	0	2	0	2	2	0	0	0	0	0	9.00	6.
25	2	0	0	0	0	2	0	0	0	0	0	0	0	8.35	75.
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
376	0	1	0	1	2	1	2	0	0	2	0	0	0	0	7.
382	0	0	0	0	0	2	0	2	0	0	0	0	0	0	7.
384	1	1	0	0	0	0	0	0	0	0	0	0	0	6.50	73.
388	2	0	0	2	0	2	2	1	0	0	0	0	0	7.78	6.
389	1	0	0	0	0	2	2	1	0	0	0	0	0	0	6.

Cluster 4



	Python	R Programming	Deep Learning	PHP	MySQL	HTML	CSS	JavaScript	AJAX	Bootstrap	MongoDB	Node.js	ReactJS	Performance_PG	Performance_U
1	2	0	0	2	2	2	2	2	0	0	0	0	0	0	85.5
3	2	0	2	1	0	2	0	0	0	0	0	0	0	0	6.8
4	2	0	0	2	0	2	1	1	0	0	2	2	2	0	6.3
9	2	0	2	0	0	2	2	2	0	0	0	0	2	71.00	60.0
11	3	0	2	1	0	2	0	0	0	0	0	0	0	0	7.6
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
369	0	0	0	1	1	1	1	0	0	0	0	0	0	67.00	56.0
374	2	0	2	0	2	2	2	1	0	2	0	0	0	7.60	7.6
380	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9.0
386	1	1	0	2	2	2	2	2	0	2	0	0	0	0	75.5
390	2	2	0	0	0	0	0	0	0	0	0	0	0	0	6.5

Conclusion

We can see from the above plots that given data set is unevenly distributed, with four clusters.

The clustering of four groups grouped according to the given rating of the languages.

The K Nearest Neighbour(KNN), Decision Tree, SVM models are built which have low accuracy because uneven data is distribution.