Assignment03

March 6, 2022

1 Assignment 3

1.1 Pandas and plotting exercises

[1]:
```
# Import the pandas library
import pandas as pd
```

In Week 2, you used a dataset from the CORGIS website. You may have used either the Python, CSV, or JSON data files.

For this assignment, use the CSV file format for the same category of data that you used previously.

[5]:
```
# Use pandas read_csv function to import the data into a dataframe variable
import graduates
df = pd.read_csv('graduates.csv')
```

[8]:
```
# How many rows and columns does the dataframe have?
df.shape
#517 rows and 50 columns
```

[8]:
```
(517, 50)
```

[7]:
```
# What are the column names of the dataframe?

df.columns
```

[7]:
```
Index(['Year', 'Demographics.Total', 'Education.Major', 'Salaries.Highest',
       'Salaries.Lowest', 'Salaries.Mean', 'Salaries.Median',
       'Salaries.Quantity', 'Salaries.Standard Deviation',
       'Demographics.Ethnicity.Asians', 'Demographics.Ethnicity.Minorities',
       'Demographics.Ethnicity.Whites', 'Demographics.Gender.Females',
       'Demographics.Gender.Males', 'Education.Degrees.Bachelors',
       'Education.Degrees.Doctorates', 'Education.Degrees.Masters',
       'Education.Degrees.Professionals',
       'Employment.Employer Type.Business/Industry',
       'Employment.Employer Type.Educational Institution',
       'Employment.Employer Type.Government',
       'Employment.Reason Working Outside Field.Career Change',
       'Employment.Reason Working Outside Field.Family-related',
```

```
       'Employment.Reason Working Outside Field.Job Location',
       'Employment.Reason Working Outside Field.No Job Available',
       'Employment.Reason Working Outside Field.Other',
       'Employment.Reason Working Outside Field.Pay/Promotion',
       'Employment.Reason Working Outside Field.Working Conditions',
       'Employment.Reason for Not Working.Family',
       'Employment.Reason for Not Working.Layoff',
       'Employment.Reason for Not Working.No Job Available',
       'Employment.Reason for Not Working.No need/want',
       'Employment.Reason for Not Working.Student',
       'Employment.Status.Employed', 'Employment.Status.Not in Labor Force',
       'Employment.Status.Unemployed',
       'Employment.Work Activity.Accounting/Finance/Contracts',
       'Employment.Work Activity.Applied Research',
       'Employment.Work Activity.Basic Research',
       'Employment.Work Activity.Computer Applications',
       'Employment.Work Activity.Design',
       'Employment.Work Activity.Development',
       'Employment.Work Activity.Human Resources',
       'Employment.Work Activity.Managing/Supervising People/Projects',
       'Employment.Work Activity.Other',
       'Employment.Work Activity.Productions/Operations/Maintenance',
       'Employment.Work Activity.Professional Service',
       'Employment.Work Activity.Qualitity/Productivity Management',
       'Employment.Work Activity.Sales, Purchasing, Marketing',
       'Employment.Work Activity.Teaching'],
      dtype='object')
```

[9]:
```
# What are the datatypes of each column?

df.dtypes
```

[9]:
```
Year                                                    int64
Demographics.Total                                      int64
Education.Major                                        object
Salaries.Highest                                      float64
Salaries.Lowest                                       float64
Salaries.Mean                                         float64
Salaries.Median                                       float64
Salaries.Quantity                                       int64
Salaries.Standard Deviation                           float64
Demographics.Ethnicity.Asians                           int64
Demographics.Ethnicity.Minorities                       int64
Demographics.Ethnicity.Whites                           int64
Demographics.Gender.Females                             int64
Demographics.Gender.Males                               int64
Education.Degrees.Bachelors                             int64
```

```
Education.Degrees.Doctorates                                    int64
Education.Degrees.Masters                                       int64
Education.Degrees.Professionals                                 int64
Employment.Employer Type.Business/Industry                      int64
Employment.Employer Type.Educational Institution                int64
Employment.Employer Type.Government                             int64
Employment.Reason Working Outside Field.Career Change           int64
Employment.Reason Working Outside Field.Family-related          int64
Employment.Reason Working Outside Field.Job Location            int64
Employment.Reason Working Outside Field.No Job Available        int64
Employment.Reason Working Outside Field.Other                   int64
Employment.Reason Working Outside Field.Pay/Promotion           int64
Employment.Reason Working Outside Field.Working Conditions      int64
Employment.Reason for Not Working.Family                        int64
Employment.Reason for Not Working.Layoff                        int64
Employment.Reason for Not Working.No Job Available              int64
Employment.Reason for Not Working.No need/want                  int64
Employment.Reason for Not Working.Student                       int64
Employment.Status.Employed                                      int64
Employment.Status.Not in Labor Force                            int64
Employment.Status.Unemployed                                    int64
Employment.Work Activity.Accounting/Finance/Contracts           int64
Employment.Work Activity.Applied Research                       int64
Employment.Work Activity.Basic Research                         int64
Employment.Work Activity.Computer Applications                  int64
Employment.Work Activity.Design                                 int64
Employment.Work Activity.Development                            int64
Employment.Work Activity.Human Resources                        int64
Employment.Work Activity.Managing/Supervising People/Projects   int64
Employment.Work Activity.Other                                  int64
Employment.Work Activity.Productions/Operations/Maintenance     int64
Employment.Work Activity.Professional Service                   int64
Employment.Work Activity.Qualitity/Productivity Management      int64
Employment.Work Activity.Sales, Purchasing, Marketing           int64
Employment.Work Activity.Teaching                               int64
dtype: object
```

[10]:
```
# Look at the first 2 rows of the dataframe

df.head(2)
```

[10]:

	Year	Demographics.Total	Education.Major	Salaries.Highest	\
0	1993	1295598	Biological Sciences	999999.0	
1	1993	211875	Chemical Engineering	999999.0	

	Salaries.Lowest	Salaries.Mean	Salaries.Median	Salaries.Quantity	\
0	0.0	160585.73	51000.0	13432	

```
1           9000.0      126176.52          56000.0                3375

   Salaries.Standard Deviation  Demographics.Ethnicity.Asians  …  \
0                    297818.25                          84495  …
1                    245705.77                          27531  …

   Employment.Work Activity.Design  Employment.Work Activity.Development  \
0                           118772                                191867
1                            82344                                 76108

   Employment.Work Activity.Human Resources  \
0                                    365049
1                                     59299

   Employment.Work Activity.Managing/Supervising People/Projects  \
0                                             539430
1                                             102248

   Employment.Work Activity.Other  \
0                           99749
1                           16361

   Employment.Work Activity.Productions/Operations/Maintenance  \
0                                             103385
1                                              30480

   Employment.Work Activity.Professional Service  \
0                                         506252
1                                          24690

   Employment.Work Activity.Qualitity/Productivity Management  \
0                                             269042
1                                              63895

   Employment.Work Activity.Sales, Purchasing, Marketing  \
0                                             215169
1                                              44780

   Employment.Work Activity.Teaching
0                             381908
1                              17718

[2 rows x 50 columns]
```

[11]:
```
# Look at the last 2 rows of the dataframe

df.tail(2)
```

[11]:

	Year	Demographics.Total	Education.Major	\
515	2015	0	Management & Administration	
516	2015	0	Political and related sciences	

	Salaries.Highest	Salaries.Lowest	Salaries.Mean	Salaries.Median	\
515	0.0	0.0	0.0	0.0	
516	0.0	0.0	0.0	0.0	

	Salaries.Quantity	Salaries.Standard Deviation	\
515	0	0.0	
516	0	0.0	

	Demographics.Ethnicity.Asians	…	Employment.Work Activity.Design	\
515	0	…	0	
516	0	…	0	

	Employment.Work Activity.Development	\
515	0	
516	0	

	Employment.Work Activity.Human Resources	\
515	0	
516	0	

	Employment.Work Activity.Managing/Supervising People/Projects	\
515	0	
516	0	

	Employment.Work Activity.Other	\
515	0	
516	0	

	Employment.Work Activity.Productions/Operations/Maintenance	\
515	0	
516	0	

	Employment.Work Activity.Professional Service	\
515	0	
516	0	

	Employment.Work Activity.Qualitity/Productivity Management	\
515	0	
516	0	

	Employment.Work Activity.Sales, Purchasing, Marketing	\
515	0	
516	0	

```
     Employment.Work Activity.Teaching
515                                  0
516                                  0

[2 rows x 50 columns]
```

[12]:
```
# Print out summary statistics about the dataframe

df.info
```

[12]:
```
<bound method DataFrame.info of      Year  Demographics.Total
Education.Major  \
0    1993             1295598             Biological Sciences
1    1993              211875           Chemical Engineering
2    1993              507616                      Chemistry
3    1993              336366             Civil Engineering
4    1993             1070111      Computer Science and Math
..    …                  …                             …
512  2015             1176525                     Sociology
513  2015               55738                    Statistics
514  2015              169991              Zoology, General
515  2015                   0    Management & Administration
516  2015                   0  Political and related sciences

     Salaries.Highest  Salaries.Lowest  Salaries.Mean  Salaries.Median  \
0            999999.0              0.0      160585.73          51000.0
1            999999.0           9000.0      126176.52          56000.0
2            999999.0           8000.0      148872.00          60000.0
3            999999.0          10000.0      129070.55          50000.0
4            999999.0              0.0      134299.53          49000.0
..                …                …             …                …
512         1223166.0              0.0       58871.70          50000.0
513         1038725.0              0.0       99210.87          88000.0
514         1223166.0              0.0       86957.98          60000.0
515               0.0              0.0           0.00              0.0
516               0.0              0.0           0.00              0.0

     Salaries.Quantity  Salaries.Standard Deviation  \
0                13432                    297818.25
1                 3375                    245705.77
2                 7834                    276000.33
3                 4035                    259543.49
4                 9996                    269323.82
..                 …                           …
512               1798                     62083.18
513                157                     90534.66
```

```
514                317                   124675.05
515                  0                        0.00
516                  0                        0.00

     Demographics.Ethnicity.Asians  …  Employment.Work Activity.Design  \
0                            84495  …                           118772
1                            27531  …                            82344
2                            49984  …                            81772
3                            37295  …                           133430
4                            83826  …                           251941
..                             …  …                              …
512                          59244  …                            85782
513                          16211  …                            13908
514                          17563  …                            15366
515                              0  …                                0
516                              0  …                                0

     Employment.Work Activity.Development  \
0                                  191867
1                                   76108
2                                  123256
3                                   62031
4                                  200490
..                                    …
512                                151538
513                                 11798
514                                 30180
515                                     0
516                                     0

     Employment.Work Activity.Human Resources  \
0                                      365049
1                                       59299
2                                      121783
3                                      108338
4                                      287405
..                                        …
512                                    261143
513                                      8419
514                                     29873
515                                         0
516                                         0

     Employment.Work Activity.Managing/Supervising People/Projects  \
0                                                   539430
1                                                   102248
2                                                   208278
```

```
3                                                 203035
4                                                 439446
..                                                   …
512                                               415552
513                                                19960
514                                                60016
515                                                    0
516                                                    0

     Employment.Work Activity.Other  \
0                              99749
1                              16361
2                              35007
3                              37940
4                              62482
..                               …
512                           204818
513                            3437
514                           29924
515                               0
516                               0

     Employment.Work Activity.Productions/Operations/Maintenance  \
0                                                 103385
1                                                  30480
2                                                  40898
3                                                  28639
4                                                  73411
..                                                   …
512                                                72050
513                                                 5489
514                                                24622
515                                                    0
516                                                    0

     Employment.Work Activity.Professional Service  \
0                                             506252
1                                              24690
2                                             129716
3                                              70727
4                                             126083
..                                               …
512                                           342903
513                                            12803
514                                            57404
515                                                0
516                                                0
```

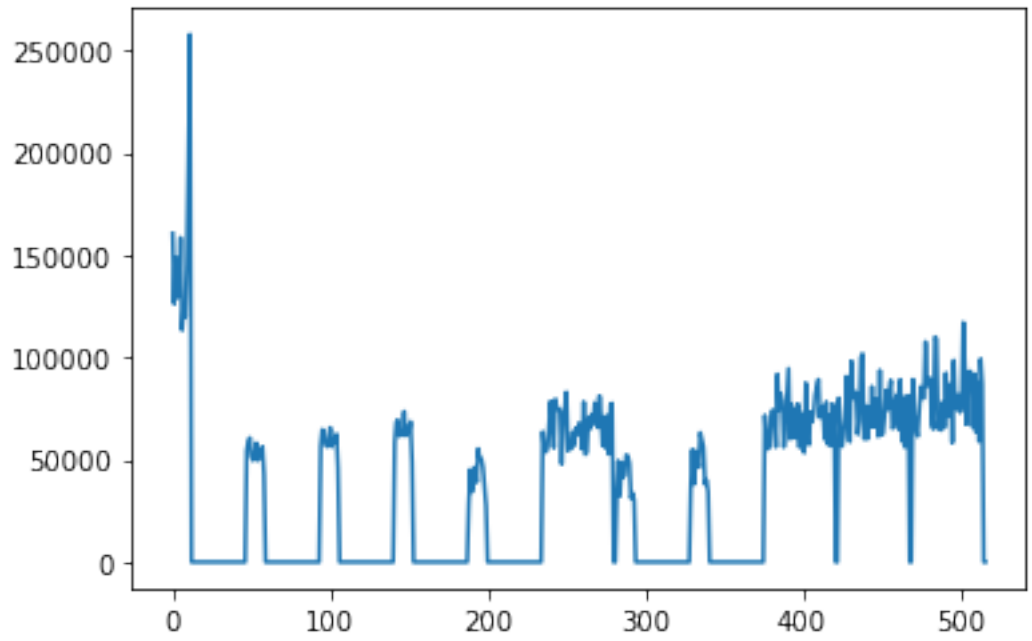
```
     Employment.Work Activity.Qualitity/Productivity Management  \
0                                                      269042
1                                                       63895
2                                                      114801
3                                                      116758
4                                                      223467
..                                                        ...
512                                                    183692
513                                                      7736
514                                                     30979
515                                                         0
516                                                         0

     Employment.Work Activity.Sales, Purchasing, Marketing  \
0                                                    215169
1                                                     44780
2                                                     78059
3                                                     73133
4                                                    168404
..                                                      ...
512                                                  345302
513                                                   10741
514                                                   35810
515                                                       0
516                                                       0

     Employment.Work Activity.Teaching
0                               381908
1                                17718
2                               104191
3                                26892
4                               234507
..                                 ...
512                             301924
513                               6290
514                              56660
515                                  0
516                                  0

[517 rows x 50 columns]>
```

[13]:
```
# Choose a column and print out the column (it's ok if the output is␣
↪abbreviated)

df['Education.Major']
```

[13]:
```
0                  Biological Sciences
1                Chemical Engineering
2                           Chemistry
3                  Civil Engineering
4           Computer Science and Math
                      …
512                        Sociology
513                       Statistics
514                  Zoology, General
515       Management & Administration
516    Political and related sciences
Name: Education.Major, Length: 517, dtype: object
```

[14]:
```
# Choose a column that has numeric values and make a line plot of the values

a= df['Salaries.Mean']

a.plot()
```

[14]:
```
<AxesSubplot:>
```



[19]:
```
# Use "loc" to print out the first 10 elements of the plotted column

a.loc[0:10]
```

[19]:
```
0     160585.73
1     126176.52
2     148872.00
3     129070.55
4     134299.53
5     158542.76
6     113262.51
7     124761.62
8     119635.90
9     145709.93
10    195036.54
Name: Salaries.Mean, dtype: float64
```

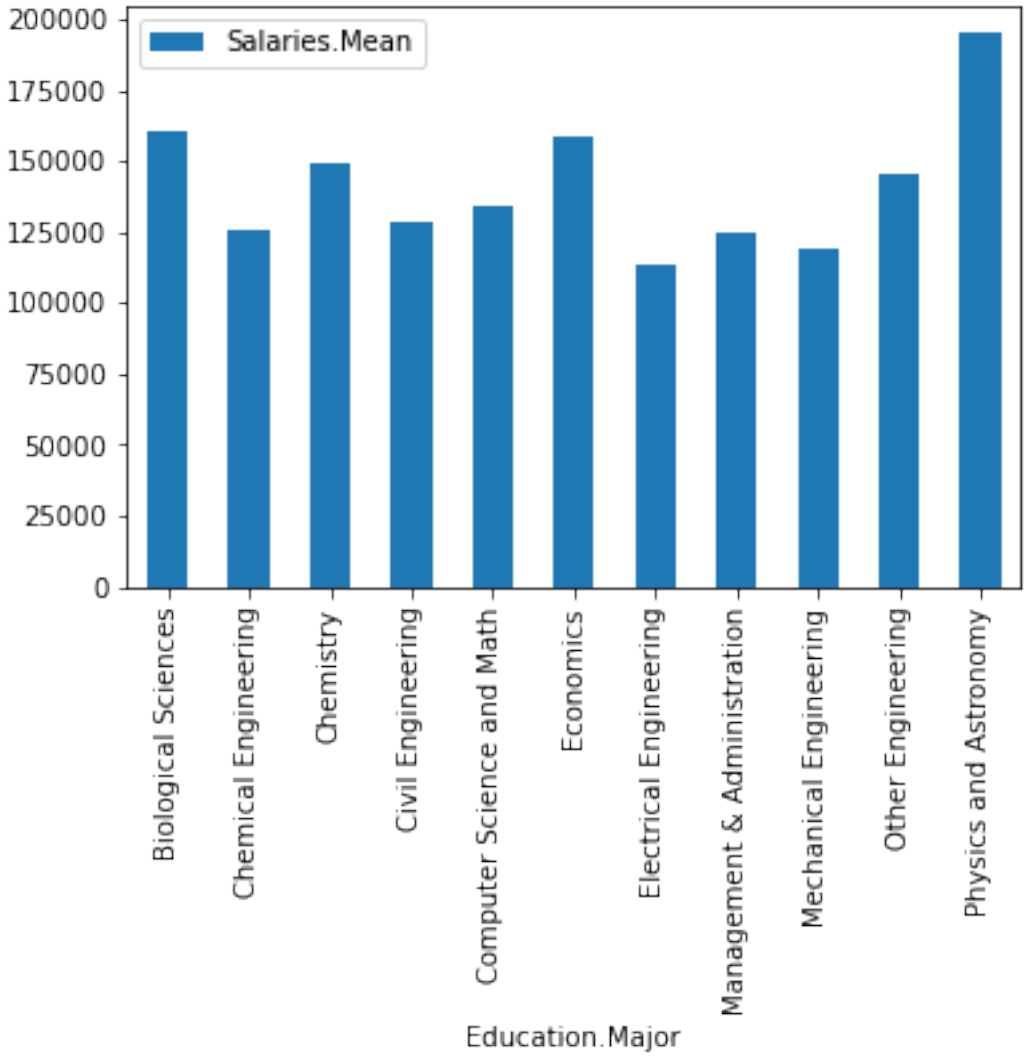
[22]:
```
# Use "loc" to print out the first 10 elements of the plotted column
# as well as the matching 10 elements of a different column that has␣
 ↪interesting text

b = df [['Education.Major', 'Salaries.Mean']]

b.loc[0:10]
```

[22]:

	Education.Major	Salaries.Mean
0	Biological Sciences	160585.73
1	Chemical Engineering	126176.52
2	Chemistry	148872.00
3	Civil Engineering	129070.55
4	Computer Science and Math	134299.53
5	Economics	158542.76
6	Electrical Engineering	113262.51
7	Management & Administration	124761.62
8	Mechanical Engineering	119635.90
9	Other Engineering	145709.93
10	Physics and Astronomy	195036.54

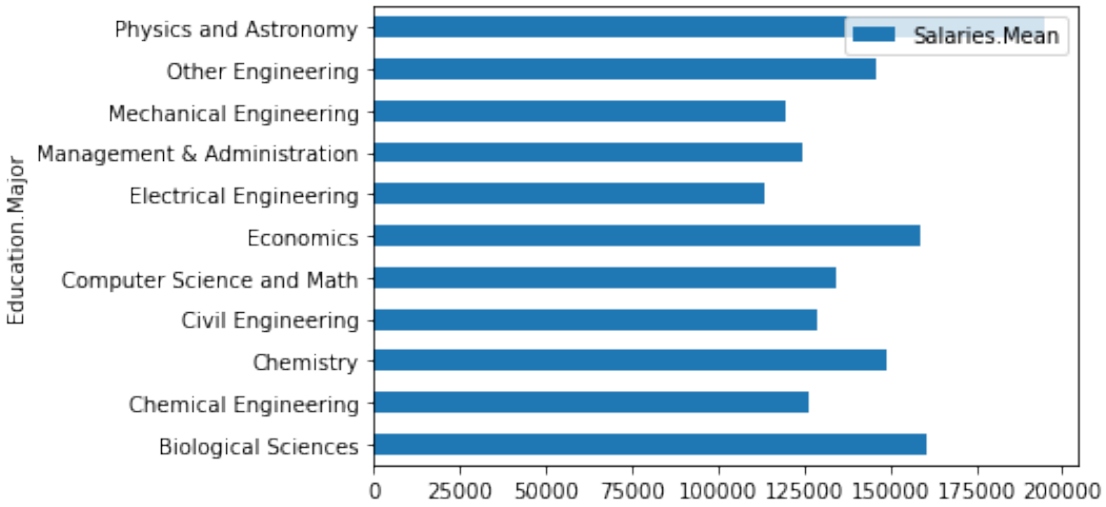
[25]:
```
# Assign the dataframe values from the previous cell into a new dataframe␣
 ↪variable
# and make a bar plot with the text values horizontally and the numeric values␣
 ↪as the bar heights

bPlot = b.loc[0:10]

bPlot.plot(kind = 'bar', x = 'Education.Major' )
```

[25]:
```
<AxesSubplot:xlabel='Education.Major'>
```



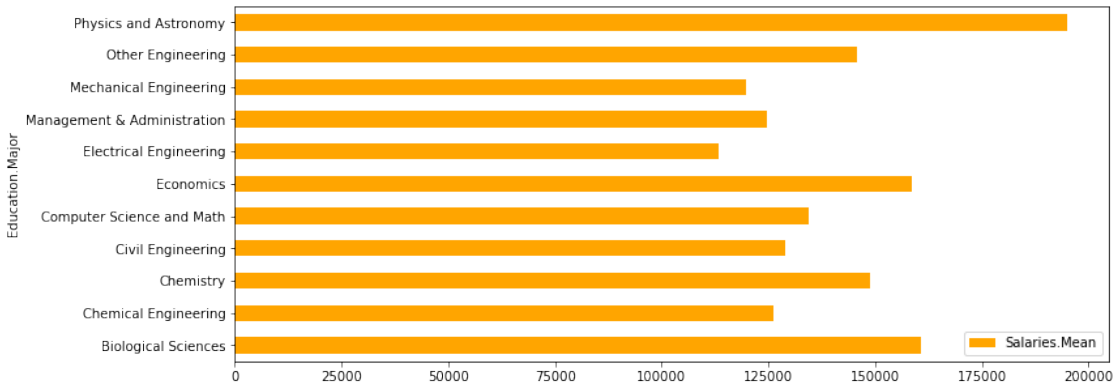
[27]:
```
# Re-do the plot from the previous cell as a horizontal bar plot

bPlot.plot(kind = 'barh', x = 'Education.Major' )
```

[27]: <AxesSubplot:ylabel='Education.Major'>



[29]:
```
# Re-do the plot from the previous cell
# and change at least two aesthetic elements (colors, labels, titles, ...)

bPlot.plot(kind = 'barh', x = 'Education.Major',
           figsize = (12,5), color = 'orange' )
```

[29]: <AxesSubplot:ylabel='Education.Major'>

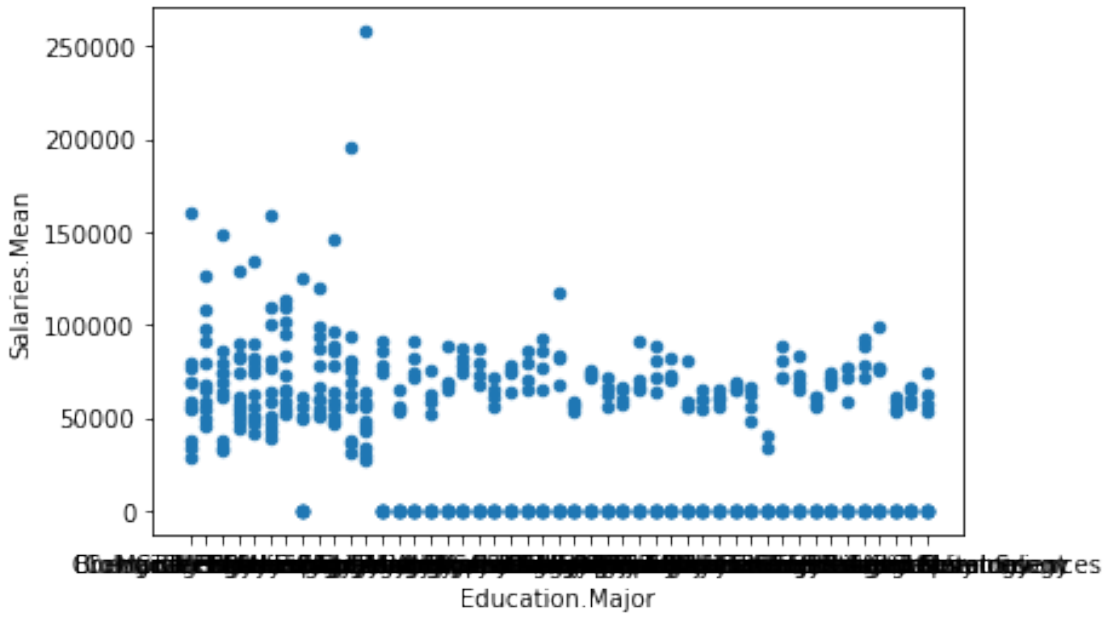


2 Free form section

- Choose another type of plot that interests you from the pandas.DataFrame.plot documentation [look at the 'kind' parameter] and make a new plot of your dataset values using the plot type

[44]:
```
b.plot(kind = 'scatter', x = 'Education.Major', y = 'Salaries.Mean' )
```

[44]: <AxesSubplot:xlabel='Education.Major', ylabel='Salaries.Mean'>

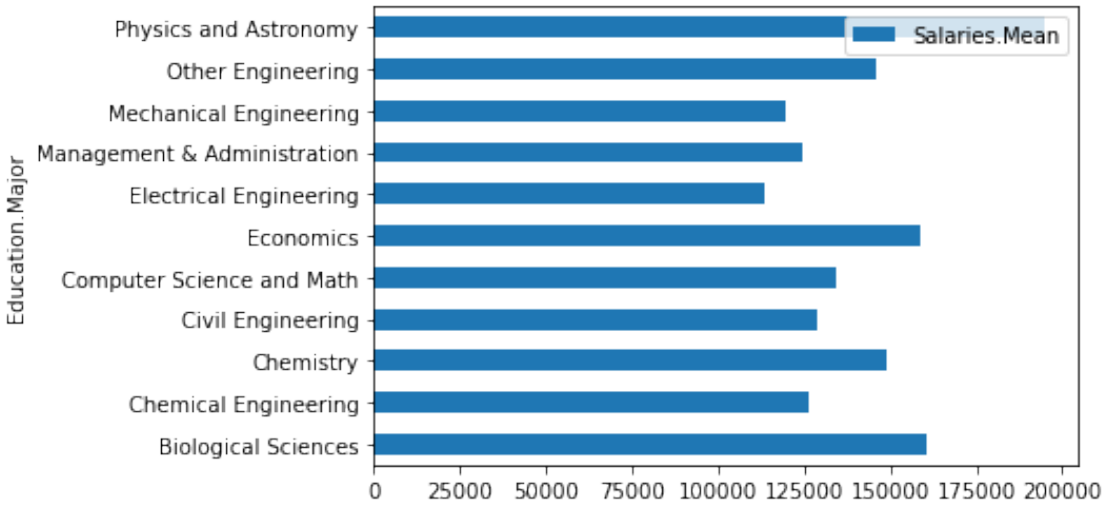


- Copy some of your analysis from the Week 2 assignment into new cells below
- Clean them up if desired, and make sure that you translate them to work with your new pandas dataframe structure here if needed
- Create several plots to complement and extend your analysis

[ ]:
```
#create a new dictionary that will contain info on each major and their mean␣
 ↪salary
majorsAndSal = {}
for val in grad_major:
    #extract the name of the major and mean salary from the dataset
    majorsAndSal.update( {(val['Education'])['Major'] :␣
 ↪(val['Salaries'])['Mean']} )

#this should now contain dictionary of majors and their mean salary
#majorsAndSal
```

[47]:
```
bPlot.plot(kind = 'barh', x = 'Education.Major' )
```

[47]: <AxesSubplot:ylabel='Education.Major'>

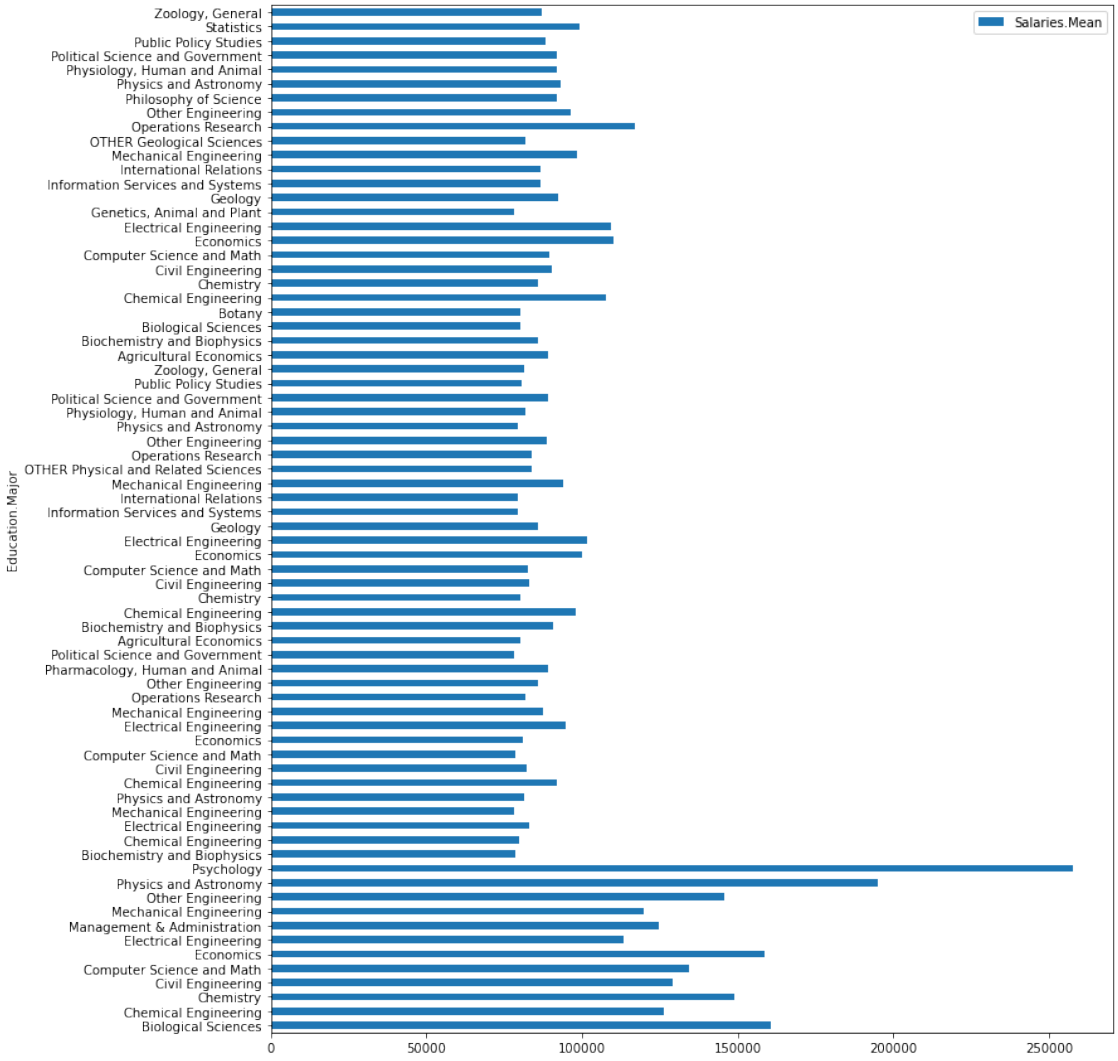


[ ]:
```
#creates a function that returns the mean of the values in a dictionary
def getMean(d):
    sum = 0
    for key in d:
        sum += d[key]
    return sum/ len(d)
```

[ ]:
```
meanSalary = getMean(majorsAndSal)
# meanSalary = 77979.24
```

[81]:
```
df['Salaries.Mean'].astype('float')

c1 = df[['Education.Major', 'Salaries.Mean']]
c2 = c1['Salaries.Mean'] >= 77979.24

c3 = c1[c2]

c3.plot(kind = 'barh', x = 'Education.Major', figsize = (12,15))
```

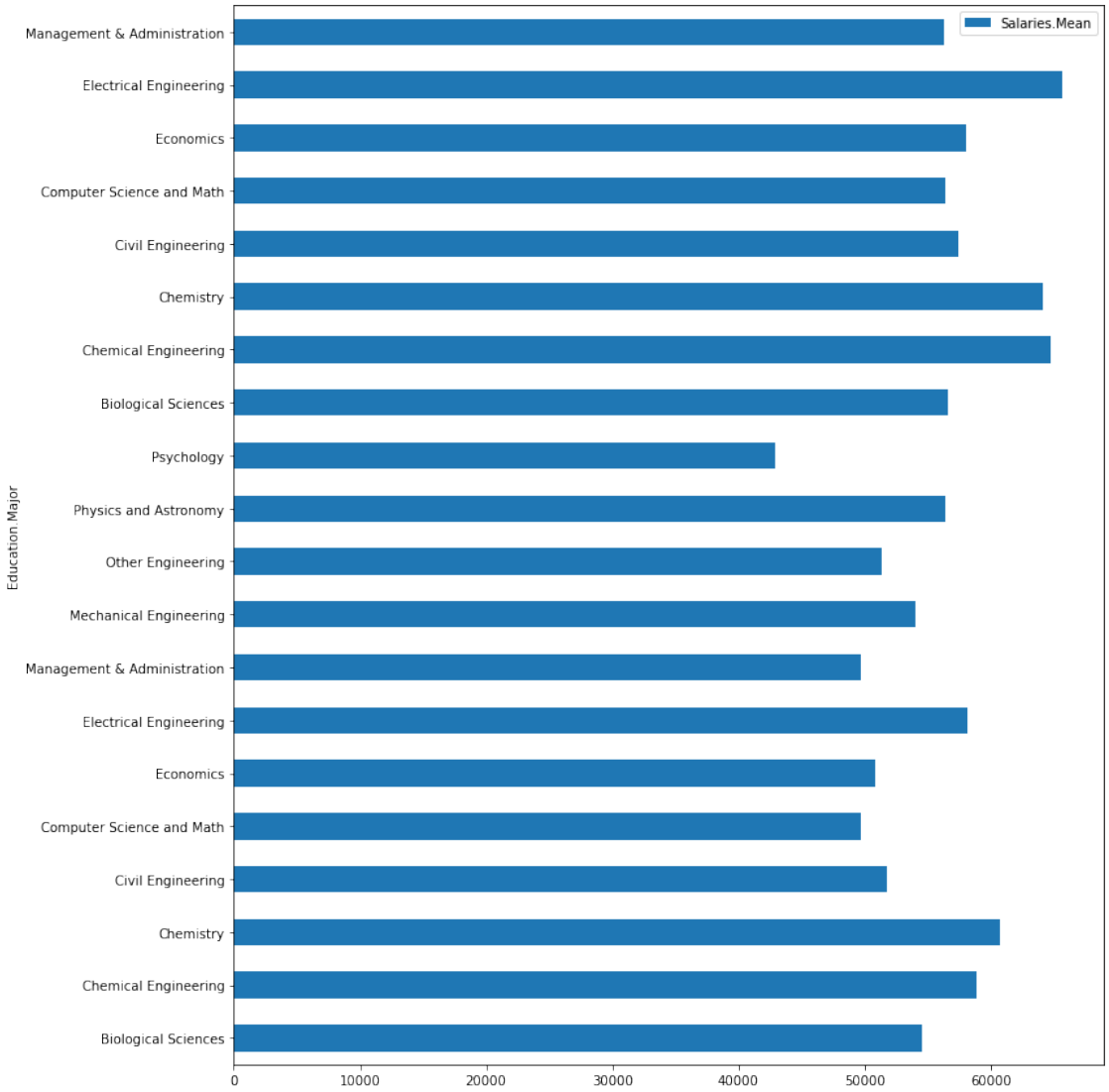
[81]:
```
<AxesSubplot:ylabel='Education.Major'>
```



[ ]:
```
def low (d):
    newDictLow = {}

    for key in d:
        if d[key] < meanSalary:
            newDictLow.update({key : d[key]})


    return newDictLow

lowSalaries = low(majorsAndSal)
lowSalaries
```

[94]:
```
df['Salaries.Mean'].astype('float')

d1 = df[['Education.Major', 'Salaries.Mean']]
d2 = d1['Salaries.Mean'] <= 77979.24

d3= d1[d2]

d4 = d3['Salaries.Mean'] != 0

d5 = d3[d4]

d6 = d5.head(20)

d6.plot(kind = 'barh', x = 'Education.Major', figsize = (12,15))
```

[94]: <AxesSubplot:ylabel='Education.Major'>



[98]:
```
c3
```

[98]:

	Education.Major	Salaries.Mean
0	Biological Sciences	160585.73
1	Chemical Engineering	126176.52
2	Chemistry	148872.00
3	Civil Engineering	129070.55
4	Computer Science and Math	134299.53
..	…	…
507	Physiology, Human and Animal	91750.58
509	Political Science and Government	92069.04
511	Public Policy Studies	88088.89
513	Statistics	99210.87
514	Zoology, General	86957.98

[72 rows x 2 columns]

[99]:
```
d5
```

[99]:

	Education.Major	Salaries.Mean
47	Biological Sciences	54523.54
48	Chemical Engineering	58896.72
49	Chemistry	60697.60
50	Civil Engineering	51758.63
51	Computer Science and Math	49672.61
..	…	…
501	Oceanography	76034.32
504	Pharmacology, Human and Animal	67193.25
508	Plant Sciences	66107.83
510	Psychology	63618.92
512	Sociology	58871.70

[192 rows x 2 columns]

[ ]:
```

```