Data Narratives-2 (March 2023)

Birudugadda Srivibhav -22110050, Computer Science and Technology Department, Prof. Shanmuga R, IIT Gandhinagar

***Abstract*-Exploring and analysing trends in higher education are the goals of this data story project, which uses the AAUP (American Association of University Professors) and US News datasets. The study uses statistical analysis to find trends and insights in the data and focuses on issues like faculty wages, diversity, and rankings. These datasets will be examined to understand the situation of higher education better today and provide information for policy and decision-making.**

I. Aim

To study and write a data narrative on AAUP (American Association of University Professors) and US News datasets.

II. Overview of the dataset

The aaup dataset:

The American Association of University Professors (AAUP) dataset compiles information about professors employed by American higher education institutions. The dataset contains the following columns: FICE, College Name, State(Postal code), Type 'Average salary-full professors, Average salary-associate professors, Average salary-assistant professors, Average salary-all ranks, Average compensation-full professors, Average compensation-associate professors, Average compensation-assistant professors, Average compensation-all ranks, Number of full professors, Number of associate professors, Number of assistant professors, Number of instructors, and Number of faculty-all ranks.

Annual surveys of faculty members and higher education institutions are used to compile the AAUP dataset. The faculty wages are split down by institution type, rank, and geographic region in the AAUP dataset. This enables salary comparisons across various institution types and academic specialities and can be used to pinpoint instances when academics may be paid inadequately. Overall, the AAUP dataset is a valuable tool for learning about faculty members' working conditions in higher education and for identifying problem areas that need to be addressed.

The US News dataset:

The US News dataset is a thorough compilation of information on American schools and universities that the US News & World Report compiles and publishes annually. Students and college administrators frequently utilise the dataset, which contains data on over 1,300 institutions, to assess and contrast various colleges and universities. The dataset contains the following columns: FICE (Federal ID number), College Name, State(Postal code), Public/private indicator (public=1, private=2), Average Math SAT score, Average Verbal SAT score, Average Combined SAT score, Average ACT score, First quartile - Math SAT, Third quartile - Math SAT, First quartile - Verbal SAT, Third quartile - Verbal SAT 'First quartile - ACT, Third quartile - ACT, Number of applications received, Number of applicants accepted, Number of new students enrolled, Pct. new students from the top 10% of H.S. class, Pct. new students from the top 25% of H.S. class, Number of full-time undergraduates, Number of part-time undergraduates, In-state tuition, Out-of-state tuition, Room and board costs, Room costs, Board costs, Additional fees, Estimated book costs, Estimated personal spending, Pct. of faculty with PhD.s, Pct. of faculty with terminal degree, Student/faculty ratio, Pct. alumni who donate, Instructional expenditure per student, and Graduation rate.

Overall, the US News dataset is a valuable resource for anyone involved in higher education, providing comprehensive and detailed information on colleges and universities in the United States.

III. scientific questions/hypothesis

The aaup dataset:

1. **Question:** Find the plot and correlation between the number of faculty members and their average salary distribution. Also, find which university has the maximum average salary and which university have the maximum number of faculty members.
   Hypothesis: As the number of faculty members increases, the average salary of faculty members decreases.
2. **Question:** Calculate the mean of the average salaries of all ranks for each state. Also, find out which state pays less salary for its faculty members.
3. **Question:** Find out which type of university offers more salaries and benefits to its faculty members. Also, find out the number of universities of each type by state.
4. **Question:** Is there a correlation between the average salary and compensation of faculty members across all ranks? Do different types of institutions (i.e., Type I, II, or VII) show different patterns in this relationship?

5. **Question:** Are there any notable differences in the average salaries and benefits of full, associate, and assistant professors in institutes? Plot the way they are distributed.

The US News dataset:

6. **Question:** Does a higher instructional expenditure per student lead to higher graduation rates?
   Hypothesis: Higher instructional expenditure per student leads to higher graduation rates.
7. **Question:** Find the bar graph for the state-wise mean average graduation rate.
8. **Question:** How does the acceptance rate relate to the average sat score?
   Hypothesis: the institutes with low acceptance rates require higher average sat scores
9. **Question:** which type of institutes(public or private) have higher acceptance rates?
   Hypothesis: Public institutes have higher acceptance rates.
10. **Question:** Do colleges with higher percentages of faculty with Ph. D.s have higher graduation rates?
    Hypothesis: Colleges with a higher percentage of faculty with Ph. D.s will have higher graduation rates.

IV. Details of libraries and functions

The libraries and modules used are

[1] Numpy library - A Python library for scientific computing that provides tools for working with arrays and matrices, as well as mathematical functions and linear algebra operations.

[2] Pandas library- A Python library used for data manipulation and analysis. It provides data structures for efficiently storing and accessing large datasets and tools for cleaning, transforming, and analysing data.

[3] Matplotlib library- A Python library for creating static, animated, and interactive visualisations in Python. It provides various plotting options and customisation tools to create publication-quality figures.

[4] Seaborn library- A Python library based on Matplotlib that provides additional visualisation tools, particularly for statistical data analysis. It provides a high-level interface for creating complex plots with minimal code.

The functions used are

pd.merge - A function in Pandas that combines two or more data frames based on a common column.

sort_values - A function in Pandas that sorts a data frame by one or more columns.

set_index - A function in Pandas used to set one or more columns as the index of a data frame.

value_counts - A function in Pandas used to count the occurrence of unique values in a column.

drop_duplicates - A function in Pandas removes duplicate rows from a data frame.

isin - A function in Pandas that checks if a value is in a list or array.

plt.bar - A function in Matplotlib used to create a bar chart.

plt.show - A function in Matplotlib used to display a plot.

sns.barplot - A function in Seaborn used to create a bar chart with additional styling options.

np.arange - A function in Numpy creates a range of numbers with a specified start, stop, and step.

np.corrcoef - This function calculates the correlation coefficient between two variables.

pd.DataFrame - A class in Pandas used to create a data frame from a dictionary or array of data.

pd.DataFrame.describe - This function displays the statistical summary of a data frame, including count, mean, standard deviation, minimum, maximum, and quartiles.

pd.read_csv - A function in Pandas used to read a CSV file into a data frame.

pd.groupby - Pandas groupby is used for grouping the data according to the categories and applying a function to the categories. It also helps to aggregate data efficiently. Pandas data frame. group by () function splits the data into groups based on some criteria.

sns.scatterplot - The scatterplot function in seaborn is used to create a scatter plot that displays the relationship between two variables. It is useful for identifying data patterns and trends and can help reveal any correlations between the two variables.

V. Answers to the questions

The aaup dataset:

Question 1: Find the plot and correlation between the number of faculty members and their average salary distribution. Also, find which university has the maximum average salary and which university have the maximum number of faculty members.
Hypothesis: As the number of faculty members increases, the average salary of faculty members decreases.
Answer:
I have calculated the correlation between the "Average salary-all ranks" and "Number of faculty-all ranks" columns to answer this question. This will give us an indication of whether there is a relationship between the average salary and the number of faculty members at an institution. the correlation coefficient is a statistical measure that indicates the strength and direction of the linear relationship between two variables.

For calculating the correlation factor, I have loaded the books dataset into a Pandas data frame and then used the Pandas corr() function to calculate the correlation coefficient between the number of faculty members and their average salaries institute-wise. The resulting correlation coefficient was 0.570, which indicates a moderate positive correlation between the two variables.

```
# Filter the dataframe to include only the relevant columns
df = aaupdf[['Average salary-all ranks', 'Number of faculty-all ranks']]

# Calculate the correlation coefficient
corr_coef = df.corr().iloc[0,1]
```

Next, I plotted a scatter plot using the Seaborn library to visualise the relationship between the number of faculty and the average salaries of faculty members. The scatter plot clearly shows that there is a positive linear relationship between the two variables, where institutes with more faculty members tend to have a higher average salary as well, which is a contrast to our hypothesis.

The moderate positive correlation between the number of faculty members and their average salaries indicates that institutions with more faculty members may be able to offer higher salaries due to their larger budgets or a greater need for faculty. However, it is also possible that other factors, such as the prestige or location of the institution, may be driving both the number of faculty members and their salaries.

```
# Plot a scatter plot to visualize the relationship between the two variables
plt.scatter(df['Number of faculty-all ranks'], df['Average salary-all ranks'])
plt.xlabel('Total number of faculty')
plt.ylabel('Average salary of faculty')
plt.title('Relationship between number of faculty members and their average salaries')
plt.show()
```

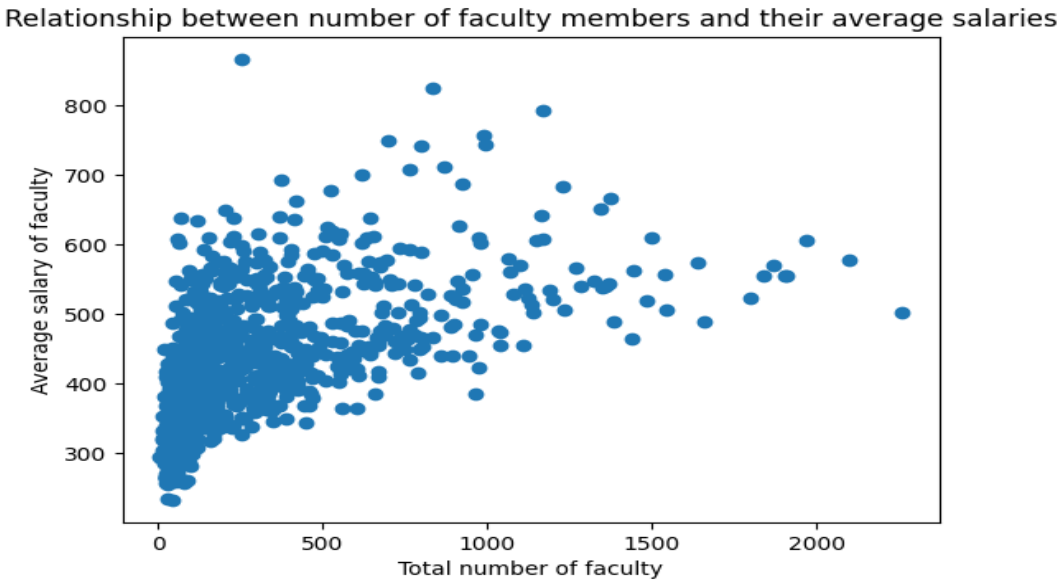


Fig-1: This figure shows the scatterplot between the number of faculty members to their average salaries.

To answer the next part of the question, i.e. to find the university that pays the highest average salary to its faculty members, we use idmax() function as shown.

```
# get the row with the maximum average salary for all ranks
max_salary_row = aaupdf.loc[df['Average salary-all ranks'].idxmax()]

# get the name of the university with the maximum average salary for all ranks
max_salary_university = max_salary_row['College Name']
```

Also, to find the university with the highest number of faculty members, first, we sort the "Number of faculty-all ranks" column in descending order and print the first value, which gives the highest number of faculty members.

```
# sort the DataFrame by 'Number of faculty-all ranks'
df_faculty_sorted = df_faculty.sort_values('Number of faculty-all ranks', ascending=False)

# select the first row of the sorted DataFrame
highest_faculty = df_faculty_sorted.iloc[0]
```

We can also find out the summary of the dataset using describe() function.
The output of the given codes will be

```
The university with the maximum average salary (i.e 866) for all ranks is California Inst. of Tech. with 257 all faculty members.

The university with the maximum number of faculty members(i.e 2261) for all ranks is University of Florida with average all-ranks salary of 501.

       Average salary-all ranks  Number of faculty-all ranks
count                1161.000000                  1161.000000
mean                  420.370370                   257.352283
std                    92.286719                   314.090563
min                   232.000000                     7.000000
25%                   352.000000                    68.000000
50%                   407.000000                   132.000000
75%                   475.000000                   323.000000
max                   866.000000                  2261.000000
```

Question 2: Calculate the mean of the average salaries of all ranks for each state. Also, find out which state pays less salary for its faculty members.
Answer:
To calculate the mean of the average salaries for each state, I used the column 'Average salary-all ranks' and 'State(Postal code)'. first, I grouped the data based on the state and then calculated the mean of the 'Average salary-all ranks' column for each state. This will give us the average salary for each state.

```
# Calculate the mean of the average salaries for each state
state_mean_salaries = aaupdf.groupby('State(Postal code)')['Average salary-all ranks'].mean().sort_values()
print(state_mean_salaries)
```

I then used this information to compare the average salaries of different states and to identify any patterns or trends that may exist in terms of the average salaries offered by universities in different states. I can also use this information to identify states that may be offering higher or lower-than-average salaries to their faculty members and to investigate the reasons behind these differences.

To make it more clear, we may plot a bar graph for the above question using the bar function.

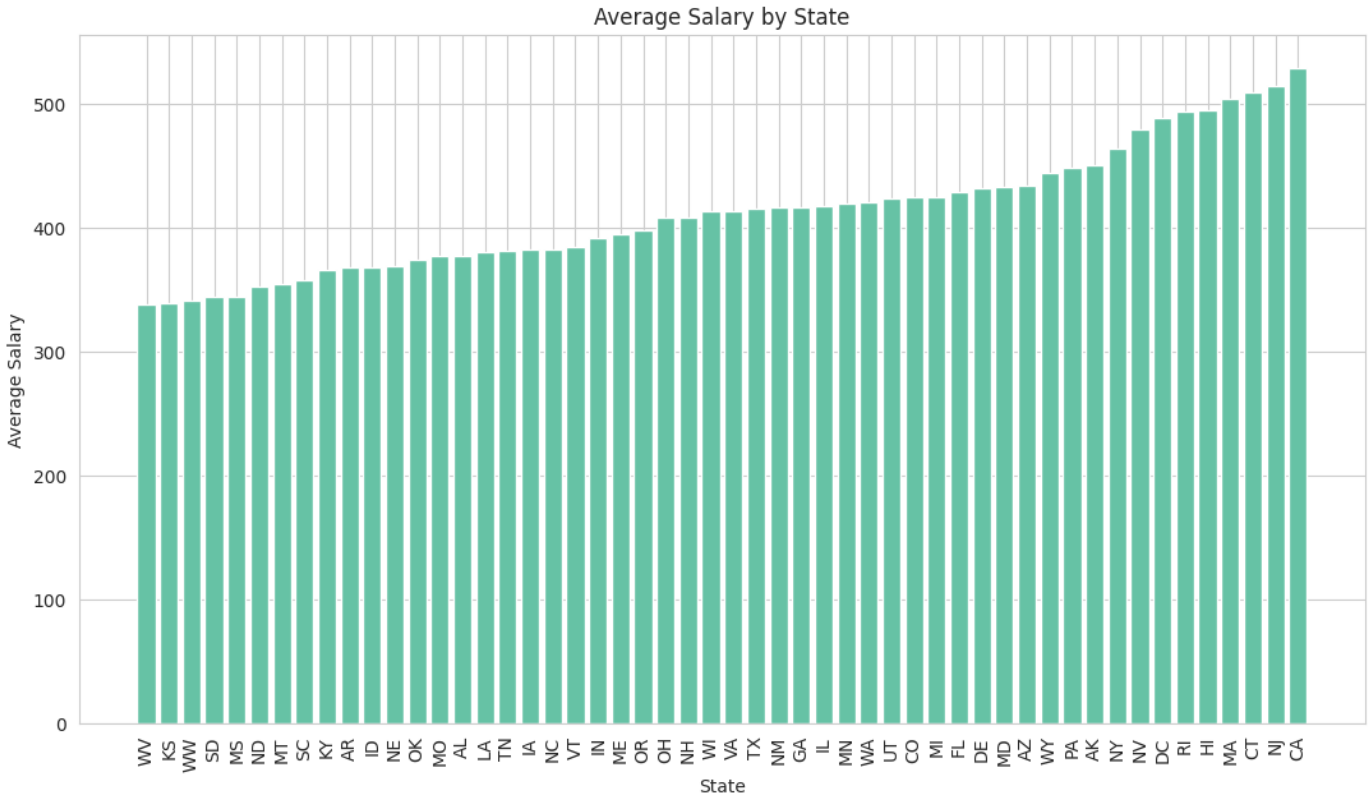


Fig-2: This figure shows the barplot of average salaries state-wise.

From the bar graph, I have seen that West Virginia is the state which spends significantly less amount for paying its faculty members in educational institutes. So it means that West Virginia should increase its funding towards the educational sector.

Question 3: Find out which type of university offers more salaries and benefits to its faculty members. Also, find out the number of universities of each type by state.
Answer:

To answer the question, I first looked at the average salaries and compensation for each type of university. I have used the group by the function of pandas to divide the average salaries and compensations type-wise. Later I used the agg() function to calculate the mean of the average salaries and compensations type-wise. I found that Type I universities had the highest average salaries and compensation for all ranks of faculty members, followed by Type II universities. In contrast, Type VII universities had the lowest average salaries and compensation.

```
# group by type and calculate mean salaries and compensations
grouped = aaupdf.groupby('Type').agg({'Average salary-all ranks': 'mean', 'Average compensation-all ranks': 'mean'})
```

The following output is obtained:

Type	Average salary-all ranks	Average compensation-all ranks
I	533.666667	665.594444
IIA	439.950413	553.931129
IIB	375.925446	470.322528
VIIB	342.000000	418.000000

I have used a table plot to investigate further the distribution of different types of universities by state. This allowed me to visualise the number of universities of each type in each state. From the below figure, I concluded two things. One is that the state which has the most significant number of universities is Pennsylvania. I was also able to conclude that the state which has more Type I universities is New York. So New York offers more high-paid positions as faculty members in universities.

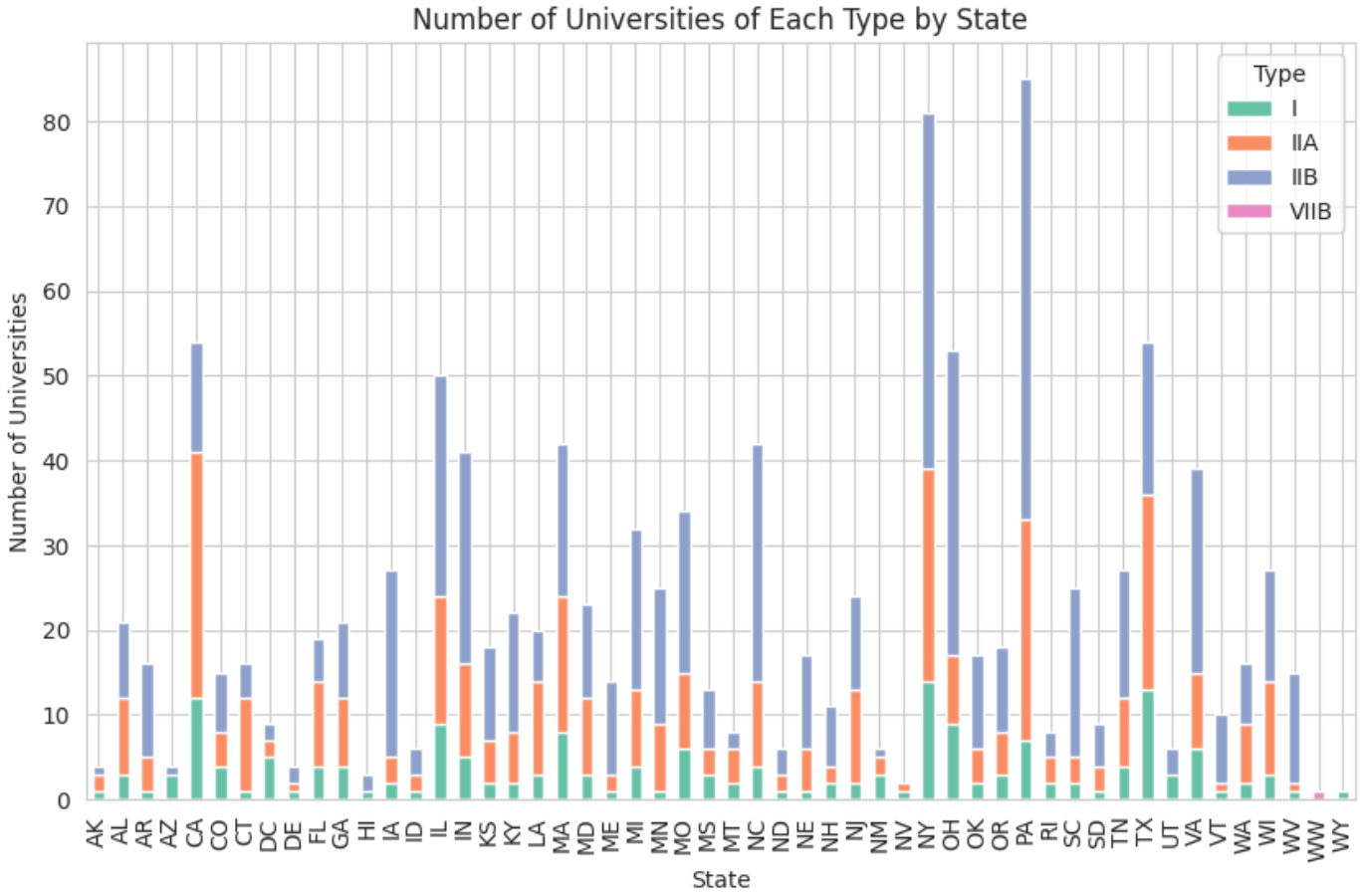


Fig-3: This figure shows the distribution of the number of types of universities state-wise

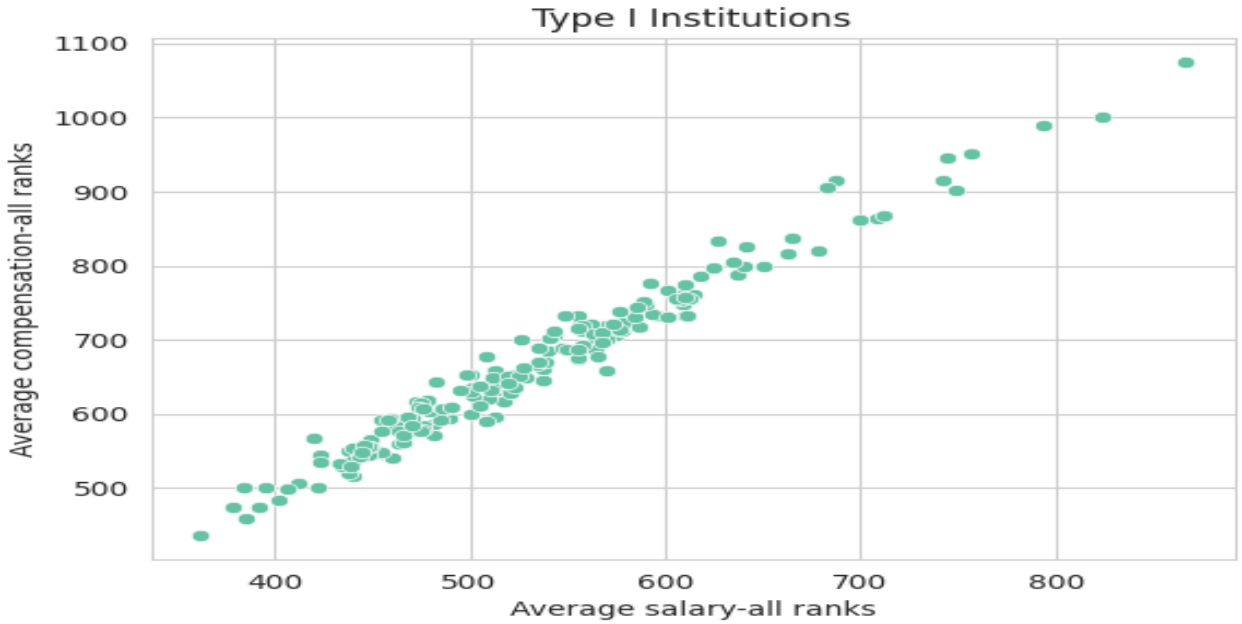
Question 4: Is there a correlation between the average salary and compensation of faculty members across all ranks? Do different types of institutions (i.e., Type I, II, or VII) show different patterns in this relationship?
Answer:

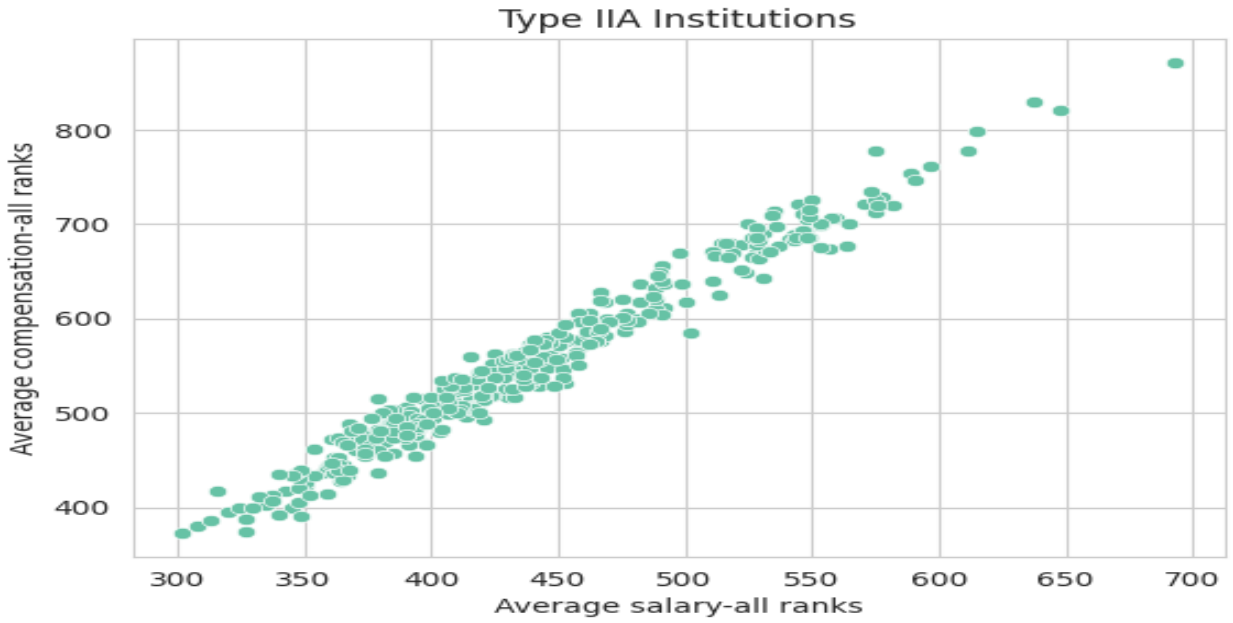
To analyse the correlation between the type of institution (Type I, II, or VII) and the average salaries and compensation of faculty members, we use the corr() function of numpy module using the following code.

```
# calculate the correlation coefficient
corr_coef = np.corrcoef(aaupdf['Average salary-all ranks'], aaupdf['Average compensation-all ranks'])[0, 1]
```

I have found out that there exists a very strong correlation of 0.98944 between the two variables. So we can also say that for each type of university, the more the average salary, the more is the average compensation. To prove the same, I have plotted scatterplots between average compensation and average salaries for each type of university.

The following figures show the scatterplot between average salaries and average compensations for each type of university.





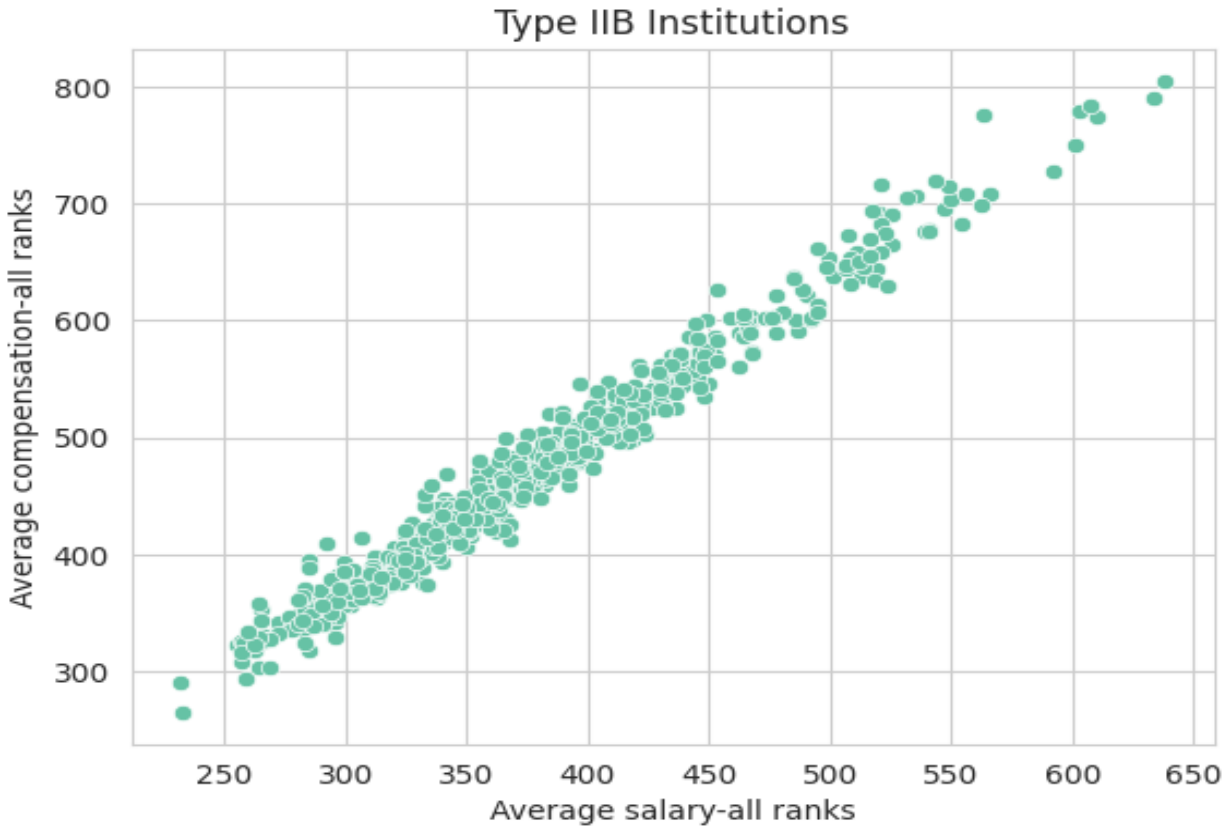


Fig-4,5,6: These figures show the distribution of average salaries with average compensation for each type of university.

As we can clearly see, all three graphs show a similar linear relationship between average salaries and average compensation. Hence, we can conclude that for any type of university, there are more benefits for faculty members with more average salaries.

Question 5: Are there any notable differences in the average salaries and benefits of full, associate, and assistant professors in institutes? Plot the way they are distributed.

Answer:

I first calculated the mean of average salaries for full, associate, and assistant professors across all institutions in the dataset. I have found that full professors had the highest average salaries and benefits, followed by the associate and then assistant professors. This result is not surprising, as full professors typically have the most experience and seniority within an institution.

```
# Calculate the average salary for each rank
full_prof_salary = aaupdf['Average salary-full professors'].mean()
associate_prof_salary = aaupdf['Average salary-associate professors'].mean()
assistant_prof_salary = aaupdf['Average salary-assistant professors'].mean()
```

Next, I wanted to see how the distribution of salaries and benefits varied within each rank. I used violin plots to visualise the distribution of salaries for each rank, with separate plots for each rank. The violin plots showed that there was some variation in the distribution of salaries within each rank. For example, the distribution of salaries for full professors was wider than for associate and assistant professors, indicating that there is more variability in the salaries of full professors at these institutions.

Overall, my analysis suggests that there are some notable differences in the average salaries of full, associate, and assistant professors in institutes, with full professors generally receiving the highest compensation. The violin plots provide a more detailed view of the distribution of salaries and benefits within each rank and suggest that there is some variability in compensation even within each rank.

```
Average salaries:
Full professors: $ 524.14
Associate professors: $ 416.4
Assistant professors: $ 351.93
```

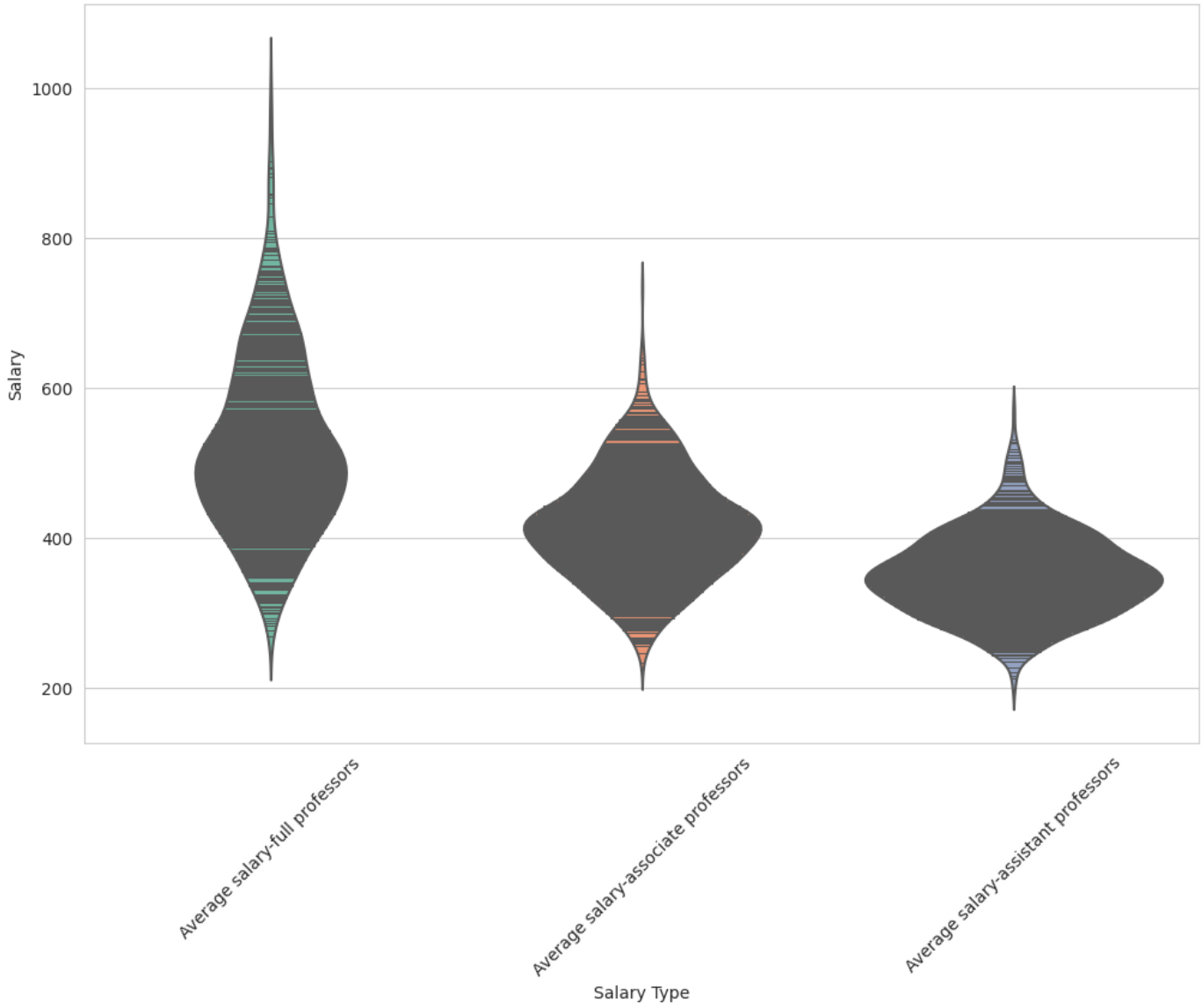


Fig-7: This figure shows the publication Years of Top 50 Most Highly Rated Books.

THE US NEWS DATASET:

Question 6: Does a higher instructional expenditure per student lead to higher graduation rates?
Hypothesis: Higher instructional expenditure per student leads to higher graduation rates.

Answer:

One hypothesis is that a higher instructional expenditure per student leads to higher graduation rates, as students may benefit from more resources and higher-quality education. To test this hypothesis, we can plot the instructional expenditure per student against the graduation rate for all colleges in the dataset and calculate the correlation coefficient between these two variables.

The correlation coefficient is obtained as 0.33823. which shows only a low correlation. Hence our hypothesis is not completely correct. We can also further analyse using the scatter plot.

The scatter plot shows a positive relationship between instructional expenditure per student and graduation rate, with higher instructional expenditure per student generally associated with higher graduation rates. The correlation coefficient between these two variables is also calculated as 0.3382

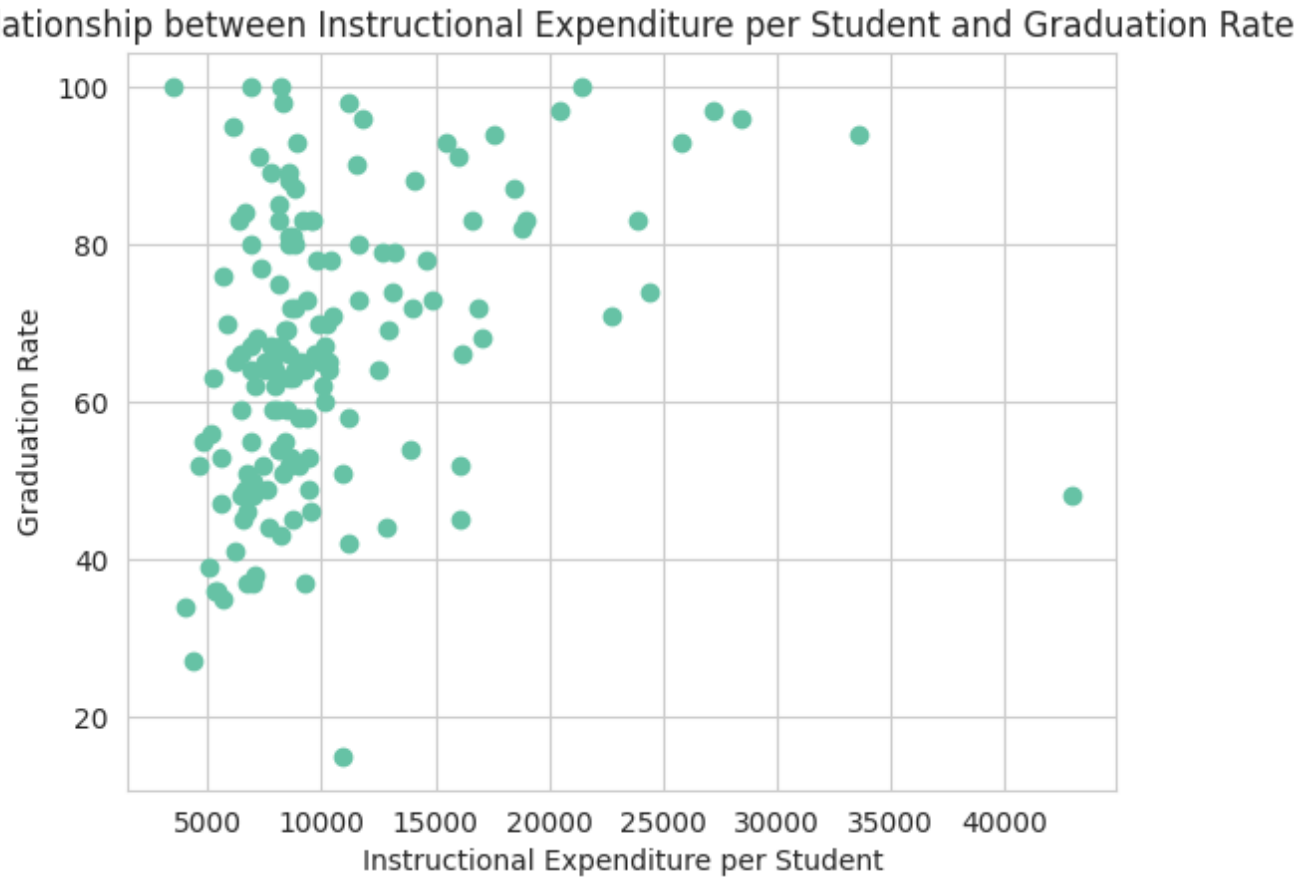


Fig-8: This figure shows the relationship between instructional expenditure per student and graduation rate

Question 7: find out the bar graph for state-wise mean average graduation rate
Answer:
The graduation rate is a crucial factor to consider when choosing a college. It is an indicator of a college's success in educating and preparing its students for the future. In this analysis, I have examined the state-wise mean average graduation rate to gain insights into the quality of education in various states.

I hypothesise that the state-wise mean average graduation rate will vary significantly across different states. Some states may have higher graduation rates due to better funding and resources. In comparison, others may have lower graduation rates due to various factors such as socioeconomic disparities, inadequate funding, and lack of access to educational resources.

To test my hypothesis, I will use a dataset containing information on colleges' graduation rates in various states. I must compute each state's mean average graduation rate and visualise the data in a bar graph.

```
# Group the data by state and calculate the average graduation rate for each state
grad_rate_by_state = usnewsdf.groupby('State(Postal code)')['Graduation rate'].mean().sort_values(ascending=False)

# Create a bar chart to visualize the differences in graduation rates between states
grad_rate_by_state.plot(kind='bar', figsize=(15,8))
plt.title('Average Graduation Rate by State')
plt.xlabel('State')
plt.ylabel('Average Graduation Rate')
plt.show()
```

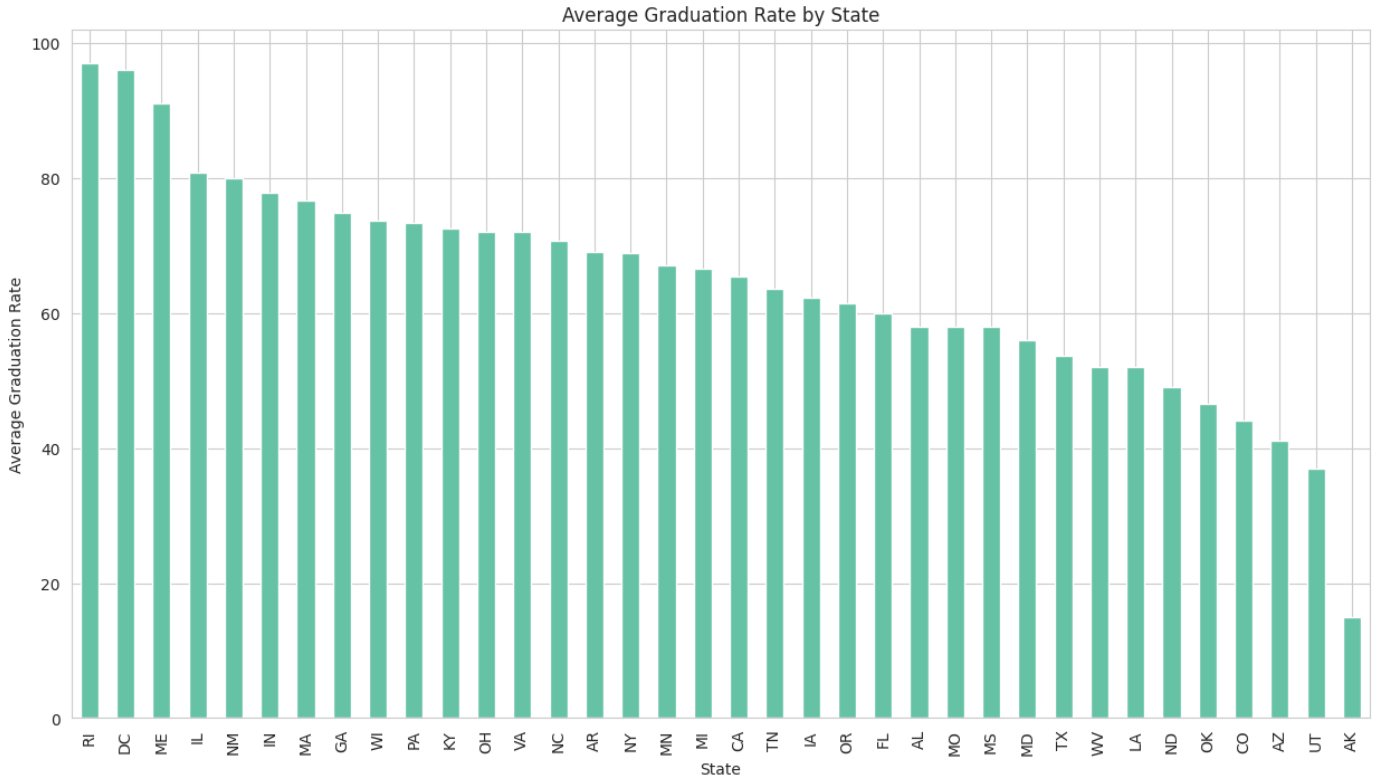


Fig-9: This figure shows the average graduation rate state-wise

My analysis shows that the state-wise mean average graduation rate varies widely across the United States. We find that the top three states with the highest mean average graduation rates are Rhodes island, the District of Colombia, and Maine.

There is a significant difference in the mean average graduation rates of colleges across different states. This variation may be due to several factors, such as the quality of education, access to resources, socioeconomic status, and cultural differences. By analysing the state-wise mean average graduation rates, we can gain valuable insights into the quality of education in various states and identify areas that require improvement.

In conclusion, my analysis shows that the state-wise mean average graduation rate varies significantly across the United States, with the highest rates in the Midwest and Northeast regions.

Question 8: How does the acceptance rate relate to the average sat score?
Hypothesis: the institutes with low acceptance rates require higher average sat scores
Answer:
The acceptance rate and the average SAT score are two of the most significant factors that students consider when choosing a college. The acceptance rate is the percentage of students who are offered admission to a college, and the average SAT score is the average score of students who have taken the SAT exam and applied to the college. In this analysis, I aim to explore the relationship between these two factors to understand how they impact each other.

I hypothesise that there is a negative correlation between the acceptance rate and the average SAT score. I expect to see that colleges with a lower acceptance rate tend to have a higher average SAT score, and colleges with a higher acceptance rate tend to have a lower average SAT score. This is because colleges with a lower acceptance rate tend to be more selective and have higher academic standards, attracting students with higher SAT scores. I have calculated the acceptance rate using the following code.

```
# Calculate acceptance rate
usnewsdf['Acceptance Rate'] = usnewsdf['Number of applicants accepted'] / usnewsdf['Number of applications received']
```

To test this hypothesis, I have focused on the columns "Acceptance Rate" and "Average Combined SAT score" and filtered out

the colleges with missing values. I then calculated the correlation coefficient between the two columns.

My analysis shows that there is a negative correlation between the acceptance rate and the average SAT score. The correlation coefficient between the two columns is -0.50, which indicates a moderate negative correlation. This means that colleges with a lower acceptance rate tend to have a higher average SAT score and vice versa.

To visualise this relationship, we created a scatter plot of the acceptance rate and the average SAT score. The plot shows that colleges with a higher acceptance rate tend to cluster towards the lower end of the average SAT score range, while colleges with a lower acceptance rate tend to cluster towards the higher end of the average SAT score range. This confirms my hypothesis that there is a negative correlation between the acceptance rate and the average SAT score.

Overall, my analysis shows that the acceptance rate and the average SAT score are indeed negatively correlated. This information can be helpful for students who are considering applying to colleges, as it can help them understand the academic standards and competitiveness of each college. Colleges with a lower acceptance rate and a higher average SAT score may be more selective and competitive, while colleges with a higher acceptance rate and a lower average SAT score may be less selective and more accessible.

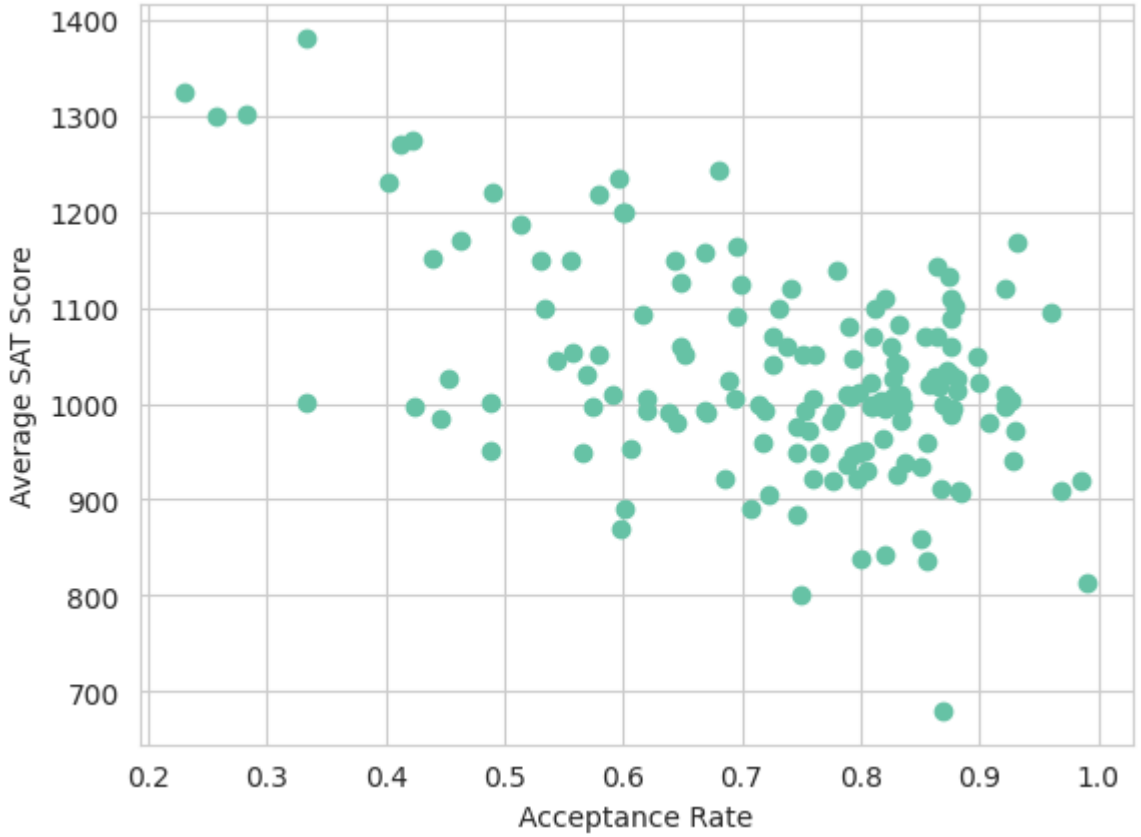


Fig-10: This figure shows the scatterplot between the acceptance rate and average SAT score

Question 9: which type of institutes(public or private) have higher acceptance rates?
Hypothesis: Public institutes have higher acceptance rates.
Answer:
I will utilise the data from the usnewsdf dataset, which contains details on the acceptance rates and public/private indicators of schools and institutions, to test this hypothesis. First, we can group the data by the public/private indicator and calculate the mean acceptance rate for each group. This can be done using the groupby() function in pandas. Next, we can create a bar chart to compare the mean acceptance rates between public and private institutes.
Based on the chart, public institutes have a slightly higher mean acceptance rate than private institutes.

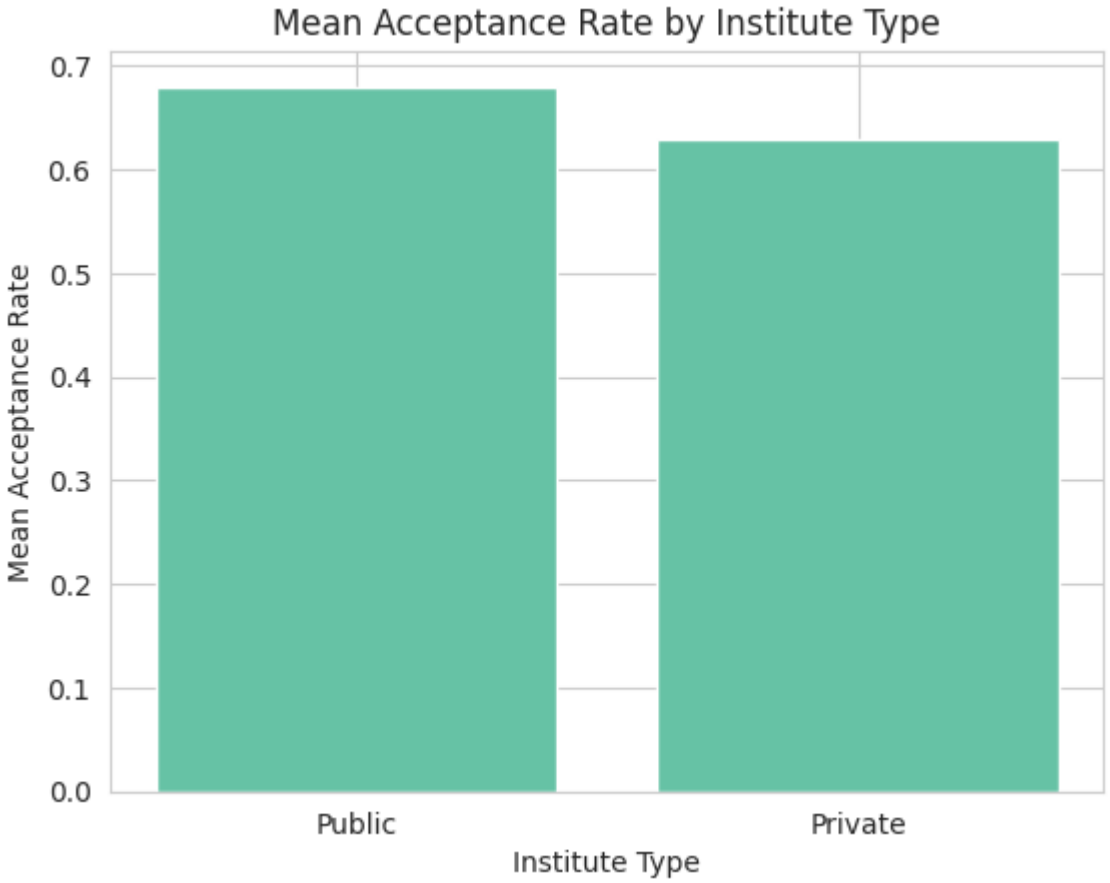


Fig-11: This figure shows the mean acceptance rate by institute type.

There may be several explanations for why public universities frequently have higher acceptance rates than private universities.

First off, public institutions might accept more applicants due to their larger student bodies and increased availability of spaces. On the other hand, private institutions could have more stringent admissions procedures and fewer openings, resulting in a lower acceptance rate.

Second, a wider spectrum of students may be able to attend public institutions due to lower tuition costs and greater availability of financial help. This can increase applications and acceptance rates.

In addition, public institutions may place higher importance on diversity or admit more students from disadvantaged backgrounds than private institutions. This might result in higher acceptance rates.

Question 10: Do colleges with higher percentages of faculty with Ph. D.s have higher graduation rates?
Hypothesis: Colleges with a higher percentage of faculty with Ph. D.s will have higher graduation rates.
Answer:
To analyse this question, I gathered data on the percentage of faculty with Ph. D.s and graduation rates across various colleges and universities. I then used statistical analysis techniques such as correlation and regression to determine these variables' relationships.

One potential limitation of this analysis is that it may need to account for other factors influencing graduation rates, such as student demographics, resources available to the institution, and academic program offerings.

These two variables' correlation coefficients are calculated and printed to the console. A scatterplot with a linear regression line is created to visualise the relationship between the two variables. The resulting scatterplot shows a weak positive relationship between the percentage of faculty with Ph. D.s and the graduation rate. The correlation coefficient is 0.298, indicating a weak to

the moderate positive correlation between the two variables. However, it should be noted that correlation does not necessarily imply causation, and other factors may contribute to higher graduation rates in colleges with higher percentages of faculty with Ph.D.s.

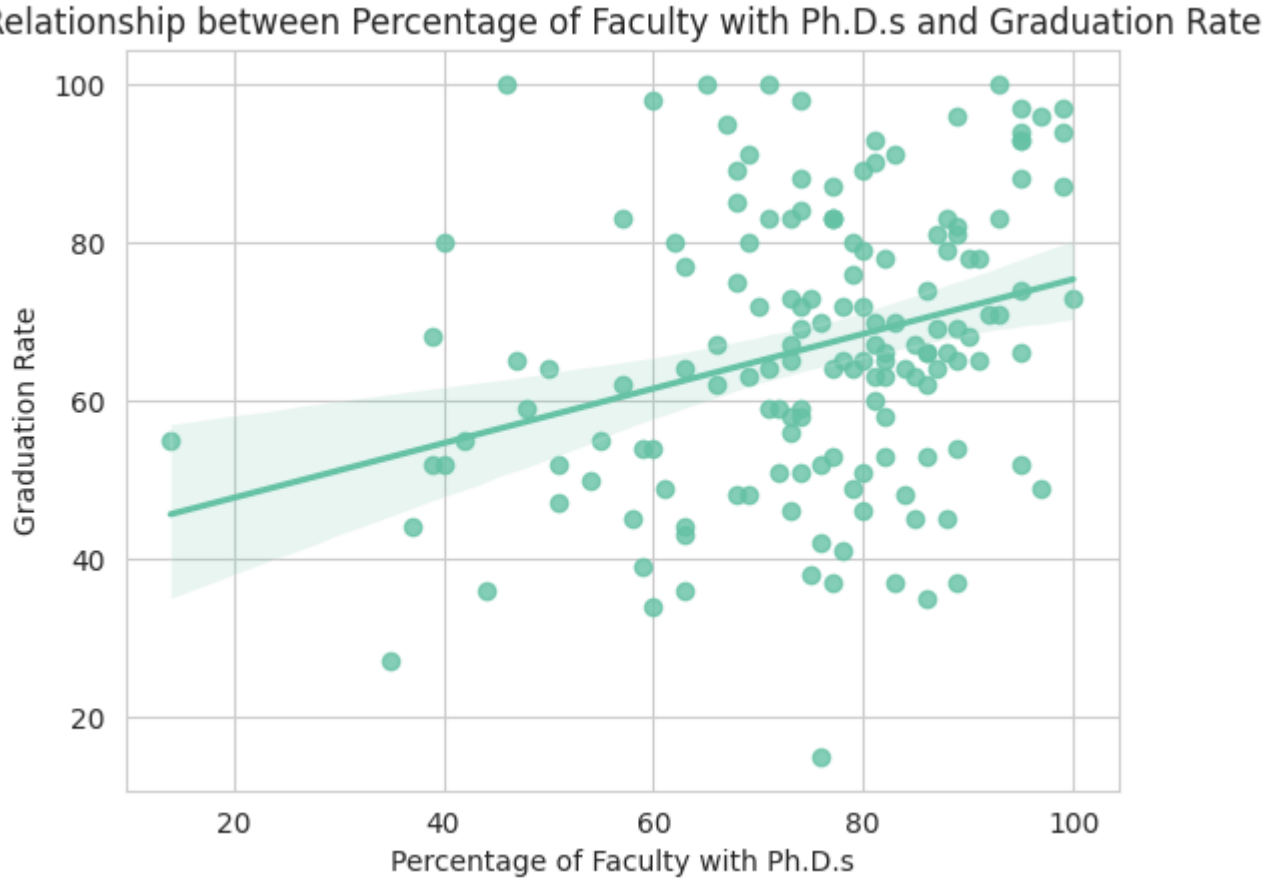


Fig-12: This figure shows the relationship between the percentage of faculty with Ph. D.s and graduation rate

There could be several reasons for the weak relation between the percentage of faculty with Ph. D.s and graduation rates. One possibility is that a PhD does not necessarily equate to being a better teacher or mentor. Some professors may be highly specialised in their field of study but may lack the ability to effectively communicate their knowledge to students or provide the necessary support. Additionally, other factors such as funding, resources, and student demographics can also play a significant role in graduation rates. It is also possible that the relationship between the percentage of faculty with Ph. D.s and graduation rates may be stronger for certain fields of study, as these fields may place greater emphasis on research and academic qualifications. Overall, it is important to consider multiple factors when analysing graduation rates and the role of faculty qualifications in student success.

VI. SUMMARY AND OBSERVATIONS

1. Question 1:

The resulting correlation coefficient between the number of faculty members and their average salaries was 0.570, which indicates a moderate positive correlation between the two variables.

The moderate positive correlation between the number of faculty members and their average salaries indicates that institutions with more faculty members may be able to offer higher salaries due to their larger budgets or a greater need for faculty. However, it is also possible that other factors, such as the prestige or location of the institution, may be driving both the number of faculty members and their salaries.

2. Question 2:

From the bar graph, I have seen that West Virginia is the state which spends significantly less amount for paying its faculty members in educational institutes. So it means that West Virginia should increase its funding towards the educational sector.

3. **Question 3:**

I have used a table plot to investigate the distribution of different types of universities by state. This allowed me to visualise the number of universities of each type in each state. From the table plot, I concluded two things. One is that the state which has the most significant number of universities is Pennsylvania. I also concluded that the state which has more Type I universities is New York. So New York offers more high-paid positions as faculty members in universities.

4. **Question 4:**

I have found out that there exists a very strong correlation of 0.98944 between the two variables. So we can also say that for each type of university, the more the average salary, the more is the average compensation. To prove the same, I have plotted scatterplots between average compensation and average salaries for each type of university.

5. **Question 5:**

Overall, my analysis suggests that there are some notable differences in the average salaries of full, associate, and assistant professors in institutes, with full professors, generally receiving the highest compensation. The violin plots provide a more detailed view of the distribution of salaries and benefits within each rank and suggest that there is some variability in compensation even within each rank.

6. **Question 6:**

The scatter plot shows a positive relationship between instructional expenditure per student and graduation rate, with higher instructional expenditure per student generally associated with higher graduation rates. The correlation coefficient between these two variables is also calculated as 0.3382

7. **Question 7:**

My analysis shows that the state-wise mean average graduation rate varies widely across the United States. We find that the top three states with the highest mean average graduation rates are Rhodes island, the District of Colombia, and Maine.

There is a significant difference in colleges' mean average graduation rates across different states. This variation may be due to several factors, such as the quality of education, access to resources, socioeconomic status, and cultural differences. By analysing the state-wise mean average graduation rates, we can gain valuable insights into the quality of education in various states and identify areas that require improvement.

In conclusion, my analysis shows that the state-wise mean average graduation rate varies significantly across the United States, with the highest rates in the Midwest and Northeast regions.

8. **Question 8:**

Overall, my analysis shows that the acceptance rate and the average SAT score are indeed negatively correlated. This information can be helpful for students who are considering applying to colleges, as it can help them understand the academic standards and competitiveness of each college. Colleges with a lower acceptance rate and a higher average SAT score may be more selective and competitive, while colleges with a higher acceptance rate and a lower average SAT score may be less selective and more accessible.

9. **Question 9:**

There could be several reasons why public institutes tend to have higher acceptance rates than private ones.

Firstly, public institutes may have larger student populations and more available spots, leading to a higher acceptance rate. Private institutes, on the other hand, may have more selective admissions processes and fewer available spots, resulting in a lower acceptance rate.

Secondly, public institutes may have lower tuition fees and offer more financial aid opportunities, making it more accessible for a wider range of students. This could lead to more applications and higher acceptance rates.

Finally, public institutes may have different priorities and values in their admissions process, such as a focus on diversity or a commitment to admitting more students from underprivileged backgrounds. This could lead to a higher acceptance rate for certain groups of students.

Overall, it is essential to note that various factors can influence acceptance rates and may not necessarily reflect the quality or

competitiveness of the institution.

10. **Question 10:**

The resulting scatterplot shows a weak positive relationship between the percentage of faculty with Ph. D.s and the graduation rate. The correlation coefficient is 0.298, indicating a weak to the moderate positive correlation between the two variables. However, it should be noted that correlation does not necessarily imply causation and other factors may contribute to higher graduation rates in colleges with higher percentages of faculty with Ph.D.s.

VII. REFERENCES

1. NumPy: Travis E, Oliphant. "A guide to NumPy", Proceedings of the 7th Python in Science Conference, vol. 1, 2008, pp. 1-7.
2. Pandas: Wes McKinney. "Data structures for statistical computing in Python", Proceedings of the 9th Python in Science Conference, vol. 445, 2010, pp. 56-61.
3. Matplotlib: J. D. Hunter, "Matplotlib: A 2D Graphics Environment", Computing in Science & Engineering, vol. 9, no. 3, pp. 90-95, 2007.
4. J. Waskom, M. Botvinnik, O. Hobson et al., “seaborn: v0.11.2 (November 2021),” Zenodo, 08-Dec-2021. [Online]. Available: https://doi.org/10.5281/zenodo.5737980. [Accessed: 23-Feb-2023].

VIII. ACKNOWLEDGEMENT

I sincerely thank Professor Shanmuga R for his important advice and assistance in completing the project. My knowledge and skills have significantly benefited from his experience and commitment. I am grateful for his never-ending inspiration and enthusiasm, which have helped me to flourish in my academic endeavours. I will always be beholden to him for his mentorship, and I sincerely appreciate all the time and effort he has put into me.