ES-114 Data Narrative 2

Chakradhar Basani

Roll No.: 22110054

Material Science and Engineering Department

Prof. Shanmuga R

IIT Gandhinagar

Gujarat, India

chakradhar.basani@iitgn.ac.in

***Abstract*—In this report, we are writing Data narrative for the dataset provided. It gives insight on how python, programming in general can make our lives easy.**

Keywords—Python, programming, Data Narrative.

I. Overview of the Dataset

The United States Colleges dataset, avaialable at http://lib.stat.cmu.edu/datasets/colleges/ , is a comprehensive resource containing information on 777 coleges and universities across the US. This dataset offers valuable insights into various aspects of higher education institutions, making it ideal for research and data analysis projects.

Key data points include college name, location, affiliation(public or private), tuttion fees, room and bord expenses, financial aid, student-to-faculty ratios, and graduation rates. Additionally, the dataset provides information on applicant numbers, acceptance rates, enrollment rates, alumni giving, and average SAT scores.

In summary, the US college dataset is an extensive resource for exploring and analyzing a widde range of aspects related to higher education institutions in the United States, offering numerous research oppurtunities and data-driven insights.

II. Scientific Questions /Hypothesis

1. What is the probability distribution of average salary for the full time professors in all type of colleges?

2. what is the covariance between the average salary of associate professors and the average salary of full professors in colleges located in Illinois and Indiana?

3. What is the probability that a randomly selected college from the dataset is of type IIA or IIB?

4. What is the probability that a randomly selected Type I college has a higher average compensation for assistant professors than the average compensation for assistant professors in Type IIA colleges?

5. What is the probability that the average salary of associate professors is greater than 350 in the colleges of Arizona state?

6. What is the probability distribution of the average MAT scores of students in both public and private sectors in Alabama and Arizona?

7. What is the probability that the room costs in private sector colleges is greater than the room costs in public sector colleges?

8. Which state has the most number of public sector universities with student/faculty ratio greater than 15?

9. Which state's universities have better average MAT and SAT scores?

10. In which of the following states is the probability of getting accepted into universities higher: California, Alabama, Florida or Georgia?

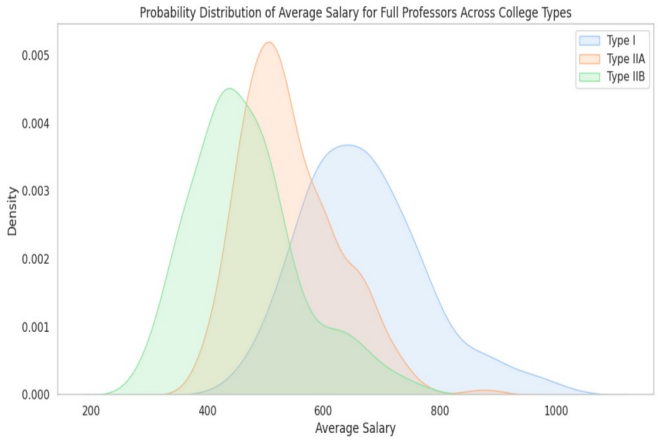
III. Details of Libraries and Functions

The following libraries and functions are used to answer the above questions.

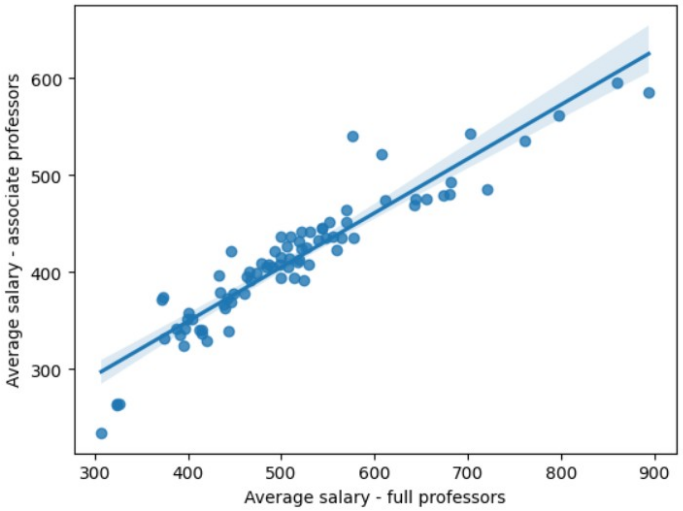
- Pandas
- Matplotlib
- Seaborn
- pd.read_excel()
- sns.regplot()
- sns.kdeplot()
- plt.pie()
- sns.histplot()
- sns.boxplot()
- sns.countplot()
- .groupby()
- sns.lineplot()
- sns.violinplot()

IV. Answers to the Questions

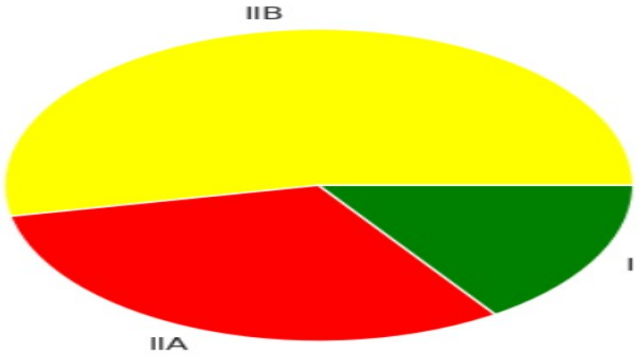
1. Probabilty distribution of average salary for the full time professors in all type of colleges.



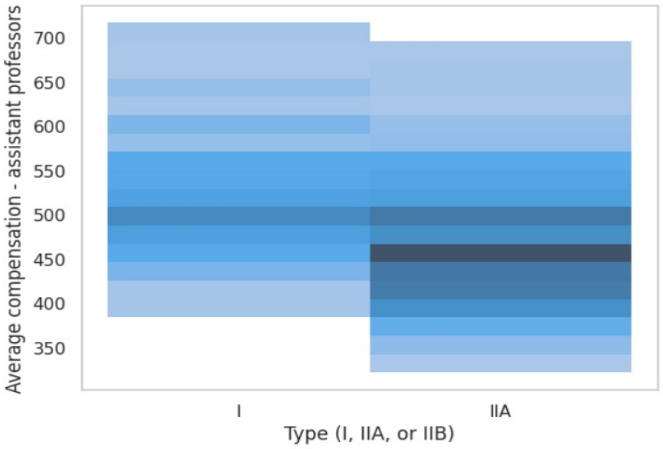
2. The covariance between the average salary of associate professors and the average salary of full professors located in Illinois and Indiana is 7509.95



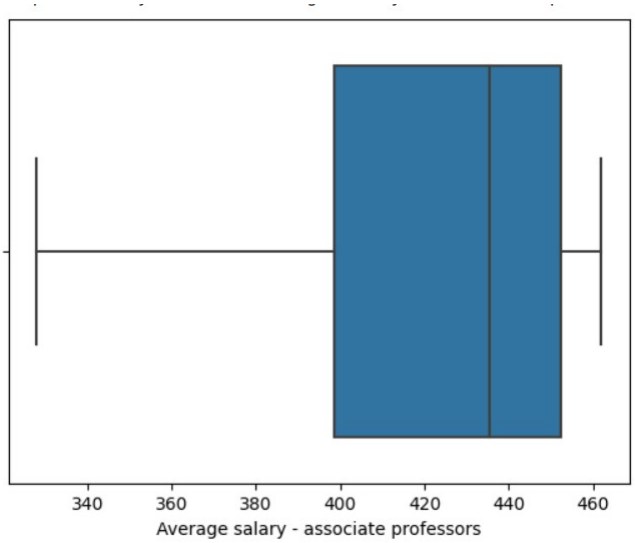
3. The probability that a randomly selected college from the dataset is of type IIA or IIB is 0.84



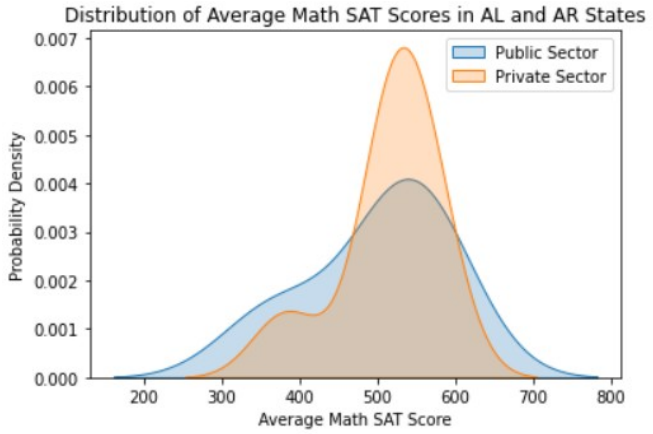
4. The probability that a randomly slected Type I college has a higher average compensation for assistant professors than the average compensation for assistant professors in type IIA colleges is 0.81



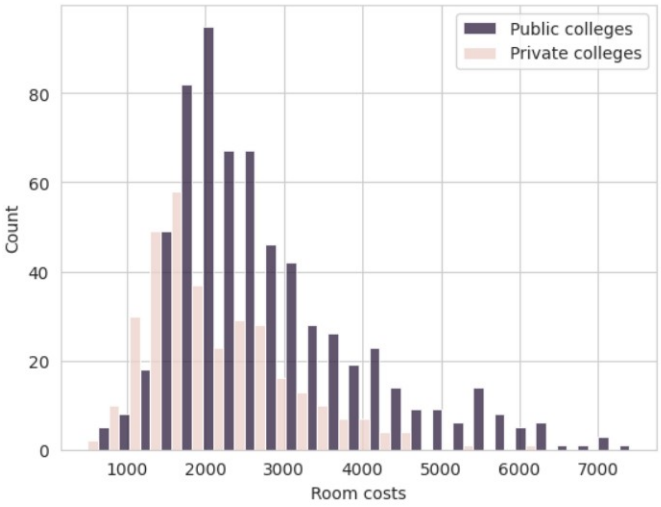
5. The probabilty that the average salary of associate professors is greater than 350 in the colleges of Arizona state is 0.75



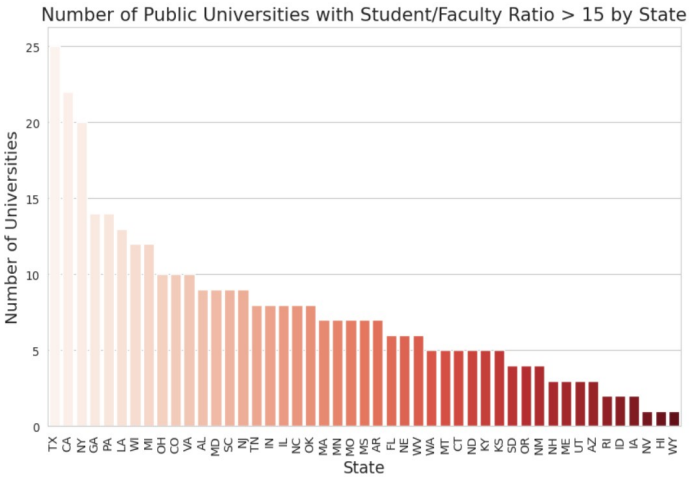
6. The probability distribution of the average MAT scores of students in both public and private sectors in Alabama and Arizona



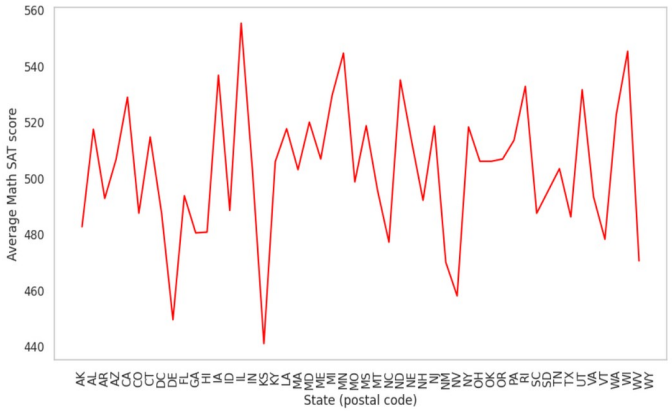
7. The probability that the room costs in private sector colleges is greater than the room costs in public sector colleges is 0.44



8. The state with the most number of public sector universities with student/faculty ratio greater than 15 is Texas with 25 universities



9. The state with the highest average Math SAT score is Illinois with a score of 555.25 and the state with the highest average Verbal SAT score is Illinois with a score of 503.12



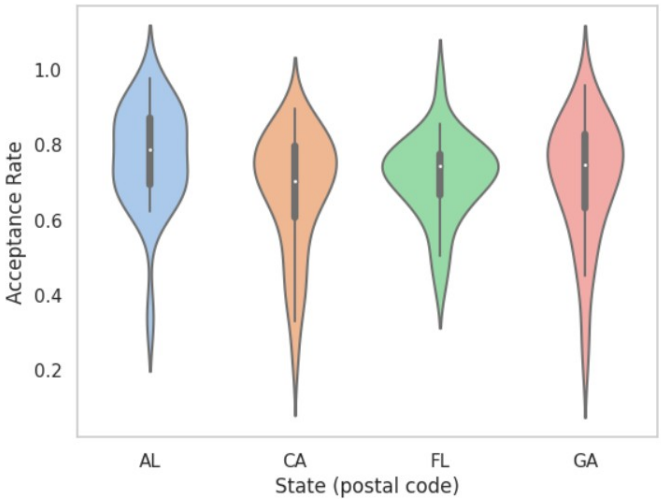
10. Probability of getting accepted into universities is higher in Alabama state

AL : 0.7627210

CA : 0.675506

FL : 0.713595

GA : 0.708093



Summary of the Observations

1. From the graph, we can observe that the distribution average salary for full professors is higher in type IIA colleges, followed by type IIB colleges and type I colleges
2. The value of covariance is 7509.95 . we can observe that there is a positive relationship between these two variables. This means that when the average salary of full professors increases, the average salary of associate professors also tends to increase.
3. We can observe that the most of the colleges in the dataset are of type IIB colleges. Second most number of colleges are of type IIA and least number of colleges are of type I
4. From the solution and graph, we can observe that type I colleges provide better average compensation for assistant professors than type IIA colleges.

5. Since the calculated probability is 0.75, it means that 75% of the colleges located in Arizona have an average salary of associate professors greater than 350. This suggests that the majority of colleges in Arizona pay their associate professors an average salary above 350.
6. From the KDE plot, we can see that the distribution of Average Math SAT scores is slightly higher in the private sector than in the public sector in both AL and AR states.
7. From the graph, we can notice that for a given room cost, there are more number of public colleges than private colleges
8. From the graph, we can observe that the state with most number of public sector universities with student/faculty ratio greater than 15 is Texas with 25 universities, followed by California with 22 universities. Nevada, Hawaii and Wyoming are last in the list with 1 university each.
9. From the graph, we can say that Illinois has the highest average math SAT score and Kansas has the least average math SAT score.
10. Universities of Alabama state has high acceptance rate, followed by universities of Florida , Georgia and California respectively.

ACKNOWLEDGMENT *(Heading 5)*

I would like to thank the developeres of python and its libraries for creating this powerful coding language and libraries which made the analysis possible. I would also like to thank Prof. Shanmuga R who provided us with the oppurtunity to learn these llibraries.

REFERENCES

- Dataset - *Index of /datasets/colleges. Accessed March 30, 2023.* http://lib.stat.cmu.edu/datasets/colleges/ .
- *Pandas - "Pandas Documentation#." pandas documentation - pandas 1.5.3 documentation. Accessed March 30, 2023.* https://pandas.pydata.org/docs/
- *Matplotlib - "Matplotlib 3.7.1 Documentation#." Matplotlib documentation - Matplotlib 3.7.1 documentation. Accessed March 30, 2023.* https://matplotlib.org/stable/index.html .
- *Seaborn - "Statistical Data Visualization#." seaborn. Accessed March 30, 2023.* https://seaborn.pydata.org/ .