

3/31/2024

36104-DATA VISUALISATION AND NARRATIVES

INDIVIDUAL ASSIGNMENT



SHASHIKANTH SENTHIL KUMAR
25218722

INTRODUCTION:

In today's dynamic job market getting a job in the same field of study is more difficult. However, the field of data science stands at the forefront, offering boundless opportunities for those who are equipped with the right skills and expertise. With the proliferation of data-driven decision-making across industries, the demand for proficient data scientists continues to surge. But this information is not delivered to the people, most of them assume that we are in recession due to COVID-19, but we are not there are plenty of opportunities available. So, I would like to show what are all the job roles that are available for us the students of Master of Data Science and Innovation, and create awareness so that none of them should have the stress about getting a job after the completion of the course. We are going to use Data visualization because it helps in understanding complex datasets, particularly in the realm of human resources and employment. In this project, we are going to use the data set based on jobs and salaries obtained from Kaggle. Some organizations use similar datasets like LinkedIn, Seek, and Indeed.

REFERENCES:

1. T. Z. Quan and M. Raheem, "Human Resource Analytics on Data Science Employment Based on Specialized Skill Sets with Salary Prediction", Int. J. Data. Science., vol. 4, no. 1, pp. 40-59, May 2023.
2. A. Kaur, D. Verma and N. Kaur, "Utilizing Quantitative Data Science Salary Analysis to Predict Job Salaries," 2022 2nd International Conference on Innovative Sustainable Computational Technologies (CISCT), Dehradun, India, 2022, pp. 1-4, doi:10.1109/CISCT55310.2022.10046491.

PROBLEM DEFINITION:

Our analysis aims to reveal the various career opportunities within the data science ecosystem, elucidating the various employment roles, employment types, experience levels, and remunerative aspects of different roles.

Through visually compelling representations and insightful data exploration, we endeavor to address key questions such as:

1. Job Role Landscape: what are all the job roles available for data science graduates? From Data Engineer to Data Science Managers, we aim to find the top 15 job roles available in the year 2023.
2. Salary Insights: What salary expectations can students anticipate across various job roles? By analyzing salary data in conjunction with job titles and experience levels, we aim to provide valuable insights into earning potential in the data science domain.
3. Industry Trends: How job roles and salary expectations vary across different employment types and experience levels. By segmenting our analysis based on work year, we aim to find out the difference between the average salaries of each employment type throughout the years.

TOOLS USED:

The tools selected for this analysis are Python and Tableau.

1. PYTHON:

Python provides robust data analysis and machine learning libraries such as NumPy, Pandas, Matplotlib, SciPy, and Seaborn, enabling advanced statistical analysis and predictive modeling. This allows for smooth data manipulation and preprocessing, enabling users to prepare data for visualization effectively.

2. TABLEAU:

Tableau, on the other hand, offers intuitive data visualization capabilities, making it ideal for creating interactive and visually appealing dashboards. Its intuitive drag-and-drop interface makes it easy for users with various levels of technical knowledge. Tableau allows for quick prototyping and iteration of visualizations. It offers seamless integration with a wide range of data sources including databases, spreadsheets, and cloud services.

DATASET:

Dataset: ds_salaries.csv

Data source: <https://www.kaggle.com/datasets/hummaamqaasim/jobs-in-data>

Date data source created: 31/12/2023.

Data Schema: Refer to the below table for the full description of variables/features and their data types.

Variable name	Type	Description
work_year	Integer	Which year the job role is posted?
experience_level	String	What experience level do you need to get the job?
employment_type	String	What type of employment is the job?
job_title	String	What is the name of the job?
salary_currency	String	What is the currency of the salary?
employee_residence	String	What should be the residence of the employee?
remote_ratio	Integer	What is the percentage of chance for a remote job?
company_location	String	What is the location of the company?
company_size	String	What is the size of the company?
salaries	Integer	What is the salary for the job?

Here in the Experience level, SE refers to the Senior level, MI refers to the Mid-level, EN refers to the Entry level and EX refers to Executive level. In Employment type FT refers to Full time, CT refers to Contract, FL refers to Freelance, and PT refers to Part time.

Importance of selected Dataset:

1. Comprehensive Representation:

This dataset gives a comprehensive representation of job roles available across various employment types, experience levels, and locations. This gives a whole view of the job landscape and helps job seekers to choose the jobs easily.

2. Salary Information:

With the salary data included, we can see the different salary trends across job roles, experience levels, employment types, and locations.

3. Geographical Variance:

With the location data included in the data set, we can see the variations in job opportunities and salary levels across the world.

4. Industry Insights:

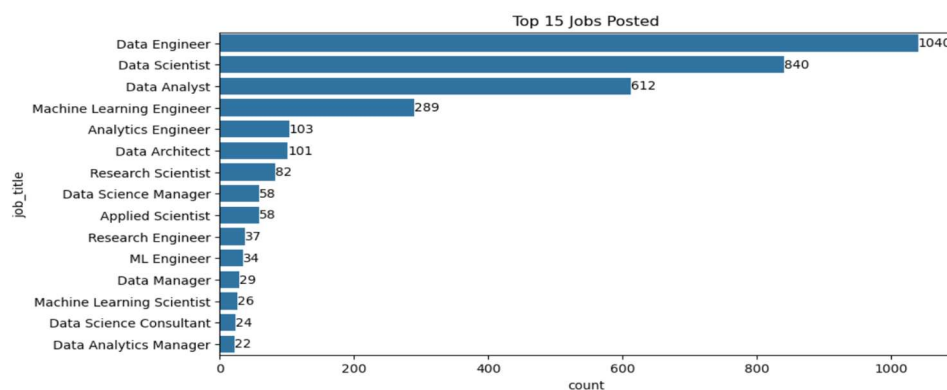
With this data, we can see the comparison between employment types and experience levels helps to find the available employment types and demanding experience levels.

Overall, the selected data set offers a valuable source of information for exploring job opportunities, salary trends, etc within the job market.

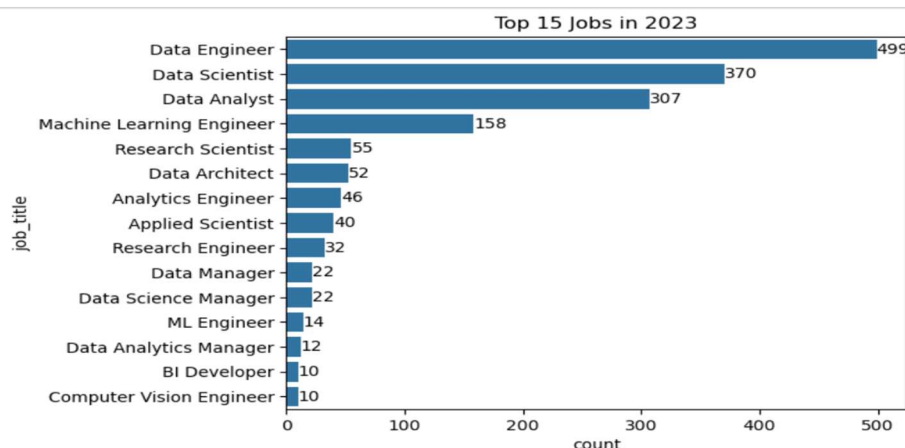
PLOTS:

PYTHON:

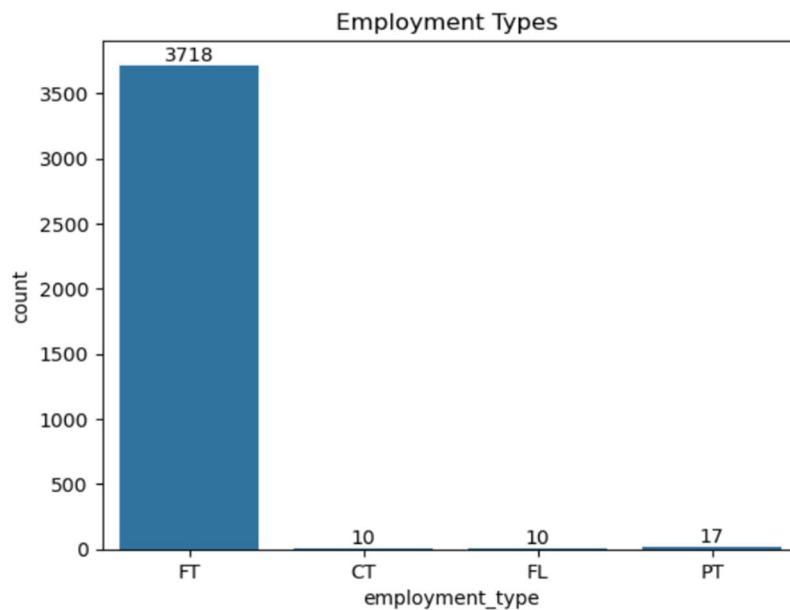
1. The below plot shows us the top 15 jobs and their count. we can see that Data Engineer, Data Scientist, and Data analyst jobs are available most, these jobs count are more than half of the total count of jobs.



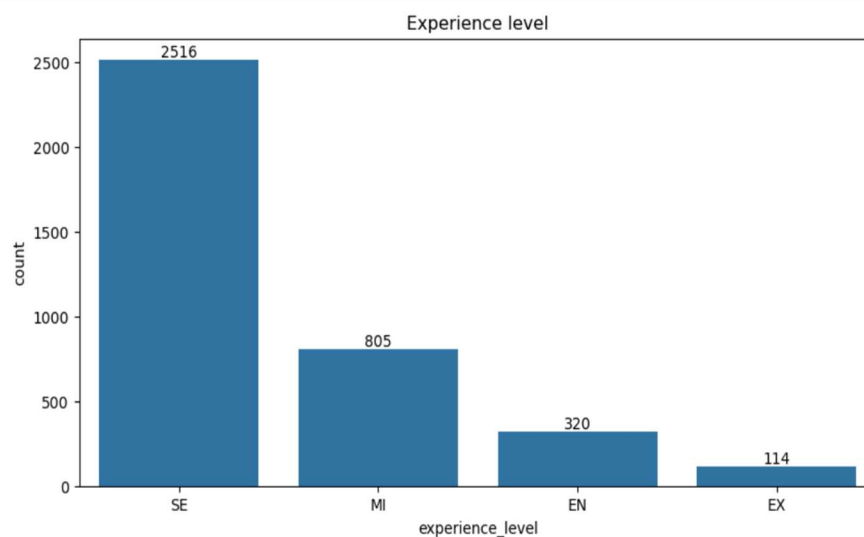
2. The below plot shows us the most 15 jobs posted in 2023 and their count. we can see that Data Engineer, Data Scientist, and Data analyst jobs are in demand consecutively.



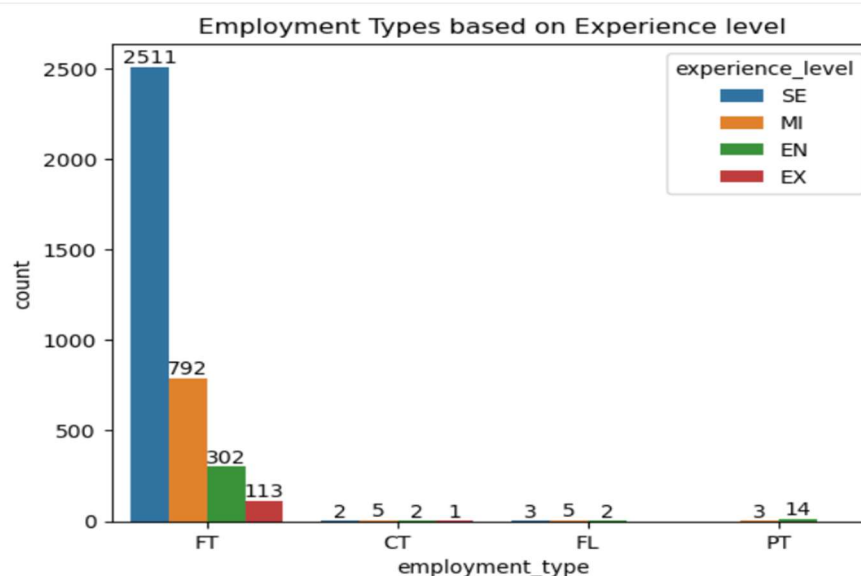
3. The below plot explains what types of employment opportunities are available. we can see that most of the jobs are available full time.



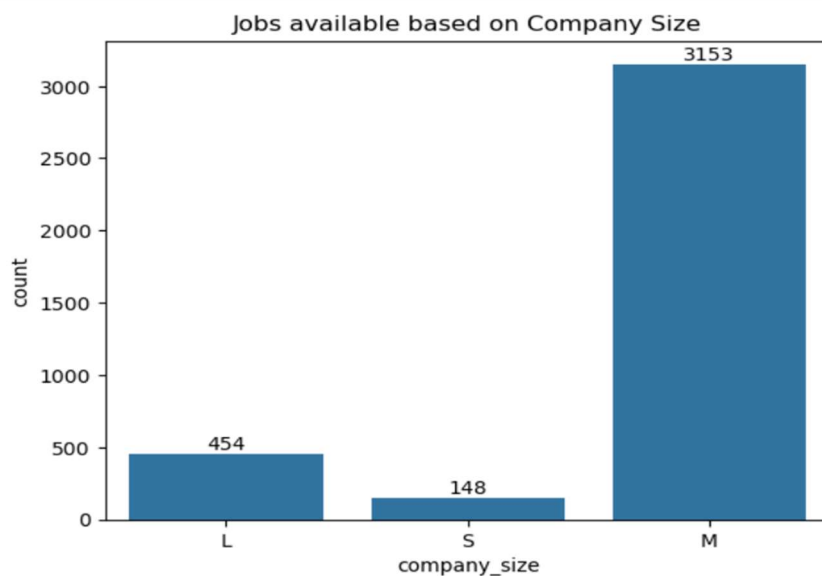
4. The below plot explains what types of Experience levels are available. we can see that senior-level experience is in demand.



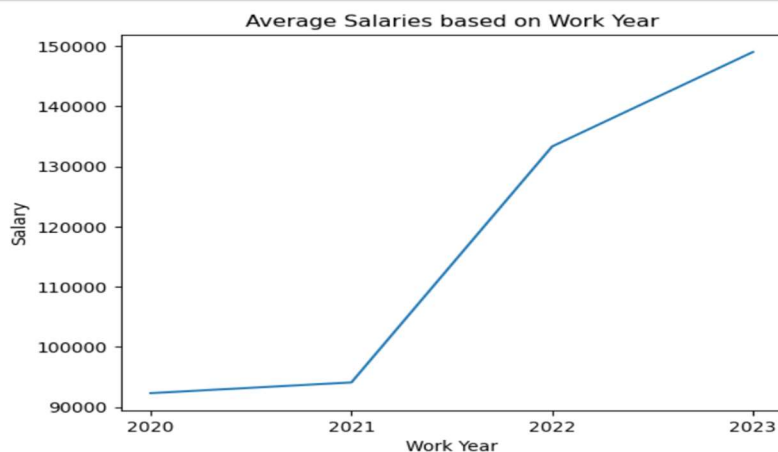
5. The below plot gives details of what are the employment types that are available based on experience levels. Mostly full-time jobs are available for all experience levels.



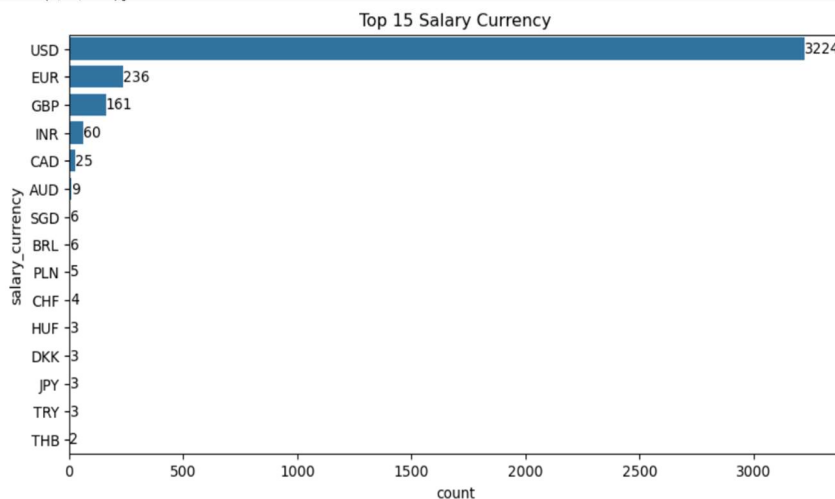
6. The below plot explains the jobs available based on company size. Here L stands for Large, M stands for Medium and S stands for Small. we can understand that a medium-sized company gives more jobs than the others.



7. The below plot gives the correlation between the job and the work year. we can understand that there will be a rapid increase in salary after the year 2021, so we under that the demand for the data science field is increased between 2021-2022.



8. The below plot gives the top 15 salary currencies that are paid. we can see that 90% of the salaries are paid in USD.



- This plot explains the availability of remote jobs in a country. we can identify the availability ratio of remote jobs with the help of the legend mentioned on the right side of the plot.

Remote Jobs Locations

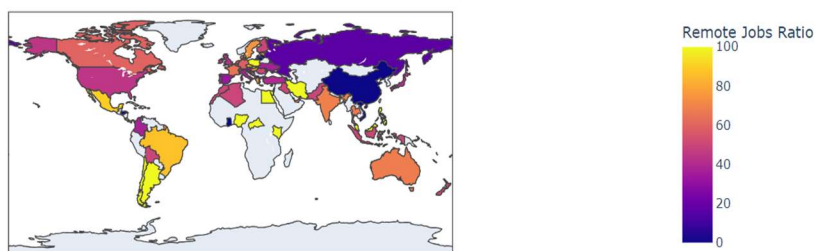
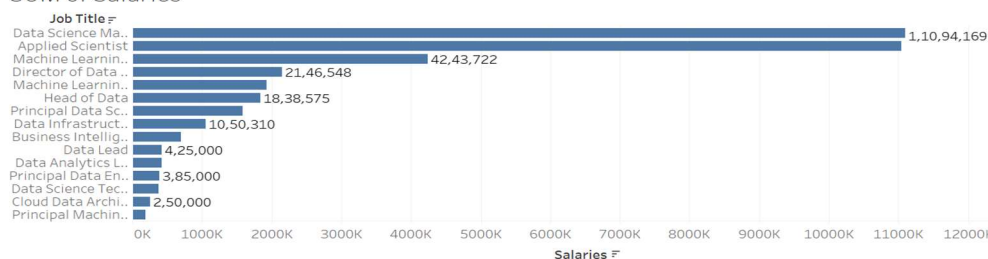


TABLEAU:

- The below plot shows us the top 15 most-paid jobs. we can see Data science manager and Applied Scientist are jobs that pay nearly 11 million.

SUM of Salaries

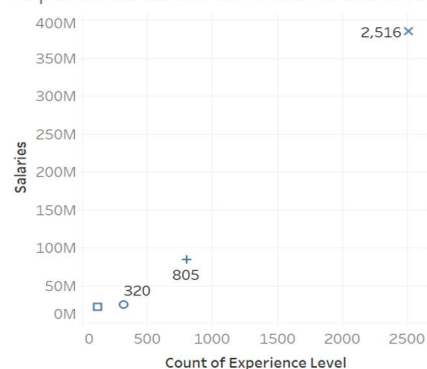


Caption

Sum of Salaries for each Job Title. The marks are labelled by sum of Salaries. The view is filtered on Job Title, which keeps top 15 of 93 members.

- The below plot shows us the relationship between the sum of salaries and the count of jobs based on experience levels.

Experience Level vs. SUM of Salaries



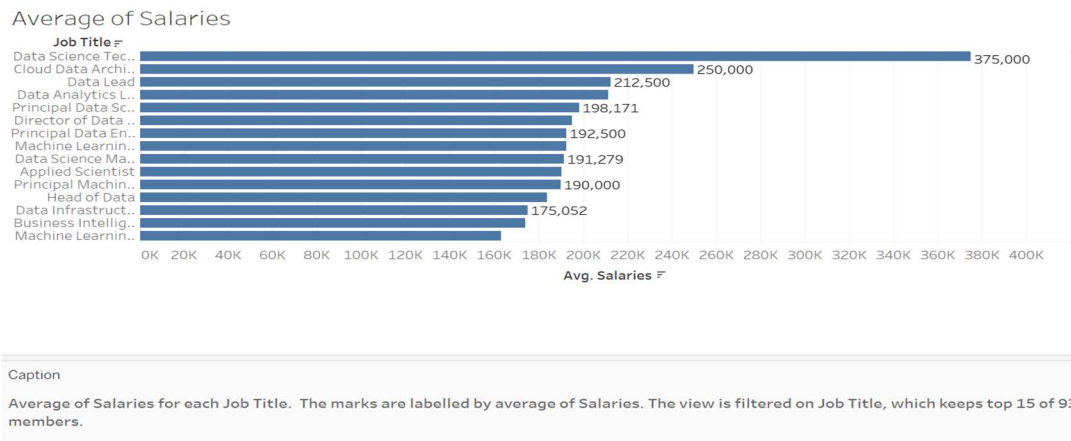
Experience Level

○ EN
□ EX
+ MI
× SE

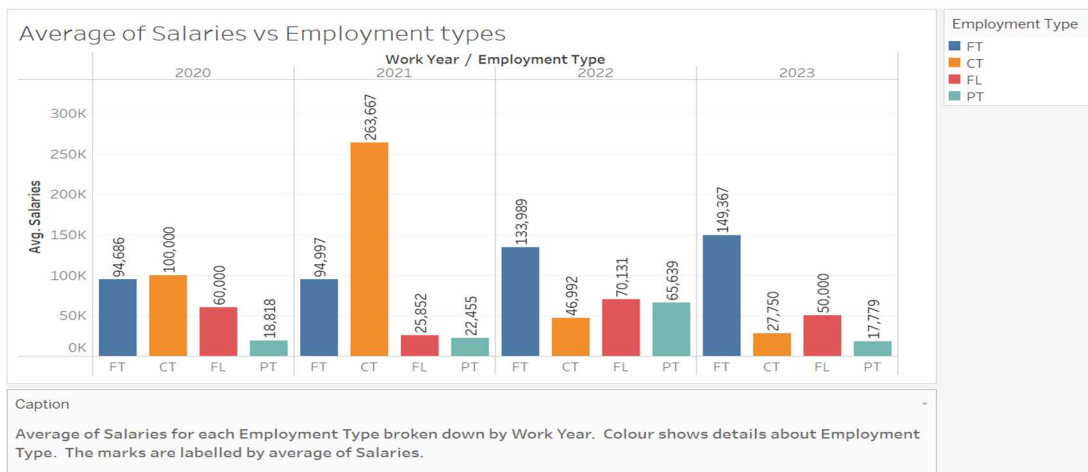
Caption

Count of jobs as per Experience Level vs. sum of Salaries. Shape shows details about Experience Level. The marks are labelled by count of jobs.

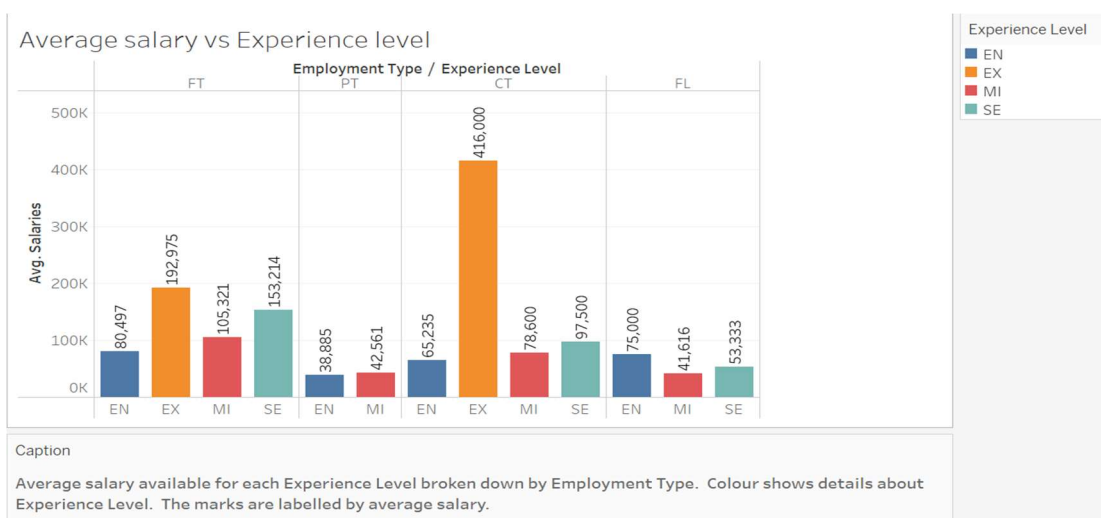
3. The below plot shows the top 15 job titles and their average salaries. we can see that the Data Science Tech Lead job came at the top in average salary.



4. The below plot explains the average salary of each employment type based on year. we can see the peak year for each employment type are FT-2023, CT-2021, and FL & PT-2022.



5. The below plot explains the availability of jobs on each employment type based on experience levels. we can see that executive-level employees get average pay more in contract than full time.



6. The below plot explains the relationship between average salary and company size based on year. we understand that apart from 2021 in all the other years medium-sized companies have given more average salaries.

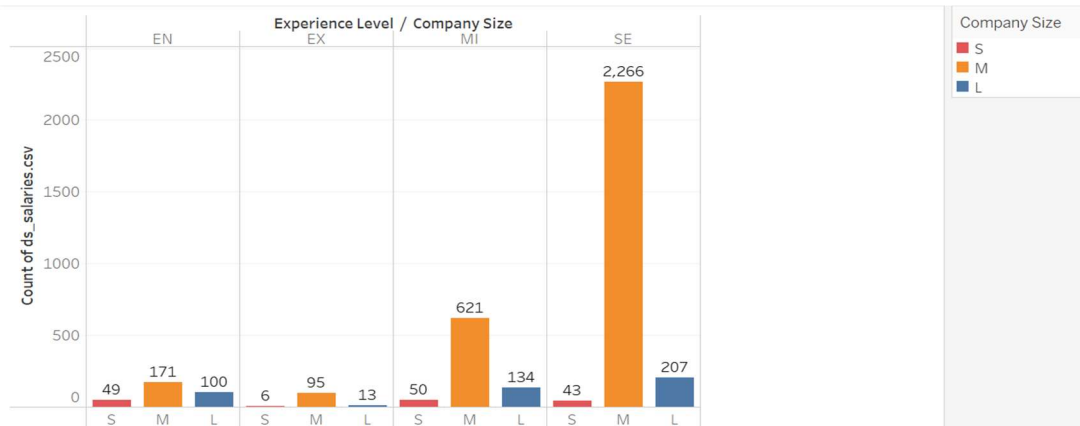
Average of salaries vs Company size for each Year.



Caption

Average of Salaries for each Company Size broken down by Work Year . Colour shows details about Company Size. The marks are labelled by average of Salaries.

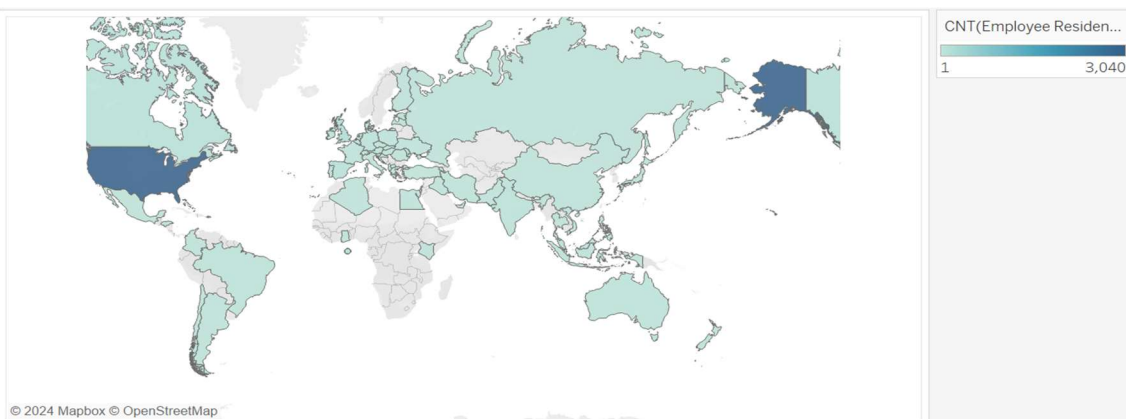
7. The below plot shows the required experience level as per company size. we can see that medium-sized companies offer more jobs than other size companies and after senior-level intermediate experience level is wanted more.



Caption

Count of Jobs available for each Company Size broken down by Experience Level. Colour shows details about Company Size. The marks are labelled by count of jobs.

8. This plot explains the employee residency for the jobs. we can identify the differences in countries with the help of the legend mentioned on the right side of the plot and we found that the US consists of more jobs in the Data science field.



Caption

Map based on Company Location and Employee Residence. Colour shows count of Employee Residence. Details are shown for Company Location.

DISCUSSION:

We have done data visualization for jobs and salaries data set and created 17 plots in total which helps us to understand the insights in the job industry including Job role landscape, Salary insights, and industry trends.

1. Job Role Landscape: We have found that there are 93 job roles available in the data science field and found the top 15 jobs available in the year 2023 in plot 2 under Python, we also found the overall top 15 jobs in plot 1.
2. Salary Insights: We have found the top 15 job roles that gave more salary in Plot 1 and the top 15 job roles based on the average salary in Plot 3 under Tableau. From plot 5 we can see that the executive level employees get more average salary than the others.
3. Industry Trends: We have found the average salaries for each experience level based on employment type in plot 5 under Tableau. we also found the difference in average salaries for each employment type throughout the years in plot 4 and Most of the job roles are available for senior experience levels only in plot 2.

Overall, we covered the job landscape and salary insights based on experience levels, employment types, company size, and work year. We also covered the remote jobs available in a country in plot 9 under Python and the location of the employee residence based on company locations in plot 8 under Tableau.

CONCLUSION:

Data science is the most in-demand and rapidly growing field in today's world. In this report, we have covered the opportunities available in the data science fields like job roles, salary trends, job location, company size, and industry trends like experience level and employment types that are available and in demand. In conclusion, with the help of this data visualization one can understand the insights in the job industry and be able to choose their desired job easily.