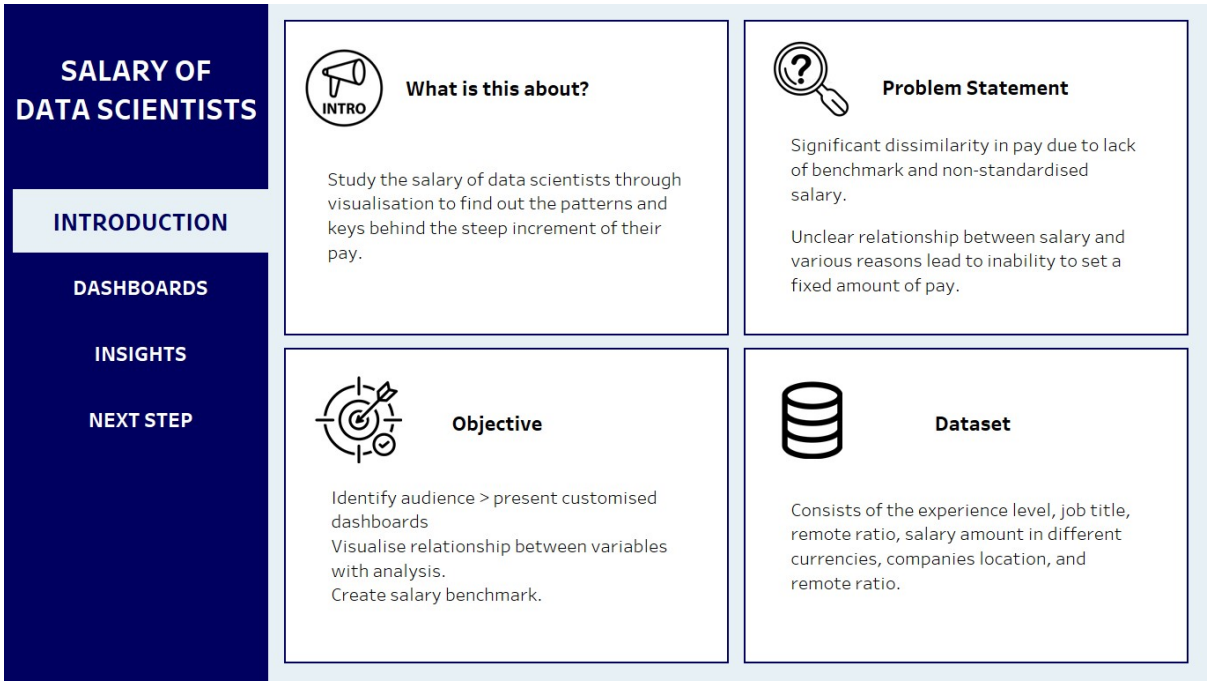
Client Report

The dashboards have 2 sections: Menu on the right, content on the left. The menu serves as a list of contents, where all titles are displayed. The title will be highlighted when its content is being presented to make sure the audience is not lost in the translation process.

In the Dashboards section, each of the diagrams has a caption as an explanation to avoid misinterpretation and misunderstanding. Several interactive elements are used throughout the dashboards, such as dropdown buttons, to act as filter to select required data and to ensure audience engagement. Teal blue colour scale is used for all of the diagrams that involve numbers, darker colour represent higher number or vice versa. The red numbers at the diagram are sequence of reading the dashboards.

Dashboard #1: Introduction



The “Introduction” dashboard is the overview of project where the background, problem statement, objectives and dataset information are covered to give a clearer picture of the study for the audience before the data visualization dashboard's presentation.

a. <u>What is this about</u>

The background of the data visualisation study focused on salary of data scientists,

where the rise in demand leads to tremendous increase in their pay within a decade. The variables in the dataset are used for the analysis of finding the patterns of the salary increment.

b. Problem Statement

Highlight the issues regarding the salary dissimilarity among data scientists. The two key issues being the absence of benchmark for candidates in separate locations, experience level, remote ratio, etc.; uncertain measure of skill sets as metric to standardise the salary amount.

c. Objective

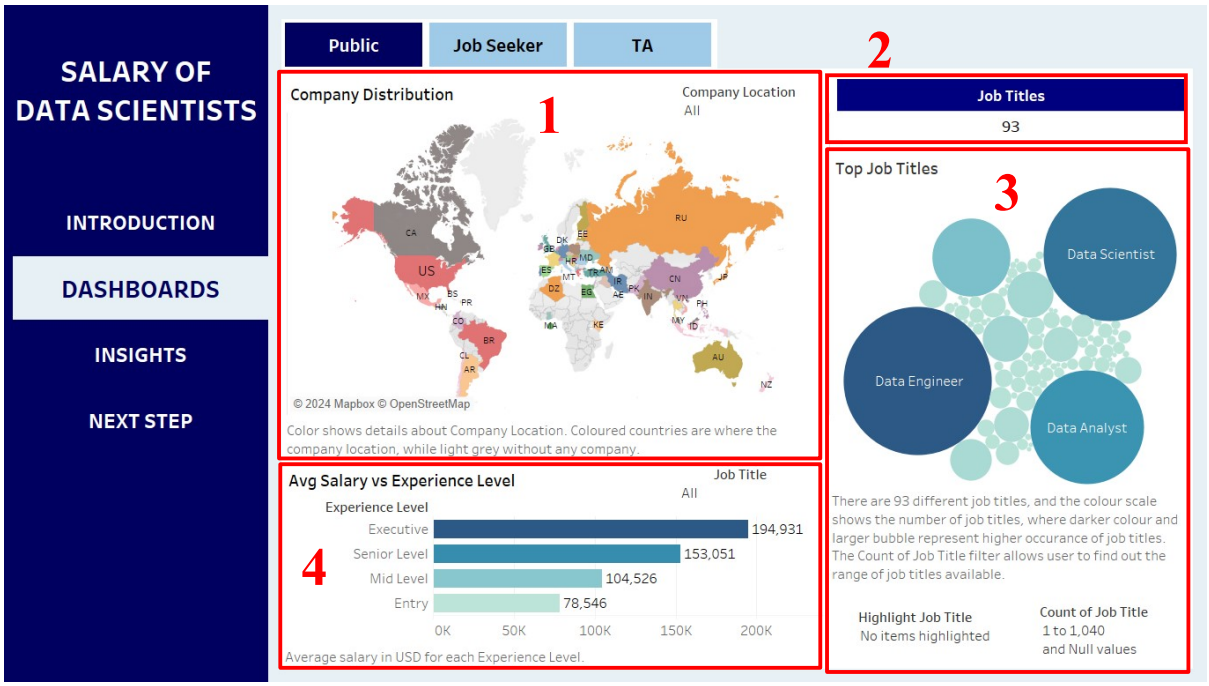
The goal of the study is to create a salary benchmark of various data scientist related roles by analyzing how salary is affected by the variables, such as location and experience. Then, create customized dashboards to feed the right and necessary information to different audience groups based on their roles, which are public, job seekers, and talent acquisitions in this case.

d. Dataset

The dataset named “2023 Data Scientists Salary” (Shan, 2023) is collected from Kaggle, with details that would help in analysing the salary amount. The fields included in the dataset are work_year, experience_level, employment_type, job_title, salary, salary_currency, salary_in_usd, employee_residence, remote_ratio, company_location, and company_size. Most of the fields would be used for analysis and visualisation.

Dashboard #2: Public, Job Seeker, Talent Acquisitions (TA)

2.1 Public



Public dashboard is created for the public. The information presented in the dashboard is straightforward for the audience to understand. In the board, the audience will get the basic information needed about the field of data science: number of roles related, location of job opportunities, and the pay in unique experience.

The data visualization techniques used are:

a. Company Distribution Map

The map is a provides a clear overview of company distribution, where the coloured countries are, where companies or potential job opportunities are located. In other words, the map displayed the population of data scientist related roles in the global scale.

Meanwhile, countries without companies recorded in the dataset are grey in colour. The audience can filter the countries with the dropdown at top right to display only the selection.

b. Job titles table

The table shows 93 distinct number of job titles in the data science field.

c. Top Job Titles Packed Bubble chart

The popularity of each job title is shown in the chart based on the size and colour of the bubbles. Darker and larger bubbles represent a higher number of titles found in the dataset or vice versa, which also indicates the popularity of those roles. The 3 titles that stand out among 93 of them according to sequence in descending order are: Data Engineer, Data Scientist and Data Analyst.

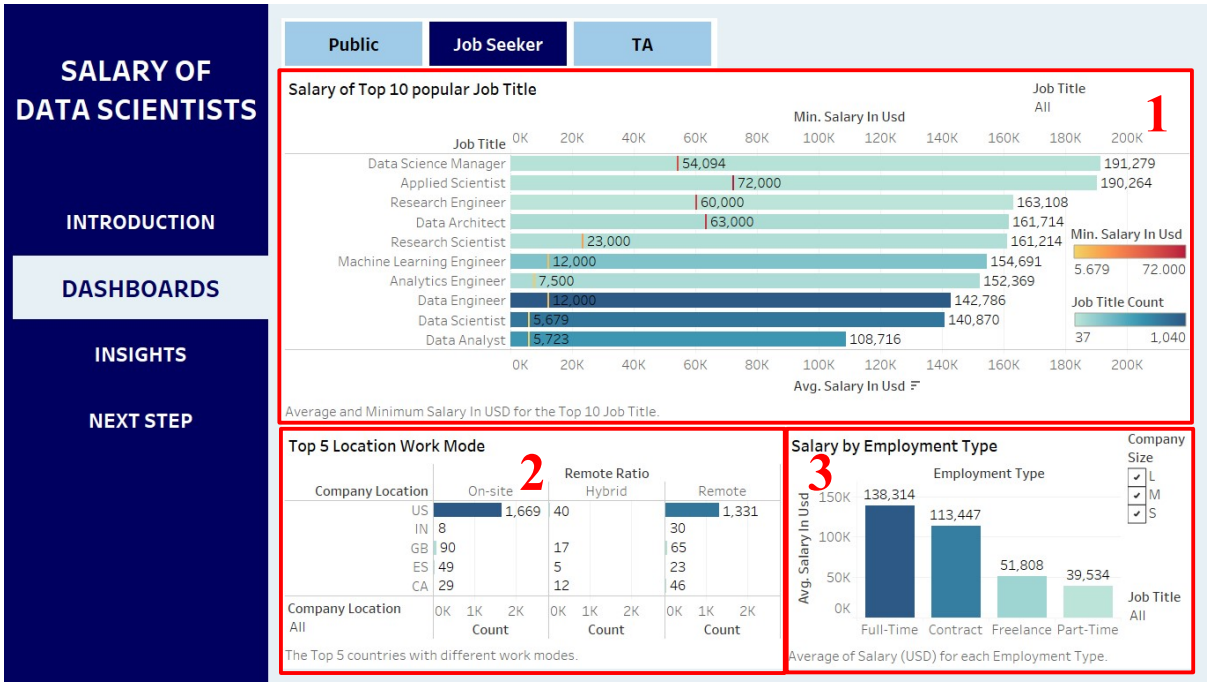
Under the bubble chart, there is a filter, “Count of Job Title” for viewers to narrow their focus for the most or less famous titles to get the details; and “Highlight Job Title” to highlight specific bubble to focus and make comparison with other bubbles.

d. <u>Avg Salary vs Experience Level bar chart</u>

The bars in chart are displayed horizontally for a better reading experience as the “Experience Level” category names are long. The average Salary in USD are shown at the end of each bar for reference without the need to hover. A “Job Title” dropdown is present at the top right for viewer to filter by the job title to see the average salary at different level of experience. If there is no record of experience of a particular job title, then the bar will not be shown in the chart.

In overall, the average salary patterns displayed as expected, where the Executive level receive the highest pay, followed by Senior Level, Mid-Level and lastly Entry Level.

2.2 Job Seeker



The Job Seeker dashboard is customized for the people who are surveying or looking for a job in the data science field. The dashboard covered the requirements from the employees' point of view, which included the minimum salary of the roles, the work modes, and the salary for different employment types. Each of the diagrams has at least 1 filter for the audience to interact with the dashboard, as they already know what the specific information is to look for at this stage.

a. Salary of Top 10 popular Job Title

The chart, by default, shows 10 most popular roles filtered by the occurrence of the job titles in dataset and sort by the Average salary in descending order, with minimum salary displayed in contrast gradient colour of red and orange Gantt bar for easy spotting. The minimum salary acts as the baseline for the audience to examine their own salary.

However, the chart can be modified by using the dropdown at the top right to display the information of specific job titles chosen.

In the chart, Applied Scientist has the second highest average salary of $190,264 and minimum salary of $72,000. Meanwhile, Data Architect's minimum salary of $63,000 is second highest despite the average ranked as the 4th in the list.

b. Top 5 Location Work Mode

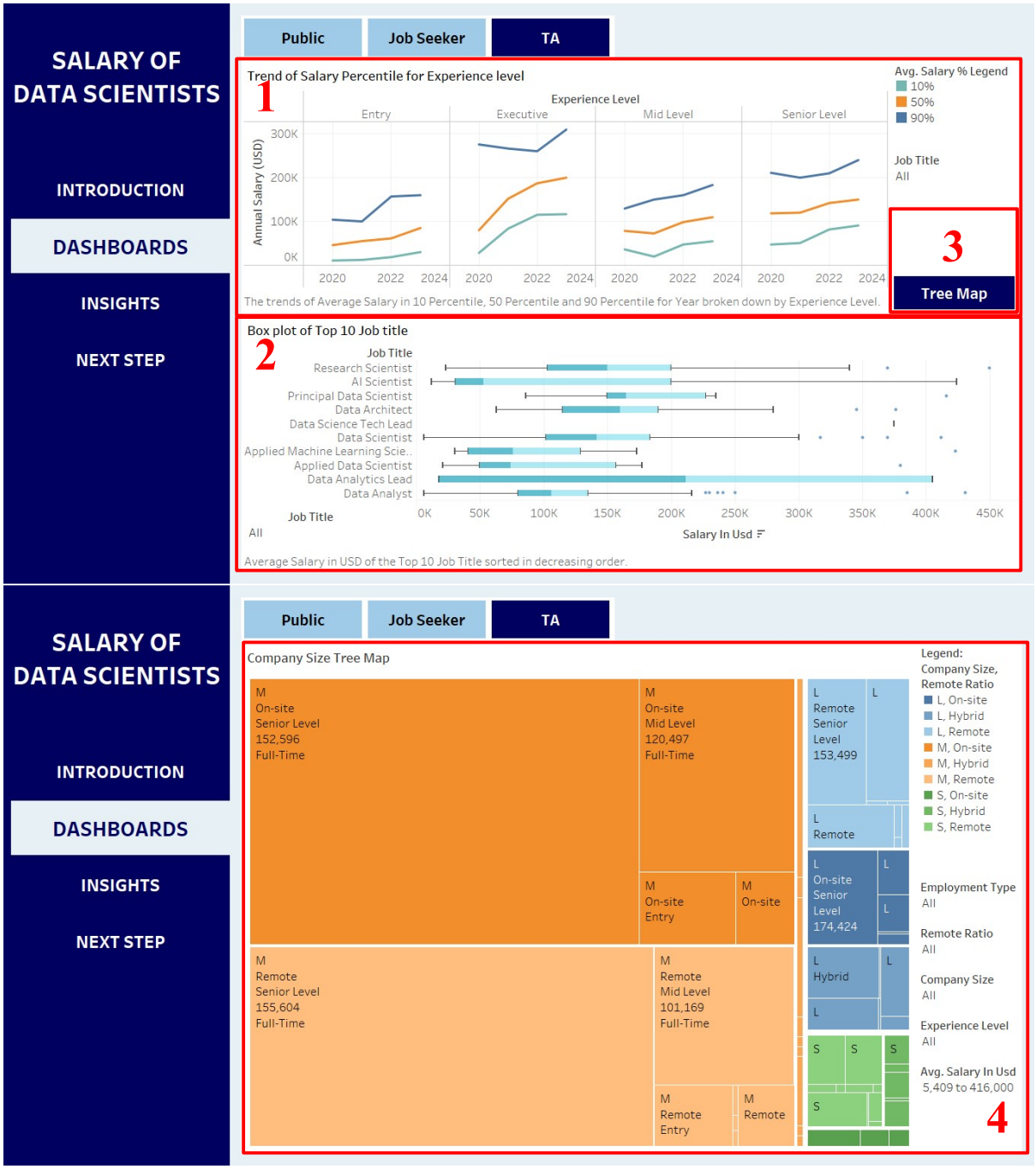
The default display of the Work Mode bar chart shows the 5 countries with the most records in different work modes. For the benefits of comparison, the chart can be altered to show the details of other countries by selecting them from the dropdown at bottom left.

From the chart, Hybrid is the lowest in number among the 3 work modes for the countries listed. This means that most of the work is done either on-site or remotely.

c. Salary by Employment Type

In overall, the full-timer salary is highest, followed by contractor, freelance and part-timer. Company Size and Job Title filters are present at the right side to display the data for filters of different combinations. This helps the job seekers in conducting their research and determine their worth.

2.3 Talent Acquisitions



The TA (Talent Acquisition) dashboard is tailored for the TA for extensive analysis and exploration. There are 2 parts and 3 diagrams in the dashboard: line graph, box plot, and tree map.

a. <u>Trend of Salary Percentile for Experience Level</u>

This is a split line graph of Average Salary versus Work Year, separated by Experience Level. The 3 lines in each Experience Level represent 10th percentile, median, and 90th percentile which can be helpful in setting up the salary standards and baselines. The

dataset contains salary data from 2020 to 2024, therefore comparison of recent market trends can be made for candidates with required experience. Legend that indicates the colours of each percentile and Job Title filter are present at the right side. In the graph above that includes all job titles, the lines do not overlap with one another, this allows a salary range to be set up between gaps of 10^{th} and 90^{th} percentile and be used to create salary guide and predict salary of potential candidates.

b. Box Plot of Top 10 Job Title

The box plot shows the salary in detail by displaying the lower whisker, lower quartile, median, upper quartile, upper whisker, and the outliers, these are helpful information in studying the salary market to hire for the specific roles related to data science. The plot displays 10 job titles with highest pay arranged in descending order by the maximum salary, with Research Scientist placed at the top and Data Analyst at last. There is also a dropdown at the bottom to filter for various roles needed.

c. Tree Map button

Navigate to tree map dashboard.

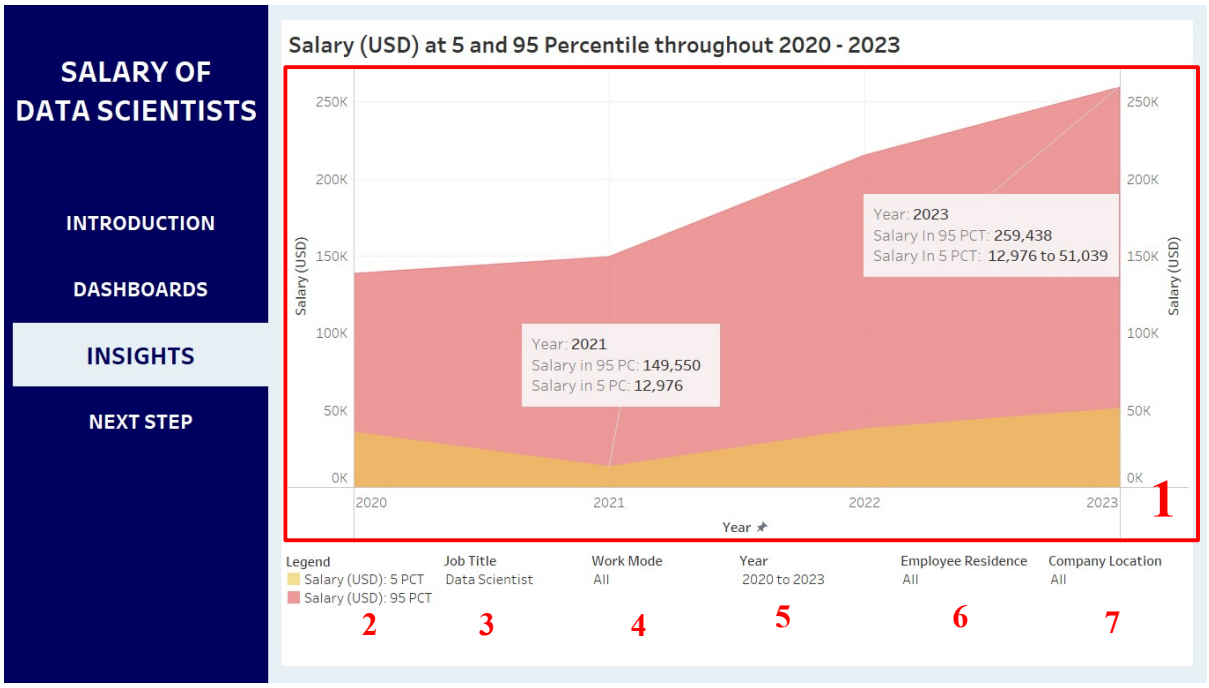
d. Tree Map

The Tree Map is a direct way to visualize the population of company size along with other variables, it is a highly space-efficient solution in presenting data, where each of the node (variable) is represented as a rectangle and would subdivide into smaller rectangle to represent the children's nodes. The benefits of tree map include its space-efficiency and high scalability which make it suitable for large hierarchical structure and practical for the expansion. The only down-side is the low visibility - difficulty to see the whole structure containing many levels (Lim Kian Long et al., 2017).

In the map, company of size M made up most of the dataset, followed by L and S. For M sized company, hybrid work mode took the least space while remote and on-site are almost equal.

To get a better understanding and clearer view of the map, there is a legend which acts as an annotation, and 5 filters for the audience to select specific details to display in the map.

Dashboard #3: Insights



The area chart shows the salary in USD at different years for 5 percentile and 95 percentiles. Salary at 95 percentile means the salary is 95% higher than the rest in the dataset, there is only 5% of records are having higher salary (Frost, 2019); while 5 percentile means the opposite.

In the chart,

a. Area Chart

The salary ranged from 12,976 USD at 5 percentiles in 2020, to 259,438 USD at 95 percentiles in 2023. In general, the salary shows an increase from 2020 to 2023. For 95 percentiles, the increment is slow in the beginning and started to speed up after 2021. Meanwhile for 5 percentiles, there is a record of a slight drop of salary in 2021 before increasing steadily.

b. Legend

Yellow for Salary in 5 percentiles, while red for 95 percentiles.

c. Filters (Job title, Work Mode, Year range, Employee Residence, Company Location)

Allow results of different combinations to be displayed for analysis. The diagram shows "Data Scientist" is being selected from the dropdown, therefore the area chart exhibits the corresponding information and result.

Dashboard #4: Next Step

From the dashboards, the main take away is to allow the audience to get a grasp of salaries by different dependent factors, but the details differ according to the audience groups, where TA would get the most information compared to the other two. The result of the study contributes to salary transparency and helps in setting the standards through the creation of customized salary guides.

Conclusion

Data Scientists roles have risen since 2010 and branched out to 93 job titles in the field. The salary of the data science related roles fascinates many people and therefore should not be a taboo of discussion when personal or community benefits are involved. The amount can be differ depending on assorted reasons, such as demographic, geographical, work mode, seniority level etc. By using the dashboards, users able to evaluate own monetary worth through the graphs and find confidence in the job opportunities in companies distribution map. Hence, the salary pattern should be studied from every aspect to encourage transparency, avoid underpayment, and set a standard range of amount based on conditions.

1965 words

Dashboard Presentation URL: https://youtu.be/WncWCGUBAXU

Reference

1. Shan, H. (2023). 2023 Data Scientists Salary. Kaggle.com. https://www.kaggle.com/datasets/henryshan/2023-data-scientists-salary
2. Lim Kian Long, Lim Chien Hui, Gim Yeong Fook, & Mohd, W. (2017). A Study on the Effectiveness of Tree-Maps as Tree Visualization Techniques. ResearchGate; Elsevier BV. https://www.researchgate.net/publication/322075128_A_Study_on_the_Effectiveness_of_Tree-Maps_as_Tree_Visualization_Techniques
3. Frost, J. (2019, April 10). Percentiles: Interpretations and Calculations. Statistics by Jim. https://statisticsbyjim.com/basics/percentiles/