REPORT ON THE ANALYSIS OF THE DATA SCIENCE JOB DATASET

Introduction

This dataset contains job listings for data science positions, providing a comprehensive view of the data science job market from year 2000 to 2002. This will help users analyze trends in data science employment.

DATA SOURCES

The primary data source for this analysis existed as ' data_science_job.csv' file. This is a 5000-row dataset that includes essential fields such as job title, company location, employee residence, work setting, salary in USD, required skills etc.

TOOLS

1. Python – Data Cleaning and EDA
2. Power BI – Data visualization and Creating of report

EXPLORATORY DATA ANALYSIS (EDA)

EDA involved exploring the 'data science job' dataset to answer key questions, such as

1. What is the most sort-after Data science job?
2. What is the average salary for the various job titles in the data science roles?
3. Which country offers most employment for the data science job?
4. What are the average salaries for the various employment types in data science?
5. Which type of work setting is most appropriate for the various data science jobs?
6. What are the demands for various skills in data science roles?

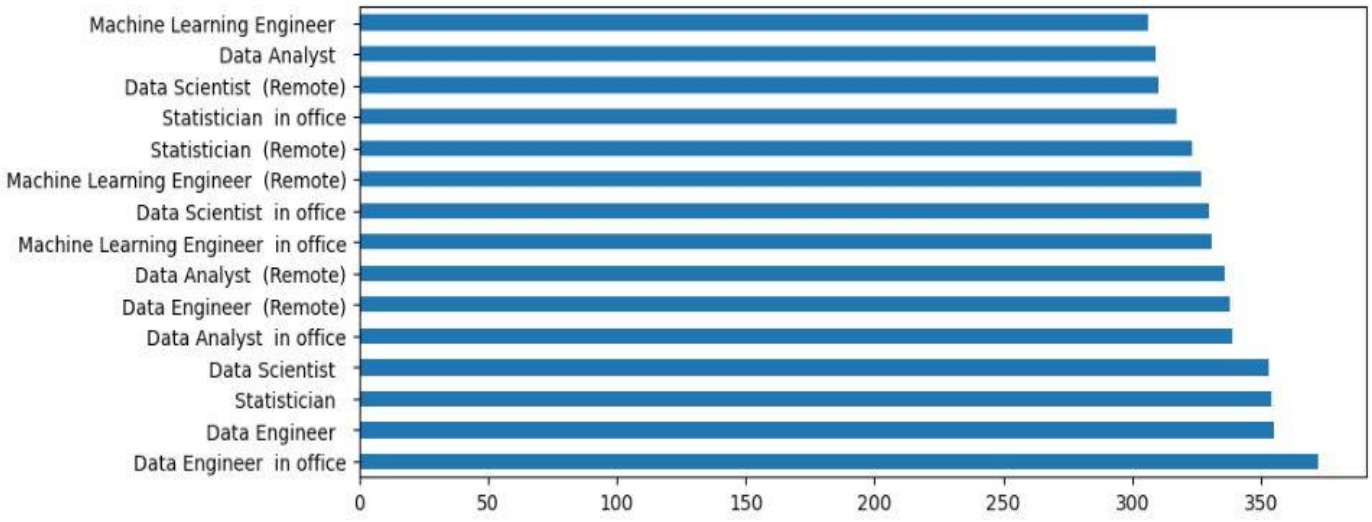
RESULTS / FINDINGS

After the analysis of the 'data science job' dataset, the following are my findings

1. There are a number of job titles in the data science job and the count of each appearances in the dataset are as follows

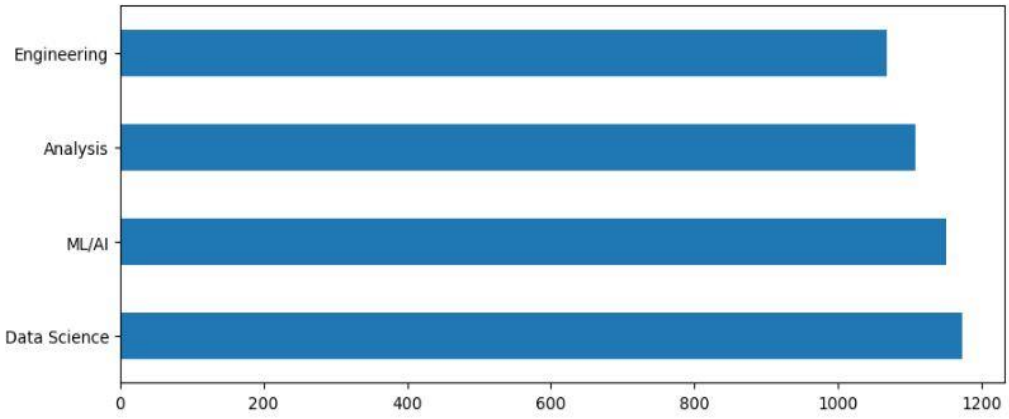
Job title	Count
Data Engineer in office	372
Data Engineer	355
Statistician	354
Data Scientist	353
Data Analyst in office	339

Data Engineer (Remote)	338
Data Analyst (Remote)	336
Machine Learning Engineer in office	331
Data Scientist in office	330
Machine Learning Engineer (Remote)	327
Statistician (Remote)	323
Statistician in office	317
Data Scientist (Remote)	310
Data Analyst	309
Machine Learning Engineer	306



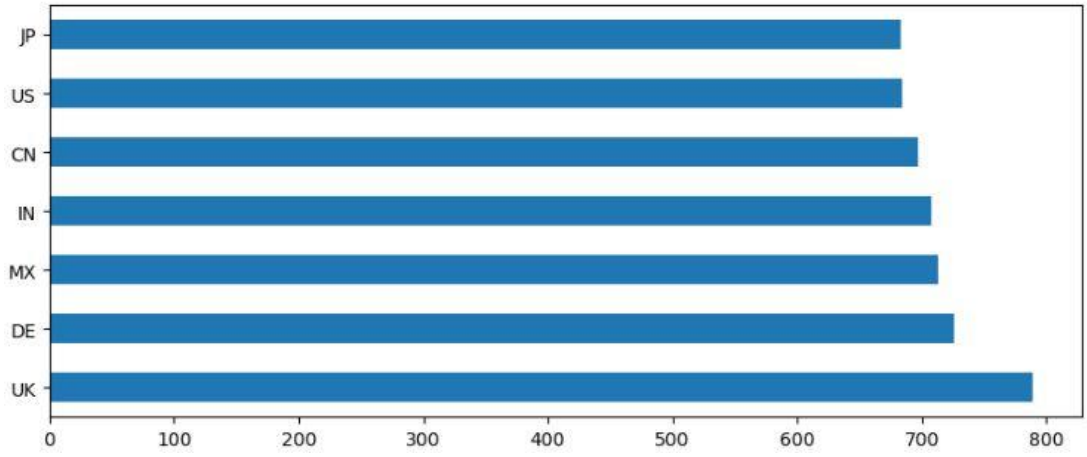
2. The number of various job categories in the data science job are below:

Job category	**Count**
Data Science	1173
ML/AI	1151
Analysis	1108
Engineering	1068



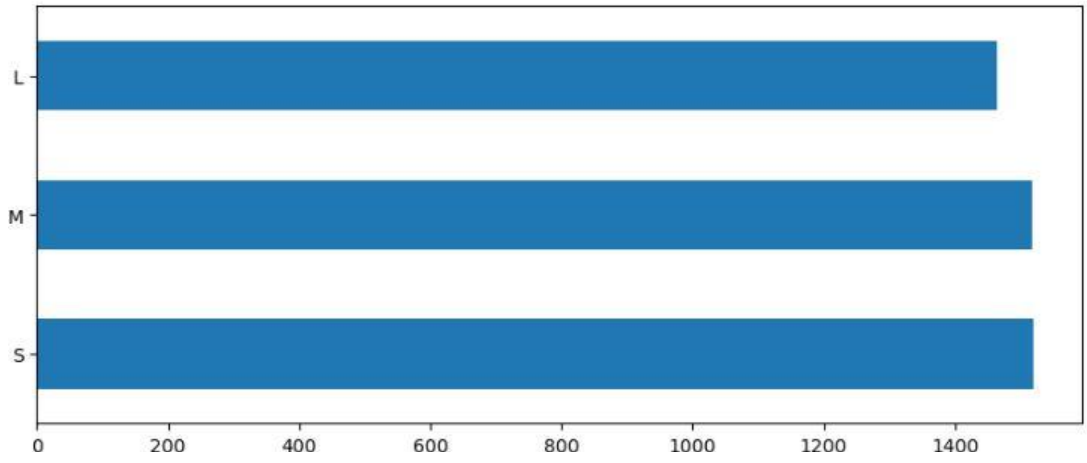
3. The country that offered the most data science jobs is UK, followed by Germany, Mexico etc. The country that offers the least data science jobs is Japan. The data for this analysis is below

Company location	Number of jobs
UK (Britain)	789
DE (Germany)	726
MX (Mexico)	713
IN (India)	708
CN (China)	697
US (America)	684
JP (Japan)	683

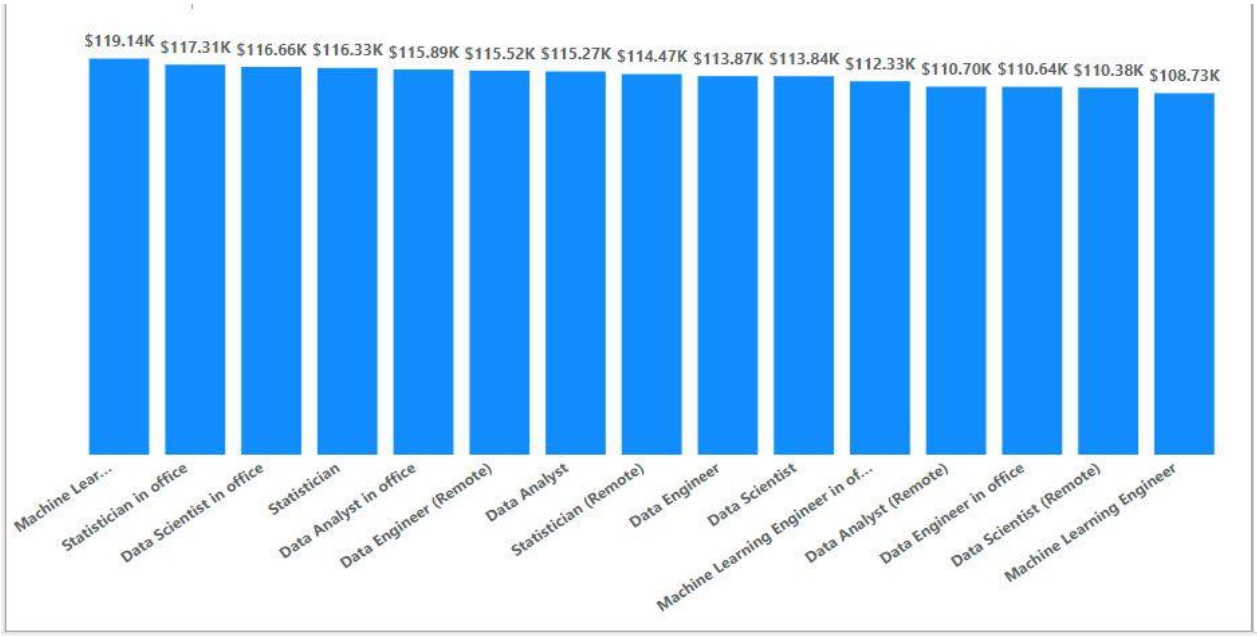


4. Small (S) company size offered the most job positions, followed by medium (M) and then large (L) company size

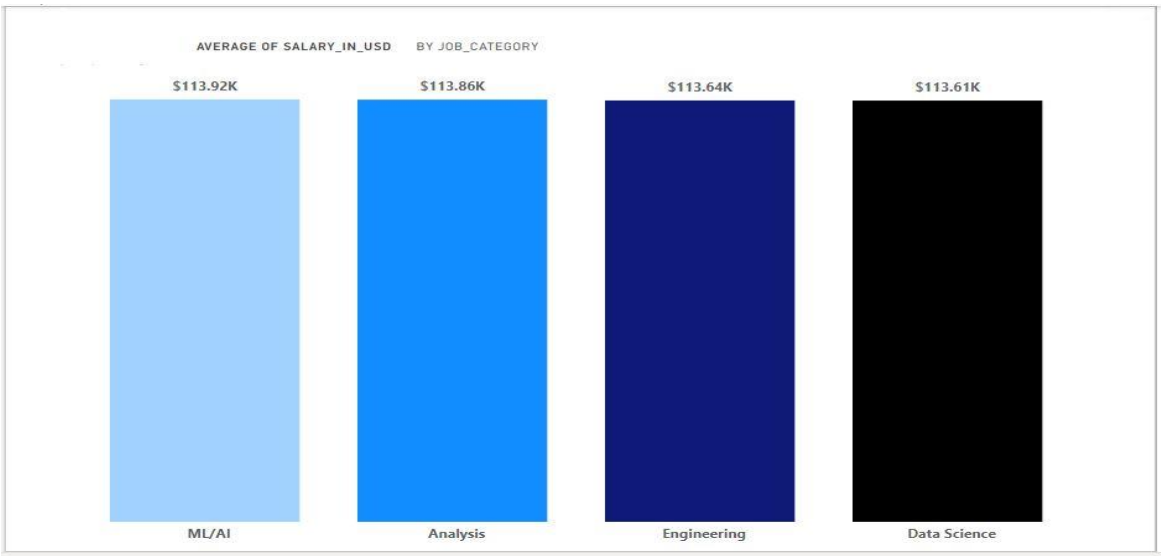
Company size	Count
S	1519
M	1517
L	1464



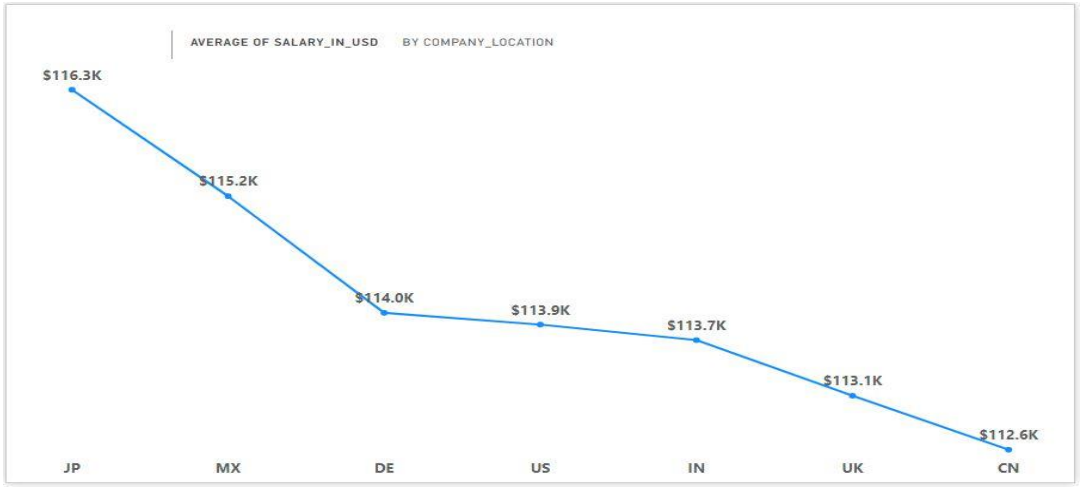
5. The analysis of average salary in USD received by the various job titles shows that 'Machine Learning Engineer Remote' receives the highest salary of **$119,143.62** followed by 'Statistician in Office' who received **$117,312.09** and the job title that received the least salary is 'Machine Learning Engineer' who also receives **$108,734.67**. The data for the average salary for the various job titles are presented below.



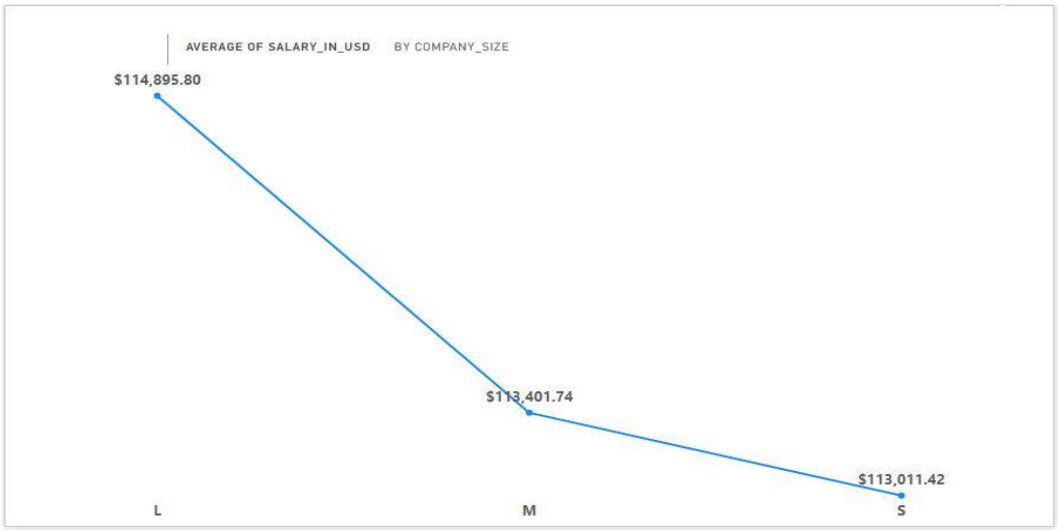
6. An analysis on the average salary in USD by the job categories also revealed that the various job categories were paid differently but with very few margins between them.
   ML/AI received the highest salary of $ 113,961.10, closely followed by the Analysis category also receiving $ 113,859.61, then the Engineering job category also receiving $113,639.09. The least salary was received by the Data Science category receiving $113,607.68



7. Analysis of company's location and the salary in USD revealed that data science jobs located in certain countries pay more than in the others. The analysis shows that data science jobs in companies located in Japan(JP) pay more with an average salary of **$116,266.27**, followed by Mexico(MX), with an average salary of **$115,175.97**; the least average salary was in companies located in China(CN) with a salary of **$112,576.92**.

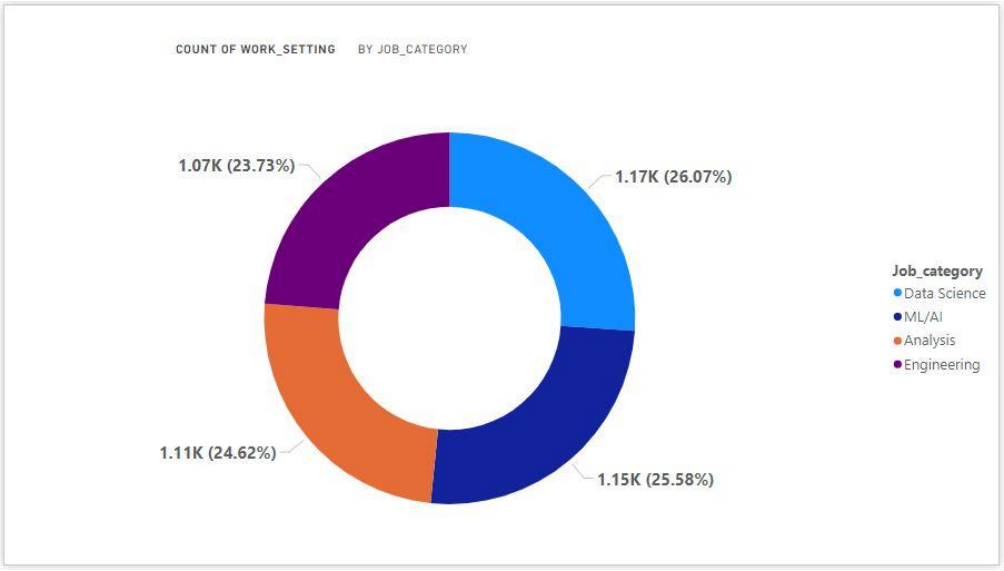


8. The analysis on the size of company and average salary paid by them showed that the Large (L) companies pays more for their data science positions, followed by the Medium (M) and then the Small(S) companies. This kind of revelation is probably due to the differences in allocated resources or the workload associated with the positions in the various company sizes.

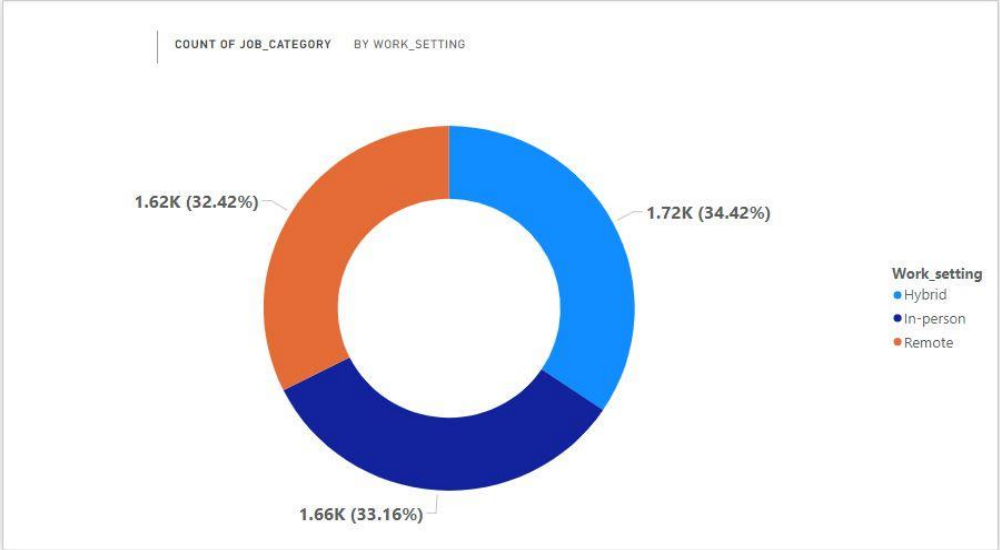


9. The analysis on the employment type and the average salary in USD was done to also know the relationship that existed between them. The data showed that employment type, 'PT' received the highest salary of $115,407.46 and then employment type 'FT' received the least salary of $111,693.79.

10. The analysis to determine the most sort-after data science job revealed that the need for the various data science jobs differ from each other but with a very narrow margin. Data science job is the most sort-after job with a count of 1173 representing 26.07%, closely followed by ML/AI also with a count of 1151(25.58%), then Analysis category with counts of 1108(24.62%) and lastly, Engineering category, with counts of 1068(23.73%)



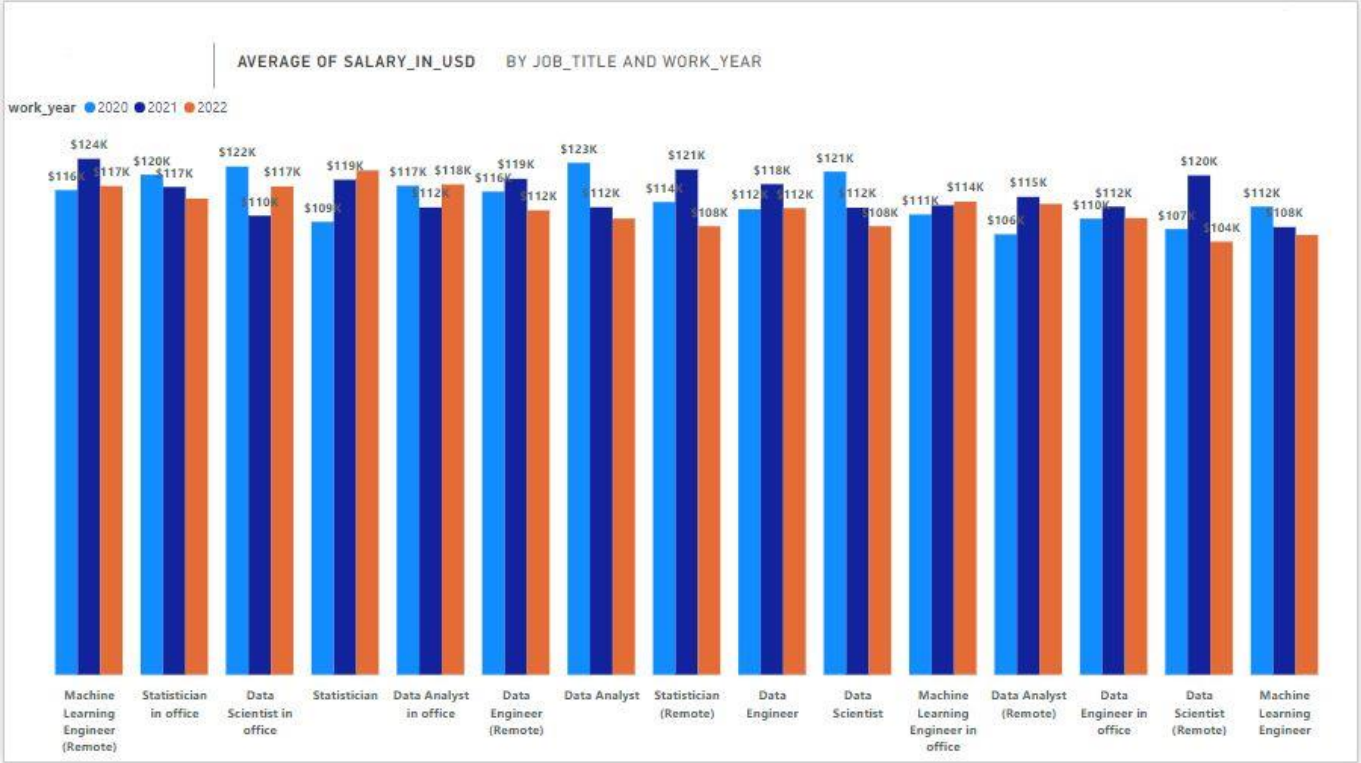
11. The work settings analysis for the various data science jobs showed that the Hybrid work setting was most observed with a count and percentage of 1721 and 34.42%, followed by In-person (1658 and 33.16%) and Remote (1621 and 32.42%)



12. In the period under study (2000 to 2002), some job titles have seen fluctuations in their average salaries, whilst others experienced increase or decrease in average salary. Machine learning engineer (Remote), Data scientist in office, Data analyst in office, Data analyst (Remote), Statistician (Remote), Data engineer, Data engineer in office and Data scientist (Remote) had experienced fluctuations in their average salaries which in my opinion were completely random.

    Statisticians in office, Data analysts, Data scientists and Machine learning engineers experienced a decline in their average salaries in the period under review. This I think may be due to the influx of these job titles in the data science job market and competing with the available job openings thereby reducing their value and salary bargain power.

    Statisticians and Machine learning engineers in office however, experienced a steady increase in their average salaries throughout the period under review. This observation, I think may be due to more job openings for statisticians and machine learning engineers and companies had to increase their average salaries either to maintain their staffs or lure other Statisticians and Machine learning engineers to fill their job openings.



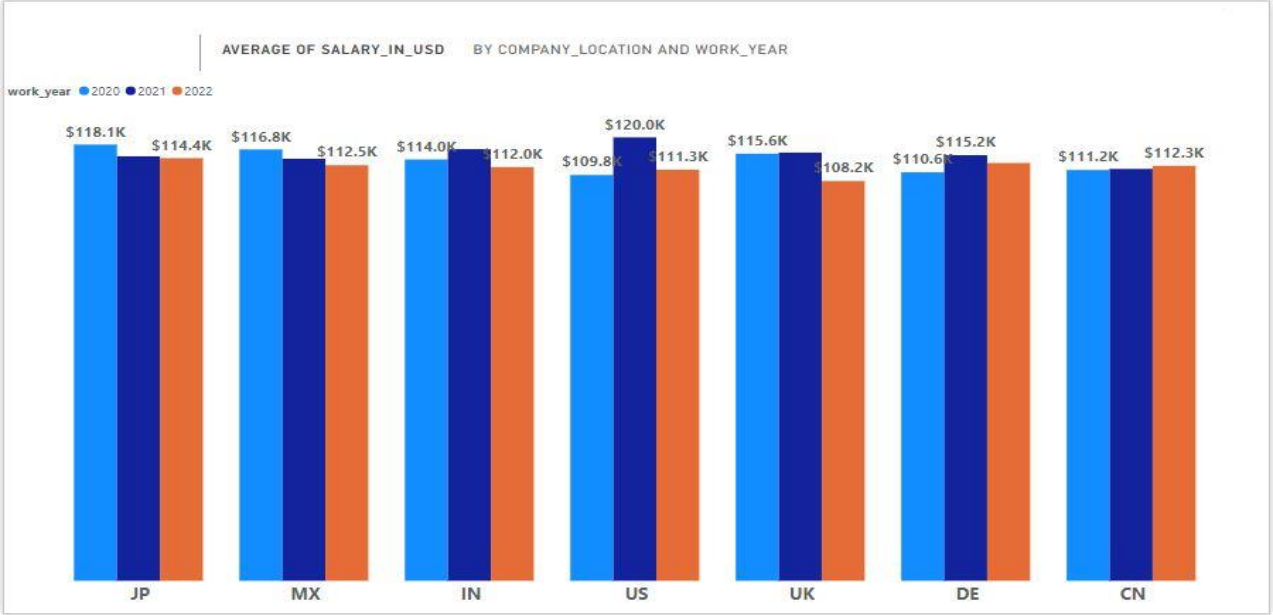
13. Analysis on the experience level of employees to their average salaries received revealed very little correlation between them. Apart from the experience level labelled ‘EN’ who had a narrow margin average salary increase all other experience levels had the same average salary. In my opinion, once the candidate have passed the job interview and the recruiting officer had found the candidate satisfactory and competent for the job position, the experience level did not play much role in the salary negotiations.

14. The average salary by company location over the three-year period revealed that Japan and Mexico experienced declining in the average salary of its data science employees. In Japan, the highest average salary ($118,705.04) was paid in 2000, which reduced to $114903.31 in 2001 and further reduced to $114,409.79 in 2002.

    In Mexico, the highest average salary ($116,762.04) was paid in 2000, which reduced to $114,244.52 in 2001 and further reduced to $112,490.28 in 2002.

    India (IN), US, UK and Germany (DE) experienced some amount of fluctuations in their salary pattern over the course of the period.

    Meanwhile, China experienced an increase in the average salary of its data science employee. In 2000, the average salary was $111,193.52 which increased to $111,493.03 in 2001 and further increased to $112,345.64



15. The analysis on the average salary by company size for the three-year period revealed that the large (L) company size experienced a decline in the average salaries of its employees while the average salaries of the medium (M) and small (S) companies kept fluctuating. In the large company, the highest salary ($116,550.32) was paid in 2000 which reduced to $116,508.55 in 2001 and further reduced to $111934.81 in 2002.

16. Analysis on average salary by company location and work setting revealed that companies located in Japan tend to pay more for those who worked in hybrid work setting ($120,540.02), followed by those who worked in-person ($118,479.53) and those who worked remotely were paid the least ($109,080.33).

    Companies located in Mexico and India paid more for those who worked in-person, followed by those who worked in hybrid work settings and those who worked remotely were the least paid. In Mexico and India, those who worked in-person were paid $119,083.31 and $ 115,230.59 respectively, those who worked in hybrid work settings received $114,238.91 and $ 114,532.16 respectively and those who worked remotely in Mexico and India were also paid $109,894.55 and $112,937.94 respectively.

    Companies located in the US payed more for those who worked in-person ($117,291.52), followed by those who worked remotely ($115,090.85) and then the hybrid workers ($109,507.34).

    Companies in UK also paid more for those who worked in-person ($115,642.53), followed by those who worked in hybrid work setting ($114,807.05) and the least for those who worked remotely ($109,697.82).

    In contrast to the companies in the above countries, Germany and China rather paid more for those who worked remotely receiving $113,545.49 and $117,823.82 respectively, followed by hybrid workers in Germany ($113,101.09) and in-person employees in China ($109,674.62) and the least paid employees in Germany were those who work in-person ($111,785.13) and in China were those who worked in the hybrid work setting ($108,218.08).

Conclusions

1. Data science job positions receives higher salaries when offered in companies located in Japan (JP), followed by Mexico (MX), then Germany (DE), US, India (IN), UK and China (CN)
2. Data science job positions offered in large companies pays higher salaries, followed by medium and then small companies.
3. The demand for the various data science jobs are almost the same but the most sort-after one is the Data Science job category.
4. All work settings (Hybrid, In-person and Remote) were preferred in the data science job openings.
5. Experience level plays little or no role in the salary negotiations for data science job position.

Limitations

1. The job experience level and employment type columns were filled with abbreviations which were not fully understood by me.
2. The null values, NA's were not included in the analysis