

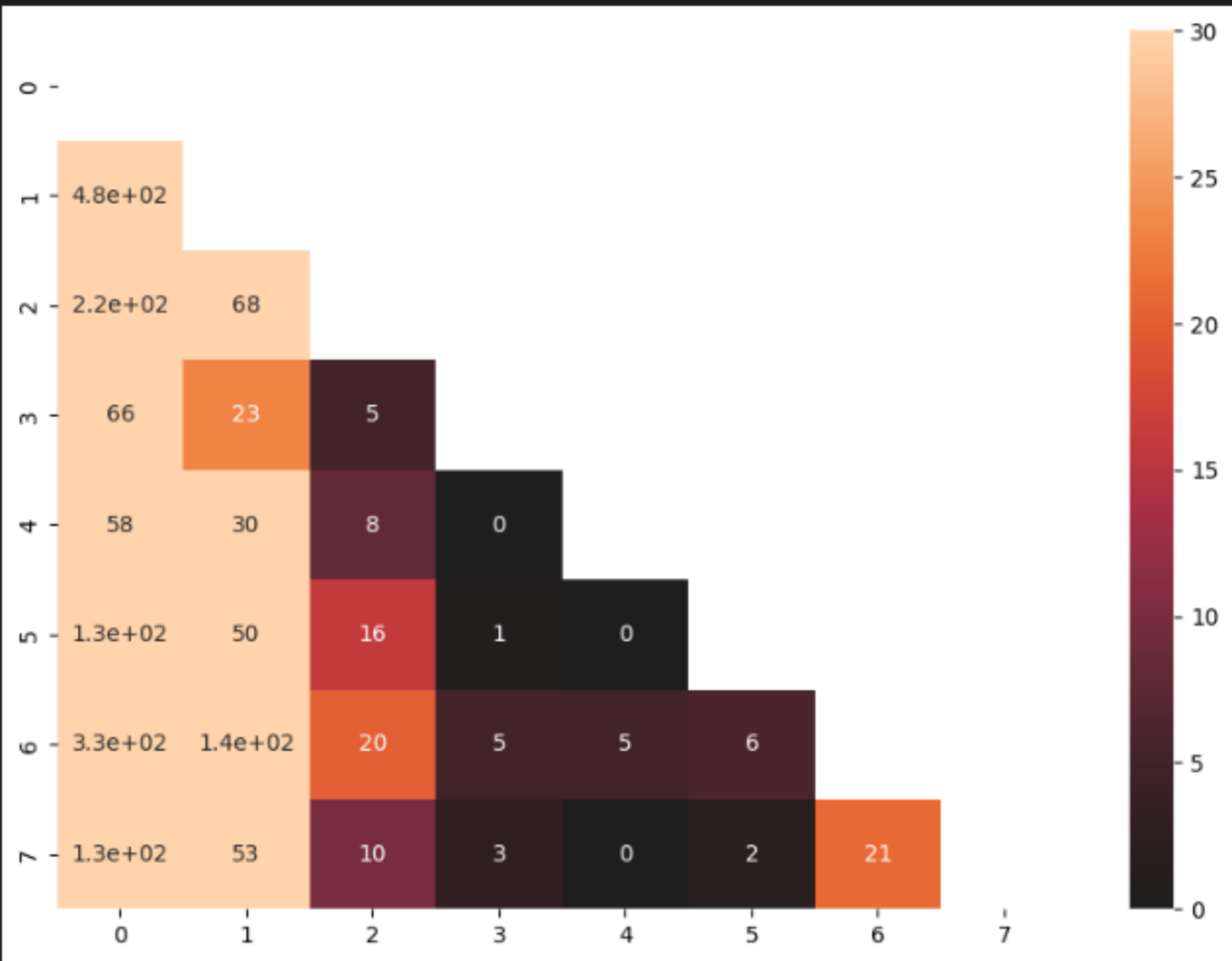
AI-Assignment-4

Preprocessing

- I read the given csv using pandas
- After checking the correlation matrix, i saw some irrelevant columns
- I removed 11-12 columns after looking at it.
- Then did one-hot encoding for all string columns and appended to the original df
- After doing normalization and standardization , I split the data into X and Y
- I also divided the jobs label into groups as shown below:
 - ◆ 'Analyst':1,
 - ◆ 'Manager':2,
 - ◆ 'Auditor':3,
 - ◆ 'Associate':4, 'Architect':5,
 - ◆ 'Administrator':6, 'Support':7,
 - ◆ 'Developer':0,'UX':0,'Testing':0, 'Engineer':0,'Designer':0
- Then did label encoding for the output also.
- Steps taken to analyze the data
 - ◆ I split the data into train and test and validation
 - ◆ After trying many grouping variations of the above job role, this type of clubbing(as mentioned above) gave the maximum accuracy
 - ◆ Then I performed the hyperparameter tuning on MLP (once manually and gridsearchCV)
 - ◆ Printed accuracy and confusion matrix
 - ◆ Then I performed the same step for the 10,20,30,40 percent of the data for the test set
- Analysis
 - ◆ If I used the default params, **I got 30% accuracy**
 - ◆ The best score I got was **40.025% accuracy.**
 - ◆ I realized the score was less because the data is very skewed.

MLP without GridsearchCV :
Accuracy score: 0.30275
Overall Precision: 0.23611365708356202
Overall Recall:0.30275

	precision	recall	f1-score	support
0	0.41	0.60	0.49	1623
1	0.19	0.21	0.20	775
2	0.07	0.03	0.05	352
3	0.00	0.00	0.00	112
4	0.00	0.00	0.00	113
5	0.08	0.02	0.03	228
6	0.14	0.10	0.12	572
7	0.06	0.02	0.03	225
accuracy			0.30	4000
macro avg	0.12	0.12	0.11	4000
weighted avg	0.24	0.30	0.26	4000



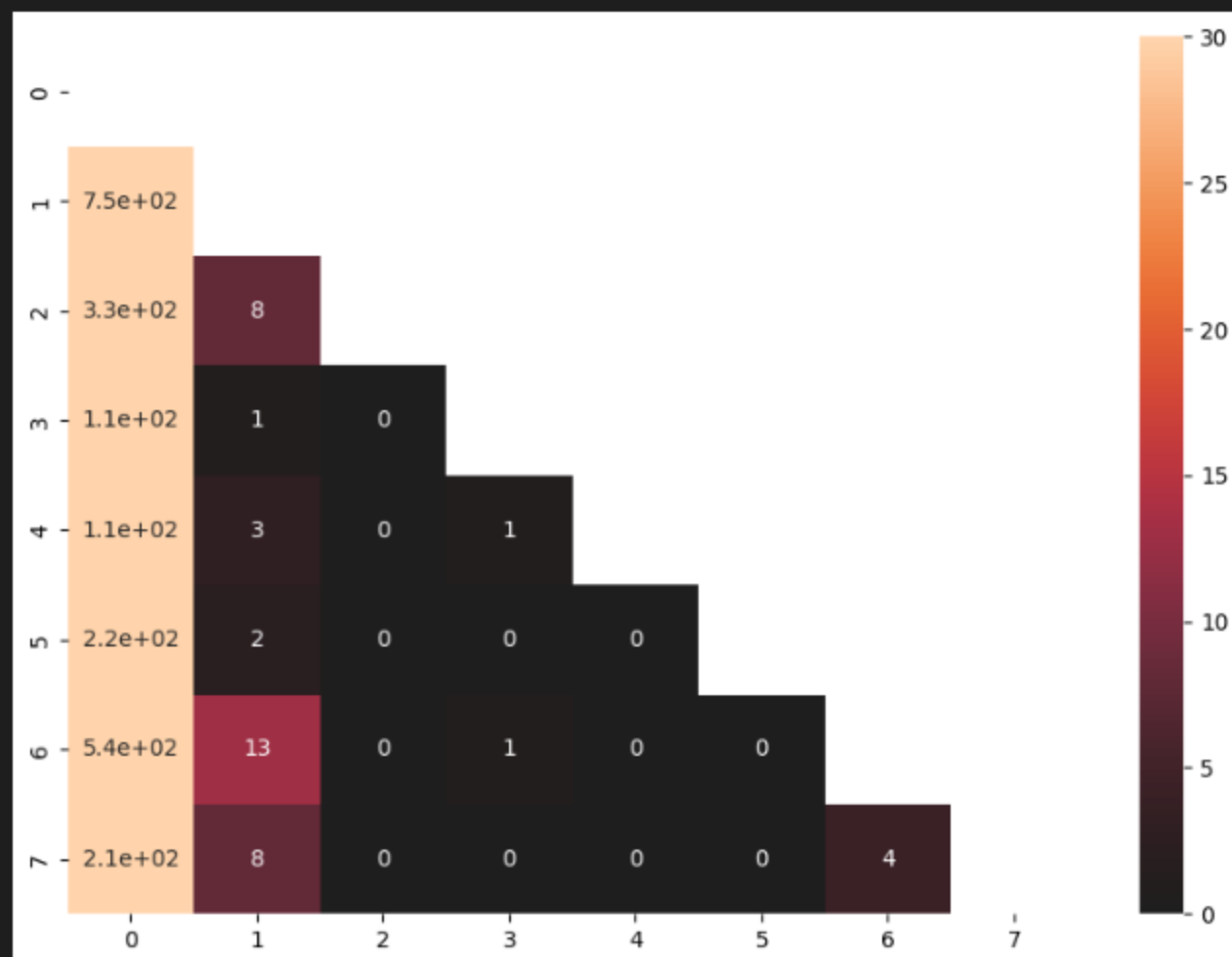
MLP with hyperparameter tuning :

Accuracy score: 0.40025

Overall Precision: 0.23260556612401684

Overall Recall:0.40025

	precision	recall	f1-score	support
0	0.41	0.97	0.57	1623
1	0.21	0.02	0.04	775
2	0.00	0.00	0.00	352
3	0.00	0.00	0.00	112
4	0.00	0.00	0.00	113
5	0.00	0.00	0.00	228
6	0.18	0.02	0.04	572
7	0.00	0.00	0.00	225
accuracy			0.40	4000
macro avg	0.10	0.13	0.08	4000
weighted avg	0.23	0.40	0.25	4000

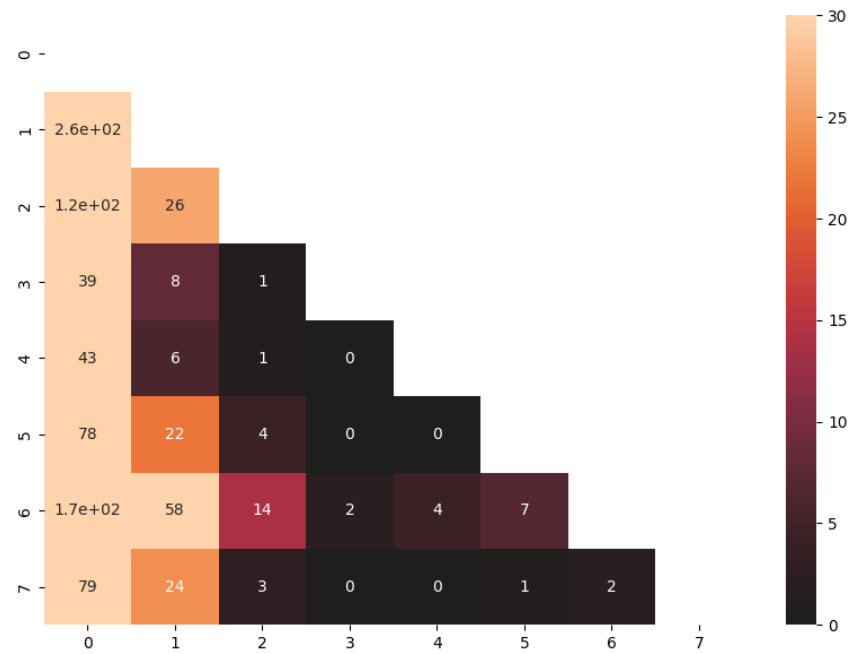


We can see here the accuracy for different splits now(Along with correlation matrix just below it):

-----Test data split is 10%-----
Model without HP :
Accuracy score: 0.315
Overall Precision: 0.2388637534262671
Overall Recall:0.315

			precision	recall	f1-score	support
		0	0.40	0.65	0.50	812
		1	0.18	0.16	0.17	387
		2	0.14	0.06	0.09	176
		3	0.08	0.02	0.03	56
		4	0.00	0.00	0.00	57
		5	0.00	0.00	0.00	114
		6	0.15	0.09	0.11	286
		7	0.09	0.03	0.04	112
	accuracy				0.32	2000
macro avg			0.13	0.13	0.12	2000
weighted avg			0.24	0.32	0.26	2000

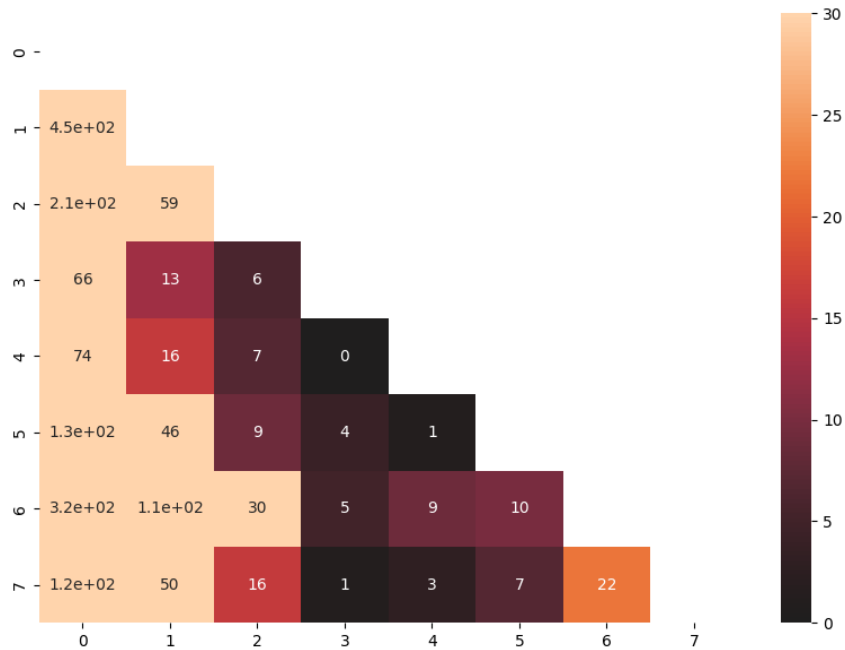
Accuracy without HP tuning is 0.315



```
-----Test data split is 20%-----
Model without HP :
Accuracy score: 0.2965
Overall Precision: 0.24101948291337538
Overall Recall:0.2965
```

		precision	recall	f1-score	support
	0	0.41	0.59	0.49	1623
	1	0.18	0.16	0.17	775
	2	0.09	0.06	0.07	352
	3	0.07	0.02	0.03	112
	4	0.01	0.01	0.01	113
	5	0.08	0.03	0.04	228
	6	0.15	0.12	0.14	572
	7	0.04	0.02	0.03	225
	accuracy			0.30	4000
	macro avg	0.13	0.13	0.12	4000
	weighted avg	0.24	0.30	0.26	4000

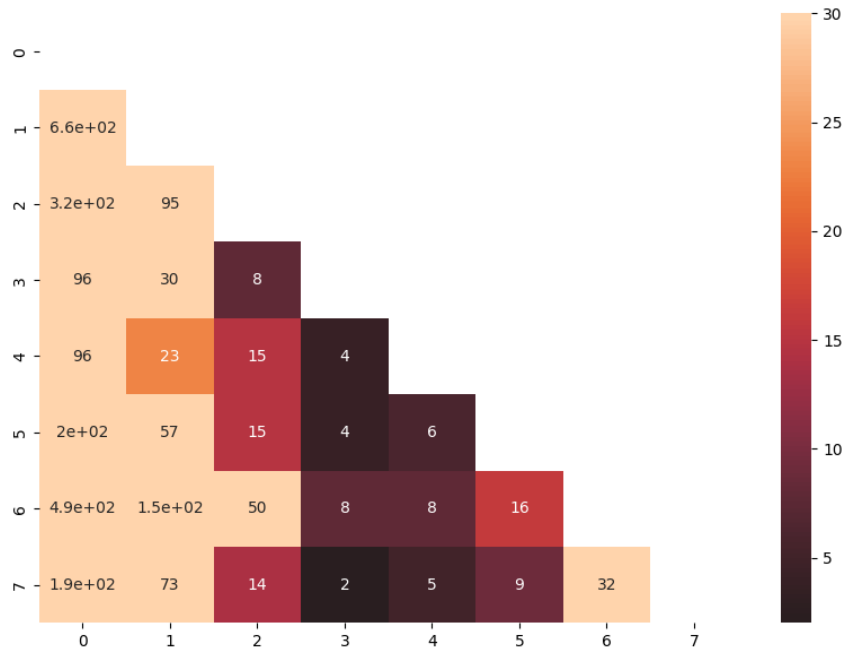
Accuracy without HP tuning is 0.2965



```
-----Test data split is 30%-----
Model without HP :
Accuracy score: 0.29645059156807196
Overall Precision: 0.24371524871074063
Overall Recall:0.296450591568072
```

		precision	recall	f1-score	support
	0	0.41	0.57	0.48	2435
	1	0.19	0.18	0.19	1162
	2	0.09	0.05	0.07	529
	3	0.00	0.00	0.00	167
	4	0.02	0.01	0.01	170
	5	0.10	0.05	0.06	343
	6	0.16	0.13	0.14	858
	7	0.09	0.04	0.05	337
	accuracy			0.30	6001
	macro avg	0.13	0.13	0.13	6001
	weighted avg	0.24	0.30	0.26	6001

Accuracy without HP tuning is 0.29645059156807196



```
-----Test data split is 40%-----
Model without HP :
Accuracy score: 0.27925
Overall Precision: 0.2317196267416508
Overall Recall:0.27925

```

		precision	recall	f1-score	support
	0	0.40	0.55	0.46	3246
	1	0.17	0.16	0.17	1550
	2	0.09	0.07	0.08	705
	3	0.02	0.01	0.01	223
	4	0.01	0.00	0.01	226
	5	0.04	0.02	0.02	457
	6	0.14	0.10	0.11	1144
	7	0.08	0.06	0.06	449
	accuracy			0.28	8000
	macro avg	0.12	0.12	0.12	8000
	weighted avg	0.23	0.28	0.25	8000

```
Accuracy without HP tuning is 0.27925
```

