Excercise_8_solutions

May 30, 2019

1 K Means Clustering Project

For this project we will attempt to use KMeans Clustering to cluster Universities into to two groups, Private and Public. #### It is very important to note, we actually have the labels for this data set, but we will NOT use them for the KMeans clustering algorithm, since that is an unsupervised learning algorithm.

1.0.1 The Data

We will use a data frame with 777 observations on the following 18 variables. * Private A factor with levels No and Yes indicating private or public university * Apps Number of applications received * Accept Number of applications accepted * Enroll Number of new students enrolled * Top10perc Pct. new students from top 10% of H.S. class * Top25perc Pct. new students from top 25% of H.S. class * F.Undergrad Number of fulltime undergraduates * P.Undergrad Number of parttime undergraduates * Outstate Out-of-state tuition * Room.Board Room and board costs * Books Estimated book costs * Personal Estimated personal spending * PhD Pct. of faculty with Ph.D.'s * Terminal Pct. of faculty with terminal degree * S.F.Ratio Student/faculty ratio * perc.alumni Pct. alumni who donate * Expend Instructional expenditure per student * Grad.Rate Graduation rate

1.0.2 Import Libraries

Import the libraries you usually use for data analysis.

```
In [1]: import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
        import seaborn as sns
        %matplotlib inline
```

1.0.3 Get the Data

Read in the College_Data file using read_csv. Figure out how to set the first column as the index.

```
In [2]: df = pd.read_csv('College_Data',index_col=0)
```

Check the head of the data

```
In [3]: df.head()
```

```
Out[3]:                               Private  Apps  Accept  Enroll  Top10perc  \
        Abilene Christian University      Yes  1660    1232     721         23
        Adelphi University                Yes  2186    1924     512         16
        Adrian College                    Yes  1428    1097     336         22
        Agnes Scott College               Yes   417     349     137         60
        Alaska Pacific University         Yes   193     146      55         16

                                      Top25perc  F.Undergrad  P.Undergrad  Outstate  \
        Abilene Christian University         52         2885          537      7440
        Adelphi University                   29         2683         1227     12280
        Adrian College                       50         1036           99     11250
        Agnes Scott College                  89          510           63     12960
        Alaska Pacific University            44          249          869      7560

                                      Room.Board  Books  Personal  PhD  Terminal  \
        Abilene Christian University        3300    450      2200   70        78
        Adelphi University                  6450    750      1500   29        30
        Adrian College                      3750    400      1165   53        66
        Agnes Scott College                 5450    450       875   92        97
        Alaska Pacific University           4120    800      1500   76        72

                                      S.F.Ratio  perc.alumni  Expend  Grad.Rate
        Abilene Christian University       18.1           12    7041         60
        Adelphi University                 12.2           16   10527         56
        Adrian College                     12.9           30    8735         54
        Agnes Scott College                 7.7           37   19016         59
        Alaska Pacific University          11.9            2   10922         15
```

Check the info() and describe() methods on the data,

```
In [4]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 777 entries, Abilene Christian University to York College of Pennsylvania
Data columns (total 18 columns):
Private        777 non-null object
Apps           777 non-null int64
Accept         777 non-null int64
Enroll         777 non-null int64
Top10perc      777 non-null int64
Top25perc      777 non-null int64
F.Undergrad    777 non-null int64
P.Undergrad    777 non-null int64
Outstate       777 non-null int64
Room.Board     777 non-null int64
```

```
Books          777 non-null int64
Personal       777 non-null int64
PhD            777 non-null int64
Terminal       777 non-null int64
S.F.Ratio      777 non-null float64
perc.alumni    777 non-null int64
Expend         777 non-null int64
Grad.Rate      777 non-null int64
dtypes: float64(1), int64(16), object(1)
memory usage: 115.3+ KB
```

In [5]: `df.describe()`

Out[5]:

	Apps	Accept	Enroll	Top10perc	Top25perc
count	777.000000	777.000000	777.000000	777.000000	777.000000
mean	3001.638353	2018.804376	779.972973	27.558559	55.796654
std	3870.201484	2451.113971	929.176190	17.640364	19.804778
min	81.000000	72.000000	35.000000	1.000000	9.000000
25%	776.000000	604.000000	242.000000	15.000000	41.000000
50%	1558.000000	1110.000000	434.000000	23.000000	54.000000
75%	3624.000000	2424.000000	902.000000	35.000000	69.000000
max	48094.000000	26330.000000	6392.000000	96.000000	100.000000

	F.Undergrad	P.Undergrad	Outstate	Room.Board	Books
count	777.000000	777.000000	777.000000	777.000000	777.000000
mean	3699.907336	855.298584	10440.669241	4357.526384	549.380952
std	4850.420531	1522.431887	4023.016484	1096.696416	165.105360
min	139.000000	1.000000	2340.000000	1780.000000	96.000000
25%	992.000000	95.000000	7320.000000	3597.000000	470.000000
50%	1707.000000	353.000000	9990.000000	4200.000000	500.000000
75%	4005.000000	967.000000	12925.000000	5050.000000	600.000000
max	31643.000000	21836.000000	21700.000000	8124.000000	2340.000000

	Personal	PhD	Terminal	S.F.Ratio	perc.alumni
count	777.000000	777.000000	777.000000	777.000000	777.000000
mean	1340.642214	72.660232	79.702703	14.089704	22.743887
std	677.071454	16.328155	14.722359	3.958349	12.391801
min	250.000000	8.000000	24.000000	2.500000	0.000000
25%	850.000000	62.000000	71.000000	11.500000	13.000000
50%	1200.000000	75.000000	82.000000	13.600000	21.000000
75%	1700.000000	85.000000	92.000000	16.500000	31.000000
max	6800.000000	103.000000	100.000000	39.800000	64.000000

	Expend	Grad.Rate
count	777.000000	777.00000
mean	9660.171171	65.46332
std	5221.768440	17.17771

```
min      3186.000000   10.00000
25%      6751.000000   53.00000
50%      8377.000000   65.00000
75%     10830.000000   78.00000
max     56233.000000  118.00000
```

1.0.4 EDA

It's time to create some data visualizations! Create a scatterplot of Grad.Rate versus Room.Board where the points are colored by the Private column.

```
In [6]: sns.set_style('whitegrid')
        sns.lmplot(y='Grad.Rate', x='Room.Board', hue='Private', palette='coolwarm', size=7, f
                   aspect=1)
```

```
/home/kamil/anaconda3/lib/python3.7/site-packages/seaborn/regression.py:546: UserWarning: The
  warnings.warn(msg, UserWarning)
```

```
Out[6]: <seaborn.axisgrid.FacetGrid at 0x7f4fdbf655c0>
```

Create a scatterplot of F.Undergrad versus Outstate where the points are colored by the Private column.

```
In [7]: sns.set_style('whitegrid')
        sns.lmplot(y='F.Undergrad', x='Outstate', hue='Private', palette='coolwarm', size=7, fi
                   aspect=1)
```

```
/home/kamil/anaconda3/lib/python3.7/site-packages/seaborn/regression.py:546: UserWarning: The
  warnings.warn(msg, UserWarning)
```

```
Out[7]: <seaborn.axisgrid.FacetGrid at 0x7f4fc00f3c50>
```

Create a stacked histogram showing Out of State Tuition based on the Private column. Try doing this using sns.FacetGrid. If that is too tricky, see if you can do it just by using two instances of pandas.plot(kind='hist').

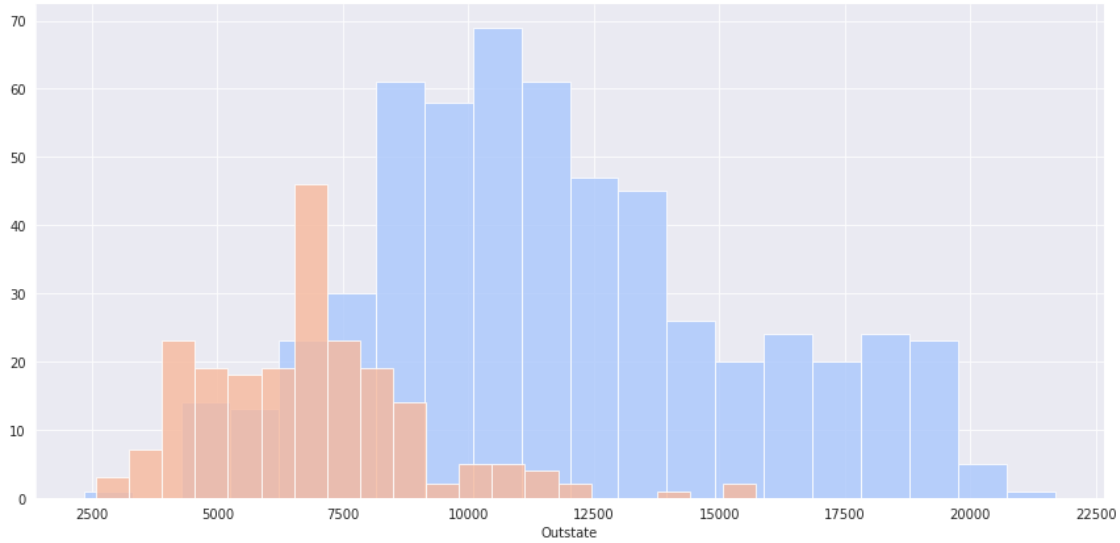
```
In [8]: df.head()
```

Out[8]:

	Private	Apps	Accept	Enroll	Top10perc	Top25perc	F.Undergrad	P.Undergrad	Outstate	Room.Board	Books	Personal	PhD	Terminal	S.F.Ratio	perc.alumni	Expend	Grad.Rate
Abilene Christian University	Yes	1660	1232	721	23	52	2885	537	7440	3300	450	2200	70	78	18.1	12	7041	60
Adelphi University	Yes	2186	1924	512	16	29	2683	1227	12280	6450	750	1500	29	30	12.2	16	10527	56
Adrian College	Yes	1428	1097	336	22	50	1036	99	11250	3750	400	1165	53	66	12.9	30	8735	54
Agnes Scott College	Yes	417	349	137	60	89	510	63	12960	5450	450	875	92	97	7.7	37	19016	59
Alaska Pacific University	Yes	193	146	55	16	44	249	869	7560	4120	800	1500	76	72	11.9	2	10922	15

```
In [15]: sns.set_style('darkgrid')
         g = sns.FacetGrid(data=df,hue='Private', palette='coolwarm', size = 6, aspect=2 )
         g = g.map(plt.hist, 'Outstate', bins=20, alpha=0.8)
```

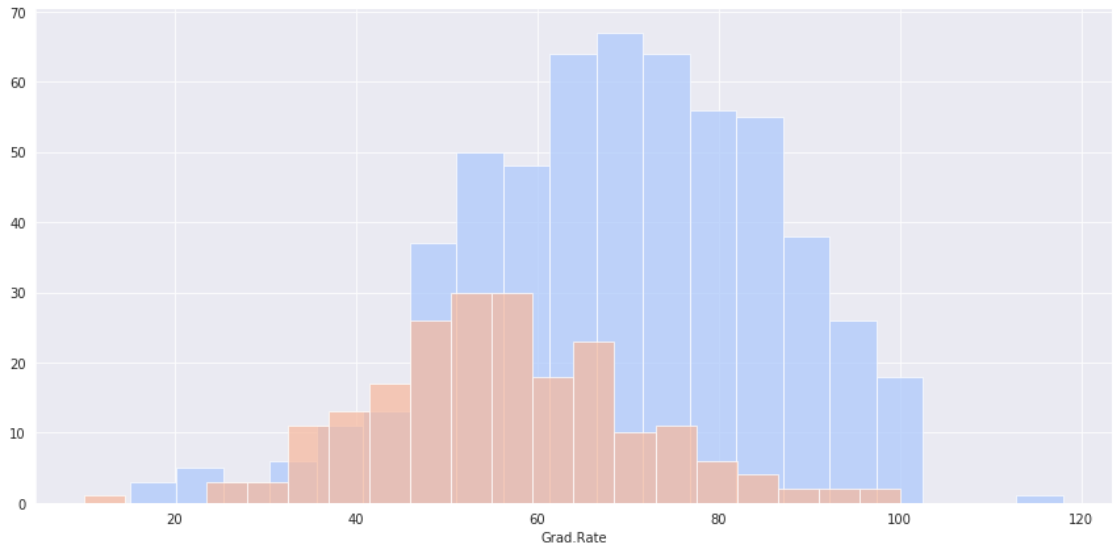
```
/home/kamil/anaconda3/lib/python3.7/site-packages/seaborn/axisgrid.py:230: UserWarning: The `si
  warnings.warn(msg, UserWarning)
```



Create a similar histogram for the Grad.Rate column.

```
In [16]: sns.set_style('darkgrid')
         g = sns.FacetGrid(df,hue="Private",palette='coolwarm',size=6,aspect=2)
         g = g.map(plt.hist,'Grad.Rate',bins=20,alpha=0.7)
```

```
/home/kamil/anaconda3/lib/python3.7/site-packages/seaborn/axisgrid.py:230: UserWarning: The `si
  warnings.warn(msg, UserWarning)
```



Notice how there seems to be a private school with a graduation rate of higher than 100%.What is the name of that school?

```
In [17]: df[df['Grad.Rate'] > 100]
```

```
Out[17]:                    Private  Apps  Accept  Enroll  Top10perc  Top25perc  \
         Cazenovia College      Yes  3847    3433     527          9         35

                            F.Undergrad  P.Undergrad  Outstate  Room.Board  Books  \
         Cazenovia College          1010           12      9384        4840    600

                            Personal  PhD  Terminal  S.F.Ratio  perc.alumni  Expend  \
         Cazenovia College       500   22        47       14.3           20    7697

                            Grad.Rate
         Cazenovia College        118
```

Set that school's graduation rate to 100 so it makes sense. You may get a warning not an error) when doing this operation, so use dataframe operations or just re-do the histogram visualization to make sure it actually went through.

```
In [18]: df['Grad.Rate']['Cazenovia College'] = 100
```

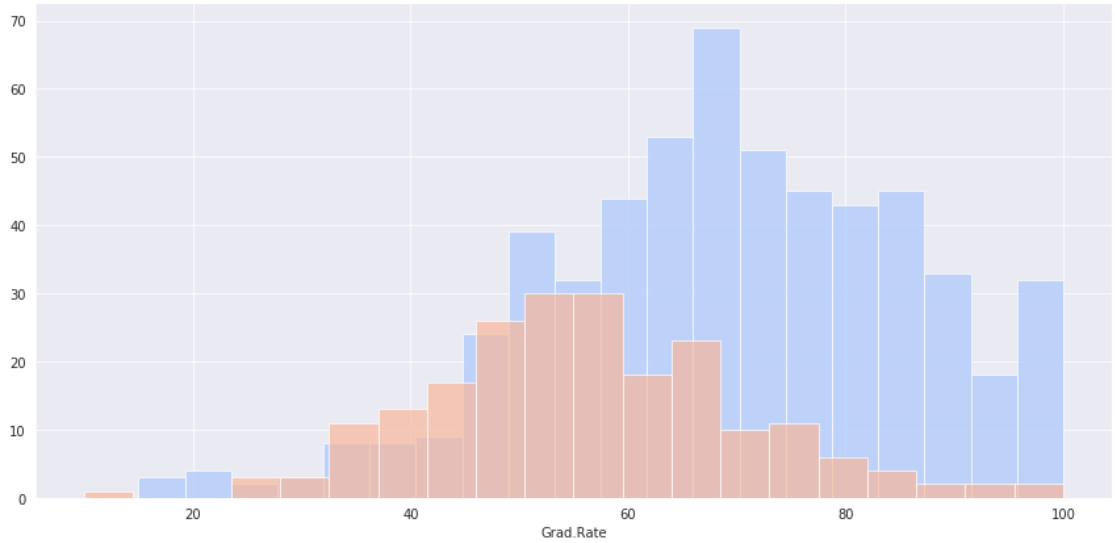
```
/home/kamil/anaconda3/lib/python3.7/site-packages/ipykernel_launcher.py:1: SettingWithCopyWarni
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html
  """Entry point for launching an IPython kernel.
```

```
In [19]: df[df['Grad.Rate'] > 100]
```

```
Out[19]: Empty DataFrame
         Columns: [Private, Apps, Accept, Enroll, Top10perc, Top25perc, F.Undergrad, P.Undergra
         Index: []
```

```
In [20]: sns.set_style('darkgrid')
         g = sns.FacetGrid(df,hue="Private",palette='coolwarm',size=6,aspect=2)
         g = g.map(plt.hist,'Grad.Rate',bins=20,alpha=0.7)
```

```
/home/kamil/anaconda3/lib/python3.7/site-packages/seaborn/axisgrid.py:230: UserWarning: The `si
  warnings.warn(msg, UserWarning)
```



1.0.5 K Means Cluster Creation

Now it is time to create the Cluster labels! - Import KMeans from SciKit Learn.

```
In [22]: from sklearn.cluster import KMeans
```

- Create an instance of a K Means model with 2 clusters.

```
In [23]: kmeans = KMeans(n_clusters=2)
```

Fit the model to all the data except for the Private label.

```
In [24]: df.head()
```

```
Out[24]:                              Private  Apps  Accept  Enroll  Top10perc  \
         Abilene Christian University     Yes  1660    1232     721         23
         Adelphi University               Yes  2186    1924     512         16
         Adrian College                   Yes  1428    1097     336         22
         Agnes Scott College              Yes   417     349     137         60
         Alaska Pacific University        Yes   193     146      55         16

                                       Top25perc  F.Undergrad  P.Undergrad  Outstate  \
         Abilene Christian University         52         2885          537      7440
         Adelphi University                   29         2683         1227     12280
         Adrian College                       50         1036           99     11250
         Agnes Scott College                  89          510           63     12960
         Alaska Pacific University            44          249          869      7560

                                       Room.Board  Books  Personal  PhD  Terminal  \
```

```
                                Abilene Christian University        3300    450      2200   70         78
                                Adelphi University                  6450    750      1500   29         30
                                Adrian College                      3750    400      1165   53         66
                                Agnes Scott College                 5450    450       875   92         97
                                Alaska Pacific University           4120    800      1500   76         72

                                                                S.F.Ratio  perc.alumni  Expend  Grad.Rate
                                Abilene Christian University         18.1           12    7041         60
                                Adelphi University                   12.2           16   10527         56
                                Adrian College                       12.9           30    8735         54
                                Agnes Scott College                   7.7           37   19016         59
                                Alaska Pacific University            11.9            2   10922         15
```

```
In [25]: kmeans.fit(df.drop('Private', axis=1))
```

```
Out[25]: KMeans(algorithm='auto', copy_x=True, init='k-means++', max_iter=300,
             n_clusters=2, n_init=10, n_jobs=None, precompute_distances='auto',
             random_state=None, tol=0.0001, verbose=0)
```

What are the cluster center vectors?

```
In [26]: kmeans.cluster_centers_
```

```
Out[26]: array([[1.81323468e+03, 1.28716592e+03, 4.91044843e+02, 2.53094170e+01,
                 5.34708520e+01, 2.18854858e+03, 5.95458894e+02, 1.03957085e+04,
                 4.31136472e+03, 5.41982063e+02, 1.28033632e+03, 7.04424514e+01,
                 7.78251121e+01, 1.40997010e+01, 2.31748879e+01, 8.93204634e+03,
                 6.50926756e+01],
                [1.03631389e+04, 6.55089815e+03, 2.56972222e+03, 4.14907407e+01,
                 7.02037037e+01, 1.30619352e+04, 2.46486111e+03, 1.07191759e+04,
                 4.64347222e+03, 5.95212963e+02, 1.71420370e+03, 8.63981481e+01,
                 9.13333333e+01, 1.40277778e+01, 2.00740741e+01, 1.41705000e+04,
                 6.75925926e+01]])
```

1.0.6 Evaluation

There is no perfect way to evaluate clustering if you don't have the labels, however since this is just an exercise, we do have the labels, so we take advantage of this to evaluate our clusters, keep in mind, you usually won't have this luxury in the real world.

Create a new column for df called 'Cluster', which is a 1 for a Private school, and a 0 for a public school.

```
In [30]: type(df['Private'][1])
```

```
Out[30]: str
```

```
In [34]: def convert(column):
             if column == 'Yes':
                 return 1
             else:
                 return 0
```

```
In [35]: df['Cluster'] = df['Private'].apply(convert)

In [37]: df
```

Out[37]:

	Private	Apps	Accept	Enroll	\
Abilene Christian University	Yes	1660	1232	721	
Adelphi University	Yes	2186	1924	512	
Adrian College	Yes	1428	1097	336	
Agnes Scott College	Yes	417	349	137	
Alaska Pacific University	Yes	193	146	55	
Albertson College	Yes	587	479	158	
Albertus Magnus College	Yes	353	340	103	
Albion College	Yes	1899	1720	489	
Albright College	Yes	1038	839	227	
Alderson-Broaddus College	Yes	582	498	172	
Alfred University	Yes	1732	1425	472	
Allegheny College	Yes	2652	1900	484	
Allentown Coll. of St. Francis de Sales	Yes	1179	780	290	
Alma College	Yes	1267	1080	385	
Alverno College	Yes	494	313	157	
American International College	Yes	1420	1093	220	
Amherst College	Yes	4302	992	418	
Anderson University	Yes	1216	908	423	
Andrews University	Yes	1130	704	322	
Angelo State University	No	3540	2001	1016	
Antioch University	Yes	713	661	252	
Appalachian State University	No	7313	4664	1910	
Aquinas College	Yes	619	516	219	
Arizona State University Main campus	No	12809	10308	3761	
Arkansas College (Lyon College)	Yes	708	334	166	
Arkansas Tech University	No	1734	1729	951	
Assumption College	Yes	2135	1700	491	
Auburn University-Main Campus	No	7548	6791	3070	
Augsburg College	Yes	662	513	257	
Augustana College IL	Yes	1879	1658	497	
...	...	...	...	...	
Westfield State College	No	3100	2150	825	
Westminster College MO	Yes	662	553	184	
Westminster College	Yes	996	866	377	
Westminster College of Salt Lake City	Yes	917	720	213	
Westmont College	No	950	713	351	
Wheaton College IL	Yes	1432	920	548	
Westminster College PA	Yes	1738	1373	417	
Wheeling Jesuit College	Yes	903	755	213	
Whitman College	Yes	1861	998	359	
Whittier College	Yes	1681	1069	344	
Whitworth College	Yes	1121	926	372	
Widener University	Yes	2139	1492	502	

Wilkes University	Yes	1631	1431	434
Willamette University	Yes	1658	1327	395
William Jewell College	Yes	663	547	315
William Woods University	Yes	469	435	227
Williams College	Yes	4186	1245	526
Wilson College	Yes	167	130	46
Wingate College	Yes	1239	1017	383
Winona State University	No	3325	2047	1301
Winthrop University	No	2320	1805	769
Wisconsin Lutheran College	Yes	152	128	75
Wittenberg University	Yes	1979	1739	575
Wofford College	Yes	1501	935	273
Worcester Polytechnic Institute	Yes	2768	2314	682
Worcester State College	No	2197	1515	543
Xavier University	Yes	1959	1805	695
Xavier University of Louisiana	Yes	2097	1915	695
Yale University	Yes	10705	2453	1317
York College of Pennsylvania	Yes	2989	1855	691

	Top10perc	Top25perc	F.Undergrad	\
Abilene Christian University	23	52	2885	
Adelphi University	16	29	2683	
Adrian College	22	50	1036	
Agnes Scott College	60	89	510	
Alaska Pacific University	16	44	249	
Albertson College	38	62	678	
Albertus Magnus College	17	45	416	
Albion College	37	68	1594	
Albright College	30	63	973	
Alderson-Broaddus College	21	44	799	
Alfred University	37	75	1830	
Allegheny College	44	77	1707	
Allentown Coll. of St. Francis de Sales	38	64	1130	
Alma College	44	73	1306	
Alverno College	23	46	1317	
American International College	9	22	1018	
Amherst College	83	96	1593	
Anderson University	19	40	1819	
Andrews University	14	23	1586	
Angelo State University	24	54	4190	
Antioch University	25	44	712	
Appalachian State University	20	63	9940	
Aquinas College	20	51	1251	
Arizona State University Main campus	24	49	22593	
Arkansas College (Lyon College)	46	74	530	
Arkansas Tech University	12	52	3602	
Assumption College	23	59	1708	
Auburn University-Main Campus	25	57	16262	

Augsburg College	12	30	2074
Augustana College IL	36	69	1950
...	...	...	...
Westfield State College	3	20	3234
Westminster College MO	20	43	665
Westminster College	29	58	1411
Westminster College of Salt Lake City	21	60	979
Westmont College	42	72	1276
Wheaton College IL	56	84	2200
Westminster College PA	21	55	1335
Wheeling Jesuit College	15	49	971
Whitman College	45	77	1220
Whittier College	35	63	1235
Whitworth College	43	70	1270
Widener University	24	64	2186
Wilkes University	15	36	1803
Willamette University	49	80	1595
William Jewell College	32	67	1279
William Woods University	17	39	851
Williams College	81	96	1988
Wilson College	16	50	199
Wingate College	10	34	1207
Winona State University	20	45	5800
Winthrop University	24	61	3395
Wisconsin Lutheran College	17	41	282
Wittenberg University	42	68	1980
Wofford College	51	83	1059
Worcester Polytechnic Institute	49	86	2802
Worcester State College	4	26	3089
Xavier University	24	47	2849
Xavier University of Louisiana	34	61	2793
Yale University	95	99	5217
York College of Pennsylvania	28	63	2988

	P.Undergrad	Outstate	Room.Board	\
Abilene Christian University	537	7440	3300	
Adelphi University	1227	12280	6450	
Adrian College	99	11250	3750	
Agnes Scott College	63	12960	5450	
Alaska Pacific University	869	7560	4120	
Albertson College	41	13500	3335	
Albertus Magnus College	230	13290	5720	
Albion College	32	13868	4826	
Albright College	306	15595	4400	
Alderson-Broaddus College	78	10468	3380	
Alfred University	110	16548	5406	
Allegheny College	44	17080	4440	
Allentown Coll. of St. Francis de Sales	638	9690	4785	

Alma College	28	12572	4552
Alverno College	1235	8352	3640
American International College	287	8700	4780
Amherst College	5	19760	5300
Anderson University	281	10100	3520
Andrews University	326	9996	3090
Angelo State University	1512	5130	3592
Antioch University	23	15476	3336
Appalachian State University	1035	6806	2540
Aquinas College	767	11208	4124
Arizona State University Main campus	7585	7434	4850
Arkansas College (Lyon College)	182	8644	3922
Arkansas Tech University	939	3460	2650
Assumption College	689	12000	5920
Auburn University-Main Campus	1716	6300	3933
Augsburg College	726	11902	4372
Augustana College IL	38	13353	4173
...	...	...	...
Westfield State College	941	5542	3788
Westminster College MO	37	10720	4050
Westminster College	72	12065	3615
Westminster College of Salt Lake City	743	8820	4050
Westmont College	9	14320	5304
Wheaton College IL	56	11480	4200
Westminster College PA	30	18460	5970
Wheeling Jesuit College	305	10500	4545
Whitman College	46	16670	4900
Whittier College	30	16249	5699
Whitworth College	160	12660	4500
Widener University	2171	12350	5370
Wilkes University	603	11150	5130
Willamette University	159	14800	4620
William Jewell College	75	10060	2970
William Woods University	120	10535	4365
Williams College	29	19629	5790
Wilson College	676	11428	5084
Wingate College	157	7820	3400
Winona State University	872	4200	2700
Winthrop University	670	6400	3392
Wisconsin Lutheran College	22	9100	3700
Wittenberg University	144	15948	4404
Wofford College	34	12680	4150
Worcester Polytechnic Institute	86	15884	5370
Worcester State College	2029	6797	3900
Xavier University	1107	11520	4960
Xavier University of Louisiana	166	6900	4200
Yale University	83	19840	6510
York College of Pennsylvania	1726	4990	3560

	Books	Personal	PhD	Terminal	\
Abilene Christian University	450	2200	70	78	
Adelphi University	750	1500	29	30	
Adrian College	400	1165	53	66	
Agnes Scott College	450	875	92	97	
Alaska Pacific University	800	1500	76	72	
Albertson College	500	675	67	73	
Albertus Magnus College	500	1500	90	93	
Albion College	450	850	89	100	
Albright College	300	500	79	84	
Alderson-Broaddus College	660	1800	40	41	
Alfred University	500	600	82	88	
Allegheny College	400	600	73	91	
Allentown Coll. of St. Francis de Sales	600	1000	60	84	
Alma College	400	400	79	87	
Alverno College	650	2449	36	69	
American International College	450	1400	78	84	
Amherst College	660	1598	93	98	
Anderson University	550	1100	48	61	
Andrews University	900	1320	62	66	
Angelo State University	500	2000	60	62	
Antioch University	400	1100	69	82	
Appalachian State University	96	2000	83	96	
Aquinas College	350	1615	55	65	
Arizona State University Main campus	700	2100	88	93	
Arkansas College (Lyon College)	500	800	79	88	
Arkansas Tech University	450	1000	57	60	
Assumption College	500	500	93	93	
Auburn University-Main Campus	600	1908	85	91	
Augsburg College	540	950	65	65	
Augustana College IL	540	821	78	83	
...	...	...	...	...	
Westfield State College	500	1300	75	79	
Westminster College MO	600	1650	66	70	
Westminster College	430	685	62	78	
Westminster College of Salt Lake City	600	2025	68	83	
Westmont College	490	1410	77	77	
Wheaton College IL	530	1400	81	83	
Westminster College PA	700	850	92	96	
Wheeling Jesuit College	600	600	66	71	
Whitman College	750	800	80	83	
Whittier College	500	1998	84	92	
Whitworth College	678	2424	80	80	
Widener University	500	1350	88	86	
Wilkes University	550	1260	78	92	
Willamette University	400	790	91	94	
William Jewell College	500	2600	74	80	

William Woods University	550	3700	39	66
Williams College	500	1200	94	99
Wilson College	450	475	67	76
Wingate College	550	1550	69	81
Winona State University	300	1200	53	60
Winthrop University	580	2150	71	80
Wisconsin Lutheran College	500	1400	48	48
Wittenberg University	400	800	82	95
Wofford College	605	1440	91	92
Worcester Polytechnic Institute	530	730	92	94
Worcester State College	500	1200	60	60
Xavier University	600	1250	73	75
Xavier University of Louisiana	617	781	67	75
Yale University	630	2115	96	96
York College of Pennsylvania	500	1250	75	75

	S.F.Ratio	perc.alumni	Expend	\
Abilene Christian University	18.1	12	7041	
Adelphi University	12.2	16	10527	
Adrian College	12.9	30	8735	
Agnes Scott College	7.7	37	19016	
Alaska Pacific University	11.9	2	10922	
Albertson College	9.4	11	9727	
Albertus Magnus College	11.5	26	8861	
Albion College	13.7	37	11487	
Albright College	11.3	23	11644	
Alderson-Broaddus College	11.5	15	8991	
Alfred University	11.3	31	10932	
Allegheny College	9.9	41	11711	
Allentown Coll. of St. Francis de Sales	13.3	21	7940	
Alma College	15.3	32	9305	
Alverno College	11.1	26	8127	
American International College	14.7	19	7355	
Amherst College	8.4	63	21424	
Anderson University	12.1	14	7994	
Andrews University	11.5	18	10908	
Angelo State University	23.1	5	4010	
Antioch University	11.3	35	42926	
Appalachian State University	18.3	14	5854	
Aquinas College	12.7	25	6584	
Arizona State University Main campus	18.9	5	4602	
Arkansas College (Lyon College)	12.6	24	14579	
Arkansas Tech University	19.6	5	4739	
Assumption College	13.8	30	7100	
Auburn University-Main Campus	16.7	18	6642	
Augsburg College	12.8	31	7836	
Augustana College IL	12.7	40	9220	
...	...	...	...	

Westfield State College	15.7	20	4222
Westminster College MO	12.5	20	7925
Westminster College	12.5	41	8596
Westminster College of Salt Lake City	10.5	34	7170
Westmont College	14.9	17	8837
Wheaton College IL	12.7	40	11916
Westminster College PA	13.2	41	22704
Wheeling Jesuit College	14.1	27	7494
Whitman College	10.5	51	13198
Whittier College	13.6	29	11778
Whitworth College	16.9	20	8328
Widener University	12.6	19	9603
Wilkes University	13.3	24	8543
Willamette University	13.3	37	10779
William Jewell College	11.2	19	7885
William Woods University	12.9	16	7438
Williams College	9.0	64	22014
Wilson College	8.3	43	10291
Wingate College	13.9	8	7264
Winona State University	20.2	18	5318
Winthrop University	12.8	26	6729
Wisconsin Lutheran College	8.5	26	8960
Wittenberg University	12.8	29	10414
Wofford College	15.3	42	7875
Worcester Polytechnic Institute	15.2	34	10774
Worcester State College	21.0	14	4469
Xavier University	13.3	31	9189
Xavier University of Louisiana	14.4	20	8323
Yale University	5.8	49	40386
York College of Pennsylvania	18.1	28	4509

	Grad.Rate	Cluster
Abilene Christian University	60	1
Adelphi University	56	1
Adrian College	54	1
Agnes Scott College	59	1
Alaska Pacific University	15	1
Albertson College	55	1
Albertus Magnus College	63	1
Albion College	73	1
Albright College	80	1
Alderson-Broaddus College	52	1
Alfred University	73	1
Allegheny College	76	1
Allentown Coll. of St. Francis de Sales	74	1
Alma College	68	1
Alverno College	55	1
American International College	69	1

Amherst College	100	1
Anderson University	59	1
Andrews University	46	1
Angelo State University	34	0
Antioch University	48	1
Appalachian State University	70	0
Aquinas College	65	1
Arizona State University Main campus	48	0
Arkansas College (Lyon College)	54	1
Arkansas Tech University	48	0
Assumption College	88	1
Auburn University-Main Campus	69	0
Augsburg College	58	1
Augustana College IL	71	1
...	...	...
Westfield State College	65	0
Westminster College MO	62	1
Westminster College	80	1
Westminster College of Salt Lake City	50	1
Westmont College	87	0
Wheaton College IL	85	1
Westminster College PA	71	1
Wheeling Jesuit College	72	1
Whitman College	72	1
Whittier College	52	1
Whitworth College	80	1
Widener University	63	1
Wilkes University	67	1
Willamette University	68	1
William Jewell College	59	1
William Woods University	52	1
Williams College	99	1
Wilson College	67	1
Wingate College	91	1
Winona State University	58	0
Winthrop University	59	0
Wisconsin Lutheran College	50	1
Wittenberg University	78	1
Wofford College	75	1
Worcester Polytechnic Institute	82	1
Worcester State College	40	0
Xavier University	83	1
Xavier University of Louisiana	49	1
Yale University	99	1
York College of Pennsylvania	99	1

[777 rows x 19 columns]

Create a confusion matrix and classification report to see how well the Kmeans clustering worked without being given any labels.

```
In [38]: from sklearn.metrics import classification_report, confusion_matrix

In [39]: print(confusion_matrix(df['Cluster'], kmeans.labels_))

[[138  74]
 [531  34]]


In [40]: print(classification_report(df['Cluster'], kmeans.labels_))

              precision    recall  f1-score   support

           0       0.21      0.65      0.31       212
           1       0.31      0.06      0.10       565

   micro avg       0.22      0.22      0.22       777
   macro avg       0.26      0.36      0.21       777
weighted avg       0.29      0.22      0.16       777
```