

# Data Visualization Process Book

**Navigating Careers in  
Data Science**

**By Data Girls**

Zaineb Tafasca, Imane Zaaroui, Tamara  
Antoun

# Introduction - Data

## Choosing Our Topic

In today's rapidly evolving data science job market, the demand for qualified professionals is rising, yet understanding what drives career transitions, salary trends and retention remains complex. As data science students ourselves, we often wonder: What factors influence whether data scientists stay in their roles or seek new opportunities? And how does salary vary across roles, locations and experience levels in this field? We wanted to create a project that not only answers these questions but also provides actionable insights for students, graduates, recruiters, and professionals navigating the job market. Thus, we set out to design an interactive data visualization project that examines the drivers of job-seeking behavior and salary trends in the data science field.

## Finding a Dataset

With this idea in mind, we began searching for suitable datasets. Our initial challenge was finding data that connects candidate profiles (like education, experience and company type) with job-seeking behavior, as well as detailed salary information across roles and locations. After exploring various sources, we identified two complementary datasets on Kaggle:

1. **Data Science Job Data Set:** This dataset provides rich information on job candidates, including demographics, education levels, years of experience, training hours and a target variable indicating whether they are actively seeking a new job.
2. **Jobs and Salaries in Data Science:** This dataset focuses on salaries in the data science field including job titles, locations, and salary ranges. It offered an excellent foundation for analyzing salary trends across roles, countries and experience levels.

# EDA and Website

## EDA

Before designing any visuals or models, we conducted an in-depth EDA to understand the structure and patterns in both datasets. We began by cleaning the data, checking for missing values, and selecting relevant features such as gender, company size, training hours, and job title. We explored distributions and relationships to better understand who is seeking new opportunities and how salary varies across roles and regions.

## Website and Visualizations from Scratch

We built our website and visualizations entirely from scratch to ensure full control over the design and data exploration process. We chose to incorporate a bubble chart because it highlights the relative size and importance of job categories at a glance and provides an intuitive entry point for deeper exploration. Additionally, we opted for simple yet clear visualizations such as histograms and line charts. These straightforward designs help users quickly interpret the data without distraction. For the visual style, we opted for a bright overall theme complemented by warm colors like orange to create a dynamic and inviting atmosphere while maintaining a professional look that enhances data interpretation.

## Tools

We built our website and visualizations entirely from scratch, using HTML, CSS, and JavaScript libraries like D3.js and Chart.js



# Website Structure

We designed a three-page website, each page focusing on a different aspect of our analysis:

- Home Page: Introduces the project, summarizes our goals, and offers an overview of career paths in data science.
- Job Seeking Page: Presents visualizations exploring what factors influence whether data science professionals are actively seeking a new job.
- Salary Page: Provides insights into salary trends across roles, experience levels, and countries.

## Implementation - Home

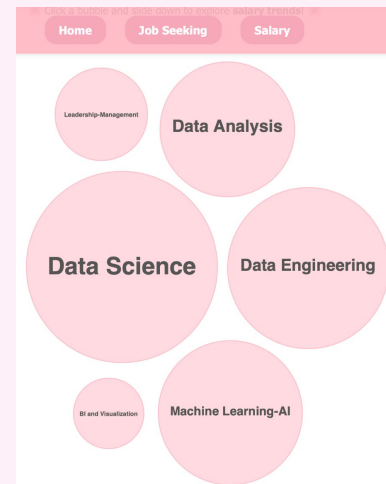
### Welcome to Data Girls Home Page !

The homepage serves as an introduction to our project featuring a clean and simple design. A relevant image adds a visual element to immediately engage visitors. Navigation links at the top of the page make it easy to move to the Job Seeking and Salary pages and a consistent color scheme ties everything together. At the bottom of the page, a contact bar highlights the team members and their email addresses to provide a direct way for users to reach out or learn more about the project.



# Implementation - Salary Page

Our Salary Page began as a simple static bubble chart created in [Draw.io](#) during EDA to explore salary trends across data science roles. We later transformed this concept into an interactive web visualization using Chart.js, where users can click on bubbles to reveal detailed salary distributions, trends over time, and experience-based salary variations.



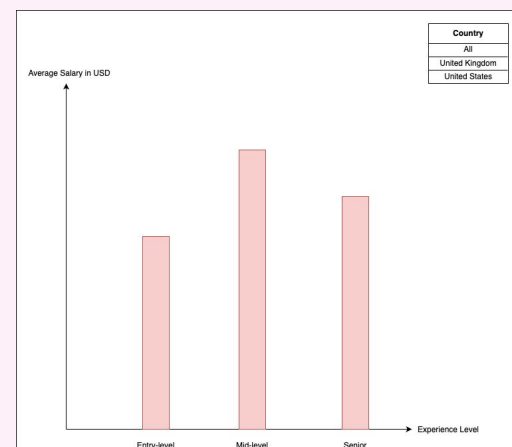
## From bubbles ... to histograms and graphs that reveals salaries, experience level ...

When a user clicks on a bubble, the page reveals detailed visualizations—such as a salary distribution histogram, a line chart of salary evolution over time, and a salary vs. experience level plot, providing a deeper dive into how compensation varies by role, experience, and location.

We expanded the visualizations to include as well:

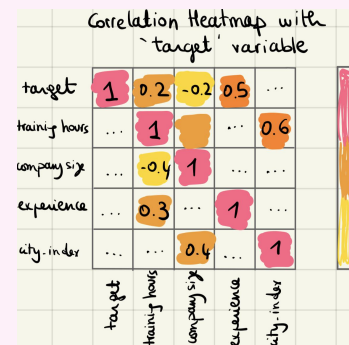
- A line chart showing salary evolution over time.
- And a bar chart comparing salary by experience level.

These changes reflect a shift from simple static diagrams to a multi-layered, interactive experience that encourages exploration and storytelling. The final design allows users to start with a broad view (bubbles) and progressively dive deeper into specific salary trends (histograms, line charts, bar charts).



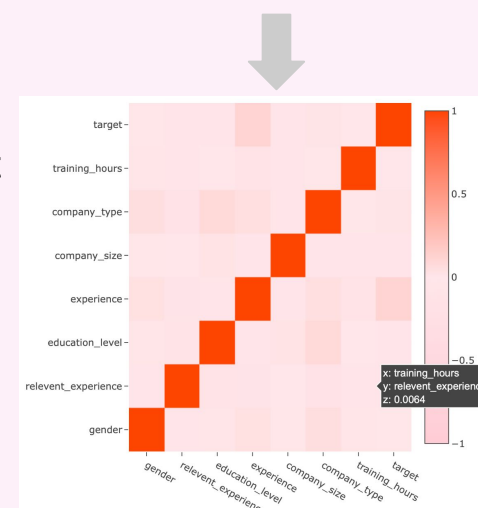
# Implementation - Job Seeking Page

In Milestone 2, our sketches for the Job Seeking Page focused on simple, static visualizations: bar charts showing the number of job seekers vs. non-seekers by experience level, a line chart for average seeking probability by training hours, a basic correlation matrix and an early concept for a job-seeking prediction tool.



## Correlation matrix as example:

We started with a simple sketch drawn on an iPad during Milestone 2, which mapped out how key features might relate to the target (job-seeking status). In the final interactive heatmap on the Job Seeking Page, we took this static idea much further. Now, users can hover over or click on any square to reveal the specific pair of features being compared and the exact correlation coefficient.



## From Job seeking analysis to Building the Prediction Model

After analyzing and displaying various graphs, such as average seeking probability, job seeking by company type, job seeking by experience level and the correlation matrix, we conclude the page with the Job Predictor, that started as a simple sketch in [Draw.io](https://draw.io) and became a real interactive tool built with HTML, CSS, and JavaScript. We trained a machine learning model in Python (scikit-learn) and deployed it as an API using Flask and Render.

Gender:

Training Hours:

Company Size:

Enter

Prediction Result

40%

Not Seeking for a Job Seeking for a Job

This setup lets users input their profile and receive predictions directly in the browser.

Try the Job Predictor

Gender: Male

Training Hours: 150 hours

Company Size: Small (S)

Predict

# Challenges

## Issues we faced ...

Here's a small list of the challenges we encountered during the project:

- Frontend development was a learning curve: At the start, none of us had extensive experience with frontend technologies used.
- Connecting data insights to visual storytelling: While we had strong insights from our analysis, translating these insights into clear, accessible visualizations was a challenge.
- Time management: It was sometimes challenging to divide tasks evenly. Coordinating frontend work, data analysis, and design required patience and collaboration.

## ... And how we solved them

- Frontend development: We tackled our technical gaps by learning as we went, experimenting with Chart.js and D3.js through tutorials, documentation and trial-errors.
- Visual storytelling: Prioritized intuitive charts through complex
- Time management: we divided the project into clear sections (website structure, virtualization, process book, ...) and supported each other's progress.

# Peer Assessment

Throughout this project, we worked collaboratively, dividing tasks based on our strengths and interests:

## **Zaineb Tafasca:**

- Exploratory data analysis for the Job Seeking dataset in M1.
- Sketched early concepts for visualizations on the Job Seeking Page in M2.
- In M3, focused on the technical implementation of the website, developed interactive visualizations, as well as writing the process book.

## **Imane Zaaraoui:**

- Conducted the exploratory analysis for the Salary dataset in M1.
- Designed early sketches for the Home and Salary pages in M2.
- In M3, worked on the technical implementation contributing significantly on integrating the desired interactive visualizations and ensuring site's functionality, as well as writing the process book

## **Tamara Antoun:**

- Explored the Salary dataset and wrote the README file in M1.
- Led writing the report for M2.
- In M3, focused on implementing the job prediction model and integrating it into the website; added some final touches to the website; created the screencast video; contributed to the process book.

