Avisha Singh
(2088371)

Data Science Job Salaries
Report 1

Objective The purpose is to analyze the average job salaries of Data Science Employees based on their experience level and job title varying from country to country. The idea is to get to know about the current market in Data Science over a period of time.

Dataset The dataset is about the Data Science Job Salaries of employees with different experience levels and employment types. It consists of Entry-level, Junior Mid-level, Intermediate Senior level, Expert Executive level, and Director positions employees. It also talks about the work done remotely by these employees in different countries. We can get to know about highly paid data scientists country-wise over a period of time. And also, analyze which job title is in more demand by just comparing the gross salary paid.

Visualization using Tufte's theory *Figure 1* is depicting the average salaries of data science employees with different experience levels and Job titles.

Conclusion- The Head of machine learning at an Expert Executive-level position is a highly paid job. The labels are denoting job titles and their average salaries with the help of the horizontal toolbar in *Figure 1* for analyzing the experience levels.

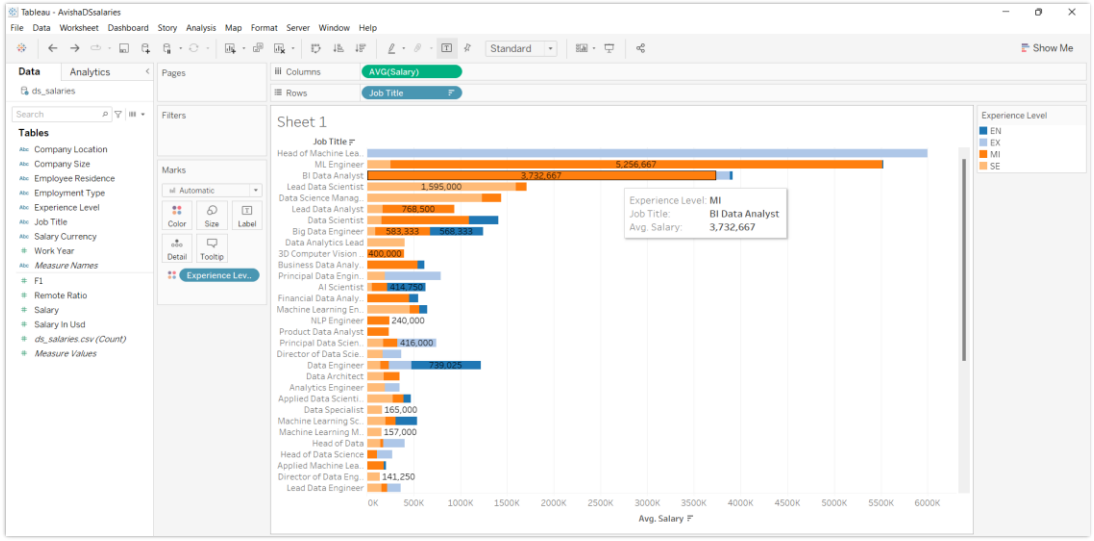

Figure 1: Average salaries of data science employees

Figure 2 is depicting the average hours that employees with different job titles spent remotely.

Conclusion- There are few job titles with employees not working remotely which is less than 20%, partially remote, and fully remote which is more than 80%.

Figure 2 is depicted by the side-by-side bars which are used to depict the average remote ratio of various job titles in different colors. And labels are also added in *Figure 2* to depict the average remote hours of various employment types.

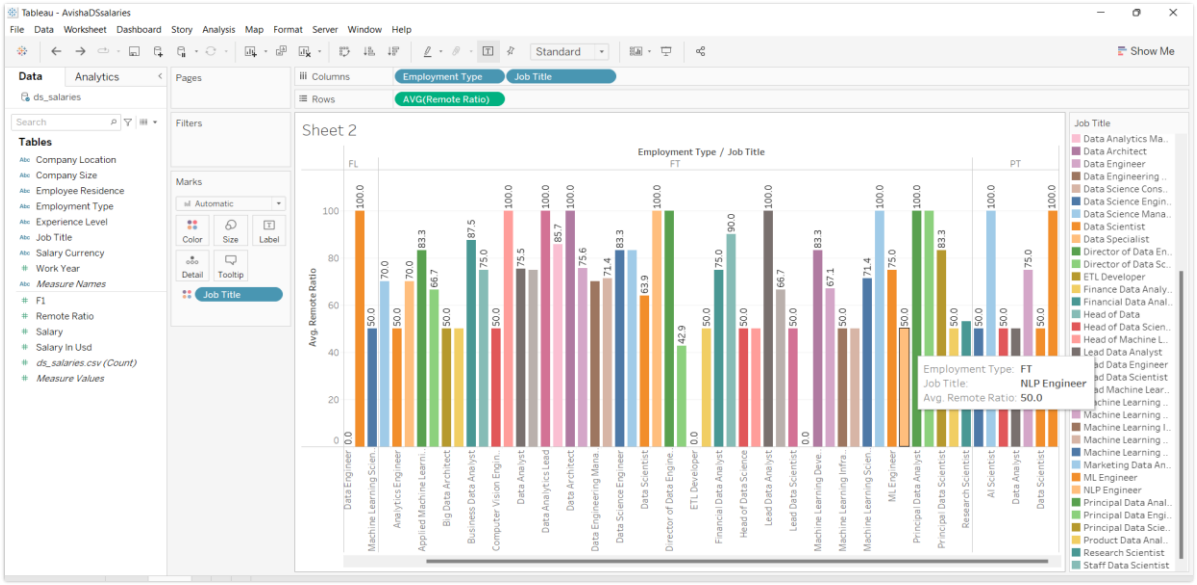

Figure 2: Average hours spent remotely by employees with different job titles

Conclusion

- By following Tufte's theory we are able to visualize the data in such a way that it can provide us with the best outcomes in an understandable way and also by maintaining graphical excellence during the visualization. Hence, by following his principles we are able to improve the quality of visualization.
- The visualization consists of large data sets as mentioned in Tufte's theory that reveals the data at several levels of detail from a broader overview.
- At last, GRAPHICS REVEAL THE DATA.

Visualization using Tufte's and Colin Ware's theory *Figure 3* is depicting the average salaries of data science employees with different experience levels and Job titles for the year 2020.

Conclusion- The average salary in USD for entry-level to executive-level positions for most of the job titles is higher than the senior-level position for the year 2020.

Figure 3 is depicted by the stacked bars toolbar used to analyze the experience levels in different colors for the year 2020.

The labels are also added in *Figure 3* for job titles depicting the average salaries in USD for the year 2020.

Avisha Singh
(2088371)

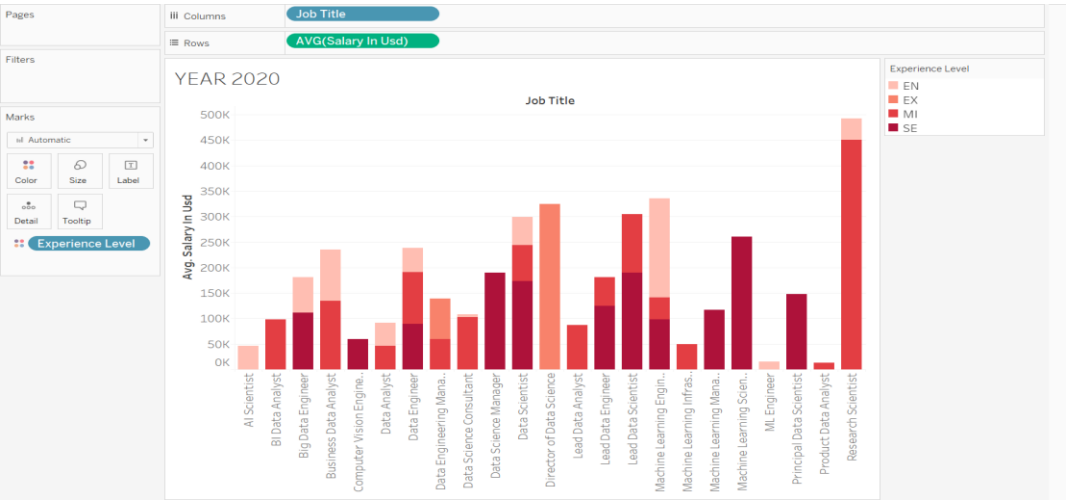

Figure 3: Average salaries for different experience levels (year 2020)

Figure 4 is depicting the average salaries of data science employees with different experience levels and Job titles for the year 2022.

Conclusion- We can see in *Figure 4,* the change in the salaries of employees for the year 2022. The average salary in USD for senior-level positions is rising.

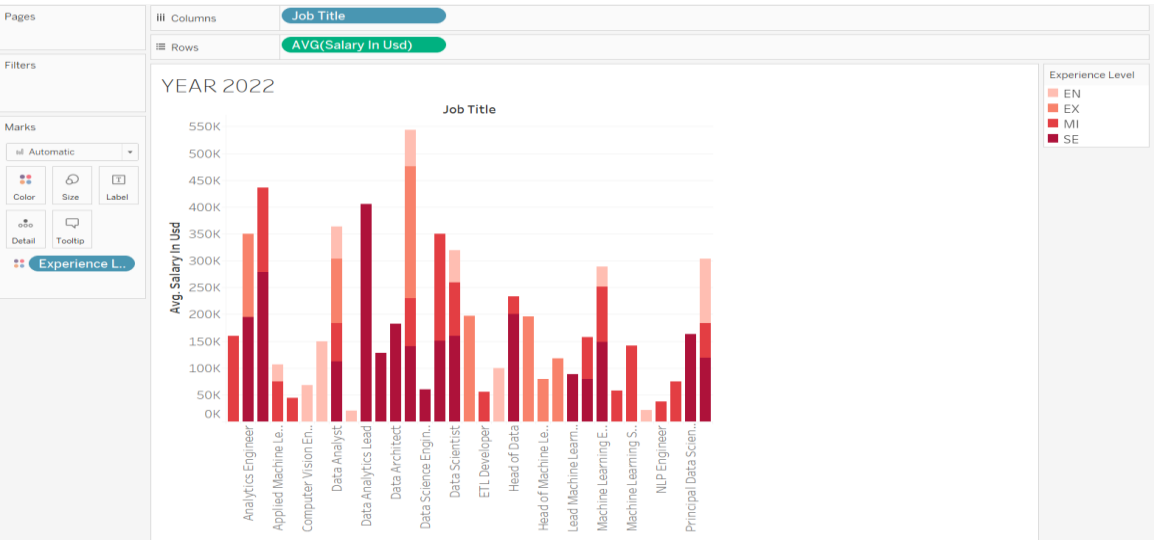

Figure 4: Average salaries for the year 2022

GLYPHS The glyphs in *Figure 5* are representing different experience level positions, each in different shapes. The circle-shaped glyph is indicating an entry-level position, the square shape is indicating executive level, plus sign is indicating mid-level and the cross sign is indicating the senior-level position for average salaries.

Figure 5: GLYPHS

Principles of Tufte's and Colin Ware's theory

- By following Tufte's theory we are able to visualize the data in such a way that it can provide us with the best outcomes in an understandable way and also by maintaining graphical excellence during the visualization. Hence, by following his principles we are able to improve the quality of visualization.
- The visualization consists of large data sets as mentioned in Tufte's theory that reveals the data at several levels of detail from a broader overview.
- By following Colin Ware's principle, we are able to apply theories of perception to the design of data visualizations by using glyphs.
- The different facets of visual perception like color, texture, and organization are used.

Hypothesis

- Hypothesis: - Gradually, there has been a rise in the average salaries of employees starting from entry-level to senior-level over the years.
- The dataset is proving the hypothesis which can be seen in the visualization of different years.
- This can be proved as the bar is rising for a senior-level position in the year 2022 depicting the average salary of employees.

Avisha Singh
(2088371)

References-

https://www.kaggle.com/

https://thedoublethink.com/tuftes-principles-for-visualizing-quantitative-information/

https://www.tableau.com/products/desktop/download

https://www.northeastern.edu/visualization/events/colin-ware-visual-queries-and-design-nuvis/

https://mdl.library.utoronto.ca/dataviz/design-principles#:~:text=One%20of%20Tufte's%20ideas%20is,else%20not%20important%20to%20understanding.