



Northeastern University

Module 5 Assignment

Case Study

Srinidheesh Ranganathan

College of Professional Studies, Northeastern University

Spring 2022'B

ALY6060: Decision Support and Business Intelligence

Prof. Michael Cavallo

June 27, 2022

Introduction

The first stage of the career planning process is self-assessment, in which we uncover our personality, interests, aptitudes, and values, while the second stage is career discovery. After using a variety of techniques to gather this data, we have a list of professions that would be a good fit for someone with our set of qualities. We as humans frequently give in to the allure and motivation of knowing the answer to the question "How much money would I make out of it?" when evaluating various disciplines. The majority of people consider salary and perks as essential yet debatable factors while searching for work opportunities across a range of industries.

While we do not have time to study how much people would make in various professions, we will discuss how much money a Data Enthusiast could make by looking at pay data for numerous positions that go under the general phrase "Data Science" in this assignment. We consider positions which I as a Data enthusiast is interested in such as the following;

- a) Analytics Engineer
- b) Applied Machine Learning Scientist
- c) Data Analyst
- d) Data Engineer
- e) ML Engineer

Our goal is to analyze and visualize the entry-level and mid-level earnings for the aforementioned designations in order to determine which job offers the best pay. We also look for differences between entry level and mid-level positions, as well as any insights that would encourage me to pursue that designation.

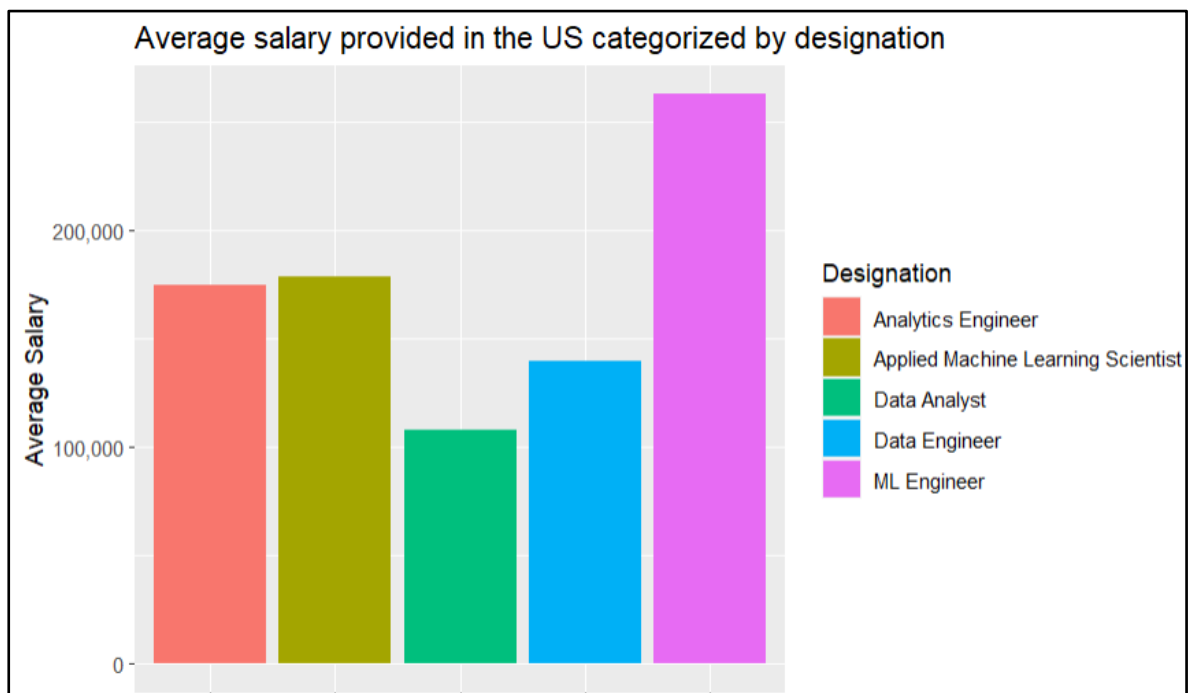
Dataset Introduction

- The dataset which we choose is the “Data Science Job Salaries”, downloaded from Kaggle.
- It has salary information on various data science related designations with about 608 rows and 11 columns
- The details of the columns are as follows:
 1. **Work year:** Year when the salary was paid
 2. **Employment type:** Full-time/Contract/Part-time
 3. **Experience level:** Entry/Mid/ Senior level
 4. **Job title:** Name of the role, i.e., ML Engineer, Data Engineer< Financial Data Analyst etc.
 5. **Salary:** Annual income paid
 6. **Salary currency:** The currency in which the salary was paid
 7. **Salary in USD:** Salary paid in another currency but converted to US Dollars.
 8. **Employee residence:** Home country of the employee
 9. **Remote ratio:** Ratio between work done remotely and on-site
 10. **Company location:** Resident country of the company
 11. **Company size:** Size of the company, Small, Medium, or Large

Basic Analysis & Visualizations

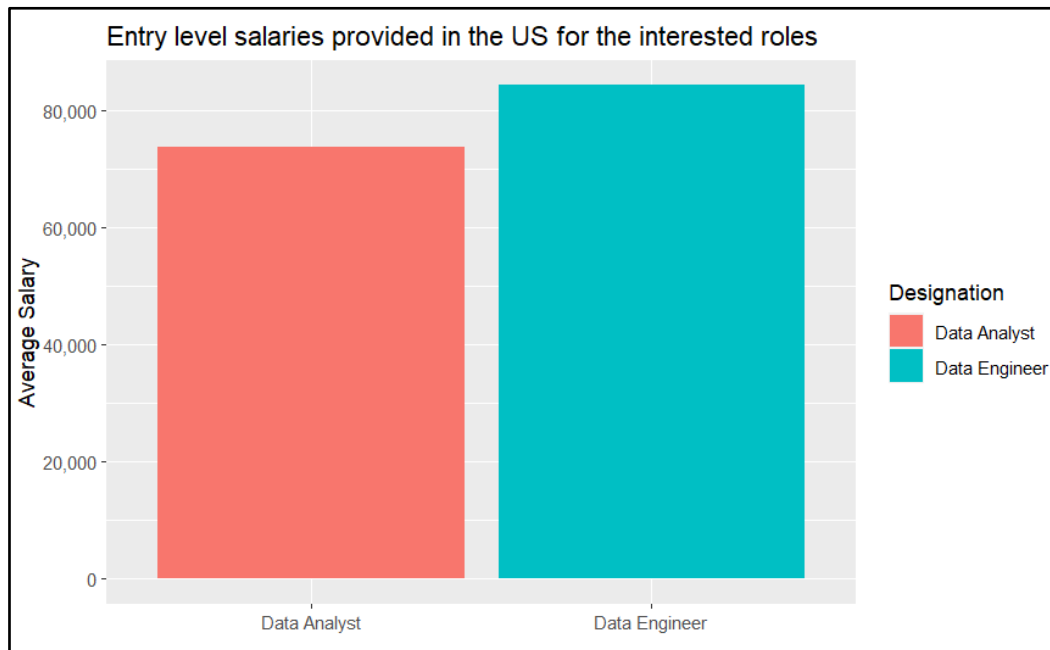
- We typically import our dataset for our fundamental analysis and then clean it to meet our criteria. When we first examined our dataset, we discovered over 100 different data science professions, including Data Production Analyst, BI Data Analyst, Data Architect, Cloud Data Engineers, and others. Another finding was that this dataset includes pay information from several additional nations. However, due to our location in the United States, we have restricted the country to the United States.

- Furthermore, by restricting the stats to only the designations I am interested in, we opt to focus only on the professions that I personally find interesting and am qualified for.
- The data was narrowed down to the following roles: Analytics Engineer, Applied Machine Learning Scientist, Data Analyst, Data Engineer, and ML Engineer. We also displayed the average salary for each position in the USA. It is seen in the following visualization.

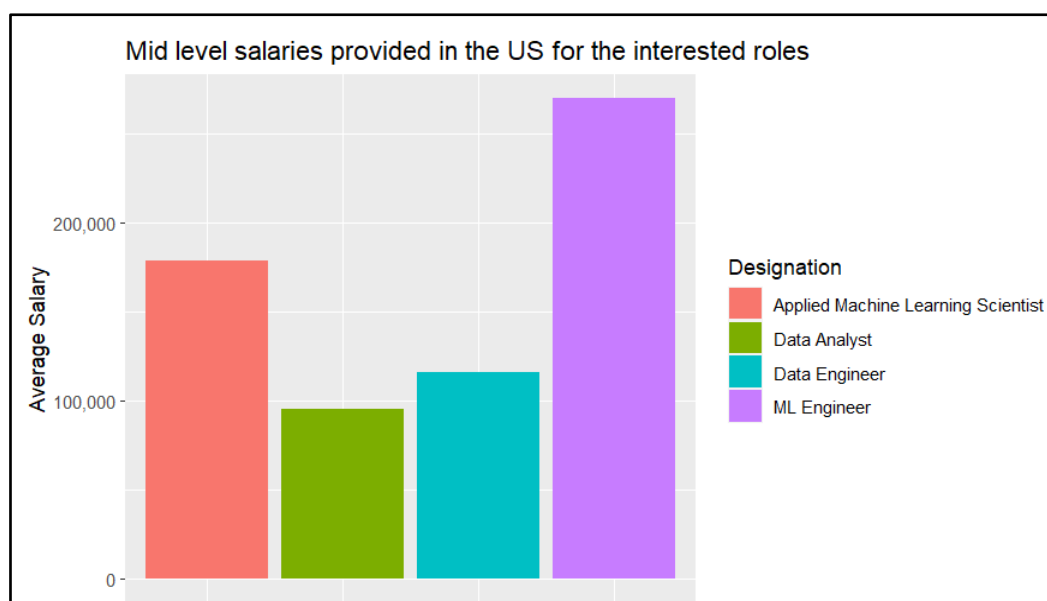


- Based on the data available above, we can see that the average income for the position of "ML Engineer" is over \$200,000 USD, whereas the last position on the list, "Data Analyst," only earns about \$100,000 USD. And the salary for the role's "Analytics engineer" and "Applied Machine learning Scientist" are closer to each other. With the aforementioned visualization, we could perceive whether we have enough evidence to decide to pursue a career as a "ML Engineer," but at this time, our evidence is insufficient because the aforementioned visualization only considers the average salary without taking into account whether the role is at the entry/mid/ senior-levels.

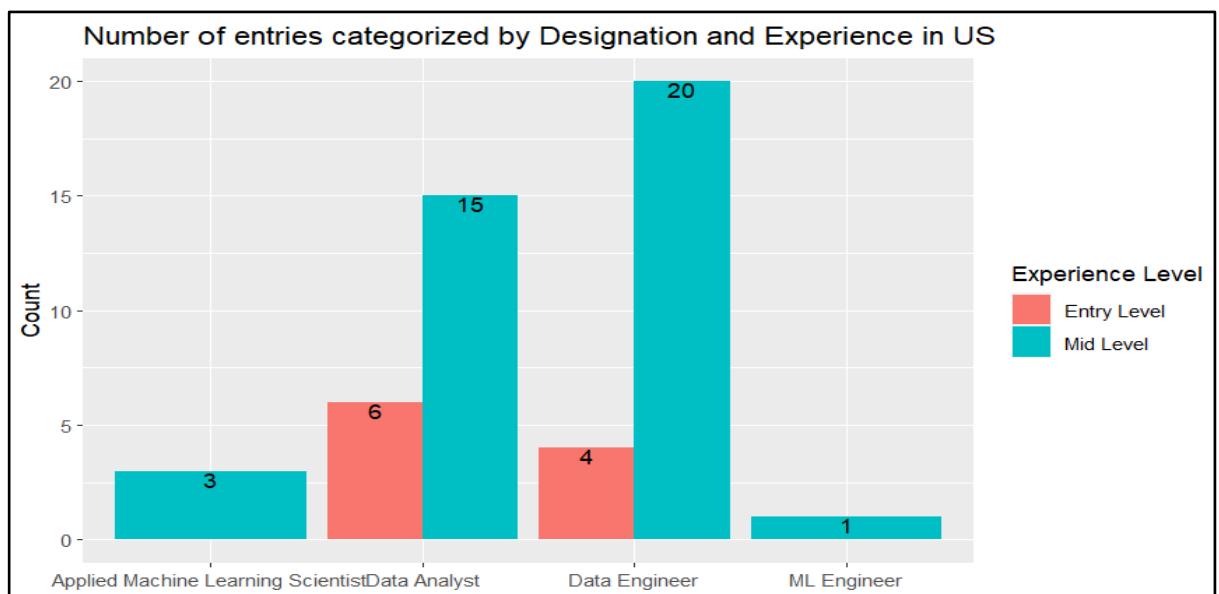
- And so, we visualize our data by categorizing it into each designation at entry level to get a clearer picture. The following visualization represents the aforementioned.



- Take note that the positions of "Analytics engineer," "ML Engineer," and "Applied Machine Learning Scientist" are absent from the aforementioned visualization. They are absent since our dataset lacks information on entry-level earnings for such roles. We next proceed to classify the pay for roles that interest us with mid-level experience. This is displayed below.



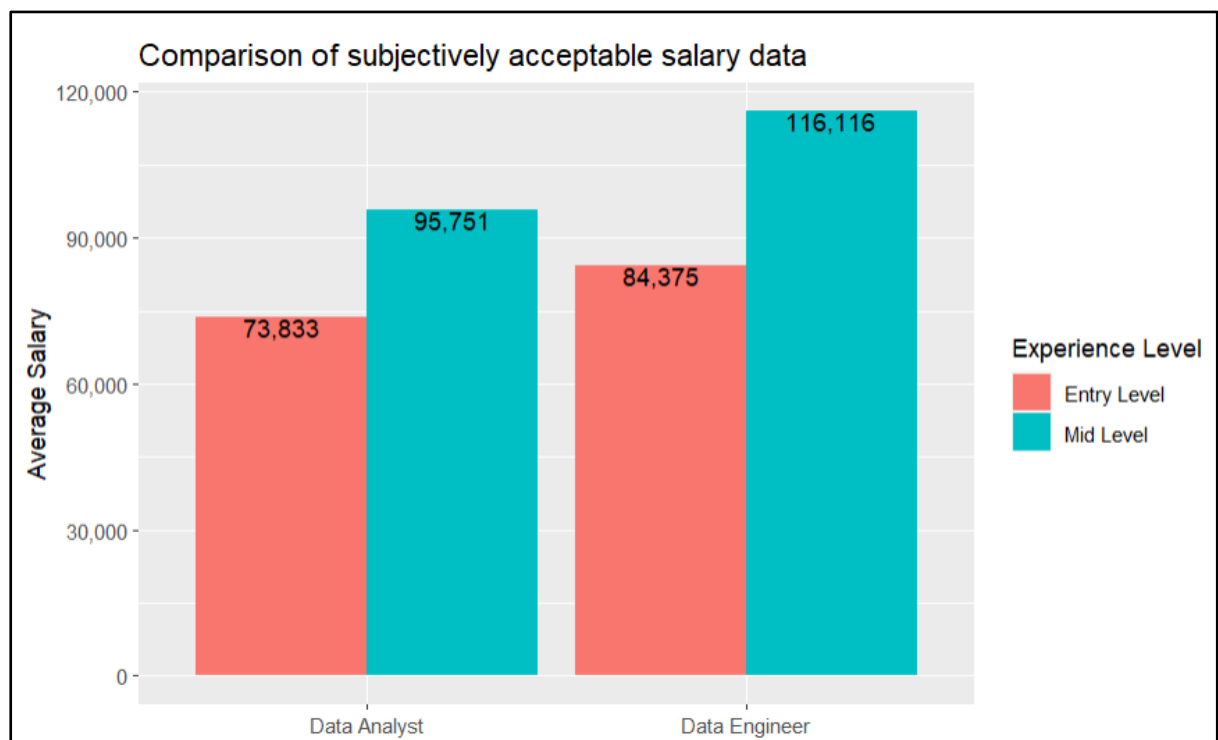
- We can see from the above diagram that the positions that were missing from our prior depiction are now there. This may mean that our dataset only includes salary information for "ML Engineers" and "Applied Machine learning scientists" with mid-levels of experience. Additionally, we see that the position of "Analytics engineer" is still missing. This occurred as a result of the position only being available at the senior level in the US, which is outside the scope of our analysis.
- In order to gather more evidence, we once again classified the data into categories based on the number of entries that each of the selected positions had in our dataset, broken down by experience level. It is illustrated in the figure below.



- The table below shows the count of data entries at the entry-level and mid-level for the chosen positions as a consequence of the aforementioned analysis.

Position	Entry level entries	Mid-level entries
Applied Machine Learning Scientist	0	3
Data Analyst	6	15
Data Engineer	4	20
ML Engineer	0	1

- Two things may be inferred from the above:
 - a) We only have salary data for "ML engineer" and "Applied Machine learning scientist" positions at the mid-level.
 - b) Since the number of entries is quite low when the focus is on the USA, we won't be able to use the data at our disposal to demonstrate that we have enough reasons for choosing to pursue that position.
- Given the remaining information, we could still run one further visualization to determine the salary disparities between entry-level and mid-level positions for the remaining two designations. The following illustrates it.



- Based on the aforementioned visualization, we can deduce that a mid-level data engineer's average compensation grows by over 27% from that of an entry-level data engineer while a mid-level data analyst's average salary only rises by about 22% from that of an entry-level data analyst.

Conclusion

And as a result of this assignment, we were able to understand the significance of having enough data when performing data analysis. We enhanced our abilities in data exploration because we had the chance to create visuals. Although our initial goal was to gather evidence to help us decide which designation to pursue, our visualizations showed that the data we had available was insufficient to draw any conclusive recommendations. However, in the end, we did find that the average US compensation for a data engineer is higher than the average US compensation for a data analyst. And it is not a bad idea nor unheard of to begin as a Data Engineer and later change roles to become a Machine Learning Engineer as we gain practical experience.

References

Bhatia, R. (2022). *Data Science Job Salaries* [Data set].

<https://www.kaggle.com/datasets/ruchi798/data-science-job-salaries>

Is data science A good career? (n.d.). Mentorcruise.Com. Retrieved June 28, 2022, from

<https://mentorcruise.com/blog/is-data-science-a-good-career-1ab9d/>

McKay, D. R. (2004, November 28). *How to explore your career options*. The Balance

Careers. <https://www.thebalancecareers.com/career-exploration-525632>