

Student Behavior Project

Hello, and thank you for taking a look. I decided to take a look at student grades to see how different factors impact them.

For example, what is their commute time, social media usage, and daily studying?

Below you'll find data analysis, visualizations, and more

Import Pandas and Numpy

It's vital to import pandas and numpy for my work, these are what's known as libraries within Python. They'll help me with really diving into my data and making sense of it.

```
#import pandas and numpy for now
```

```
import pandas as pd
import numpy as np
```

Import data from the csv

```
#import csv
student_behavior = pd.read_csv('Student_Behavior.csv')
student_behavior
```

	Certification	Course	Gender	Department	Height(CM)	Weight(KG)	\
0		No	Male	BCA	100.0	58.0	
1		No	Female	BCA	90.0	40.0	
2		Yes	Male	BCA	159.0	78.0	
3		Yes	Female	BCA	147.0	20.0	
4		No	Male	BCA	170.0	54.0	
...		
230		Yes	Male	B.com ISM	170.0	76.0	
231		Yes	Male	B.com ISM	172.0	52.0	
232		Yes	Female	BCA	139.0	33.0	
233		Yes	Female	Commerce	153.0	58.0	
234		No	Female	B.com ISM	155.0	39.0	

	10th Mark	12th Mark	college mark	hobbies	daily studing
time \					
0	79.0	64.0	80.0	Video Games	0 - 30
minute					
1	70.0	80.0	70.0	Cinema	30 - 60
minute					
2	71.0	61.0	55.0	Cinema	1 - 2
Hour					
3	70.0	59.0	58.0	Reading books	1 - 2

Hour					
4	40.0	65.0	30.0	Video Games	30 - 60
minute					
..	
...					
230	72.0	67.0	65.0	Video Games	30 - 60
minute					
231	72.0	70.0	76.0	Cinema	2 - 3
hour					
232	90.0	75.0	70.0	Cinema	30 - 60
minute					
233	85.0	74.0	75.0	Cinema	0 - 30
minute					
234	45.0	45.0	50.0	Sports	3 - 4
hour					

	prefer to study in	salary expectation	Do you like your degree?	\
0	Morning	40000	No	
1	Morning	15000	Yes	
2	Anytime	13000	Yes	
3	Anytime	1500000	No	
4	Morning	50000	Yes	
..	
230	Morning	7000	Yes	
231	Anytime	25000	Yes	
232	Night	20000	Yes	
233	Anytime	20000	Yes	
234	Night	10	No	

	willingness to pursue a career based on their degree	\
0	50%	
1	75%	
2	50%	
3	50%	
4	25%	
..	...	
230	50%	
231	75%	
232	50%	
233	75%	
234	75%	

	social media & video	Travelling Time	Stress Level	Financial
Status \				
0	1.30 - 2 hour	30 - 60 minutes	Bad	
Bad				
1	1 - 1.30 hour	0 - 30 minutes	Bad	
Bad				
2	More than 2 hour	30 - 60 minutes	Awful	
Bad				

```

3          1.30 - 2 hour    0 - 30 minutes          Bad
good
4          1.30 - 2 hour    30 - 60 minutes          Good
good
..          ...          ...          ...
..
230        More than 2 hour    30 - 60 minutes          Bad
Bad
231          30 - 60 Minute    30 - 60 minutes          Good
good
232          30 - 60 Minute    2.30 - 3 hour          Bad
good
233        More than 2 hour    2.30 - 3 hour          Awful
good
234          30 - 60 Minute    2.30 - 3 hour          fabulous
good

```

```

part-time job
0          No
1          No
2          No
3          No
4          No
..          ...
230         No
231         No
232         No
233         No
234         No

```

```
[235 rows x 19 columns]
```

Column Replacements

Here I'm replacing the blanks in my column names with a dash which is standard in Python practice and more readable

```

#replacing column with dashes for the column name

for col in student_behavior.columns:
    student_behavior.rename(columns = {col: col.lower().replace(' ', '_')}, inplace = True)

```

Viewing the dataframe

Viewing the dataframe to ensure the dashes are there for the column names. Notice daily_studing_time as an example. As you can see this is what's known as "messy data" since there are misspellings, missing data, and more. We can clean it up!

student_behavior

	certification_course	gender	department	height(cm)	weight(kg)	\
0	No	Male	BCA	100.0	58.0	
1	No	Female	BCA	90.0	40.0	
2	Yes	Male	BCA	159.0	78.0	
3	Yes	Female	BCA	147.0	20.0	
4	No	Male	BCA	170.0	54.0	
..	
230	Yes	Male	B.com ISM	170.0	76.0	
231	Yes	Male	B.com ISM	172.0	52.0	
232	Yes	Female	BCA	139.0	33.0	
233	Yes	Female	Commerce	153.0	58.0	
234	No	Female	B.com ISM	155.0	39.0	

	10th_mark	12th_mark	college_mark	hobbies	
daily_studing_time \					
0	79.0	64.0	80.0	Video Games	0 - 30
minute					
1	70.0	80.0	70.0	Cinema	30 - 60
minute					
2	71.0	61.0	55.0	Cinema	1 - 2
Hour					
3	70.0	59.0	58.0	Reading books	1 - 2
Hour					
4	40.0	65.0	30.0	Video Games	30 - 60
minute					
..	
...					
230	72.0	67.0	65.0	Video Games	30 - 60
minute					
231	72.0	70.0	76.0	Cinema	2 - 3
hour					
232	90.0	75.0	70.0	Cinema	30 - 60
minute					
233	85.0	74.0	75.0	Cinema	0 - 30
minute					
234	45.0	45.0	50.0	Sports	3 - 4
hour					

	prefer_to_study_in	salary_expectation	do_you_like_your_degree?	\
0	Morning	40000	No	
1	Morning	15000	Yes	
2	Anytime	13000	Yes	
3	Anytime	1500000	No	
4	Morning	50000	Yes	
..	
230	Morning	7000	Yes	
231	Anytime	25000	Yes	
232	Night	20000	Yes	

233	Anytime	20000	Yes
234	Night	10	No

	willingness_to_pursue_a_career_based_on_their_degree__ \
0	50%
1	75%
2	50%
3	50%
4	25%
..	...
230	50%
231	75%
232	50%
233	75%
234	75%

	social_medai_&_video	travelling_time_	stress_level_
	financial_status \		
0	1.30 - 2 hour	30 - 60 minutes	Bad
Bad			
1	1 - 1.30 hour	0 - 30 minutes	Bad
Bad			
2	More than 2 hour	30 - 60 minutes	Awful
Bad			
3	1.30 - 2 hour	0 - 30 minutes	Bad
good			
4	1.30 - 2 hour	30 - 60 minutes	Good
good			
..
..			
230	More than 2 hour	30 - 60 minutes	Bad
Bad			
231	30 - 60 Minute	30 - 60 minutes	Good
good			
232	30 - 60 Minute	2.30 - 3 hour	Bad
good			
233	More than 2 hour	2.30 - 3 hour	Awful
good			
234	30 - 60 Minute	2.30 - 3 hour	fabulous
good			

	part-time_job
0	No
1	No
2	No
3	No
4	No
..	...
230	No
231	No

```
232      No
233      No
234      No
```

```
[235 rows x 19 columns]
```

Stress Levels

Here I'm viewing the different stress levels, data type(object, so it's not a number), the length, and the name.

```
#stress level
stress = student_behavior['stress_level_']
stress

0      Bad
1      Bad
2     Awful
3      Bad
4      Good
...
230     Bad
231     Good
232     Bad
233     Awful
234  fabulous
Name: stress_level_, Length: 235, dtype: object
```

Machine Learning

I enjoy machine learning so decided to import it in case I'd like to use it at some point.

```
#machine learning imports
import sys
from packaging import version
import sklearn
```

First few rows

The .head() function allows me to see the first 5 rows of the dataframe. I also have the option to place a number inside of the head function to state my preference.

```
#use .head() for first few rows
student_behavior.head()

   certification_course  gender department  height(cm)  weight(kg)
10th_mark \
0              No      Male      BCA      100.0      58.0
```

79.0					
1	No	Female	BCA	90.0	40.0
70.0					
2	Yes	Male	BCA	159.0	78.0
71.0					
3	Yes	Female	BCA	147.0	20.0
70.0					
4	No	Male	BCA	170.0	54.0
40.0					

	12th_mark	college_mark	hobbies	daily_studying_time	\
0	64.0	80.0	Video Games	0 - 30 minute	
1	80.0	70.0	Cinema	30 - 60 minute	
2	61.0	55.0	Cinema	1 - 2 Hour	
3	59.0	58.0	Reading books	1 - 2 Hour	
4	65.0	30.0	Video Games	30 - 60 minute	

	prefer_to_study_in	salary_expectation	do_you_like_your_degree?	\
0	Morning	40000	No	
1	Morning	15000	Yes	
2	Anytime	13000	Yes	
3	Anytime	1500000	No	
4	Morning	50000	Yes	

willingness_to_pursue_a_career_based_on_their_degree__		
social_medai_&_video	\	
0	50%	1.30 -
2 hour		
1	75%	1 -
1.30 hour		
2	50%	More than
2 hour		
3	50%	1.30 -
2 hour		
4	25%	1.30 -
2 hour		

	travelling_time_	stress_level_	financial_status	part-time_job
0	30 - 60 minutes	Bad	Bad	No
1	0 - 30 minutes	Bad	Bad	No
2	30 - 60 minutes	Awful	Bad	No
3	0 - 30 minutes	Bad	good	No
4	30 - 60 minutes	Good	good	No

Dataframe info

.info(). This is important since objects and numbers don't work the same. Float64 and int64 means it's a number, and object as stated above means it's a non-number.

```
#check the info
```

```
student_behavior.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 235 entries, 0 to 234
```

```
Data columns (total 19 columns):
```

#	Column	Count	Dtype	Non-Null
---	--------	-------	-------	----------

```
---  ---
```

```
-----
```

0	certification_course	235	non-
---	----------------------	-----	------

```
null object
```

1	gender	235	non-
---	--------	-----	------

```
null object
```

2	department	235	non-
---	------------	-----	------

```
null object
```

3	height(cm)	235	non-
---	------------	-----	------

```
null float64
```

4	weight(kg)	235	non-
---	------------	-----	------

```
null float64
```

5	10th_mark	235	non-
---	-----------	-----	------

```
null float64
```

6	12th_mark	235	non-
---	-----------	-----	------

```
null float64
```

7	college_mark	235	non-
---	--------------	-----	------

```
null float64
```

8	hobbies	235	non-
---	---------	-----	------

```
null object
```

9	daily_studing_time	235	non-
---	--------------------	-----	------

```
null object
```

10	prefer_to_study_in	235	non-
----	--------------------	-----	------

```
null object
```

11	salary_expectation	235	non-
----	--------------------	-----	------

```
null int64
```

12	do_you_like_your_degree?	235	non-
----	--------------------------	-----	------

```
null object
```

13	willingness_to_pursue_a_career_based_on_their_degree__	235	non-
----	--	-----	------

```
null object
```

14	social_medai_&_video	235	non-
----	----------------------	-----	------

```
null object
```

15	travelling_time_	235	non-
----	------------------	-----	------

```
null object
```

16	stress_level_	235	non-
----	---------------	-----	------

```
null object
```

17	financial_status	235	non-
----	------------------	-----	------

```
null object
```

18	part-time_job	235	non-
----	---------------	-----	------

```
null object
```

```
dtypes: float64(5), int64(1), object(13)
```

```
memory usage: 35.0+ KB
```


Change information to numbers

In Pandas I need to change the information to numbers. See how it states "Bad" "Good" etc? I'd like to gain access to the data for better visualization later on through numerical data.

So for stress levels I'm changing fabulous to 1, good to 2, bad to 3, and awful to 4

Then I'm taking that information and placing it into a new column called stress_level_number.

```
#stress map dictionary
```

```
stress_mapping = {'fabulous': 1, 'good': 2, 'bad': 3, 'awful': 4}
```

```
#creating a new dataframe
```

```
student_behavior['stress_level_number'] =
```

```
student_behavior['stress_level_'].str.lower().replace(stress_mapping)
```

```
student_behavior
```

	certification_course	gender	department	height(cm)	weight(kg)	\
0	No	Male	BCA	100.0	58.0	
1	No	Female	BCA	90.0	40.0	
2	Yes	Male	BCA	159.0	78.0	
3	Yes	Female	BCA	147.0	20.0	
4	No	Male	BCA	170.0	54.0	
..	
230	Yes	Male	B.com ISM	170.0	76.0	
231	Yes	Male	B.com ISM	172.0	52.0	
232	Yes	Female	BCA	139.0	33.0	
233	Yes	Female	Commerce	153.0	58.0	
234	No	Female	B.com ISM	155.0	39.0	

	10th_mark	12th_mark	college_mark	hobbies	
daily_studing_time \					
0	79.0	64.0	80.0	Video Games	0 - 30
minute					
1	70.0	80.0	70.0	Cinema	30 - 60
minute					
2	71.0	61.0	55.0	Cinema	1 - 2
Hour					
3	70.0	59.0	58.0	Reading books	1 - 2
Hour					
4	40.0	65.0	30.0	Video Games	30 - 60
minute					
..	
...					
230	72.0	67.0	65.0	Video Games	30 - 60
minute					
231	72.0	70.0	76.0	Cinema	2 - 3
hour					
232	90.0	75.0	70.0	Cinema	30 - 60
minute					

233	85.0	74.0	75.0	Cinema	0 - 30
minute					
234	45.0	45.0	50.0	Sports	3 - 4
hour					

	prefer_to_study_in	salary_expectation	do_you_like_your_degree?	\
0	Morning	40000	No	
1	Morning	15000	Yes	
2	Anytime	13000	Yes	
3	Anytime	1500000	No	
4	Morning	50000	Yes	
..	
230	Morning	7000	Yes	
231	Anytime	25000	Yes	
232	Night	20000	Yes	
233	Anytime	20000	Yes	
234	Night	10	No	

	willingness_to_pursue_a_career_based_on_their_degree__	\
0	50%	
1	75%	
2	50%	
3	50%	
4	25%	
..	...	
230	50%	
231	75%	
232	50%	
233	75%	
234	75%	

	social_medai_&_video	travelling_time_	stress_level_	financial_status_	\
0	1.30 - 2 hour	30 - 60 minutes	Bad		
Bad					
1	1 - 1.30 hour	0 - 30 minutes	Bad		
Bad					
2	More than 2 hour	30 - 60 minutes	Awful		
Bad					
3	1.30 - 2 hour	0 - 30 minutes	Bad		
good					
4	1.30 - 2 hour	30 - 60 minutes	Good		
good					
..
..					
230	More than 2 hour	30 - 60 minutes	Bad		
Bad					
231	30 - 60 Minute	30 - 60 minutes	Good		
good					
232	30 - 60 Minute	2.30 - 3 hour	Bad		

```

good
233      More than 2 hour      2.30 - 3 hour      Awful
good
234      30 - 60 Minute      2.30 - 3 hour      fabulous
good

      part-time_job      stress_level_number
0              No              3
1              No              3
2              No              4
3              No              3
4              No              2
..          ...
230          No              3
231          No              2
232          No              3
233          No              4
234          No              1

[235 rows x 20 columns]

```

Stress Level information

This tells me how many students have good, bad, awful, and fabulous stress levels.

```

#values in the stress_level_column
student_behavior["stress_level_"].value_counts()

stress_level_
Good      137
Bad       68
Awful     19
fabulous  11
Name: count, dtype: int64

```

Salary Expectations

Notice the salary expectations below. You'll see how many students per salary have that expectation.

```

#values of salary expectations
student_behavior["salary_expectation"].value_counts()

salary_expectation
20000      64
15000     49

```

```
25000    22
10000    16
30000    13
18000    13
50000    10
12000     5
40000     5
17000     4
13000     4
100000     3
5000      3
10        2
35000     2
16000     2
22        1
15        1
100       1
45000     1
500       1
700000    1
500000    1
23000     1
21        1
120000    1
11000     1
12500     1
8000      1
0         1
60000     1
17        1
1500000   1
7000      1
Name: count, dtype: int64
```

Import Matplotlib plus visualization

Here I'm importing matplotlib for visualization purposes. Visualization really helps with taking a look at the data and seeing any connections.

It shows the height, weight, grade, and more. It's a great way to see what's the most common.

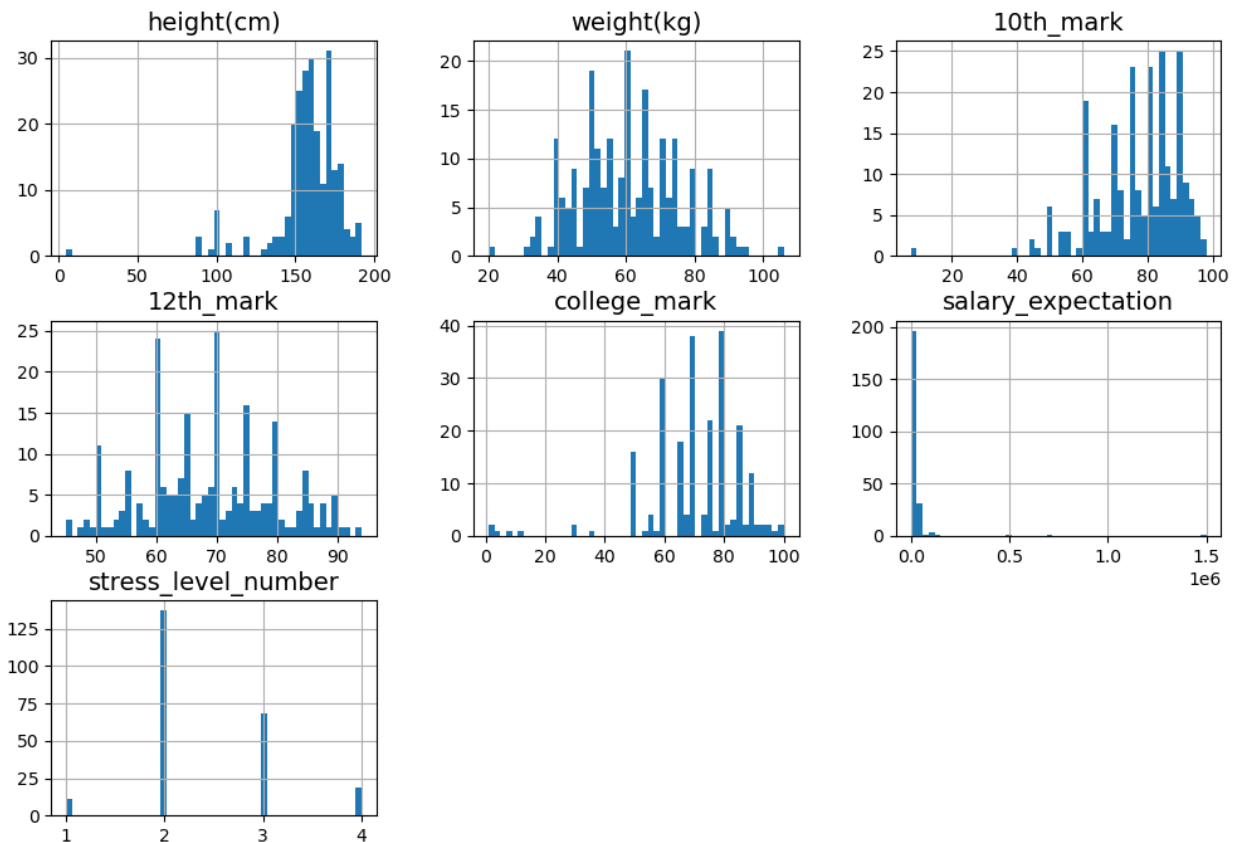
For example, most of the salary expectations are low, below \$20,000!

```
#import matplotlib as plt
import matplotlib.pyplot as plt

# visualization time
plt.rc('font', size=14)
plt.rc('axes', labelsz=14, titlesize=14)
```

```
plt.rc('legend', fontsize=14)
plt.rc('xtick', labels=10)
plt.rc('ytick', labels=10)

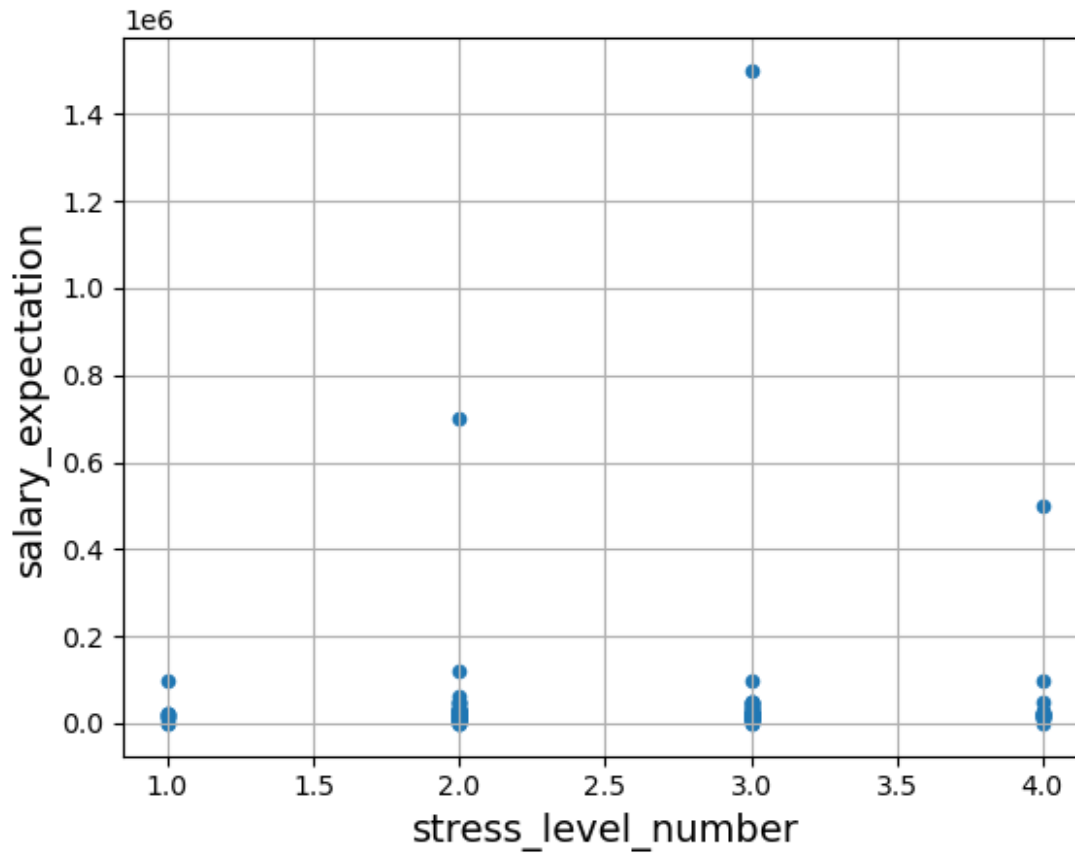
# plots histogram
student_behavior.hist(bins=50, figsize=(12, 8))
plt.show()
```



Salary expectation vs stress level

Is there a correlation? Here it's interesting that many students seem to have a lower salary expectation overall.

```
#scatter plot of salary and stress level
student_behavior.plot(kind="scatter", x= "stress_level_number",
y="salary_expectation", grid=True)
plt.show()
```

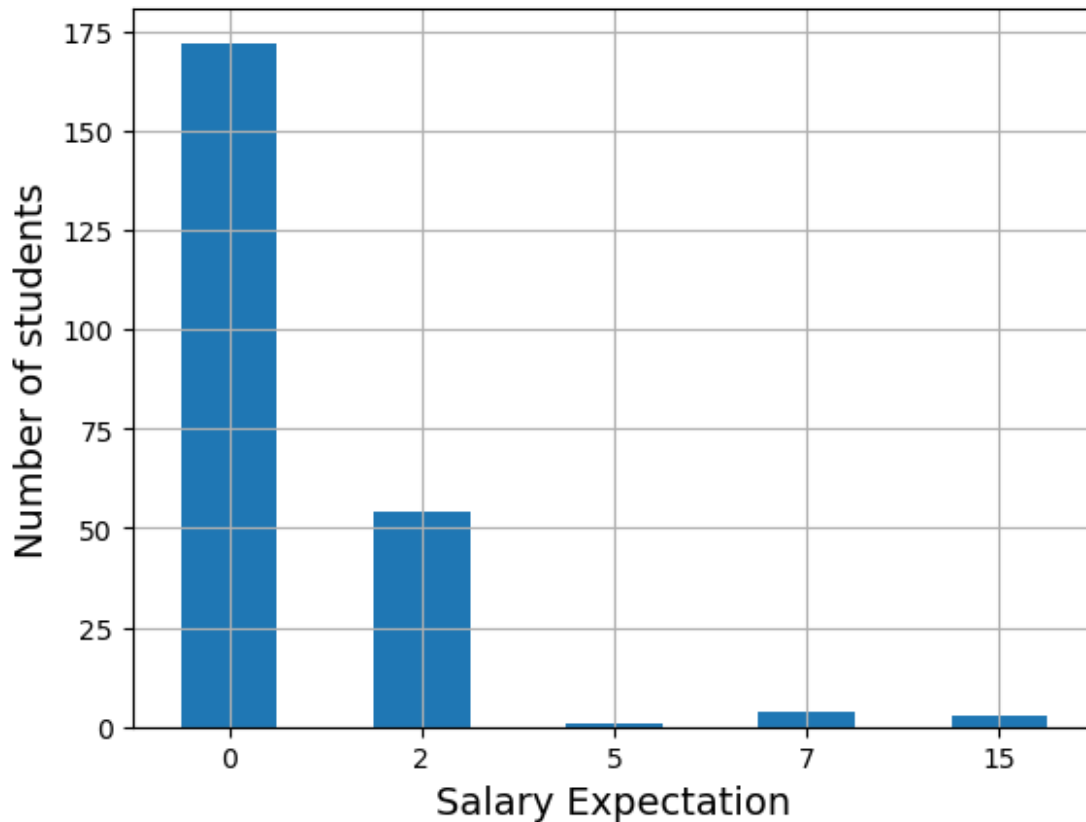


Salary Expectations

Here you can see the number of students and their salary expectations. 0 is under 20,000, 2 is 20,000 but under 50,000, etc. As you can see, many of the students are under \$20,000 for their salary expectations.

```
#salary expectations for each student with binning
student_behavior["salary_expectations"] =
pd.cut(student_behavior["salary_expectation"],
        bins=[0., 20000, 50000, 70000, 150000,
np.inf],
        labels=[0, 2, 5, 7, 15])

student_behavior["salary_expectations"].value_counts().sort_index().plot.bar(rot=0, grid=True)
plt.xlabel("Salary Expectation")
plt.ylabel("Number of students")
plt.show()
```



Column renaming

I'm renaming columns to make them easier to read. Inplace = True assigns them back to the original dataframe (student_behavior)

```
#renaming columns for readability
student_behavior.rename(columns = {'social_medai_&_video':
'social_media'}, inplace = True)
student_behavior

student_behavior.rename(columns =
{'willingness_to_pursue_a_career_based_on_their_degree__':
'pursue_career_in_field'}, inplace = True)
student_behavior.rename(columns = {'do_you_like_your_degree?':
'degree_satisfaction'}, inplace = True)

student_behavior.rename(columns = {'daily_studing_time':
'daily_studying'}, inplace = True)

student_behavior.rename(columns = {'travelling_time_': 'commute'},
inplace = True)
student_behavior
```

	certification_course	gender	department	height(cm)	weight(kg)	\
0	No	Male	BCA	100.0	58.0	
1	No	Female	BCA	90.0	40.0	
2	Yes	Male	BCA	159.0	78.0	
3	Yes	Female	BCA	147.0	20.0	
4	No	Male	BCA	170.0	54.0	
...	
230	Yes	Male	B.com ISM	170.0	76.0	
231	Yes	Male	B.com ISM	172.0	52.0	
232	Yes	Female	BCA	139.0	33.0	
233	Yes	Female	Commerce	153.0	58.0	
234	No	Female	B.com ISM	155.0	39.0	

	10th_mark	12th_mark	college_mark	hobbies	daily_studying
...	\				
0	79.0	64.0	80.0	Video Games	0 - 30 minute
...					
1	70.0	80.0	70.0	Cinema	30 - 60 minute
...					
2	71.0	61.0	55.0	Cinema	1 - 2 Hour
...					
3	70.0	59.0	58.0	Reading books	1 - 2 Hour
...					
4	40.0	65.0	30.0	Video Games	30 - 60 minute
...					
...
...					
230	72.0	67.0	65.0	Video Games	30 - 60 minute
...					
231	72.0	70.0	76.0	Cinema	2 - 3 hour
...					
232	90.0	75.0	70.0	Cinema	30 - 60 minute
...					
233	85.0	74.0	75.0	Cinema	0 - 30 minute
...					
234	45.0	45.0	50.0	Sports	3 - 4 hour
...					

	salary_expectation	degree_satisfaction	pursue_career_in_field	\
0	40000	No	50%	
1	15000	Yes	75%	
2	13000	Yes	50%	
3	1500000	No	50%	
4	50000	Yes	25%	
...	
230	7000	Yes	50%	
231	25000	Yes	75%	
232	20000	Yes	50%	
233	20000	Yes	75%	
234	10	No	75%	

	social_media	commute	stress_level_	financial_status
0	1.30 - 2 hour	30 - 60 minutes	Bad	Bad
1	1 - 1.30 hour	0 - 30 minutes	Bad	Bad
2	More than 2 hour	30 - 60 minutes	Awful	Bad
3	1.30 - 2 hour	0 - 30 minutes	Bad	good
4	1.30 - 2 hour	30 - 60 minutes	Good	good
..
230	More than 2 hour	30 - 60 minutes	Bad	Bad
231	30 - 60 Minute	30 - 60 minutes	Good	good
232	30 - 60 Minute	2.30 - 3 hour	Bad	good
233	More than 2 hour	2.30 - 3 hour	Awful	good
234	30 - 60 Minute	2.30 - 3 hour	fabulous	good

	part-time_job	stress_level_number	salary_expectations
0	No	3	2
1	No	3	0
2	No	4	0
3	No	3	15
4	No	2	2
..
230	No	3	0
231	No	2	2
232	No	3	0
233	No	4	0
234	No	1	0

[235 rows x 21 columns]

#checking info again
student_behavior.info

<bound method DataFrame.info of				certification_course	gender
department	height(cm)	weight(kg)	\		
0	No	Male	BCA	100.0	58.0
1	No	Female	BCA	90.0	40.0
2	Yes	Male	BCA	159.0	78.0
3	Yes	Female	BCA	147.0	20.0
4	No	Male	BCA	170.0	54.0

...
230	Yes	Male	B.com	ISM	170.0	76.0
231	Yes	Male	B.com	ISM	172.0	52.0
232	Yes	Female		BCA	139.0	33.0
233	Yes	Female		Commerce	153.0	58.0
234	No	Female	B.com	ISM	155.0	39.0

	10th_mark	12th_mark	college_mark	hobbies	daily_studying
...	\				
0	79.0	64.0	80.0	Video Games	0 - 30 minute
...					
1	70.0	80.0	70.0	Cinema	30 - 60 minute
...					
2	71.0	61.0	55.0	Cinema	1 - 2 Hour
...					
3	70.0	59.0	58.0	Reading books	1 - 2 Hour
...					
4	40.0	65.0	30.0	Video Games	30 - 60 minute
...					
...
...					
230	72.0	67.0	65.0	Video Games	30 - 60 minute
...					
231	72.0	70.0	76.0	Cinema	2 - 3 hour
...					
232	90.0	75.0	70.0	Cinema	30 - 60 minute
...					
233	85.0	74.0	75.0	Cinema	0 - 30 minute
...					
234	45.0	45.0	50.0	Sports	3 - 4 hour
...					

	salary_expectation	degree_satisfaction	pursue_career_in_field	\
0	40000		No	50%
1	15000		Yes	75%
2	13000		Yes	50%
3	1500000		No	50%
4	50000		Yes	25%
...
230	7000		Yes	50%
231	25000		Yes	75%
232	20000		Yes	50%
233	20000		Yes	75%
234	10		No	75%

	social_media	commute	stress_level_	financial_status
\				
0	1.30 - 2 hour	30 - 60 minutes	Bad	Bad
1	1 - 1.30 hour	0 - 30 minutes	Bad	Bad

2	More than 2 hour	30 - 60 minutes	Awful	Bad
3	1.30 - 2 hour	0 - 30 minutes	Bad	good
4	1.30 - 2 hour	30 - 60 minutes	Good	good
..
230	More than 2 hour	30 - 60 minutes	Bad	Bad
231	30 - 60 Minute	30 - 60 minutes	Good	good
232	30 - 60 Minute	2.30 - 3 hour	Bad	good
233	More than 2 hour	2.30 - 3 hour	Awful	good
234	30 - 60 Minute	2.30 - 3 hour	fabulous	good

	part-time_job	stress_level_number	salary_expectations
0	No	3	2
1	No	3	0
2	No	4	0
3	No	3	15
4	No	2	2
..
230	No	3	0
231	No	2	2
232	No	3	0
233	No	4	0
234	No	1	0

[235 rows x 21 columns]>

Stress levels, college marks, daily studying

Here I'm viewing daily studying, college marks, and stress levels. It seems like the people with the lowest stress tend to study for longer. Makes sense, they probably feel better going into exams.

```
#stress levels, college marks are the groupby compared to daily studying
```

```
sum_students = student_behavior.groupby(['stress_level',
'college_mark'], as_index = False)['daily_studying'].sum()
sum_students
```

	stress_level	college_mark	daily_studying
0	Awful	30.0	0 - 30 minute
1	Awful	55.0	1 - 2 Hour

2	Awful	60.0	30 - 60 minute	30 - 60 minute
3	Awful	65.0	30 - 60 minute	1 - 2 Hour
4	Awful	70.0	30 - 60 minute	3 - 4 hour
...
67	fabulous	75.0		30 - 60 minute
68	fabulous	80.0	2 - 3 hour	0 - 30 minute
69	fabulous	85.0		More Than 4 hour
70	fabulous	92.0		30 - 60 minute
71	fabulous	97.0		0 - 30 minute

[72 rows x 3 columns]

Awful stress levels

I wanted to see all the awful stress levels and how much they stressed along with their studying.

```
#viewing awful stress levels
```

```
unhappy_students = sum_students[sum_students['stress_level_'] ==  
'Awful']  
unhappy_students
```

	stress_level_	college_mark	
daily_studying			
0	Awful	30.0	0 -
30 minute			
1	Awful	55.0	1
- 2 Hour			
2	Awful	60.0	30 - 60 minute
60 minute			
3	Awful	65.0	30 - 60 minute
30 minute			
4	Awful	70.0	30 - 60 minute
- 2 Hour			
5	Awful	75.0	0 - 30 minute
30 minute			
6	Awful	80.0	0 - 30 minute
60 minute			
7	Awful	90.0	0 - 30 minute
Than 4 hour			

Fabulous stress levels

I wanted to see fabulous stress levels compared to studying and college marks. It does seem like they put in more study time than those who say their stress levels are awful.

```
#viewing fabulous stress levels
```

```
happy_students = sum_students[sum_students['stress_level_'] ==
```

```
'fabulous']  
happy_students
```

	stress_level_	college_mark	daily_studying
64	fabulous	50.0	3 - 4 hour3 - 4 hour
65	fabulous	60.0	1 - 2 Hour2 - 3 hour
66	fabulous	65.0	0 - 30 minute
67	fabulous	75.0	30 - 60 minute
68	fabulous	80.0	2 - 3 hour0 - 30 minute
69	fabulous	85.0	More Than 4 hour
70	fabulous	92.0	30 - 60 minute
71	fabulous	97.0	0 - 30 minute

Mean stress level and college marks -Awful

Here is the mean for college marks for those with awful stress levels

```
#awful stress mean  
filtered_students = sum_students[sum_students['stress_level_'] ==  
'Awful']  
mean_awful = filtered_students['college_mark'].mean()  
mean_awful  
  
65.625
```

Mean stress level and college marks -Fabulous

Here is the mean grade for students with fabulous stress levels

```
#mean for fabulous stress  
filtered_students = sum_students[sum_students['stress_level_'] ==  
'fabulous']  
mean_happy = filtered_students['college_mark'].mean()  
mean_happy  
  
75.5
```

Checking unique stress levels again

All the stress levels are awful, bad, good, fabulous

```
#viewing valus in stress_level_  
sum_students['stress_level_'].unique()  
  
array(['Awful', 'Bad', 'Good', 'fabulous'], dtype=object)
```

Mean stress level and college marks-good

Here is the mean for those students with a good stress level

```
# mean for good stress
filtered_students = sum_students[sum_students['stress_level_'] ==
'Good']
mean_good = filtered_students['college_mark'].mean()
mean_good
```

65.66472222222222

filtered_students

	stress_level_	college_mark	\
28	Good	1.00	
29	Good	2.00	
30	Good	3.00	
31	Good	7.50	
32	Good	12.00	
33	Good	30.00	
34	Good	35.00	
35	Good	50.00	
36	Good	55.00	
37	Good	60.00	
38	Good	65.00	
39	Good	66.30	
40	Good	67.00	
41	Good	68.00	
42	Good	69.00	
43	Good	70.00	
44	Good	74.00	
45	Good	75.00	
46	Good	76.00	
47	Good	79.00	
48	Good	79.50	
49	Good	80.00	
50	Good	80.87	
51	Good	82.00	
52	Good	82.96	
53	Good	84.00	
54	Good	85.00	
55	Good	86.00	
56	Good	88.00	
57	Good	89.00	
58	Good	90.00	
59	Good	91.00	
60	Good	92.80	
61	Good	93.00	
62	Good	95.00	

63	Good	100.00
----	------	--------

	daily_studying
28	1 - 2 Hour
29	0 - 30 minute
30	1 - 2 Hour
31	1 - 2 Hour
32	More Than 4 hour
33	30 - 60 minute
34	3 - 4 hour
35	0 - 30 minute1 - 2 Hour0 - 30 minute30 - 60 mi...
36	0 - 30 minute0 - 30 minute
37	1 - 2 Hour30 - 60 minute1 - 2 Hour30 - 60 minu...
38	1 - 2 Hour0 - 30 minute0 - 30 minute30 - 60 mi...
39	2 - 3 hour
40	30 - 60 minute
41	1 - 2 Hour0 - 30 minute
42	0 - 30 minute
43	0 - 30 minute30 - 60 minute1 - 2 Hour3 - 4 hou...
44	0 - 30 minute
45	1 - 2 Hour1 - 2 Hour30 - 60 minute30 - 60 minu...
46	2 - 3 hour
47	0 - 30 minute
48	0 - 30 minute
49	1 - 2 Hour30 - 60 minute2 - 3 hour30 - 60 minu...
50	1 - 2 Hour
51	1 - 2 Hour
52	0 - 30 minute
53	30 - 60 minute30 - 60 minute
54	1 - 2 Hour1 - 2 Hour1 - 2 Hour1 - 2 Hour3 - 4 ...
55	1 - 2 Hour30 - 60 minute
56	30 - 60 minute
57	0 - 30 minute30 - 60 minute
58	1 - 2 Hour30 - 60 minute2 - 3 hour2 - 3 hour30...
59	1 - 2 Hour
60	3 - 4 hour
61	1 - 2 Hour
62	2 - 3 hour
63	1 - 2 Hour3 - 4 hour

student_behavior

	certification_course	gender	department	height(cm)	weight(kg)	\
0	No	Male	BCA	100.0	58.0	
1	No	Female	BCA	90.0	40.0	
2	Yes	Male	BCA	159.0	78.0	
3	Yes	Female	BCA	147.0	20.0	
4	No	Male	BCA	170.0	54.0	
..	
230	Yes	Male	B.com ISM	170.0	76.0	

231	Yes	Male	B.com	ISM	172.0	52.0
232	Yes	Female		BCA	139.0	33.0
233	Yes	Female	Commerce		153.0	58.0
234	No	Female	B.com	ISM	155.0	39.0

	10th_mark	12th_mark	college_mark	hobbies	daily_studying
...	\				
0	79.0	64.0	80.0	Video Games	0 - 30 minute
...					
1	70.0	80.0	70.0	Cinema	30 - 60 minute
...					
2	71.0	61.0	55.0	Cinema	1 - 2 Hour
...					
3	70.0	59.0	58.0	Reading books	1 - 2 Hour
...					
4	40.0	65.0	30.0	Video Games	30 - 60 minute
...					
..
...					
230	72.0	67.0	65.0	Video Games	30 - 60 minute
...					
231	72.0	70.0	76.0	Cinema	2 - 3 hour
...					
232	90.0	75.0	70.0	Cinema	30 - 60 minute
...					
233	85.0	74.0	75.0	Cinema	0 - 30 minute
...					
234	45.0	45.0	50.0	Sports	3 - 4 hour
...					

	salary_expectation	degree_satisfaction	pursue_career_in_field	\
0	40000		No	50%
1	15000		Yes	75%
2	13000		Yes	50%
3	1500000		No	50%
4	50000		Yes	25%
..
230	7000		Yes	50%
231	25000		Yes	75%
232	20000		Yes	50%
233	20000		Yes	75%
234	10		No	75%

	social_media	commute	stress_level_	financial_status
\				
0	1.30 - 2 hour	30 - 60 minutes	Bad	Bad
1	1 - 1.30 hour	0 - 30 minutes	Bad	Bad
2	More than 2 hour	30 - 60 minutes	Awful	Bad

3	1.30 - 2 hour	0 - 30 minutes	Bad	good
4	1.30 - 2 hour	30 - 60 minutes	Good	good
..
230	More than 2 hour	30 - 60 minutes	Bad	Bad
231	30 - 60 Minute	30 - 60 minutes	Good	good
232	30 - 60 Minute	2.30 - 3 hour	Bad	good
233	More than 2 hour	2.30 - 3 hour	Awful	good
234	30 - 60 Minute	2.30 - 3 hour	fabulous	good

	part-time_job	stress_level_number	salary_expectations
0	No	3	2
1	No	3	0
2	No	4	0
3	No	3	15
4	No	2	2
..
230	No	3	0
231	No	2	2
232	No	3	0
233	No	4	0
234	No	1	0

[235 rows x 21 columns]

Import and use seaborn

I'm importing seaborn as sns which is a common nickname for it. I decided to use a lineplot to view the data. It certainly does look like those with fabulous stress levels study more.

```
#more visualization with a line plot
```

```
import seaborn as sns
```

```
sns.relplot(x='stress_level_', y='daily_studying', kind='line',
data=student_behavior)
```

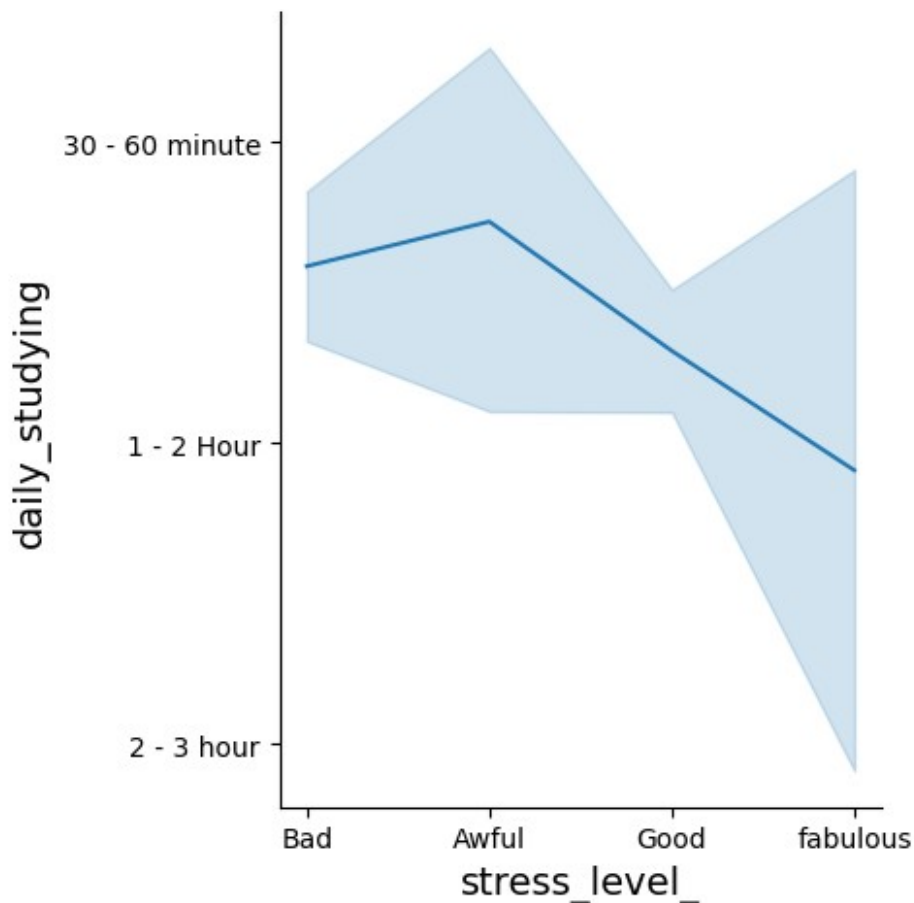
```
C:\Users\itali\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
```

```
with pd.option_context('mode.use_inf_as_na', True):
C:\Users\itali\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
```

```
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
```

```
with pd.option_context('mode.use_inf_as_na', True):
```

```
<seaborn.axisgrid.FacetGrid at 0x2db3b4e6790>
```



```
#line plot
```

```
sns.relplot(x='stress_level_', y='financial_status', kind='line',
data=student_behavior)
```

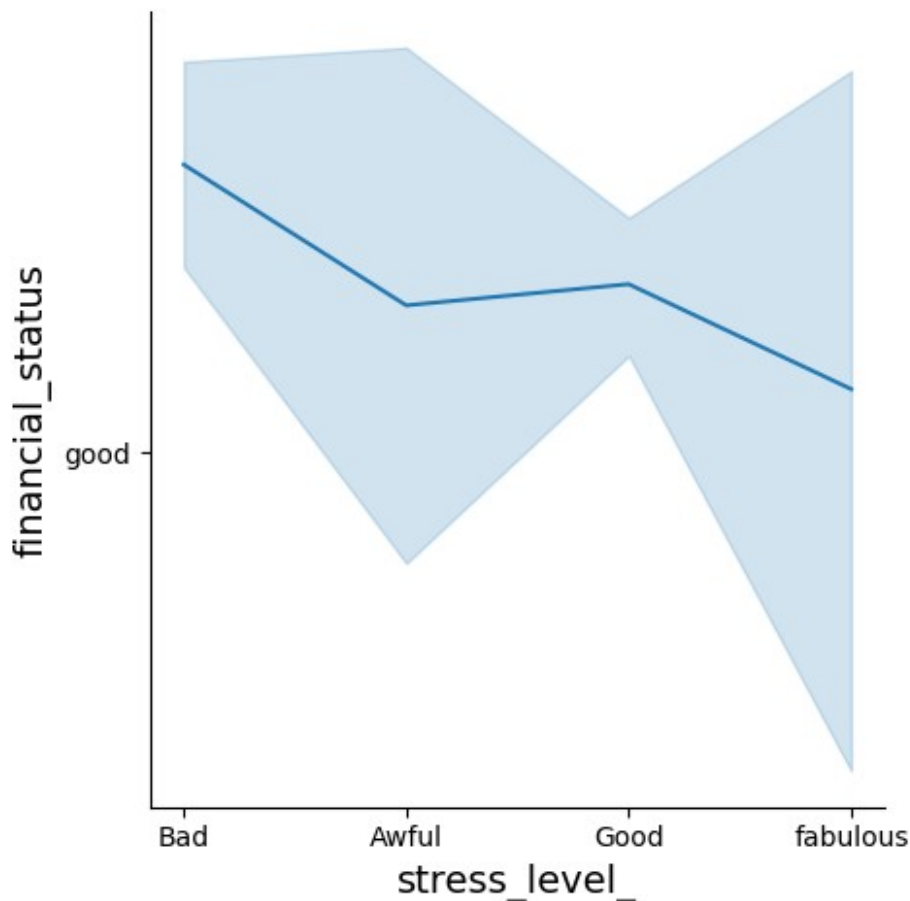
```
C:\Users\itali\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
```

```
with pd.option_context('mode.use_inf_as_na', True):
```

```
C:\Users\itali\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
```

```
with pd.option_context('mode.use_inf_as_na', True):
```

```
<seaborn.axisgrid.FacetGrid at 0x2db3b624150>
```



Counting the stress levels

I wanted to see the students and how long they studied overall. Most students studied between 30-60 minutes. Few studies more than 4 hours (8).

```
#groupby daily_studying and stress level
#TODO - this just counts stress level column entries, reagrdless of
actual value, so below data isn't meaningful in regards to stress
level
groupby_mention =
pd.DataFrame(student_behavior.groupby(['stress_level_',
'daily_studying'])['stress_level_'].count())
groupby_mention
```

		stress_level_
stress_level_	daily_studying	
Awful	0 - 30 minute	7
	1 - 2 Hour	5
	3 - 4 hour	1

Bad	30 - 60 minute	5
	More Than 4 hour	1
	0 - 30 minute	10
	1 - 2 Hour	17
	2 - 3 hour	4
Good	3 - 4 hour	3
	30 - 60 minute	33
	More Than 4 hour	1
	0 - 30 minute	26
	1 - 2 Hour	38
fabulous	2 - 3 hour	18
	3 - 4 hour	9
	30 - 60 minute	41
	More Than 4 hour	5
	0 - 30 minute	3
	1 - 2 Hour	1
	2 - 3 hour	2
	3 - 4 hour	2
	30 - 60 minute	2
	More Than 4 hour	1
student_behavior		

Conclusions

It's interesting how study time could have an impact on stress levels. It makes sense, if you study longer, you might feel better going in to take an exam or write a paper.

There's more for future analysis options. Consider asking students other poll questions. I'd love to dig deeper about commute time and grades.

Thank you for taking a look.