Data Visualization
Project Proposal

1. Basic Info

Project Title: Job Market Explorer: Interactive Visualization of Job Opportunities in the USA

Group Members:
- Vidhi Agarwal, Email: vidhi.agarwal@utah.edu, UID: u1448738
- Pavitra Mehra, Email: pavitra.mehra@utah.edu, UID: u1472969

Project Repository:
https://github.com/dataviscourse2024/group-project-jobquest-visualization

2. Background and Motivation

The tech industry in the USA is booming, with constant demand for skilled workers across various roles such as software engineers, data scientists, and cybersecurity experts. However, understanding which specific skills are in demand, which regions offer the most job opportunities, and what salary trends look like can be overwhelming due to the sheer volume of job postings and industry data available. This project aims to visualize tech job opportunities in an interactive way to make this data more accessible.

We chose this project because of our personal and academic interest in data visualization and the tech industry. As aspiring software engineers, understanding the dynamics of the tech job market will not only benefit us but also provide insights to other job seekers, recruiters, and policymakers. Additionally, visualizing this data will help us hone our skills in working with real-world data, modern web frameworks, and interactive visualization techniques.

3. Project Objectives

This project seeks to answer the following primary questions through interactive visualization:

- What tech job roles are most in demand across the USA?

- How do job opportunities vary across different states and cities?
- What are the salary ranges for various tech roles across different regions?
- What trends exist in job postings over time, and how do they change based on skills?

Benefits:

- Provide job seekers with easy-to-navigate data on job availability and salary trends in the tech sector.
- Enable recruiters and employers to identify areas with a high demand for tech talent.
- Offer a visual comparison of demand across skills, roles, and locations.

4. Data

The data will include information such as job titles, locations, salaries, company details, and skill requirements. We are currently aiming for data between 2023-2024. We will collect data from various sources, including:

- **LinkedIn**: We will gather job posting data related to tech roles across different cities and states in the U.S. LinkedIn's API provides access to job listings, though usage may require approval. In cases where direct API access is not available, we will employ web scraping techniques to extract relevant job data from publicly available job postings.
- **Indeed**: We will either use the Indeed API or web scraping to collect job listings specific to tech roles. The data will cover job titles, locations, company names, and job descriptions.
- **Glassdoor**:For salary information and job trend data, we will use both the Glassdoor API (if accessible) and web scraping. Glassdoor provides valuable insights into company reviews, salary distributions, and job satisfaction ratings. We will scrape salary data related to tech roles to analyze pay differences across regions and job titles.
- **Bureau of Labor Statistics (BLS)**: This website provides employment statistics, job outlook, and wage data across various industries, including tech. We will download and process public datasets from their Occupational Employment and Wage Statistics (OEWS) section.
- **Kaggle Datasets:** We discovered a dataset on Kaggle scraped from LinkedIn for the years 2023-2024, containing over 124,000 job postings. We can utilize this dataset to complement the real-time data collected through APIs or web scraping. By leveraging this dataset, we can fill in any gaps in data from current job postings.

Data Sources:

- LinkedIn Job Search
- Indeed Job Search API
- Glassdoor API
- Bureau of Labor Statistics (BLS)
- Kaggle Dataset

5. Data Processing

We expect significant data cleanup and transformation steps, including:

- **Deduplication**: Removing duplicate job listings across different sources.
- **Normalization**: Converting salary data into a consistent format (e.g., annual salary).
- **Categorization**: Grouping job roles into broad categories (e.g., software development, data science).
- **Derived Quantities**:
    - Total number of job postings per role.
    - Average salary per role and per region.
    - Job trends over time (based on job posting dates).

Implementation:

- We will use **Python** and **Pandas** for cleaning and processing data.
- **APIs** or **web scraping** tools like **BeautifulSoup** and **Scrapy** will be used for data collection.

6. Visualization Design

Design Goals:

We aim to create an intuitive, interactive visualization that allows users to:

- Explore job opportunities geographically (e.g., job density on a US map).
- Compare salary ranges and job demand for different roles.
- View trends in job postings over time.

Prototype Designs:

- **Prototype 1: Choropleth Map with Pop-ups**
    - **Description**: This design features a choropleth map of the U.S., color-coded by the density of tech jobs in each state. The darker the

shade, the more job opportunities in that area. Users can hover over states or cities to view a pop-up with more detailed information (e.g., job titles, salary ranges, number of job openings). There will also be filters to narrow down the data by job type, role, or salary.
    - **Justification**: This design leverages geographic encoding to visually represent data density, making it intuitive for users to identify high-concentration areas for tech jobs. It's particularly effective for users interested in location-based job opportunities.

- **Prototype 2: Bar Chart with Dynamic Filters**
    - **Description**: This design features multiple bar charts comparing the demand for various tech roles across different regions. Users can filter by city, state, job role, or salary. Bar charts will dynamically update based on user selection, showing side-by-side comparisons of job openings and salary ranges.
    - **Justification**: Bar charts provide a straightforward way to compare multiple categories simultaneously. This design is ideal for users looking to compare specific roles or locations, enabling a more detailed analysis of job demand and salaries.

- **Prototype 3: Line Graph**
    - **Description**: A line graph will be used to show trends over time, such as the number of job postings or salary changes for specific tech roles or regions. The x-axis will represent time (e.g., months or years), and the y-axis will display metrics like job counts or average salaries. Multiple lines can be used to compare trends across different job categories (e.g., software engineering, data science) or geographic locations (e.g., California vs. New York).
    - **Justification**: This visualization will help users easily identify patterns, seasonal trends, and the growth or decline of opportunities over time in the tech job market.
- **Prototype 4: Pie Chart for Job Share**
    - **Description**: A pie chart that displays the distribution of different job types within a selected region or the entire dataset. Each slice of the pie represents a job type, with the size of each slice proportional to the number of job openings or the share of each job type relative to the total.
    - **Justification**:The pie chart allows users to quickly grasp the proportion of different job types, making it easier to identify which roles dominate the job market within a given context.

Final Design

The final design will combine the best elements from all three prototypes to create a rich, interactive visualization:

- **Choropleth Map as the Main View**: The map will serve as the central element, allowing users to visually explore tech job opportunities across the U.S. by state and city. It will use color encoding to represent job density and offer hover-over pop-ups with detailed information on job roles, salaries, and openings.

- **Bar Charts for Comparison**: Beneath the map, bar charts will allow users to compare job roles and salaries across various regions. These charts will dynamically update based on the selected location on the map, providing users with a clear comparison of different tech roles and their corresponding demand.

- **Line Graph for Trends Over Time**: In a separate panel or section, a line graph will show trends in job postings or salaries over time. Users can filter the graph by job type or region to track how job demand or salaries have changed over months or years.

- **Pie Chart for proportion of jobs in different sectors:** For overall job opportunities across the United States we will have a pie chart which will show the job proportions for different types of job opportunities in the technology domain.

- **Filters and Controls:** Dropdowns for filtering by job type, location, salary, experience level, and required skills.

Justification

1. Purpose and Audience

The final design aims to address the needs of users looking for detailed, actionable insights into the tech job market. By combining multiple visualizations, it offers a holistic view that caters to different aspects of job searching and career planning.

2. Data Types and Relationships

The final design uses geographic encoding (choropleth map), categorical comparisons (bar charts), time series (line graphs), and relational data (bubble charts) to capture the different dimensions of job market data. This ensures comprehensive coverage of the data's various relationships.

3. Visual Encoding Choices

- **Choropleth Map with Pop-ups:** Uses color encoding to represent job density, with interactive pop-ups providing detailed information.
    - **Justification:** Geographic encoding is ideal for visualizing location-based data, allowing users to quickly identify high-density areas. Hover-over pop-ups add context without cluttering the map.
- **Bar Charts with Dynamic Filters:** Displays comparative data on job roles and salaries.
    - **Justification:** Bar charts effectively compare different categories side by side. Dynamic filters enable users to customize their view based on their preferences or needs.
- **Line Graph for Trends:** Shows changes over time with multiple lines for comparison.
    - **Justification:** Line graphs are excellent for illustrating trends and changes over periods, allowing users to track patterns and growth in job postings or salaries.

4. Clarity and Simplicity

Each component is designed to be easily understandable and intuitive. The separation of different data aspects into distinct visualizations helps avoid clutter and ensures that users can focus on specific types of information.

5. Interactivity and Context

- The choropleth map allows users to explore different regions and view detailed information via pop-ups.
- Bar charts and line graphs have dynamic filters for user customization.
- The bubble chart provides interactive elements for a detailed analysis of job roles and salaries.
- Clear labels, tooltips, and legends will be built to ensure that users understand the data being presented.
- Filters and controls enable users to tailor the visualization to their specific needs and interests.

7. Must-Have Features

- Ability to filter job listings by role, skill, and location.
- Clean and well-processed data that accurately represents the tech job market.
- Interactive map showing job density and salary distribution across states.
- Responsive web design for different screen sizes.

8. Optional Features

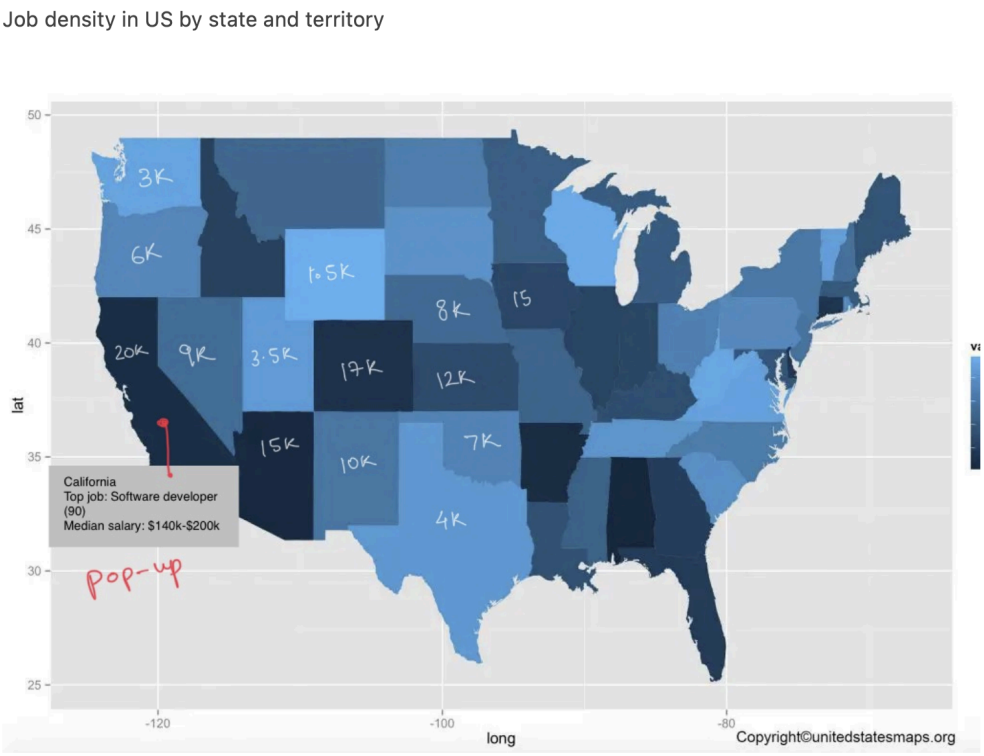
- Predictive analytics: Displaying future job demand based on current trends.
- Skill matching: Recommending tech roles based on user-entered skills.
- Incorporation of remote job opportunities.
- Real-time updates from job APIs like LinkedIn or GitHub.
- A feature that shows the most in-demand skills for each job role by region.
- Salary comparisons over time, showing trends in tech job compensation.
- Filters for job title, location, and salary.

9. Project Schedule

Vidhi and Pavitra both will be working simultaneously on each task every week.

Week	Tasks
Week 1/Week 2	Finalize data sources and begin data collection using APIs/web scraping.
Week 3	Clean and preprocess the collected data, removing duplicates and categorizing roles. Normalizing salaries and other data types
Week 4	Design wireframes for the visualizations and finalize the UI/UX approach.
Week 5	start building the frontend using javascript and maybe React.js.
Week 6	Create visualizations for the important information to provided like bar plots/ line graph
Week 7/Week 8	Integrate visualizations using D3.js and create interactive features.
Week 9	Complete the map-based visualization with interactive filters.
Week 10	Add Features like toggling for different types of visualization. Create Process book and Readme file
Week 11	Make final adjustments based on user feedback and update the process book
Week 12	Final presentation preparation and video.

Sketches:

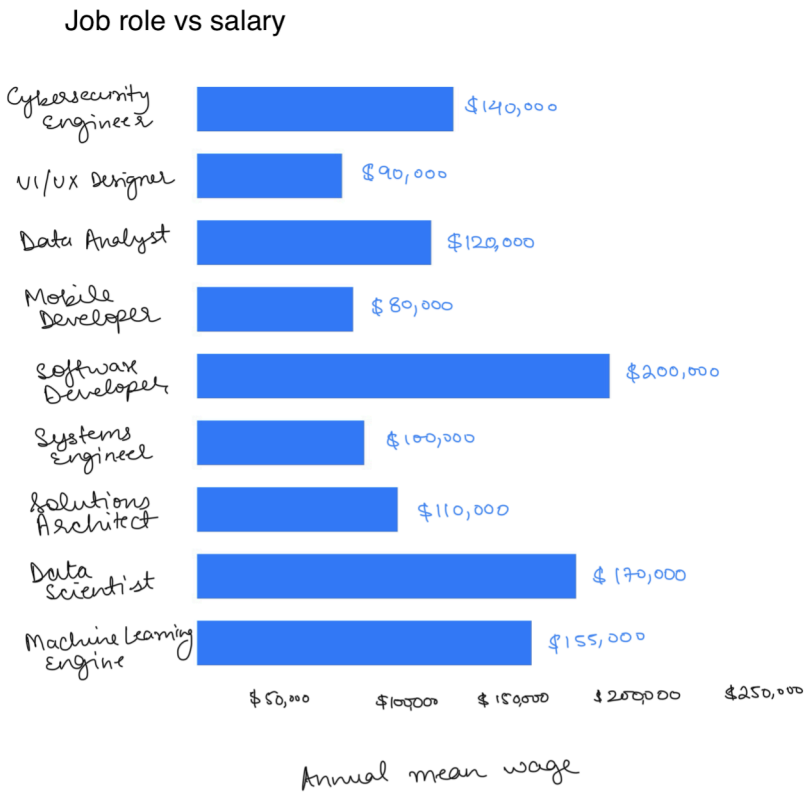


This sketch represents a **choropleth map of the United States**, where states are color-coded to indicate job density. Darker blue shades represent states with more job opportunities, while lighter blue indicates fewer job openings.

In this example:

- California shows the highest density of job opportunities, particularly for **Software Developer** roles, with a median salary ranging from **$140k to $200k**.
- States like Texas, New York, and Illinois also show significant job availability in various tech roles.
- A pop-up window provides additional information for each state, displaying the top job, number of openings, and salary range when users hover over a specific region.

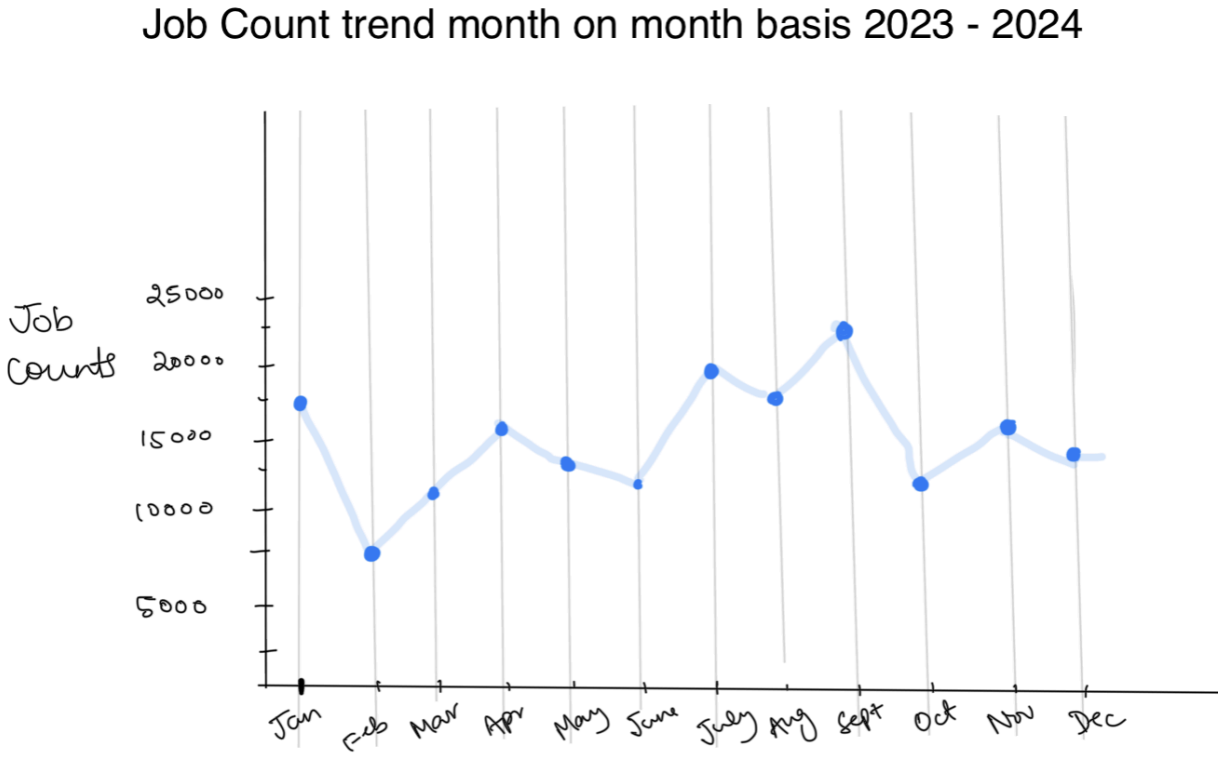
The map is an interactive tool that helps users explore job opportunities by state, with key details displayed in hover-over pop-ups.



This sketch is for a bar chart for **Job Role vs. Salary Comparison Chart**. This chart will be designed to provide users with a clear comparison of the average annual salaries across different tech job roles. The visualization will be created using a **horizontal bar chart**, where each bar represents a specific job role, and the length of the bar corresponds to the average salary for that role. This design allows users to visually compare salaries across a wide range of job roles in a simple, straightforward manner.

Explanation of Design

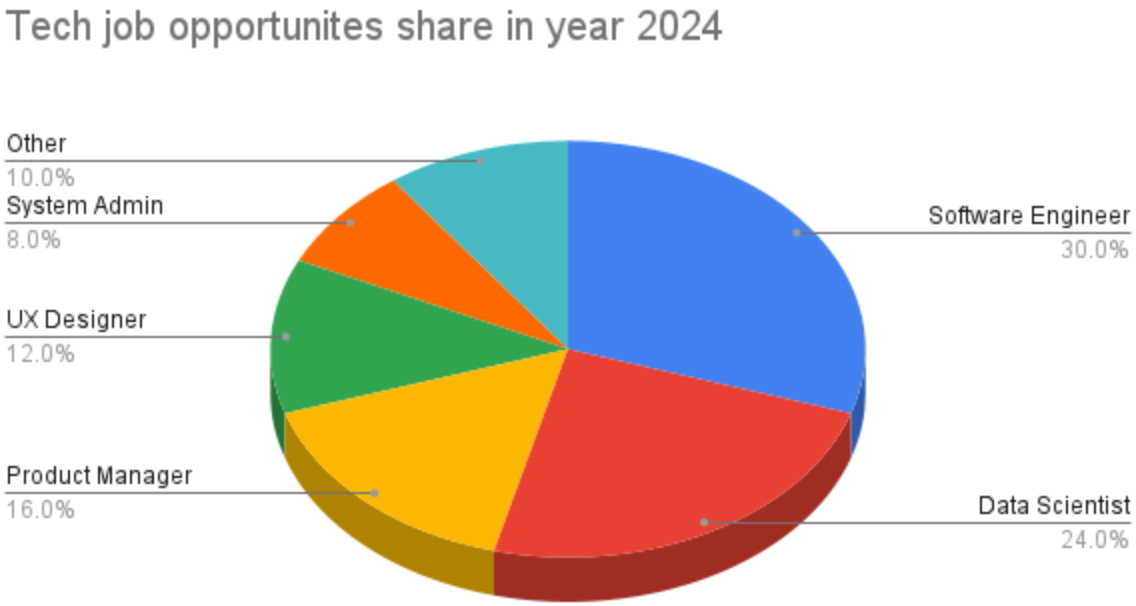
- **X-Axis**: The x-axis will represent the salary range, from a lower to a higher range based on the data. This will allow users to quickly compare salaries across multiple roles.
- **Bars (Job Roles)**: Each bar in the chart will represent a different tech job role, such as Software Engineer, Data Scientist, UI/UX Designer, etc. The length of the bar will indicate the average salary for that particular role, making it easy to visualize which roles command higher or lower wages.
- **Labels**: Each bar will be labeled with the exact salary value to provide more precise information. The chart will also allow for hover-over tooltips in the interactive version to show additional details such as the number of job openings and in-demand skills for that role.



This sketch represents a **line graph** illustrating the **job count trend** on a **month-to-month basis** for the years **2023 to 2024**.

- The **x-axis** shows the months from January to December.
- The **y-axis** represents the **number of job openings** ranging from 0 to 25,000.
- Each point on the line corresponds to the job count for that particular month, and the connecting lines highlight the overall trend in job postings throughout the year.

The graph reveals how the job count fluctuates, increasing and decreasing over time. Peaks and troughs are clearly visible, showing when job postings were higher (e.g., in months like May and November) and when they were lower (e.g., in February and June). This visualization can be used to analyze seasonal trends in job availability over the year.



This chart drawn represents a **pie chart** illustrating the **distribution of tech job opportunities** across various roles for the year **2024**. Each slice of the pie corresponds to a different tech job role, with the size of each slice indicating the relative proportion of that role in the overall job market.

The chart highlights key tech roles such as **Software Engineers**, **Data Scientists**, **Product Managers**, and **UX Designers**, along with other categories like **System Admins** and an "Other" section for less common roles. This visualization provides an overview of how tech job opportunities are spread across different roles, helping users understand which roles dominate the job market and how opportunities are distributed. We will be including multiple different job opportunities based on the data we get from the job postings and this sketch shows the general view for the pie chart.