DATA ANALYSIS PORTFOLIO

AUTHOR: ABIKOYE JESUTOMISIN E.

TABLE OF CONTENTS

UDEMY COURSES PROJECT DESCRIPTION

In a bid to determine the potential moves that have the tendency of increasing the revenue of Udemy, the report gives surgical visualizations of various interactions of features on the dataset. This report is to examine the data on courses from different topics and subjects from Udemy.

It is believed that Web development courses are more popular in the range of courses offered by Udemy, however, this report will explain the performances of courses and potential steps to increase revenue. Four different datasets containing information on different subjects will be combined to carry out the analysis.

Based on the analysis carried out, web development courses tend to be the most prevalent with a high number of subscribers. More visualized information on the findings will be seen in the later part of the report.

DESIGN

CLEANING THE DATA:

First, duplicated values were highlighted and subsequently removed

Next, blank cells were removed with the filter option.

Lastly, the headers were made clear and consistent. This involved changing all the headers to lower case.

TOOLS USED IN THE ANALYSIS:

Google sheet: Pivot tables on google Sheets were used to create various visualizations that include: Total number of subscribers, the average number of subscribers, the average cost of subjects at each level, etc.

Tableau: This was used to make interactive visualizations that show the relationship between each subject and other features in the dataset.

FINDINGS

Figure 1: Total Number of subscribers for each subject

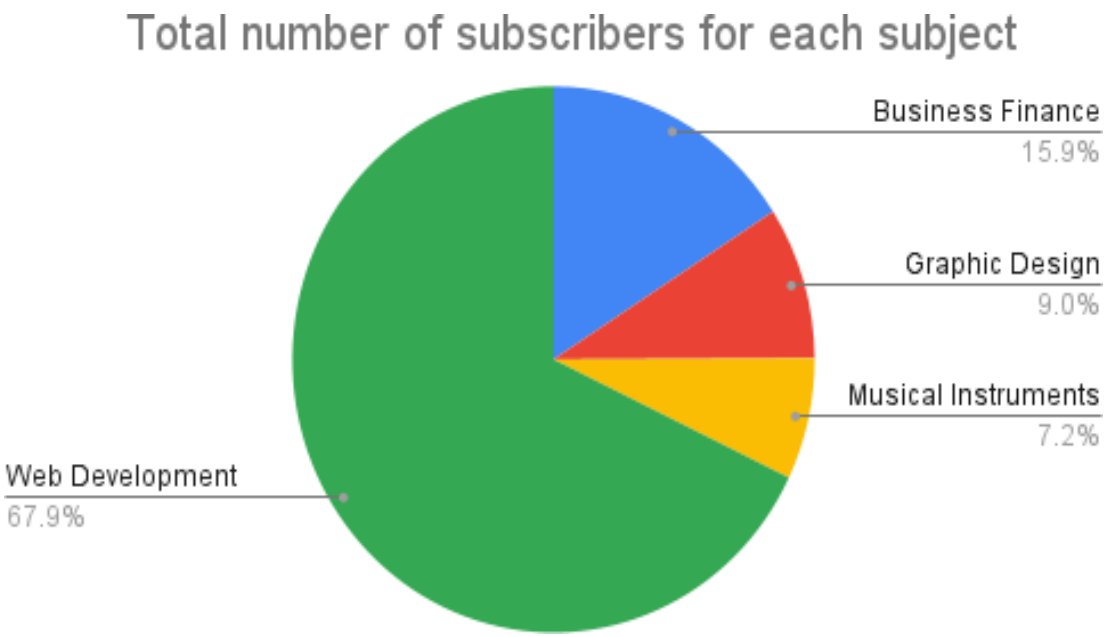


Figure 2: Average number of subscribers for each subject

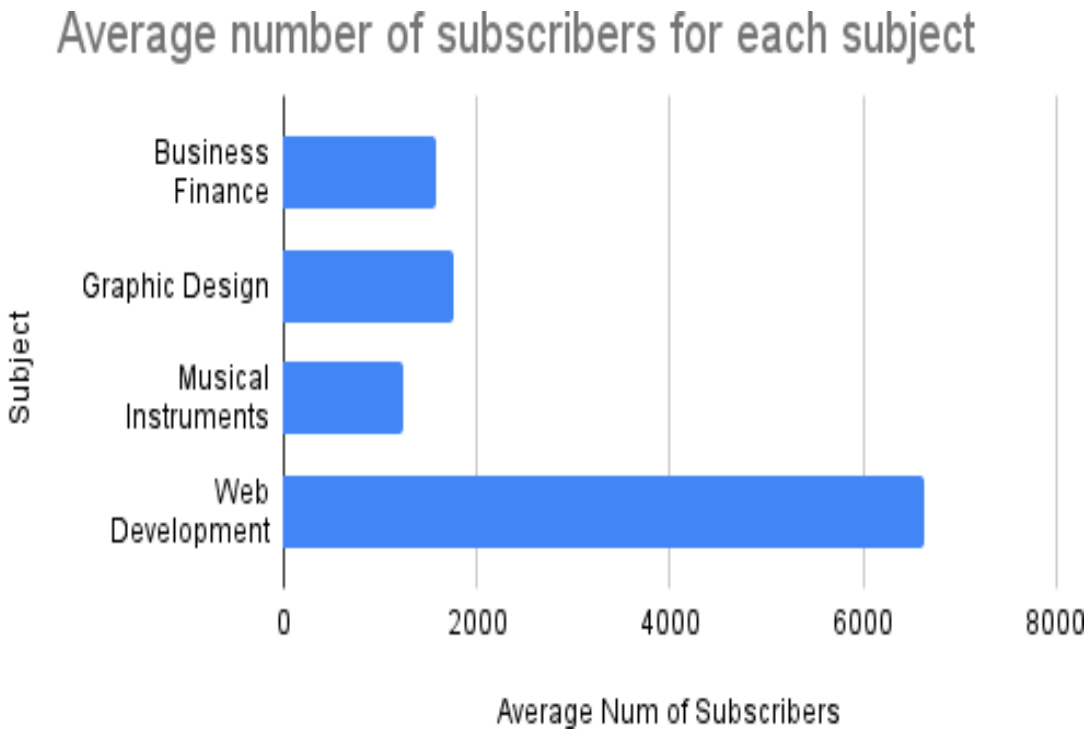


Figure 3: Average cost per subjects at each level

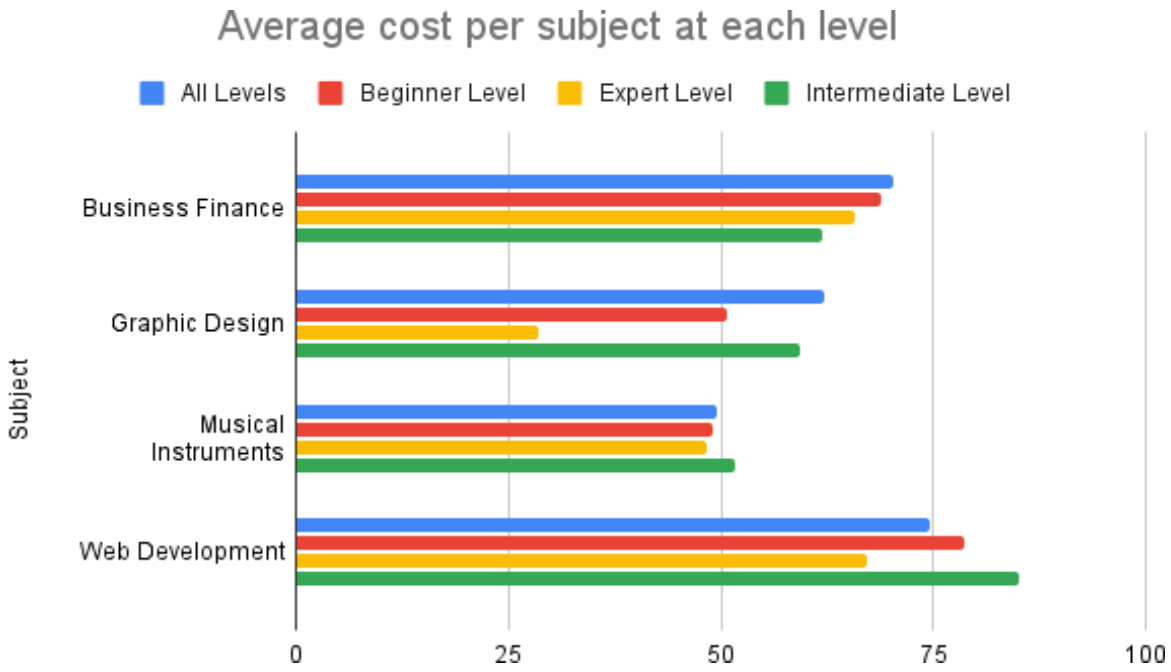


Figure 4: Average content duration for each subject

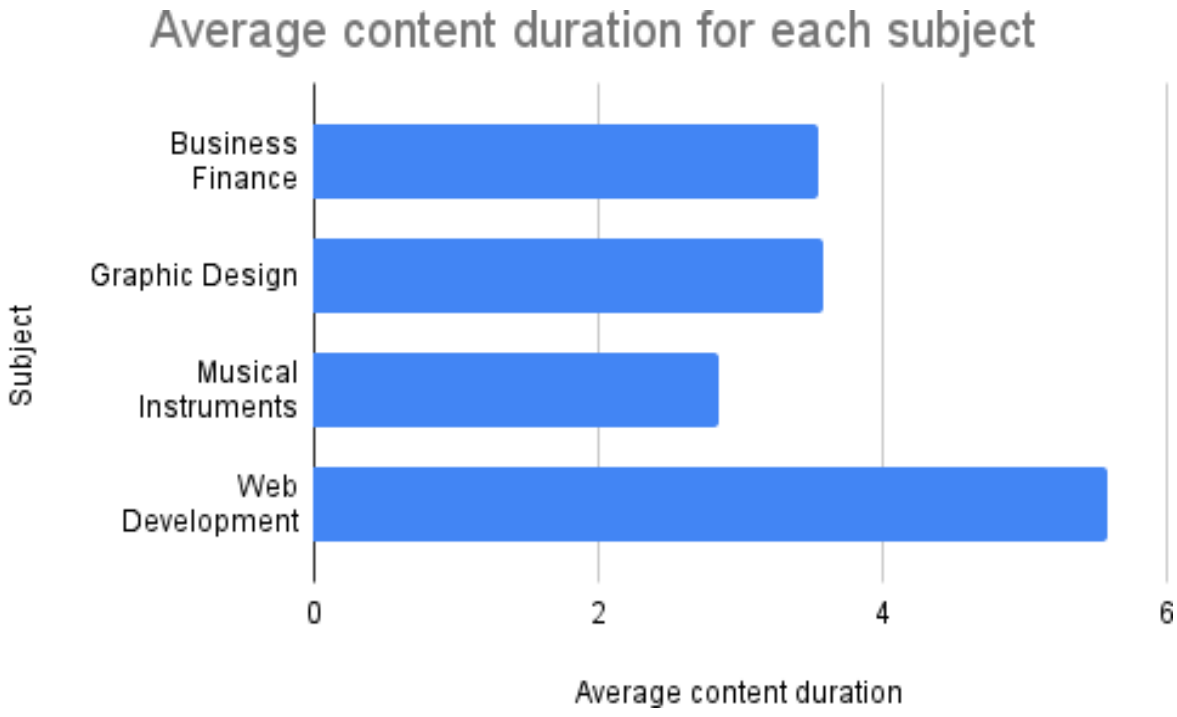


Figure 5: Average rating per subject for each level

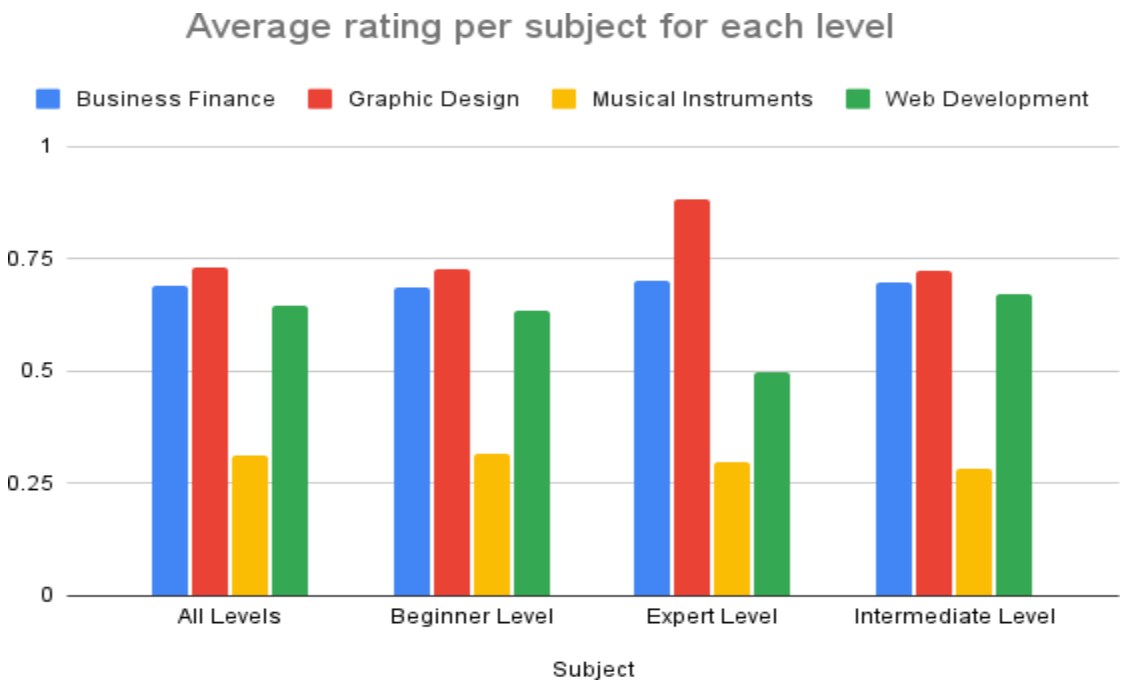


Figure 6: Total number of free and paid courses for each subject

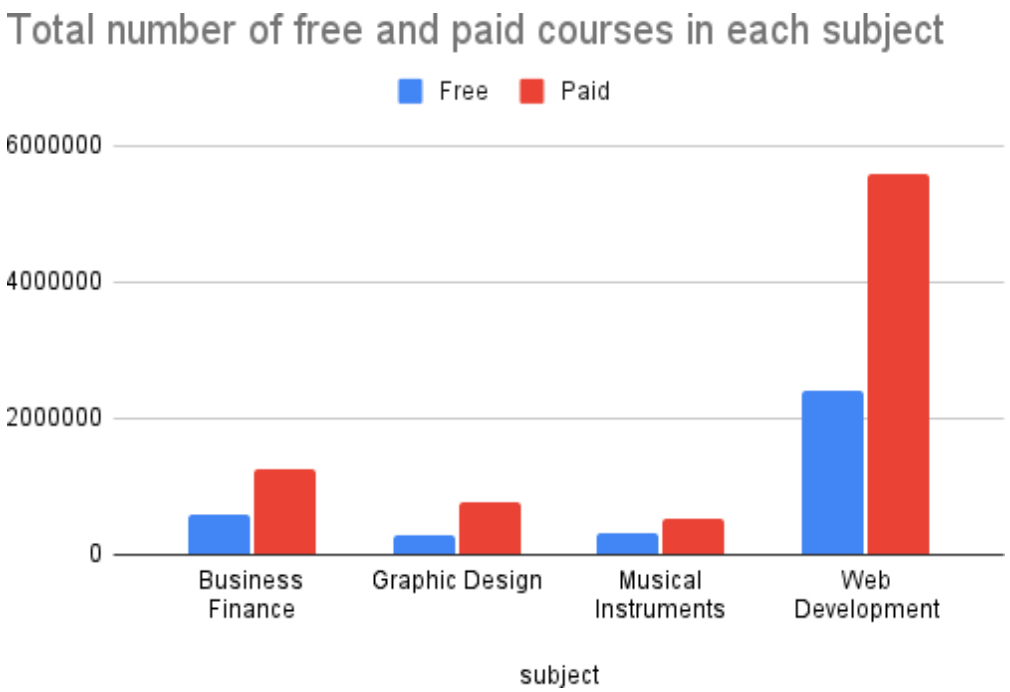
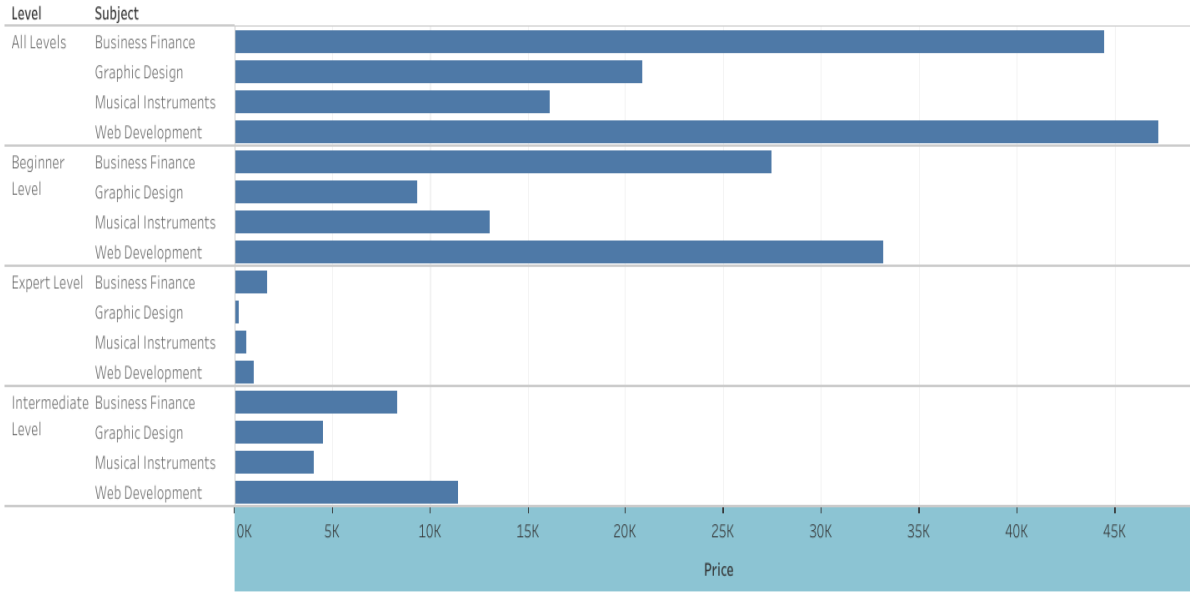


Figure 7: Total price of subjects per level



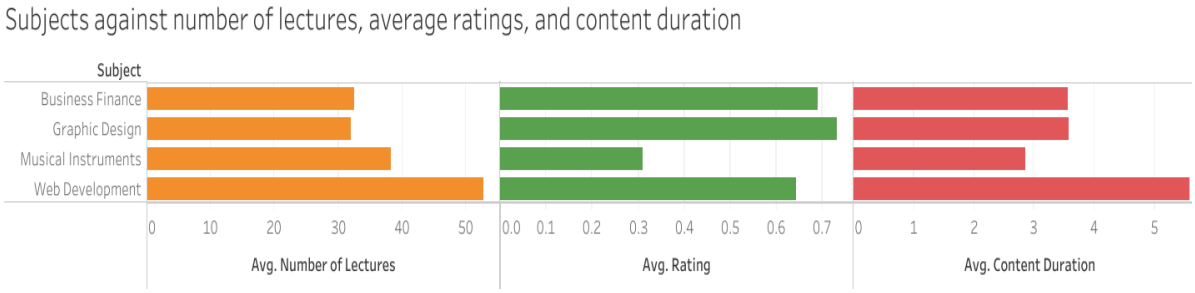


Figure 8: Average number of lectures, rating, and content duration for all subjects

ANALYSIS

From the findings above, it is revealed that Web development has the highest number of subscribers. Although Finance is the subject with the second highest total number of subscribers, however, Graphics design, displaced it as the second most subscribed on average. Additionally, Web development, in general, is the most expensive of the subjects followed by Business Finance.

To understand the root cause of the predominance of Web development, it is observed that it has the highest content duration. To investigate why it has the highest content duration, it was discovered that it has the highest number of paid and free courses. To understand this, it was discovered that the large availability of free Web development can be attributed to the high demand for people's hunger to learn about it. The reason for this can be because of some evolving demand for the application of Technology in all sectors for better problem-solving.

NOTE: A point to note is that at the expert level, Business Finance has the highest total prices of courses paid for, displacing Web development to the second spot. An explanation for this is that most people are hungry for knowledge on emerging technology at the beginner and intermediate stage, while self-learning on jobs to augment skillsets. However, Business Finance requires a steady learning process even till the expert level. Also, Graphics design has the highest rating followed by Business Finance. This can be attributed to the large course contents available for Web development, hence, users are not satisfied on average with the courses taken. This can be a result of high curiosity before taking the course on Web development or the thought that they should start building websites right from the beginner stage which is not realistic. Consequently, they gave courses on web development a considerably low rating. On contrary, after taking even beginner courses on Graphics design, the users can get enough knowledge for little freelancing jobs. Hence, the reasons for high ratings.

CONLCUSION AND RECOMMENDATION

This well-detailed report shows that increasing the prices of courses on Web development, based on its high demand, could be a good way to augment revenue. Due to the escalated rate at which technology is evolving, even in a few years, the demand for Web development courses will still be on the increase. It implies that when the price of courses on Web development rises marginally, there will still be an observed rise in the number of subscribers.

This report recommends that to increase revenue, the price of web development courses should be increased.

DATA SCIENCE JOBS PROJECT DESCRIPTION

Due to the growing significance of Tech jobs in the modern world, especially in the era where data generation has increased immensely, young talents are a bit uncertain about what career path to choose to be instrumental in solving global problems. This report addresses the curiosity of Tech enthusiasts who are interested in starting a career in Data science.

The report draws insights from a dataset on Kaggle collected from 2020 to 2022 on various job niches in the data technology space. While many desire to become full-stack data scientists, the report shows which job specialties have been in exponential demand and highlights their future potential

DESIGN

CLEANING THE DATA:

The dataset has no missing data points. However, there were some abbreviations used that were changed for better communication. For example, in the employment type; FT was changed to Full-time. Similarly, PT, CT, and FL were changed to Part-time, Contract, and Freelance respectively. Similar cleaning was done in the experience level column as well.

TOOLS USED:

Google sheet was used in visualizing various insights not limited to increasing the number of data scientists across the year, salary earned across the year based on experience, Number of data scientists based on job title and experience, etc.

Tableau was used to show how the salaries of data scientist change across the year. Also, how it changes with company size, experience level, job title, and employment type.

FINDINGS

Figure 9: Number of data science across years

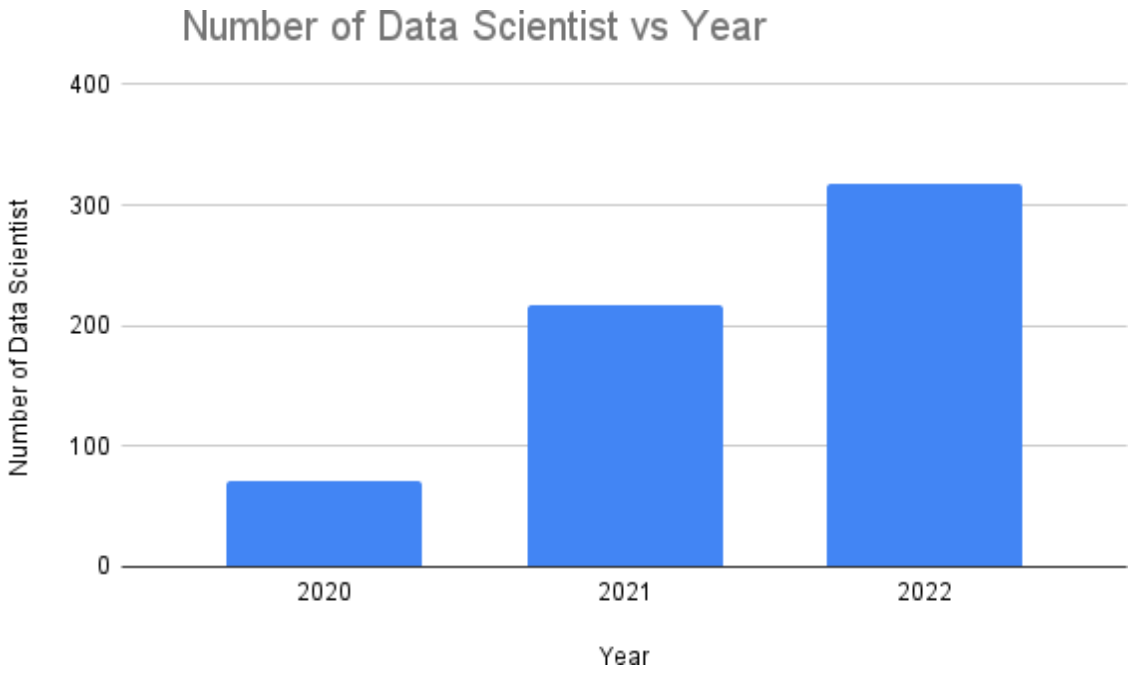


Figure 10: Total number of company size

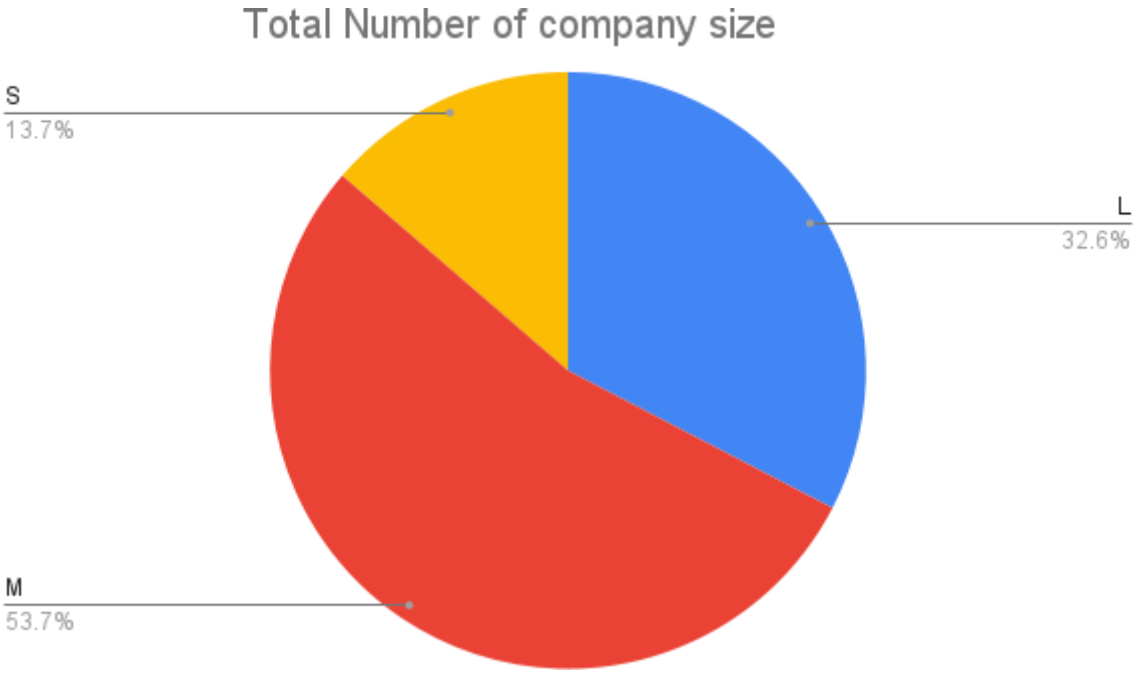


Figure 11: Average salary of each job titles in USD

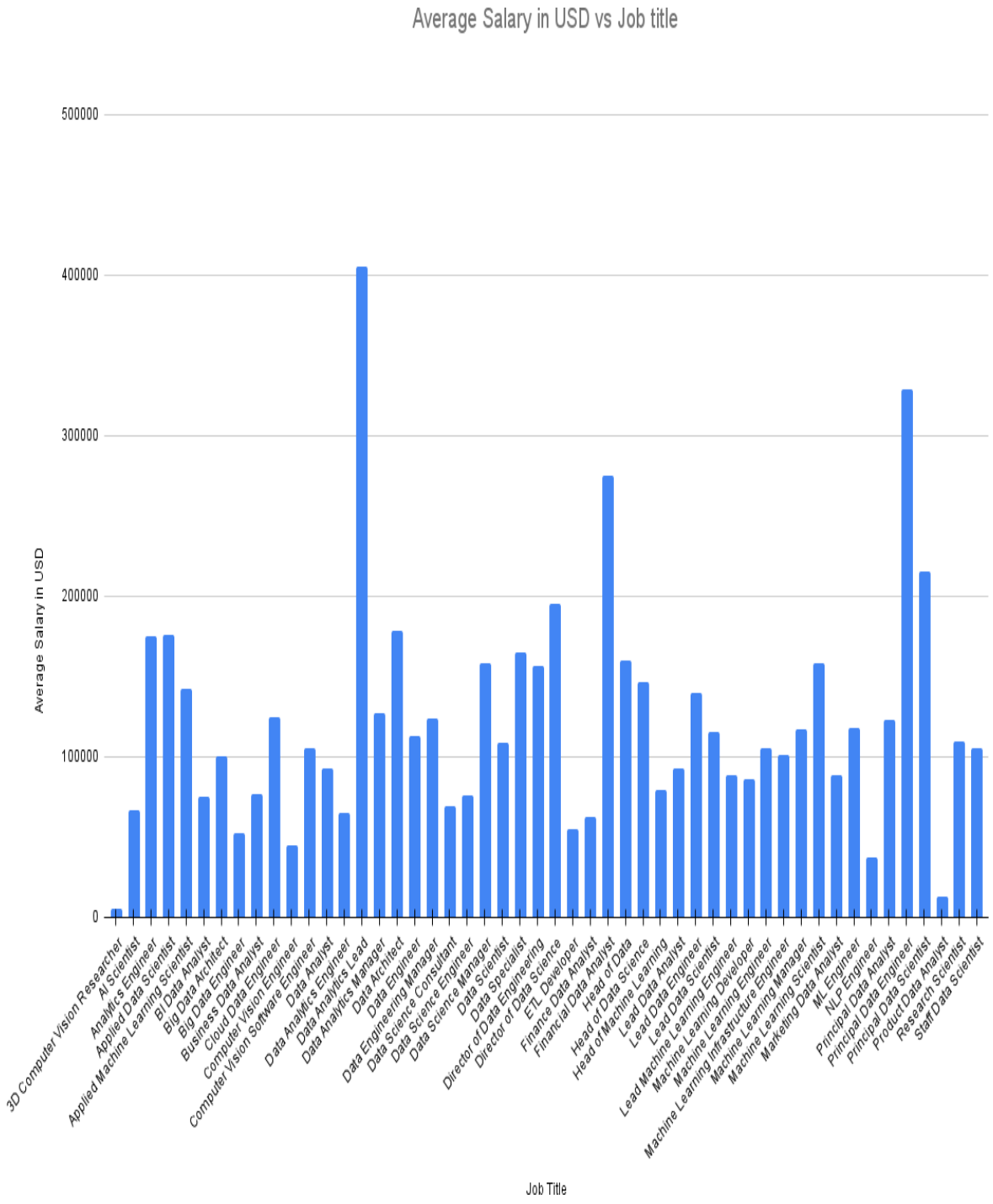


Figure 12: Salary earned by employees based on company location

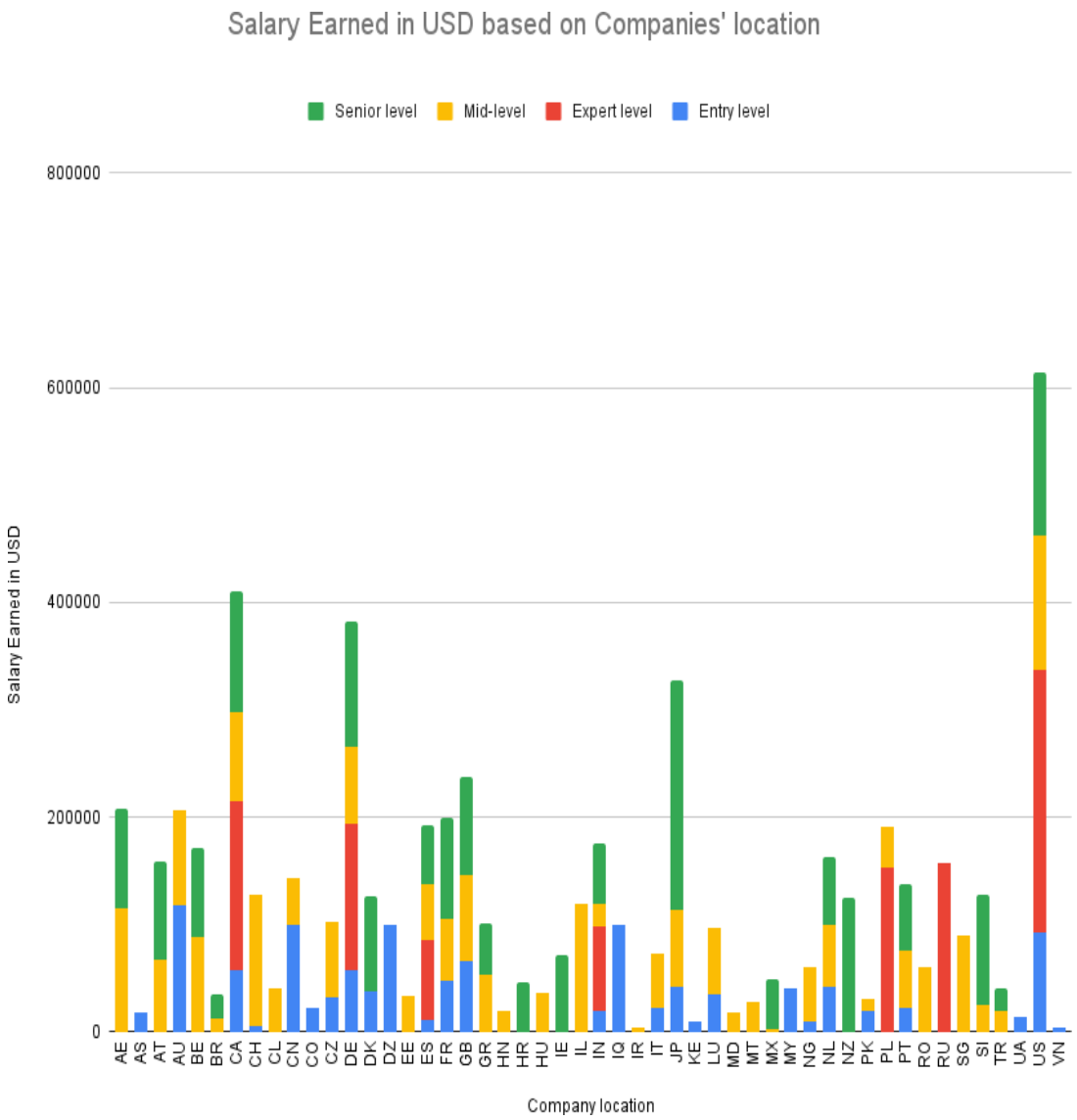


Figure 13: Number of employees for each job titles per experience level

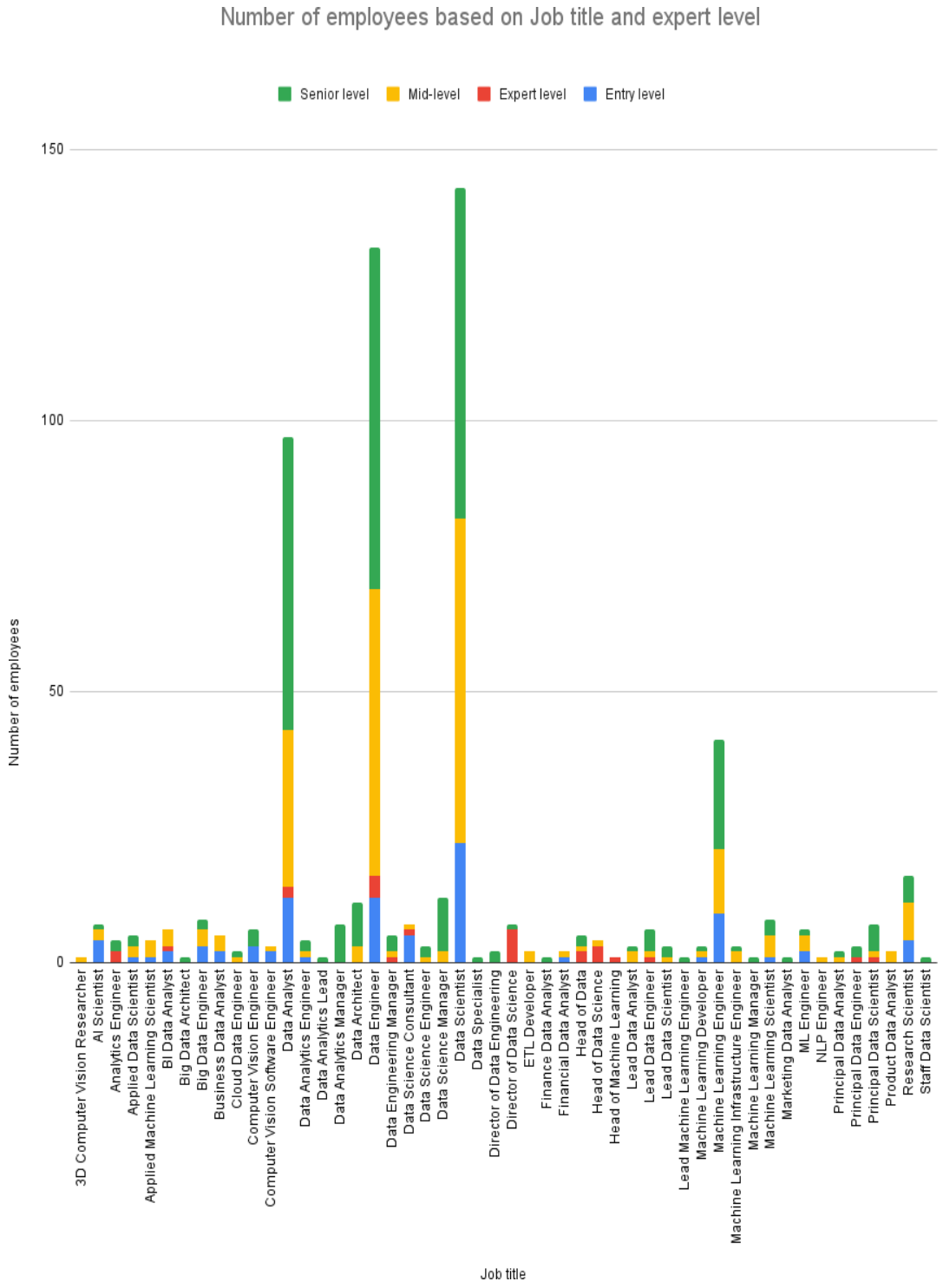


Figure 14: Number of employees working remotely for each job title

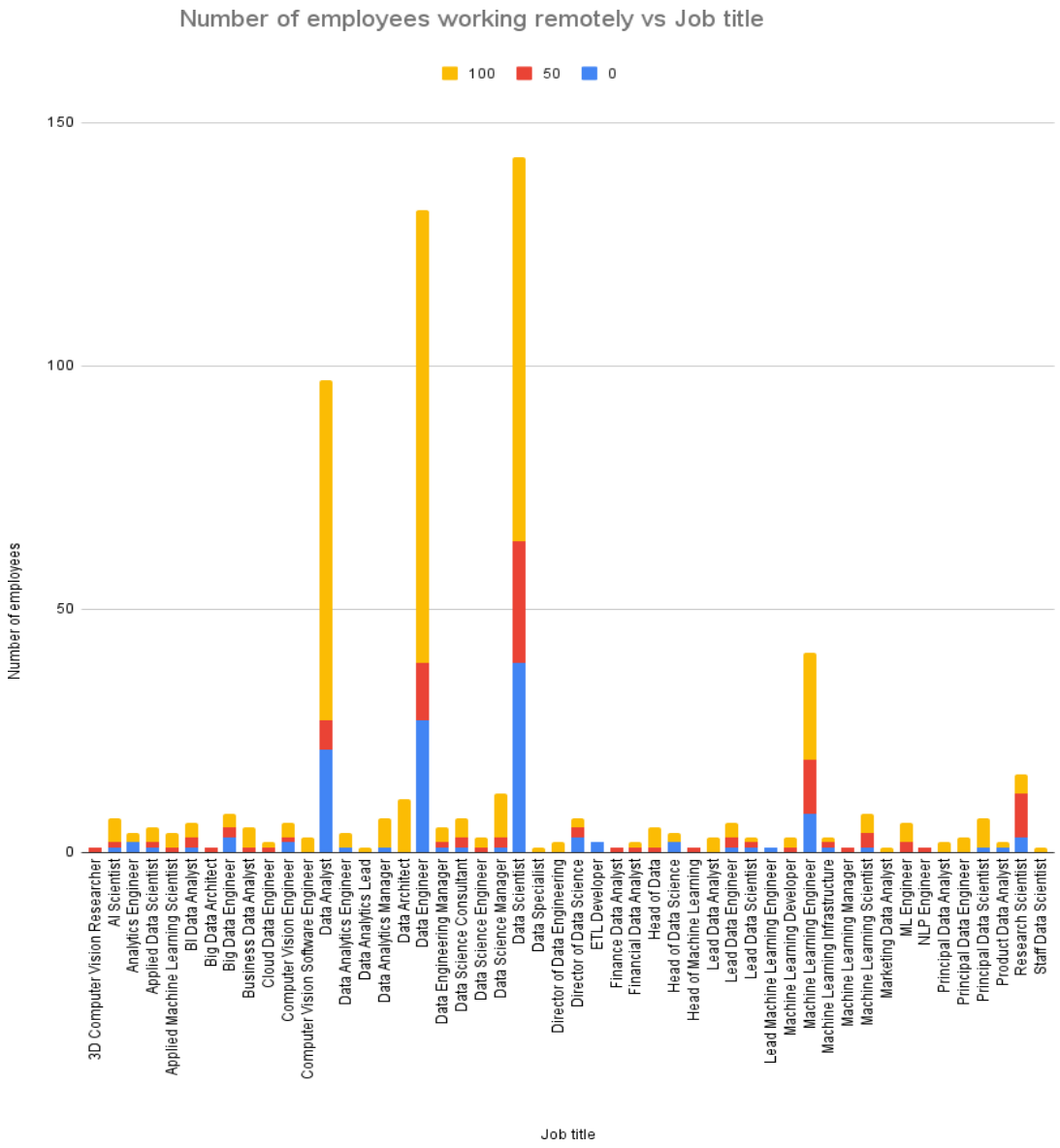
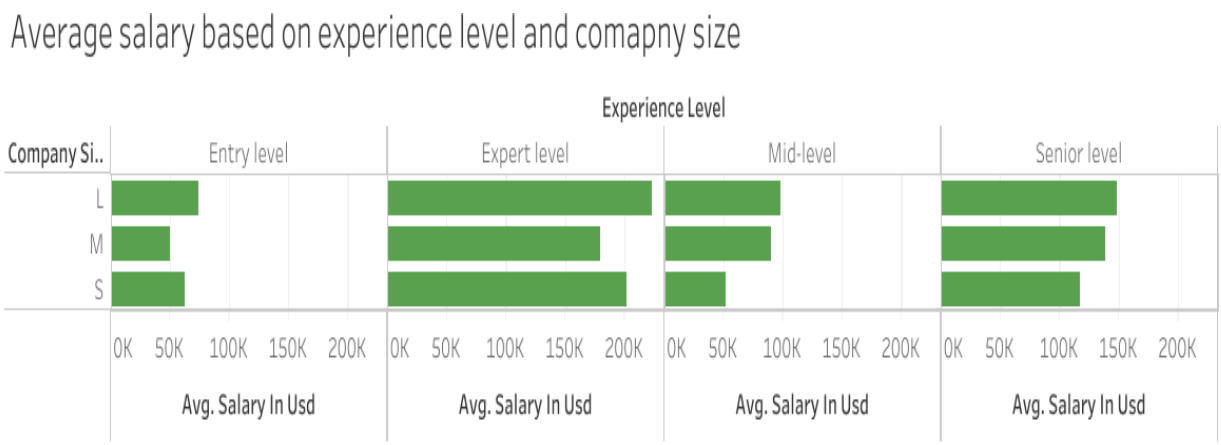


Figure 15: Average salary of employees for each experience level based on company size



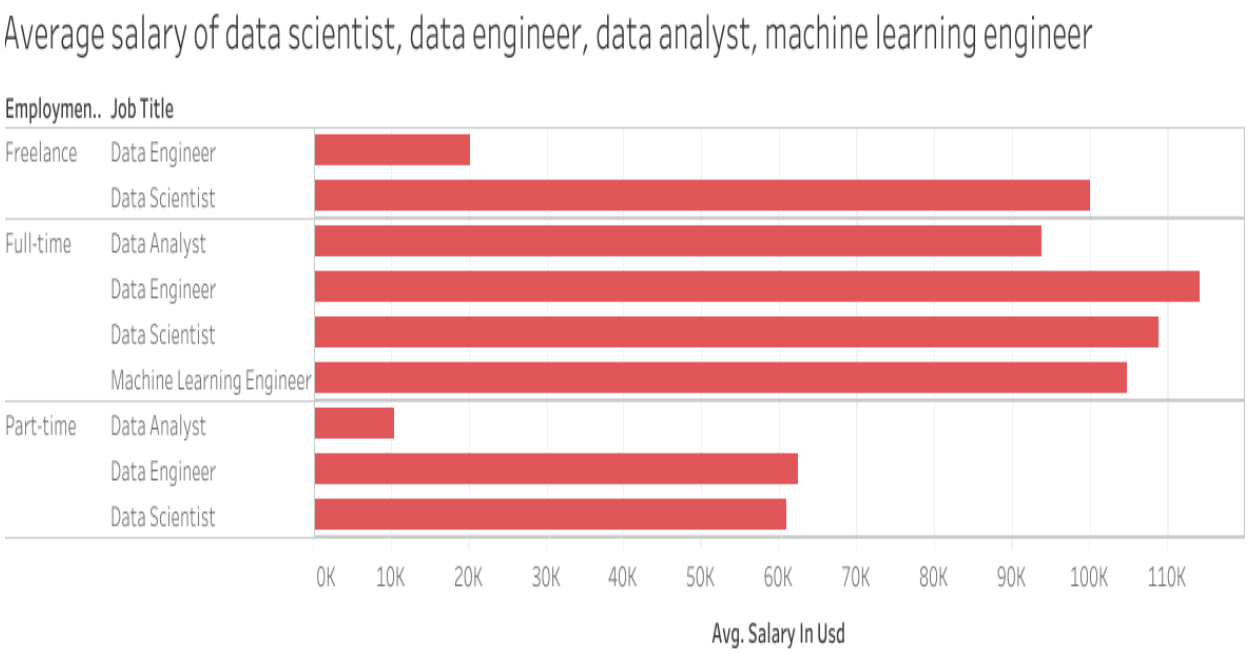


Figure 16: Average salary of the four in-demand job titles per employment type

ANALYSIS

Based on the findings above, data scientists are the most represented, followed by data engineers, data analysts, and machine learning engineers. However, the most paid job specialty is Data Analytics lead, followed by Principal Data Engineer, Finance Data Analyst, Principal Data Scientist, and Director of Data Science.

Looking at the root cause of why data scientists, data engineers, data analysts, and ML engineers are the most sought after, it was seen that they provide the largest opportunity for remote work. This could be attributed to the fact that the demand for the services is in high demand especially at the entry level, with research scientists completing the top 5 job titles at the entry-level in high demand. The reason for this is due to the zillions of data that are now generated every day. Hence, companies are now ready to pay for their services. Why are companies ready to pay for their services? This can be traced to the significant roles they play in minimizing risks and increasing an organization's revenue.

It was also discovered that the United States (US), followed by Canada (CA) pays the most salary on average in USD for the professionals in the data technology space.

CONCLUSION AND RECOMMENDATION

This report revealed that there is a growing demand for professionals in the data technology space. Interestingly, the progressive increase will continue for some years with the central focus on full-stack data scientists, followed by data engineers, data analysts, and machine learning engineers.

This report recommends that students and recent graduates should learn one of the four most rated job titles which are: data scientist, data engineer, data analyst, or machine learning engineer. This should be learnt in line with their course of study to be instrumental in the technologically advancing world where data have proven pivotal in pivoting into the future while being in the present to minimize risk, maximize customers' satisfaction, and exponentially increase revenue.