Data

The objective of this analysis is to gain insights into the characteristics of colleges and answer key questions related to the educational landscape. By understanding the data, we aim to inform strategies for improving the quality of education and enhancing the overall college experience. The analysis will provide valuable insights and recommendations for stakeholders in the education sector.

Data Description

- Names: Names of various university and colleges
- Apps: Number of applications received
- Accept: Number of applications accepted
- Enroll: Number of new students enrolled
- Top10perc: Percentage of new students from top 10% of Higher Secondary class
- Top25perc: Percentage of new students from top 25% of Higher Secondary class
- F.Undergrad: Number of full-time undergraduate students
- P.Undergrad: Number of part-time undergraduate students
- Outstate: Number of students for whom the particular college or university is Out-of-state tuition
- Room.Board: Cost of Room and board
- Books: Estimated book costs for a student
- Personal: Estimated personal spending for a student
- PhD: Percentage of faculties with Ph.D.'s
- Terminal: Percentage of faculties with terminal degree
- S.F.Ratio: Student/faculty ratio
- perc.alumni: Percentage of alumni who donate
- Expend: The Instructional expenditure per student

- Grad.Rate: Graduation rate

Basic Steps

In [6]:
```
#1-     Display the top 5 rows.
#2-     Display the last 5 rows
#3-     Check the shape of dataset.
#4-     Check the datatypes of each feature.
#5-     Check the Statistical summary
#6-     Check the null values
#7-     Check the duplicate values
#8-     Check the anomalies or wrong entries.
#9-     Check the outliers and their authenticity.
#10-    Do the necessary data cleaning steps like dropping duplicates, unnecessary
```

import Libraries

In [33]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

Loading dataset

In [34]:
```
df=pd.read_excel('1-Education_Post_12th_Standard_New.xlsx')
```

Basic Data Exploration

1- Display the top 5 rows.

In [9]:
```

```

Out[9]:

	Names	Apps	Accept	Enroll	Top10perc	Top25perc	F.Undergrad	P.Undergrad	Outstate
0	Abilene Christian University	1660.0	1232	721.0	23.0	52	2885	537	7440
1	Adelphi University	2186.0	1924	512.0	16.0	29	2683	1227	12280
2	Adrian College	1428.0	1097	336.0	22.0	50	1036	99	11250
3	Agnes Scott College	417.0	349	NaN	60.0	89	510	63	12960
4	Alaska Pacific University	193.0	146	55.0	16.0	44	249	869	7560

Observation

*There are non integer values in S.F.Ratio in our dataset.*There are NaN(Non null values) in Eroll column.

2- Display the last 5 rows

In [10]:

Out[10]:

	Names	Apps	Accept	Enroll	Top10perc	Top25perc	F.Undergrad	P.Undergrad	Outs
772	Worcester State College	2197.0	1515	543.0	4.0	26	3089	2029	
773	Xavier University	1959.0	1805	695.0	24.0	47	2849	1107	1
774	Xavier University of Louisiana	2097.0	1915	695.0	34.0	61	2793	166	
775	Yale University	10705.0	2453	1317.0	95.0	99	5217	83	1
776	York College of Pennsylvania	2989.0	1855	691.0	28.0	63	2988	1726	

Observations:

- There are no NaN value in our dataset at the bottom in both S.F.Ratio and Enroll.

3- Check the shape of dataset.

In [11]:

Out[11]:
```
(777, 18)
```

Observations:

- The number of columns in our dataset is 777.
- The number of rows in our dataset is 18.

4- Check the datatypes of each feature.

In [12]:

Out[12]:
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 777 entries, 0 to 776
Data columns (total 18 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   Names        777 non-null    object 
 1   Apps         775 non-null    float64
 2   Accept       777 non-null    int64  
 3   Enroll       775 non-null    float64
 4   Top10perc    773 non-null    float64
 5   Top25perc    777 non-null    int64  
 6   F.Undergrad  777 non-null    int64  
 7   P.Undergrad  777 non-null    int64  
 8   Outstate     777 non-null    int64  
 9   Room.Board   777 non-null    int64  
 10  Books        777 non-null    int64  
 11  Personal     774 non-null    float64
 12  PhD          777 non-null    int64  
 13  Terminal     777 non-null    int64  
 14  S.F.Ratio    777 non-null    object 
 15  perc.alumni  777 non-null    int64  
 16  Expend       777 non-null    int64  
 17  Grad.Rate    777 non-null    int64  
dtypes: float64(4), int64(12), object(2)
memory usage: 109.4+ KB
Names          object
Apps          float64
Accept          int64
Enroll        float64
Top10perc     float64
Top25perc       int64
F.Undergrad     int64
P.Undergrad     int64
Outstate        int64
Room.Board      int64
Books           int64
Personal      float64
PhD             int64
Terminal        int64
S.F.Ratio      object
perc.alumni     int64
Expend          int64
Grad.Rate       int64
dtype: object
```

Observations:

- There are Integer , float and object data types in the given dataset.

5- Check the Statistical summary

In [13]:

Out[13]:

	Apps	Accept	Enroll	Top10perc	Top25perc	F.Undergrad	P.Undergra
count	775.000000	777.000000	775.000000	773.000000	777.000000	777.000000	777.00000
mean	3007.592258	2018.804376	780.961290	27.620957	55.796654	3699.907336	855.29858
std	3873.414660	2451.113971	930.077779	17.645470	19.804778	4850.420531	1522.43188
min	81.000000	72.000000	35.000000	1.000000	9.000000	139.000000	1.00000
25%	778.000000	604.000000	242.500000	15.000000	41.000000	992.000000	95.00000
50%	1561.000000	1110.000000	434.000000	23.000000	54.000000	1707.000000	353.00000
75%	3635.000000	2424.000000	902.500000	35.000000	69.000000	4005.000000	967.00000
max	48094.000000	26330.000000	6392.000000	96.000000	100.000000	31643.000000	21836.00000

Obeservations:

- Describing the data with the given attributes like count, mean, std,min,25%,50%,75%,max.

6- Check the null values

In [14]:

Out[14]:

	Names	Apps	Accept	Enroll	Top10perc	Top25perc	F.Undergrad	P.Undergrad	Outsta
3	Agnes Scott College	417.0	349	NaN	60.0	89	510	63	129
9	Alderson-Broaddus College	NaN	498	172.0	21.0	44	799	78	104
41	Bellarmine College	NaN	707	308.0	39.0	63	1198	605	88
81	Campbell University	2087.0	1339	NaN	20.0	54	3191	1204	75
102	Central Connecticut State University	4158.0	2532	902.0	NaN	24	6394	3881	59
103	Central Missouri State University	4681.0	4101	1436.0	NaN	35	8094	1596	46
128	College of Notre Dame	344.0	264	97.0	NaN	42	500	331	126
129	College of Notre Dame of Maryland	457.0	356	177.0	NaN	61	667	1983	111
166	Dillard University	1998.0	1376	651.0	41.0	88	1539	45	67
175	Earlham College	1358.0	1006	274.0	35.0	63	1028	13	150
177	East Tennessee State University	3330.0	2730	1303.0	15.0	36	6706	2640	58

Observation

*There is NaN values in Apps and Enroll.

7- Check the duplicate values

In [15]:

```
```

Out[15]:

```
0
```

Obersevation

- No duplicated values in our dataset.

8- Check the anomalies or wrong entries

In [17]:

Out[17]:

	Names	Apps	Accept	Enroll	Top10perc	Top25perc	F.Undergrad	P.Undergrad	Outstate
1	Adelphi University	2186.0	1924	512.0	16.0	29	2683	1227	12280
81	Campbell University	2087.0	1339	NaN	20.0	54	3191	1204	7550
241	Gwynedd Mercy College	380.0	237	104.0	30.0	56	716	1108	11000

Observation

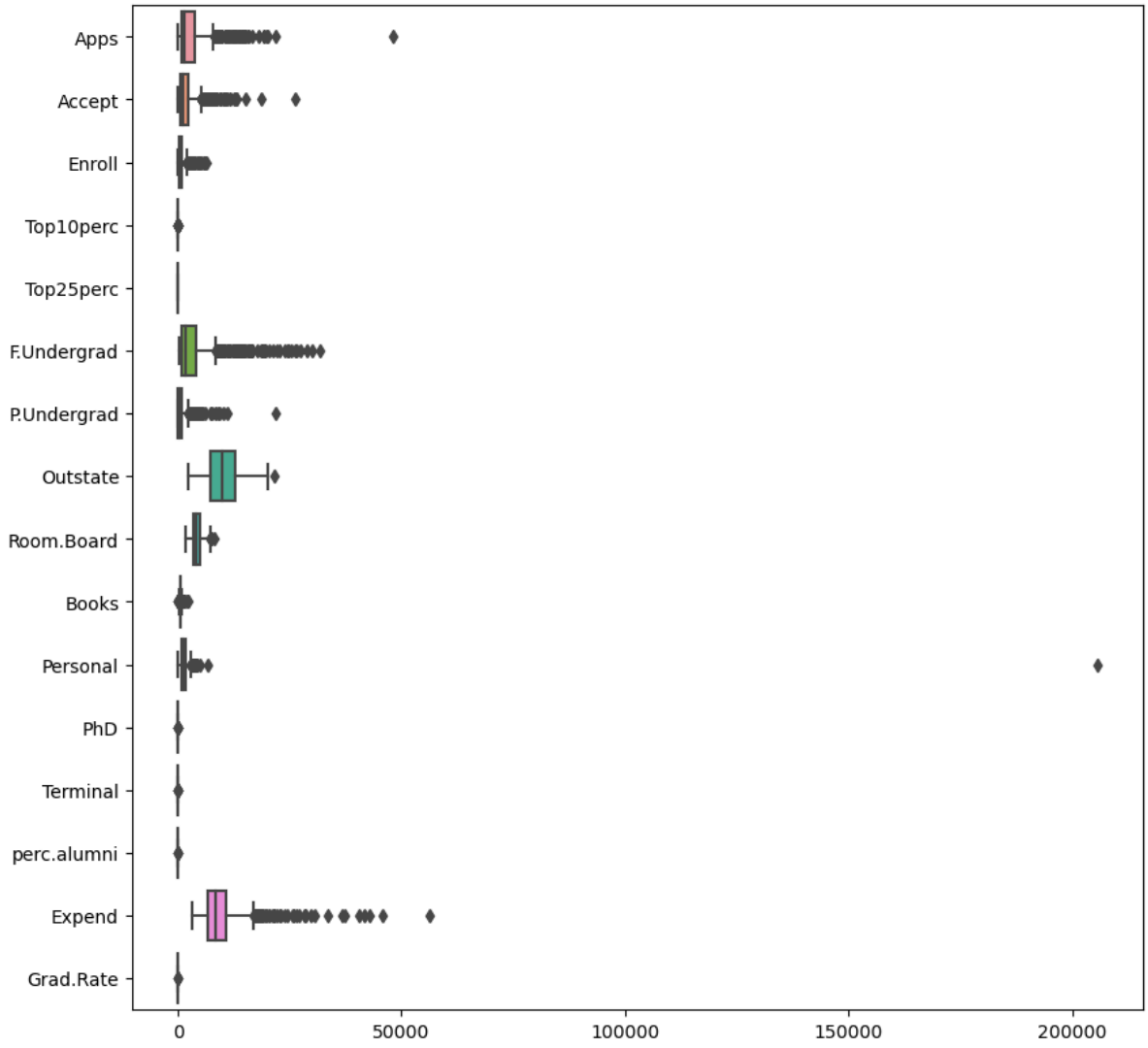
- Shows that in S.F.Ratio row no:1,81 and 241 contains '?'(Non integer value).

9- Check the outliers and their authenticity.

In [18]:

Out[18]:

```
<Axes: >
```



In [19]:

Observation

*Shows the outliners in Expend and F.Undergrad.

10- Do the necessary data cleaning steps like dropping duplicates, unnecessary columns, null value imputation, outliers treatment etc.

In [20]:

In [21]:

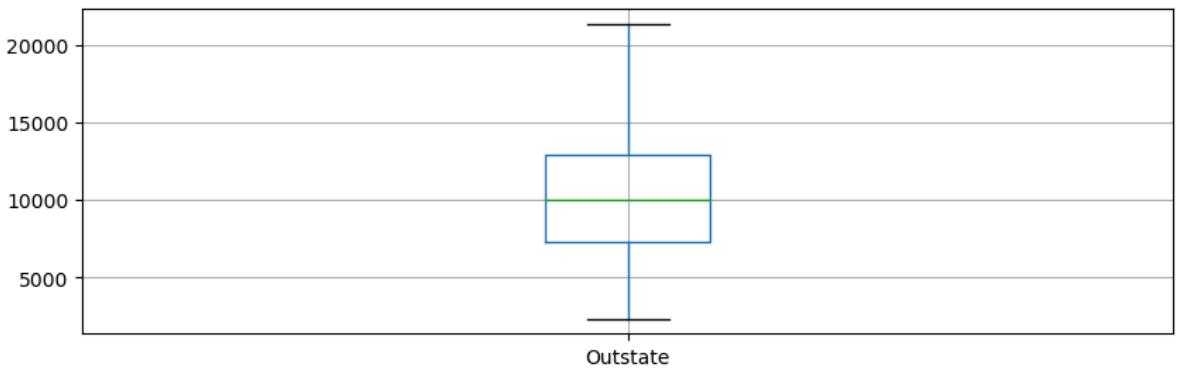
Out[21]:

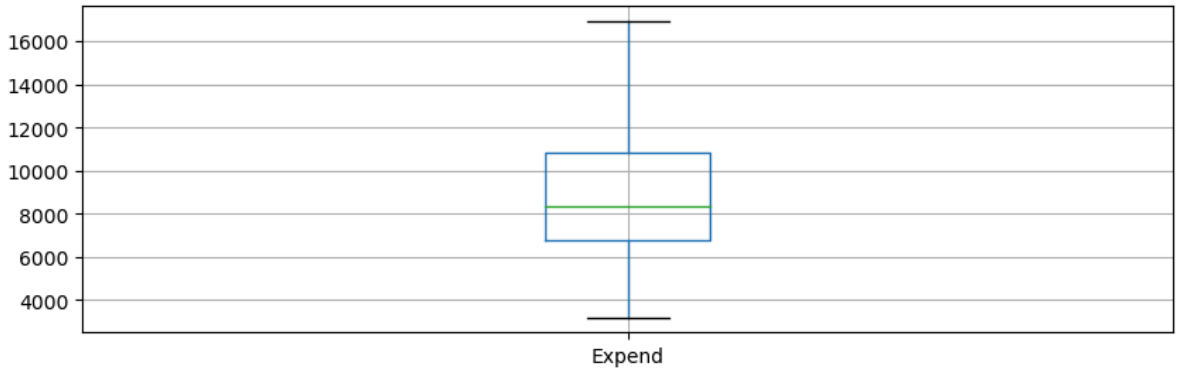
	Names	Apps	Accept	Enroll	Top10perc	Top25perc	F.Undergrad	P.Undergrad	Outstate
0	Abilene Christian University	1660.0	1232.0	721.0	23.0	52.0	2885.0	537.0	7440.0
1	Adelphi University	2186.0	1924.0	512.0	16.0	29.0	2683.0	1227.0	12280.0
2	Adrian College	1428.0	1097.0	336.0	22.0	50.0	1036.0	99.0	11250.0
3	Agnes Scott College	417.0	349.0	NaN	60.0	89.0	510.0	63.0	12960.0
4	Alaska Pacific University	193.0	146.0	55.0	16.0	44.0	249.0	869.0	7560.0

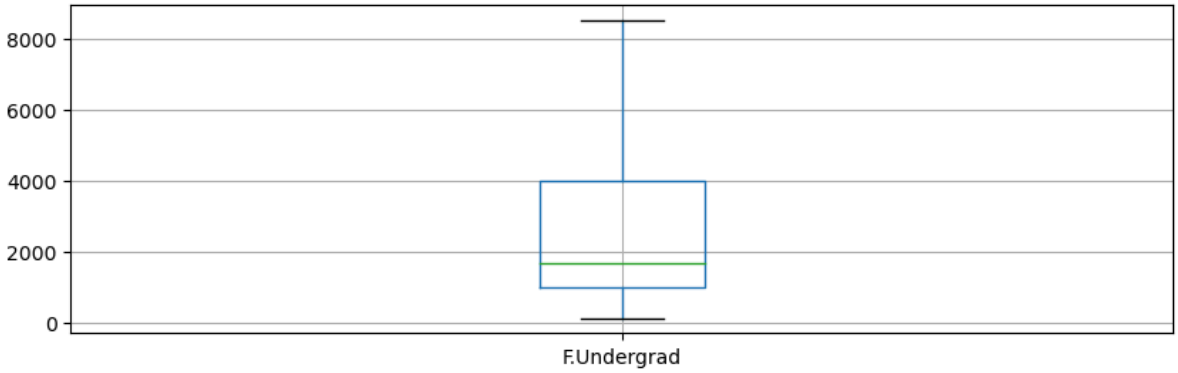
In [22]:

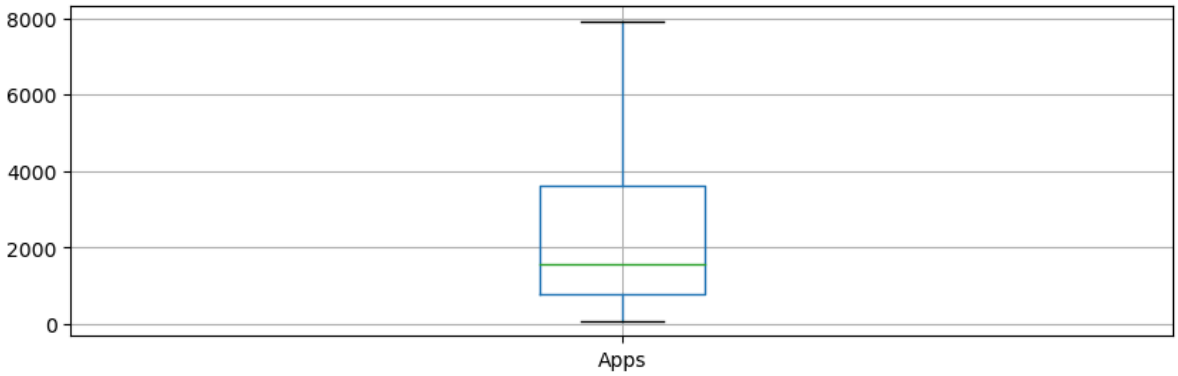
Out[22]:
```
Names          0
Apps           2
Accept         0
Enroll         2
Top10perc      4
Top25perc      0
F.Undergrad    0
P.Undergrad    0
Outstate       0
Room.Board     0
Books          0
Personal       3
PhD            0
Terminal       0
S.F.Ratio      3
perc.alumni    0
Expend         0
Grad.Rate      0
dtype: int64
```

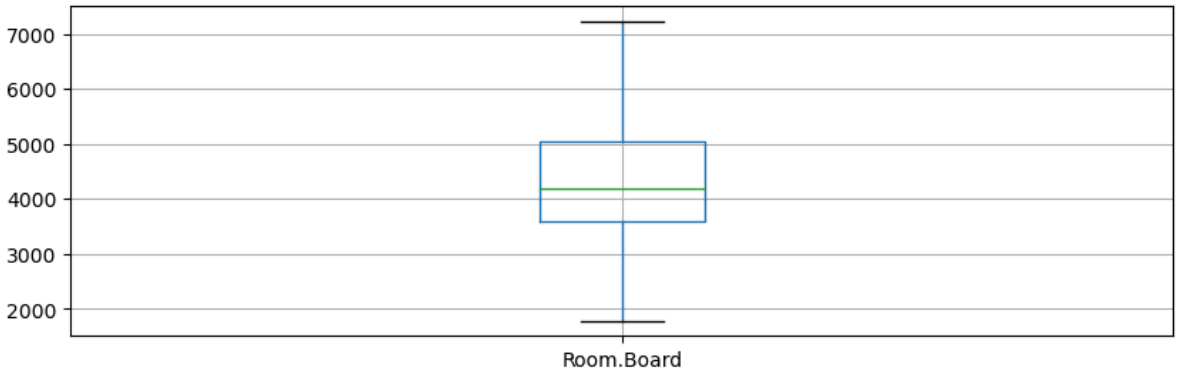
In [23]:











Observation

- Exchanged values with NaN.

1. Application and Enrollment Analysis

- What is the average number of applications received by colleges?
- What percentage of applications are accepted on average across all colleges?
- What is the average enrollment rate (number of students enrolled divided by number of applications accepted)?

- Which college has the highest number of applications received?

In [7]:

```
Average number of applications received by colleges: 3007.592258064516
```

In [8]:

```
Average percentage of applications accepted across all colleges: 74.6620740888845
8%
```

In [9]:

```
Average enrollment rate: 1165.6270717382552%
```

In [11]:

```
College with the highest number of applications received: Rutgers at New Brunswick
```

2. Academic Excellence

- What is the average percentage of new students from the top 10% of their higher secondary class across all colleges?
- What is the average percentage of new students from the top 25% of their higher secondary class?
- Is there a correlation between the percentage of students from the top 10% and the top 25% of their higher secondary class?

In [12]:

```
Average percentage of new students from the top 10% of their higher secondary clas
s: 27.620957309184995%
```

In [13]:

```
Average percentage of new students from the top 25% of their higher secondary clas
s: 55.7966537966538%
```

In [14]:

```
Correlation between the percentage of students from the top 10% and the top 25% of
their higher secondary class: 0.8916010734346559
```

3. Student Demographics

- What is the average number of full-time undergraduate students per college?
- What is the average number of part-time undergraduate students per college?
- Which college has the highest number of out-of-state students?

In [15]:

```
Average number of full-time undergraduate students per college: 3699.907335907336
```

In [16]:

```
Average number of part-time undergraduate students per college: 855.2985842985843
```

In [18]:

```
College with the highest number of out-of-state students: Bennington College
```

4. Cost and Spending

- What is the average cost of room and board across all colleges?
- What is the average estimated book cost for a student?
- What is the average estimated personal spending for a student?
- How does the instructional expenditure per student vary across colleges?

In [19]:

```
Average cost of room and board across all colleges: $4357.526383526383
```

In [20]:

```
Average estimated book cost for a student: $547.8751608751609
```

In [21]:

```
Average estimated personal spending for a student: $1601.5077519379845
```

In [22]:

```
Instructional expenditure per student across colleges:
                             Names  Personal
0      Abilene Christian University    2200.0
1                 Adelphi University    1500.0
2                     Adrian College    1165.0
3                Agnes Scott College     875.0
4          Alaska Pacific University    1500.0
..                               ...       ...
772          Worcester State College    1200.0
773                Xavier University    1250.0
774  Xavier University of Louisiana     781.0
775                  Yale University    2115.0
776    York College of Pennsylvania    1250.0

[777 rows x 2 columns]
```

5. Faculty Qualifications

- What is the average percentage of faculties with Ph.D.s across all colleges?
- What is the average percentage of faculties with terminal degrees?
- Is there a correlation between the percentage of faculties with Ph.D.s and the graduation rate?

In [3]:

```
The average percentage of faculties with Ph.D.s across all colleges is: 72.66%
```

In [4]:

```
The average percentage of faculties with terminal degrees across all colleges is:
79.70%
```

In [5]:

```
The correlation between the percentage of faculties with Ph.D.s and graduation rat
e is: 0.31
```

6. Student-Faculty Interaction

- What is the average student/faculty ratio across all colleges?
- Which college has the lowest student/faculty ratio?
- Is there a correlation between the student/faculty ratio and the graduation rate?

In [6]:

```
The average student/faculty ratio across all colleges is: 10.34
```

In [43]:

```
The college with the lowest student/faculty ratio is: Christendom College
```

In [14]:

```
The correlation between student/faculty ratio and graduation rate is: -0.07
```

7. Alumni Engagement

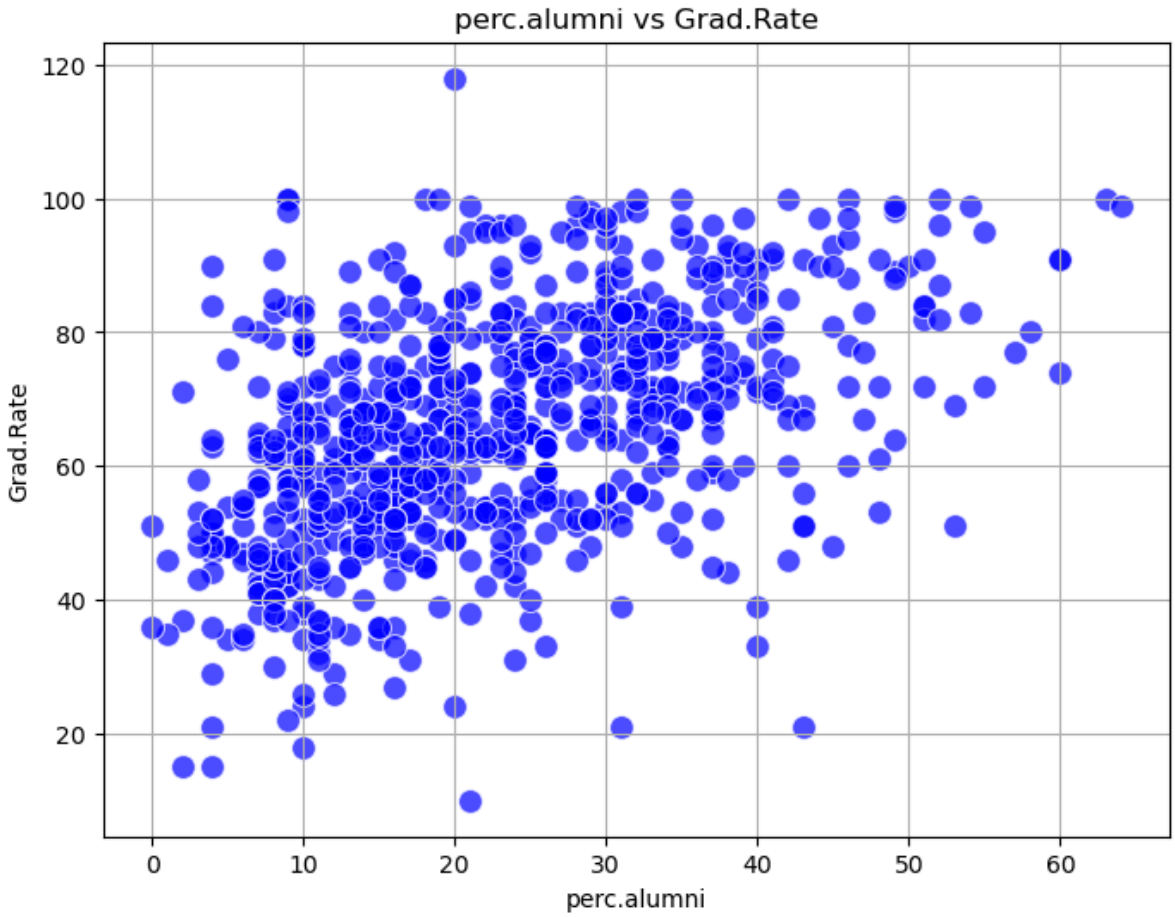
- What is the average percentage of alumni who donate across all colleges?
- Is there a correlation between the percentage of alumni who donate and the graduation rate?

In [16]:

```
Average donation percentage across all colleges: 22.74%
```

In [18]:

```
Correlation between donation percentage and graduation rate: 0.49
```



8. Graduation Rates

- What is the average graduation rate across all colleges?
- Which college has the highest graduation rate?
- Is there a correlation between the instructional expenditure per student and the graduation rate?

In [22]:

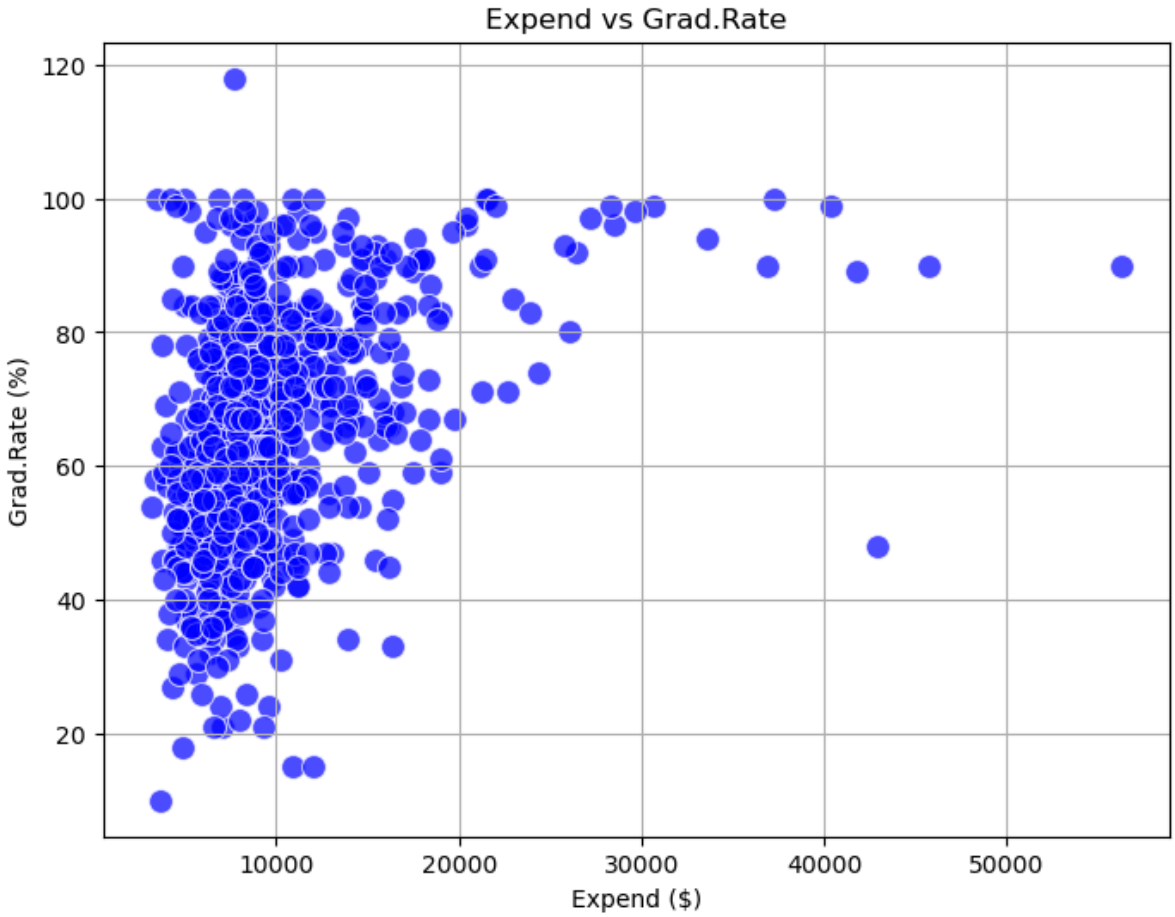
```
Average graduation rate across all colleges: 65.46%
```

In [23]:

```
College with the highest graduation rate: Cazenovia College (118%)
```

In [24]:

```
Correlation between instructional expenditure per student and graduation rate: 0.3
9
```



9. Overall Insights

- Which factors (applications, acceptance rate, enrollment, academic excellence, costs, faculty qualifications, student/faculty ratio, alumni donations, expenditures) are most strongly associated with higher graduation rates?
- What recommendations can be made to colleges to improve their graduation rates based on the data analysis?

In [39]:

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 764 entries, 0 to 776
Data columns (total 18 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   Names        764 non-null    object 
 1   Apps         764 non-null    float64
 2   Accept       764 non-null    int64  
 3   Enroll       764 non-null    float64
 4   Top10perc    764 non-null    float64
 5   Top25perc    764 non-null    int64  
 6   F.Undergrad  764 non-null    int64  
 7   P.Undergrad  764 non-null    int64  
 8   Outstate     764 non-null    int64  
 9   Room.Board   764 non-null    int64  
 10  Books        764 non-null    int64  
 11  Personal     764 non-null    float64
 12  PhD          764 non-null    int64  
 13  Terminal     764 non-null    int64  
 14  S.F.Ratio    764 non-null    float64
 15  perc.alumni  764 non-null    int64  
 16  Expend       764 non-null    int64  
 17  Grad.Rate    764 non-null    int64  
dtypes: float64(5), int64(12), object(1)
memory usage: 113.4+ KB
None
                Apps    Accept    Enroll  S.F.Ratio  Top10perc  Room.Board  \
Apps         1.000000  0.943363  0.846368   0.090700   0.344302    0.169436   
Accept       0.943363  1.000000  0.911336   0.171654   0.197501    0.095105   
Enroll       0.846368  0.911336  1.000000   0.233052   0.185971   -0.036486   
S.F.Ratio    0.090700  0.171654  0.233052   1.000000  -0.383485   -0.358022   
Top10perc    0.344302  0.197501  0.185971  -0.383485   1.000000    0.375651   
Room.Board   0.169436  0.095105 -0.036486  -0.358022   0.375651    1.000000   
PhD          0.393806  0.359035  0.333713  -0.138108   0.537429    0.343354   
perc.alumni -0.088466 -0.158103 -0.179209  -0.406118   0.451758    0.277001   
Expend       0.262631  0.127074  0.066373  -0.583576   0.661679    0.503798   
Grad.Rate    0.150009  0.070710 -0.019744  -0.307118   0.494489    0.430285   

                 PhD  perc.alumni    Expend  Grad.Rate  
Apps         0.393806    -0.088466  0.262631   0.150009  
Accept       0.359035    -0.158103  0.127074   0.070710  
Enroll       0.333713    -0.179209  0.066373  -0.019744  
S.F.Ratio   -0.138108    -0.406118 -0.583576  -0.307118  
Top10perc    0.537429     0.451758  0.661679   0.494489  
Room.Board   0.343354     0.277001  0.503798   0.430285  
PhD          1.000000     0.245065  0.434793   0.312200  
perc.alumni  0.245065     1.000000  0.417524   0.488046  
Expend       0.434793     0.417524  1.000000   0.394980  
Grad.Rate    0.312200     0.488046  0.394980   1.000000  
Grad.Rate      1.000000
Top10perc      0.494489
perc.alumni    0.488046
Room.Board     0.430285
Expend         0.394980
PhD            0.312200
Apps           0.150009
Accept         0.070710
Enroll        -0.019744
S.F.Ratio     -0.307118
Name: Grad.Rate, dtype: float64
```

What recommendations can be made to colleges to improve their graduation rates based on the data

analysis?

*Recommendations for Colleges*Based on the correlation analysis and domain knowledge, make recommendations to colleges to potentially improve their graduation rates:

*Focus on Factors with Positive Correlation: Prioritize factors that have a strong positive correlation with graduation rates, such as improving academic excellence metrics, enhancing faculty qualifications, optimizing student/faculty ratios, and increasing alumni donations.

*Address Factors with Negative Correlation: Identify factors negatively correlated with graduation rates (if any) and develop strategies to mitigate their impact, such as reducing costs without compromising educational quality or improving acceptance rates while maintaining academic standards.

*Data-Driven Decision Making: Encourage colleges to regularly analyze their data and adapt strategies based on ongoing trends and correlations observed in their specific context.

In [ ]: