Untitled11

January 19, 2020

1 DS SURVEY ANALYSIS

DS.jpg

This is a document written based on our study of the data science survey data which was conducted a few months back, to know how data science is thought about throughout the college. Our analysis comprises of plotting ideas and derived facts which provide evidence to various possibilities.

```
[1]: import pandas as pd
     df = pd.read_csv('ASE-B Survey.csv')
```

Importing pandas and reading the file.

1.1 Basic Information

```
[2]: df.head(2)
```

```
[2]:              Timestamp  Which year are you in currently?  \
     0  11/5/2019 21:51:27                                 2
     1  11/5/2019 21:53:49                                 2

        On a scale of 1-10, how excited do you get when you hear the term "Data
     Science" ?  \
     0                                                    10
     1                                                     9

       Which of the following options best describes you during these holidays?  \
     0                 Working on Data Science Projects
     1                          Learning Data Science

       If you are learning Data Science currently, kindly mention the sources from
     which you are learning(Separate each source with a comma)  \
     0                                  Udemy , blog posts.
     1  I watch YouTube videos because I like visual s...

       What are your sources to learn Python? (You can select more than one)  \
     0  Codecademy, Solo Learn, Whatever I can find on...
```

```
1                             Whatever I can find on Google

   Are you aware of the ACM Data Science SIG ?  \
0                                             Yes
1                                             Yes

   Are you aware of the "The Winter CodeX 2018 - Python Edition" beginner's
   guide?  \
0                                             Yes
1                                             No

   What kind of learning do you prefer?  \
0                        Online Courses
1                        Online Courses

    On a scale of 1-10, how serious are you about a career in Data Science?  \
0                                                      10
1                                                       8

    On a scale of 1-10, how much do you fear math and statistics?  \
0                                                       2
1                                                       3

   Have you worked on Data Science projects before OR are you currently working
   on any?  \
0                                             No
1                                             No

    Would you like to share a link of your data science project? It can be a
   website or your github repo or a blog article. If yes, kindly paste the link
   beneath.  \
0                                            NaN
1                                            NaN

   If you would like to receive the ACM ASE-B Data Science SIG related news,
   kindly provide us with your email id (We hate spam as much as you do, so don't
   worry about that)
0                         ritwiklal2000@gmail.com
1                         meghanarao.99@gmail.com
```

I'm carefully selecting a few attributes which do not have unique values for analysing some numbers.

```python
count = df.columns[1:4]
count = count.append(df.columns[6:12])
for i in count:
    print(str(df[i].value_counts())+'\n')
```

```
1     27
```

```
2    26
3    11
4     5
Name: Which year are you in currently?, dtype: int64

10    20
9     17
8     16
7      9
6      3
4      2
5      1
3      1
Name: On a scale of 1-10, how excited do you get when you hear the term "Data
Science" ?, dtype: int64

Learning Python
26
Learning Data Science
20
I am stuck with other work and have been unable to do any Data Science related
work    18
Working on Data Science Projects
2
I don't want to do Data Science
2
I decline to answer
1
Name: Which of the following options best describes you during these holidays?,
dtype: int64

Yes                                                            41
No                                                             25
Heard of it, but I thought it was like a "Unicorns Exist" scenario    3
Name: Are you aware of the ACM Data Science SIG ?, dtype: int64

Yes    36
No     33
Name: Are you aware of the "The Winter CodeX 2018 - Python Edition" beginner's
guide?, dtype: int64

Project-based         38
Online Courses        23
Book-based             7
Prayer to the Gods     1
Name: What kind of learning do you prefer?, dtype: int64

8     19
```

```
10      18
9       11
7        8
6        6
5        4
3        2
2        1
Name: On a scale of 1-10, how serious are you about a career in Data Science?,
dtype: int64

3       13
1       12
7       11
2       11
5        7
4        7
8        4
10       3
6        1
Name: On a scale of 1-10, how much do you fear math and statistics?, dtype:
int64

No      60
Yes      9
Name: Have you worked on Data Science projects before OR are you currently
working on any?, dtype: int64
```

Well, these numbers give me a good idea of where to head for plots.

1.2 Cleaning our data

Firstly, the column names are too long. So we are going to save them (into a variable 'Default')
and shorten them into suitable keywords.

```
[4]: new_columns = ['Timestamp','Year','Excitement scale','Holiday work','Learning␣
     ↪sources(DS)','Learning sources(Py)','SIG Aware','WinterCodex␣
     ↪Aware','Learning preference','Seriousness scale','Mathstatfear','Project␣
     ↪exp','Project link','E-mail ID']
```

```
[5]: Default = df.columns
     df.columns = new_columns
```

After cleaning, the datatype and name of the attributes are as follows.

```
[6]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 69 entries, 0 to 68
Data columns (total 14 columns):
```

```
Timestamp            69 non-null object
Year                 69 non-null int64
Excitement scale     69 non-null int64
Holiday work         69 non-null object
Learning sources(DS) 27 non-null object
Learning sources(Py) 59 non-null object
SIG Aware            69 non-null object
WinterCodex Aware    69 non-null object
Learning preference  69 non-null object
Seriousness scale    69 non-null int64
Mathstatfear         69 non-null int64
Project exp          69 non-null object
Project link          0 non-null float64
E-mail ID            50 non-null object
dtypes: float64(1), int64(4), object(9)
memory usage: 7.6+ KB
```

[7]:
```python
df['Timestamp'] = pd.to_datetime(df['Timestamp'])
```

Changing the timestamp into datetime datatype to unlock a lot of functionality.

[8]:
```python
df['time'] = df['Timestamp'].dt.hour + df['Timestamp'].dt.minute/100
df['day'] = df['Timestamp'].dt.day
```

Here, we take the values into a slightly different form for ease of plotting later on.

[9]:
```python
def rep_year(x):
    if (x is 4):
        return 'IV'
    else:
        return 'I'*x

df['Year'] = df['Year'].apply(rep_year)
```

[10]:
```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 69 entries, 0 to 68
Data columns (total 16 columns):
Timestamp            69 non-null datetime64[ns]
Year                 69 non-null object
Excitement scale     69 non-null int64
Holiday work         69 non-null object
Learning sources(DS) 27 non-null object
Learning sources(Py) 59 non-null object
SIG Aware            69 non-null object
WinterCodex Aware    69 non-null object
Learning preference  69 non-null object
Seriousness scale    69 non-null int64
Mathstatfear         69 non-null int64
```

```
Project exp             69 non-null object
Project link            0 non-null float64
E-mail ID               50 non-null object
time                    69 non-null float64
day                     69 non-null int64
dtypes: datetime64[ns](1), float64(2), int64(4), object(9)
memory usage: 8.7+ KB
```

It's better we keep the values of year attribute in the roman format because they signify a particular group of people and exclude them from numerical analysis.

[11]: `df.iloc[55]`

```
[11]: Timestamp                2019-11-07 12:35:01
      Year                                       I
      Excitement scale                          10
      Holiday work                 Learning Python
      Learning sources(DS)                     NaN
      Learning sources(Py)                     NaN
      SIG Aware                                 No
      WinterCodex Aware                         No
      Learning preference       Prayer to the Gods
      Seriousness scale                         10
      Mathstatfear                               5
      Project exp                               No
      Project link                             NaN
      E-mail ID                                NaN
      time                                   12.35
      day                                        7
      Name: 55, dtype: object
```

An enthusiastic 1st year who is towards learning python but is headed in the wrong path as he is a belief-based learner in grasping concepts. Considering his seriuosness of 10 and excitement too of 10, he seems to be one of the ideal candidate who is supposed to know the existence of the SIG. This would definitely help him out with his mathstat fear of 5 and guide him in the right way, he is a 1st year student which is great as he has enough time to work properly.

Such a case has possibly come out because he is in a hurry to know everything about the subject based on his excitement and seriuosness, and obviuosly it's never going to work. If he becomes a part of the DS community there are high chances of him being a success story who can inspire many others to reach heights with just a change in mindset.

He is definitely an outlier to the study preference feature but is still the most special throughout the dataset.

1.3 Let's get plotting

```python
import matplotlib.pyplot as plt
import seaborn as sns
import matplotlib.style as mplstyle
mplstyle.use(['bmh','fast','seaborn-dark-palette'])
```
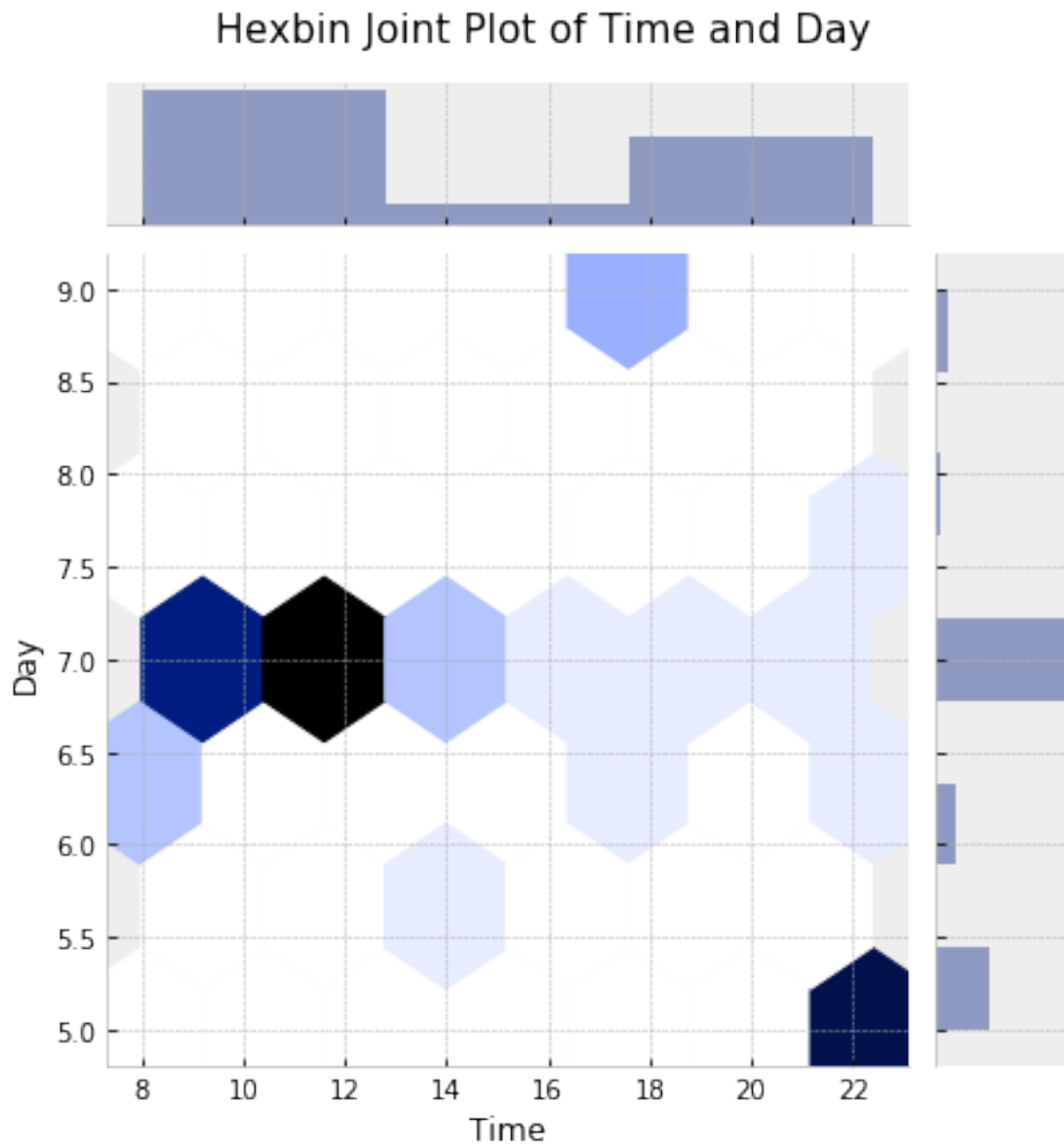
6

Importing matplotlib and seaborn. Also, customizing plot style.

1.4 Timestamp

Q) When do we release the survey to increase the number of responses?

```
[13]: hexbin = sns.jointplot(x = 'time', y = 'day', data = df, kind="hex")
      hexbin.set_axis_labels(xlabel='Time', ylabel='Day')
      hexbin.fig.suptitle('Hexbin Joint Plot of Time and Day',fontsize=15, y=1.04)
```

[13]: Text(0.5, 1.04, 'Hexbin Joint Plot of Time and Day')

The survey was released on 5th of November 2019 to the DS SIG whatsapp groups. 93% entries have been filled up in the span of 5th to 7th.

```
[14]: kde = sns.jointplot(x = 'time', y = 'day', data = df, kind="kde")
      kde.set_axis_labels(xlabel='Time', ylabel='Day')
      kde.fig.suptitle('KDE Joint Plot of Time and Day',fontsize=15, y=1.04)
```

[14]: Text(0.5, 1.04, 'KDE Joint Plot of Time and Day')
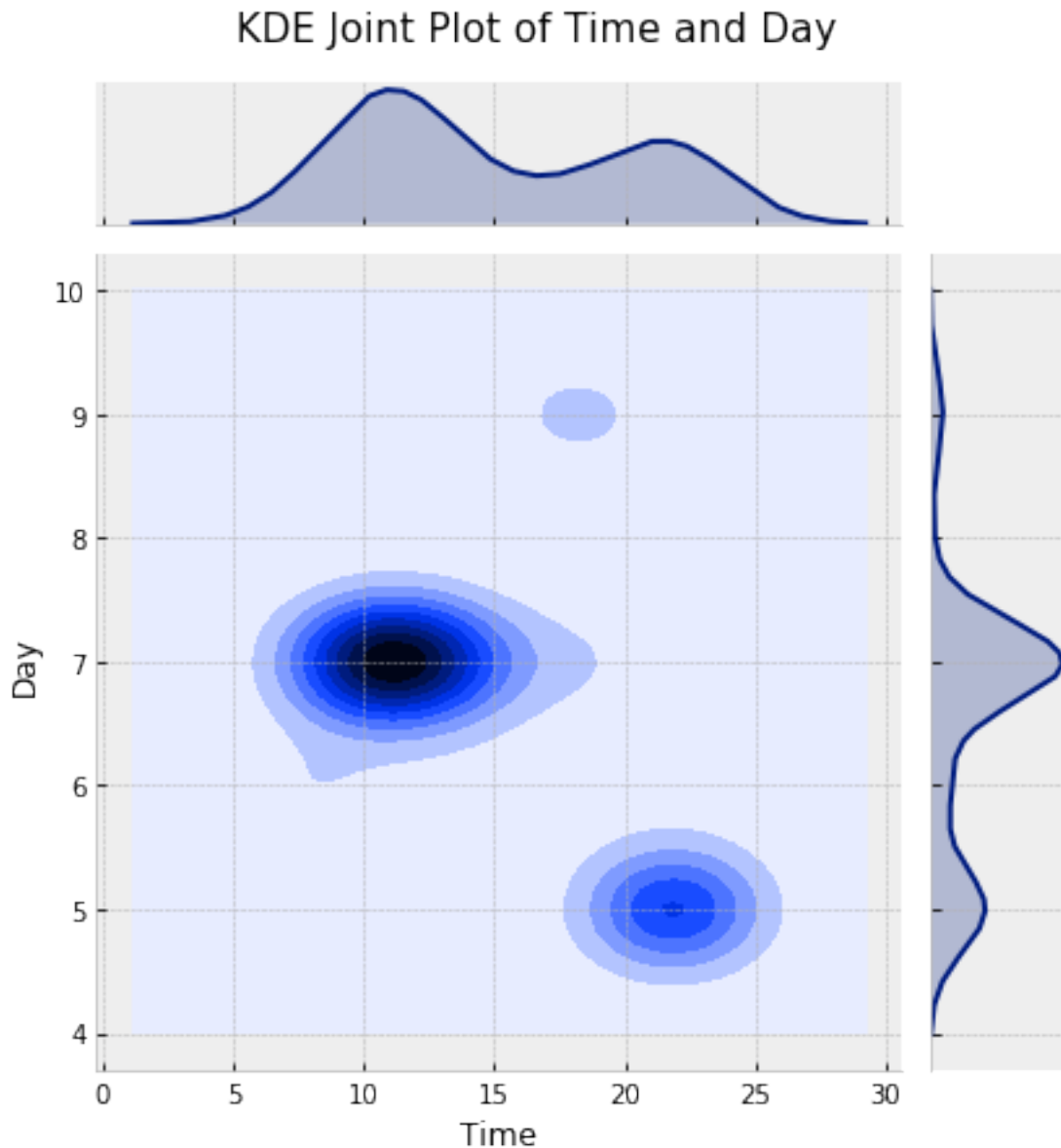
About 60% entries were accepted on 7th of November. This was because the SURVEY reached many class groups on this date(we were allowed to share the survey link on this date) and as there were holidays going on students didn't mind answering the same.

Now since sharing the survey was an option to the members, it is best to get in contact with the

CR's of every class to put it on their class groups. This will make it a 100% reach to every class group.

1.5 How excited are students across years

```
[15]: fig, ax = plt.subplots(figsize = (6,8))
      ax = sns.countplot(y = 'Excitement scale', hue = 'Year', data = df)
      ax.legend(loc = 1)
      ax.set_title('Excitement of students across years')
      ax.set_xlabel('Number of students')
      ax.set_ylabel('Scale(1-10)')
```

[15]: Text(0, 0.5, 'Scale(1-10)')

Excitement of students across years
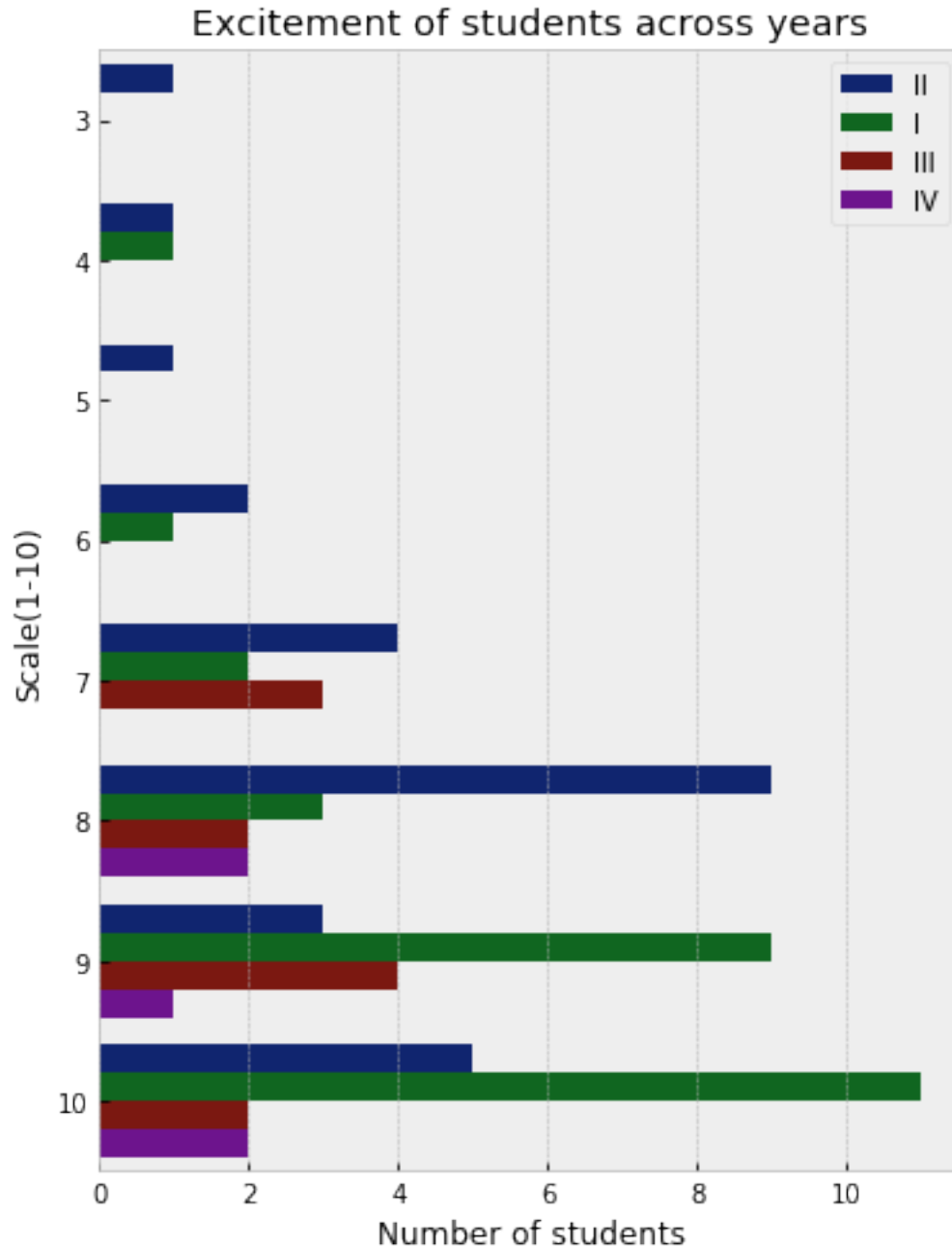
We see two tall blue lines representing the first years with an excitement at scale 9 and 10 implying that first years are trying to explore fields one of which is 'Data Science'.

We see a tall green line at excitement scale 8 representing second years who are not too sure on 'Data Science' yet.

All the fourth years have an excitement scale above 7 giving a sense of the statement made earlier that all the fourth years have decided their field of expertise which they will be considering. All the other fourth years didn't bother to take the survey. Same goes with third years in a bit smaller scale.

1.6 Holidays

```
[16]: count, ax = plt.subplots()
      ax = sns.countplot(y = 'Holiday work', data = df)
      ax.set_title('This is what suits students best in holidays.')
      ax.set_xlabel('Number of students')
      ax.set_ylabel('Work done by the students')
```

[16]: Text(0, 0.5, 'Work done by the students')
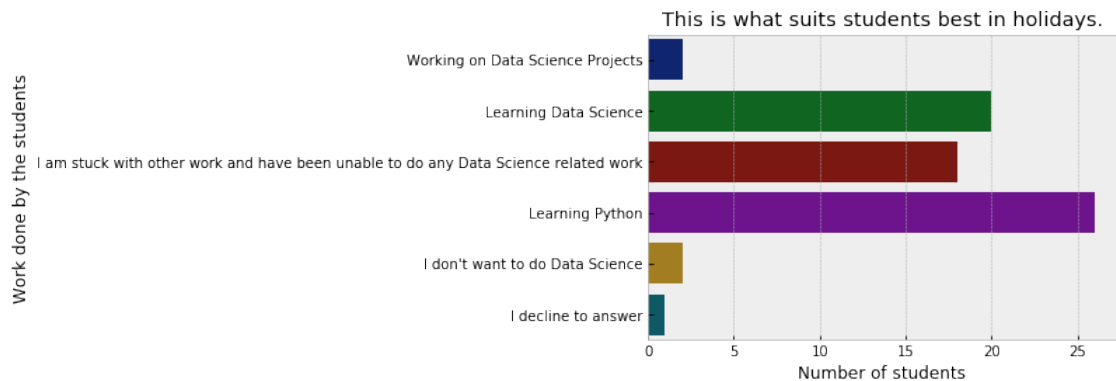
Majority of our students are learning python or data science. - (46 students) A few are not interested while another few are working on projects. - (4 students) A major part of our students is also stuck with other work and unable to work on data science. - (18 students)

1.7 Maths and Statistics fear

```
[17]: fig, ax = plt.subplots(figsize = (9,6))
      ax = sns.countplot(x = 'Mathstatfear', hue = 'Year', data = df)
      ax.set_title('Fear of maths and stats for year')
      ax.set_xlabel('Scale(1-10)')
      ax.set_ylabel('Number of students')
      plt.legend(loc = 1)
```

[17]: <matplotlib.legend.Legend at 0xb8824e0>

Fear of maths and stats for year
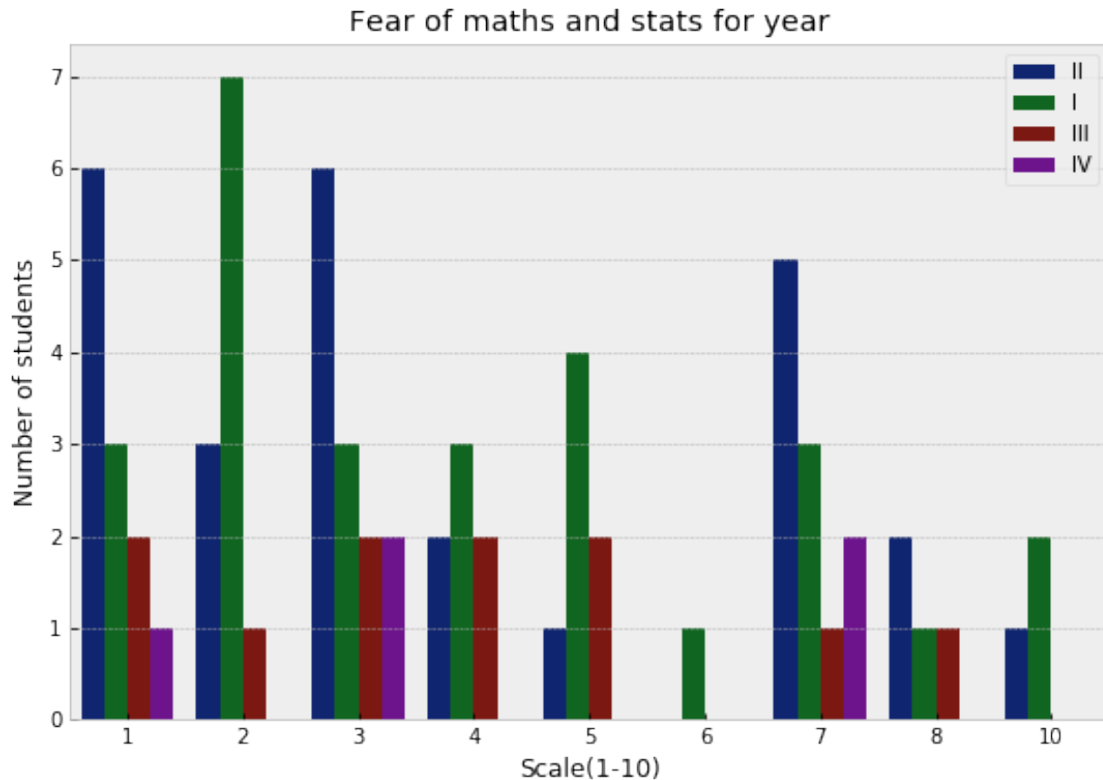
1) 74% of the 1st years fear maths and statistics equal to or less than 5. This can be based on their performances up till 12th standard and hence, they are confident enough that they can handle the further complexities in the subject. This could be the reason for high participation of the 1st years in projects.

2) If the awareness of the DS SIG can be increased from 2/3rd to close to full, we can expect more participation as Data Science involves a lot of maths and statistics.

3) 69% of the 2nd years fear maths and statistics less than or equal to 5. There is an increase in the number of students who fear the subject more than 5, this can be based on the college courses they have studied related to maths and stats and might be hesitant to join the SIG (as a result). The 73% students of 2nd years can play an important here.

4) 81% of third years and 60% of the fourth years fear math and stats <=5. All these students are either project-based learners or learners through online courses. The below plot also proves the kind of learners the SIG existence has reached to.

```
[18]: c, ax = plt.subplots()
      ax = sns.countplot(y = 'Learning preference', data = df[df['SIG Aware']=='Yes'])
      ax.set_title('How do students like to learn?')
      ax.set_ylabel('Option')
      ax.set_xlabel('Number of students')
```

[18]: Text(0.5, 0, 'Number of students')

How do students like to learn?
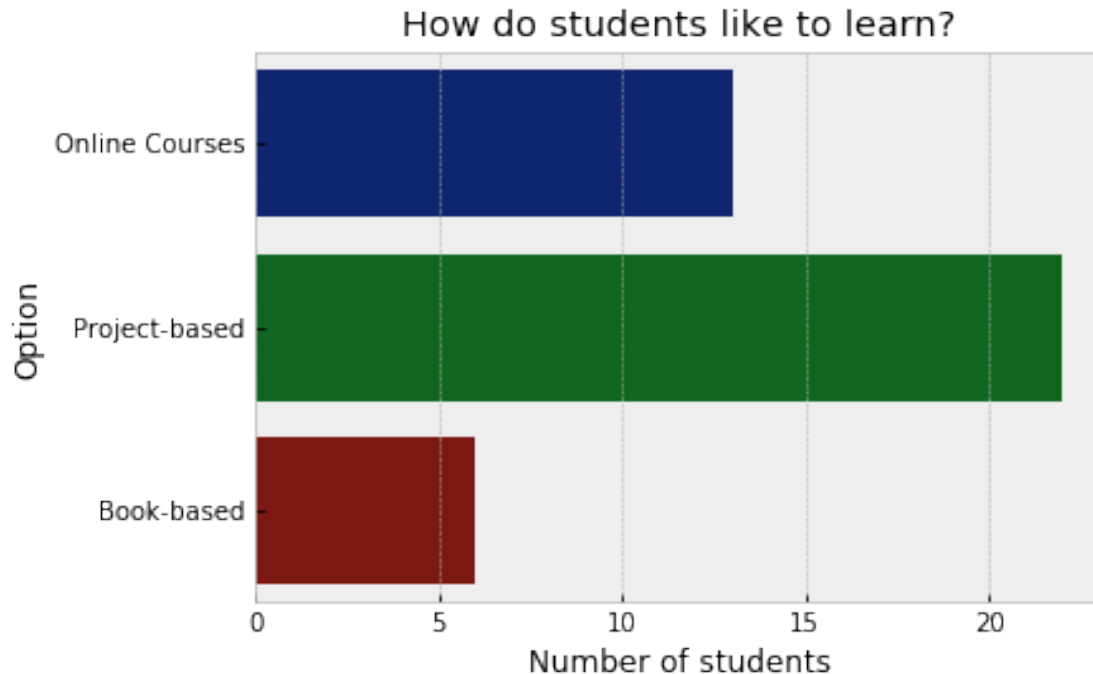
Q) Maths and stats can be handled but what about the skill of analysis?

R) Analysis comes from practice whether you do it through studies or games. It is an experience you develop over the years, it is just that data science is a different way to analyse leading to productive outcomes. So a start towards analyse for a change is never a harm, this will definitely develop interest for DS.

```
[19]: m, ax = plt.subplots(figsize = (6,9))
      ax = sns.violinplot(y = 'Mathstatfear', data = df)
      ax.set_title('Fear on Maths and Stats')
      ax.set_xlabel('Density of students')
      ax.set_ylabel('Scale(1-10)')
```

```
[19]: Text(0, 0.5, 'Scale(1-10)')
```

Fear on Maths and Stats
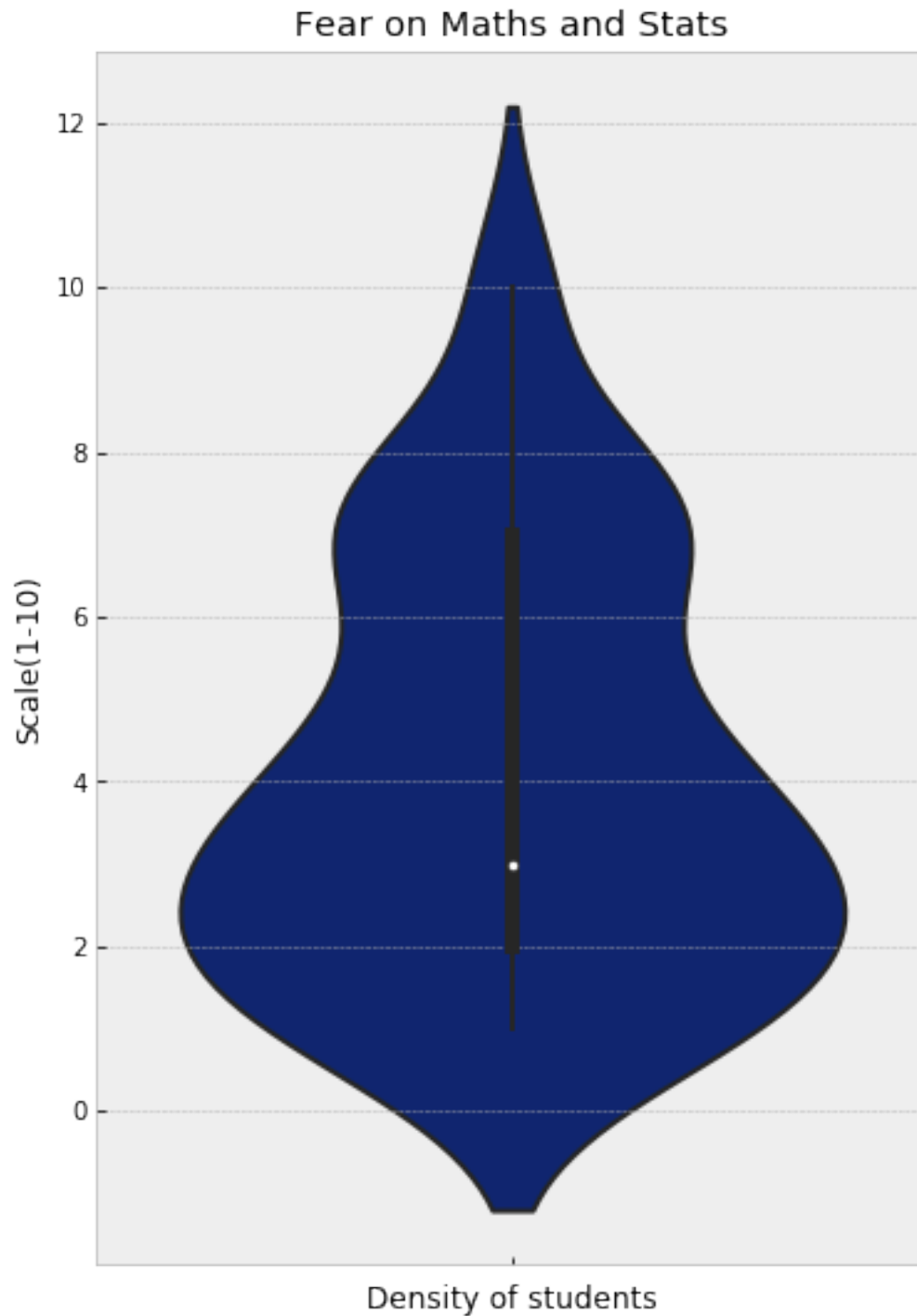
Don't feel the graph is wrong because you see a value 12 which is out of index. The small black line inside ending at 10 implies that 10 is the maximum value taken by any student.
We have a nice amount of density between 6 and 8 implying that there are students who are scared of maths and statistics.
We get a diverse plot implying that most students are moderately scared of maths and statistics.

1.8 Links

```
[20]: del df['Project link']
```

    Q) Why did we delete the attribute 'project links'?

    Project based preferring students are 56.5% of the crowd but the ones with actual project experience are just 13% (9/69) in total.

```
[21]: fig = sns.PairGrid(df[df.columns[0:13]] , hue = 'Year')
      fig = fig.map_upper(sns.regplot, fit_reg = False)
      fig = fig.map_lower(sns.distplot, hist = False)
      fig = fig.map_diag(sns.distplot, hist = False)
```
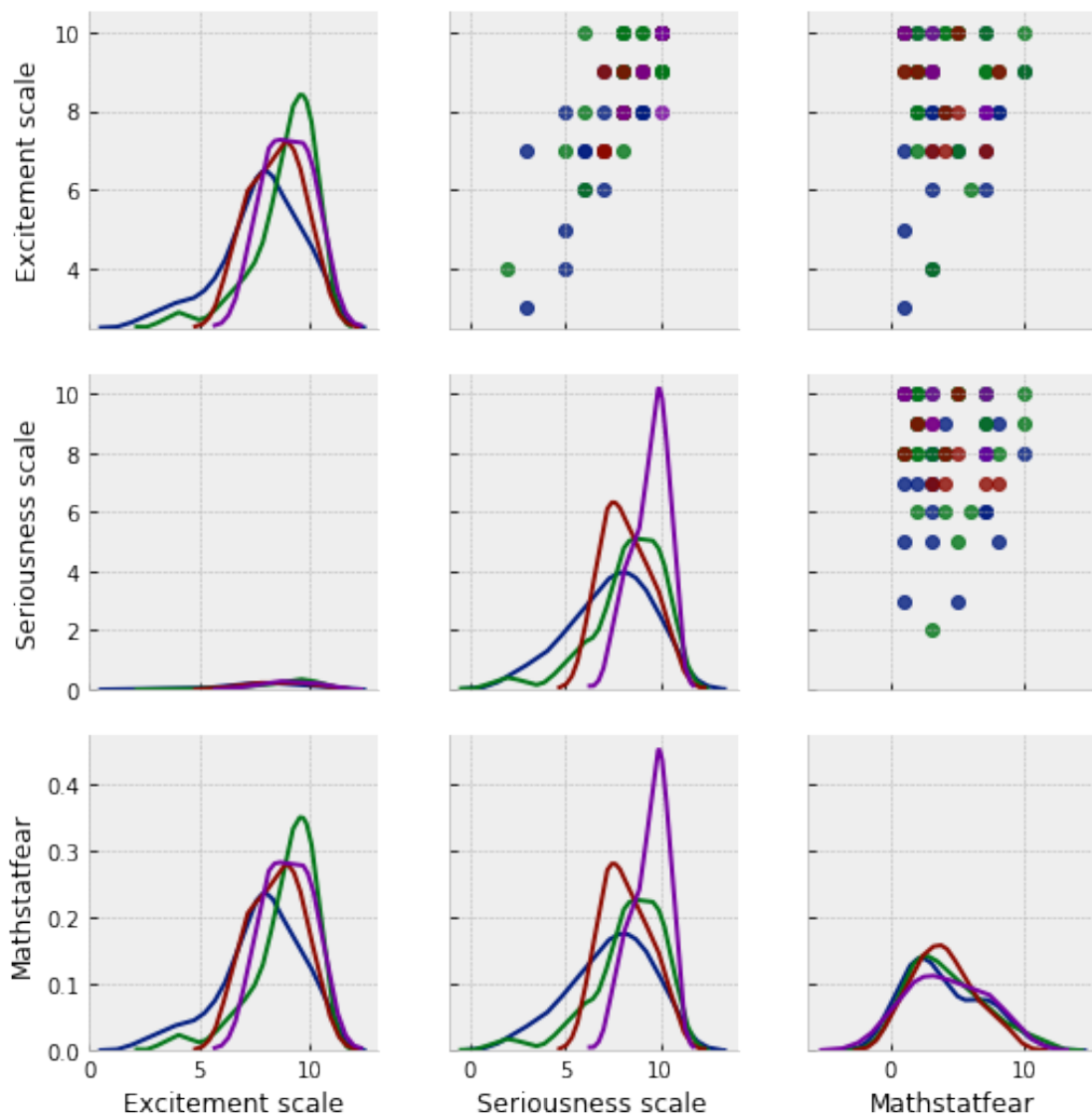
1) Most of the crowd is very excited except for a few 1st and 2nd years.

2) 3rd years are the most math fearing people.

3) All the 4th years are pretty serious which makes sense.

4) First years are not yet decided with their career, hence they are the least serious group.

```
[22]: fig, ax = plt.subplots(figsize = (6,5))
      ax = sns.countplot(y = 'Seriousness scale', hue = 'Year', data = df[df['Project␣
       ↪exp']=='Yes'])
      ax.set_title('Seriousness scale of students')
      ax.set_ylabel('Scale(1-10)')
      ax.set_xlabel('Number of students')
      plt.legend(loc = 1)
```

[22]: <matplotlib.legend.Legend at 0xbc2b0b8>
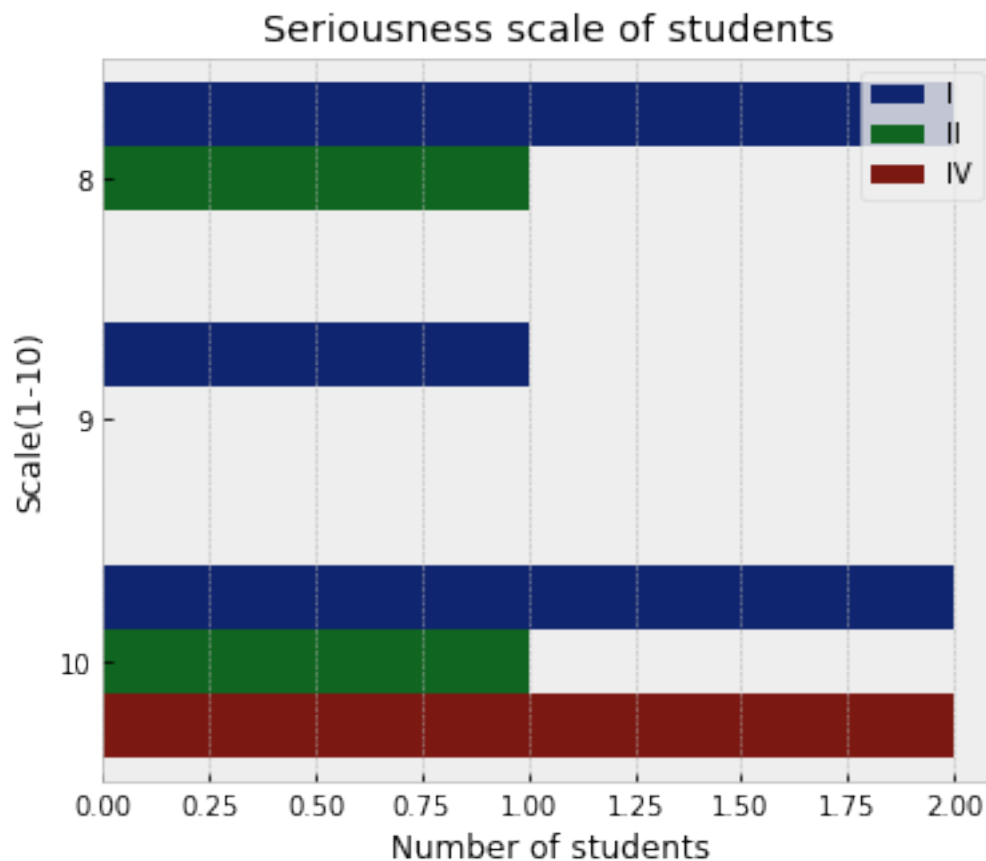
Seriousness factor of the project experienced students:
1st years avg. 9
2nd years avg. 9

No 3rd years
4th years avg. 10

```
[23]: k = df[df['Learning preference']=='Project-based']
```

We are grouping the students to be analysed into the variable 'k'.

```
[24]: sns.lmplot(x = 'Mathstatfear', y = 'Seriousness scale', data = df, hue =␣
      ↪'Year', fit_reg = False, markers = ['*','s','>','D'])
      plt.title('How serious and scared of maths and stats are students across years')
```

```
[24]: Text(0.5, 1, 'How serious and scared of maths and stats are students across
      years')
```

How serious and scared of maths and stats are students across years
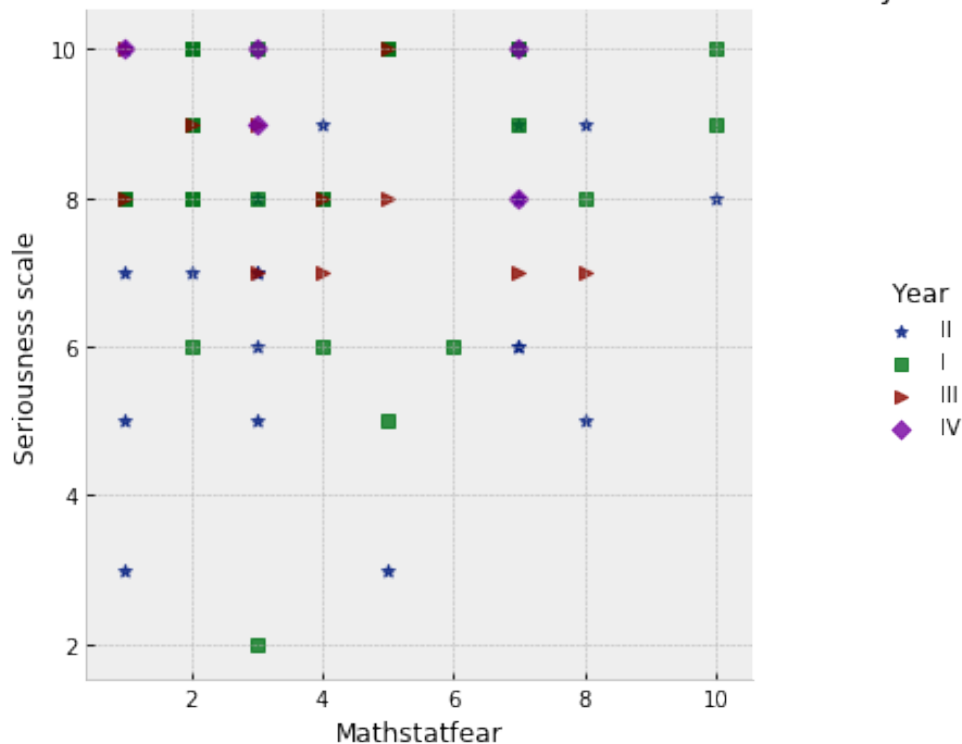
Project based preferring students are 56.5% of the crowd but the ones with actual project experience are just 13% (9/69) in total.

```
[25]: kde_joint = sns.jointplot(x ='Excitement scale', y ='Seriousness scale',data =␣
      ↪k, kind = 'kde')
      kde_joint.fig.suptitle('Kernel Density Plot', fontsize = 15, y = 1.04)
      kde_joint.set_axis_labels(xlabel = 'Excitement', ylabel = 'Seriousness')
```

```
[25]: <seaborn.axisgrid.JointGrid at 0xc3c8f60>
```

Kernel Density Plot
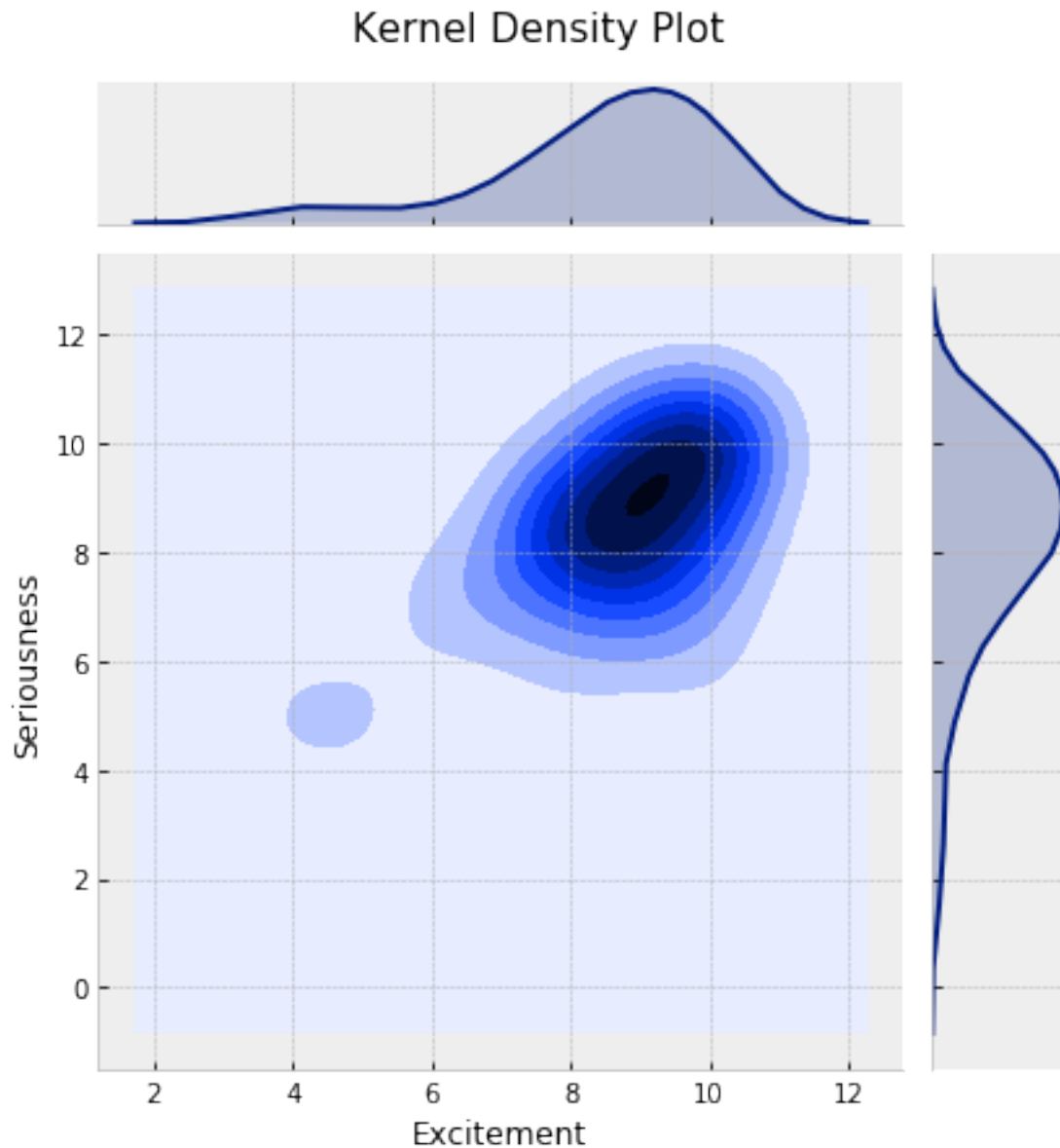
The 1st years can be considered to be learners who took part in projects to gain knowledge and experience. They have just started exploring, so they don't want to share their project links but have done projects of some level. Their definition of a project would be different from the other year students.

The first years comprise of 55.5% of the once who have done projects and all of them know the DS SIG. The SIG can play an important role by focussing on making them do projects collectively to achieve confidence.

```
[26]: kde_joint = sns.jointplot(x ='Excitement scale', y ='Seriousness scale',data =
      ↪k[k['Year']=='II'], kind = 'kde')
      kde_joint.fig.suptitle('Kernel Density Plot', fontsize = 15, y = 1.04)
      kde_joint.set_axis_labels(xlabel = 'Excitement', ylabel = 'Seriousness')
```

Kernel Density Plot
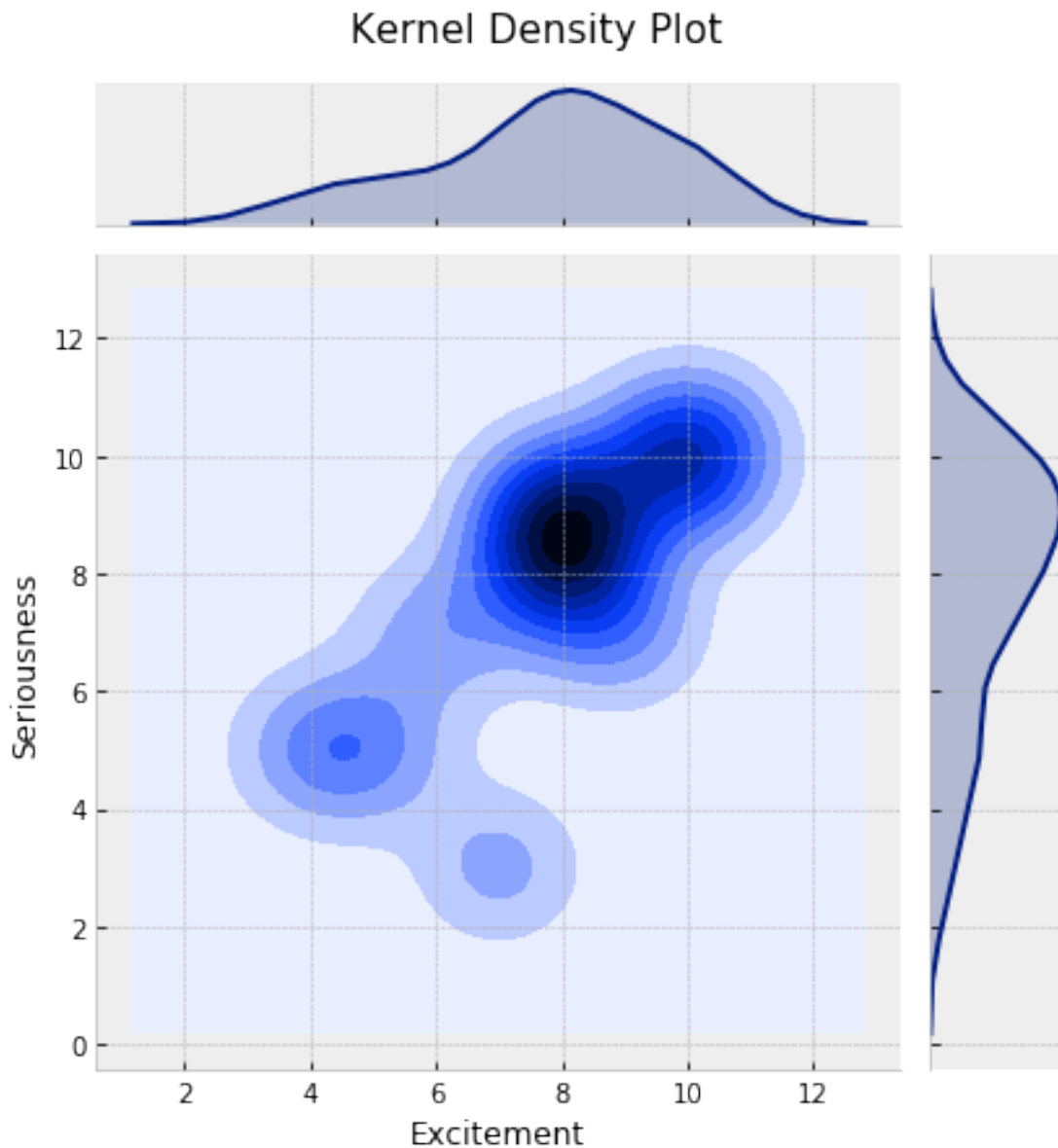
About 61% of the 2nd year students are project-based learners and are aware of the DS SIG. They comprise of 22.2% of the project experienced who are aware of the DS SIG, with every other 2nd years preferring project-based learning. Also 73% of the 2nd years know the DS SIG.

Considering the numbers above it is clear that a lot of 2nd year students are still exploring the other career options and trying them out. Also most of them here have a seriousness scale close to 9 which says that they have Data Science as an option but are not sure enough to pursue the same. There is a huge number of students who know the SIG but are not data driven, maybe they had started in their 1st year but kept it aside for the time being.

This is also the reason why there are just 2/9 students who are project experienced and as a result they are not interested in sharing their project links.

[27]:
```
kde_joint = sns.jointplot(x ='Excitement scale', y ='Seriousness scale',data =␣
  ↪k[k['Year']=='III'], kind = 'kde')
kde_joint.fig.suptitle('Kernel Density Plot', fontsize = 15, y = 1.04)
kde_joint.set_axis_labels(xlabel = 'Excitement', ylabel = 'Seriousness')
```

[27]: <seaborn.axisgrid.JointGrid at 0xb982978>

Kernel Density Plot
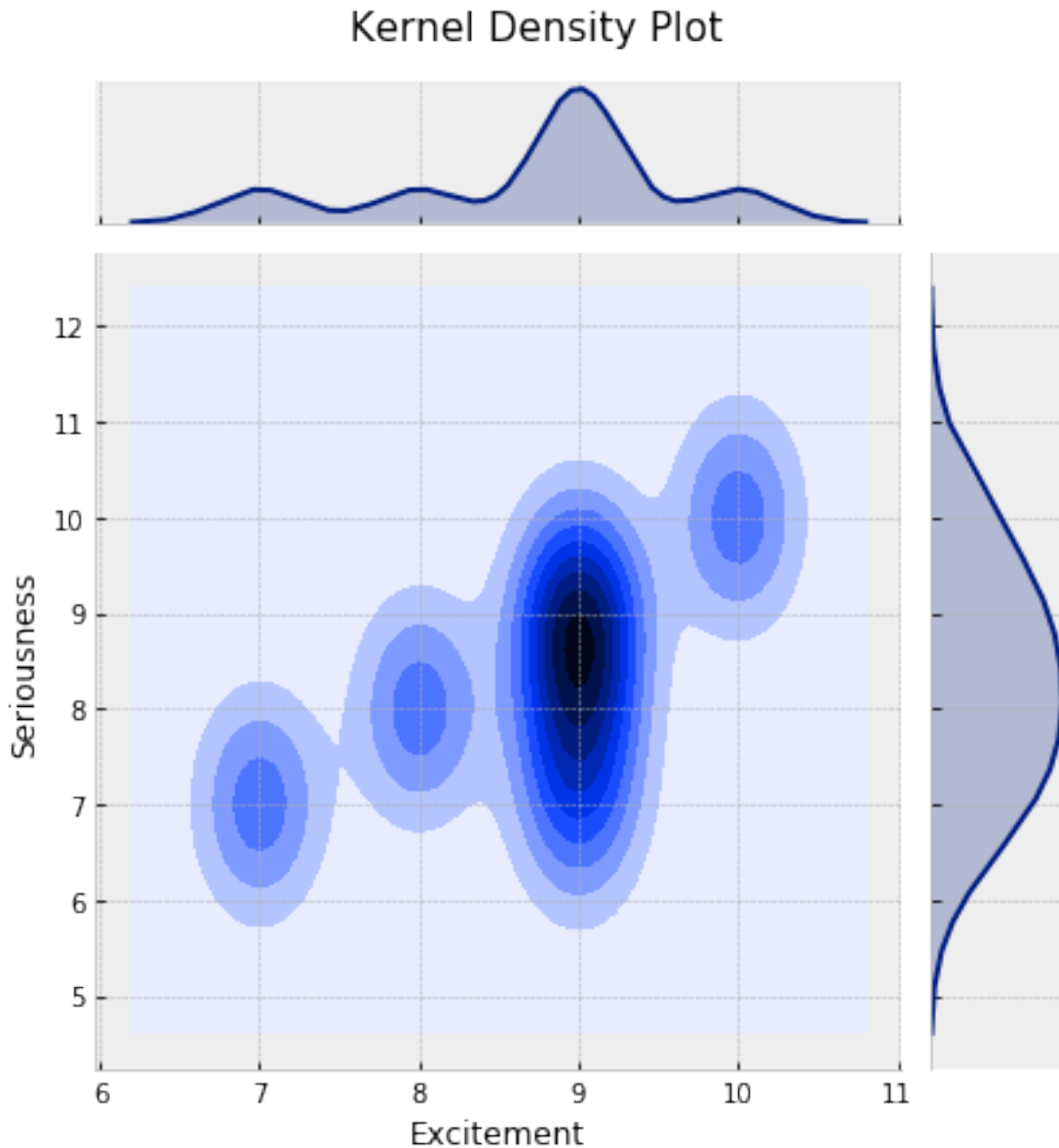
There are no 3rd year students who are project experienced, as a result there are no project links from their side. There are only around 28.6% of the students who are project-based and are aware of the DS SIG. The possibility is that these students understand project experience in a different way, they are still excited about the subject but have a seriousness dramatically varying from 7 to 10. This proves that the 3rd years are pretty unsure of a career in DS and many of them

are disconnected from the DS SIG, this could be the reason of a 35.6% drop from 2nd to 3rd year on DS SIG awareness.

Considering the dropping seriousness of the 3rd years with a very few who still stick to DS, it is best to direct their 7.5 (approx) seriousness rate into a productive progress by helping them explore the possibility and ways to succeed in Data Science.

None of the 4th years who are project experienced know the SIG and there are none who prefer project-based learning too, this proves their disinterest towards doing DS projects and sharing their project links as they are well trained in the field with 100% seriousness but are away from the students with similar interests, this is a considerable demotivation factor.

Also the Mathstat Fear of the 4th years is as good as 4.2 which is clear from the fact that all of them prefer online courses and have a mean seriousness of 9.4.

Conclusion : Also the Mathstat Fear of the 4th years is as good as 4.2 which is clear from the fact that all of them prefer online courses and have a mean seriousness of 9.4.

The DS SIG has to be publicised in a more effective way. At this point it is crucial for the DS SIG that they direct the knowledge of the 4th years into actual practicals i.e. Through projects. This will definitely encourage the whole of the data science community of the college to come forward and work, this will also direct the high excitement (8.37) to some meaning and also develop interest (especially for the current 3rd year status).

Surprisingly, 6 out of these 9 people know or are part of DS SIG. Well, this is again advantageous as 7/9 people are project and online based. Why I say this is that because these students can be easily reached by the SIG and these students are in turn an encouragement to their kind of crowd (which is huge), this definitely can direct an excitement rate that is 8.37 throughout the college.

1.9 Learning Python

```
[28]: py = df['Learning sources(Py)'][0].split(',')
      for i in range(1,69):
          if (str(df['Learning sources(Py)'][i]) == 'nan'):
              continue
          py += str(df['Learning sources(Py)'][i]).split(',')
```

```
[29]: for i in range(len(py)):
          py[i] = py[i].strip()
          py[i] = py[i].lower()
          py[i] = py[i].capitalize()
```

```
[30]: py.sort()
      for i in range(4,13):
          py[i] = 'Others'
      py[48] = 'Others'
      py[57] = 'Others'
```

```
[31]: for i in range(len(py)):
          if py[i] == 'The in-house "winter codex 2018-python edition" beginner\'s␣
      ↪guide':
              py[i] = 'Winter Codex'
          if py[i] == 'Whatever i can find on google':
              py[i] = 'Google'
```

The attribute had multiple entries which were separated by a comma.So, I made a list of the seperated values which i later converted to a Series object in pandas to use various functions of pandas upon these. For plotting, I'm considering all the values which have a frequency more then two and making all the other objects as 'Others'. And also renaming the first two as Winter Codex and Google.

```python
bar, ax = plt.subplots()
ax = sns.countplot(y = 0, data = pd.DataFrame(py))
ax.set_title('Learning Python')
ax.set_ylabel('Sources')
ax.set_xlabel('Number of students')
```

[32]: Text(0.5, 0, 'Number of students')
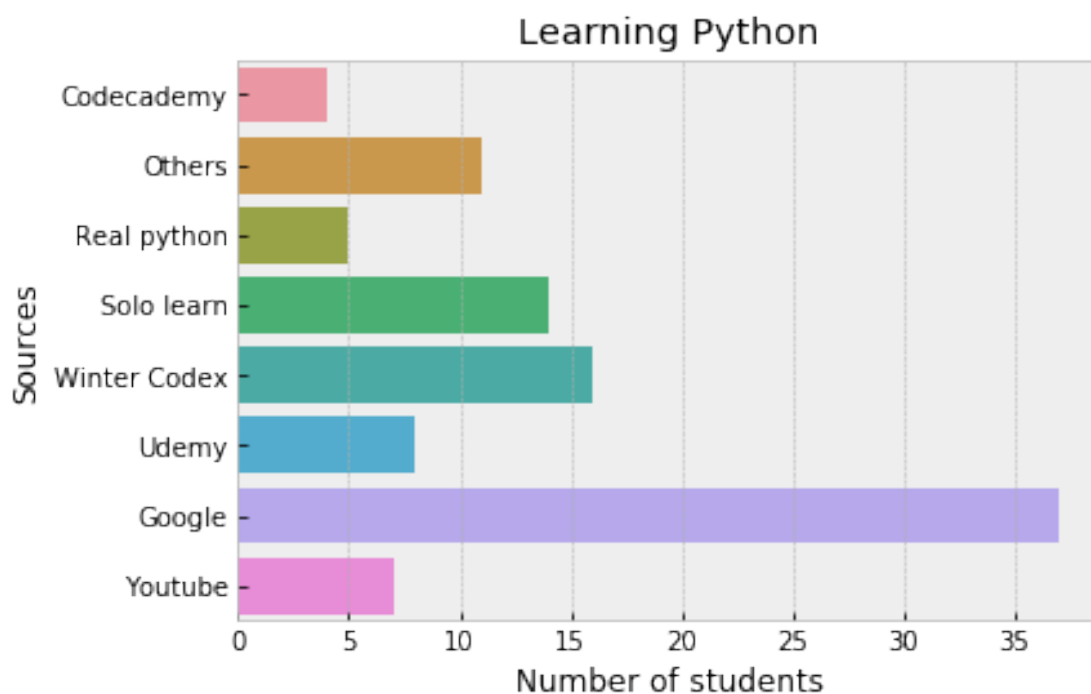
Most students learn python from Google. - (37 students) Youtube is also pretty popular due to real python in visual format leading to saving time and stuff. - (7 students) Winter codex is the second most popular source to learn python because all the basic essentials are in a single place and nicely documented too. - (16 students) Few of the lot is also going through the codes on real time or also learning by themselves. - (19 students) All the others are learning through online courses through various websites like Udemy, Codcademy etc. - (23 students)

1.10 Learning preference

Q) What kind of students has the DS SIG reached the most?

```
[33]: fig, ax = plt.subplots()
      ax = sns.countplot(y = 'Learning preference', data = df[df['SIG Aware']=='Yes'])
      ax.set_title('People aware of the SIG :')
      ax.set_xlabel('Number of students')
      ax.set_ylabel('Options')
```

[33]: Text(0, 0.5, 'Options')
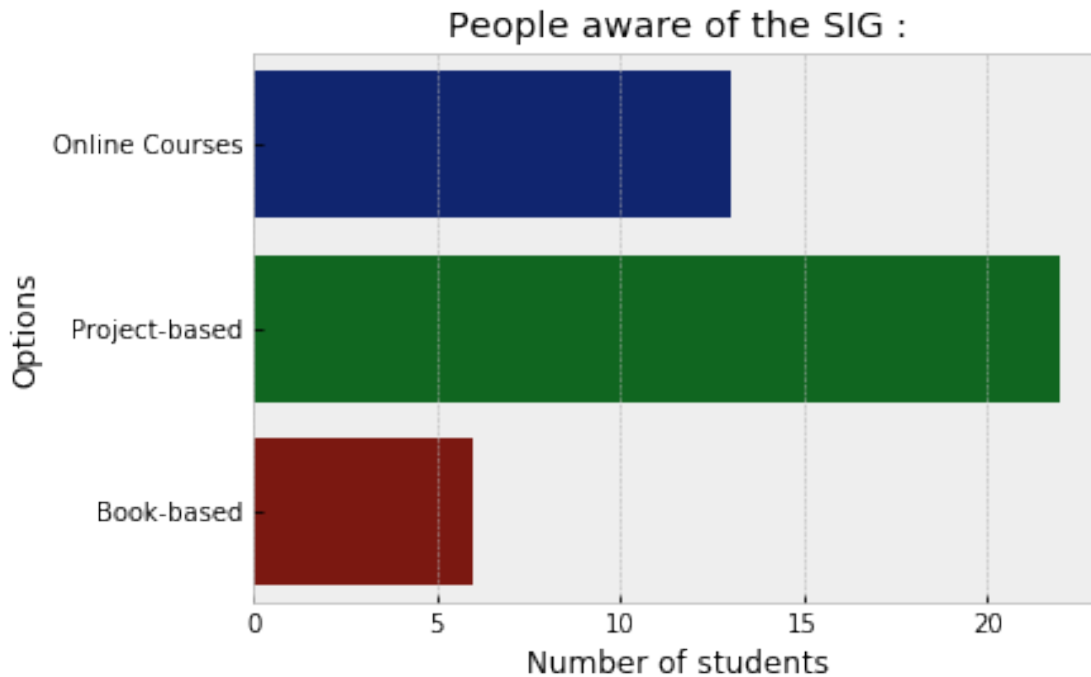
To answer this question we consider the studying preference feature which comprises of three essential values i.e. project-based, online courses and book-based. The above plot depicts that the SIG has been known most, by project-based and online course learners. Project and online course learners : 88.4%

Book-based : 10%

Belief-based : 1%
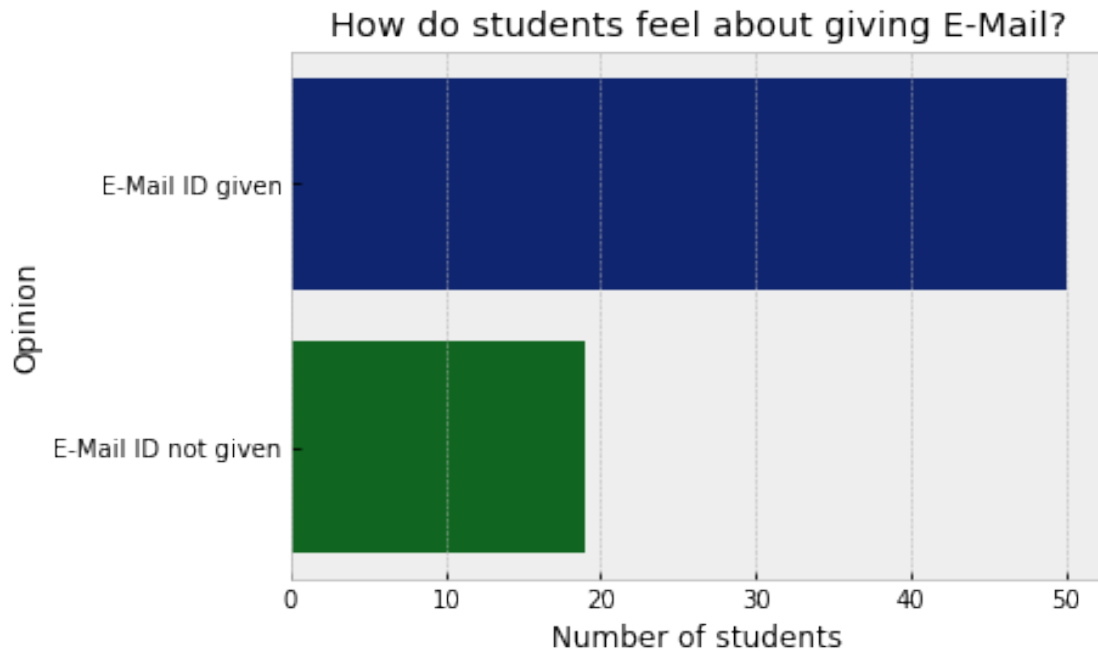
1.11 E-Mail ID

Q) Why hasn't everyone shared their E-Mail ID?

```
[34]: e = []
      for i in range(69):
          if str(df['E-mail ID'][i]) == 'nan':
              e.append('E-Mail ID not given')
          else:
              e.append('E-Mail ID given')
```

```
df['SS'] = pd.DataFrame(e)
```

[35]:
```
em, ax = plt.subplots()
ax = sns.countplot(y = 'SS', data = df)
ax.set_title('How do students feel about giving E-Mail?')
ax.set_xlabel('Number of students')
ax.set_ylabel('Opinion')
```

[35]: `Text(0, 0.5, 'Opinion')`

50 students agree that the SIG wouldn't spam their E-Mails.

19 students feel its a scam and do not enter their E-Mails.

[36]:
```
Ex = [df[df['SS']!='E-Mail ID not given']['Excitement scale'].
 →mean(),df['Excitement scale'].mean(),df[df['SS']=='E-Mail ID not␣
 →given']['Excitement scale'].mean()]
Ss = [df[df['SS']!='E-Mail ID not given']['Seriousness scale'].
 →mean(),df['Seriousness scale'].mean(),df[df['SS']=='E-Mail ID not␣
 →given']['Seriousness scale'].mean()]
x=[2,4,6]
plt.plot(x,Ex,label='Excitement means')
plt.plot(x,Ss,label='Seriousness means')
plt.xlabel('x-axis')
plt.ylabel('y-axis')
plt.title("E-Mail Id and factors")
plt.legend()
```

[36]: `<matplotlib.legend.Legend at 0xbaedf28>`

E-Mail Id and factors
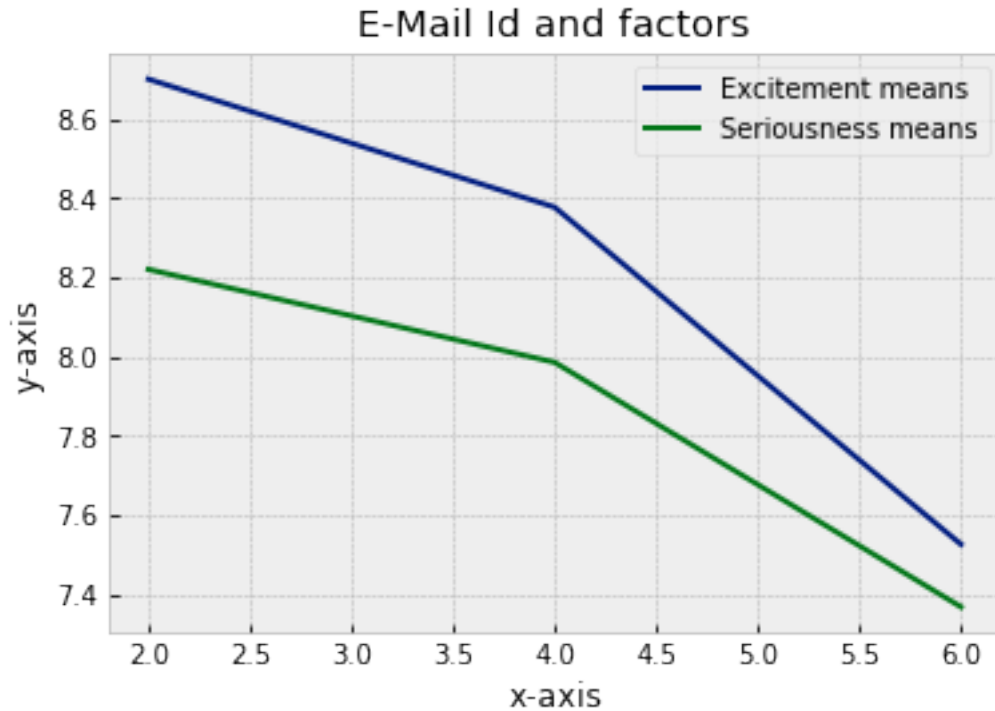
On analysing the data set there are some self-explanatory numbers -

1) The mean excitement and seriousness of the students who have shared their e-mail are 8.7 and 8.22 respectively. These comprise of 50/69 students, which proves the excitement of the students and to great extent the reputation of the DS SIG. This when compared to the overall rates (exc.- 8.37 and ser.- 8) depict that the people who have shared their emails are highly excited as well as serious enough, they want the information to be shared irrespective of what they actually do with it (discussed above).

2) 38% of the students who have shared their emails are not aware or are doubtful about the existence of the DS SIG. It is worth grouping this lot (with half the students being part of 1st or 2nd year) separately and carefully sending them quality content considering their above average seriousness towards DS (also the fact that the SIG has been able to reach more project-based and online course students). This increases the scope of high participation in the coming years as half the students are part of 1st or 2nd year.

3) The ones who haven't shared the email have an excitement rate of 7.53 and seriousness of 7.3 much away from 8.37 and 8. They are not interested and neither are they excited enough to gain any related information. About 53% students who haven't shared their mails know the DS SIG, which gives a clear picture that these are the ones who don't want to SPAM their inbox.

4) The remaining 9 students (47%) are not aware of the DS SIG. This brings in the Publicity of the SIG where the SIG can connect to these students which could probably affect their seriousness towards the subject. On deeper analysis it is found that about 78% students

25

comprise of 2nd and 4th years. The 2nd years can definitely change tracks and get back to DS, this can partly be dependent on the kind of participation of the 3rd and 4th years.

5) The conclusion is that the SIG should focus on bringing in the experienced 3rd and 4th years into quality projects, for the overall popularity and participation of the DS SIG.

6) There are cases where the seriousness is as less as 2 or 3, but they have still shared their emails, these students could be excitement driven where they are interested in going through courses, projects etc.

Done by - Aditya and Kishan