

## Decision Tree Regression Model Predicted\_GPA Analysis.....

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from matplotlib import rcParams
import seaborn as sb
%matplotlib inline
rcParams['figure.figsize']=10,8
sb.set_style('whitegrid')

df=pd.read_csv("C:/Users/Vivek Vishan Jetani/Documents/Models/New
Prediction/PredictionOriginal.csv")

df
```

	classification	resource	place_of_study	sleep_time	time_spend_on_social_media	SSC	HSC	attendance	self_study	Predicted_GPA
0	0	0	0	0	0.0	0	0	0	3	2.0
1	0	0	0	0	0.0	0	0	0	8	3.3
2	0	0	0	0	0.0	0	0	0	13	3.3
3	0	0	0	0	0.0	0	0	0	18	3.3
4	0	0	0	0	0.0	0	0	21-40%	0	2.0
5	0	0	0	0	0.0	0	0	41-60%	0	2.0
6	0	0	0	0	0.0	0	0	61-80%	0	2.0
7	0	0	0	0	0.0	0	0	81-100%	0	2.0
8	0	0	0	0	0.0	0	65%	0	0	2.0
9	0	0	0	0	0.0	0	70%	0	0	2.0

## Classification vs Predicted\_GPA.....

```
pd.crosstab(df.classification, df.Predicted_GPA,normalize='index').plot(kind='bar')
pd.crosstab(df.classification, df.Predicted_GPA,normalize='index').plot(kind='line')
pd.crosstab(df.classification, df.Predicted_GPA,normalize='index')
```

Predicted_GPA	2.0	2.35	2.7	3.0	3.25	3.3	3.8
---------------	-----	------	-----	-----	------	-----	-----

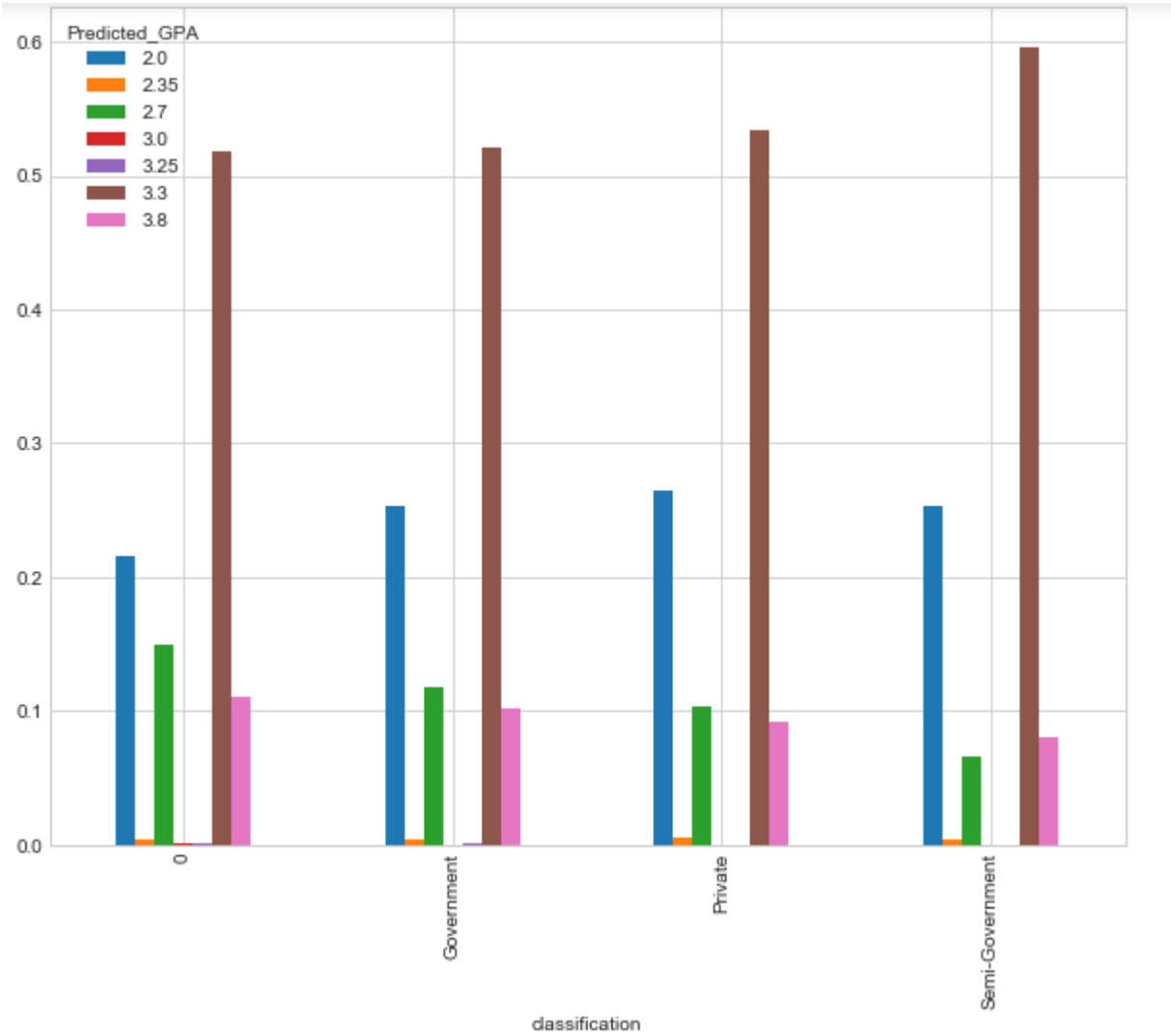
classification

	0	0.216108	0.003772	0.148880	0.001331	0.000666	0.518305	0.110939
--	---	----------	----------	----------	----------	----------	----------	----------

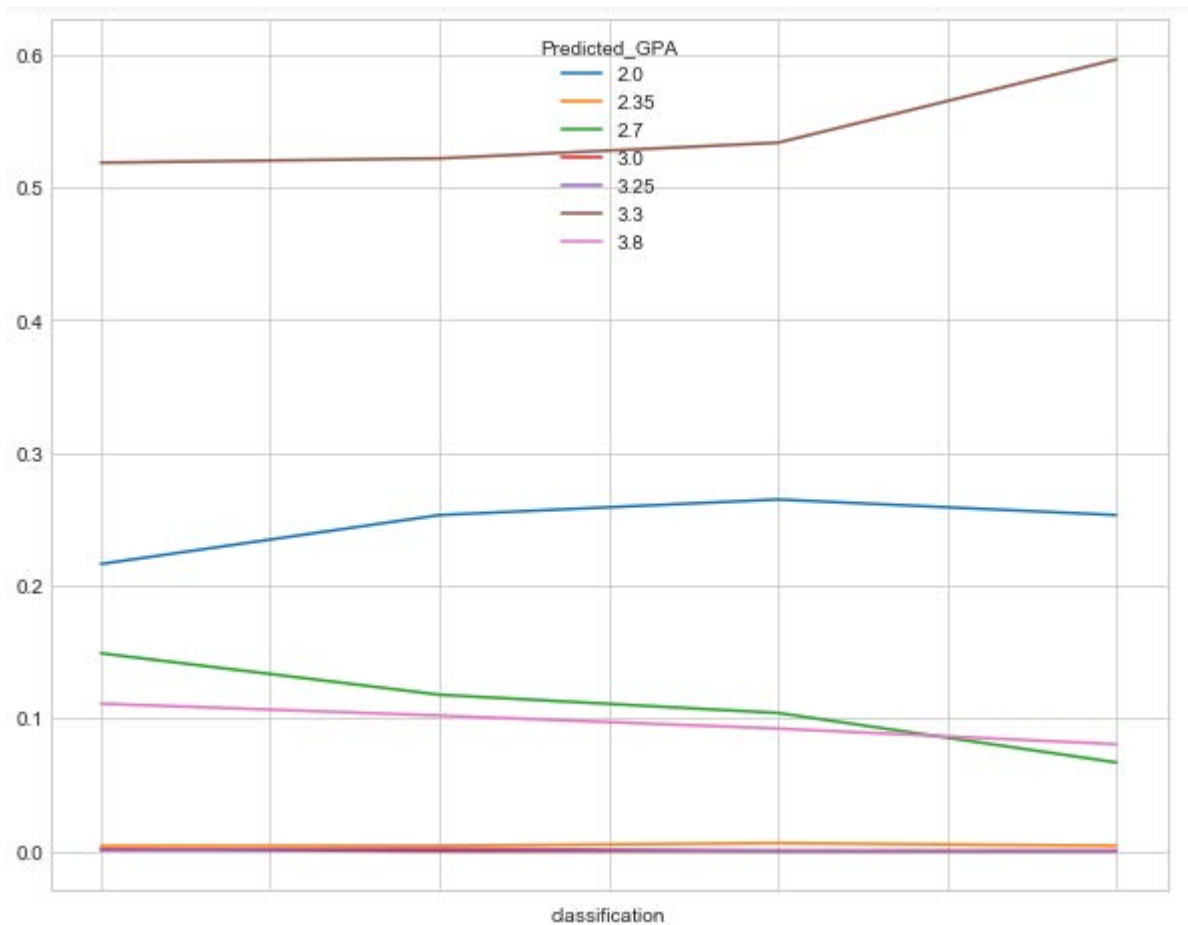
Government		0.252941	0.003922	0.117647	0.000000	0.001961	0.521569	0.101961
------------	--	----------	----------	----------	----------	----------	----------	----------

Private		0.264706	0.005882	0.103922	0.000000	0.000000	0.533333	0.092157
---------	--	----------	----------	----------	----------	----------	----------	----------

Semi-Government		0.252941	0.003922	0.066667	0.000000	0.000000	0.596078	0.080392
-----------------	--	----------	----------	----------	----------	----------	----------	----------



In this figure all parameters are slightly equal but 3.3 CGPA is increasing in each parameter. ("0" is showing that values in which these three(Government, Private, Semi-Government) parameters are not depending.... )



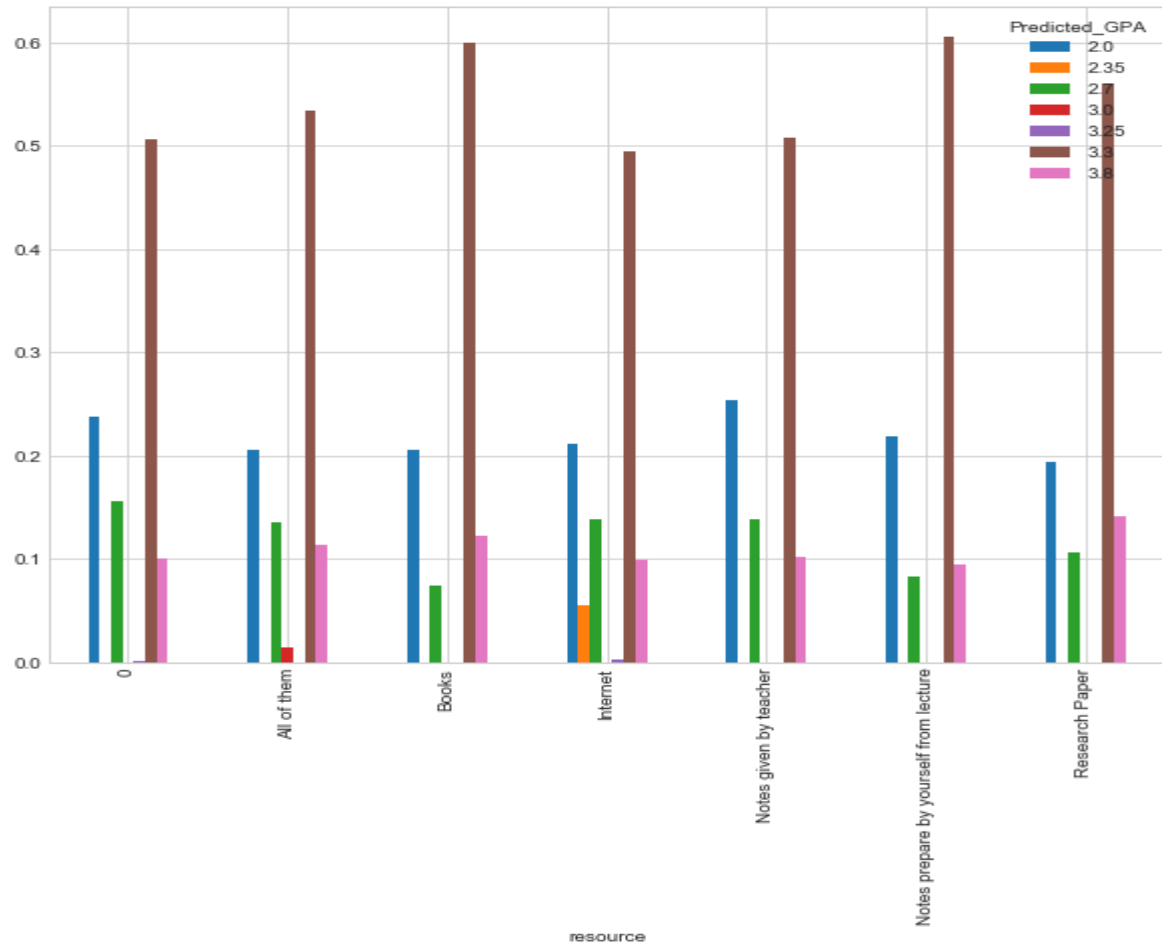
Resource vs Predicted\_GPA.....

```
pd.crosstab(df.resource, df.Predicted_GPA, normalize='index').plot(kind='Bar')
```

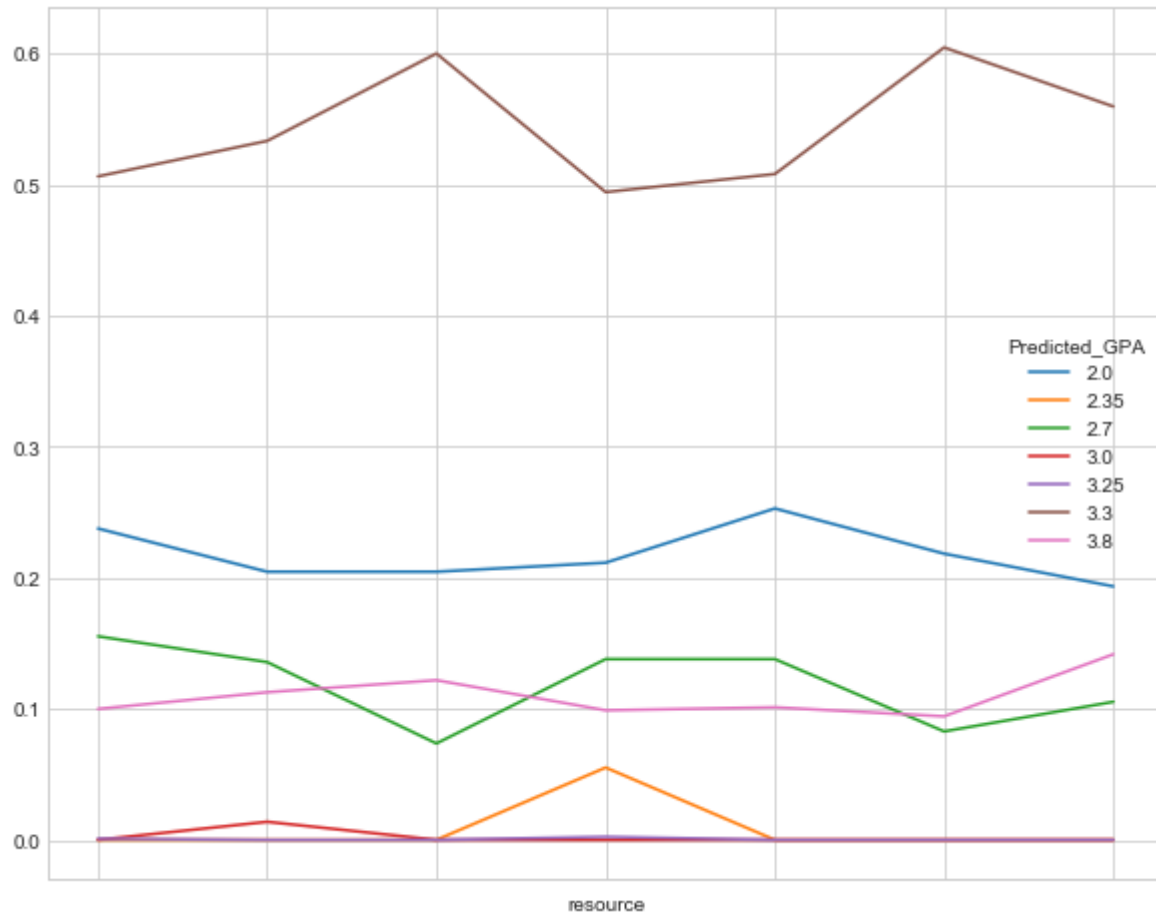
```
pd.crosstab(df.resource, df.Predicted_GPA, normalize='index').plot(kind='line')
```

```
pd.crosstab(df.resource, df.Predicted_GPA, normalize='index')
```

Predicted_GPA	2.0	2.35	2.7	3.0	3.25	3.3	3.8
resource							
0	0.237588	0.000000	0.155346	0.000000	0.000914	0.506244	0.099909
All of them	0.204598	0.000000	0.135632	0.013793	0.000000	0.533333	0.112644
Books	0.204598	0.000000	0.073563	0.000000	0.000000	0.600000	0.121839
Internet	0.211494	0.055172	0.137931	0.000000	0.002299	0.494253	0.098851
Notes given by teacher	0.252874	0.000000	0.137931	0.000000	0.000000	0.508046	0.101149
Notes prepare by yourself from lecture	0.218391	0.000000	0.082759	0.000000	0.000000	0.604598	0.094253
Research Paper	0.193437	0.000000	0.105354	0.000000	0.000000	0.559585	0.141623



In this figure all parameters are slightly equal but 3.3 CGPA is increasing in each parameter. This figure shows those students refers “Books”, “Research paper” and “Notes repair by yourself from lecture” are around 60 % which have 3.3 CGPA (“0” is showing the values in which all Resource parameters are not depending.... )



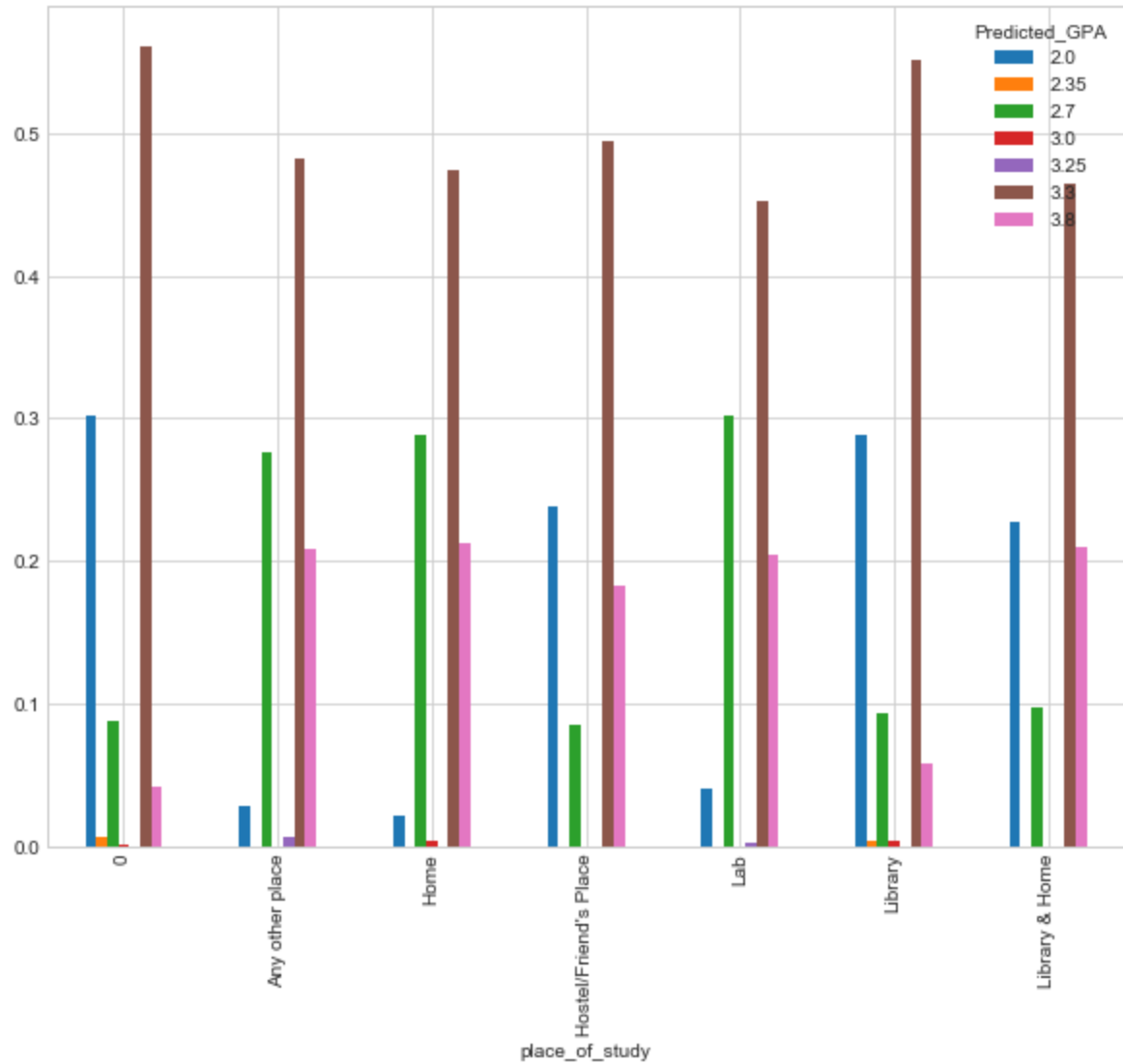
**Place\_of\_study vs Predicted\_GPA.....**

```
pd.crosstab(df.place_of_study, df.Predicted_GPA, normalize='index').plot(kind='Bar')
```

```
pd.crosstab(df.place_of_study, df.Predicted_GPA, normalize='index').plot(kind='line')
```

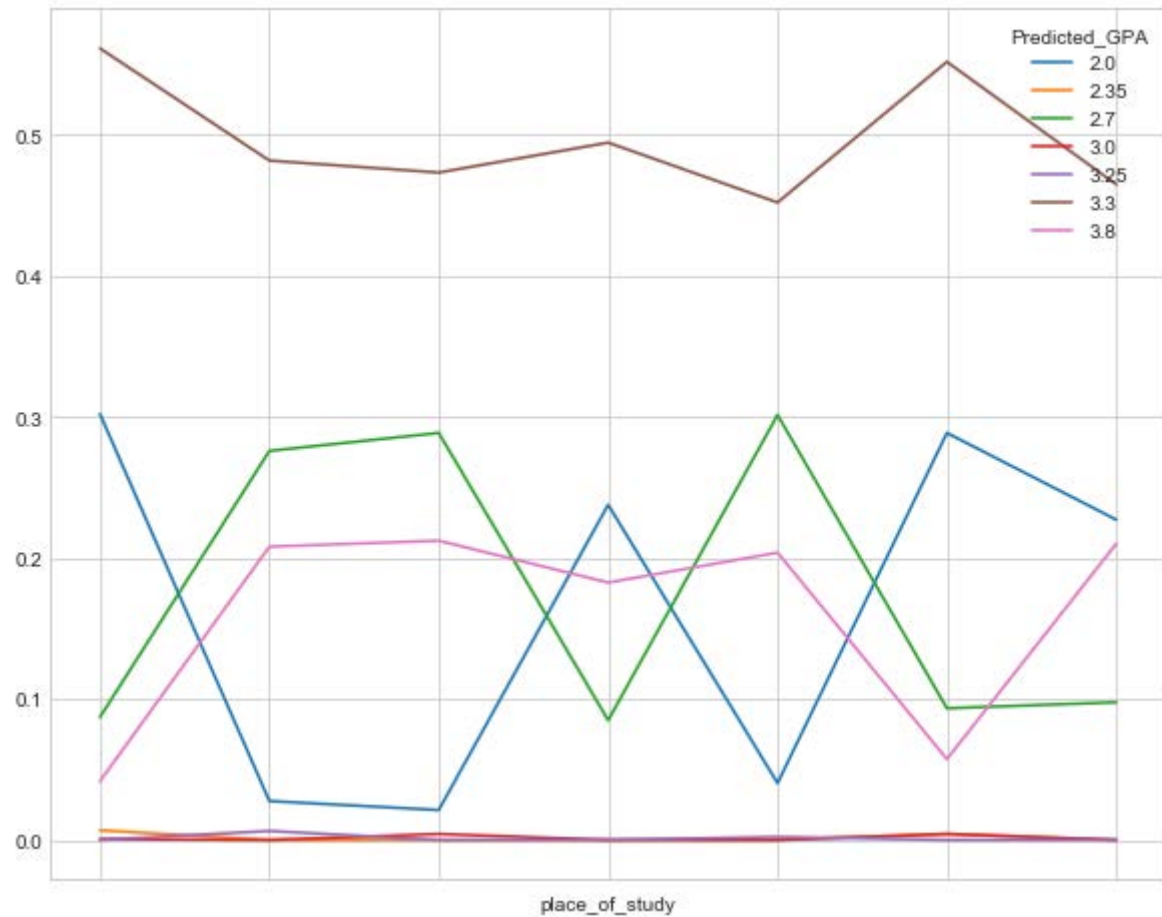
```
pd.crosstab(df.place_of_study, df.Predicted_GPA, normalize='index')
```

Predicted_GPA	2.0	2.35	2.7	3.0	3.25	3.3	3.8
place_of_study							
0	0.302087	0.006851	0.087200	0.000623	0.000000	0.561507	0.041732
Any other place	0.027601	0.000000	0.276008	0.000000	0.006369	0.481953	0.208068
Home	0.021231	0.000000	0.288747	0.004246	0.000000	0.473461	0.212314
Hostel/Friend's Place	0.237792	0.000000	0.084926	0.000000	0.000000	0.494692	0.182590
Lab	0.040340	0.000000	0.301486	0.000000	0.002123	0.452229	0.203822
Library	0.288747	0.004246	0.093418	0.004246	0.000000	0.552017	0.057325
Library & Home	0.227176	0.000000	0.097665	0.000000	0.000000	0.464968	0.210191



In this figure all parameters are slightly equal. But those students which use Library are 55% which have 3.3 CGPA ("0" is showing the values in which all Resource parameters are not depending.... )





**Sleep\_time vs Predicted\_GPA.....**

```
pd.crosstab(df.sleep_time, df.Predicted_GPA, normalize='index').plot(kind='Bar')
pd.crosstab(df.sleep_time, df.Predicted_GPA, normalize='index').plot(kind='line')
print('Average Sleep Time in a day {4 hours, 7 hours, 10 hours}')
pd.crosstab(df.sleep_time, df.Predicted_GPA, normalize='index')
```

Predicted\_GPA            2.0            2.35            2.7            3.0            3.25            3.3            3.8

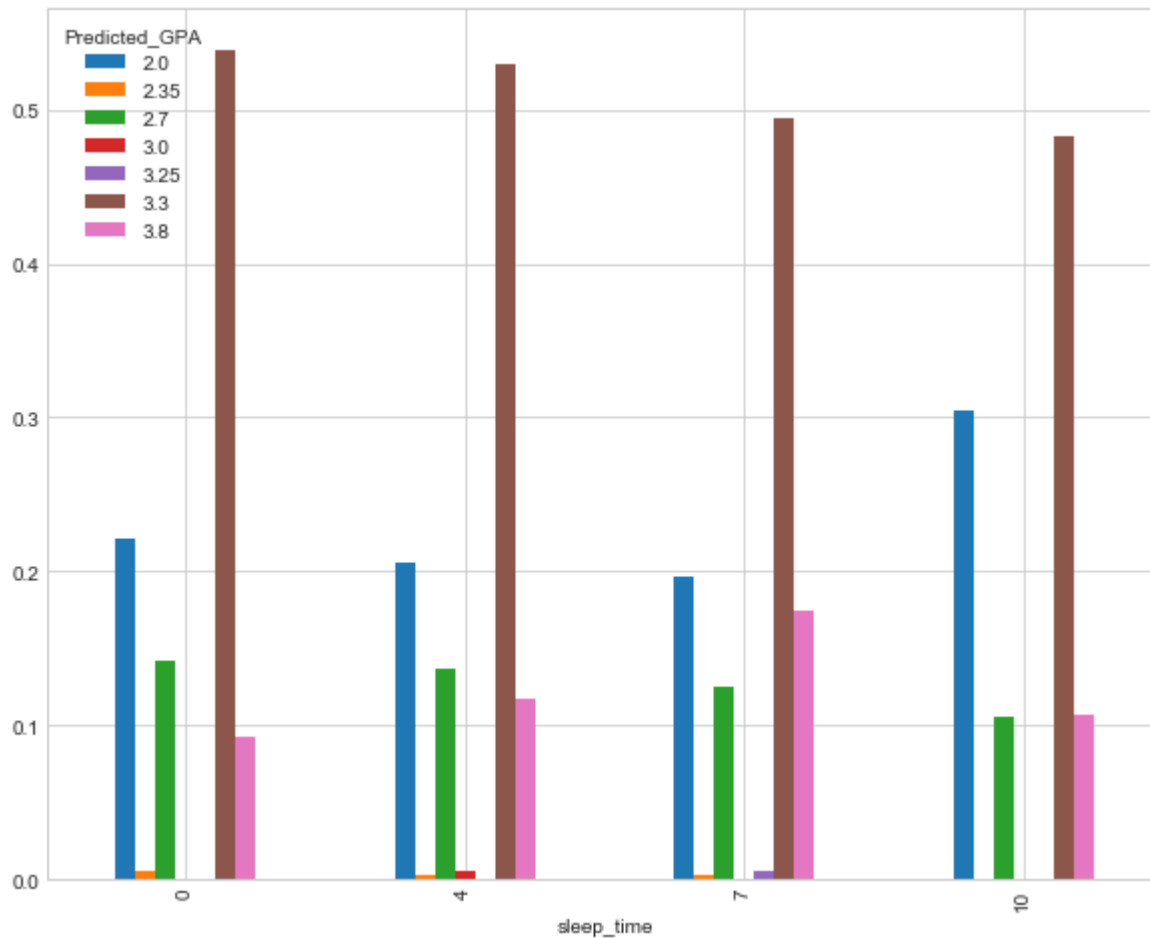
sleep\_time

0    0.221753    0.004966    0.141793    0.000497    0.000000    0.538614    0.092376

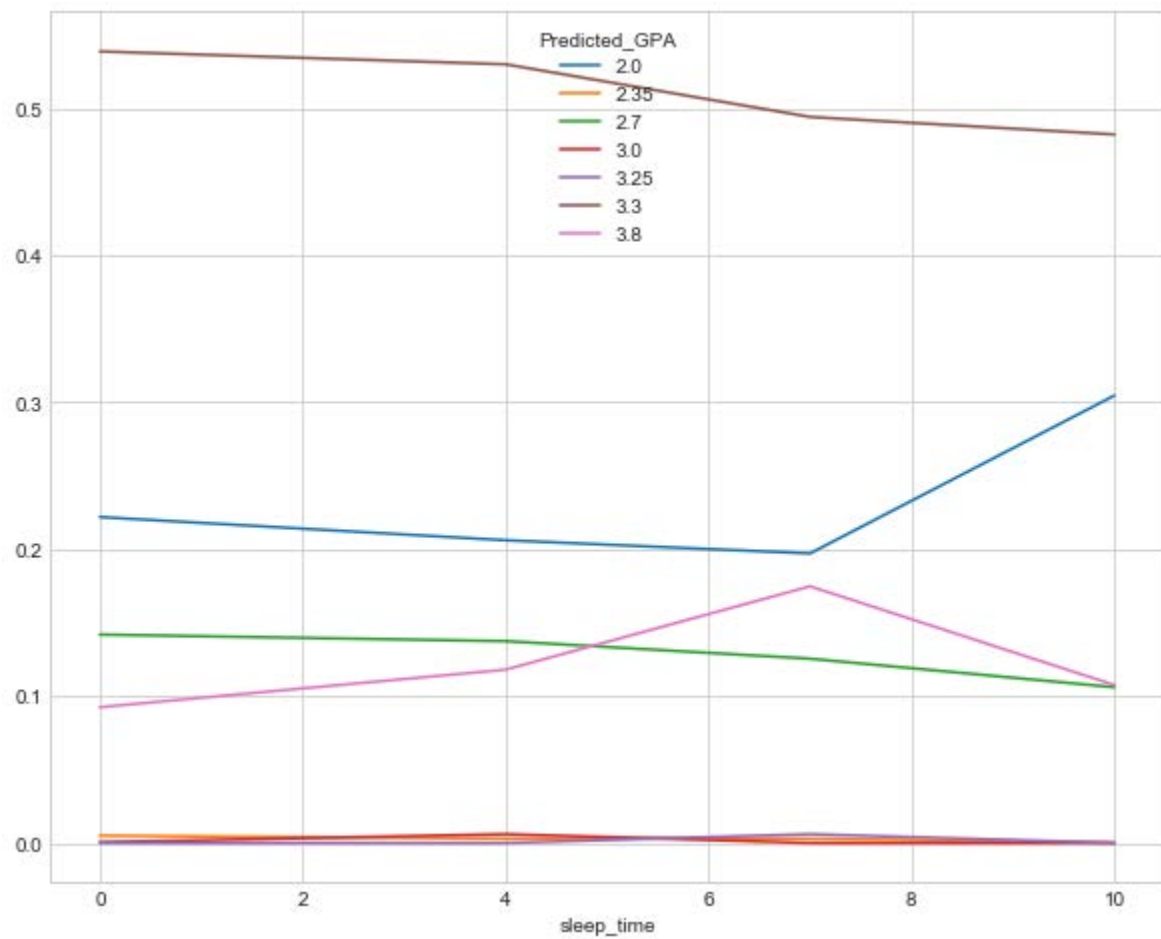
4    0.205970    0.002985    0.137313    0.005970    0.000000    0.529851    0.117910

7    0.197015    0.002985    0.125373    0.000000    0.00597    0.494030    0.174627

10    0.304478    0.000000    0.105970    0.000000    0.000000    0.482090    0.107463



In this figure as Sleep time is increasing 3.3 CGPA is decreasing and at average 7 hours sleep 3.8 CGPA is increasing and at average 10 hours sleep 2.0 CGPA is increasing and 3.8 is decreasing. ("0" is showing the values in which all Sleep\_time parameters are not depending....)



**Time\_spend\_on\_social\_media vs Predicted\_GPA.....**

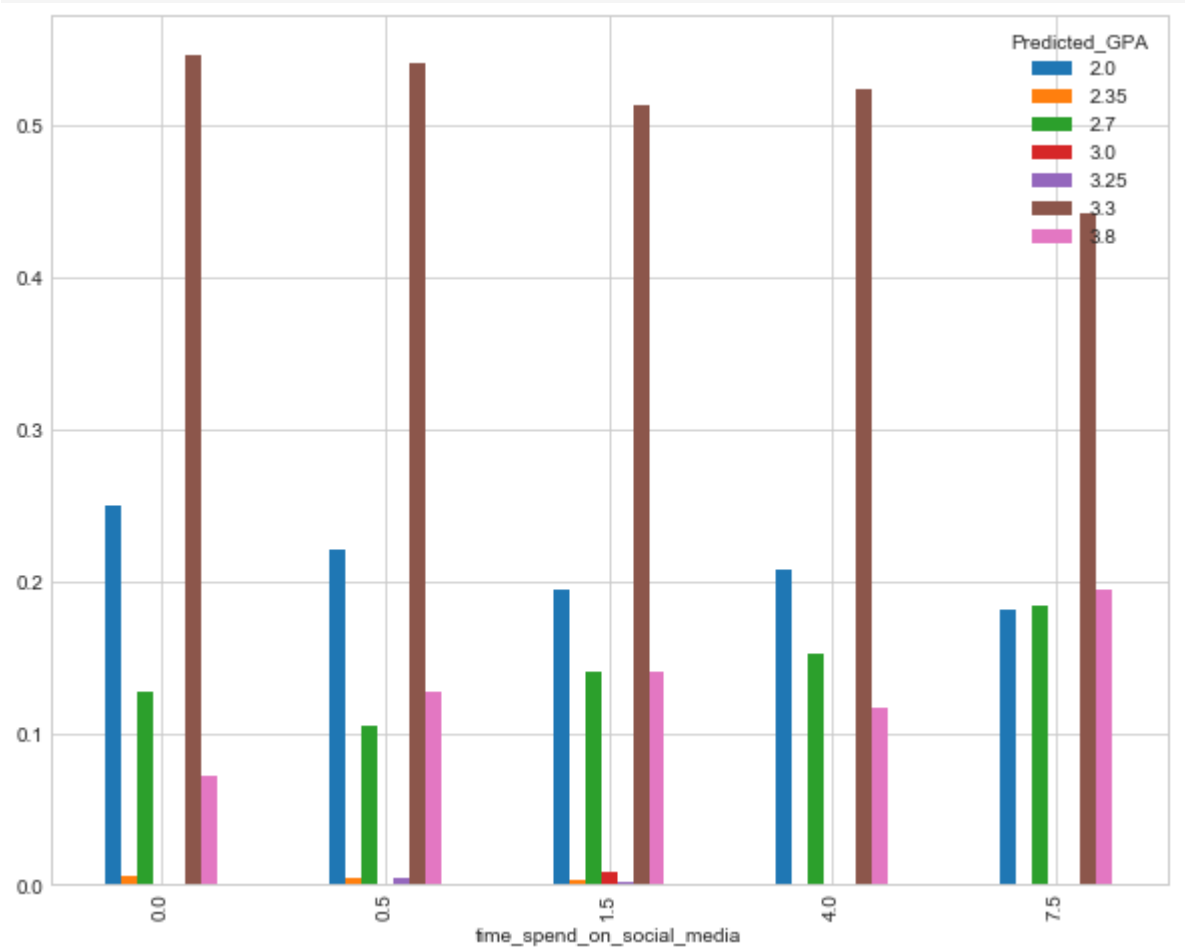
```
pd.crosstab(df.time_spend_on_social_media, df.Predicted_GPA, normalize='index').plot(kind='Bar')
```

```
pd.crosstab(df.time_spend_on_social_media, df.Predicted_GPA, normalize='index').plot(kind='line')
```

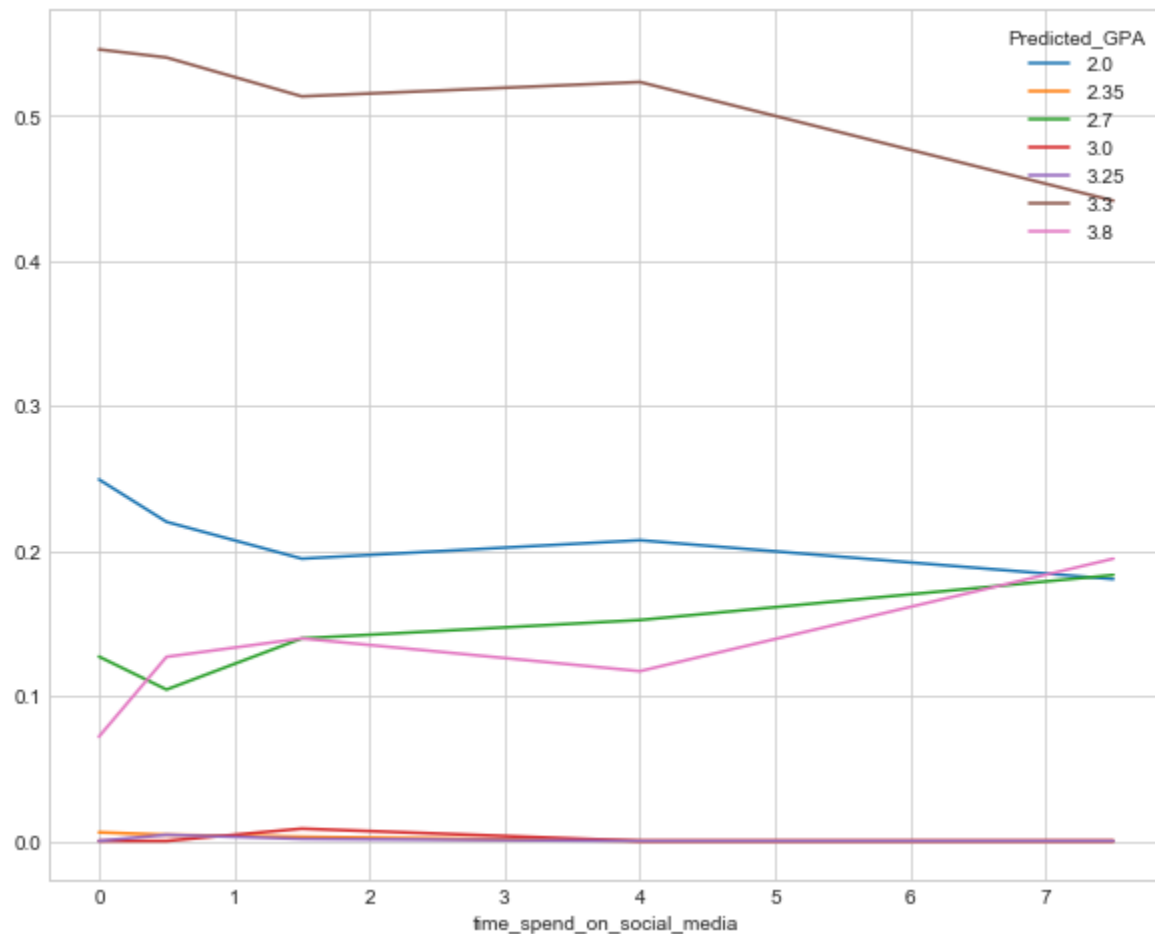
```
print('Place of Study {0.5 hours, 1.5 hours, 4 hours, 7.5 hours}')
```

```
pd.crosstab(df.time_spend_on_social_media, df.Predicted_GPA, normalize='index')
```

Predicted_GPA	2.0	2.35	2.7	3.0	3.25	3.3	3.8
0.0	0.249297	0.005936	0.127148	0.000000	0.000000	0.545767	0.071853
0.5	0.220028	0.004231	0.104372	0.000000	0.004231	0.540197	0.126939
1.5	0.194640	0.002821	0.139633	0.008463	0.001410	0.513399	0.139633
4.0	0.207334	0.000000	0.152327	0.000000	0.000000	0.523272	0.117066
7.5	0.180536	0.000000	0.183357	0.000000	0.000000	0.441467	0.194640



In this figure as time spend on social media is increasing 3.3 CGPA is decreasing and 2.7 CGPA is increasing ("0" is showing the values in which all time spend on social media parameters are not depending....)



**SSC vs Predicted\_GPA.....**

```
pd.crosstab(df.SSC, df.Predicted_GPA, normalize='index').plot(kind='Bar')
```

```
pd.crosstab(df.SSC, df.Predicted_GPA, normalize='index').plot(kind='line')
```

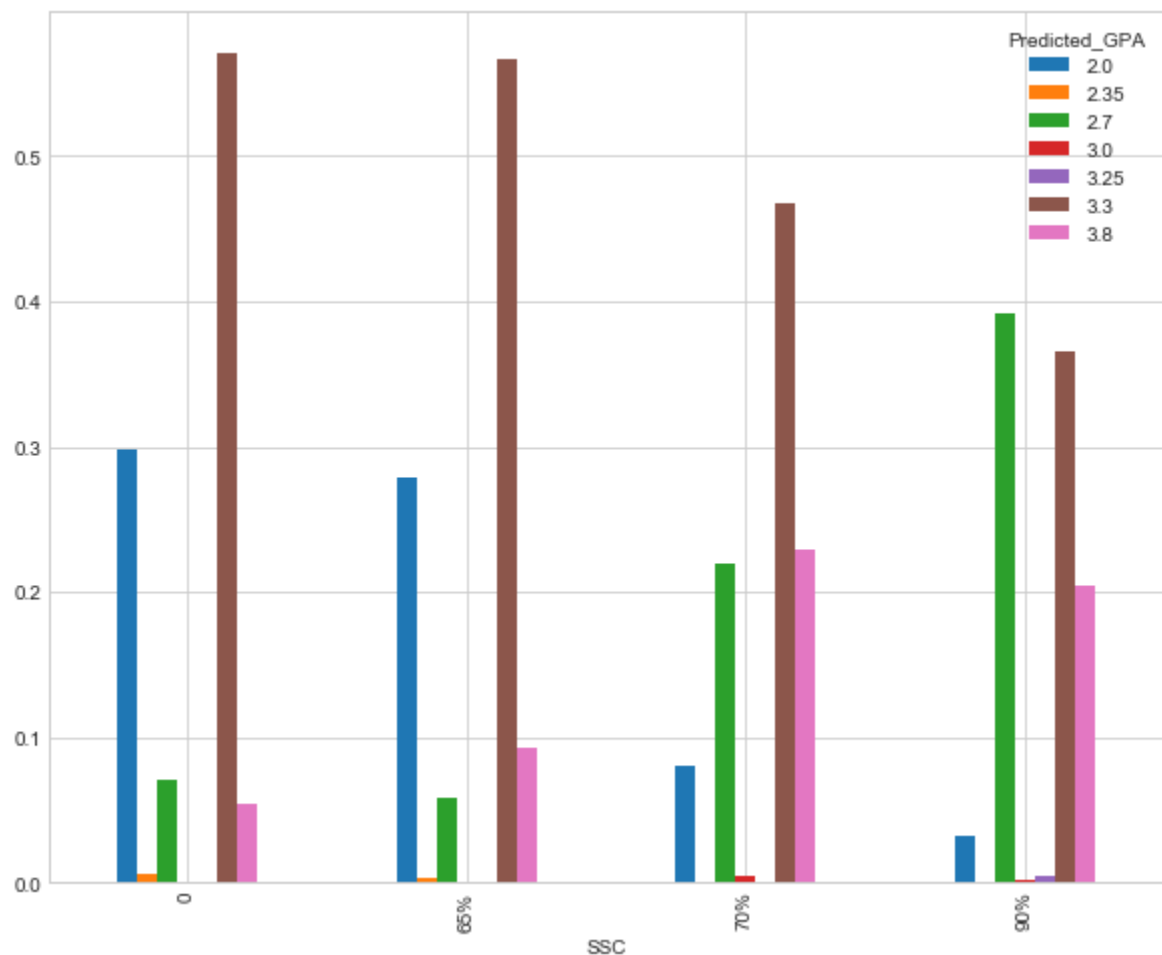
```
print('SSC Percentage{65%, 70%, 90%}')
```

```
pd.crosstab(df.SSC, df.Predicted_GPA, normalize='index')
```

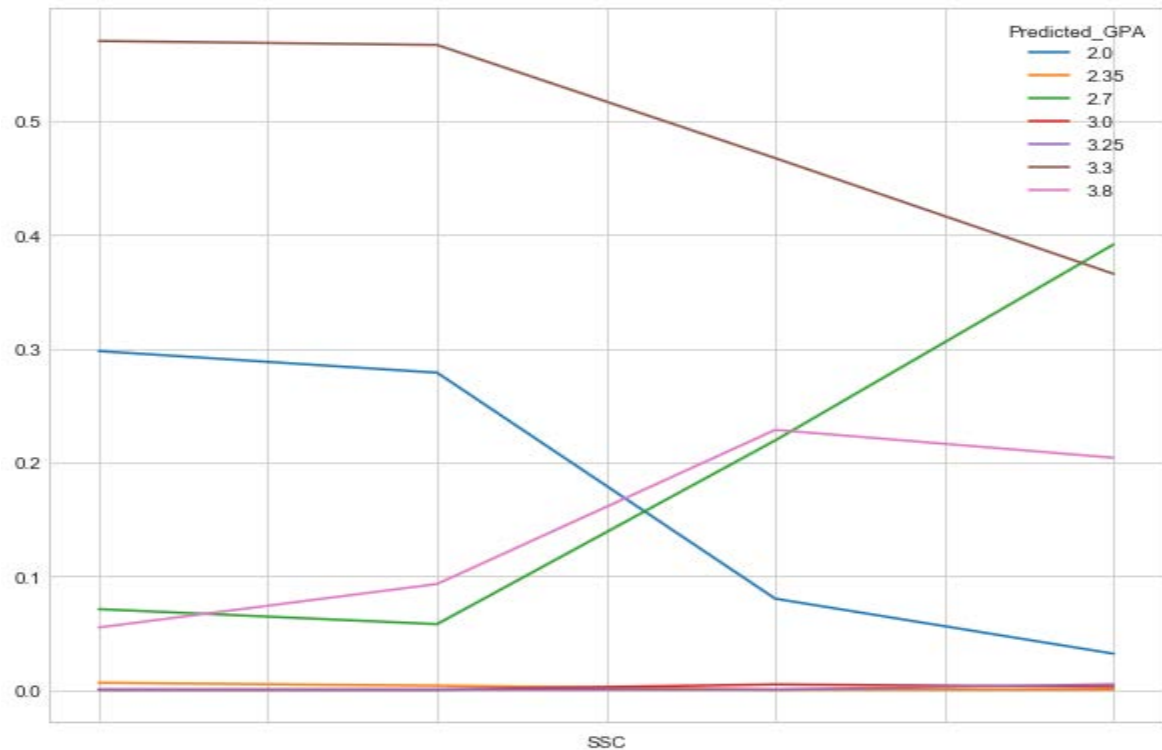
Predicted\_GPA            2.0            2.35            2.7            3.0            3.25            3.3            3.8

SSC

0	0.297811	0.006048	0.070853	0.000000	0.000000	0.570565	0.054724
65%	0.278824	0.003529	0.057647	0.000000	0.000000	0.567059	0.092941
70%	0.080046	0.000000	0.219258	0.004640	0.000000	0.467517	0.228538
90%	0.031653	0.000000	0.391559	0.002345	0.004689	0.365768	0.203986



In this figure as SSC percentage is increasing 3.3 CGPA is decreasing, 2.7 and 3.8 CGPA is increasing. ("0" is showing the values in which all SSC parameters are not depending....)



## HSC vs Predicted\_GPA.....

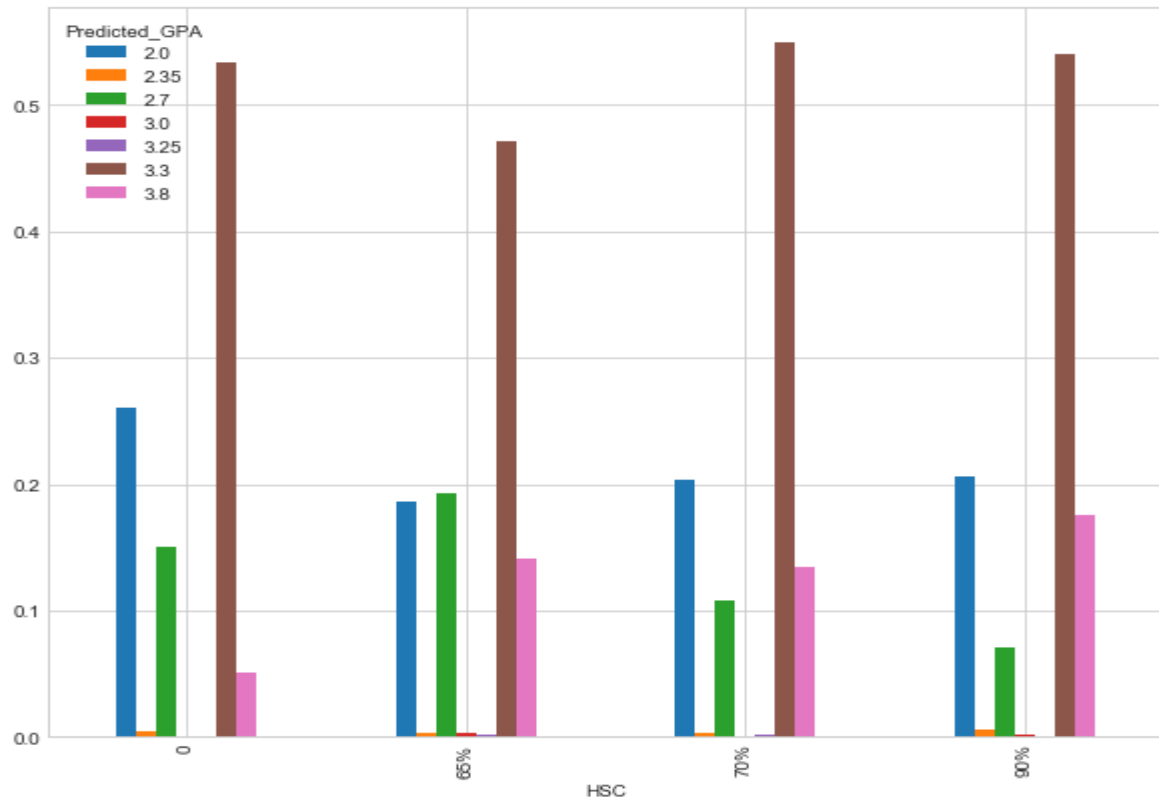
```
pd.crosstab(df.HSC, df.Predicted_GPA, normalize='index').plot(kind='Bar')
```

```
pd.crosstab(df.HSC, df.Predicted_GPA, normalize='index').plot(kind='line')
```

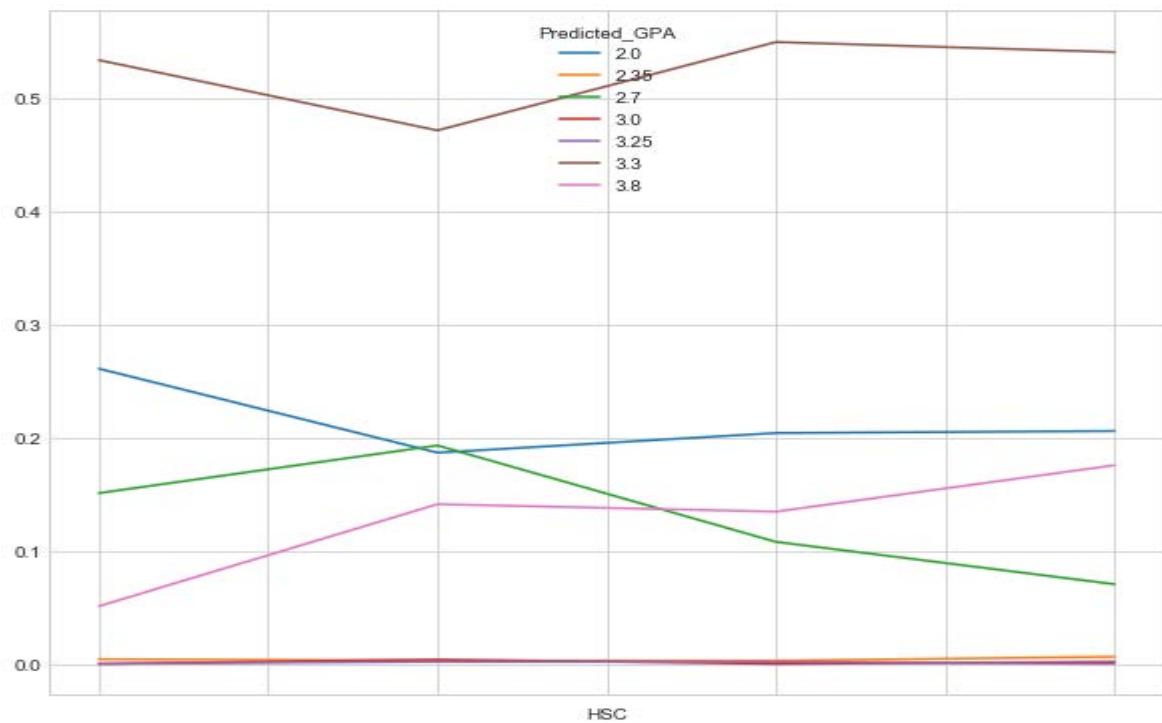
```
print('HSC Percentage{65%, 70%, 90%}')
```

```
pd.crosstab(df.HSC, df.Predicted_GPA, normalize='index')
```

Predicted_GPA	2.0	2.35	2.7	3.0	3.25	3.3	3.8
HSC							
0	0.260773	0.004086	0.150817	0.000000	0.000000	0.533432	0.050892
65%	0.186533	0.002730	0.192903	0.003640	0.001820	0.471338	0.141037
70%	0.203918	0.002671	0.107747	0.000000	0.001781	0.549421	0.134461
90%	0.205699	0.006233	0.070347	0.001781	0.000000	0.540516	0.175423



In this figure as HSC percentage is increasing 3.3 and 3.8 CGPA is also increasing. ("0" is showing the values in which all HSC parameters are not depending....)

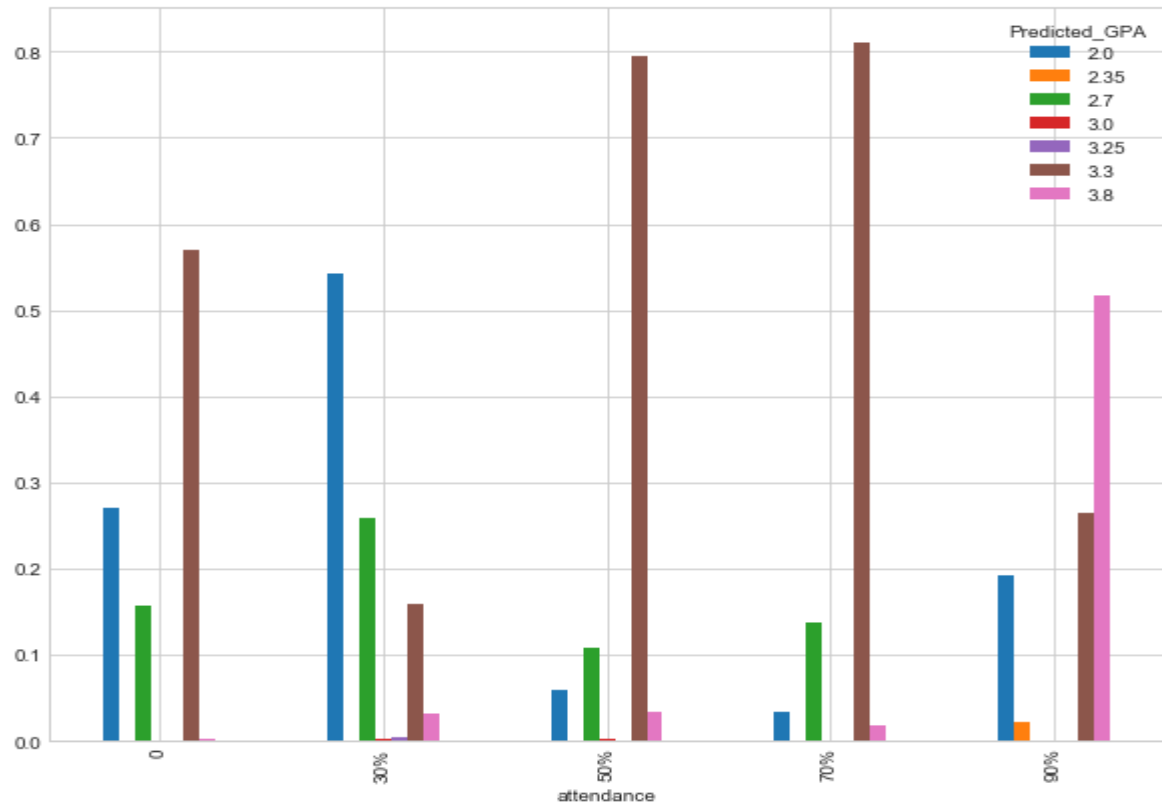




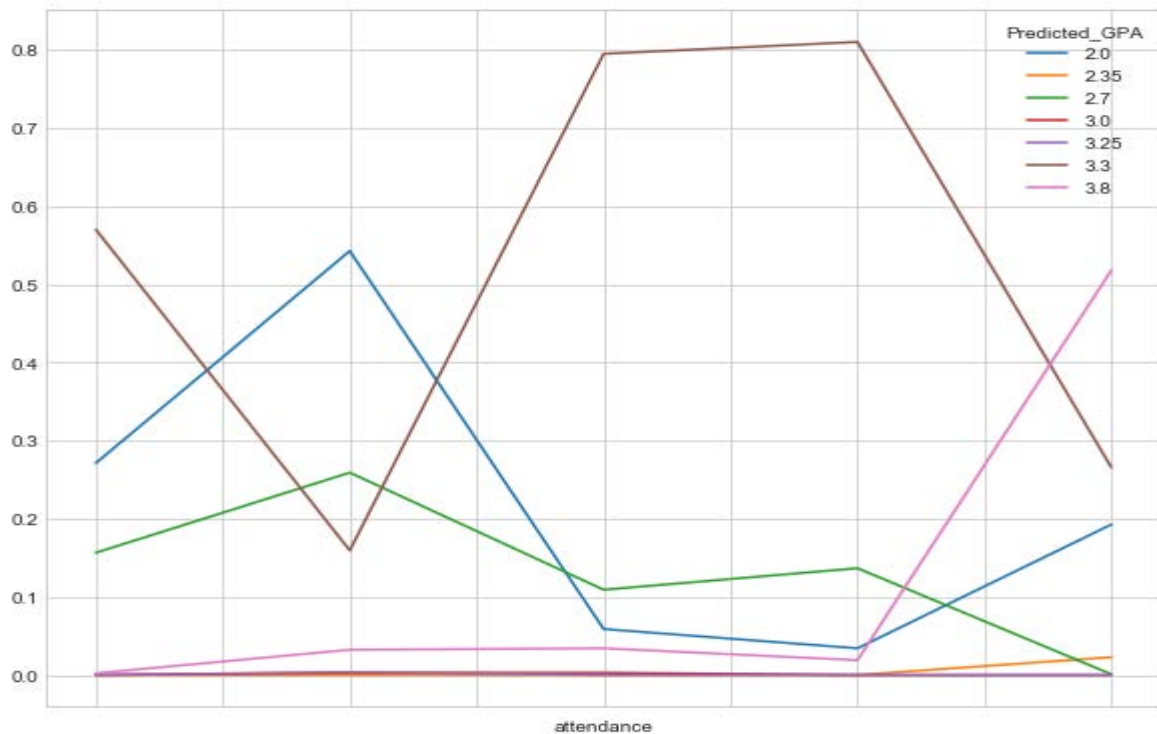
## Attendance vs Predicted\_GPA.....

```
pd.crosstab(df.attendance, df.Predicted_GPA, normalize='index').plot(kind='bar')
pd.crosstab(df.attendance, df.Predicted_GPA, normalize='index').plot(kind='line')
print('Average attendance{30%, 50%, 70%, 90%}')
pd.crosstab(df.attendance, df.Predicted_GPA, normalize='index')
```

Predicted_GPA	2.0	2.35	2.7	3.0	3.25	3.3	3.8
attendance							
0	0.271280	0.000000	0.156507	0.000000	0.000000	0.570016	0.002197
30%	0.542694	0.000000	0.259013	0.002846	0.003795	0.159393	0.032258
50%	0.058824	0.000000	0.109108	0.002846	0.000000	0.795066	0.034156
70%	0.034156	0.000000	0.136622	0.000000	0.000000	0.810247	0.018975
90%	0.192600	0.02277	0.000949	0.000000	0.000000	0.265655	0.518027



In this figure as Attendance is increasing the 3.3 CGPA is also increasing at 50% and 70% attendance, and at 90% attendance 3.8 CGPA is increasing. ("0" is showing the values in which all Attendance parameters are not depending....)



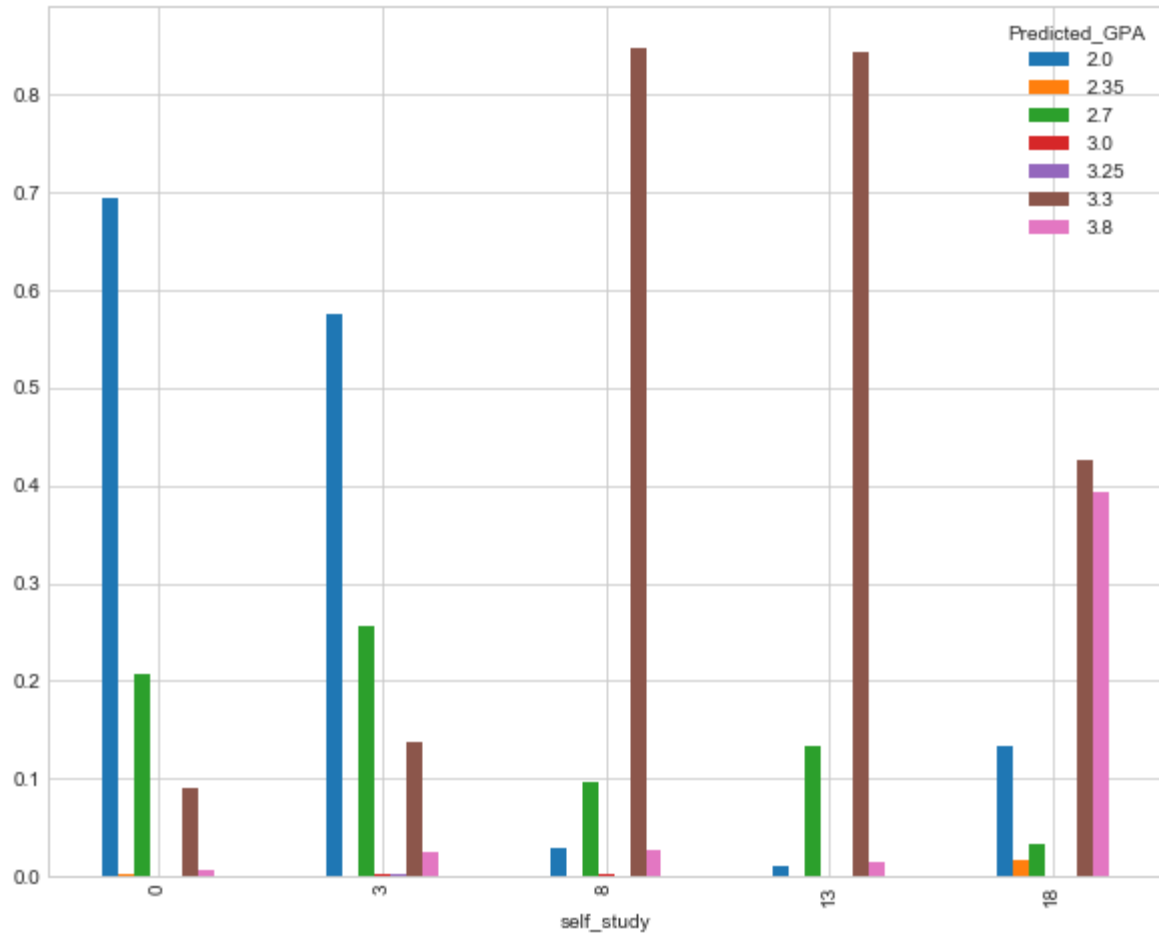
## Self\_study vs Predicted\_GPA.....

```
pd.crosstab(df.self_study, df.Predicted_GPA, normalize='index').plot(kind='Bar')
```

```
pd.crosstab(df.self_study, df.Predicted_GPA, normalize='index').plot(kind='line')
```

```
pd.crosstab(df.self_study, df.Predicted_GPA, normalize='index')
```

Predicted_GPA	2.0	2.35	2.7	3.0	3.25	3.3	3.8
self_study							
0	0.693920	0.002096	0.207547	0.000000	0.000000	0.090147	0.006289
3	0.574820	0.000000	0.256835	0.002158	0.002878	0.138129	0.025180
8	0.028058	0.000000	0.095683	0.002158	0.000000	0.848201	0.025899
13	0.009353	0.000000	0.132374	0.000000	0.000000	0.843885	0.014388
18	0.133094	0.016547	0.032374	0.000000	0.000000	0.425180	0.392806



In this figure as if average Self-study Time is 3 hours per week the CGPA is decreasing to 2 CGPA and if average Self-study Time is 8 and 13 hours per week the CGPA is increasing to 3.3 CGPA and if average Self-study Time is 18 hours per week the CGPA is increasing from 3.3 to 3.8 CGPA. ("0" is showing the values in which all Self-study Time parameters are not depending....)

