In [2]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
plt.rcParams['figure.figsize'] = (10.0, 8.0)
import seaborn as sns
from scipy import stats
from scipy.stats import norm
```

In [3]:
```
df = pd.read_csv('most.csv')
df.head(5)
```

```
C:\Users\Shakena Ford\Anaconda3\lib\site-packages\IPython\core\interactiveshel
l.py:3049: DtypeWarning: Columns (6,9,31,1608,1619,1620,1621,1622,1623,1624,162
5,1626,1627,1628,1629,1688,1689,1690,1691,1692,1703,1704,1725,1726,1727,1728,17
29,1743,1815,1816,1817,1818,1823,1824,1830,1831,1879,1880,1881,1882,1883,1884,1
885,1886,1887,1888,1889,1890,1891,1892,1893,1894,1895,1896,1897,1898,1909,1910,
1911,1912,1913,1957,1958,1959,1960,1961,1962,1963,1964,1965,1966,1967,1968,196
9,1970,1971,1972,1973,1974,1975,1976) have mixed types. Specify dtype option on
import or set low_memory=False.
  interactivity=interactivity, compiler=compiler, result=result)
```

Out[3]:

	UNITID	OPEID	OPEID6	INSTNM	CITY	STABBR	ZIP	ACCREDAGENCY	
0	100654	100200	1002	Alabama A & M University	Normal	AL	35762	Southern Association of Colleges and Schools C...	
1	100663	105200	1052	University of Alabama at Birmingham	Birmingham	AL	35294-0110	Southern Association of Colleges and Schools C...	
2	100690	2503400	25034	Amridge University	Montgomery	AL	36117-3553	Southern Association of Colleges and Schools C...	www.am
3	100706	105500	1055	University of Alabama in Huntsville	Huntsville	AL	35899	Southern Association of Colleges and Schools C...	
4	100724	100500	1005	Alabama State University	Montgomery	AL	36104-0271	Southern Association of Colleges and Schools C...	

5 rows × 1977 columns

In [5]:
```
#question 1
df[['INSTNM', 'COSTT4_A']].sort_values('INSTNM', ascending=False).nlargest(10, '
#Answer the most costly college Aviator College of Aeronautical Science and Te
```

Out[5]:

	INSTNM	COSTT4_A
5161	Aviator College of Aeronautical Science and Te...	93704.0
925	University of Chicago	72717.0
2290	Jewish Theological Seminary of America	72120.0
2222	Columbia University in the City of New York	71972.0
324	Harvey Mudd College	71917.0
1022	Northwestern University	70317.0
1948	Washington University in St Louis	69754.0
492	University of Southern California	69547.0
2027	Dartmouth College	69474.0
2971	Drexel University	69462.0

for cost for full program college

CIPTFBS1

In [6]:
```
#question 1
df[['INSTNM', 'COSTT4_A']].sort_values('COSTT4_A', ascending=0).nsmallest( 10,'C
#Answer the cheapest college is Instituto Tecnologico de Puerto Rico-Recinto d..

#question 1 answer pt2
#how to find college that cost most
#df[["INSTNM", "COSTT4_A"]].min()

#question 1 answer
#The most costly college is eClips school of cosmetology
#df[["INSTNM", "COSTT4_A"]].max()
```

Out[6]:

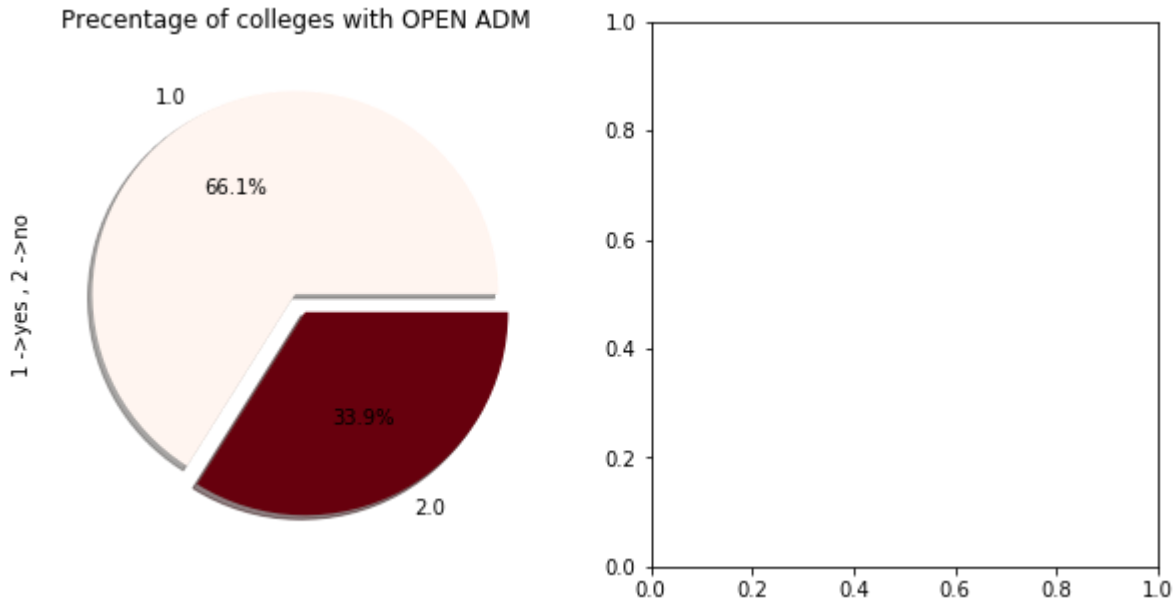
	INSTNM	COSTT4_A
6589	Purdue University Northwest	0.0
3921	Instituto Tecnologico de Puerto Rico-Recinto d...	3930.0
3922	Instituto Tecnologico de Puerto Rico-Recinto d...	4007.0
3912	Instituto Tecnologico de Puerto Rico-Recinto d...	5025.0
2481	Cleveland Community College	5185.0
5377	Escuela De Troqueleria Y Herramentaje	5481.0
26	J F Ingram State Technical College	5496.0
3906	Colegio Universitario de San Juan	5950.0
3951	Palau Community College	6085.0
705	Indian River State College	6276.0

In [ ]:

In [ ]:

In [7]:
```
#question 4 answer (percenatge of open admission)
fig,ax = plt.subplots(1,2,figsize=(10,5))
ax[0].set_title('Precentage of colleges with OPEN ADM')
df['OPENADMP'].value_counts().plot.pie(explode=[0.0,0.1],autopct='%.1f%%',shadow=
ax[0].set_ylabel("1 ->yes , 2 ->no")
```

Out[7]: Text(0, 0.5, '1 ->yes , 2 ->no')

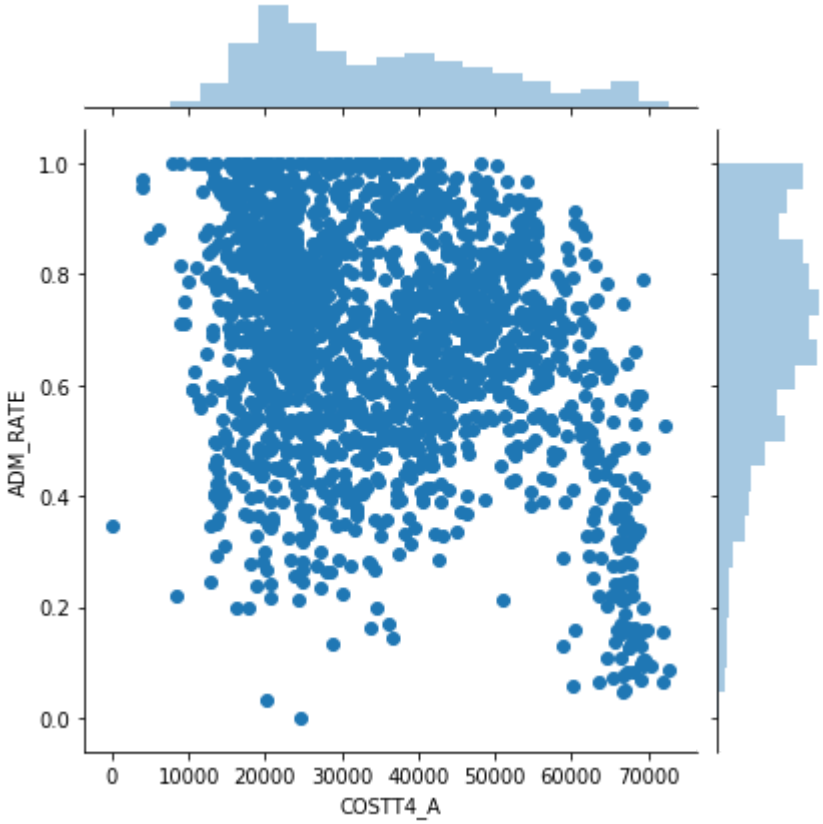


In [8]:
```
#how to compare two columns that have correlations
#question 6 answer
df['ADM_RATE'].corr(df['SAT_AVG'])
```

Out[8]: -0.4132664998444278

In [9]:
```
#question 5 answer
sns.jointplot(x=df['COSTT4_A'], y=df['ADM_RATE'])
```

Out[9]:
```
<seaborn.axisgrid.JointGrid at 0x24be97f2630>
```



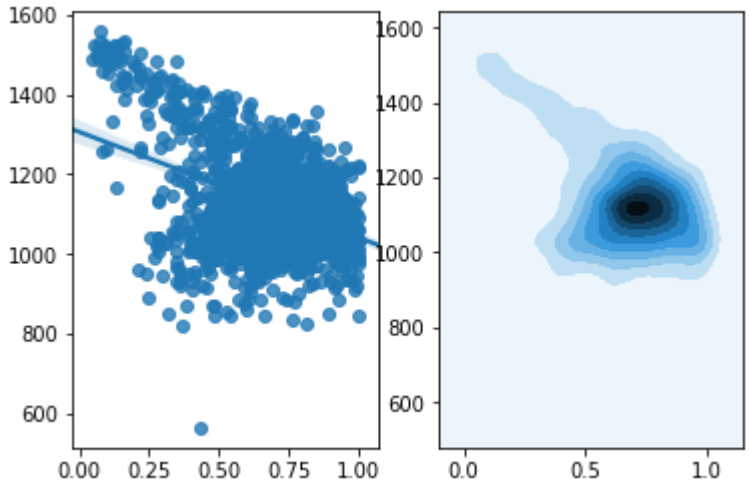
In [10]:
```
#question 5 answer  for better indepth look
fig = plt.figure();
ax1 = fig.add_subplot(121);
ax2 = fig.add_subplot(122);

sns.jointplot("ADM_RATE", "SAT_AVG", data=df, kind='reg', ax=ax1)
sns.jointplot("ADM_RATE", "SAT_AVG", data=df, kind='kde', ax=ax2)

JG1 = sns.jointplot("ADM_RATE", "SAT_AVG", data=df, kind='reg')
JG2 = sns.jointplot("ADM_RATE", "SAT_AVG", data=df, kind='kde')

#subplots migration
f = plt.figure()
for J in [JG1, JG2]:
    for A in J.fig.axes:
        f._axstack.add(f._make_key(A), A)

#subplots size adjustment
f.axes[0].set_position([0.05, 0.05, 0.4,  0.4])
f.axes[1].set_position([0.05, 0.45, 0.4,  0.05])
f.axes[2].set_position([0.45, 0.05, 0.05, 0.4])
f.axes[3].set_position([0.55, 0.05, 0.4,  0.4])
f.axes[4].set_position([0.55, 0.45, 0.4,  0.05])
f.axes[5].set_position([0.95, 0.05, 0.05, 0.4])
```



In [11]:
```
#income of the familes
df["FAMINC"].head(10)
```

Out[11]:
```
0    32362.826114
1    51306.674306
2    21079.472973
3    61096.588949
4    31684.382188
5    91846.749624
6    30767.764486
7     38479.38737
8    41987.986288
9    92148.626516
Name: FAMINC, dtype: object
```

In [12]:
```
#
df["MD_FAMINC"].head(10)
```

Out[12]:
```
0      23553
1      34489
2    15033.5
3      44787
4    22080.5
5    66733.5
6      22217
7    29645.5
8    29671.5
9      72031
Name: MD_FAMINC, dtype: object
```

In [22]:
```
#question 2
df[['STABBR', 'FAMINC']].mean
#Answer the most costly college Aviator College of Aeronautical Science and Te
```

Out[22]:
```
<bound method DataFrame.mean of      STABBR            FAMINC
0         AL       32362.826114
1         AL       51306.674306
2         AL       21079.472973
3         AL       61096.588949
4         AL       31684.382188
5         AL       91846.749624
6         AL       30767.764486
7         AL        38479.38737
8         AL       41987.986288
9         AL       92148.626516
10        AL       86672.871041
11        AL       26584.324786
12        AL       20920.501247
13        AL       29130.921239
14        AL       30336.675829
15        AL       36117.937329
16        AL       36952.206116
17        AL       19449.725019
18        AL        22762.19469
19        AL       19125.651634
20        AL       36705.769088
21        AL       15431.714286
22        AL       26184.228503
23        AL       53792.633136
24        AL           33173.64
25        AL       16557.179487
26        AL        321.3853211
27        AL       48404.346827
28        AL       19863.224599
29        AL       36652.478692
...      ...                ...
7028      NJ  PrivacySuppressed
7029      LA  PrivacySuppressed
7030      DE  PrivacySuppressed
7031      TN  PrivacySuppressed
7032      TX  PrivacySuppressed
7033      IL  PrivacySuppressed
7034      LA  PrivacySuppressed
7035      PA  PrivacySuppressed
7036      CO  PrivacySuppressed
7037      TN  PrivacySuppressed
7038      TX                NaN
7039      TX                NaN
7040      DC                NaN
7041      CA       27233.927419
7042      OH       52239.642336
7043      CA       31730.632653
7044      KS       35850.098826
7045      UT                NaN
7046      AR                NaN
```

```
7047     NY               NaN
7048     OH      35710.470588
7049     PA      22422.557252
7050     AZ               NaN
7051     AZ               NaN
7052     AZ               NaN
7053     AZ               NaN
7054     CA      25486.253589
7055     CA      21764.237265
7056     MA      15602.133333
7057     FL               NaN

[7058 rows x 2 columns]>
```

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: