# Data Narrative Assignment - 2

Course : Probability, Statistics and Data Visualisation (ES-114)

Nimitt
*Computer Science and Engineering*
*IIT Gandhinagar*
Gandhinagar, India
nimitt.nimitt@iitgn.ac.in

***Abstract*—This document is a data narrative based on datasets containing information about various universities in the USA. It answers various questions raised during analysis of these datasets.**

***Index Terms*—Data Narrative**

## I. Overview of the Dataset

Theere are two datasets in this narrative- a) AAUP and b) USNews. Both of these datasets contain information about the US universites. Below is the overview of both these datasets.

### A. Dataset 1 : AAUP

The various ideas discussed in this dataset are:

- State codes of the universities
- Salaries of Professors
- Number of Professors

| | FICE (Federal ID number) | College name | State (postal code) | Type (I IA or IIB) | Average salary - full professors | Average salary - associate professors | Average salary - assistant professors | Average salary - all ranks | Average compensation - full professors | Average compensation - associate professors |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1061 | Alaska Pacific University | AK | IIB | 454 | 382 | 362 | 382 | 567 | 485 |
| 1 | 1063 | Univ.Alaska-Fairbanks | AK | I | 686 | 560 | 432 | 508 | 914 | 753 |
| 2 | 1065 | Univ.Alaska-Southeast | AK | IIA | 533 | 494 | 329 | 415 | 716 | 663 |
| 3 | 11462 | Univ.Alaska-Anchorage | AK | IIA | 612 | 507 | 414 | 498 | 825 | 681 |
| 4 | 1002 | Alabama Agri.&Mech. Univ. | AL | IIA | 442 | 369 | 310 | 360 | 530 | 444 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 1166 | 3825 | West Virginia Inst. of Tech. | WV | IIB | 408 | 340 | 297 | 349 | 502 | 419 |
| 1167 | 3827 | West Virginia University | WV | I | 535 | 431 | 361 | 439 | 642 | 521 |
| 1168 | 3830 | West Virginia Wesleyan Coll. | WV | IIB | 441 | 383 | 339 | 383 | 572 | 494 |

Fig. 1. dataset 1

### B. Dataset 2 : USNews

This dataset contains information about dimensions like:

- Examinations' scores for entrance to university
- Various expenses of students during their stay at university
- Alumni information
- Student to faculty ratio
- Graduation and Acceptance rates

| | 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 | ... | Board costs |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1061 | Alaska Pacific University | AK | 2 | 490 | 482 | 972 | 20 | 440 | 530 | ... | 2500 |
| 1 | 1063 | University of Alaska at Fairbanks | AK | 1 | 499 | 462 | 961 | 22 | * | * | ... | 1790 |
| 2 | 1065 | University of Alaska Southeast | AK | 1 | * | * | * | * | * | * | ... | 2250 |
| 3 | 11462 | University of Alaska at Anchorage | AK | 1 | 459 | 422 | 881 | 20 | * | * | ... | 2520 |
| 4 | 1002 | Alabama Agri. & Mech. Univ. | AL | 1 | * | * | * | 17 | * | * | ... | 1442 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 1297 | 3826 | West Virginia State College | WV | 1 | * | * | * | 18 | * | * | ... | 1700 |
| 1298 | 3827 | West Virginia University | WV | 1 | 507 | 439 | 946 | 22 | 450 | 560 | ... | 2026 |
| 1299 | 3830 | West Virginia Wesleyan College | WV | 2 | 489 | 439 | 928 | 23 | 420 | 560 | ... | 2025 |

Fig. 2. Dataset 2

## II. Questions to be addressed

The dataset is multi-dimensional and consists varied ideas. The following questions were analyzed:

### A. Dataset 1 : AAUP

- Is there any relationship between salaries and compensation given to professors?
- Which universities have higher expenditure for salaries of professors?
- Which professors have highest salaries ?
- Which type of universities have better faculty records ?
- Which professors are more in number ?

### B. Dataset 2 : USNEWS

- What is the relationship between type of the university and state it is located ?
- Which universities are more expensive and why ?
- How are acceptance and graduation rates effected by the type of university ?
- Which universities have lower student to faculty ratio and how it is effected by the fees of the university ?
- Which part of the SAT is more scoring ?

## III. Details of Libraries and Functions

Python was used as programming language for the computations and analysis. Various libraries used in this project have been listed below:

- Pandas
- Numpy
- Matplotlib

- Seaborn
- Circlify
- Plotly

```
#importing the libraries used
import matplotlib.pyplot as plt
%matplotlib inline
import numpy as np
import pandas as pd
import seaborn as sns
import plotly.express as px
import circlify
```

Fig. 3. Libraries used

All the functions used are part of the Python's built-in functions package and libraries mentioned above. The below list enlists the functions used in this assignment:

- pandas.DataFrame()
- pandas.Series()
- pandas.read-csv()
- len()
- pandas.Series.describe()
- pandas.Series.value-counts()
- pandas.DataFrame.plot()
- pandas.DataFrame.where()
- matplotlib.pyplot.scatter()
- matplotlib.pyplot.bar()
- pandas.rename()
- pandas.head()
- pandas.merge()
- pandas.DataFrame.replace()
- pandas.DataFrame.keys()
- floatinisation()
- integerization()
- pandas.DataFrame.apply()
- pandas.DataFrame.dropna()
- matplotlib.pyplot.scatter()
- numpy.polyfit()
- pandas.DataFrame.sum()
- matplotlib.pyplot.bar()
- matplotlib.pyplot.plot()
- pandas.concat()
- pandas.DataFrame.groupby()
- pandas.DatatFrame.index
- plotly.express.chloropleth
- seaborn.color_palette
- matplotlib.pyplot.pie()
- circlify.circlify()
- circlify.circlify.reverse()
- type()
- plotly.express.scatter()
- matplotlib.subplots()
- matplotlib.pyplot.hist()
- pandas.Serise.name
- pandas.DataFrame.mean()

## IV. Answers to Questions

### AAUP Dataset

#### *A. Relationship between compensation to professors and their salaries*

Compensation is a component of the pay a professor gets along with salary. The salary and compensation seem very related. On plotting a scatter plot between average salaries and compensation of professors was plotted to observe any effects. On observing that the scatter plots approach linearity line of best fit was plotted alongside the scatters and their parameters were observed.

```
The parameters of the line: [  1.29401604 -17.27263496]
```

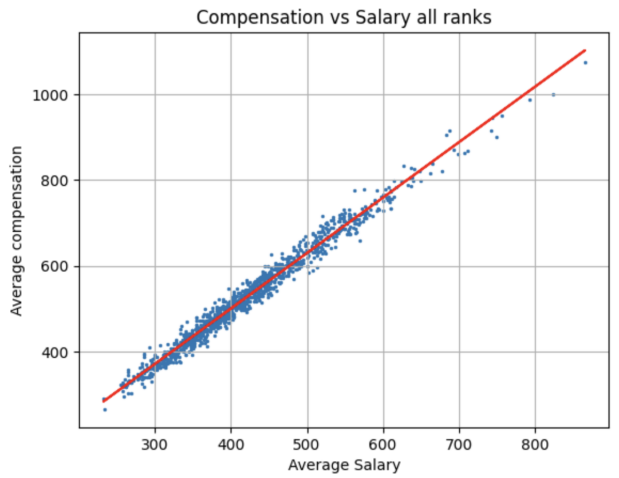


Fig. 4. Compensation V/S salary

```
The parameters of the line: [ 1.26260473 -8.12124525]
```

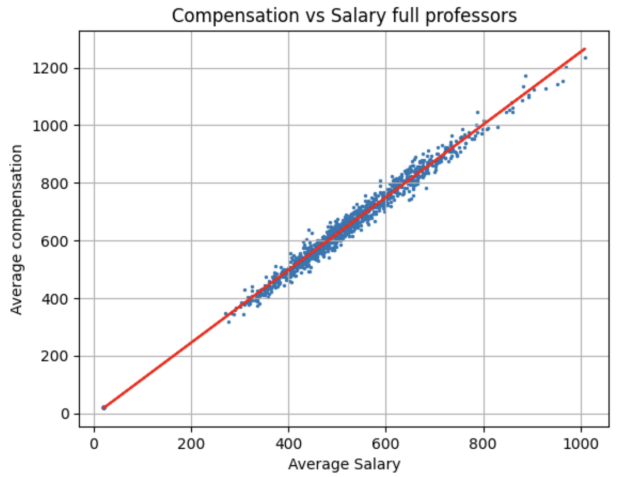


Fig. 5. Compensation V/S salary

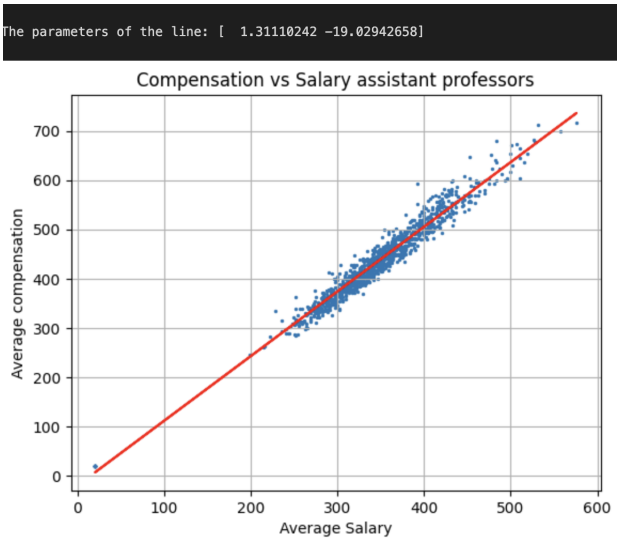


Fig. 6. Compensation V/S salary

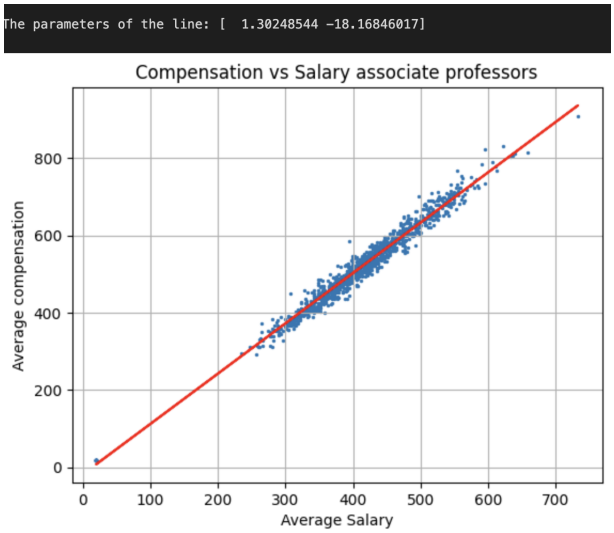


Fig. 7. Compensation V/S salary

From these plots it is evident that almost all professors get a certain factor of salary as compensation. In general, it can be inferred that a professor with high salary is subject to get a high compensation and vice-versa.

### *B. Expenditure of Universities on professors*

As the dataset provides information about salaries and compensations given to professors, the expenditure of universities on professors can be easily analyzed. To do this the following approach was used: Approach:

- University, State - code data, Type and corresponding data of Average Salary and compensation calculated for all professors, number of professors was taken into account
- To calculate average total salary Average Salary and compensation were aggregated.
- Now to come up with total expenditure of university on professor, average total salary was multiplied with total number of professors in the university.
- Thus, we arrive at data of total expenditure of each university on professors.

*1) Total Expenditure - State wise:* The total expenditure of universities reflects how prosperous the university is. It is remarkably important to understand the state-wise distribution of universities and their expenditure. To do this, the expenditure data of universities was grouped according to their state and further data of "Sum of Expenditures of all universities in the state", "Average Expenditure of Universities in the state". The same data has been represented as below:

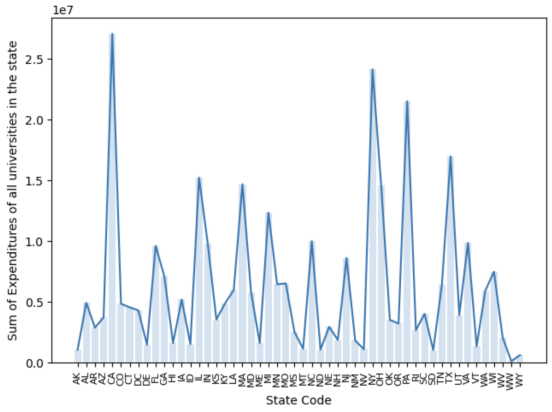


Fig. 8. Expenditure of universities based on states

We can observe some peaks in the Sum of Expenditures of all universities in the state v/s State plot. This shows universities of these states have maximum expenditure for professors. But probably, this is due to the fact these states have more number of universities evident by the following plot.

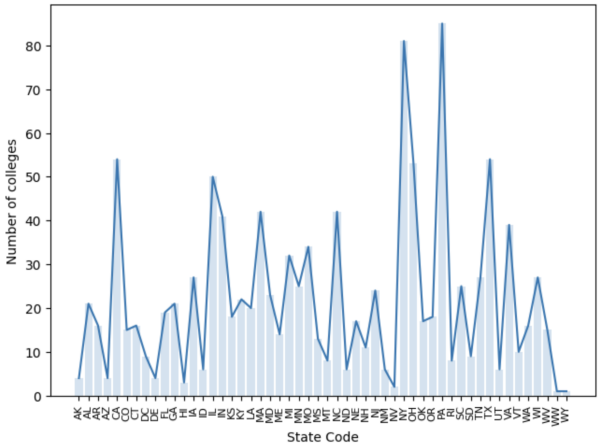


Fig. 9. Output

For better analysis, we need to look at "Average expenditure of Universities in the v/s State". It gives better visualization of

which has more universities with greater expenditures for professors. The chloropleth shows the regions with universities

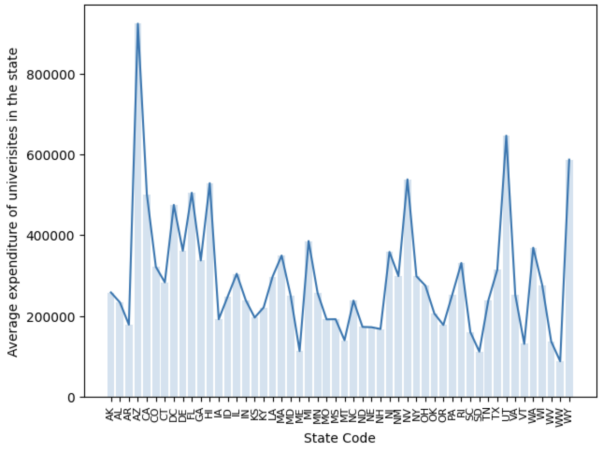


Fig. 10. Number of Universities in each state

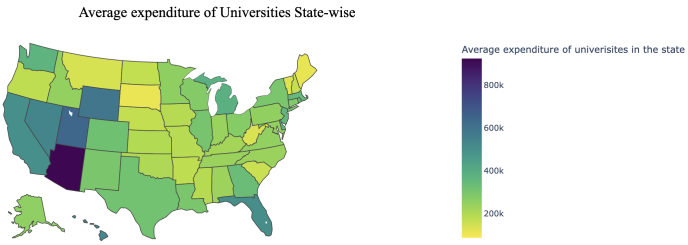


Fig. 11. Average Expenditure of universities based on states

having higher expenditure on professors.

*2) Total Expenditure Type-wise:* Now, let's look at Universities with higher budgets according to their type. So, on grouping the universities according to their type and then finding the average budget of universities of each type the following was obtained:

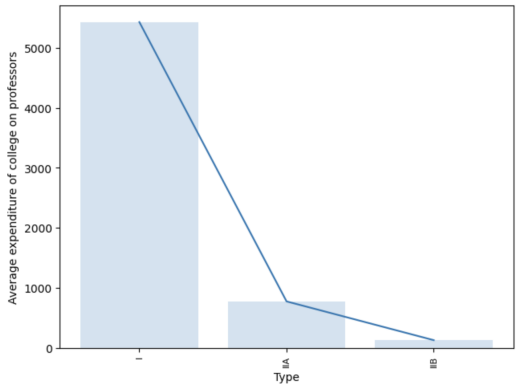


Fig. 12. Average Expenditure of universities based on type

This gives a very prominent fact that type - I universities typically have higher professor budgets than the other type of universities. But, to consolidate this fact let's look into number of universities in each type.

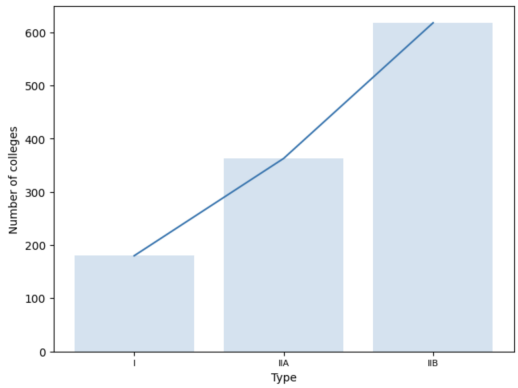


Fig. 13. Number of Universities based on type.

The type - I has least number of universities but have highest average expenditure for professors.

### *C. Salary distribution of professors*

There are three categories of professors : a) Full professors b) Associate professors and c) Assistant professors. In order to look at the data of which professors have higher salary share, the aggregate of compensation and salary was taken into account. To compare the shares of each the mean total salary was obtained for each category of professors and then the total share of each category was evaluated as below:

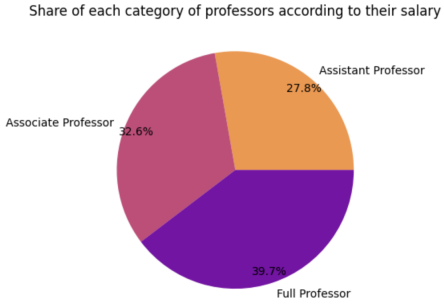


Fig. 14. Percentage share of professors in expenditure

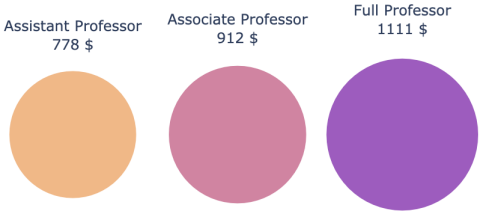


Fig. 15. Average salaries of professors

Clearly, it is evident that in general, Full professors have highest salaries followed by Associate professors and then Assistant professors.

### *D. Faculty size - Type-wise*

It is interesting to note that Type - I universities had the highest budget for professors. Further, it is important to extrapolate the similar kind of strategy to analyze the number of professor in each university. In order to this, the distribution of number of professors type-wise was plotted according to their type.

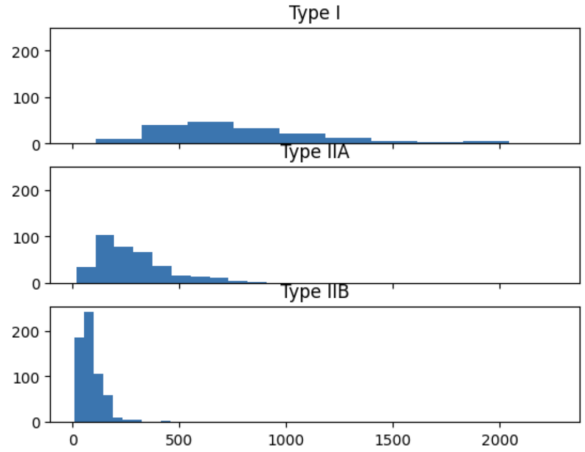


Fig. 16. Distribution of faculty size type wise

These distribution gives some prominent conclusions :

- The Type - IIB universities have the highest number of professors followed by type - IIA and I universities
- The Type - IIB universities are maximum in number while type - I universities are least in number

From the previous questions, it was concluded that generally type - I universities have higher expenditure for professors. But, it is very interesting to note that even a type - I university has lesser number of professors it spends more on salaries of professors. Clearly, a type - I university has better number of professors, moreover, who are paid more. It shows that type - I universities are generally more prosperous than the other type of universities.

### *E. Share of each category of professors in total number of professors*

To analyze the share of each category of professors in total number of professors we need to look at Number of total professors against each category of professors and use linear regression to approximately measure the proportion of total number of faculty members and number of professors in each category and then use this factors data to know the share of each category of professor in total number of faculty.

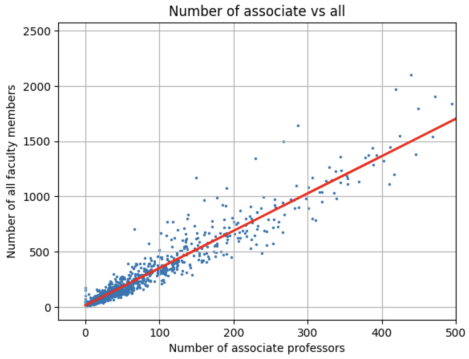


Fig. 17. All vs associate

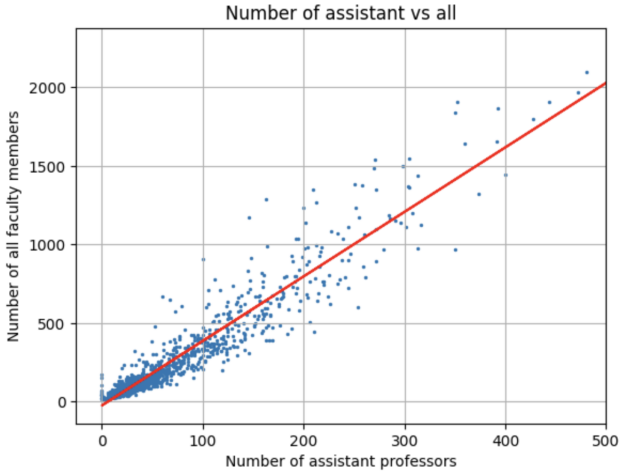


Fig. 18. All vs assistant

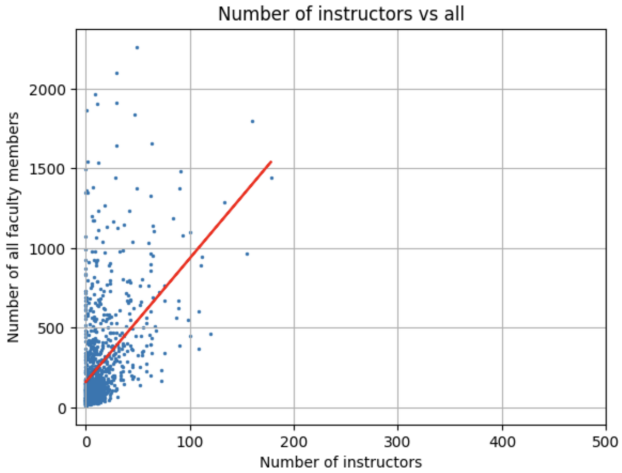


Fig. 19. All vs instructors

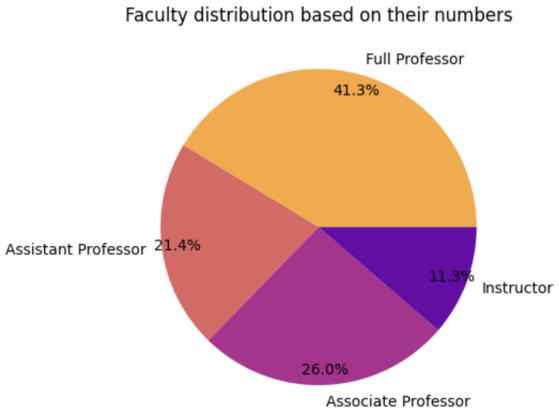


Fig. 20. Output

It shows that full professors are maximum in number and then followed by associate and then assistant. Also, it was shown in one of the previous questions that the salary distribution of professors also follows the same order. So, full professors lead the way in size as well as salary.

*USNews*

### *F. Number of Universities in each state*

It is possible that the location of private or public universities is concentrated in only some of the states. To verify this fact, the data Universities was grouped according to type of university and their state to obtain the following.

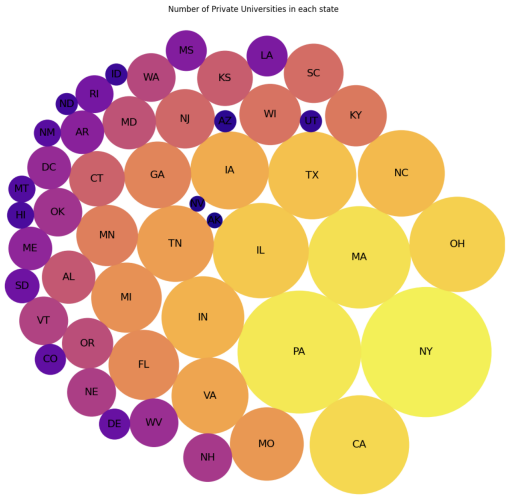


Fig. 21. Private Universities

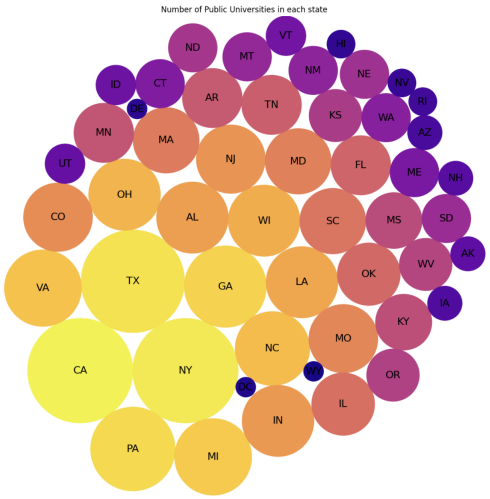


Fig. 22. Public Universities

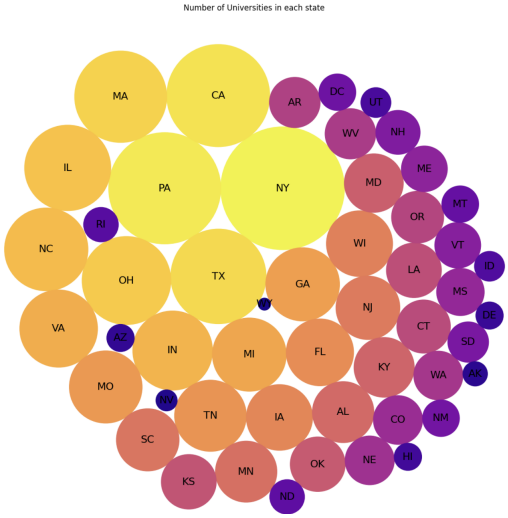


Fig. 23. Total Universities

The same set of universities have more private universities and public universities. It shows that a certain states have higher number of universities and the number of private universities and public universities is comparable.

### *G. Comparison of Universities on the basis of the expense of studying*

To get the expense of studying in a university all the fees were added to get the aggregated amount. This data was then studied by grouping according to type of university.

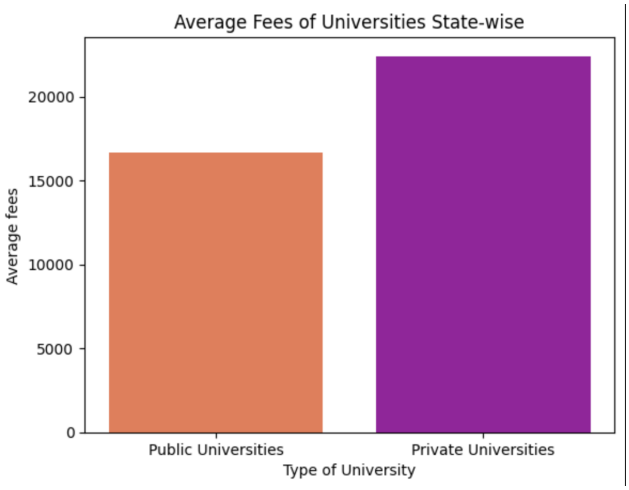


Fig. 24. Student Expenses according to type of university

It is conclusive that private universities in US are more expensive than public universities.

### *H. Comparison of private and public universities on the basis of Acceptance rate*

Acceptance rate is a good measure to describe a university. To calculate the acceptance rate, the data of number of applications and acceptance was taken into account and the same was calculated. The following shows the required analysis.

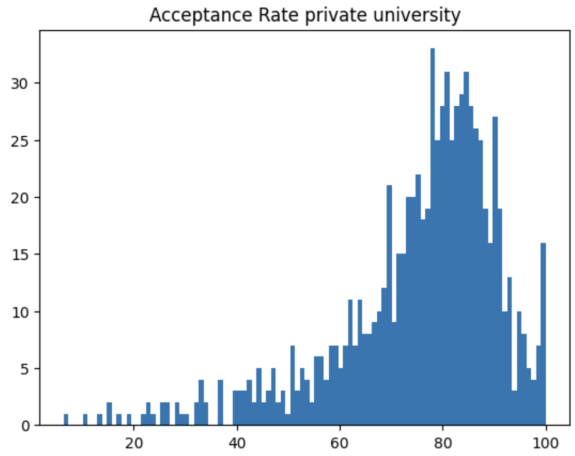


Fig. 25. Acceptance rate - Private

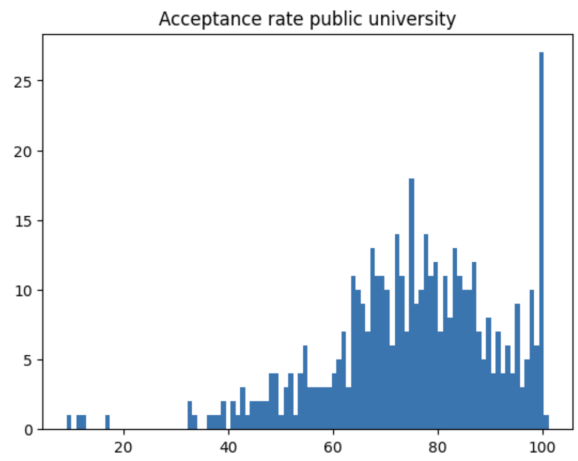


Fig. 26. Acceptance rate - Public

```
Private : 75.49370909354673
Public  : 75.31902409968704
```

Fig. 27. Output.

The acceptance rates of private and public universities are quite comparable. This is in contrast with the graduation rates:

```
Graduation rates:
Private : 65.7079326923077
Public : 50.93404255319149
```

Fig. 28. Output.

This difference between graduation rates predominantly explains that private universities have higher competence and seriousness among students.

### *I. Relationship between Student to faculty ratio and Fees of the college*

The idea of Student of faculty ratio is deterministic to determine prosperity of a university. So, to look into this factor. Student to faculty ratio was plotted against total fees of the university.

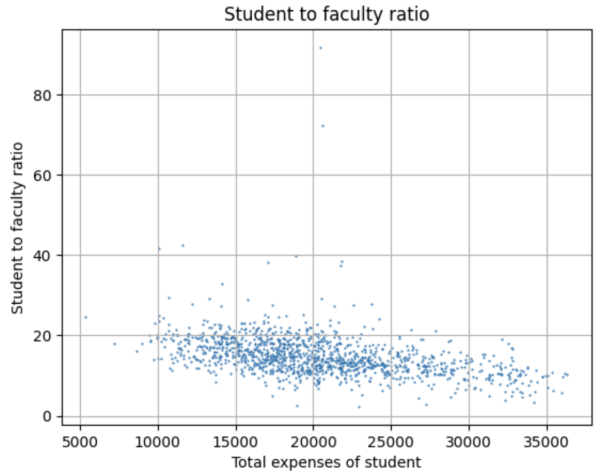


Fig. 29. Student to faculty ratio for universities according to their fees

The universities with high total expenses have a slightly smaller student to faculty ratio. Probably, this infact explains the high expenses in these universities!

### J. Math SAT marks v/s Verbal SAT marks

SAT exam is one of the most popular exams for taking entrance in US universities. SAT has two parts : a) Math SAT and b) Verbal SAT. To look into which section student score much, a plot between Math SAT marks and Verbal SAT marks was linearly regressed to get the factor the scores of two sections differ.

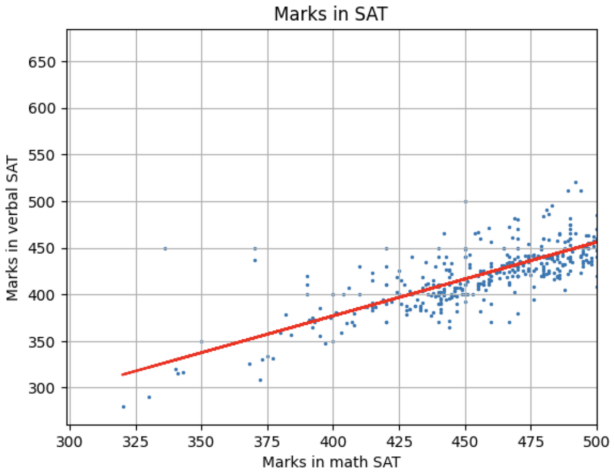


Fig. 30. Math Marks v/s verbal marks

So, roughly the Math section of the SAT is more scoring than the verbal part. This also proves that the students score less in verbal section which is probably due to the fact that a good proportion of people non-natives and less fluent in the language.

## V. Summary of Observations

On analysis of these datasets following observations were inferred:

### A. Dataset 1 : AAUP

- The salaries and compensation awarded to the professors are related. The professors with higher salary are being awarded higher compensation.
- It was noted that type - I universities have lower faculty size yet bigger expenditure for the professors when comparing with other type of universities. This predominantly explains the superiority of type - I universities.
- Full professor have higher salaries than the associate professors while the latter have higher salaries than assistant professors.
- Type - I universities have very good faculty records with them spending the highest on professors when compared with other universities.
- Full professors constitute the largest part of the faculty, as they make up about 40 percent of the total faculty members.

### B. Dataset 2 : USNews

- Certain states have greater number of universities than the other states. But the distribution of private and public universities is uniform in states.
- Private universities are more expensive than Public universities
- On average, private and public universities have comparable acceptance rates. But, in contrast to this public universities have much lower graduation rates probably, due to lack of competency in these universities.
- Colleges with higher student expenses and fees spend more on their infrastructure and faculty. Due to this, the student to faculty ration is lower in these universities.
- Generally, the Math and Verbal scores of SAT are comparable, but all the times Math scores are higher than verbal scores. This proves the toughness of verbal SAT over Math SAT.

## VI. Acknowledgment

I would like to thank Prof. Shanmuga R for helping me through this assignment. He taught the concepts used for this assignment. His teachings have a great contribution in this assignment. My friends and the teaching assistants also shaped my ideas and thoughts to help me complete this assignment. I would like to thanks all of them.

## VII. References

## References

[1] Pandas Documentation, Available: Pandas Documentation
[2] Numpy Documentation, Available: Numpy Documentation
[3] W3 Schools,W3 Schools website
[4] The dataset was taken from Github online Github dataset
[5] Stack Exschange, Stack Excahnge Site