

**MULTIMEDIA UNIVERSITY**

**FACULTY OF COMPUTING AND  
INFORMATICS**

**BACHELOR IN COMPUTER SCIENCE  
(HONS)**

**DATA VISUALIZATION - CDS6324**

**TRIMESTER II, SESSION 2024/2025**

**D3 Visualization of Data Science Job Trends in  
the Early 2020s**

**By:**

**Ooi Li Yoong (1211306826)**

### **Acknowledgment**

We are immensely thankful to our instructor, Ts. Dr. Noramiza Binti Hashim, whose insightful guidance has been pivotal throughout this project's journey.

## Table of Content

Acknowledgment .....	2
Table of Content .....	3
Application Overview .....	4
Dataset Description .....	5
Visualization Charts .....	6
Design Principles .....	8

## **Application Overview**

Our interactive visualization application provides information about data science jobs in the early 2020s. Data science is an emerging field with rewarding salary, thus attracting many working people. However, further context should be provided for them to understand more about this exciting field.

To support our analysis, we use a related dataset available at <https://www.kaggle.com/datasets/willianoliveiragibin/data-jobs-salaries>.

We selected a dark theme for our application to offer a modern user experience. We also enable tooltips, which display key information whenever applicable and are activated by mouse hovering actions.

In this small project, we aimed to answer only one question through our displayed visualizations: What are the top 10 data science jobs by salary increase from 2020 to 2023 inclusively? This question is important for viewers to identify the fast-growing data science jobs and the current market demands.

## **Dataset Description**

The utilized dataset includes the necessary attributes as follows:

1. Worker's working experience level
2. Worker's employment type
3. Worker's job title
4. Worker's working company size
5. Worker's working company location
6. Worker's residence
7. Worker's salary

The raw dataset is inspected and found to be relatively “clean” and “edible”. There were no missing data, duplicated rows, nor invalid values such as a negative salary value. Therefore, no data cleaning tasks were performed. All calculations needed to visualize relationships between variables such as computing mean or median values were performed using d3.js instead of directly changing the raw dataset. Relationships between attributes found are presented in a few visualization charts which will be explained in the subsequent section.

## Visualization Charts

Our application reveals some must-know insights about the data science field in the early 2020s. Through exploring the visualizations, the viewers understand a few valuable insights.

Our application consists of two main screens, namely “Introduction” and “Top 10 Data Science Jobs”. Users can navigate between screens easily by clicking on the corresponding buttons found at the top panel. In the “Introduction” screen, a boxplot visualizing the salary range of data science workers is presented. From the boxplot as shown in Figure 1, we can see that the lowest salary is \$15,000, while the highest record is \$450,000. The median salary is \$142,200. The calculated upper fence or upper boundary of the boxplot is \$307,250. Therefore, higher values are considered as outliers and presented in red dots. Users might not be familiar with a boxplot; thus, we provide a key message guiding them to aim at a higher salary visualized rightwards the boxplot.

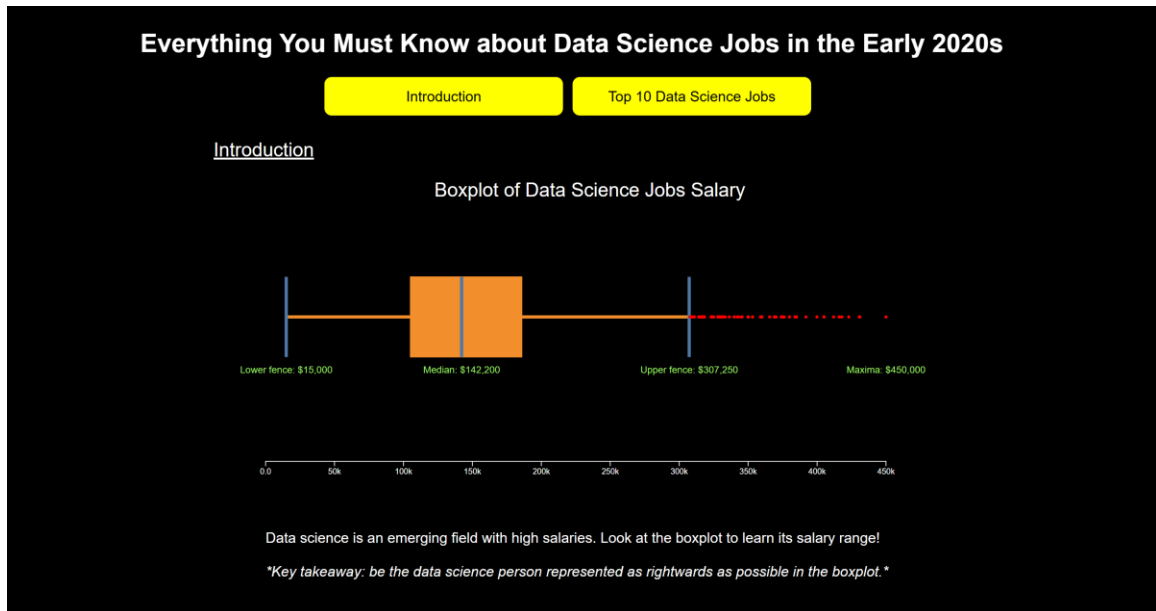


Figure 1: “Introduction” Screen

In the “Top 10 Data Science Jobs” screen as shown in Figure 2, a column chart visualizing salary increments of different data science jobs such as machine learning engineer is shown. However, only the top 10 jobs by salary increments from 2020 to 2023 are shown, due to the extensive amounts of categories available. To avoid information overload, we filter it out for easier and faster information processing, especially for users from a non-data-science related background. We apply a consistent color for all the “horizontal rectangles” to minimize the complexity of the chart. Users should be able to find key points rapidly, such as computer vision engineer has the highest salary increment from 2020 to 2023 inclusively (\$125,000).

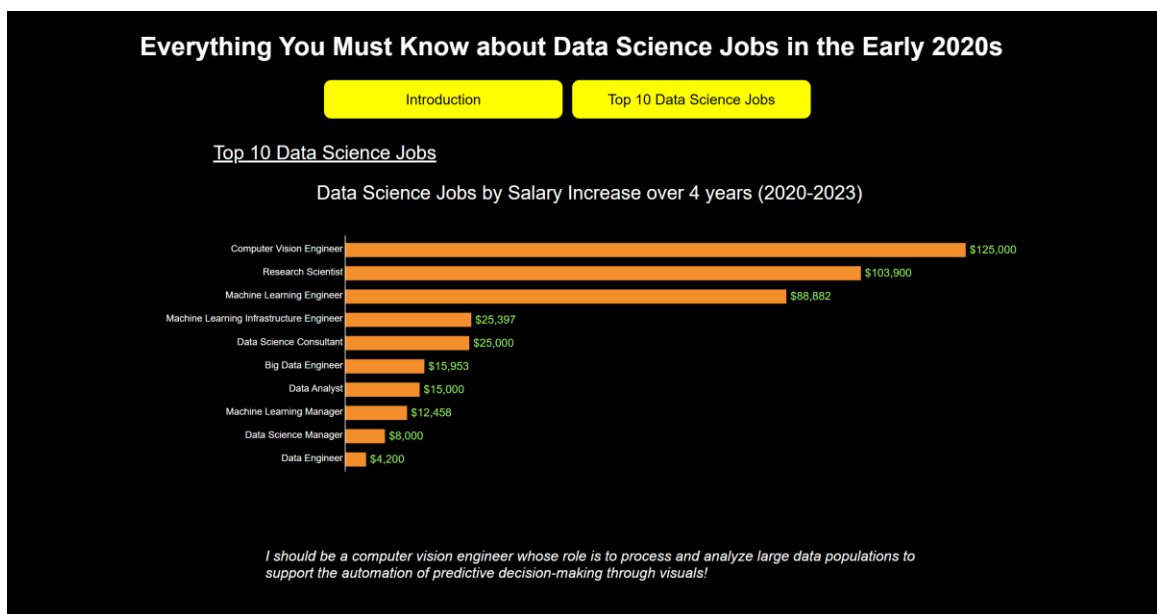


Figure 2: “Top 10 Data Science Jobs” Screen

### **Design Principles**

The design principles used in our application includes the “Clear”, “Clean”, “Concise”, and “Captivating” from the Shaffer **4C**’s of Data Visualization:

- **Clear:** The charts are easy to see, and the message is clearly shown to the audience.
- **Clean:** The charts are complete and use the right labels, axes, and chart types.
- **Concise:** The captions for each of the graphs are brief but comprehensive.
- **Captivating:** This storytelling is captivating because the discussed topic is up to date.