**Report on the Business Analyst Careers Data**

**Overview of the tool used:**

* For the analysis of this data we have used the R programming along with several packages such as **“ggplot2”** and **“dplyr”** packages.
* R is a widely used language for statistical computing and graphics. It has an extensive collection of packages that offers various functionalities to perform data manipulation, and statistical analyses.
* The ggplot2 package is a data visualization package in R. With ggplot2 you can create a wide range of plots, including scatter plots, bar plots, line plots, etc.
* The dplyr package is another essential package for data manipulation and data wrangling in R. With this package you can easily perform common data manipulation tasks, such as filtering rows, selecting columns, aggregating data, and creating new variables.
* Firstly, the data was checked for duplicates or errors which were then fixed and the data was cleaned before moving on to data exploration.
* Then we conducted data exploration to identify patterns, relationships, and insights within the dataset. We used various techniques such as bar plots and histograms to visualize the data better and understand its characteristics.

A graph of a graph

Description automatically generated**Data Exploration**

The above graph depicts the salary range distribution in the form of bar plots. We can clearly see that the most number of companies have a salary in the range of 48k-87k. We can also see that the number of unknown values in the dataset is pretty large.

A screenshot of a computer

Description automatically generated

The above image depicts the job titles in the dataset, there are a lot more job titles but I have depicted only some in the image as it was difficult to fit every job title in a single image.

A grid of black numbers

Description automatically generated

The above image shows the rating of the companies according to the job title as well. Most of the rating is above 4 whereas there are rating that are very low as well about 1 or below 2. Again, there are a lot of missing values which need to be addressed.

A screenshot of a computer

Description automatically generatedThe above image depicts the location of the companies and we can see that most of the companies are situated in New York City, NY.

A screenshot of a computer

Description automatically generated

The above image depicts the sectors of the jobs are available in and we can see that most of the jobs are in the sector of Information Technology, Finance and Business Services.

A screenshot of a computer program

Description automatically generated

The above image represents the top 20 sectors the jobs are available in. We can see that healthcare is the most sought-after sector followed by Finance, Manufacturing, and Information Technology. I have used the Dplyr package this find the top 20 distinct sectors.

A screenshot of a computer code

Description automatically generated Just like how the top 20 distinct sectors were represented, similarly, the top 20 types of ownership have been represented in the above picture using the same Dplyr package.

A screenshot of a computer code

Description automatically generated

The above image gives the top 20 distinct headquarters of the companies, we can see Zurich is at the top which means most of the companies' headquarters could be in Zurich or the above picture could also depict the top 20 headquarters in terms of how the data is constructed.

A screenshot of a computer code

Description automatically generated

The above picture depicts the top 15 job titles according to the rating. We can see that all the top 15 have the same rating which is 5 and most of the job titles are related to business analysis. I first arranged the rating in the descending order and then extracted the top 15 job title according to the rating.

A screenshot of a computer

Description automatically generated

The above picture gives the top 15 job title according to the rating in the information technology sector. I have used a filter function from the dplyr package to select the job titles in the information technology sector. The most sought-after job title in the information technology sector is again the Business analyst position.

A screenshot of a computer

Description automatically generated

The above image represents the top 10 companies that have a rating greater than 3 and are in the information technology sector. In this as well, I have used the filter function and then extracted the company name along with the rating.

A screenshot of a computer

Description automatically generated

Just like the previous visualization, this image represents the list of top 10 companies according to the rating in the accounting and legal sector.

A screenshot of a computer

Description automatically generated

Similarly, the above visualization represents the top 10 companies according to the rating in the oil, gas, and energy sector.

A screenshot of a computer screen

Description automatically generated

In contrast to all the visualization above, this image above represents the least sought-after job titles according to the rating. The bottom 15 jobs has been shown above in the image. We can see that Business Analyst is present here as well. Earlier, We had said that Business Analyst is the most sought-after job looking at the rating but this could be also depending on the salary or the company that is offering this job. Hence, it could be both most and least sought after depending on the company and the location.