**Hiranmai Vallabhaneni**

**U25850442**

Business Analytics and Information System

**Final Project Report**

Data Science Job Analysis

**Data Visualization**

Prof.Johannes Reichgelt, PhD

**Data Science Job Analysis**

**1.Introduction to the dataset**

Data science is a branch of study that uses cutting-edge technologies and processes to analyse vast amounts of data to find hidden patterns, gather useful information, and make business decisions. To create prediction models, data scientists use sophisticated machine learning algorithms.

The data used for analysis can be presented in a variety of ways and come from a wide range of sources. Data science is one of the most fiercely debated issues in IT circles and is essential to many sectors given the large volumes of data being produced today.

Through these datasets, I would like to analyse different types of jobs in the field of Data Science from the year 2020 to the year 2023, categorized by Salary, Remote Ratio, Location, Experience Level, and company size.

The datasets used for the job analysis combinedly consists of 4800 entries with columns work\_year,experience\_level,employment\_type,job\_title,salary,salary\_currency,salary\_in\_usd, employee\_residence,remote\_ratio,company\_location,andcompany\_size.

These datasets provide valuable information for analysing the job market and understanding the factors that influence salary levels for different positions.

The work\_year column indicates the number of years of work experience of an employee, while the experience\_level column indicates their level of experience. The employment\_type column specifies whether the job is full-time, part-time, or contractual. The job\_title column provides information on the employee's job title, while the salary column lists the employee's salary in their local currency.

These datasets also include a column indicating the employee's residence, which can provide insights into regional variations in salaries. The remote\_ratio column indicates the percentage of remote work allowed for the job, while the company\_location column provides information on where the company is based. Finally, the company\_size column provides information on the size of the company, which can also affect salary levels.

As a result, this dataset can be utilized to answer a variety of research questions on income trends and job market studies. For instance, we can examine the link between work experience and pay scales or investigate whether full-time versus part-time employment results in different earnings. The dataset can also be used to examine geographical differences in pay as well as whether factors like company size or the availability of remote employment affect pay scales.

**Dataset Reference Links**

Here are the references for the datasets that I have chosen to draw out all these visualizations.

**https://www.kaggle.com/datasets/niyalthakkar/data-science-jobs-analysis**

**https://ai-jobs.net/salaries/download/**

**2.Methodology**

Here are a few research questions that have been answered.

**Salary analysis**: We can create visualizations and tables to analyse salary based on various factors such as work year, experience level, employment type, job title, employee residence, remote ratio, company location, and company size. We can also create calculations to compare salaries in different currencies or convert them to a common currency.

**Employee retention analysis**: We can analyse the factors that contribute to employee retention, such as work year, experience level, employment type, and salary. We can create visualizations to show the retention rate by different factors and identify trends and patterns.

**Job market analysis**: We can analyse the job market by studying the demand for different types of jobs, the average salary for different job titles, and the distribution of jobs by location, company size, and employment type. This can help to identify the most lucrative job opportunities and industries.

**Remote work analysis**: We can analyse the impact of remote work on various factors such as employee satisfaction, retention, and job performance. We can create visualizations to compare the performance of remote workers versus office workers and identify best practices for remote work.

Based on the above discussion following are the research questions.

1. What is the frequency of employees countrywide and patterns?
2. What is the Remote ratio those organizations are following within their organizations?
3. What are the average salaries that are being offered in all the years 2020 to 2023?
4. Which jobs are highly paid?
5. What is the impact of experience on salary?
6. Is there any relation between salary and the company size?

**3.Analysing the dataset**

For analysing the dataset, we have used a visualisation tool known as “Power BI Desktop”. This tool helps to analyse and draw meaningful insights and visualizations which represents the data in both the datasets. For analysing the datasets, we have created a link between both the datasets in the Power BI software in Microsoft excel format.

1. **Country-wise Employees and their Frequency:**



**Figure 1**: Company Locations.

The first visualisation, which is made by using Power BI explains about the location of the company, and for this visualization we have used map to showcase various locations where all the companies are located which is shown using the image. In this we have displayed the data of companies which are offering data science jobs based on various geographical locations.

1. **Remote Jobs By title:**

**Chart, pie chart

Description automatically generated**

**Figure 2**: Remote Jobs By title

The second visualization was designed using the remote jobs and job title which are being offered mostly as remote. There are various countries in which remote job facility is available. There are a lot of countries where the ratio is as high as 100% and there are also places where the remote job ratio is only 50% and null which means no remote jobs are offered to employees, which means they are totally in work from home mode which is the need of the hour considering the continuous use of resources we are making now a days with the developing technologies and improvement in availability of resources.

1. **Country Wise Employees:**

**Chart, bar chart

Description automatically generated**

**Figure 3:** Country Wise Employee

For this visualization we have used a clustered bar graph displaying the results of locations where the highest number of employees are working and as we can see that the visualization shows that USA is the place where most of the people are working in the data science jobs and in the next place comes Canada with only a minimum difference when we compare the numbers and then comes India which is almost equal to Canada where people are working in Data science field.

1. **Average Salary offered in US currency:**

**Chart

Description automatically generated**

**Figure 4**: Average Salary in US currency

In order to showcase this visualization we have used stacked bar chart, where we studied the datasets for average salary as per the title of the job being provided to the employee and it can be observed that “Data Analytics Lead” is getting the highest salary which is followed by “Principal Data Engineer” and “Financial Data Analyst “followed by rest of all the other jobs as per the currency of United States of America.

1. **Average salary by country and Location:**

****

**Figure 5:** Average salary In USD by company Size

In this visualization we have drawn a tree map diagram in which we can see that the USA leads the chart with average salary. This is followed by Japan and Canada is at number three position followed by rest of the other countries.

1. **Employee And Their Residence:**

Map

Description automatically generated**Figure 6:** Employee and their Residence

In the above visualization we have shown the residence of employee according to the geographical location. In this it is clearly visible that as per the bubble size depicted in the map US is the country where most of the people are living which is followed by “GB”(Great Britain) and “IN”(India) respectively.

1. **Experience level wise Average salary:**

**Chart, bar chart

Description automatically generated**

**Figure 7.** Experience level wise Average salary

The above visualization shows the impact of experience on salary. As we can see from the chart, there is a clear correlation between experience and salary and this visualization shows that pays that are offered to employees according to their level of experience and we can observe that middle level jobs are taking up most of the salaries when we compare with the other experience levels such as entry level, senior and expert level in the field of Data science.

1. **Year Wise Job Title and Average Salary in USD:**Table

   Description automatically generated

**Figure 8.** Year Wise Job and Average salary

The above visualization table, shows the data of various jobs and salaries offered in all the years from 2020 to 2023 and most highest paid job was Research scientist in all the consecutive years 2020, 2021, 2022, 2023 and then followed by Principal Data Scientist in comparison with all other jobs such as data engineer and data scientist and this shows how the trend is shifting towards data science and the need of data is changing day by day making it new fuel.

**Dashboard:**

**Dashboard 1:**

Graphical user interface, application, map

Description automatically generated

In the above dashboard we have displayed the results of various categories such as salary that ie being offered as per the experience level and how much is offered as per US currency, ratio of remote jobs by title and location of the company, and we are filtering the results of each one of factor as per the year and company size employee worked.

**Dashboard 2:**

Map

Description automatically generated

In this dashboard we have shown the number of employees residing in each one of the geographical locations and what is the average salary that the employees were paid as per the company location and the year in which the job had been offered to the employee.

**Conclusion**

• The highest paid jobs in the data science field are Research Scientist, Principal Data Scientist and Principal Data Analyst.

• Experience has a clear impact on salary. Salaries increase with experience and the highest paid employees are those with 10+ years of experience.

• Larger companies tend to pay higher salaries than smaller companies. The highest paid employees are those working in large companies with more than 10,000 employees.

• Remote job opportunities are available in many countries and companies with remote work options have a higher employee retention rate.

• The demand for data science jobs is high and the average salary for data science jobs has been increasing over the years.

• Companies in the United States, India, and the United Kingdom are the largest employers in the data science field.

Overall, the dataset provides valuable insights into the job market trends in the data science field and can be used to guide job seekers and employers in making informed decisions about job opportunities and compensation. In general, we can say that the concept of average income is an important economic indicator that can provide insights into the overall economic growth of a country. The salary structure gets impacted by the experience level and the firm size as well. The location of the company also affects the salary. Nowadays, data science-related jobs are tending, and they will remain to be on top for some time until and unless some record-breaking techniques of data handling do not arrive just like AI.

**Further Research Questions:**

1.How do data science job requirements and salaries vary across different sectors, such as healthcare, finance, and e-commerce?

2.What is the gender and racial diversity of the data science workforce, and how has it changed over time?