

# SUMMER BOOTCAMP 2024

## Data Science Project KESHAV BISHT Education Post 12th Data Analysis

### PROBLEM STATEMENT / OBJECTIVE

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.

### BASIC EXPLORATION

#### 1. FIRST FIVE ROWS

	Names	Apps	Accept	Enroll	Top10perc	Top25perc	F.Undergrad	P.Undergrad	Outstate	Room.Board	Books	Personal	PhD	Terminal	S.F.Ratio	perc
0	Abilene Christian University	1660.0	1232	721.0	23.0	52	2885	537	7440	3300	450	2200.0	70	78	18.1	2
1	Adelphi University	2186.0	1924	512.0	16.0	29	2683	1227	12280	6450	750	1500.0	29	30	?	1
2	Adrian College	1428.0	1097	336.0	22.0	50	1036	99	11250	3750	400	1165.0	53	66	12.9	1
3	Agnes Scott College	417.0	349	NaN	60.0	89	510	63	12960	5450	450	875.0	92	97	7.7	1
4	Alaska Pacific University	193.0	146	55.0	16.0	44	249	869	7560	4120	800	1500.0	76	72	11.9	1

#### 2. INFO

```
<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
```

#### 3. LAST 5 ROWS

	Names	Apps	Accept	Enroll	Top10perc	Top25perc	F.Undergrad	P.Undergrad	Outstate	Room.Board	Books	Personal	PhD	Terminal	S.F.Ratio	perc
772	Worcester State College	2197.0	1515	543.0	4.0	26	3089	2029	6797	3900	500	1200.0	60	60	2	2
773	Xavier University	1959.0	1805	695.0	24.0	47	2849	1107	11520	4960	600	1250.0	73	75	13	1
774	Xavier University of Louisiana	2097.0	1915	695.0	34.0	61	2793	166	6900	4200	617	781.0	67	75	14	1
775	Yale University	10705.0	2453	1317.0	95.0	99	5217	83	19840	6510	630	2115.0	96	96	5	1
776	York College of Pennsylvania	2989.0	1855	691.0	28.0	63	2988	1726	4990	3560	500	1250.0	75	75	18	1

#### 4. SHAPE

(777, 18)

#### 5. STATISTICAL SUMMARY

	Apps	Accept	Enroll	Top10perc	Top25perc	F.Undergrad	P.Undergrad	Outstate	Room.Board	Books	Personal	perc
count	775.000000	777.000000	775.000000	773.000000	777.000000	777.000000	777.000000	777.000000	777.000000	777.000000	777.000000	774.000000
mean	3007.592258	2018.804376	780.961290	27.620957	55.796654	3699.907336	855.298584	10440.669241	4357.526384	547.875161	1601.507752	1601.507752
std	3873.414660	2451.113971	930.077779	17.645470	19.804778	4850.420531	1522.431887	4023.016484	1096.696416	167.426237	7369.594038	7369.594038
min	81.000000	72.000000	35.000000	1.000000	9.000000	139.000000	1.000000	2340.000000	1780.000000	0.000000	50.000000	50.000000
25%	778.000000	604.000000	242.500000	15.000000	41.000000	992.000000	95.000000	7320.000000	3597.000000	465.000000	855.000000	855.000000
50%	1561.000000	1110.000000	434.000000	23.000000	54.000000	1707.000000	353.000000	9990.000000	4200.000000	500.000000	1200.000000	1200.000000
75%	3635.000000	2424.000000	902.500000	35.000000	69.000000	4005.000000	967.000000	12925.000000	5050.000000	600.000000	1687.500000	1687.500000
max	48094.000000	26330.000000	6392.000000	96.000000	100.000000	31643.000000	21836.000000	21700.000000	8124.000000	2340.000000	205500.000000	205500.000000

#### 6. NULL VALUES

	Names	Apps	Accept	Enroll	Top10perc	Top25perc	F.Undergrad	P.Undergrad	Outstate	Room.Board	Books	Personal	PhD	Terminal	S.F.Ratio	perc
0	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
3	False	False	False	True	False	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
772	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
773	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
774	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
775	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
776	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False

777 rows x 18 columns

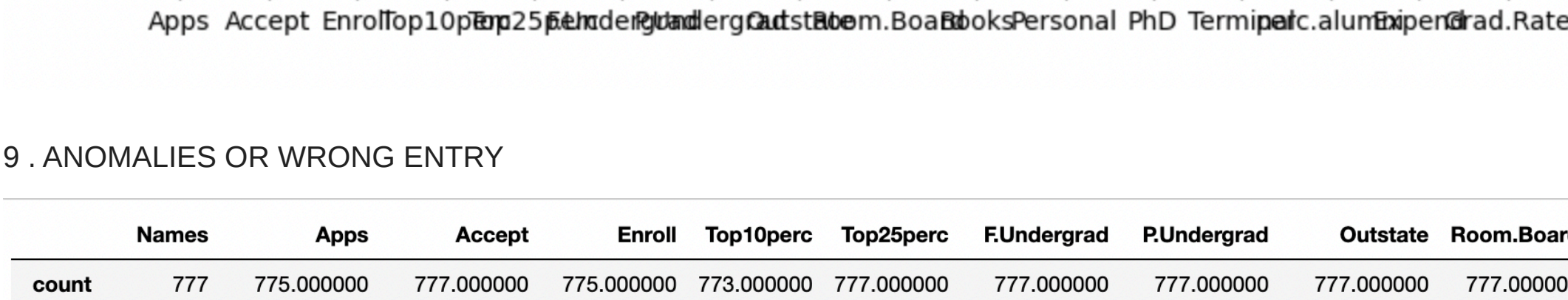
#### 7. DUPLICATE VALUES

```
0      False
1      False
2      False
3      False
4      False

...
772     False
773     False
774     False
775     False
776     False
Length: 777, dtype: bool
```

#### 8. OUTLIERS AND THEIR AUTHITICITY

<Axes: >



#### 9. ANOMALIES OR WRONG ENTRY

	Names	Apps	Accept	Enroll	Top10perc	Top25perc	F.Undergrad	P.Undergrad	Outstate	Room.Board	Books	perc
count	777	775.000000	777.000000	775.000000	773.000000	777.000000	777.000000	777.000000	777.000000	777.000000	777.000000	777.000000
unique	777	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
top	Abilene Christian University	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
freq	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
mean	NaN	3007.592258	2018.804376	780.961290	27.620957	55.796654	3699.907336	855.298584	10440.669241	4357.526384	547.875161	1601.507752
std	NaN	3873.414660	2451.113971	930.077779	17.645470	19.804778	4850.420531	1522.431887	4023.016484	1096.696416	167.426237	7369.594038
min	NaN	81.000000	72.000000	35.000000	1.000000	9.000000	139.000000	1.000000	2340.000000	1780.000000	0.000000	50.000000
25%	NaN	778.000000	604.000000	242.500000	15.000000	41.000000	992.000000	95.000000	7320.000000	3597.000000	465.000000	855.000000
50%	NaN	1561.000000	1110.000000	434.000000	23.000000	54.000000	1707.000000	353.000000	9990.000000	4200.000000	500.000000	1200.000000
75%	NaN	3635.000000	2424.000000	902.500000	35.000000	69.000000	4005.000000	967.000000	12925.000000	5050.000000	600.000000	1687.500000
max	NaN	48094.000000	26330.000000	6392.000000	96.000000	100.000000	31643.000000	21836.000000	21700.000000	8124.000000	2340.000000	205500.000000

### APPLICATION AND ENROLLMENT ANALYSIS

#### 1. WHAT IS THE AVERAGE NUMBER OF APPLICATIONS RECEIVED BY COLLEGES?

3007.592258064516

#### 2. WHAT PERCENTAGE OF APPLICATIONS ARE ACCEPTED ON AVERAGE ACROSS ALL COLLEGES?

74.6620740888846

#### 3. What is the average enrollment rate (number of students enrolled divided by number of applications accepted)?

41.19383221625094

#### 4. Which college has the highest number of applications received?

'Rutgers at New Brunswick'

#### 5. Average percentage of new students from the top 10% of their high school class

27.620957309184995

### ACADEMIC EXCELLENCE

#### 1. Average percentage of new students from the top 10% of their high school class

27.620957309184995

#### 2. Average percentage of new students from the top 25% of their high school class

55.7966537966538

#### 3. Correlation between the percentage of students from the top 10% and the top 25% of their high school class

0.8916010734346558

### STUDENT DEMOGRAPHICS

#### 1. What is the average number of full-time undergraduate students per college?

3699.907335907336

#### 2. Average number of part-time undergraduate students per college

855.2985842985843

#### 3. College with the highest number of out-of-state students

'Bennington College'

### COST OF SPENDING

#### 1. What is the average cost of room and board across all colleges?

4357.526383526383

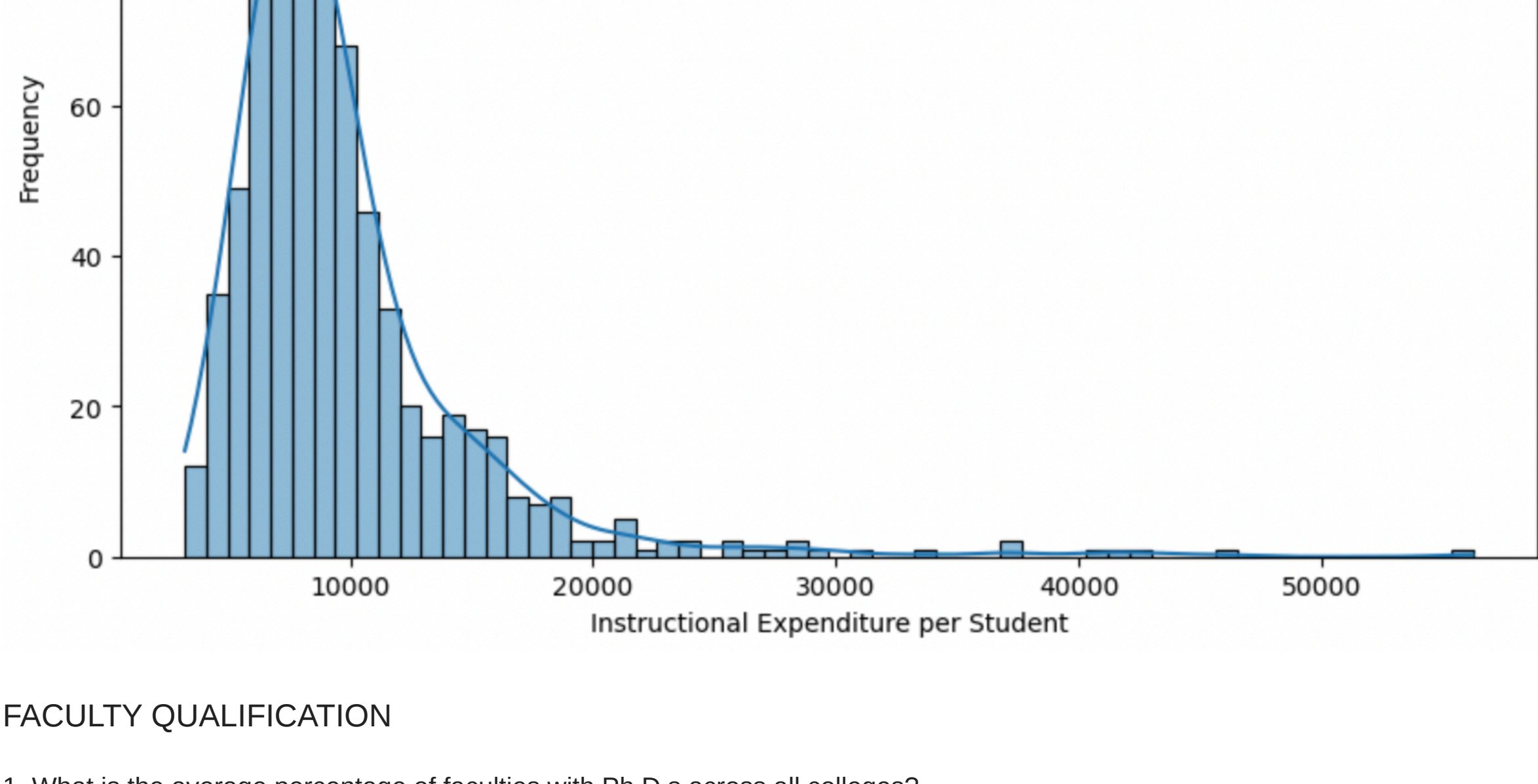
#### 2. What is the average estimated book cost for a student?

547.8751608751609

#### 3. What is the average estimated personal spending for a student?

1601.507751979845

#### 4. How does the instructional expenditure per student vary across colleges?



### FACULTY QUALIFICATION

#### 1. What is the average percentage of faculties with Ph.D.s across all colleges?

72.66023166023166

#### 2. What is the average percentage of faculties with terminal degrees?

79.70270270270271

#### 3. Is there a correlation between the percentage of faculties with Ph.D.s and the graduation rate?

0.30503785002341166

### STUDENT - FACULTY INTERACTION

#### 1. Which college has the lowest student/faculty ratio?

'Saint Olaf College'

### ALUMNI ENGAGEMENT

#### 1. What is the average percentage of alumni who donate across all colleges?

22.743886743886744

#### 2. Is there a correlation between the percentage of alumni who donate and the graduation rate?

0.4908975622847202

### GRADUATION RATES

#### 1. What is the average graduation rate across all colleges?

65.46332046332046

#### 2. Which college has the highest graduation rate?

'Cazenovia College'

#### 3. Is there a correlation between the instructional expenditure per student and the graduation rate?

0.3903426958593891

### OVERALL INSIGHTS

1. 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?

- Out-of-state Tuition (Outstate):** Strong positive correlation (0.571). Higher tuition often reflects better resources.
- Top 10% of High School Class (Top10perc):** Strong positive correlation (0.493). Attracting high-achieving students leads to higher graduation rates.
- Alumni Donations (perc.alumni):** Strong positive correlation (0.491). Alumni support often enhances resources and institutional quality.
- Top 25% of High School Class (Top25perc):** Strong positive correlation (0.477). High-achieving students contribute to higher graduation rates.
- Room and Board Costs (Room.Board):** Moderate positive correlation (0.425). Better living conditions support student success.

#### 2. What recommendations can be made to colleges to improve their graduation rates based on the data analysis?

- Recruit High-Achieving Students:** Offer scholarships and honors programs.
- Enhance Alumni Engagement:** Strengthen alumni relations and encourage donations.
- Improve Campus Living Conditions:** Invest in quality housing and facilities.
- Increase Instructional Spending:** Allocate more resources to teaching and support services.
- Improve Faculty Qualifications:** Support faculty in obtaining higher degrees and professional development.