In [119]:
```
import pandas as pd
df=pd.read_csv('ass1.csv')
df.head(5)
```

Out[119]:

	Undergraduate Major	Starting Median Salary	Mid-Career Median Salary	Percent change from Starting to Mid-Career Salary	Mid-Career 10th Percentile Salary	Mid-Career 25th Percentile Salary	Mid-Career 75th Percentile Salary	Mid-Car 9(Percent Sal
0	Accounting	$46,000.00	$77,100.00	67.6	$42,200.00	$56,100.00	$108,000.00	$152,000
1	Aerospace Engineering	$57,700.00	$101,000.00	75.0	$64,300.00	$82,100.00	$127,000.00	$161,000
2	Agriculture	$42,600.00	$71,900.00	68.8	$36,300.00	$52,100.00	$96,300.00	$150,000
3	Anthropology	$36,800.00	$61,500.00	67.1	$33,800.00	$45,500.00	$89,300.00	$138,000
4	Architecture	$41,600.00	$76,800.00	84.6	$50,600.00	$62,200.00	$97,000.00	$136,000

In [120]:
```
df.columns
```

Out[120]:
```
Index(['Undergraduate Major', 'Starting Median Salary',
       'Mid-Career Median Salary',
       'Percent change from Starting to Mid-Career Salary',
       'Mid-Career 10th Percentile Salary',
       'Mid-Career 25th Percentile Salary',
       'Mid-Career 75th Percentile Salary',
       'Mid-Career 90th Percentile Salary'],
      dtype='object')
```

In [121]:
```
df.size
```

Out[121]:
```
400
```

In [122]:
```
names=['major','Start Med Sal', 'Mid Career Med Sal', 'Start-Mid Career
 Sal Change', '10%tile Sal','25%tile Sal','75%tile Sal','90%tileSal']
df=pd.read_csv('ass1.csv',names=names,skiprows=1)
df.columns
```

Out[122]:
```
Index(['major', 'Start Med Sal', 'Mid Career Med Sal',
       'Start-Mid Career Sal Change', '10%tile Sal', '25%tile Sal',
       '75%tile Sal', '90%tileSal'],
      dtype='object')
```

In [125]:
```
df.dtypes
```

Out[125]:
```
major                          object
Start Med Sal                  object
Mid Career Med Sal             object
Start-Mid Career Sal Change    float64
10%tile Sal                    object
25%tile Sal                    object
75%tile Sal                    object
90%tileSal                     object
dtype: object
```

In [126]:
```
names=names[1:]
for i in nam:
    df[i]=df[i].str.replace('\$','').str.replace(',','').astype('float'
)
df.dtypes
```

Out[126]:
```
major                          object
Start Med Sal                  float64
Mid Career Med Sal             float64
Start-Mid Career Sal Change    float64
10%tile Sal                    float64
25%tile Sal                    float64
75%tile Sal                    float64
90%tileSal                     float64
dtype: object
```

In [127]:
```
df.loc[(df['Start-Mid Career Sal Change']>70.0),'major']
```

Out[127]:
```
1       Aerospace Engineering
4                Architecture
5                 Art History
9                   Chemistry
11              Communications
12       Computer Engineering
13           Computer Science
17                   Economics
20                     English
21                        Film
22                     Finance
25                     Geology
28                     History
33    International Relations
34                  Journalism
36                   Marketing
37                        Math
42                  Philosophy
44                     Physics
45           Political Science
Name: major, dtype: object
```

In [135]:
```
df.loc[(df['90%tileSal']-df['10%tile Sal'])>40000.0,'Start Med Sal'].min()
```

Out[135]:
```
34000.0
```

In [141]:
```
df2=pd.read_csv('a2.csv')
na=df2.columns[2:]
for i in na:
    df2[i]=df2[i].str.replace('\$','').str.replace(',','').astype('float')
df2.dtypes
```

Out[141]:
```
School Name                    object
School Type                    object
Starting Median Salary        float64
Mid-Career Median Salary      float64
```

```
Mid-Career 10th Percentile Salary    float64
Mid-Career 25th Percentile Salary    float64
Mid-Career 75th Percentile Salary    float64
Mid-Career 90th Percentile Salary    float64
dtype: object
```

In [142]:
```
df2.isnull().sum()
```

Out[142]:
```
School Name                          0
School Type                          0
Starting Median Salary               0
Mid-Career Median Salary             0
Mid-Career 10th Percentile Salary    38
Mid-Career 25th Percentile Salary     0
Mid-Career 75th Percentile Salary     0
Mid-Career 90th Percentile Salary    38
dtype: int64
```

In [143]:
```
df2['Mid-Career 10th Percentile Salary']
```

Out[143]:
```
0      76800.0
1          NaN
2          NaN
3      66800.0
4          NaN
        ...
264    32200.0
265    25600.0
266    30700.0
267    22600.0
268    27000.0
Name: Mid-Career 10th Percentile Salary, Length: 269, dtype: float64
```

In [159]:
```
df2['Mid-Career 90th Percentile Salary'] = df2.groupby('School Type').t
ransform(lambda x: x.fillna(x.mean()))
df2['Mid-Career 10th Percentile Salary'] = df2.groupby('School Type').t
```

```
ransform(lambda x: x.fillna(x.mean()))
df2.isnull().sum()
```

Out[159]:
```
School Name                          0
School Type                          0
Starting Median Salary               0
Mid-Career Median Salary             0
Mid-Career 10th Percentile Salary    0
Mid-Career 25th Percentile Salary    0
Mid-Career 75th Percentile Salary    0
Mid-Career 90th Percentile Salary    0
dtype: int64
```

In [149]:
```
df2['Starting Median Salary'].loc[(df2['School Type']=='Liberal Arts')]
.mean()
```

Out[149]:
```
45746.8085106383
```

In [162]:
```
temp= df2['Mid-Career 25th Percentile Salary'].max()
df2.loc[(df2['Mid-Career 25th Percentile Salary'] == temp)]
```

Out[162]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary	Mid-Career 10th Percentile Salary	Mid-Career 25th Percentile Salary	Mid-Career 75th Percentile Salary	Mid-Career 90th Percentile Salary
1	California Institute of Technology (CIT)	Engineering	75500.0	123000.0	75500.0	104000.0	161000.0	75500.0

In [165]:
```
a=df2.loc[(df2['School Type'] == 'Party')]
b=df2.loc[(df2['School Type'] == 'Ivy League')]
a['Starting Median Salary'].min()
```

Out[165]:
```
41300.0
```

In [166]:
```
a['Starting Median Salary'].max()
```

Out[166]:
```
52900.0
```

In [168]:
```
b['Starting Median Salary'].min()
```

Out[168]:
```
56200.0
```

In [169]:
```
b['Starting Median Salary'].max()
```

Out[169]:
```
66500.0
```

In [172]:
```
(df2.loc[df2['Starting Median Salary'] == df2['Starting Median Salary']
.max()])['School Type']
```

Out[172]:
```
1    Engineering
Name: School Type, dtype: object
```

In [173]:
```
(df2.loc[(df2['Starting Median Salary'] > 65000)])['School Name'].uniqu
e()
```

Out[173]:
```
array(['Massachusetts Institute of Technology (MIT)',
       'California Institute of Technology (CIT)', 'Harvey Mudd Colleg
e',
       'Princeton University'], dtype=object)
```

In [ ]:
```

```