Earnings_Visuals_Based_On_College_Major

February 28, 2020

```
[1]: import pandas as pd
     import matplotlib as plt
     %matplotlib inline

     recent_grads = pd.read_csv('recent-grads.csv')
     recent_grads.iloc[0]
     recent_grads.head()
```

```
[1]:    Rank  Major_code                                    Major    Total  \
     0     1        2419                      PETROLEUM ENGINEERING   2339.0
     1     2        2416                MINING AND MINERAL ENGINEERING    756.0
     2     3        2415                  METALLURGICAL ENGINEERING    856.0
     3     4        2417  NAVAL ARCHITECTURE AND MARINE ENGINEERING   1258.0
     4     5        2405                       CHEMICAL ENGINEERING  32260.0

            Men    Women Major_category  ShareWomen  Sample_size  Employed  … \
     0   2057.0    282.0    Engineering    0.120564           36      1976  …
     1    679.0     77.0    Engineering    0.101852            7       640  …
     2    725.0    131.0    Engineering    0.153037            3       648  …
     3   1123.0    135.0    Engineering    0.107313           16       758  …
     4  21239.0  11021.0    Engineering    0.341631          289     25694  …

        Part_time  Full_time_year_round  Unemployed  Unemployment_rate  Median  \
     0        270                  1207          37           0.018381  110000
     1        170                   388          85           0.117241   75000
     2        133                   340          16           0.024096   73000
     3        150                   692          40           0.050125   70000
     4       5180                 16697        1672           0.061098   65000

        P25th   P75th  College_jobs  Non_college_jobs  Low_wage_jobs
     0  95000  125000          1534               364            193
     1  55000   90000           350               257             50
     2  50000  105000           456               176              0
     3  43000   80000           529               102              0
     4  50000   75000         18314              4440            972

     [5 rows x 21 columns]
```

1

```
[2]: recent_grads.tail()
```

```
[2]:       Rank  Major_code                    Major   Total     Men   Women  \
     168   169        3609                    ZOOLOGY  8409.0  3050.0  5359.0
     169   170        5201    EDUCATIONAL PSYCHOLOGY  2854.0   522.0  2332.0
     170   171        5202       CLINICAL PSYCHOLOGY  2838.0   568.0  2270.0
     171   172        5203    COUNSELING PSYCHOLOGY  4626.0   931.0  3695.0
     172   173        3501          LIBRARY SCIENCE  1098.0   134.0   964.0

                    Major_category  ShareWomen  Sample_size  Employed  …  \
     168      Biology & Life Science    0.637293           47      6259  …
     169   Psychology & Social Work    0.817099            7      2125  …
     170   Psychology & Social Work    0.799859           13      2101  …
     171   Psychology & Social Work    0.798746           21      3777  …
     172                   Education    0.877960            2       742  …

          Part_time  Full_time_year_round  Unemployed  Unemployment_rate  Median  \
     168       2190                  3602         304           0.046320   26000
     169        572                  1211         148           0.065112   25000
     170        648                  1293         368           0.149048   25000
     171        965                  2738         214           0.053621   23400
     172        237                   410          87           0.104946   22000

          P25th  P75th  College_jobs  Non_college_jobs  Low_wage_jobs
     168  20000  39000          2771              2947            743
     169  24000  34000          1488               615             82
     170  25000  40000           986               870            622
     171  19200  26000          2403              1245            308
     172  20000  22000           288               338            192

     [5 rows x 21 columns]
```

```
[3]: recent_grads.describe()
```

```
[3]:              Rank   Major_code          Total            Men          Women  \
     count  173.000000   173.000000     172.000000     172.000000     172.000000
     mean    87.000000  3879.815029   39370.081395   16723.406977   22646.674419
     std     50.084928  1687.753140   63483.491009   28122.433474   41057.330740
     min      1.000000  1100.000000     124.000000     119.000000       0.000000
     25%     44.000000  2403.000000    4549.750000    2177.500000    1778.250000
     50%     87.000000  3608.000000   15104.000000    5434.000000    8386.500000
     75%    130.000000  5503.000000   38909.750000   14631.000000   22553.750000
     max    173.000000  6403.000000  393735.000000  173809.000000  307087.000000

            ShareWomen  Sample_size      Employed      Full_time     Part_time  \
     count  172.000000   173.000000    173.000000     173.000000    173.000000
     mean     0.522223   356.080925   31192.763006   26029.306358   8832.398844
```

```
std        0.231205     618.361022    50675.002241    42869.655092    14648.179473
min        0.000000       2.000000        0.000000      111.000000        0.000000
25%        0.336026      39.000000     3608.000000     3154.000000     1030.000000
50%        0.534024     130.000000    11797.000000    10048.000000     3299.000000
75%        0.703299     338.000000    31433.000000    25147.000000     9948.000000
max        0.968954    4212.000000   307933.000000   251540.000000   115172.000000
```

```
       Full_time_year_round    Unemployed   Unemployment_rate         Median  \
count          173.000000      173.000000          173.000000     173.000000
mean         19694.427746     2416.329480            0.068191   40151.445087
std          33160.941514     4112.803148            0.030331   11470.181802
min            111.000000        0.000000            0.000000   22000.000000
25%           2453.000000      304.000000            0.050306   33000.000000
50%           7413.000000      893.000000            0.067961   36000.000000
75%          16891.000000     2393.000000            0.087557   45000.000000
max         199897.000000    28169.000000            0.177226  110000.000000
```

```
              P25th          P75th   College_jobs   Non_college_jobs  \
count    173.000000     173.000000     173.000000         173.000000
mean   29501.445087   51494.219653   12322.635838       13284.497110
std     9166.005235   14906.279740   21299.868863       23789.655363
min    18500.000000   22000.000000       0.000000           0.000000
25%    24000.000000   42000.000000    1675.000000        1591.000000
50%    27000.000000   47000.000000    4390.000000        4595.000000
75%    33000.000000   60000.000000   14444.000000       11783.000000
max    95000.000000  125000.000000  151643.000000      148395.000000
```

```
        Low_wage_jobs
count      173.000000
mean      3859.017341
std       6944.998579
min          0.000000
25%        340.000000
50%       1231.000000
75%       3466.000000
max      48207.000000
```

```python
[4]: raw_data_count = recent_grads.shape
     raw_data_count
```

```
[4]: (173, 21)
```

```python
[5]: recent_grads = recent_grads.dropna()
     cleaned_data_count = recent_grads.shape
     cleaned_data_count
```

```
[5]: (172, 21)
```

```
[6]: recent_grads.plot(x='Sample_size', y='Median', kind='scatter')
```

[6]: <matplotlib.axes._subplots.AxesSubplot at 0x117221ed0>
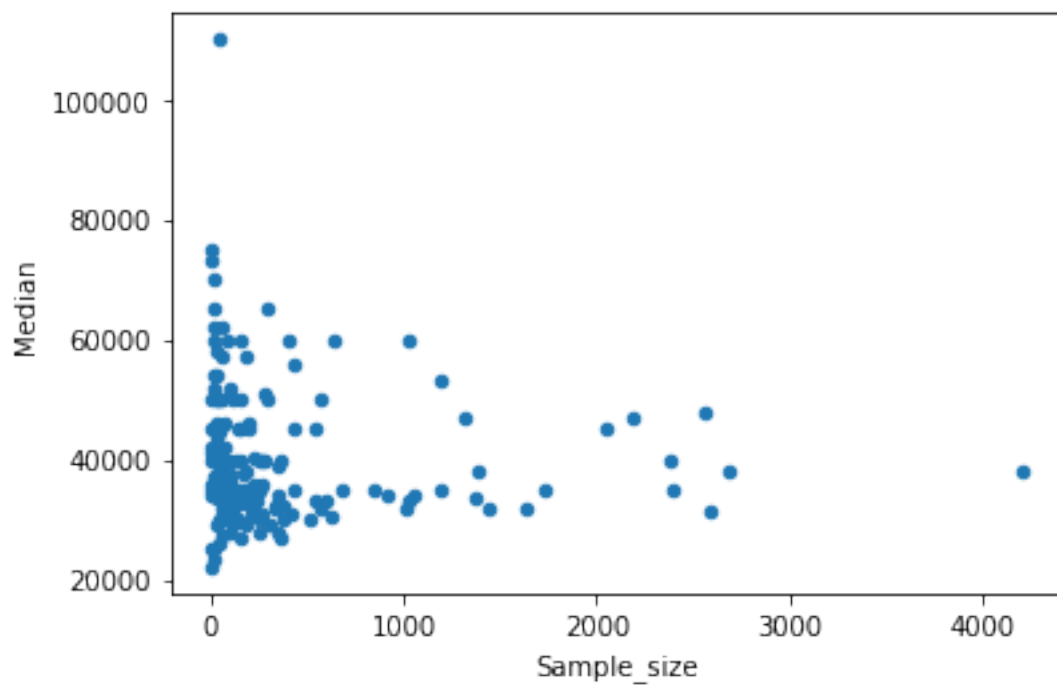

```
[7]: recent_grads.plot(x='Sample_size', y='Unemployment_rate', kind='scatter')
```
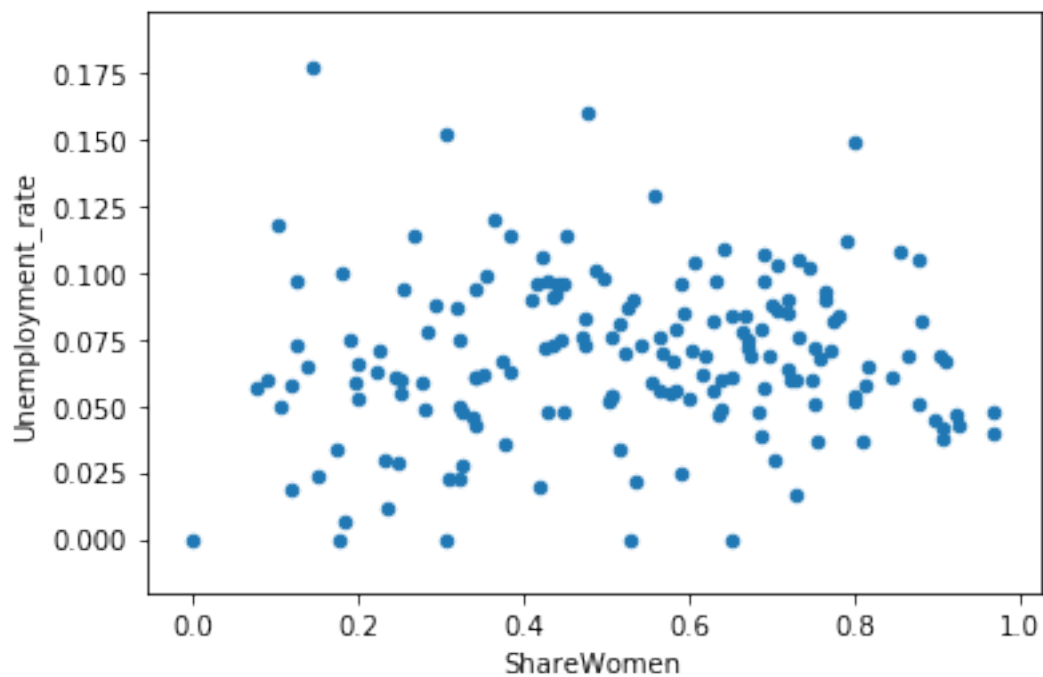
[7]: <matplotlib.axes._subplots.AxesSubplot at 0x117d0f790>
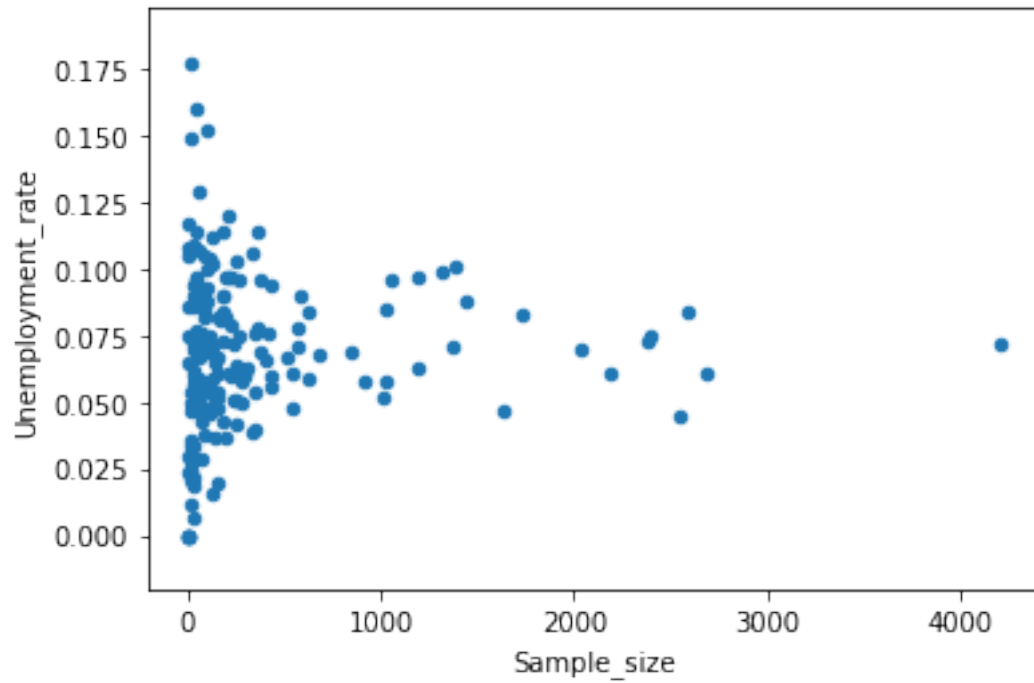
```
[8]: recent_grads.plot(x='Full_time', y='Median', kind='scatter')
```
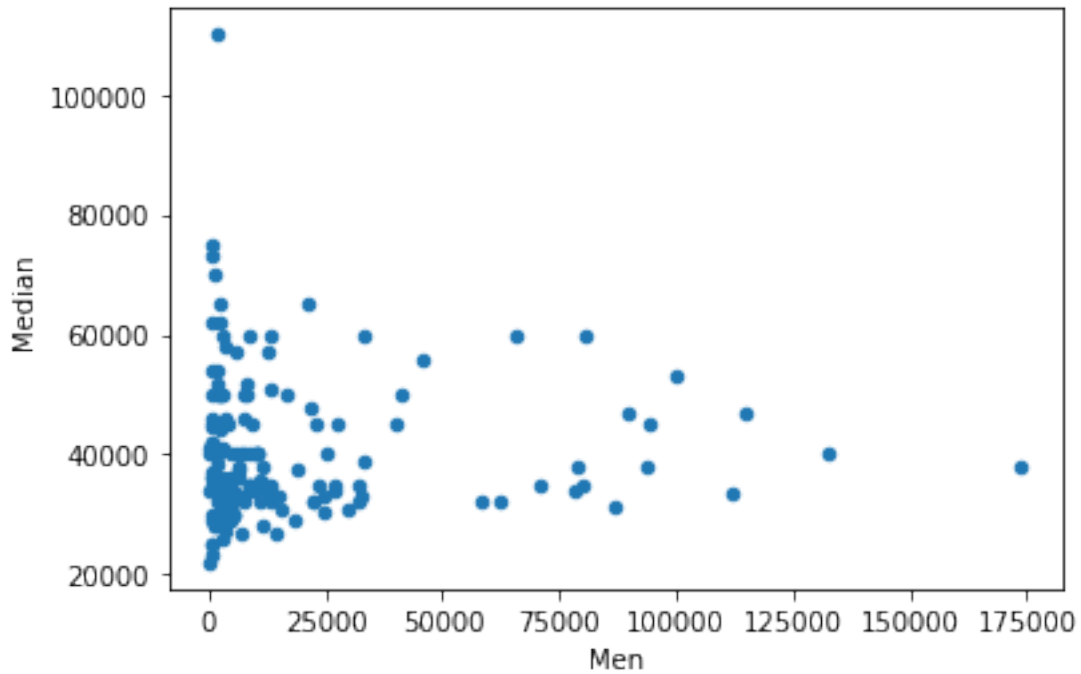
[8]: <matplotlib.axes._subplots.AxesSubplot at 0x117e232d0>
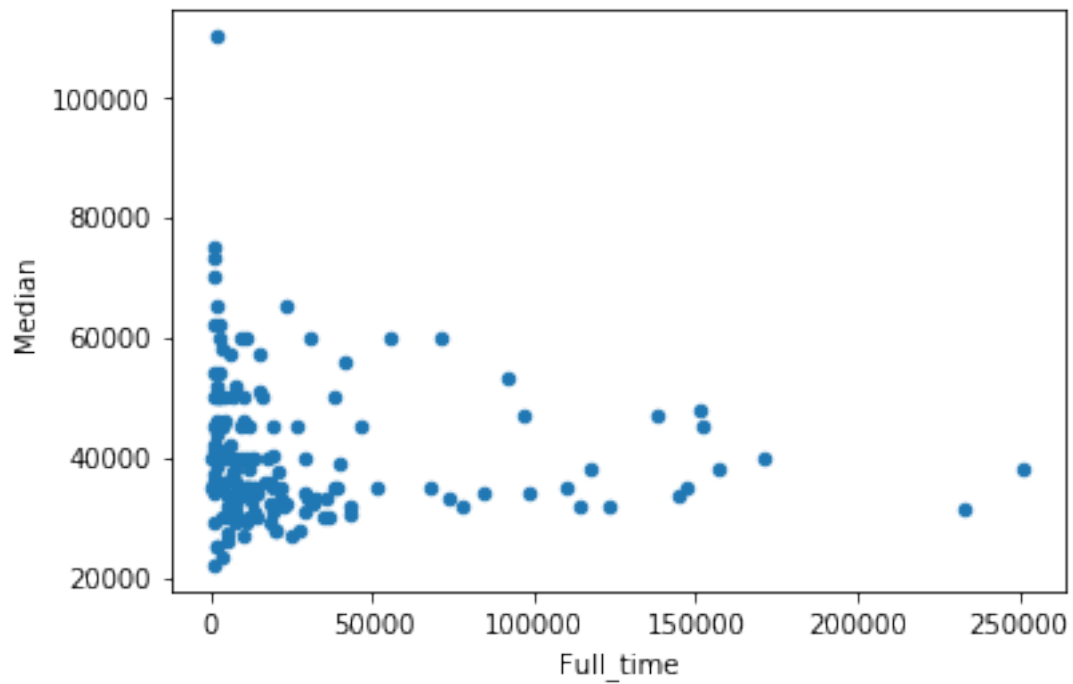
```
[9]: recent_grads.plot(x='ShareWomen', y='Unemployment_rate', kind='scatter')
```

[9]: <matplotlib.axes._subplots.AxesSubplot at 0x117f34410>


```
[10]: recent_grads.plot(x='Men', y='Median', kind='scatter')
```

[10]: <matplotlib.axes._subplots.AxesSubplot at 0x11801be10>

Dependent variables: values on the y-axis Independent variables: values on x-axis - Things to lookk at in a scatter plot: Direction - positive or negative, Form - linear or non-linear, Strength - weak or moderate or strong, outliers

[11]: `recent_grads.plot(x='Women', y='Median', kind='scatter')`

[11]: `<matplotlib.axes._subplots.AxesSubplot at 0x1180e7850>`
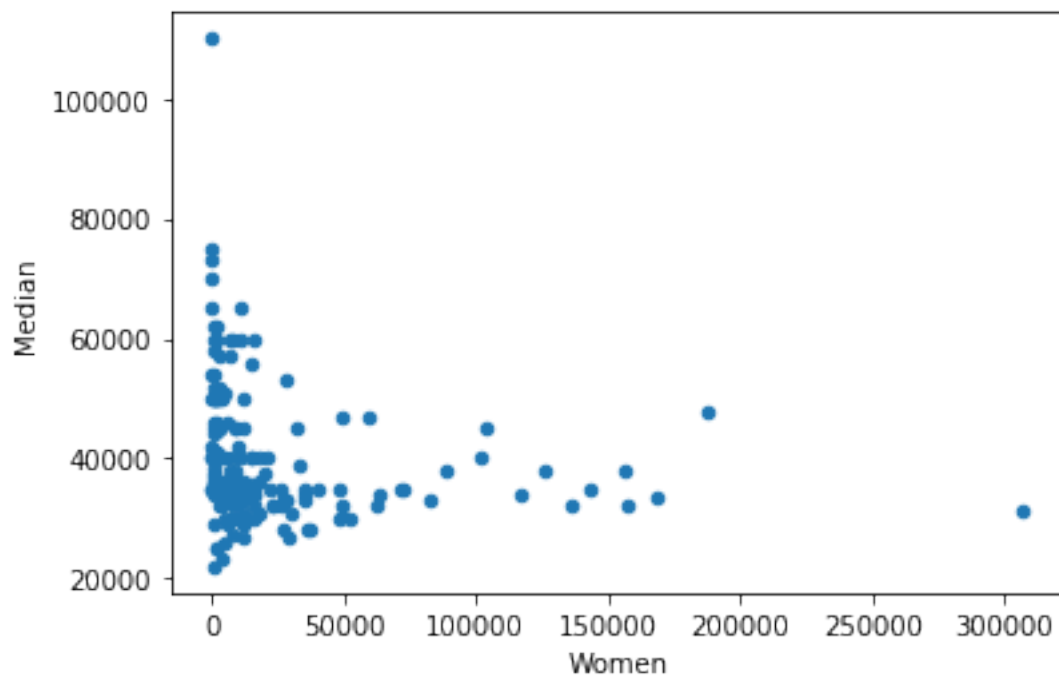
```
[12]: recent_grads['Sample_size'].hist(bins=25, range=(0,5000))
```

[12]: <matplotlib.axes._subplots.AxesSubplot at 0x1181d6a10>
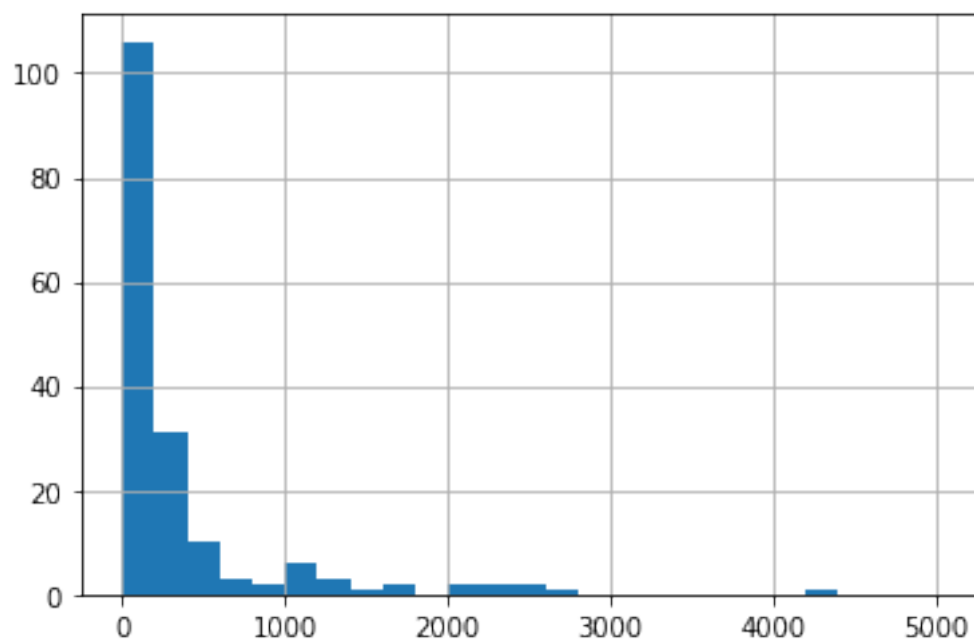
[13]: `recent_grads['Median'].hist(bins=25, range=(0,100000))`

[13]: `<matplotlib.axes._subplots.AxesSubplot at 0x1171f1590>`
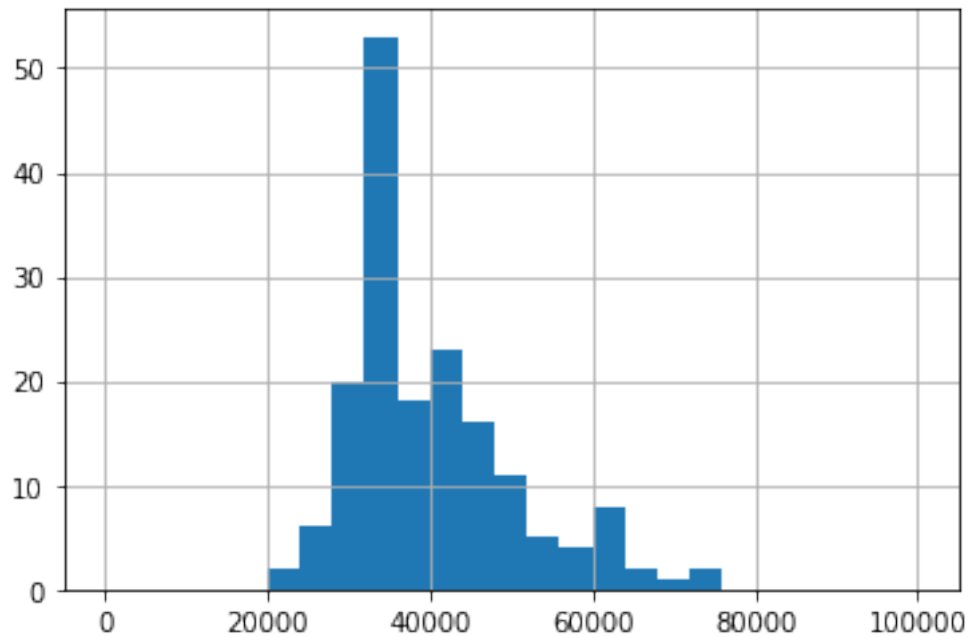
As we can see from the graph above, the most common median salary ranges anywhere between 30,000 and 35,000 dollars.

[14]: `recent_grads['Employed'].hist(bins=25, range=(0,100000))`

[14]: `<matplotlib.axes._subplots.AxesSubplot at 0x11842f510>`
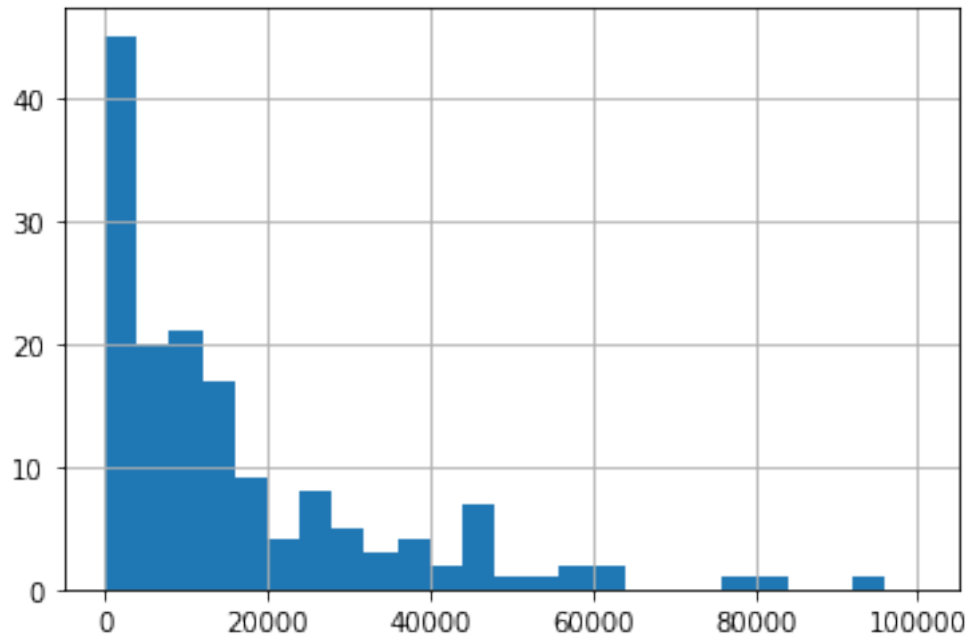
```
[15]: recent_grads['Full_time'].hist(bins=25, range=(0,180000))
```

```
[15]: <matplotlib.axes._subplots.AxesSubplot at 0x118549410>
```
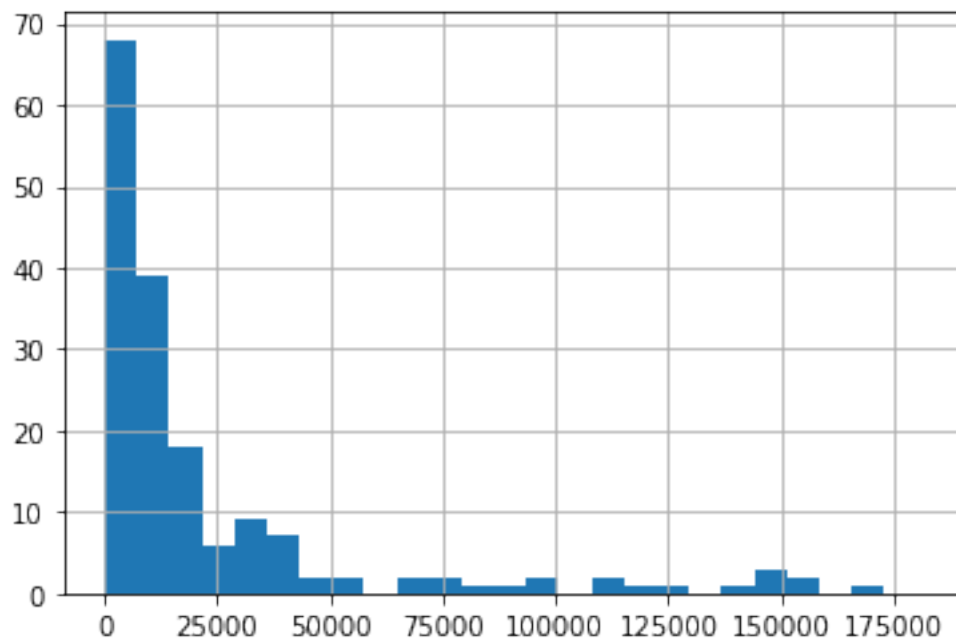
```
[16]: recent_grads['ShareWomen'].hist(bins=25, range=(0,1))
```
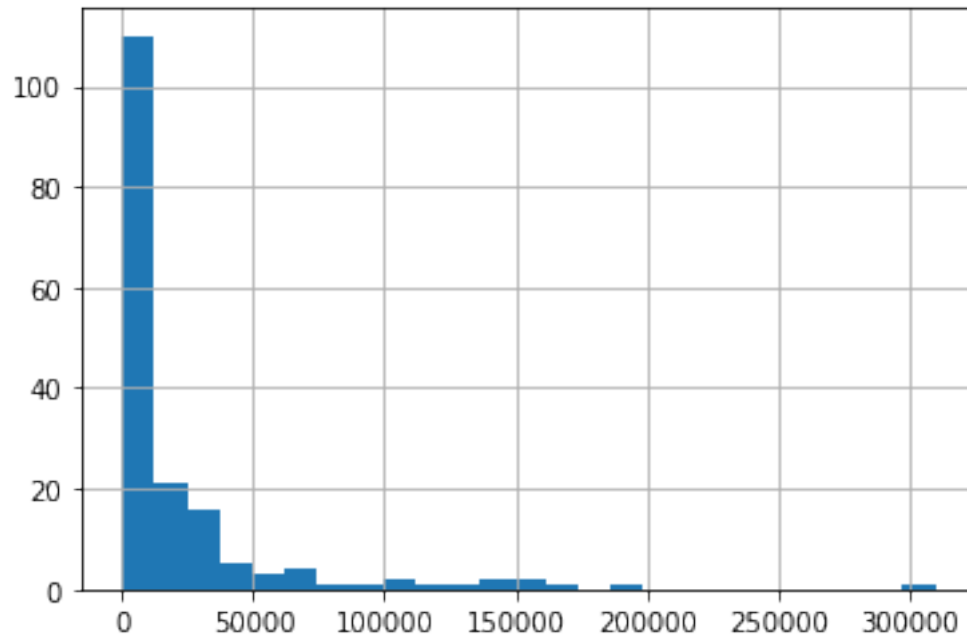
[16]: <matplotlib.axes._subplots.AxesSubplot at 0x118664c10>
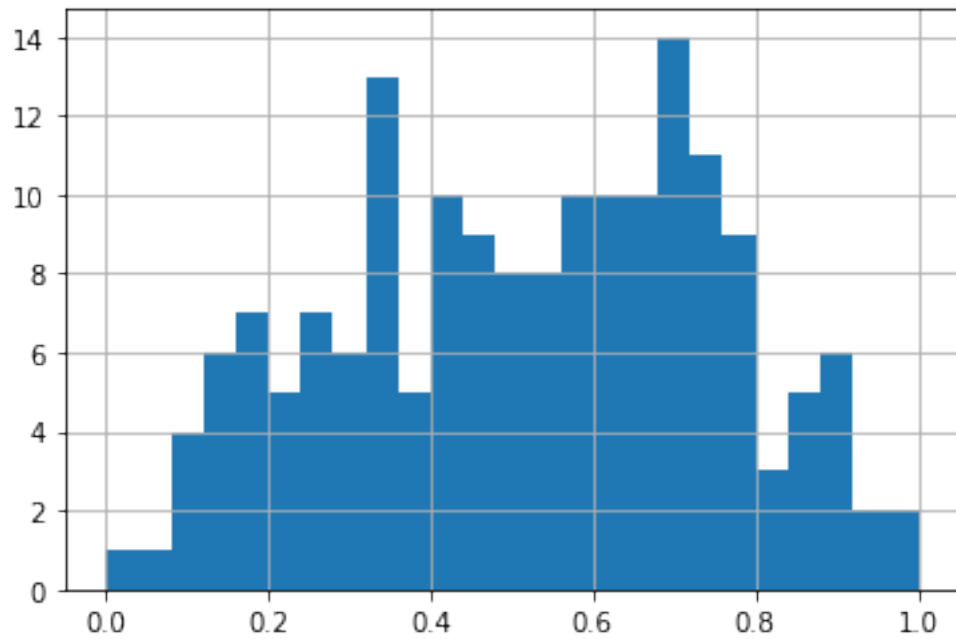

```
[17]: recent_grads['Unemployment_rate'].hist(bins=25, range=(0,0.2))
```

[17]: <matplotlib.axes._subplots.AxesSubplot at 0x11878f7d0>
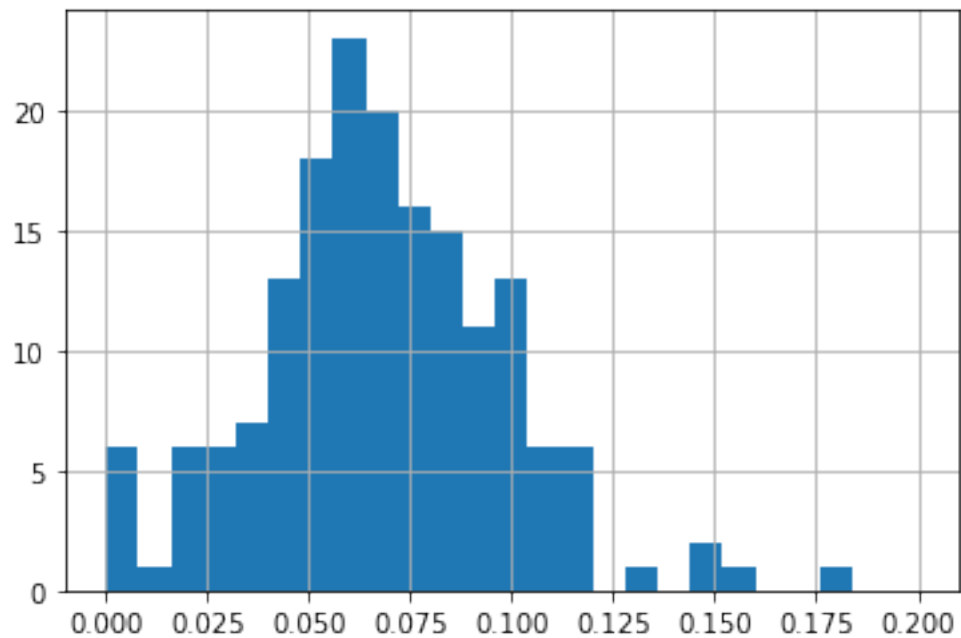
```
[18]: recent_grads['Men'].hist(bins=25, range=(0,175000))
```

```
[18]: <matplotlib.axes._subplots.AxesSubplot at 0x1188a3e90>
```
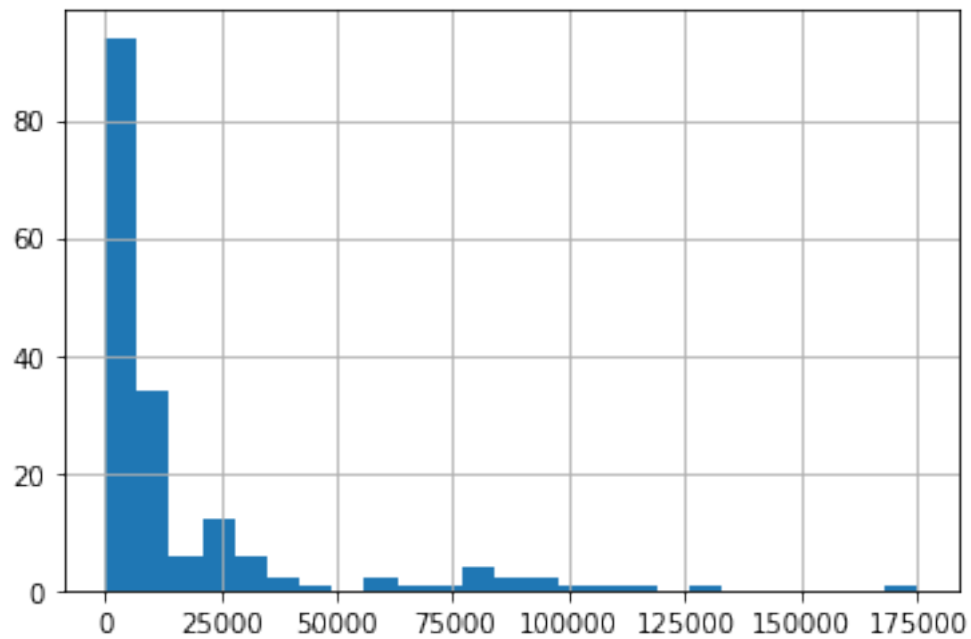

As you can observe from the graph above, about 55% of the majors were predominantly men.

```
[19]: recent_grads['Women'].hist(bins=25, range=(0,310000))
```

```
[19]: <matplotlib.axes._subplots.AxesSubplot at 0x1189d0490>
```

As you can observe from the graph above, about 66-70% of majors were predominantly women.

```
[26]: from pandas.plotting import scatter_matrix
      scatter_matrix(recent_grads[['Sample_size','Median']], figsize=(6,6))
```
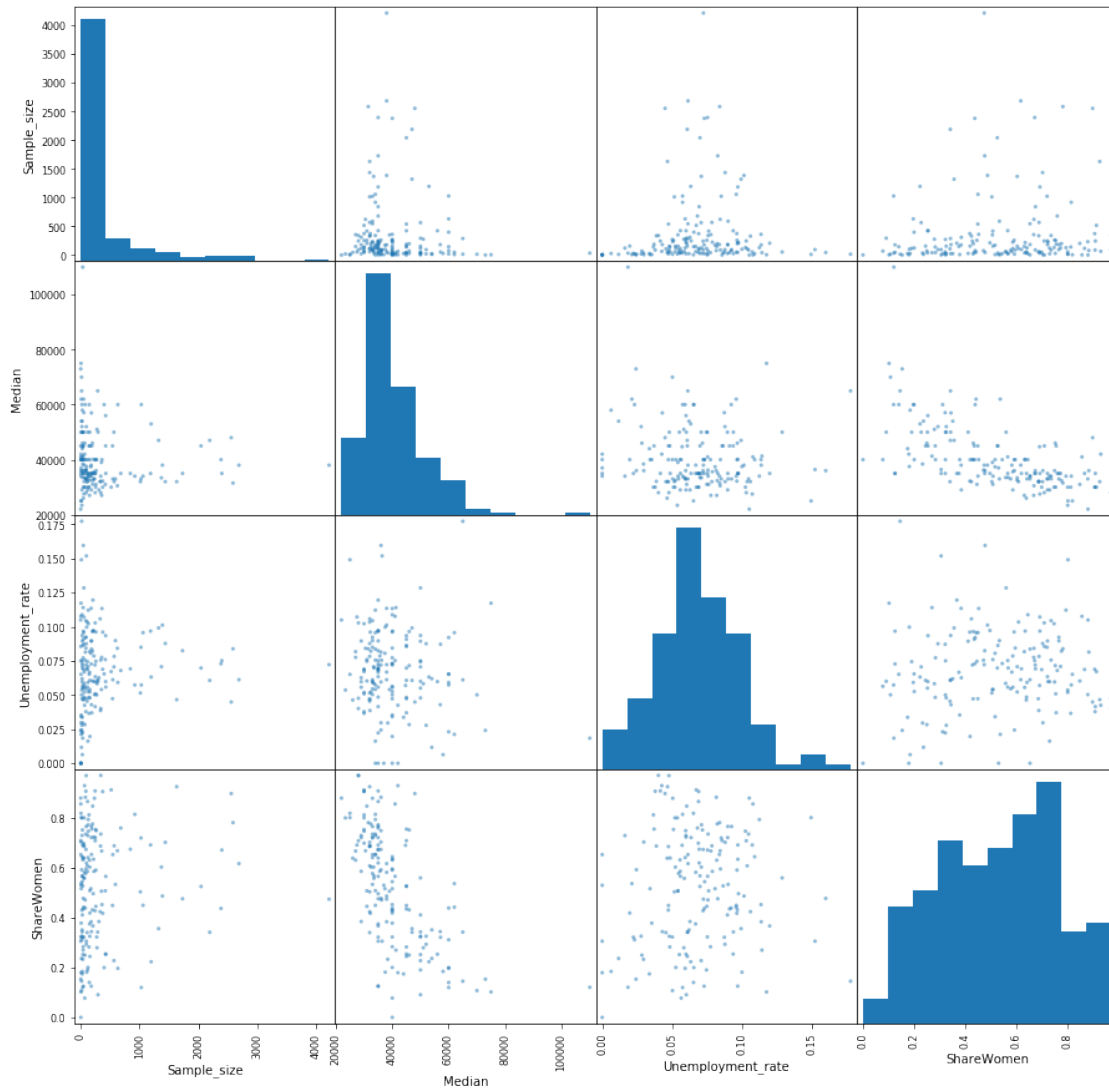
```
[26]: array([[<matplotlib.axes._subplots.AxesSubplot object at 0x118f1e890>,
              <matplotlib.axes._subplots.AxesSubplot object at 0x118fd0710>],
             [<matplotlib.axes._subplots.AxesSubplot object at 0x1190b7f10>,
              <matplotlib.axes._subplots.AxesSubplot object at 0x1190ec710>]],
            dtype=object)
```
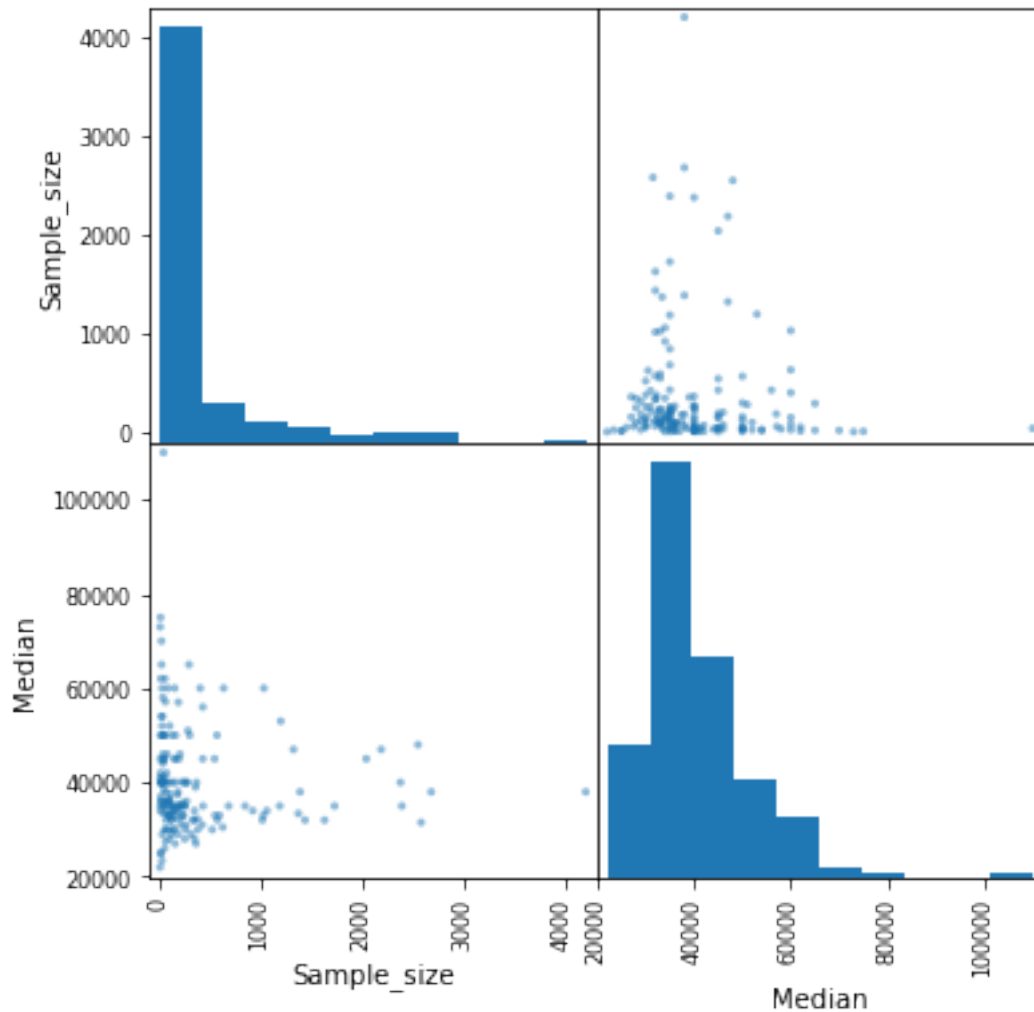
```
[30]: scatter_matrix(recent_grads[['Sample_size','Median','Unemployment_rate',
      ↪'ShareWomen']], figsize=(15,15))
```

```
[30]: array([[<matplotlib.axes._subplots.AxesSubplot object at 0x11a5d7390>,
              <matplotlib.axes._subplots.AxesSubplot object at 0x104ec5610>,
              <matplotlib.axes._subplots.AxesSubplot object at 0x104ef0dd0>,
              <matplotlib.axes._subplots.AxesSubplot object at 0x104f31610>],
             [<matplotlib.axes._subplots.AxesSubplot object at 0x105c4be10>,
              <matplotlib.axes._subplots.AxesSubplot object at 0x105c8c650>,
              <matplotlib.axes._subplots.AxesSubplot object at 0x105cc0e50>,
              <matplotlib.axes._subplots.AxesSubplot object at 0x105d01690>],
             [<matplotlib.axes._subplots.AxesSubplot object at 0x105d0b6d0>,
              <matplotlib.axes._subplots.AxesSubplot object at 0x105d4d090>,
              <matplotlib.axes._subplots.AxesSubplot object at 0x119a45f10>,
              <matplotlib.axes._subplots.AxesSubplot object at 0x119a7b710>],
```

```
[<matplotlib.axes._subplots.AxesSubplot object at 0x119abaf50>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x119af0750>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x119b24f50>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x119b67790>]],
dtype=object)
```


- In a sample size of about 500 people, most earn a Median salary between 20,000 to 40,000 dollars.
- In the majors that were majority women, those majors earned significantly less than majors that were majority men.
- In the first sample size of about 1000 people, most of them were women.

```
[39]: recent_grads[:10].plot.bar(x ='Major', y='ShareWomen', legend=False)
      recent_grads[163:].plot.bar(x ='Major', y='ShareWomen', legend=False)
```
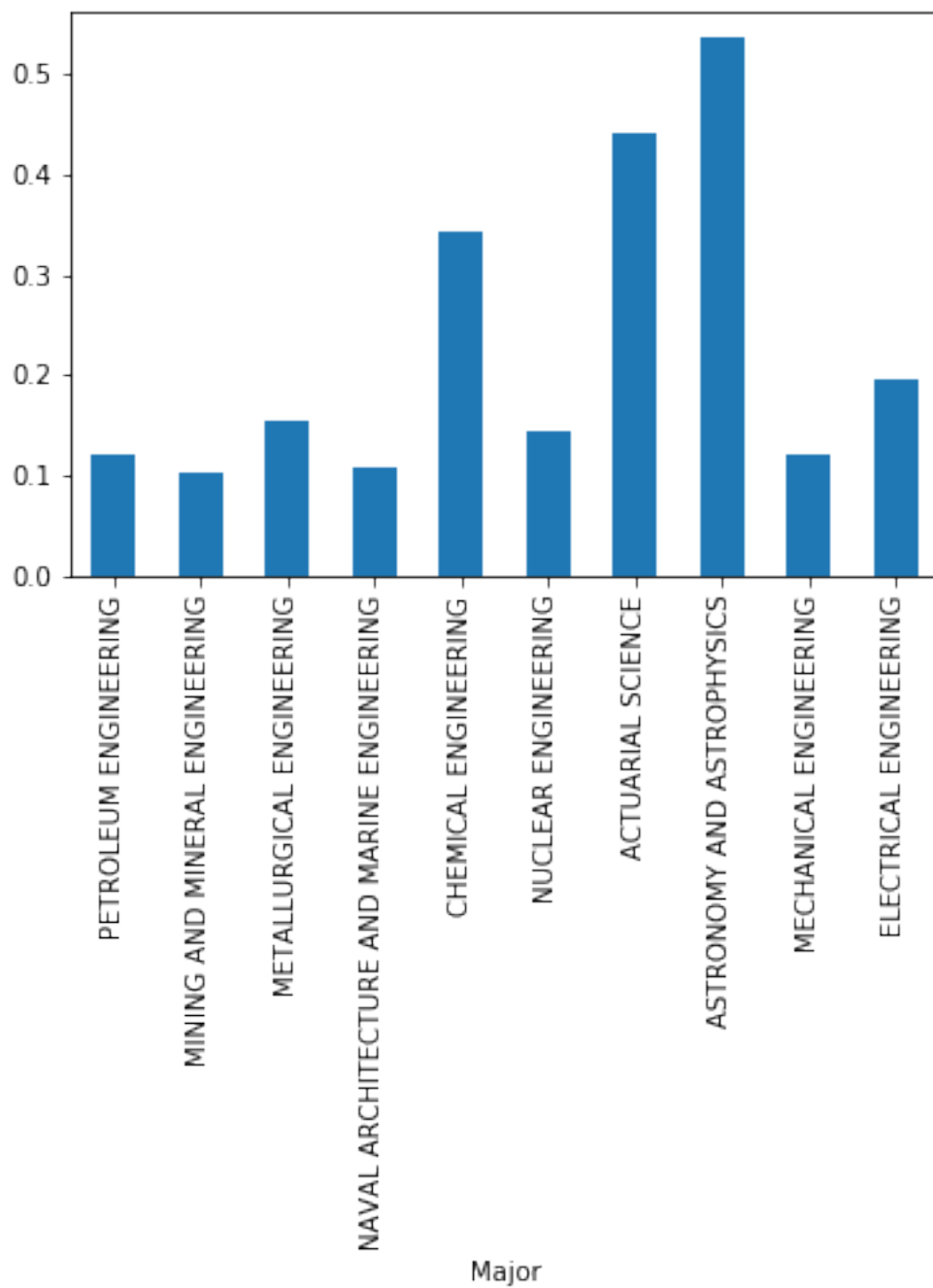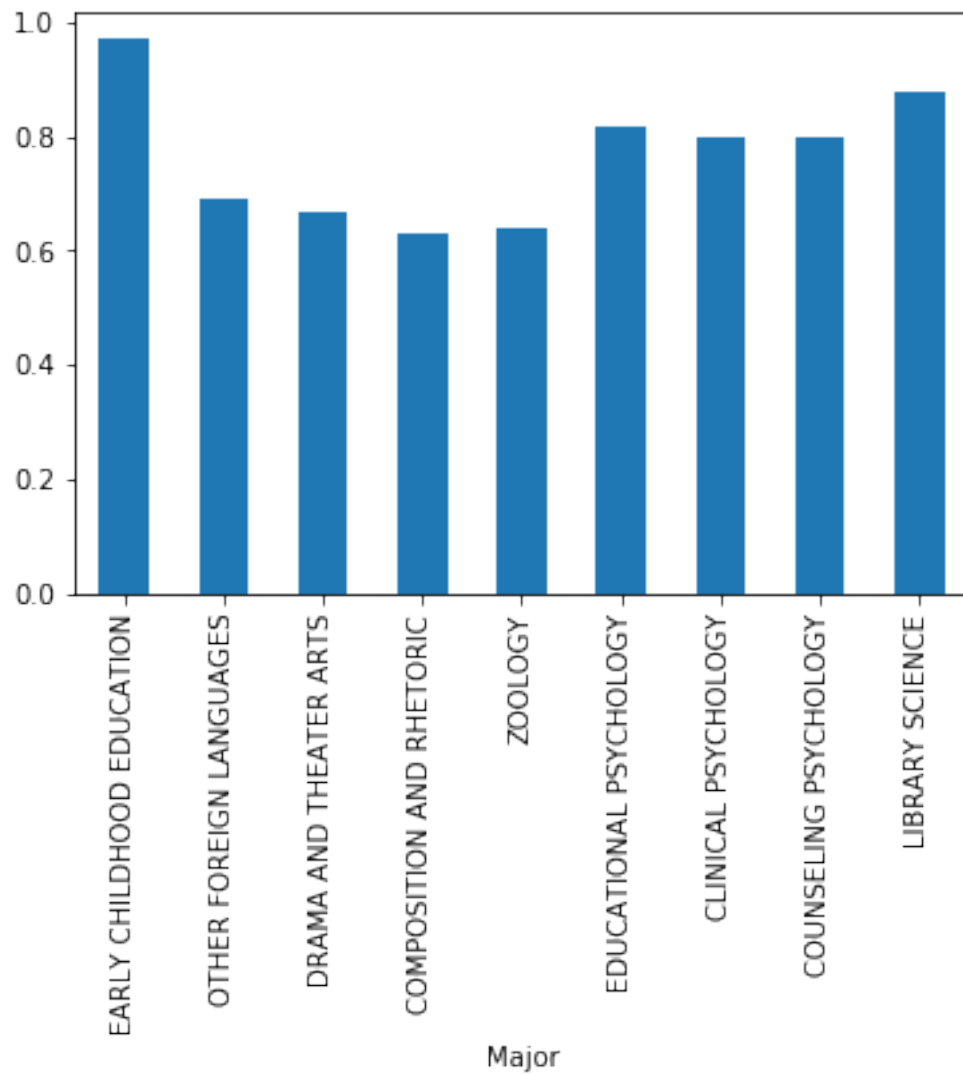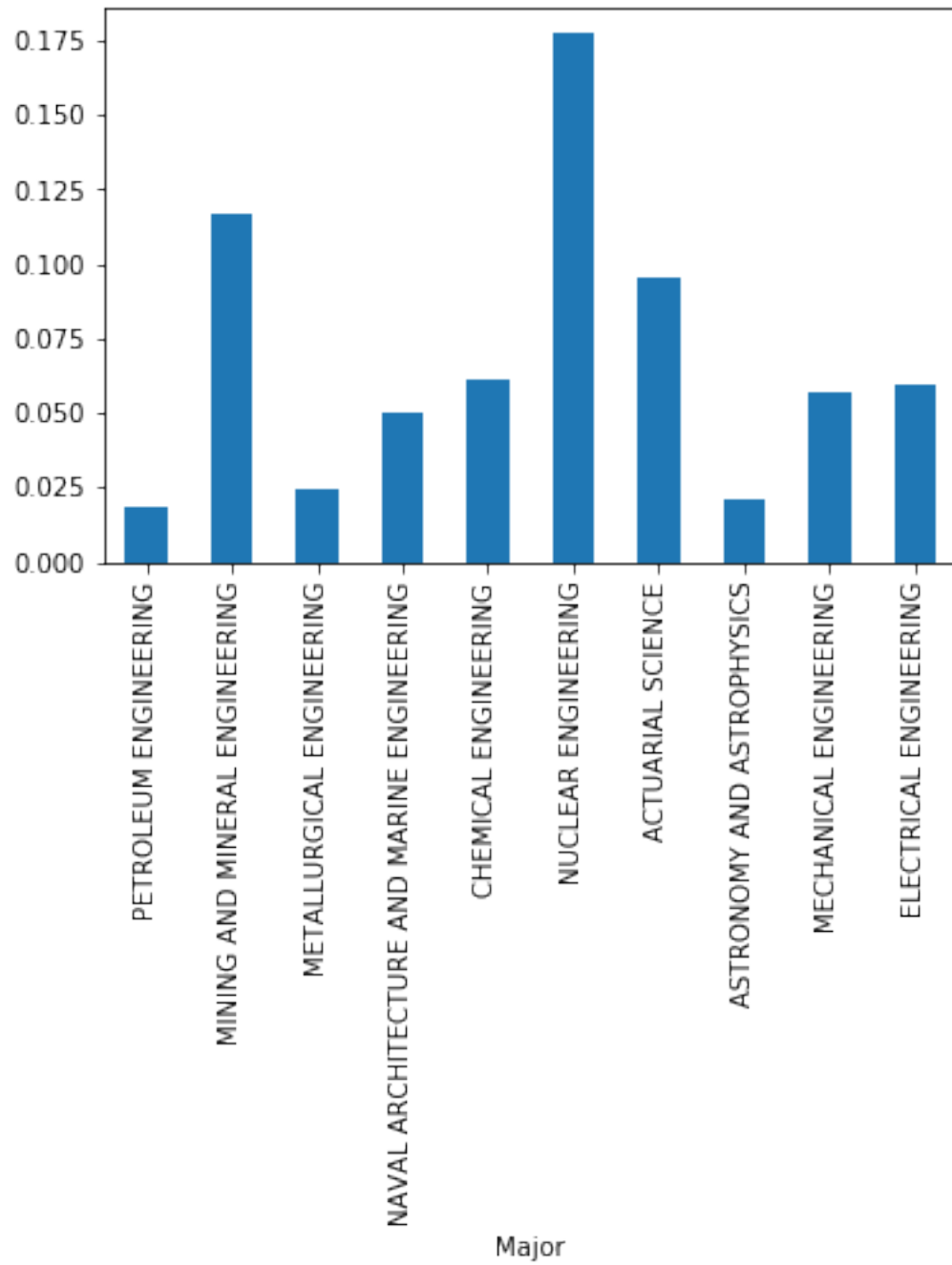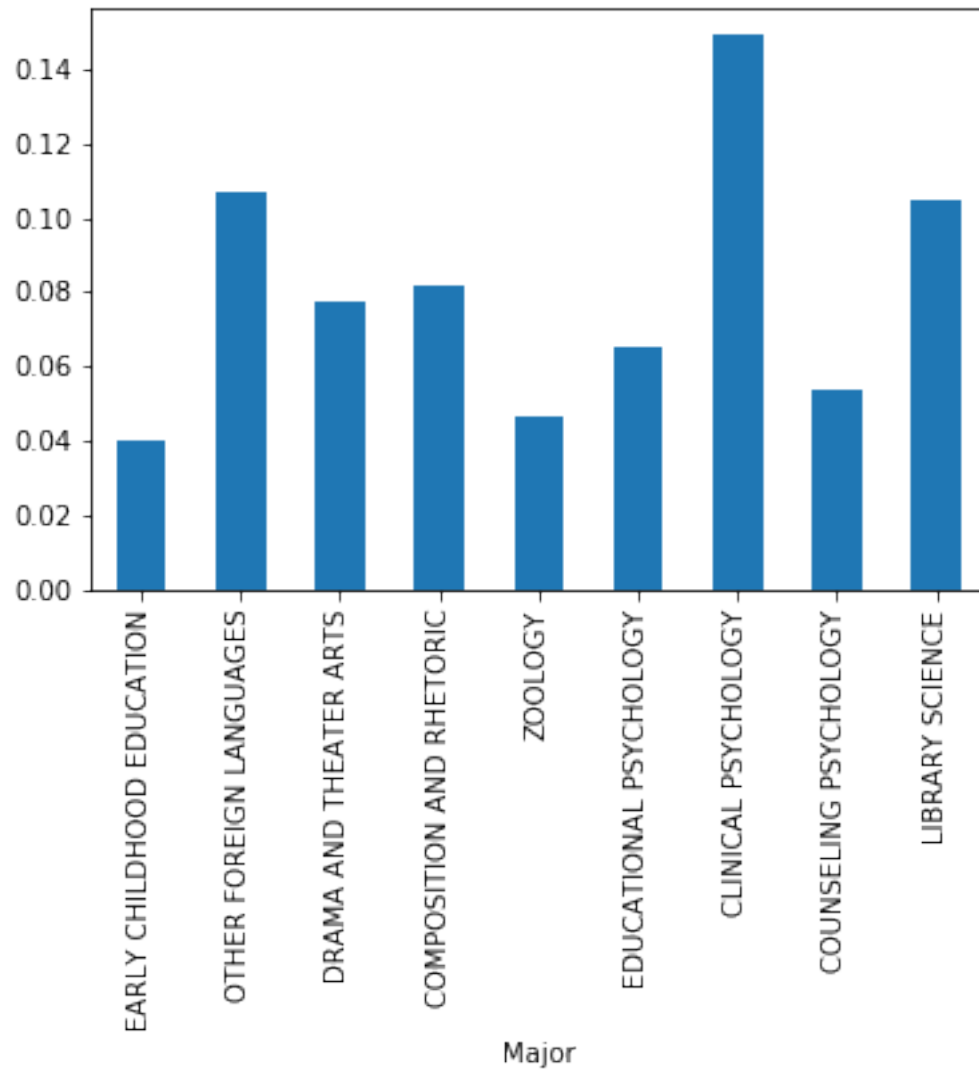
```
[41]: recent_grads[:10].plot.bar(x ='Major', y='Unemployment_rate', legend=False)
      recent_grads[163:].plot.bar(x ='Major', y='Unemployment_rate', legend=False)
```

```
[41]: <matplotlib.axes._subplots.AxesSubplot at 0x11b7d5a90>
```

Major

Major

[ ]: