Big Data Applications

Student Survey Analysis

Kavish Kothari

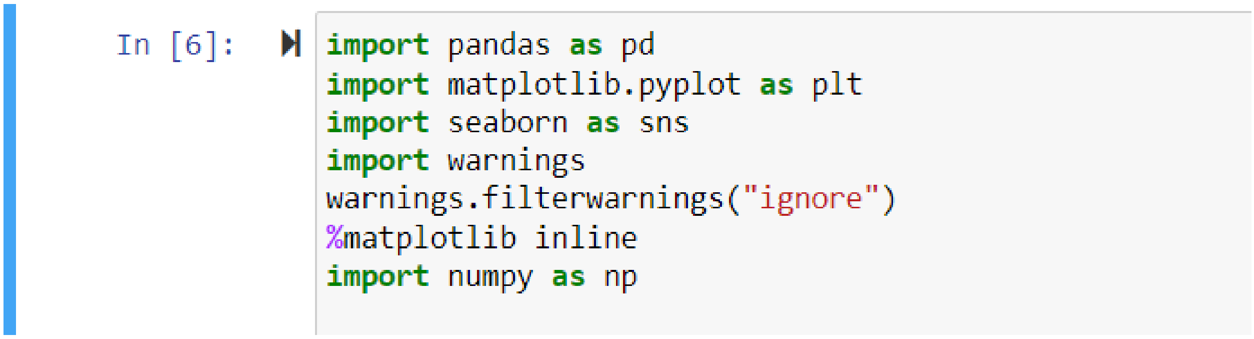
ID - 2001102160

Data Analysis Track

Introduction

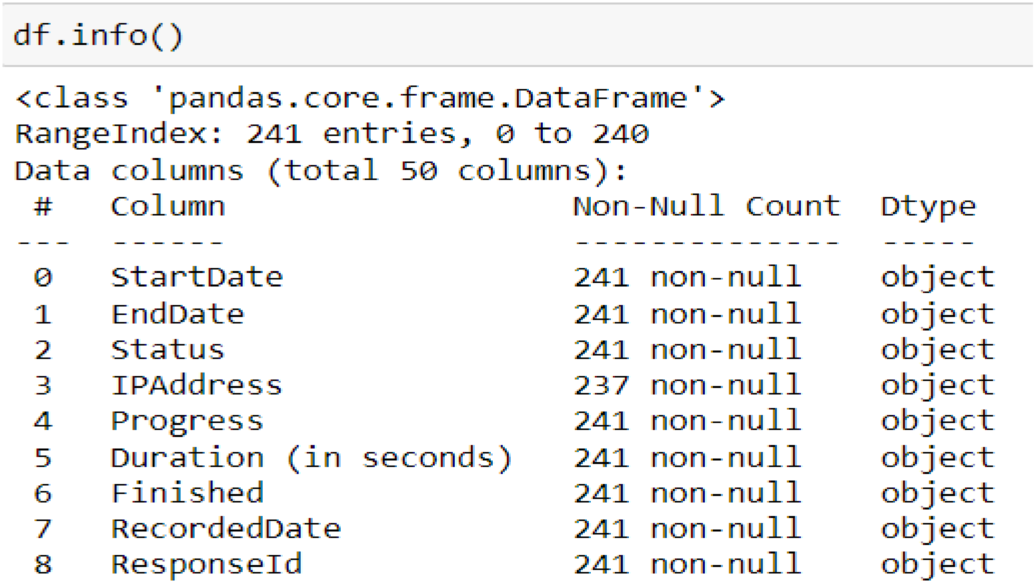
In this project, Dataset is provided which consist of details regarding the survey of Luddy Students in which data of almost 240 respondents is present. For analysis of data .

The Libraries used for performing data analysis in this project are Numy , matplotlib, Seaborn, Pandas.



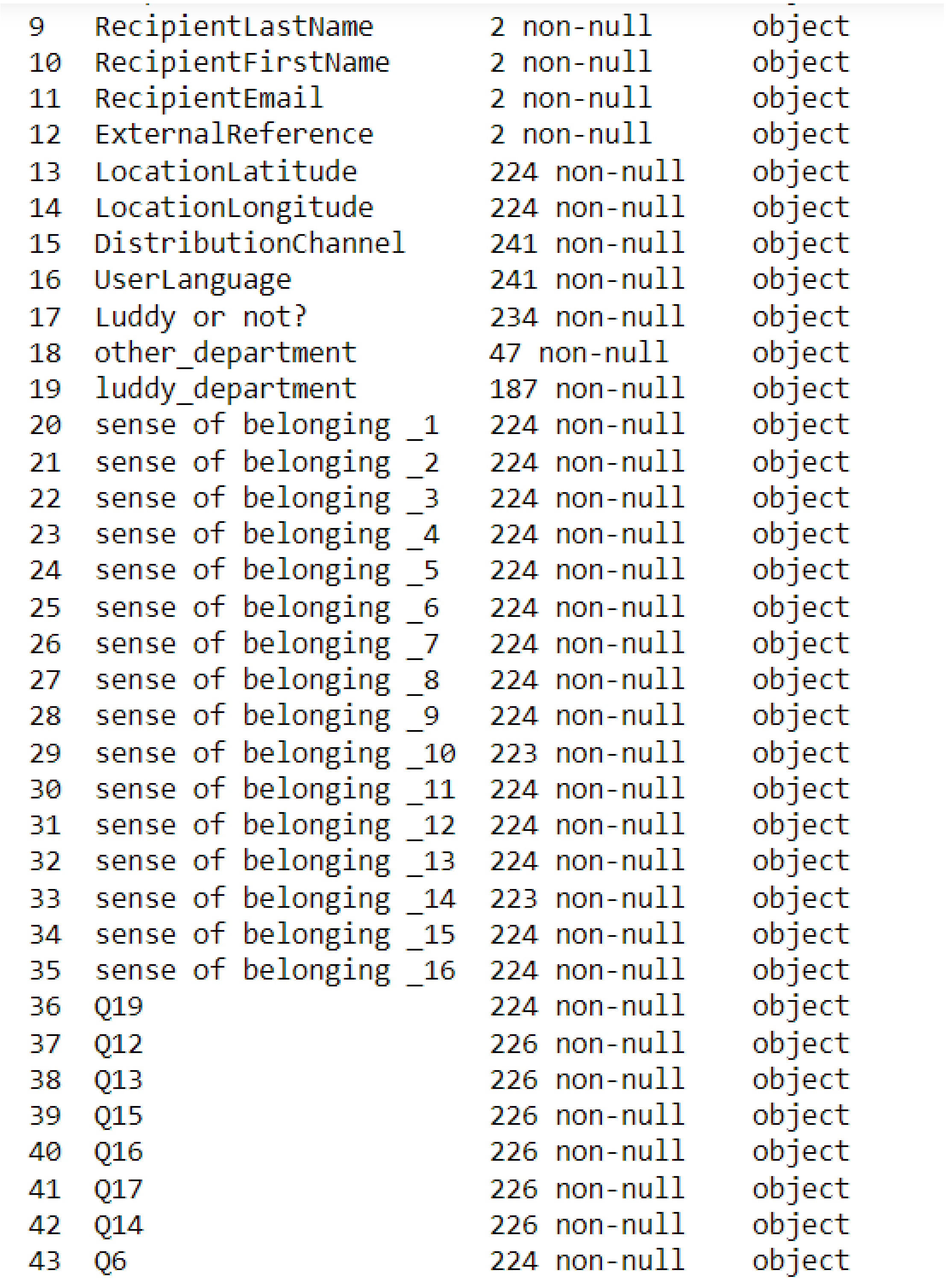
```
In [6]:    import pandas as pd
           import matplotlib.pyplot as plt
           import seaborn as sns
           import warnings
           warnings.filterwarnings("ignore")
           %matplotlib inline
           import numpy as np
```

Now, to get data from dataset.csv, I have used pandas library to extract data from dataset.csv file . In dataset There are 50 columns with 241 entries in it.



```
df.info()
```

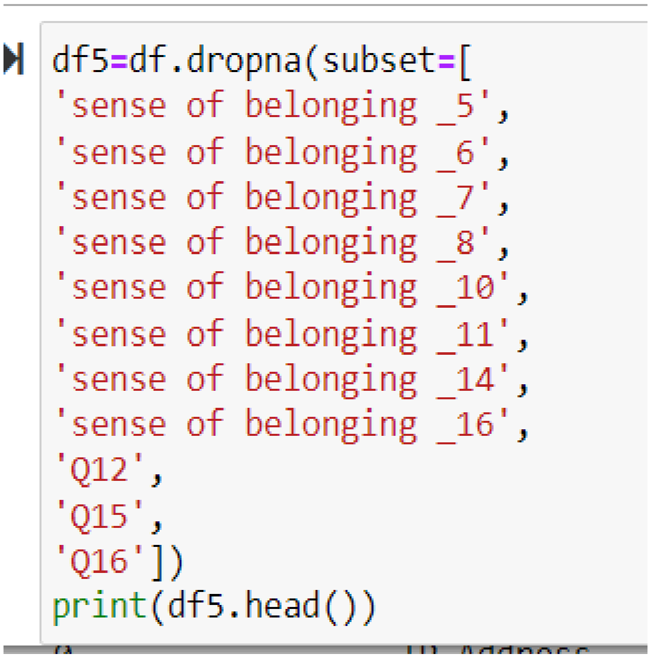
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 241 entries, 0 to 240
Data columns (total 50 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   StartDate             241 non-null    object
 1   EndDate               241 non-null    object
 2   Status                241 non-null    object
 3   IPAddress             237 non-null    object
 4   Progress              241 non-null    object
 5   Duration (in seconds) 241 non-null    object
 6   Finished              241 non-null    object
 7   RecordedDate          241 non-null    object
 8   ResponseId            241 non-null    object
```



```
9   RecipientLastName      2 non-null      object
10  RecipientFirstName     2 non-null      object
11  RecipientEmail         2 non-null      object
12  ExternalReference      2 non-null      object
13  LocationLatitude       224 non-null    object
14  LocationLongitude      224 non-null    object
15  DistributionChannel    241 non-null    object
16  UserLanguage           241 non-null    object
17  Luddy or not?          234 non-null    object
18  other_department       47 non-null     object
19  luddy_department       187 non-null    object
20  sense of belonging _1  224 non-null    object
21  sense of belonging _2  224 non-null    object
22  sense of belonging _3  224 non-null    object
23  sense of belonging _4  224 non-null    object
24  sense of belonging _5  224 non-null    object
25  sense of belonging _6  224 non-null    object
26  sense of belonging _7  224 non-null    object
27  sense of belonging _8  224 non-null    object
28  sense of belonging _9  224 non-null    object
29  sense of belonging _10 223 non-null    object
30  sense of belonging _11 224 non-null    object
31  sense of belonging _12 224 non-null    object
32  sense of belonging _13 224 non-null    object
33  sense of belonging _14 223 non-null    object
34  sense of belonging _15 224 non-null    object
35  sense of belonging _16 224 non-null    object
36  Q19                    224 non-null    object
37  Q12                    226 non-null    object
38  Q13                    226 non-null    object
39  Q15                    226 non-null    object
40  Q16                    226 non-null    object
41  Q17                    226 non-null    object
42  Q14                    226 non-null    object
43  Q6                     224 non-null    object
```

Methodology

The Methodology that I employed to clean the dataset (what you looked for, and technique you used to address a particular issue) Firstly I observed all the questions carefully and after that I extracted some of the questions which can be useful for the analysis and then I removed NaN values from the dataset from that questions as to clean data and I stored cleaned data into new dataframe, Moreover, while looking for the important questions I concentrated more on the questions in which faculty and students are involved in that question.

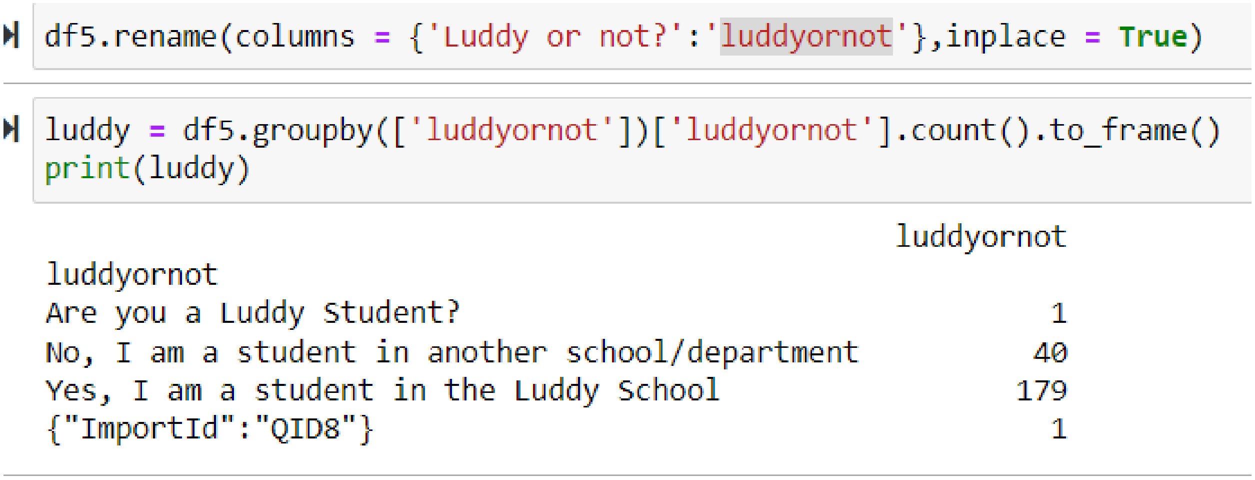


```
df5=df.dropna(subset=[
'sense of belonging _5',
'sense of belonging _6',
'sense of belonging _7',
'sense of belonging _8',
'sense of belonging _10',
'sense of belonging _11',
'sense of belonging _14',
'sense of belonging _16',
'Q12',
'Q15',
'Q16'])
print(df5.head())
```

After cleaning the dataset only 222 rows left so basically for 18 rows the Value were NaN.

Question “ **Luddy or Not” ‘**

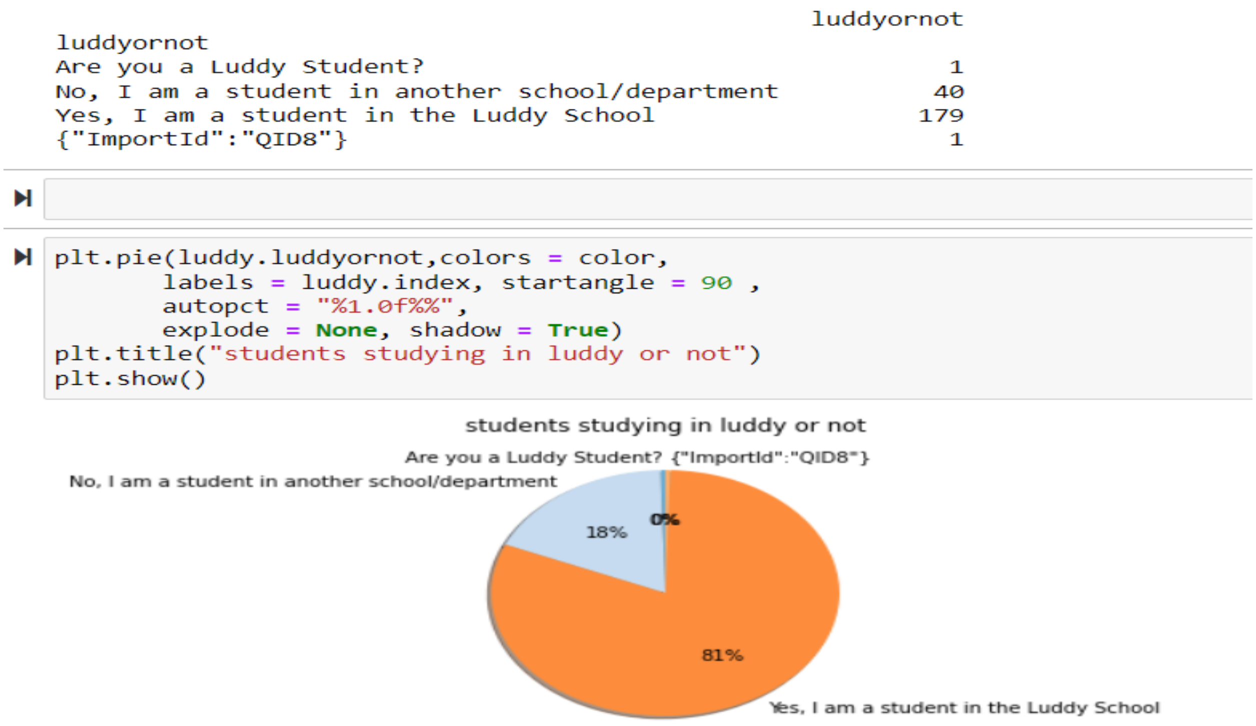
While looking at the dataset for Data cleaning, I found out one interesting questions that “students studying in Luddy or Not” and after doing analysis on that question I found out that 40 students which is equals to 18% are not the students of luddy school but still they are filling out survey questions which are related to luddy school I Think that this can effect the Data Quality. However, it is also possible that for instance, The students who are studying in Kelly school rather than luddy school might have taken few courses which are taught in luddy school by the professors of luddy as I don’t know How the Data is collected so I am not removing the rows in which students have selected that they are not studying in Luddy. But at the same time, It is a dilemma whether to remove the students who are not studying in luddy but are still filling all the questions in the survey form. As this can impact the Data quality and may be we will not be able to get the precise statistics. The statistics are described below.



```
df5.rename(columns = {'Luddy or not?':'luddyornot'},inplace = True)
```

```
luddy = df5.groupby(['luddyornot'])['luddyornot'].count().to_frame()
print(luddy)
```

```
                                                 luddyornot
luddyornot
Are you a Luddy Student?                                  1
No, I am a student in another school/department          40
Yes, I am a student in the Luddy School                 179
{"ImportId":"QID8"}                                       1
```

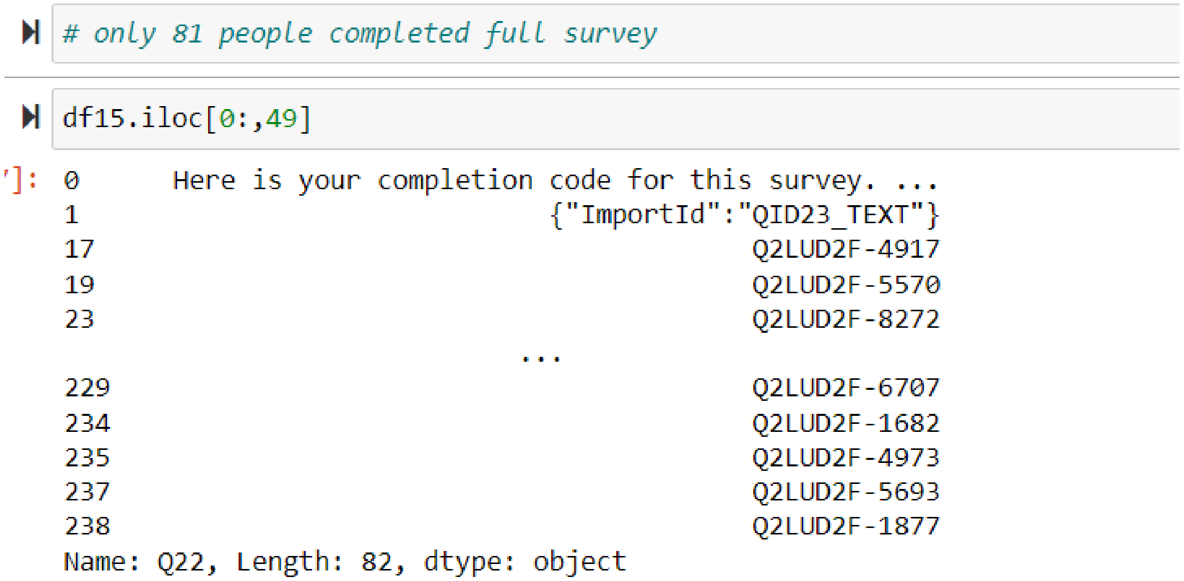


```
                                                      luddyornot
luddyornot
Are you a Luddy Student?                                       1
No, I am a student in another school/department               40
Yes, I am a student in the Luddy School                      179
{"ImportId":"QID8"}                                            1
```

```
plt.pie(luddy.luddyornot,colors = color,
       labels = luddy.index, startangle = 90 ,
       autopct = "%1.0f%%",
       explode = None, shadow = True)
plt.title("students studying in luddy or not")
plt.show()
```

students studying in luddy or not
Are you a Luddy Student? {"ImportId":"QID8"}
No, I am a student in another school/department
18%
0%
81%
Yes, I am a student in the Luddy School

While cleaning the data I also found one more interesting thing that out of 240 students only 81 students only completed the full survey as 81 people only got completion code of survey.



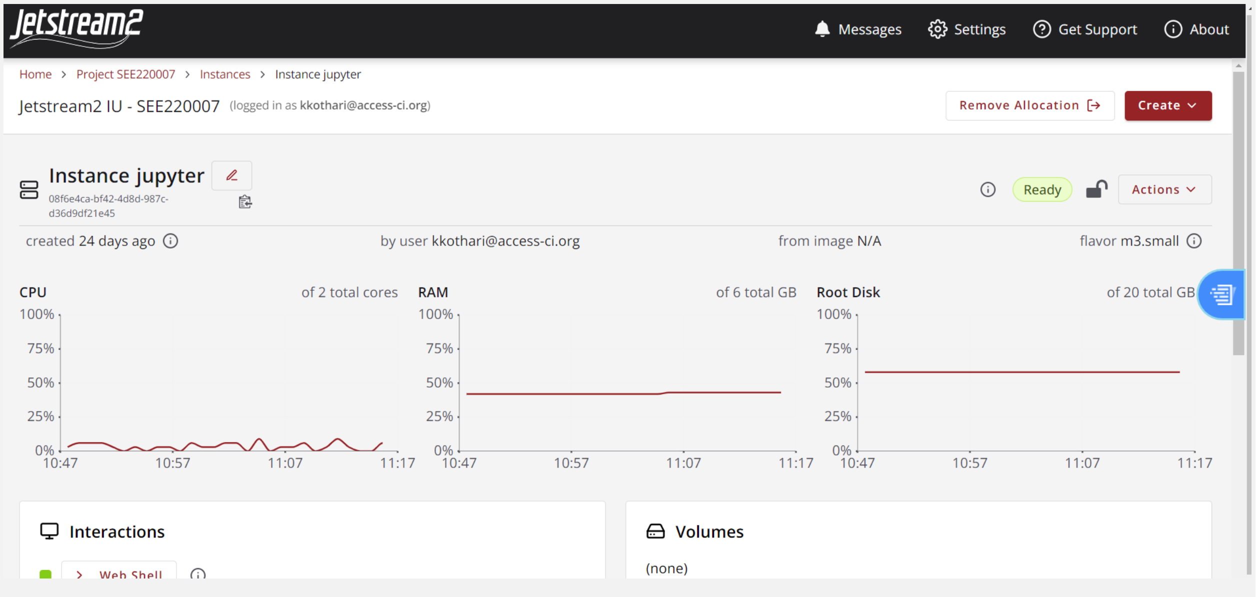
```
# only 81 people completed full survey
```

```
df15.iloc[0:,49]
```

```
]:  0       Here is your completion code for this survey. ...
    1                                 {"ImportId":"QID23_TEXT"}
    17                                             Q2LUD2F-4917
    19                                             Q2LUD2F-5570
    23                                             Q2LUD2F-8272
                                    ...
    229                                            Q2LUD2F-6707
    234                                            Q2LUD2F-1682
    235                                            Q2LUD2F-4973
    237                                            Q2LUD2F-5693
    238                                            Q2LUD2F-1877
    Name: Q22, Length: 82, dtype: object
```

Jetstream VM

I used jupyter notebook for performing the data analysis on jetstream2 by launching with the web desktop enabled. As it makes it very easy to access a jupyter notebook.



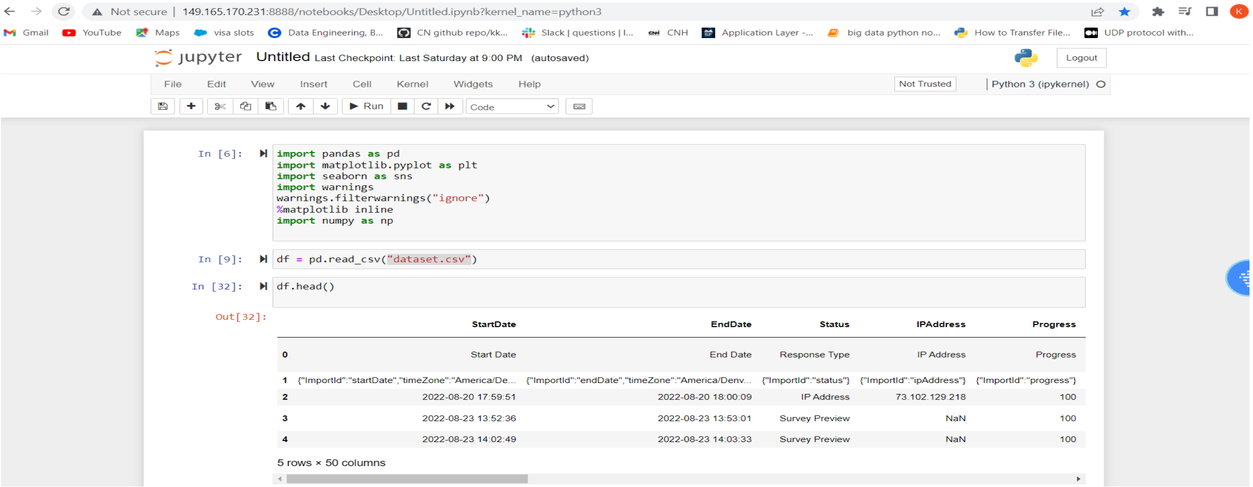
Jetstream2 IU - SEE220007 (logged in as kkothari@access-ci.org)

Instance jupyter
08f6e4ca-bf42-4d8d-987c-d36d9df21e45

created 24 days ago | by user kkothari@access-ci.org | from image N/A | flavor m3.small

Volumes: (none)

Jupyter accessed by jetstream VM



Untitled Last Checkpoint: Last Saturday at 9:00 PM (autosaved)

```
In [6]:  import pandas as pd
         import matplotlib.pyplot as plt
         import seaborn as sns
         import warnings
         warnings.filterwarnings("ignore")
         %matplotlib inline
         import numpy as np

In [9]:  df = pd.read_csv("dataset.csv")

In [32]: df.head()

Out[32]:
```

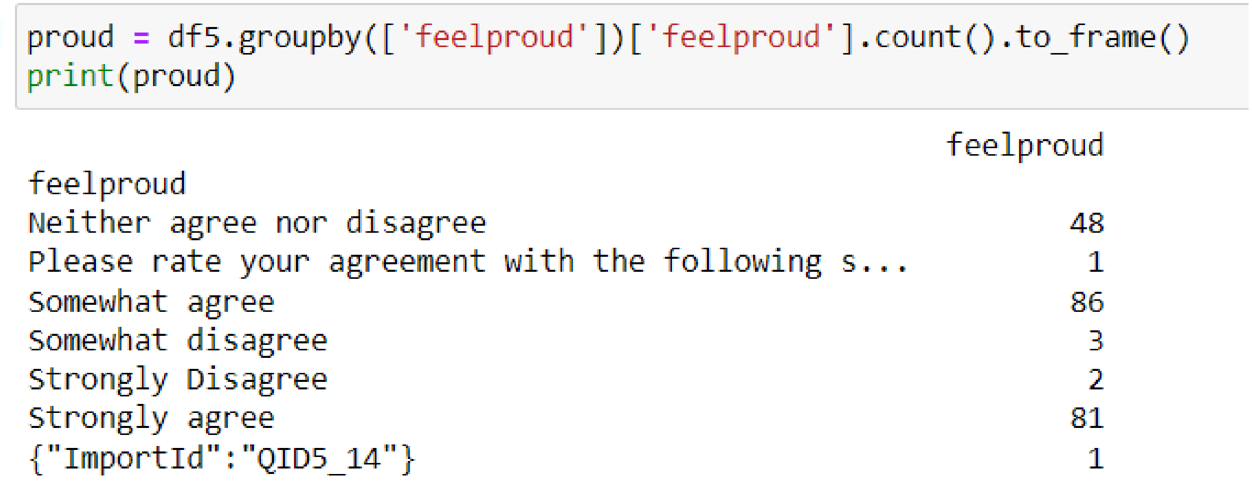
	StartDate	EndDate	Status	IPAddress	Progress
0	Start Date	End Date	Response Type	IP Address	Progress
1	{"ImportId":"startDate","timeZone":"America/De...	{"ImportId":"endDate","timeZone":"America/Denv...	{"ImportId":"status"}	{"ImportId":"ipAddress"}	{"ImportId":"progress"}
2	2022-08-20 17:59:51	2022-08-20 18:00:09	IP Address	73.102.129.218	100
3	2022-08-23 13:52:36	2022-08-23 13:53:01	Survey Preview	NaN	100
4	2022-08-23 14:02:49	2022-08-23 14:03:33	Survey Preview	NaN	100

5 rows × 50 columns

Methodology used in project for Data analysis.

Firstly, I observed all the questions which are given in a dataset for a survey in which I selected some of the questions which can be more useful for analysis and in that questions there were many NaN values so for cleaning data I used dropna function to remove NaN values from the dataset. Secondly, I analyzed some of the questions one by one to get some useful statistics from it. On a primary level, I have focused on the questions which include both students and faculty. As from those statements, we will be able to analyze how much students are compatible with faculty.

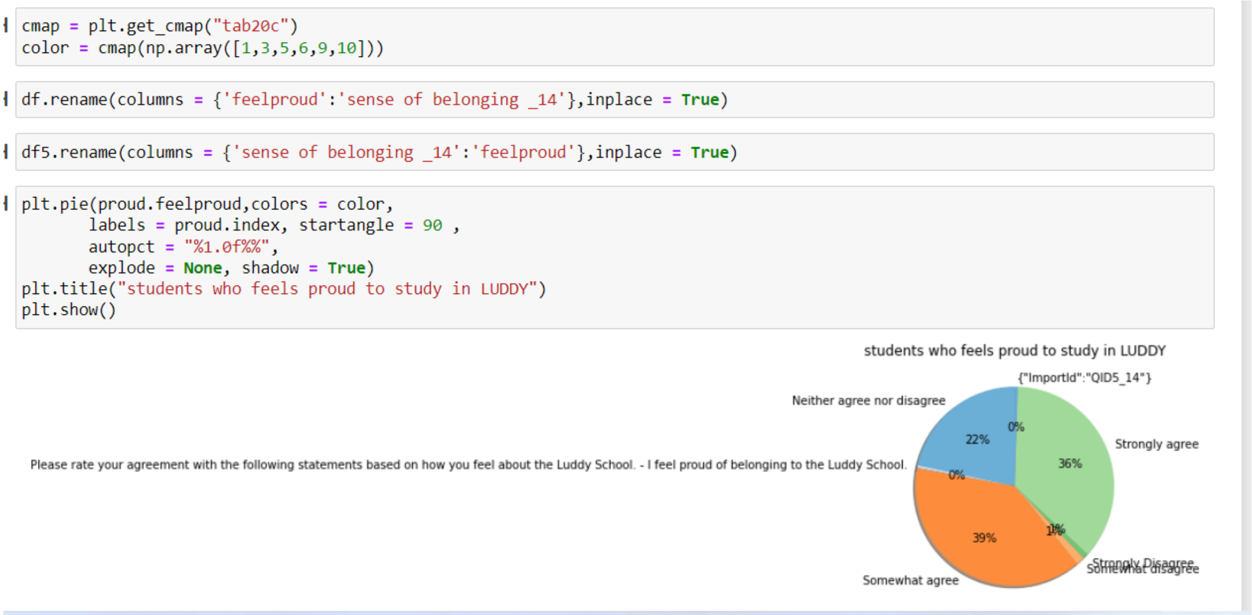
The First question which I found interesting is 'Please rate your agreement with the following statements based on how you feel about the Luddy School. - I feel proud of belonging to the Luddy School.'. To analyze that question first I used groupby function and count function to get counts for responses. For instance, to get the total number of students who totally agreed, who agreed partially. After that, I visualized it by using a pie chart.



```
proud = df5.groupby(['feelproud'])['feelproud'].count().to_frame()
print(proud)
```

```
                                                   feelproud
feelproud
Neither agree nor disagree                                48
Please rate your agreement with the following s...         1
Somewhat agree                                            86
Somewhat disagree                                          3
Strongly Disagree                                          2
Strongly agree                                            81
{"ImportId":"QID5_14"}                                     1
```

Question - 'Please rate your agreement with the following statements based on how you feel about the Luddy School. - I feel proud of belonging to the Luddy School.'



```
cmap = plt.get_cmap("tab20c")
color = cmap(np.array([1,3,5,6,9,10]))
```

```
df.rename(columns = {'feelproud':'sense of belonging _14'},inplace = True)
```

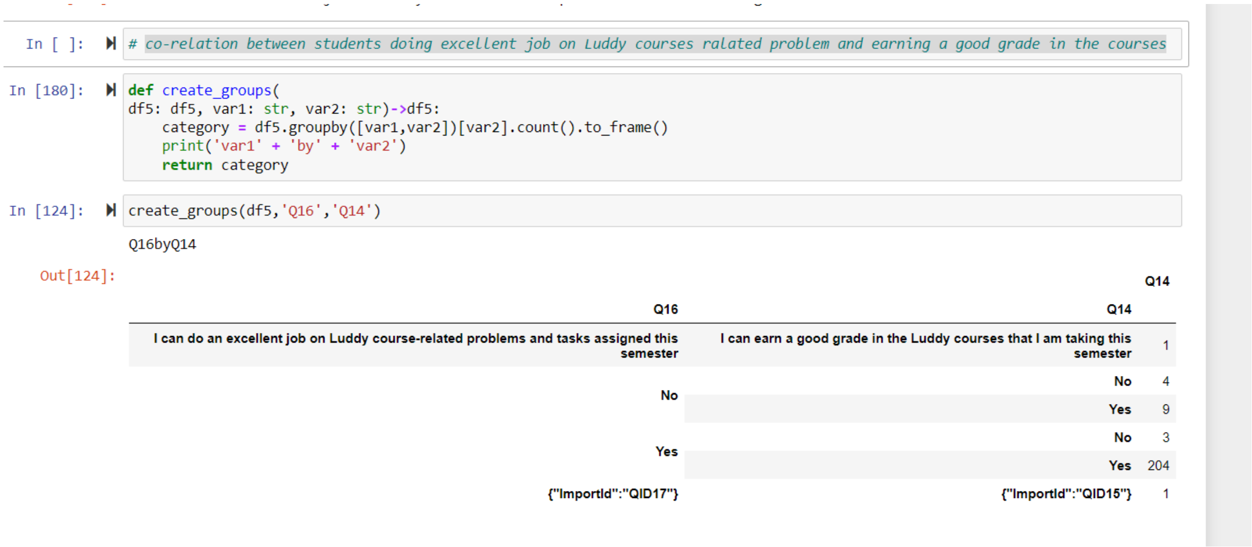
```
df5.rename(columns = {'sense of belonging _14':'feelproud'},inplace = True)
```

```
plt.pie(proud.feelproud,colors = color,
       labels = proud.index, startangle = 90 ,
       autopct = "%1.0f%%",
       explode = None, shadow = True)
plt.title("students who feels proud to study in LUDDY")
plt.show()
```

students who feels proud to study in LUDDY
{"ImportId":"QID5_14"}
Neither agree nor disagree
Strongly agree
Please rate your agreement with the following statements based on how you feel about the Luddy School. - I feel proud of belonging to the Luddy School.
Somewhat agree
Strongly Disagree
Somewhat disagree
0%
22%
36%
0%
39%
1%
1%

From this pie chart, we can clearly observe that only 36% of the students strongly agree with the statement that they feel proud to be students of Luddy school. Moreover, at the same time 39% of the students partially agree with this statement.However, on the other hand 22% people remained Neutral as They neither agree nor disagreed with the statement. It can be concluded that most of the students either agreed strongly or partially that they feel proud as a student of luddy.

Correlation between students doing excellent job on Luddy courses related problems and earning a good grade in the courses

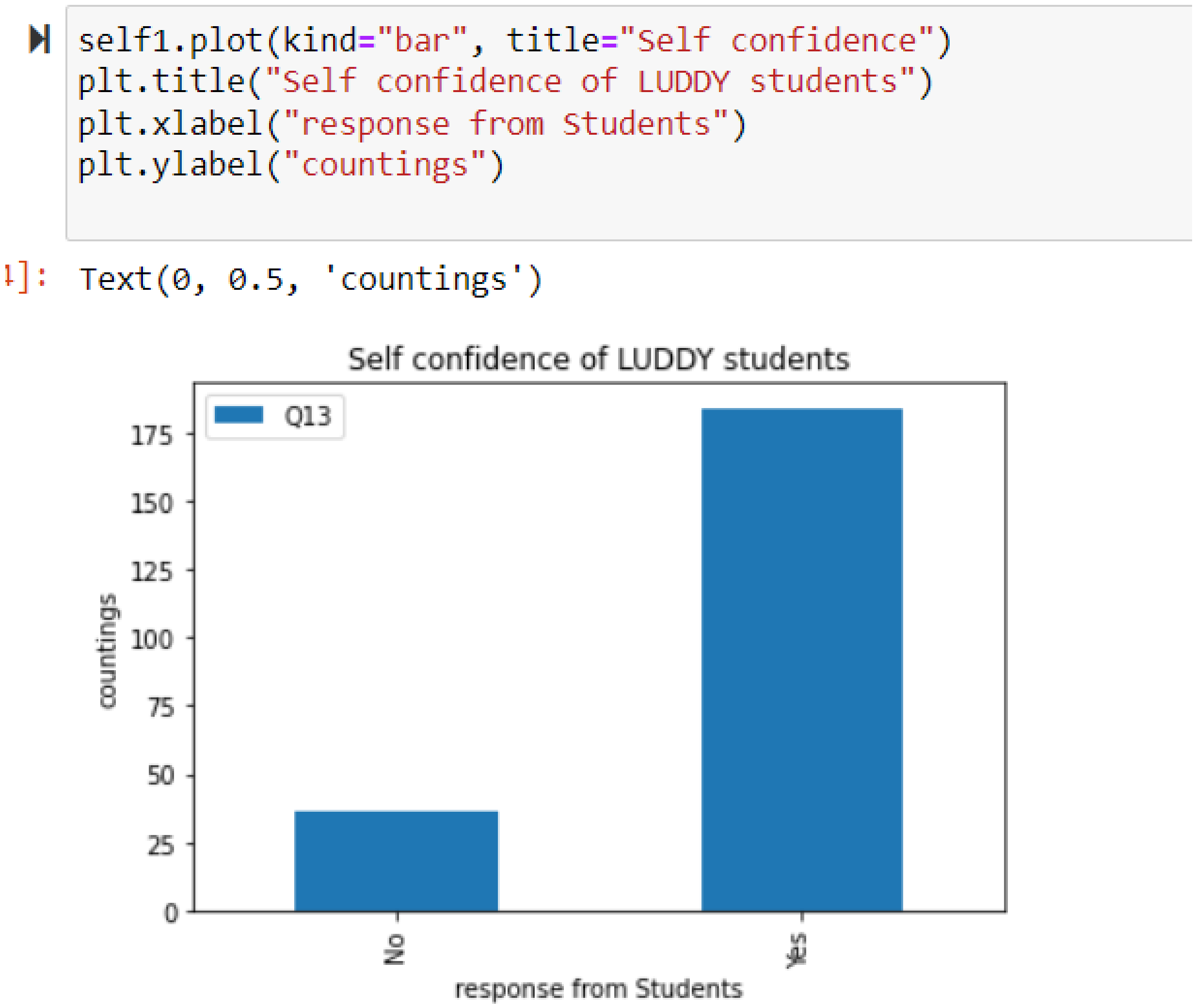


```
In [ ]:   # co-relation between students doing excellent job on Luddy courses ralated problem and earning a good grade in the courses

In [180]: def create_groups(
          df5: df5, var1: str, var2: str)->df5:
              category = df5.groupby([var1,var2])[var2].count().to_frame()
              print('var1' + 'by' + 'var2')
              return category

In [124]: create_groups(df5,'Q16','Q14')

          Q16byQ14
Out[124]:
                                                                                                                                  Q14
                                                                             Q16                                             Q14
   I can do an excellent job on Luddy course-related problems and tasks assigned this   I can earn a good grade in the Luddy courses that I am taking this   1
                                                                        semester                                        semester
                                                                                                                              No     4
                                                                              No
                                                                                                                             Yes     9
                                                                                                                              No     3
                                                                             Yes
                                                                                                                             Yes   204
                                                            {"ImportId":"QID17"}                            {"ImportId":"QID15"}     1
```

From this correlation we can conclude that 204 (majority) of the students who do excellent job on Luddy Courses related problems and tasks assigned in this semester are also able to earn good grades in the luddy courses this semester. On the other hand there are a few students who believe that if they don't do an excellent job but still they will be able to earn a good grade in the luddy courses.

reflection of self confidence by Luddy Students (Q13)

'I could master the content of even the most challenging Luddy courses if I try'



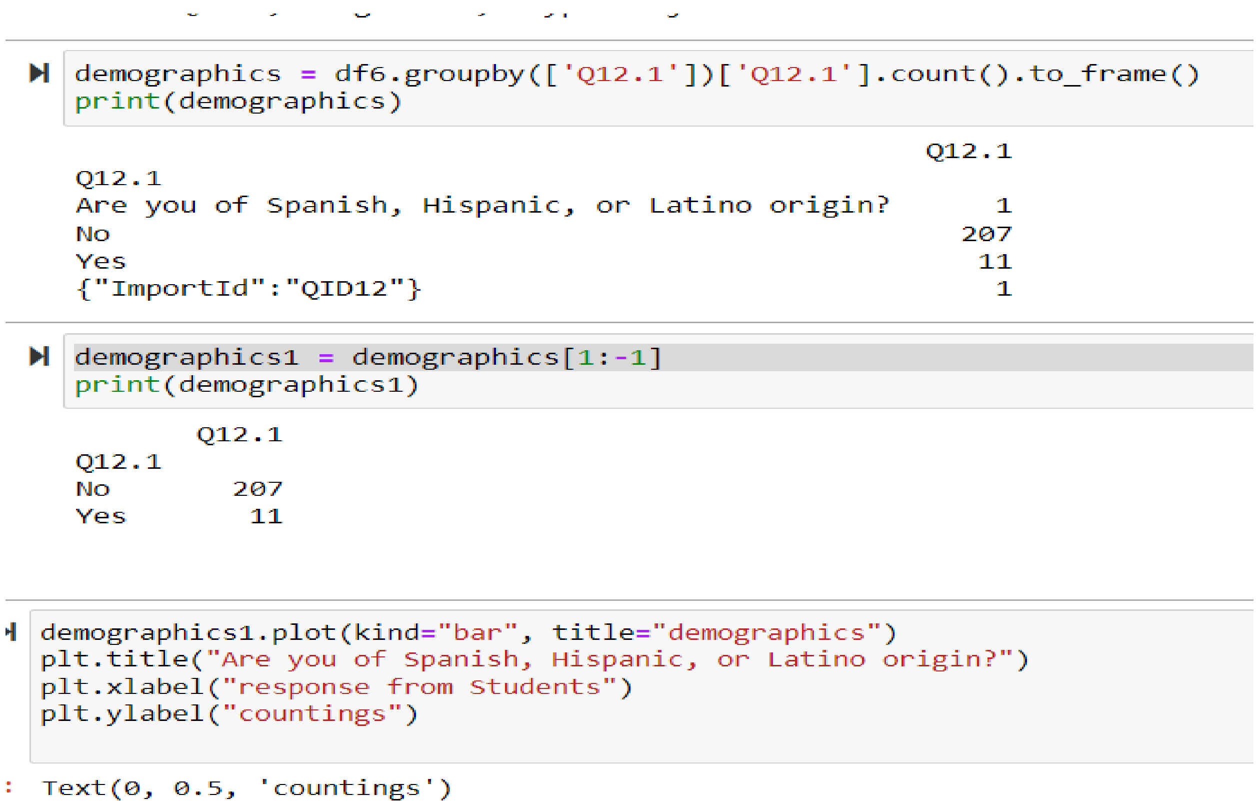
```
self1.plot(kind="bar", title="Self confidence")
plt.title("Self confidence of LUDDY students")
plt.xlabel("response from Students")
plt.ylabel("countings")
```

```
]: Text(0, 0.5, 'countings')
```

From the bar chart, we can clearly observe that the majority of the students feel confident as they can master the content of even the most difficult luddy courses if they try.

Demographics analytics

Q12.1 Are you of Spanish, Hispanic, or Latino origin?



```
demographics = df6.groupby(['Q12.1'])['Q12.1'].count().to_frame()
print(demographics)
```

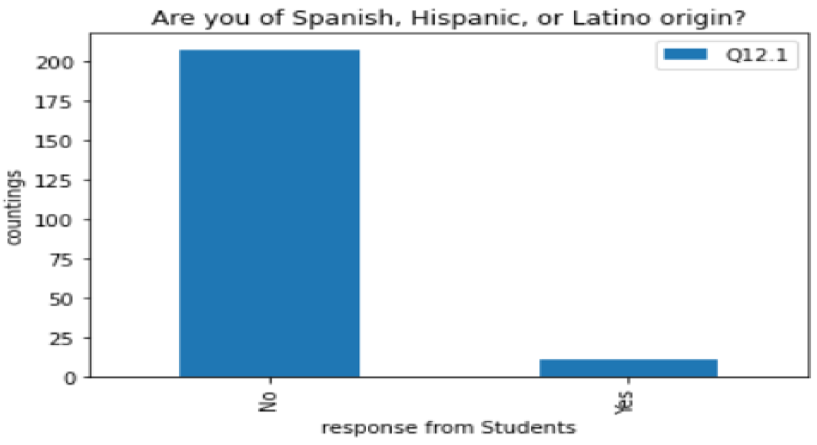
```
                                                  Q12.1
Q12.1
Are you of Spanish, Hispanic, or Latino origin?       1
No                                                  207
Yes                                                  11
{"ImportId":"QID12"}                                  1
```

```
demographics1 = demographics[1:-1]
print(demographics1)
```

```
       Q12.1
Q12.1
No       207
Yes       11
```

```
demographics1.plot(kind="bar", title="demographics")
plt.title("Are you of Spanish, Hispanic, or Latino origin?")
plt.xlabel("response from Students")
plt.ylabel("countings")
```

```
: Text(0, 0.5, 'countings')
```



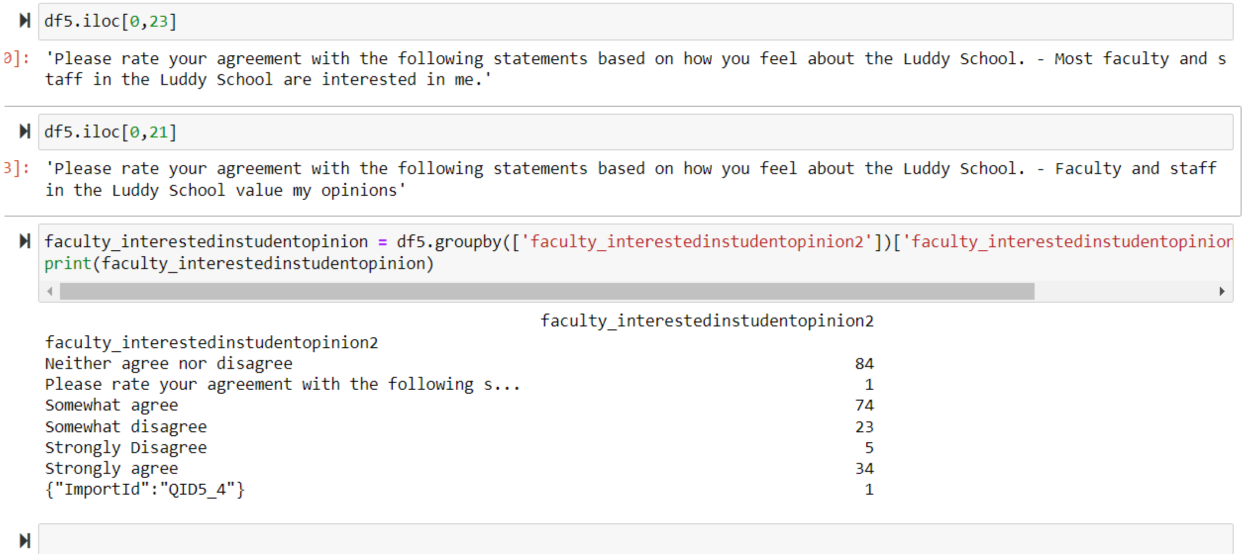
From the Bar graph it is clearly observed that Majority of the Students do not belong to Spanish, Hispanic or Latino origin.

Correlation between how faculty and students are interested in students and value students opinion

While doing data analysis on a data set, I Found out Two questions.

'Please rate your agreement with the following statements based on how you feel about the Luddy School. - Most faculty and staff in the Luddy School are interested in me.' and

'Please rate your agreement with the following statements based on how you feel about the Luddy School. - Faculty and staff in the Luddy School value my opinions'.



```
df5.iloc[0,23]
```

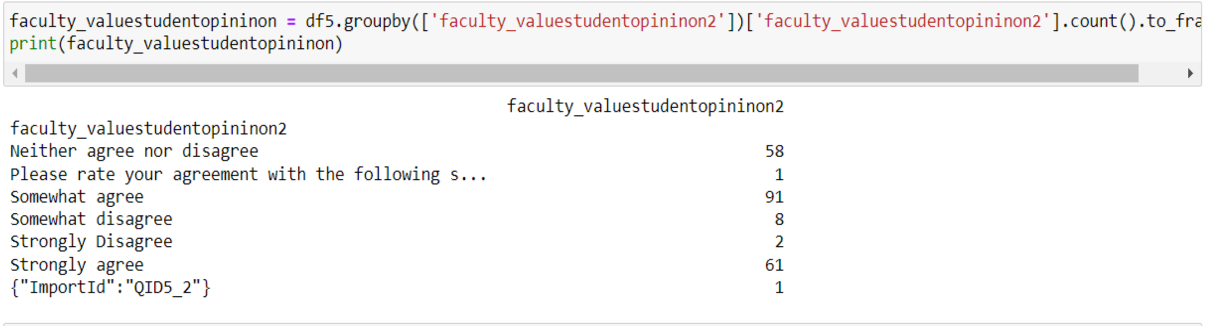
```
0]: 'Please rate your agreement with the following statements based on how you feel about the Luddy School. - Most faculty and s
    taff in the Luddy School are interested in me.'
```

```
df5.iloc[0,21]
```

```
3]: 'Please rate your agreement with the following statements based on how you feel about the Luddy School. - Faculty and staff
    in the Luddy School value my opinions'
```

```
faculty_interestedinstudentopinion = df5.groupby(['faculty_interestedinstudentopinion2'])['faculty_interestedinstudentopinior
print(faculty_interestedinstudentopinion)
```

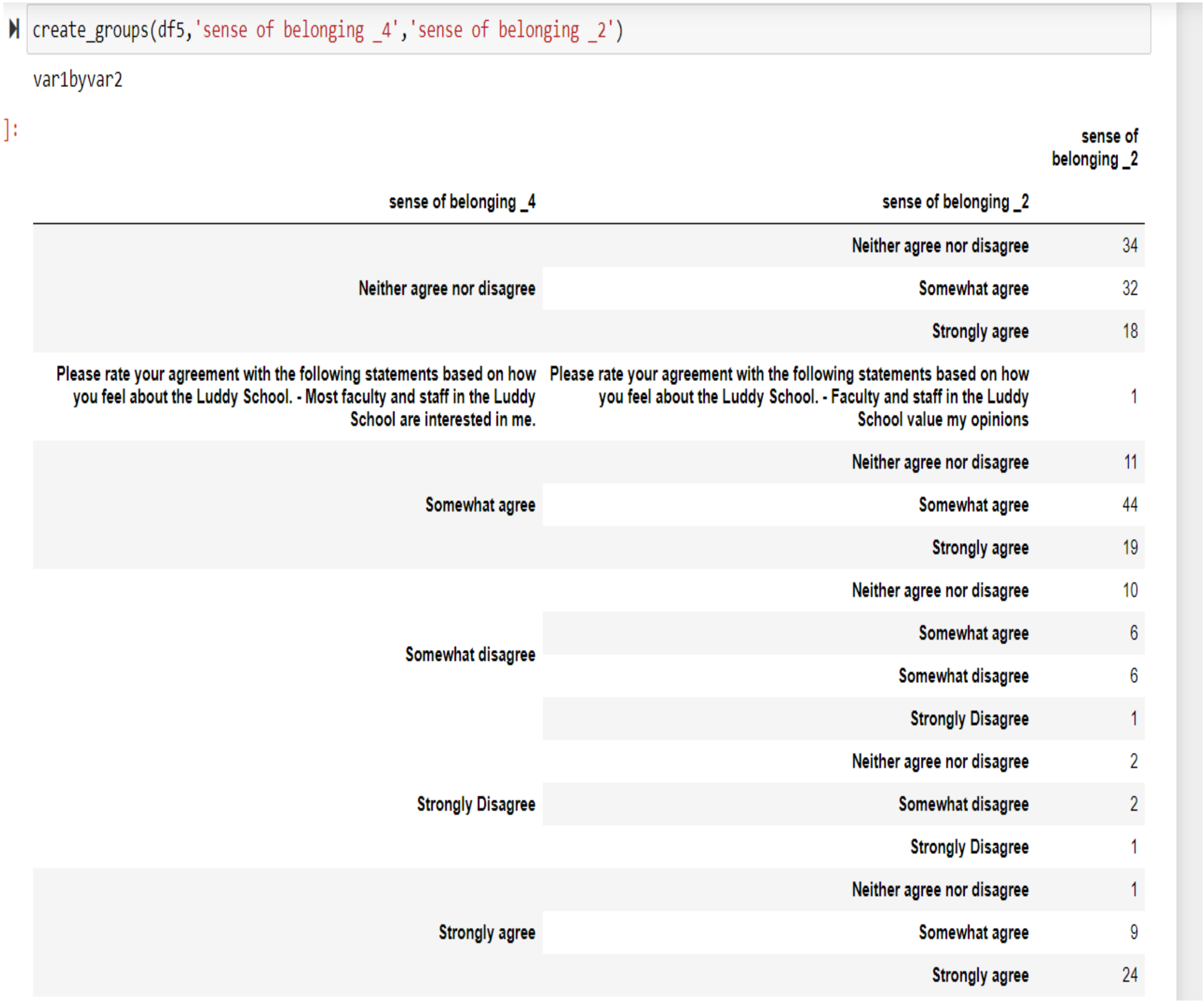
```
                                                    faculty_interestedinstudentopinion2
faculty_interestedinstudentopinion2
Neither agree nor disagree                                                           84
Please rate your agreement with the following s...                                    1
Somewhat agree                                                                       74
Somewhat disagree                                                                    23
Strongly Disagree                                                                     5
Strongly agree                                                                       34
{"ImportId":"QID5_4"}                                                                 1
```



```
faculty_valuestudentopininon = df5.groupby(['faculty_valuestudentopininon2'])['faculty_valuestudentopininon2'].count().to_fra
print(faculty_valuestudentopininon)
```

```
                                                    faculty_valuestudentopininon2
faculty_valuestudentopininon2
Neither agree nor disagree                                                     58
Please rate your agreement with the following s...                              1
Somewhat agree                                                                 91
Somewhat disagree                                                               8
Strongly Disagree                                                               2
Strongly agree                                                                 61
{"ImportId":"QID5_2"}                                                           1
```

After reading statistics from both the questions, I decided to find correlation between both the questions and by finding correlation between these two questions I found some interesting statistics about student which are described below.



```
create_groups(df5,'sense of belonging _4','sense of belonging _2')
```

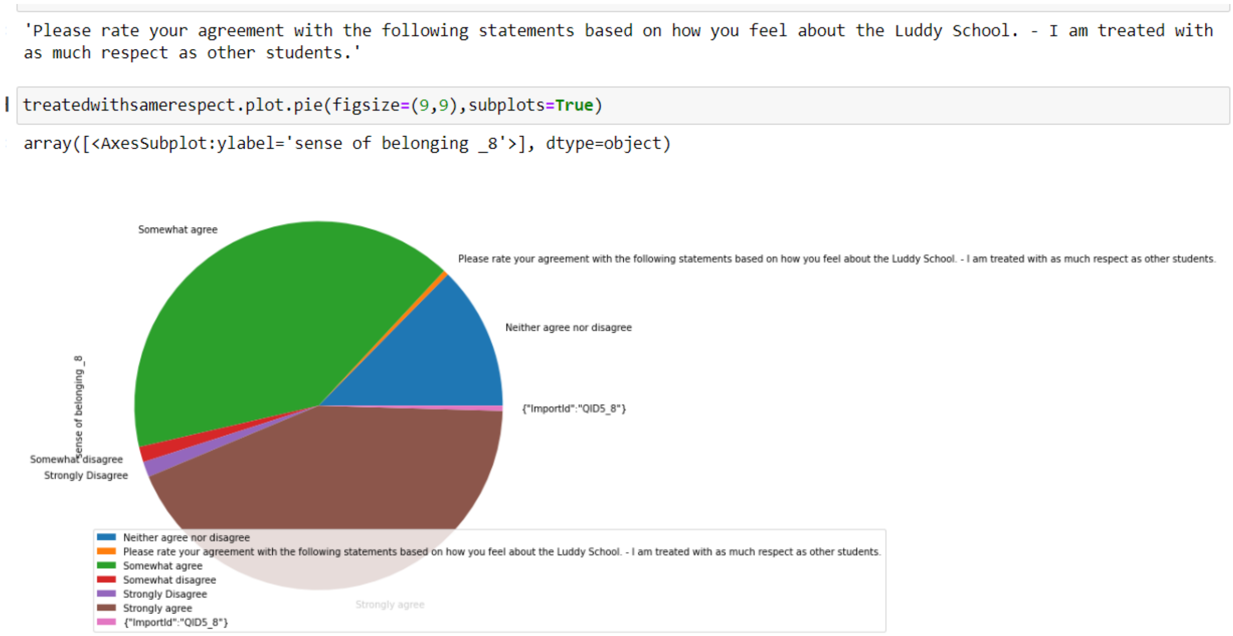
```
var1byvar2
```

		sense of belonging _2
sense of belonging _4	sense of belonging _2	
Neither agree nor disagree	Neither agree nor disagree	34
	Somewhat agree	32
	Strongly agree	18
Please rate your agreement with the following statements based on how you feel about the Luddy School. - Most faculty and staff in the Luddy School are interested in me.	Please rate your agreement with the following statements based on how you feel about the Luddy School. - Faculty and staff in the Luddy School value my opinions	1
Somewhat agree	Neither agree nor disagree	11
	Somewhat agree	44
	Strongly agree	19
Somewhat disagree	Neither agree nor disagree	10
	Somewhat agree	6
	Somewhat disagree	6
	Strongly Disagree	1
Strongly Disagree	Neither agree nor disagree	2
	Somewhat disagree	2
	Strongly Disagree	1
Strongly agree	Neither agree nor disagree	1
	Somewhat agree	9
	Strongly agree	24

By finding correlation between both questions, I found out that There are 32 students who didn't agree or disagree with the question “Most faculty and staff in the Luddy School are interested in me” but partially agreed with the other question “ Faculty and staff value my opinions ”. It’s quite strange because they don’t know exactly that whether the faculty and staff are interested in them or not and in addition 18 students confidently selected strongly agree option in question “Faculty and staff in the luddy value my opinion ” without knowing whether the faculty and staff are invested in them.

Student treated with as much respect as other students

From this graph we can conclude that almost 45 % of the students strongly agreed that they strongly believe that equality is maintained by faculty and other staff of luddy. At the same time almost 30% people selected some what agree option so it seems that they are not sure whether all the students are treated in same manner or not. However, almost 15 % remained neutral as they selected option neither agree nor disagree.



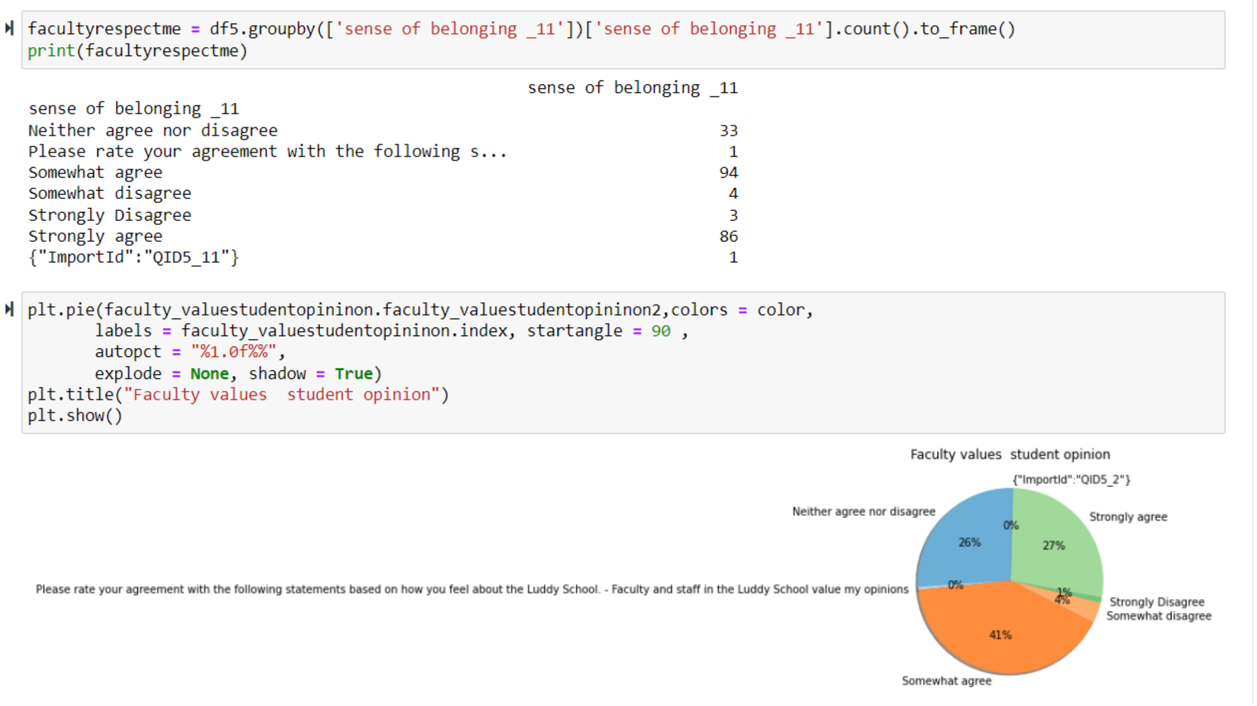
```
'Please rate your agreement with the following statements based on how you feel about the Luddy School. - I am treated with
as much respect as other students.'
```

```
treatedwithsamerespect.plot.pie(figsize=(9,9),subplots=True)
```

```
array([<AxesSubplot:ylabel='sense of belonging _8'>], dtype=object)
```

Somewhat agree
Please rate your agreement with the following statements based on how you feel about the Luddy School. - I am treated with as much respect as other students.
Neither agree nor disagree
{"ImportId":"QID5_8"}
sense of belonging _8
Somewhat disagree
Strongly Disagree
Strongly agree
Neither agree nor disagree
Please rate your agreement with the following statements based on how you feel about the Luddy School. - I am treated with as much respect as other students.
Somewhat agree
Somewhat disagree
Strongly Disagree
Strongly agree
{"ImportId":"QID5_8"}

Sense of belonging_11 “ Faculty valuing Student opinion ”



```
facultyrespectme = df5.groupby(['sense of belonging _11'])['sense of belonging _11'].count().to_frame()
print(facultyrespectme)
```

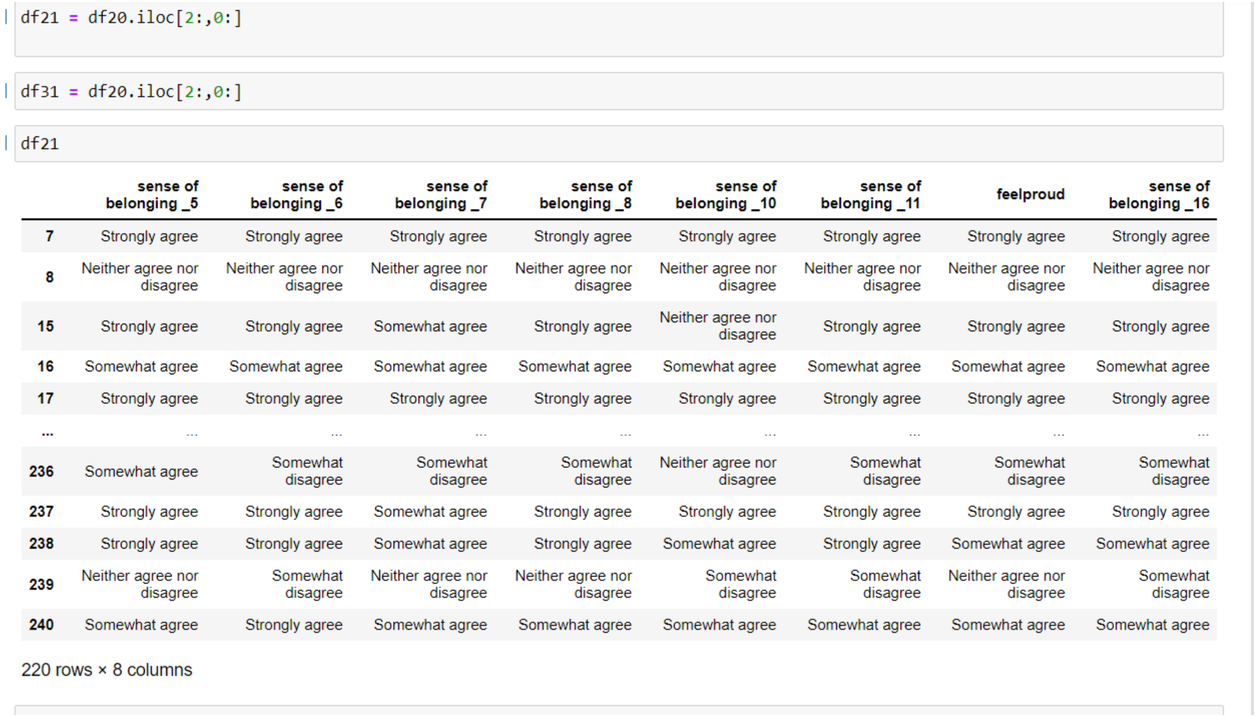
```
                                                    sense of belonging _11
sense of belonging _11
Neither agree nor disagree                                              33
Please rate your agreement with the following s...                       1
Somewhat agree                                                          94
Somewhat disagree                                                        4
Strongly Disagree                                                        3
Strongly agree                                                          86
{"ImportId":"QID5_11"}                                                   1
```

```
plt.pie(faculty_valuestudentopininon.faculty_valuestudentopininon2,colors = color,
       labels = faculty_valuestudentopininon.index, startangle = 90 ,
       autopct = "%1.0f%%",
       explode = None, shadow = True)
plt.title("Faculty values  student opinion")
plt.show()
```

Faculty values student opinion
{"ImportId":"QID5_2"}
Neither agree nor disagree
Strongly agree
0%
26%
27%
Please rate your agreement with the following statements based on how you feel about the Luddy School. - Faculty and staff in the Luddy School value my opinions
0%
1%
4%
Strongly Disagree
Somewhat disagree
41%
Somewhat agree

From this Pie Chart we can conclude that most of the staff and faculty of luddy school listen’s the students opinion and also value the students as we can see statistics that almost 80% student are selecting option somewhat agree and strongly agree so from this it is clear that faculty and staff of luddy values student opinion.

Hierarchical clustering



```
df21 = df20.iloc[2:,0:]
```

```
df31 = df20.iloc[2:,0:]
```

```
df21
```

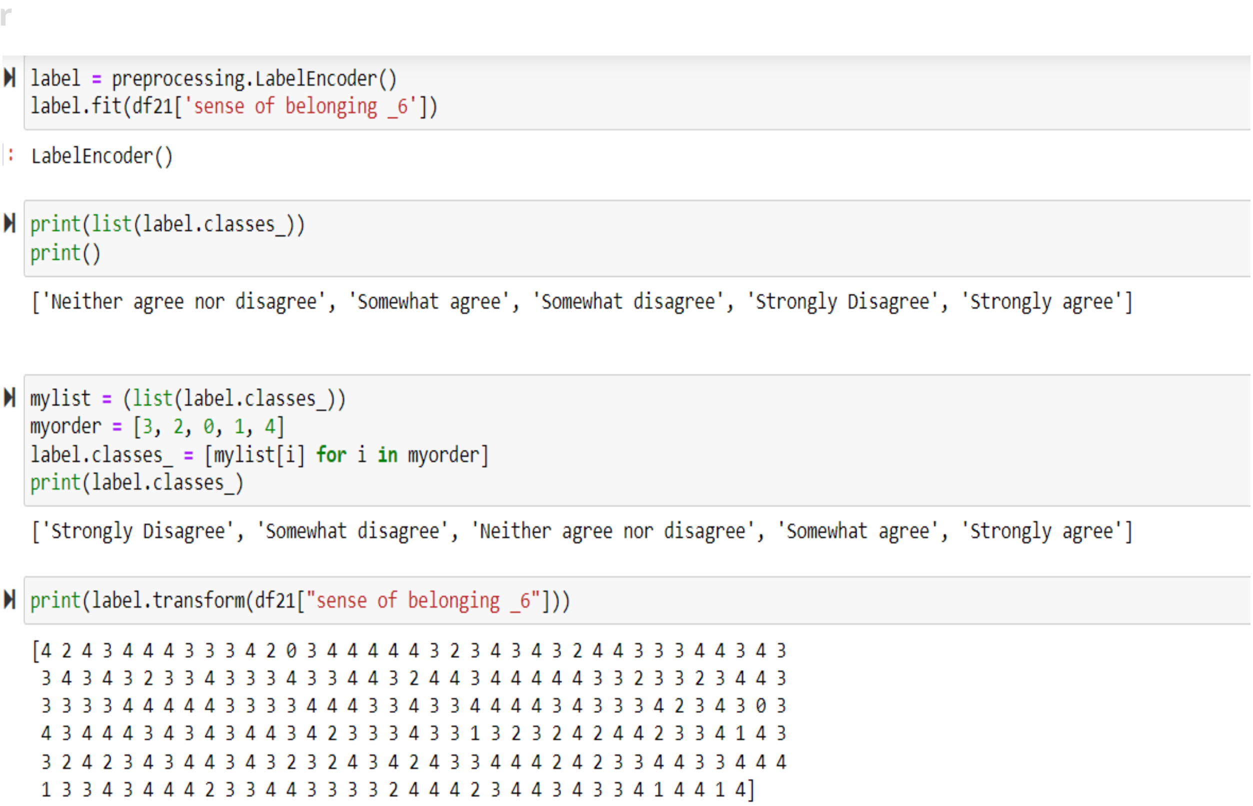
	sense of belonging _5	sense of belonging _6	sense of belonging _7	sense of belonging _8	sense of belonging _10	sense of belonging _11	feelproud	sense of belonging _16
7	Strongly agree	Strongly agree	Strongly agree	Strongly agree	Strongly agree	Strongly agree	Strongly agree	Strongly agree
8	Neither agree nor disagree	Neither agree nor disagree	Neither agree nor disagree	Neither agree nor disagree	Neither agree nor disagree	Neither agree nor disagree	Neither agree nor disagree	Neither agree nor disagree
15	Strongly agree	Strongly agree	Somewhat agree	Strongly agree	Neither agree nor disagree	Strongly agree	Strongly agree	Strongly agree
16	Somewhat agree	Somewhat agree	Somewhat agree	Somewhat agree	Somewhat agree	Somewhat agree	Somewhat agree	Somewhat agree
17	Strongly agree	Strongly agree	Strongly agree	Strongly agree	Strongly agree	Strongly agree	Strongly agree	Strongly agree
...	...	...	...	...	...	...	...	...
236	Somewhat agree	Somewhat disagree	Somewhat disagree	Somewhat disagree	Neither agree nor disagree	Somewhat disagree	Somewhat disagree	Somewhat disagree
237	Strongly agree	Strongly agree	Somewhat agree	Strongly agree	Strongly agree	Strongly agree	Strongly agree	Strongly agree
238	Strongly agree	Strongly agree	Somewhat agree	Strongly agree	Somewhat agree	Strongly agree	Somewhat agree	Somewhat agree
239	Neither agree nor disagree	Somewhat disagree	Neither agree nor disagree	Neither agree nor disagree	Somewhat disagree	Somewhat disagree	Neither agree nor disagree	Somewhat disagree
240	Somewhat agree	Strongly agree	Somewhat agree	Somewhat agree	Somewhat agree	Somewhat agree	Somewhat agree	Somewhat agree

220 rows × 8 columns

Created a new data frame for doing more detailed analysis in which I selected 8 questions which involved both faculty of luddy and students both. I think from this analysis which is described below we will be able to conclude how much proportion of students are comfortable with the faculty of luddy and how comfortable they are with studying a luddy curriculum.

Now to get more insights I will be using Hierarchical clustering which is an unsupervised learning method for clustering data points. The algorithm builds clusters by measuring the dissimilarities between data. Unsupervised learning means that a model does not have to be trained.we do not need a “target” variable, This method can be used on any data to visualize and interpret the relationship between individual data points.

Here in this project I have used hierarchical clustering to group data points and visualize the clustering using dendrogram and statistics.



```
label = preprocessing.LabelEncoder()
label.fit(df21['sense of belonging _6'])
```

```
LabelEncoder()
```

```
print(list(label.classes_))
print()
```

```
['Neither agree nor disagree', 'Somewhat agree', 'Somewhat disagree', 'Strongly Disagree', 'Strongly agree']
```

```
mylist = (list(label.classes_))
myorder = [3, 2, 0, 1, 4]
label.classes_ = [mylist[i] for i in myorder]
print(label.classes_)
```

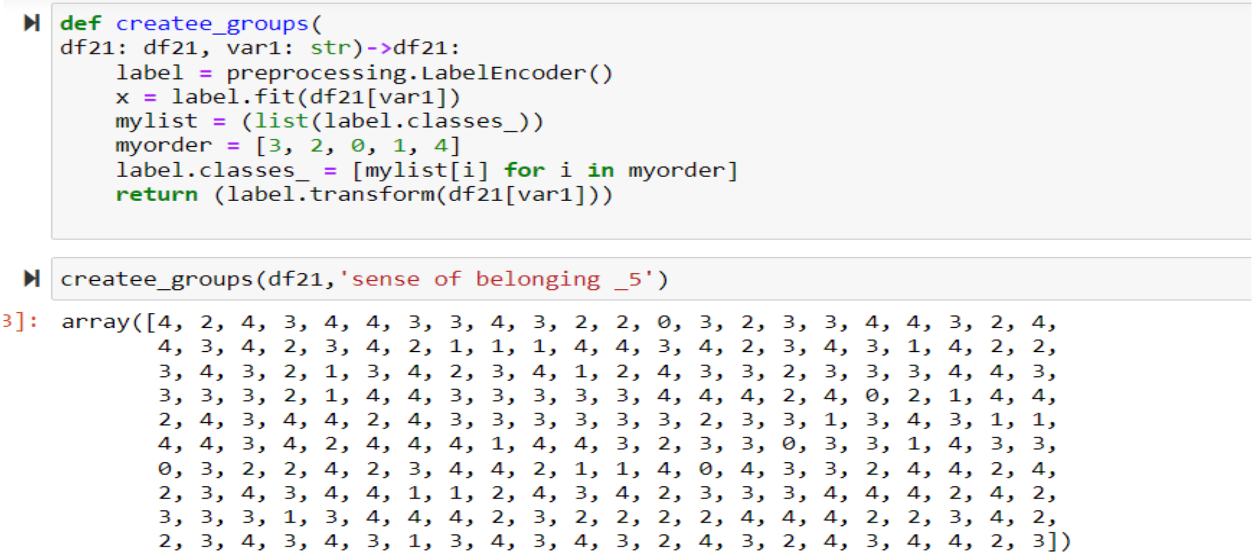
```
['Strongly Disagree', 'Somewhat disagree', 'Neither agree nor disagree', 'Somewhat agree', 'Strongly agree']
```

```
print(label.transform(df21["sense of belonging _6"]))
```

```
[4 2 4 3 4 4 4 3 3 3 4 2 0 3 4 4 4 4 4 3 2 3 4 3 4 3 2 4 4 3 3 3 4 4 3 4 3
 3 4 3 4 3 2 3 3 4 3 3 3 4 3 3 4 4 3 2 4 4 3 4 4 4 4 4 3 3 2 3 3 2 3 4 4 3
 3 3 3 3 4 4 4 4 4 3 3 3 3 4 4 4 3 3 4 3 3 4 4 4 4 3 4 3 3 3 4 2 3 4 3 0 3
 4 3 4 4 4 3 4 3 4 3 4 4 3 4 2 3 3 3 4 3 3 1 3 2 3 2 4 2 4 4 2 3 3 4 1 4 3
 3 2 4 2 3 4 3 4 4 3 4 3 2 3 2 4 3 4 2 4 3 3 4 4 4 2 4 2 3 3 4 4 3 3 4 4 4
 1 3 3 4 3 4 4 4 2 3 3 4 4 3 3 3 3 2 4 4 4 2 3 4 4 3 4 3 3 4 1 4 4 1 4]
```

Now for performing Hierarchical clustering firstly, I had converted all the Answers which are in categories into ordinal values. For converting categories into ordinal values I have used label encoder and reordered the indexes of labels For instance, as it is described in screenshot 0th position is for Strongly Disagree, 1st position is for Somewhat disagree, ,2nd is for Neither agree nor disagree, 3rd is for somewhat disagree , 4th is for Strongly agree.

Now for labeling the categorical values and transforming it into ordinal values I have created one python Createe_groups function in which data will be converted from categorical values into ordinal values. And formed a new data frame which includes all the ordinal values.

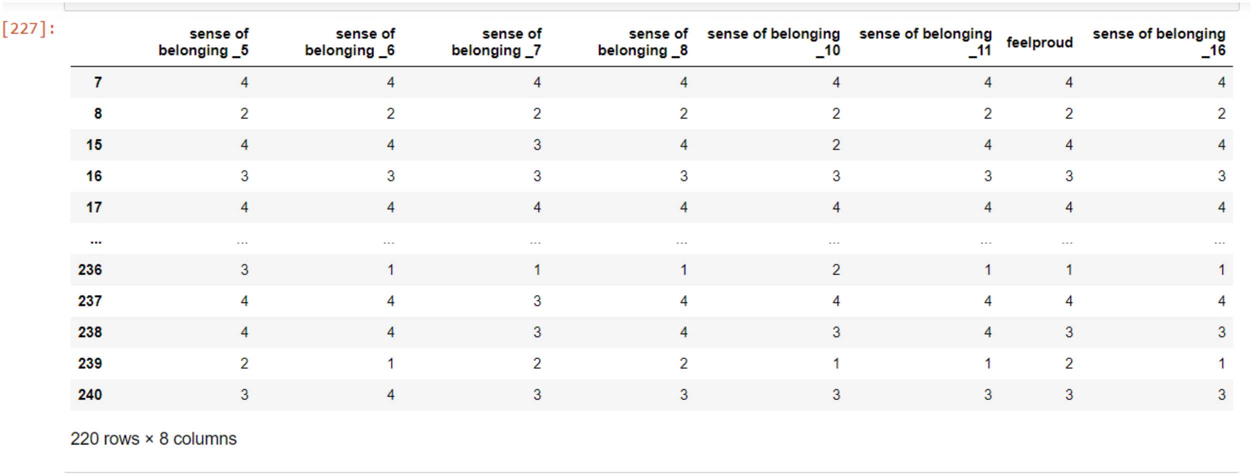


```
def createe_groups(
df21: df21, var1: str)->df21:
    label = preprocessing.LabelEncoder()
    x = label.fit(df21[var1])
    mylist = (list(label.classes_))
    myorder = [3, 2, 0, 1, 4]
    label.classes_ = [mylist[i] for i in myorder]
    return (label.transform(df21[var1]))
```

```
createe_groups(df21,'sense of belonging _5')
```

```
3]: array([4, 2, 4, 3, 4, 4, 3, 3, 4, 3, 2, 2, 0, 3, 2, 3, 3, 4, 4, 3, 2, 4,
           4, 3, 4, 2, 3, 4, 2, 1, 1, 1, 4, 4, 3, 4, 2, 3, 4, 3, 1, 4, 2, 2,
           3, 4, 3, 2, 1, 3, 4, 2, 3, 4, 1, 2, 4, 3, 3, 2, 3, 3, 3, 4, 4, 3,
           3, 3, 3, 2, 1, 4, 4, 3, 3, 3, 3, 3, 4, 4, 4, 2, 4, 0, 2, 1, 4, 4,
           2, 4, 3, 4, 4, 2, 4, 3, 3, 3, 3, 3, 3, 2, 3, 3, 1, 3, 4, 3, 1, 1,
           4, 4, 3, 4, 2, 4, 4, 4, 1, 4, 4, 3, 2, 3, 3, 0, 3, 3, 1, 4, 3, 3,
           0, 3, 2, 2, 4, 2, 3, 4, 4, 2, 1, 1, 4, 0, 4, 3, 3, 2, 4, 4, 2, 4,
           2, 3, 4, 3, 4, 4, 1, 1, 2, 4, 3, 4, 2, 3, 3, 3, 4, 4, 4, 2, 4, 2,
           3, 3, 3, 1, 3, 4, 4, 4, 2, 3, 2, 2, 2, 2, 4, 4, 4, 2, 2, 3, 4, 2,
           2, 3, 4, 3, 4, 3, 1, 3, 4, 3, 4, 3, 2, 4, 3, 2, 4, 3, 4, 4, 2, 3])
```

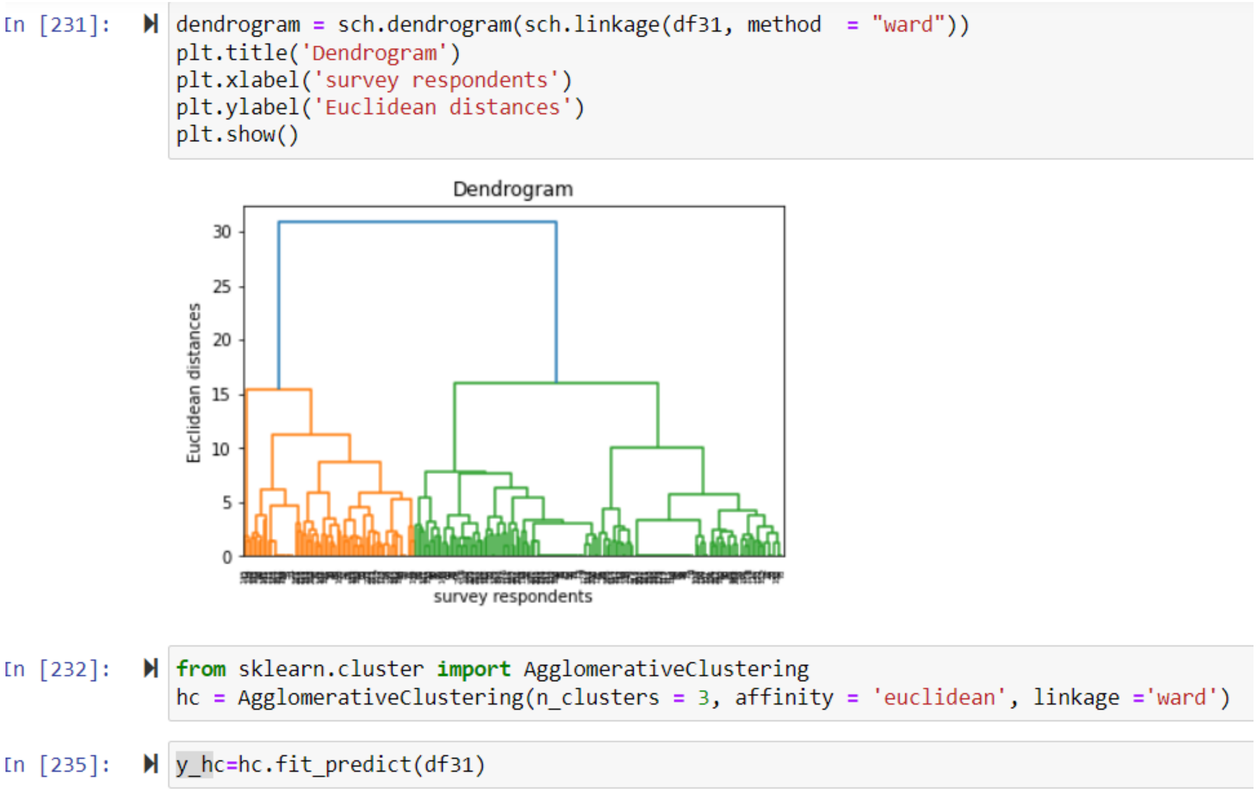
New Dataframe which includes all ordinal values



[227]:

	sense of belonging _5	sense of belonging _6	sense of belonging _7	sense of belonging _8	sense of belonging _10	sense of belonging _11	feelproud	sense of belonging _16
7	4	4	4	4	4	4	4	4
8	2	2	2	2	2	2	2	2
15	4	4	3	4	2	4	4	4
16	3	3	3	3	3	3	3	3
17	4	4	4	4	4	4	4	4
...	...	...	...	...	...	...	...	...
236	3	1	1	1	2	1	1	1
237	4	4	3	4	4	4	4	4
238	4	4	3	4	3	4	3	3
239	2	1	2	2	1	1	2	1
240	3	4	3	3	3	3	3	3

220 rows × 8 columns

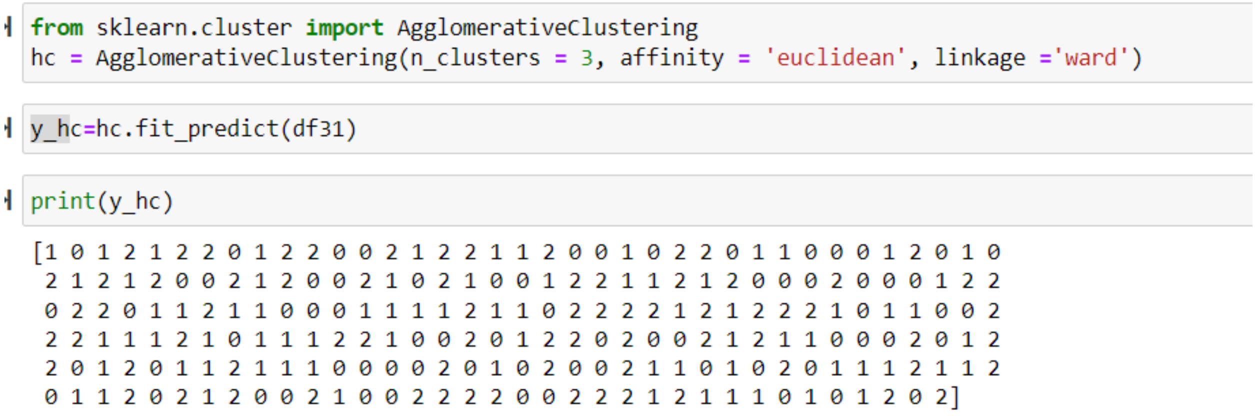


```
In [231]:  dendrogram = sch.dendrogram(sch.linkage(df31, method  = "ward"))
           plt.title('Dendrogram')
           plt.xlabel('survey respondents')
           plt.ylabel('Euclidean distances')
           plt.show()
```

```
In [232]:  from sklearn.cluster import AgglomerativeClustering
           hc = AgglomerativeClustering(n_clusters = 3, affinity = 'euclidean', linkage ='ward')
```

```
In [235]:  y_hc=hc.fit_predict(df31)
```

A dendrogram is **a diagram that shows the hierarchical relationship between objects**. It is most commonly created as an output from hierarchical clustering. The main use of a dendrogram is to work out the best way to allocate objects to clusters.



```
from sklearn.cluster import AgglomerativeClustering
hc = AgglomerativeClustering(n_clusters = 3, affinity = 'euclidean', linkage ='ward')
```

```
y_hc=hc.fit_predict(df31)
```

```
print(y_hc)
```

```
[1 0 1 2 1 2 2 0 1 2 2 0 0 2 1 2 2 1 1 2 0 0 1 0 2 2 0 1 1 0 0 0 1 2 0 1 0
 2 1 2 1 2 0 0 2 1 2 0 0 2 1 0 2 1 0 0 1 2 2 1 1 2 1 2 0 0 0 2 0 0 0 1 2 2
 0 2 2 0 1 1 2 1 1 0 0 0 1 1 1 1 2 1 1 0 2 2 2 2 1 2 1 2 2 2 1 0 1 1 0 0 2
 2 2 1 1 1 2 1 0 1 1 1 2 2 1 0 0 2 0 1 2 2 0 2 0 0 2 1 2 1 1 0 0 0 2 0 1 2
 2 0 1 2 0 1 1 2 1 1 1 0 0 0 0 2 0 1 0 2 0 0 2 1 1 0 1 0 2 0 1 1 1 2 1 1 2
 0 1 1 2 0 2 1 2 0 0 2 1 0 0 2 2 2 2 0 0 2 2 2 1 2 1 1 1 0 1 0 1 2 0 2]
```

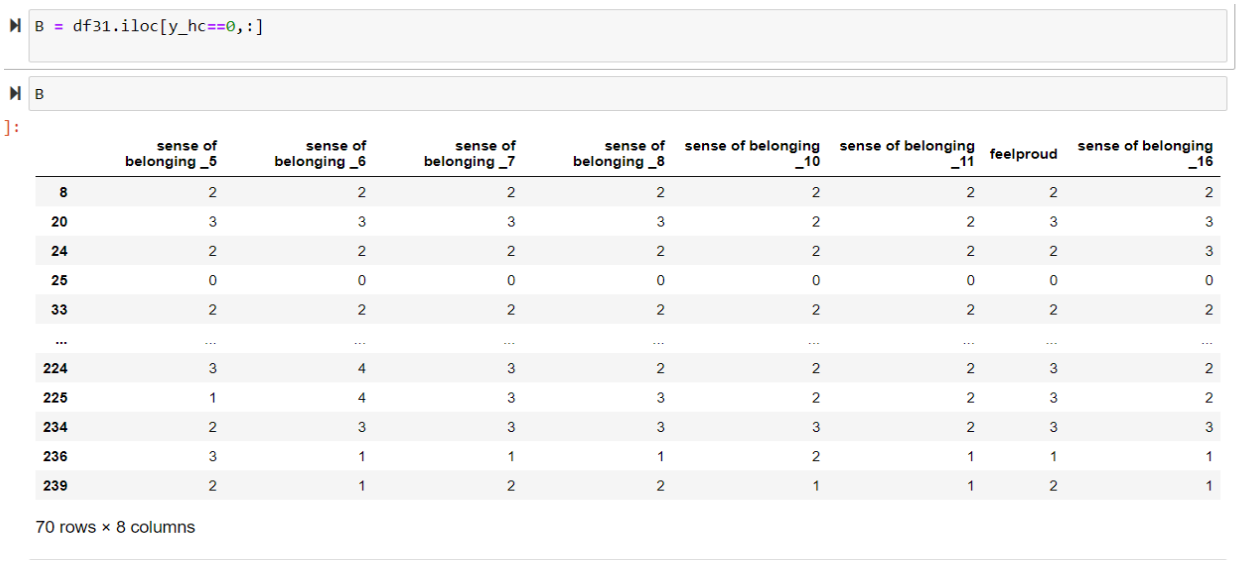
Now by observing the dendrogram, I have decided to divide clusters into Three categories, So for making clusters I have used sklearn library and also used euclidean and wark function in it for getting 3 clusters.

- ## 3 Clusters

B - For value Y_hc = = 0

C - For value Y_hc == 1

D - For value Y_hc == 2

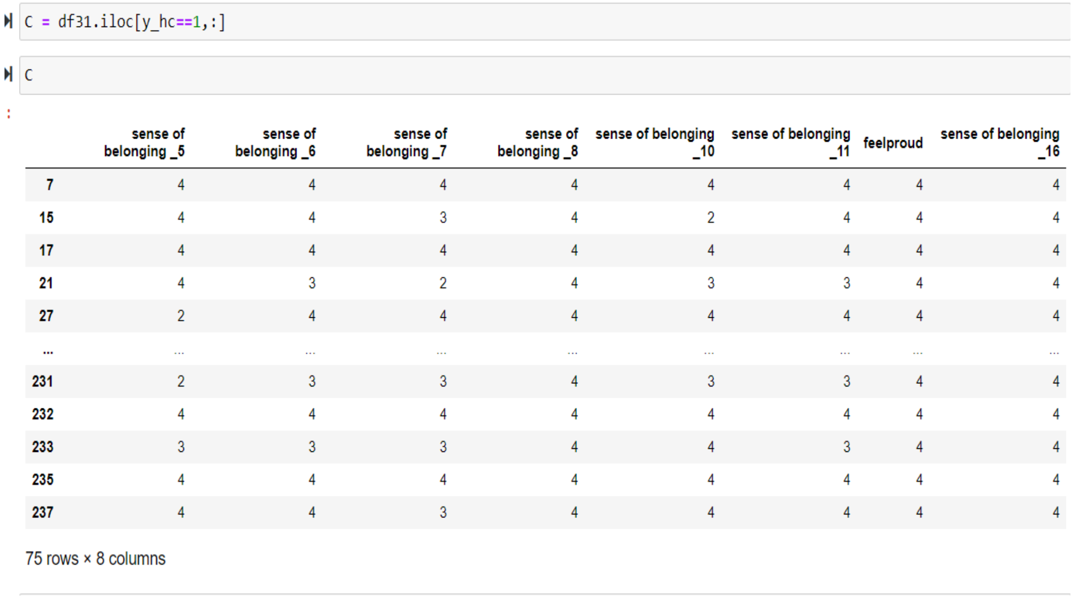


```
B = df31.iloc[y_hc==0,:]
```

```
B
```

	sense of belonging _5	sense of belonging _6	sense of belonging _7	sense of belonging _8	sense of belonging _10	sense of belonging _11	feelproud	sense of belonging _16
8	2	2	2	2	2	2	2	2
20	3	3	3	3	2	2	3	3
24	2	2	2	2	2	2	2	3
25	0	0	0	0	0	0	0	0
33	2	2	2	2	2	2	2	2
...	...	...	...	...	...	...	...	...
224	3	4	3	2	2	2	3	2
225	1	4	3	3	2	2	3	2
234	2	3	3	3	3	2	3	3
236	3	1	1	1	2	1	1	1
239	2	1	2	2	1	1	2	1

70 rows × 8 columns

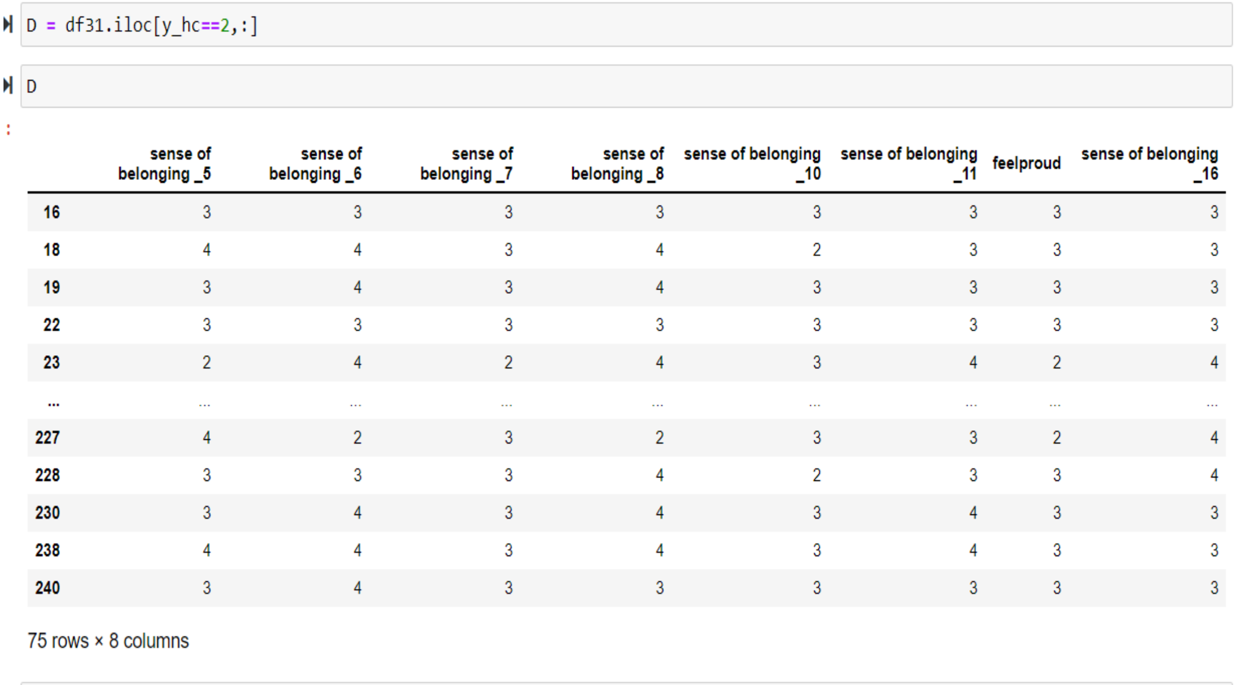


```
C = df31.iloc[y_hc==1,:]
```

```
C
```

	sense of belonging _5	sense of belonging _6	sense of belonging _7	sense of belonging _8	sense of belonging _10	sense of belonging _11	feelproud	sense of belonging _16
7	4	4	4	4	4	4	4	4
15	4	4	3	4	2	4	4	4
17	4	4	4	4	4	4	4	4
21	4	3	2	4	3	3	4	4
27	2	4	4	4	4	4	4	4
...	...	...	...	...	...	...	...	...
231	2	3	3	4	3	3	4	4
232	4	4	4	4	4	4	4	4
233	3	3	3	4	4	3	4	4
235	4	4	4	4	4	4	4	4
237	4	4	3	4	4	4	4	4

75 rows × 8 columns



```
D = df31.iloc[y_hc==2,:]
```

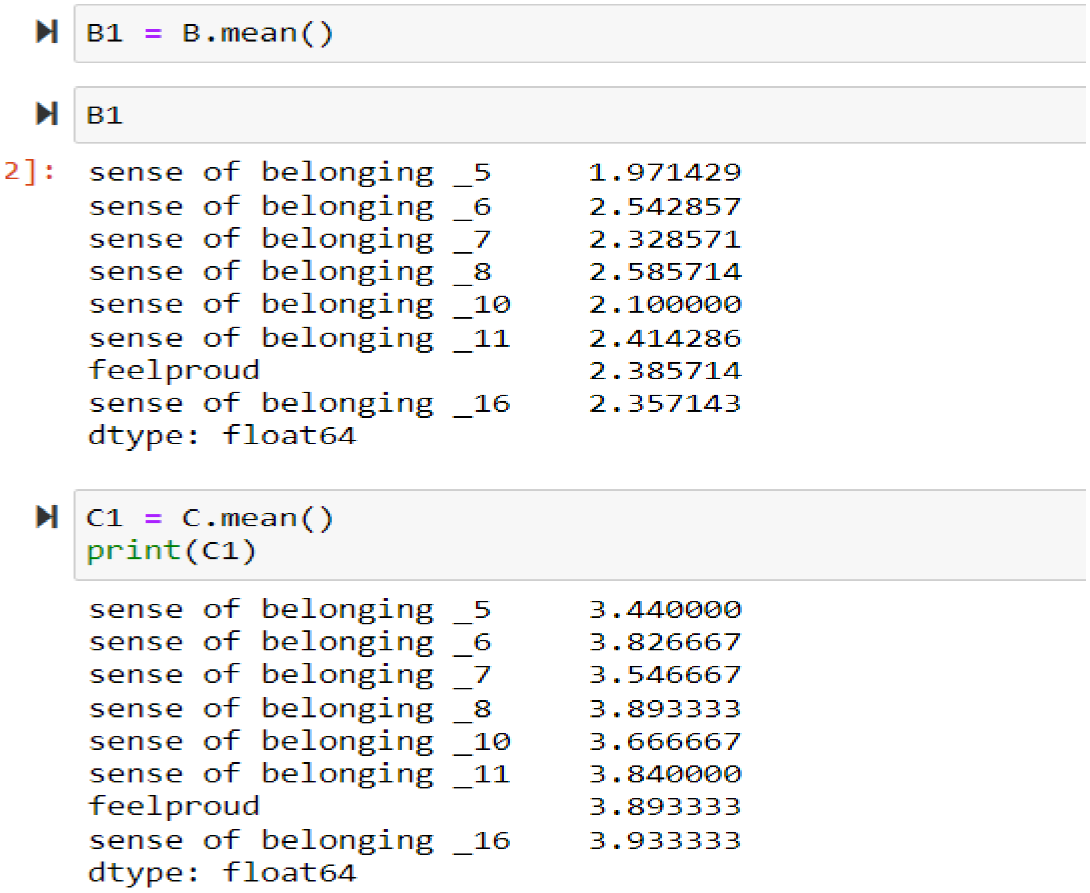
```
D
```

	sense of belonging _5	sense of belonging _6	sense of belonging _7	sense of belonging _8	sense of belonging _10	sense of belonging _11	feelproud	sense of belonging _16
16	3	3	3	3	3	3	3	3
18	4	4	3	4	2	3	3	3
19	3	4	3	4	3	3	3	3
22	3	3	3	3	3	3	3	3
23	2	4	2	4	3	4	2	4
...	...	...	...	...	...	...	...	...
227	4	2	3	2	3	3	2	4
228	3	3	3	4	2	3	3	4
230	3	4	3	4	3	4	3	3
238	4	4	3	4	3	4	3	3
240	3	4	3	3	3	3	3	3

75 rows × 8 columns

After making a new Dataframe which consists of cordial values, I decided to form 3 different clusters for pattern recognition. From the Hierarchical clustering model , from Dendrogram figure I understood that I should have 3 different clusters to recognize patterns.

From the 3 different clusters, I found out the mean for each of the columns in each of 3 clusters. I found out that in the B cluster there are 70 students who selected   options like strongly Disagree, somewhat Disagree , Neither agree nor disagree as the mean is around 2. At the same time, in C cluster there are 75 students who are comfortable with the courses and faculty of luddy as majority of the student selected options like somewhat agree, strongly disagree as the mean is almost around 3.5. However, in the D cluster there are 75 students who have mixed opinions regarding faculty and courses as most of the students have selected options like neither agree nor disagree, somewhat agree, somewhat disagree. So from this It can be concluded that only 75 students are completely optimistic about the courses and faculty of Luddy,



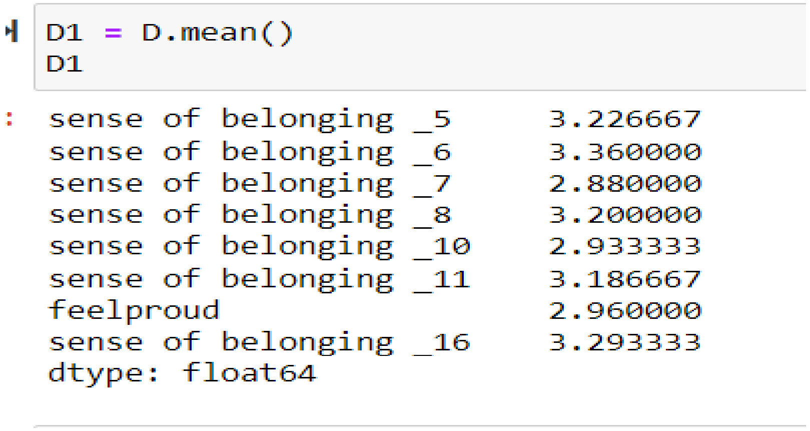
```
B1 = B.mean()
```

```
B1
```

```
2]: sense of belonging _5     1.971429
    sense of belonging _6     2.542857
    sense of belonging _7     2.328571
    sense of belonging _8     2.585714
    sense of belonging _10    2.100000
    sense of belonging _11    2.414286
    feelproud                 2.385714
    sense of belonging _16    2.357143
    dtype: float64
```

```
C1 = C.mean()
print(C1)
```

```
sense of belonging _5     3.440000
sense of belonging _6     3.826667
sense of belonging _7     3.546667
sense of belonging _8     3.893333
sense of belonging _10    3.666667
sense of belonging _11    3.840000
feelproud                 3.893333
sense of belonging _16    3.933333
dtype: float64
```



```
D1 = D.mean()
D1
```

```
: sense of belonging _5     3.226667
  sense of belonging _6     3.360000
  sense of belonging _7     2.880000
  sense of belonging _8     3.200000
  sense of belonging _10    2.933333
  sense of belonging _11    3.186667
  feelproud                 2.960000
  sense of belonging _16    3.293333
  dtype: float64
```