Project Report For CS661: BIG DATA VISUAL ANALYTICS
2022-2023 Semester II

Project Title: **CAREER COMPASS**

Team members: Avinash Prasad; Bugada Ashritha; Jhaansi Reddy; Mansi Singh; Nikhil Verma; Prattipati Mokshagna; Priyal Agrawal; Vaishnavi; Vartika
Member emails: avinashp20; bugada20; jhaansir20; mansisingh21; nikhilv20; mokshagna21; priyalag20; vaishnavi21; vartikag20

IIT Kanpur

1 Introduction

Choosing a career path is undoubtedly a monumental decision, as it not only shapes one's professional journey but also deeply influences personal fulfillment, financial stability, and overall well-being. Navigating the complexities of today's job market is a formidable task, characterized by uncertainty and intricacy. Success in this endeavor hinges on a combination of self-awareness, strategic planning, and adaptability, ensuring that career choices resonate with personal passions, skills, and the evolving demands of the professional landscape.

In the digital era, platforms like LinkedIn provide a wealth of data on job opportunities, skill requirements, and company cultures. This abundance of information, augmented by additional insights such as company ratings, average salaries, and geographical factors, empowers individuals to make well-informed career decisions. By leveraging these resources, individuals can align their aspirations with market trends, paving the way for fulfilling and rewarding career paths that align with their goals and offer ample opportunities for growth and development.

Indeed, harnessing this vast amount of data for decision-making can be daunting. Addressing this challenge necessitates the creation of Career Compass, a data-driven visualization platform. Such a platform would facilitate users in navigating the overwhelming abundance of available information. Through this solution, users could effectively filter sectors based on required skills, scrutinize industry growth across different periods, and evaluate various conditions such as employment rates and the proportion of educated populations.

Furthermore, users should have access to explore the attributes that companies and industries prioritize, including ratings of top-performing companies within their desired sectors. This comprehensive analysis should offer users tailored insights, enabling them to confidently make informed decisions as they navigate their career paths.Moreover, companies could also leverage insights from this data to expand operations, adjust their company structure, or refine their talent acquisition strategies in alignment with evolving market trends.

2 Tasks

2.1 Data Sourcing

We utilized datasets about jobs from Kaggle for our data. The following are the datasets that we exploited for data. Refer References

- 1.3M Linkedin Jobs Skills (2024)
  LinkedIn, a prevalent professional networking platform, hosts millions of job postings. Compiled from LinkedIn in 2024, this dataset comprises 1.3 million job listings. It serves as a valuable resource for conducting research tasks such as analyzing the job market, mapping skills, developing job recommendation systems, and more.

- Top world's companies
  The dataset comprises data from Ambition Box, offering insights into 10,000 distinct companies. Ambition Box is a platform renowned for providing comprehensive information on company reviews, ratings, salaries, interviews, and other related details. This dataset can be used to analyze company's ratings and help users to gain information based on the reviews.

- LinkedIn Digital Data
  The datasets featured here are the foundational datasets powering the visualizations found at `linkedindata.worldbank.org`. These datasets encompass four key metric categories: 1) Industry Employment Shifts, 2) Talent Migration, 3) Industry Skills Needs, and 4) Skill Penetration. These datasets consist of annual time series data dating back to 2015 till 2019. This dataset will prove to be helpful in analyzing global employment shifts, industry growth and industries with their skill needs.

- Jobs data world wide
  The Global Jobs dataset encompasses nearly all countries (243 in total) from 1990 to 2016. It offers a comprehensive resource for assessing various aspects that influence career choices and job distribution within each country, including GDP, unemployment rates, employment figures, gender participation, and more.

- LinkedIn Tech Jobs
  This dataset comprises over 500 job listings extracted from the job section of LinkedIn, focusing exclusively on tech positions within India. The dataset includes various columns providing essential information about each job, such as the location, job designation, company name, industry, employee count, LinkedIn followers, total number of applicants, and required skills. These listings offer insights into the tech job market in India, allowing analysis of factors such as job distribution, skill requirements, and company profiles.

2.2 Data preparation

During the data preparation stage, we undertook tasks such as data cleaning, handling missing values and processing of strings to extract meaningful information. Additionally, we employed methods such as averaging and aggregation to streamline the data. Conducting exploratory data analysis (EDA) among ourselves allowed us to gain further insights into the dataset.

2.3 Data visualization

In our visualization and analysis process, we employed a diverse range of plots and graphs to present and examine the data comprehensively. This included pie charts to illustrate distribution patterns, bar graphs to visualize values effectively, and line graphs to depict timelines accurately. Additionally, we utilized chloropleth maps to visualize geographical data based on metrics such as value density, and scatter plots to showcase relationships between variables. Furthermore, we utilized network-based graphs to reveal intricate patterns and connections within the data.

2.4 Website layout

We applied CSS styling to ensure the visualizations are aesthetically pleasing and user-friendly. Furthermore, we enhanced our visualizations by incorporating interactive features, such as dropdown menus and search tabs, which allow users to filter and search for specific information, providing them with freedom and flexibility in exploring the data. Additionally, we utilized select options for toggles to further enhance the interactive experience for users.

3 Proposed Solution

To execute our tasks efficiently, we utilized Jupyter Notebooks, allowing for seamless integration and ease of collaboration. Leveraging Git and GitHub facilitated effective teamwork and version control, ensuring smooth collaboration throughout the project. For generating visualizations, we employed Python's Plotly library, enabling us to create dynamic and interactive plots with ease. We utilized the Dash framework to develop a website for our dashboard, facilitating the integration of visual plots.
The following is detailed descriptions for intended information dissemination through visualization.

3.1 Jobs across various locations

1. We utilized pie charts to illustrate the distribution of job postings across different cities in four countries, as obtained from dataset (1). This offers insights for job seekers and recruiters on geographical employment distribution.

2. Using data from dataset (5), we displayed the number of applicants for each company in India across various states on a map. The visualization included color coding on a color scale to represent the numerical values enhancing data interpretation. This aids recruiters in understanding candidate engagement dynamics.

3.2 Your favourite companies and jobs

1. We've generated a horizontal bar chart showcasing the top 20 job positions and their availability within a chosen company in a specific country using dataset (1). This tool offers users a streamlined way to explore job opportunities, providing insight into the current openings and the roles in demand within their preferred company and location.

2. We have made a bar graph which gives a comprehensive overview of top 20 companies within the chosen sector using dataset (2).Through organizing companies in descending order of average salary and integrating essential considerations like job security and company culture etc., our visualization empowers users to make informed decisions aligned with their career aspirations. Alongside, there is a scatterplot of those filtered out companies on a number of interviews v/s number of job postings. By visualizing these two metrics, users can discern patterns and trends related to company recruitment practices and candidate engagement.

3.3 Top job positions/titles searched

1. We have created a bar graph using dataset (1) which provides a view of the number of searched positions and their corresponding job postings within a specified company. This tool simplifies the process of assessing job opportunities by highlighting which companies have more openings and potential career prospects and make directed job search efforts.

2. Using the same dataset, we've generated a bar graph that presents the count of searched positions alongside their associated job postings within a designated location. It helps users understand the demand and supply dynamics of the job market in that location, allowing them to make informed decisions about their job search or recruitment strategies.

3.4 Country Wise Evaluation

1. A line chart depicting the evolution of various aspects of a country's condition over time was made using dataset (4). By focusing on important factors like the percentage of women with advanced education, access to electricity, and employment in services etc. , it provides valuable insights into each country's development path.

2. Another line chart, generated using dataset (3) on each country shows the growth rate of industry sectors over the years. By identifying sectors experiencing rapid growth, job seekers can target industries with ample opportunities for employment and career advancement.

3. We employed a scatter plot on a chloropleth world map using dataset (3) to show migration trends in countries for an industry or a skill. This visualization can provide insights into global labor market dynamics. Understanding migration trends can inform decisions regarding relocation or international job opportunities

3.5 Industry and the skills

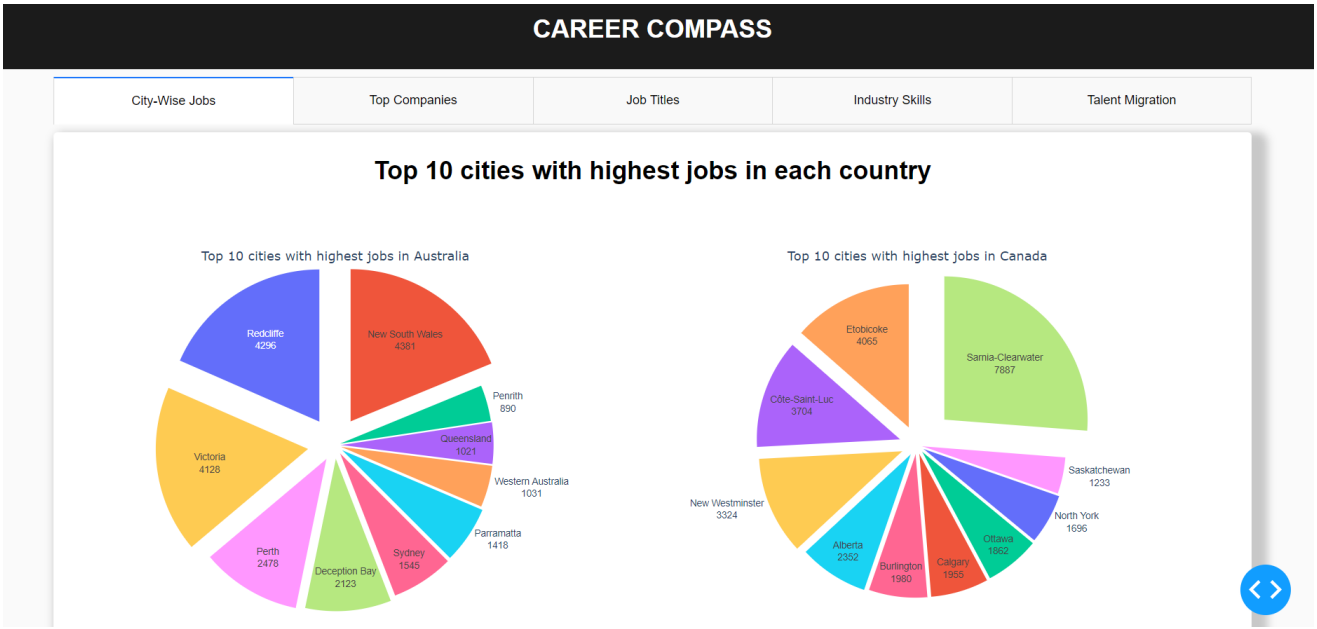
1. Using cytoscape, we also made an interactive network graph where each node represents either an industry (pink color coded) or a skill (grey). An edge exists between a skill node and an industry node if that skill is relevant to that industry. A node can be selected that also highlights its immediate neighbors. This helps in finding meaningful connections and patterns within industry and skills. One can use it to match their skills with the industries that they are relevant or inculcate certain skills to prepare for their target jobs in industry. We used dataset (3) to make this possible.

2. We have also made a pie chart showing the frequency of demand of certain technical skill using jobs on LinkedIn as gained from data from dataset (5). This can be used to find the most sought after skill for a particular tech sector.

4 Results

Our analysis using the career compass platform yielded valuable insights into various aspects of the job market. Overall, the Visual Career Guide platform equipped users with actionable insights to navigate the job market effectively and make informed career decisions aligned with their aspirations.

Main Dashboard Interface:-



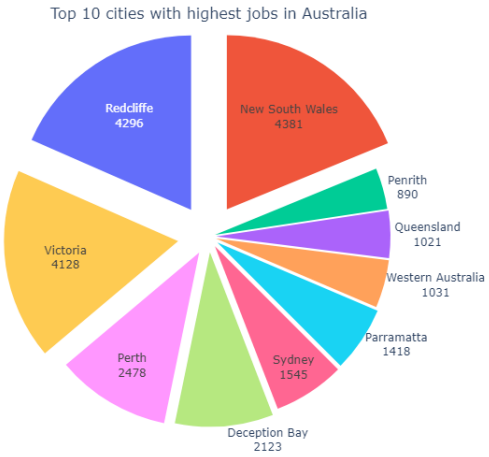
The main dashboard image showcases an interactive interface where users can seamlessly switch between tabs to explore various aspects of our web app. The tabs include:

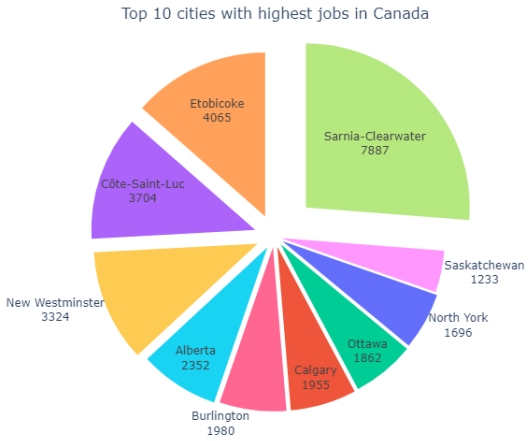
- City-Wise Jobs
- Top Companies
- Job Titles
- Industry Skills
- Talent Migration

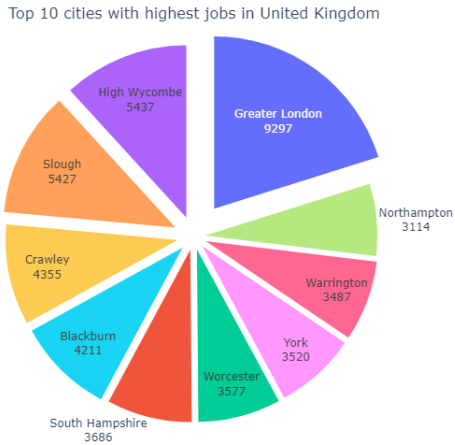
Each tab offers unique insights and functionalities, enabling users to engage with different web app features and access relevant information based on their preferences and requirements. This intuitive interface enhances user experience and facilitates efficient navigation throughout the platform.

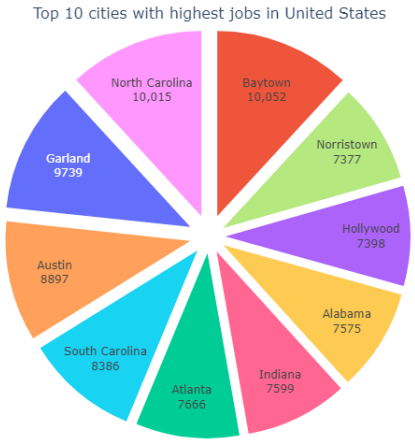
- **Top 10 cities with the highest jobs in each country**
  We've depicted the distribution of job opportunities across each country's top 10

cities using a pie chart. This visual succinctly illustrates the proportion of job postings in each city, allowing users to quickly identify where the highest concentration of job opportunities lies within a country.

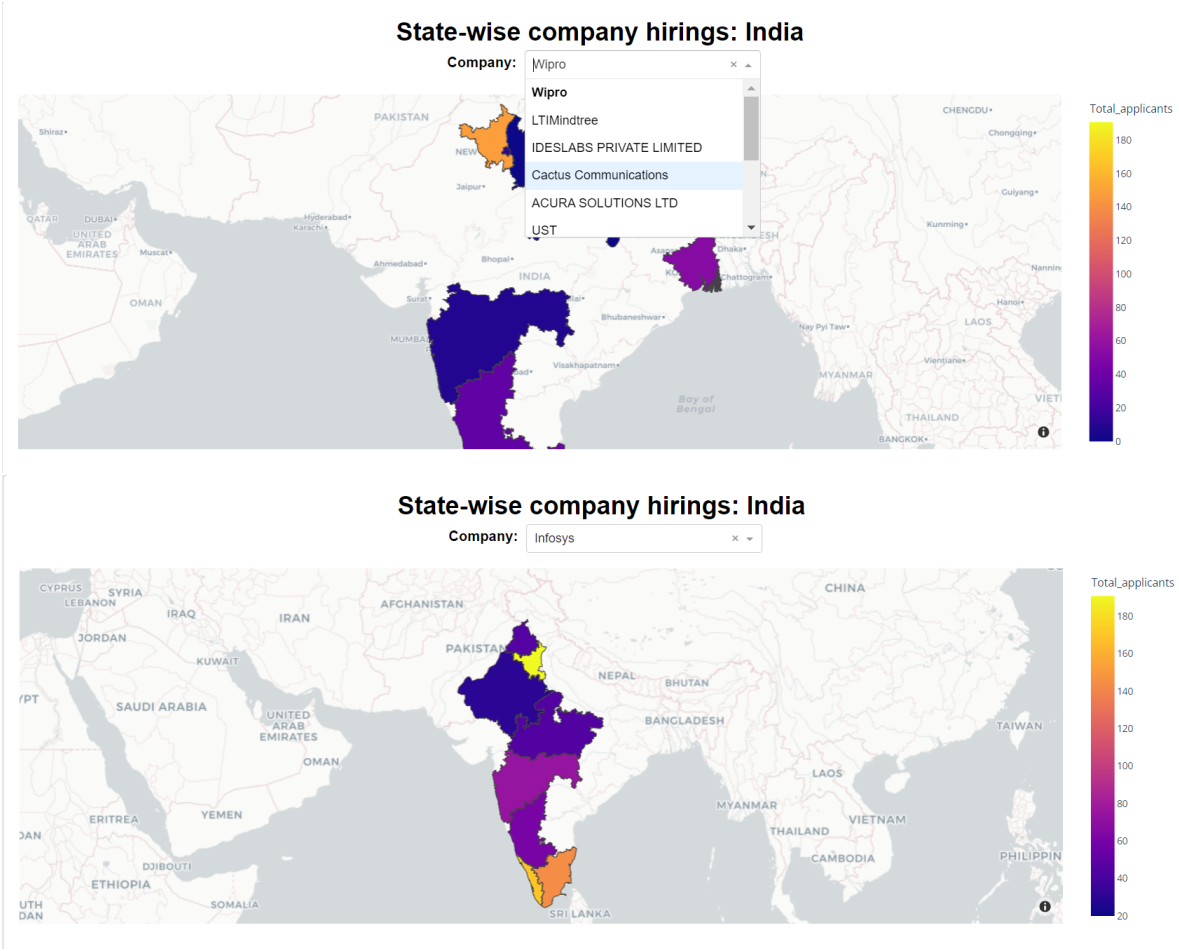








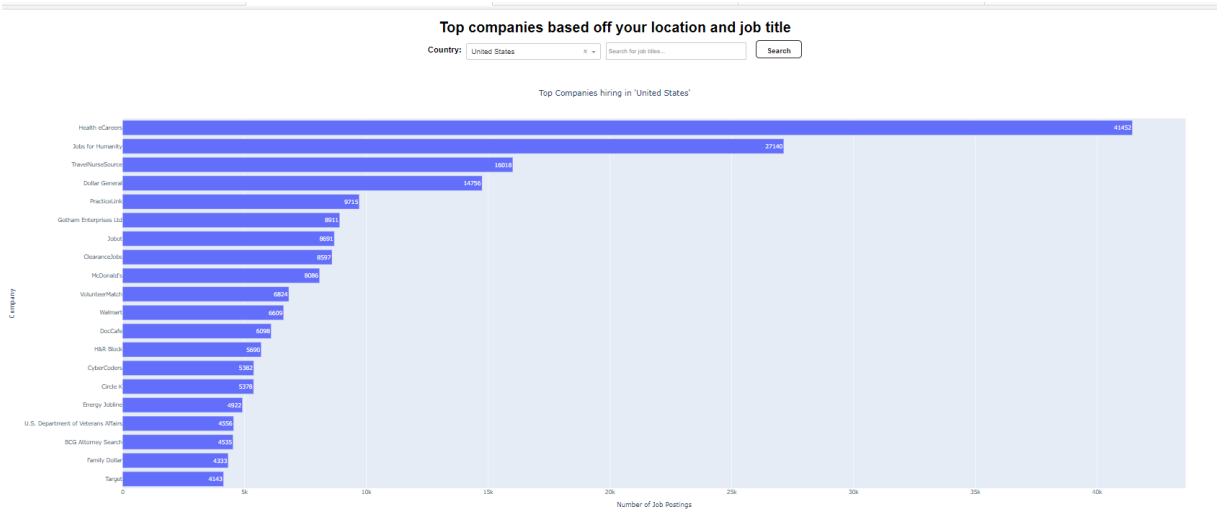
- **State-wise company hirings in India**
  Our next visualization focuses on depicting the distribution of company hirings across different states in India. Users can interact with a dropdown box to select a specific company, and the visualization dynamically updates to display the total number of applicants for that company in each state. The visualization utilizes a color range from 0 to 180 to represent the total number of applicants in each state. This color-coded representation allows users to easily identify states with higher or lower numbers of applicants for the selected company.



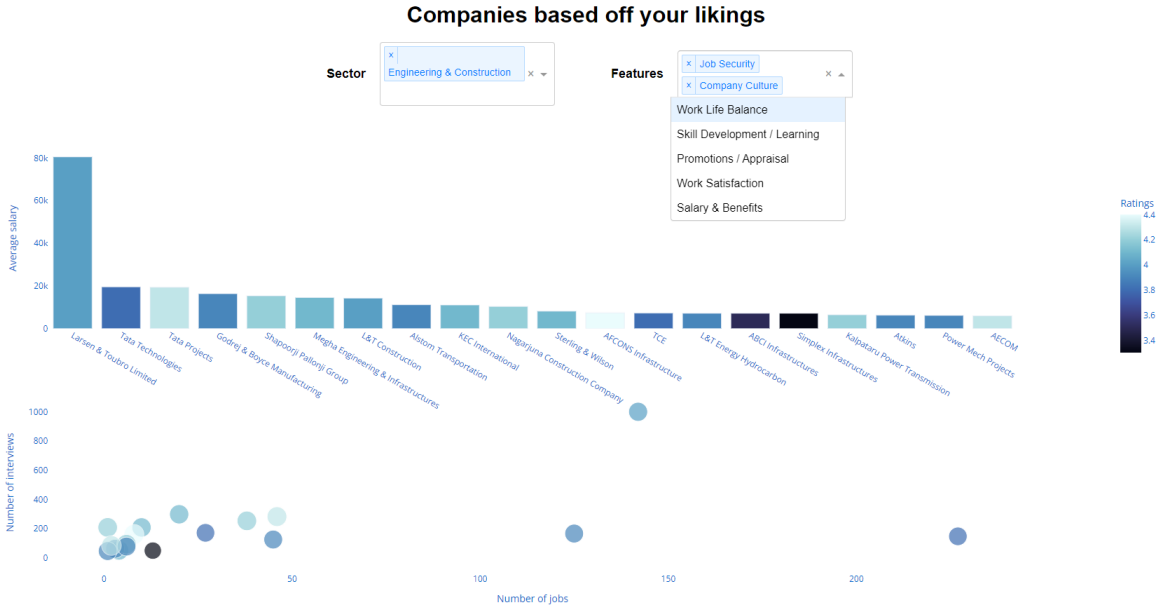
- **Top companies based on the location and job title**
  Our next visualization allows users to explore the top companies based on their location and job title preferences. Users interact with two dropdown menus—one for selecting the country and the other for choosing the job title. Additionally, a search button initiates the query.

  Upon selecting the desired country and job title, the visualization generates a bar graph illustrating the top companies corresponding to the user's criteria. Each bar represents a company, with the length of the bar indicating its prominence in hiring for the specified job title within the chosen country.

  This interactive feature empowers users to identify prominent companies hiring for their desired job title in a specific location, facilitating informed decision-making and job search strategies.



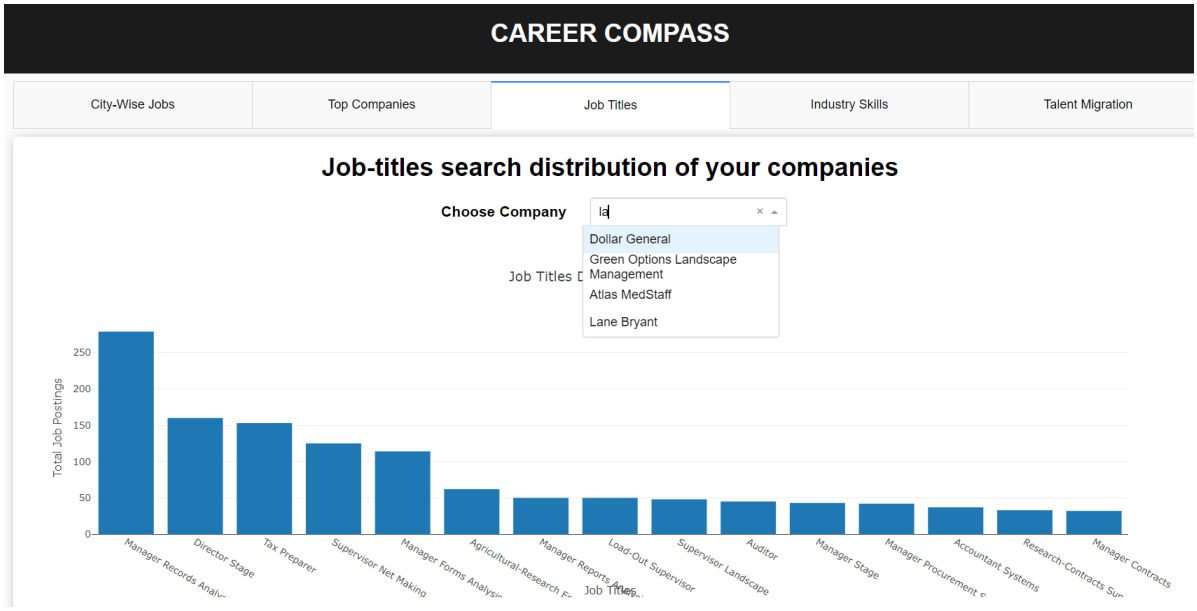
- **Companies based off likings**
  In this visualization, users can customize their search by selecting their preferred sector (e.g., banking) and desired feature (e.g., job security) from two dropdown menus. Upon selection, the visualization generates a bar graph depicting the average salary offered by different companies in the chosen sector, colour-coded according to their rating. This lets users compare salary offerings and company ratings simultaneously, helping them make informed decisions aligned with their preferences and priorities.



- **Job title distribution of your companies**

  This visualization simplifies the exploration of job title distributions within specific companies. Users can select a company from the dropdown menu, and the visualization generates a bar graph showing the total number of job postings for each title within that company. This lets users quickly grasp the distribution of job roles within their selected company, facilitating insights into potential career paths and

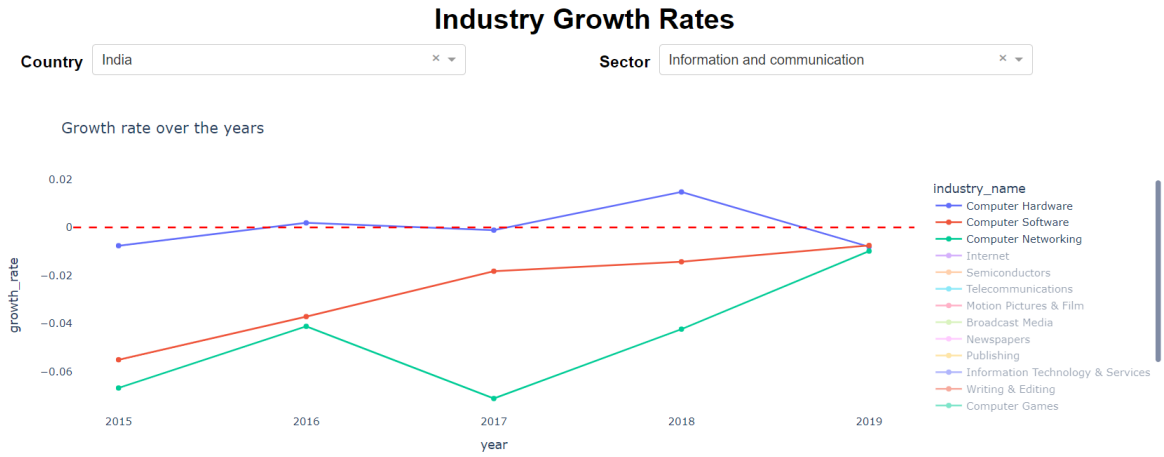
opportunities within the organization.



- **Top 20 job positions searched in the city**

  This visualization offers insights into the most sought-after job positions within a specific city. Users can utilize two dropdown menus to select the country and city of interest. Upon selection, the visualization generates a bar graph displaying the total number of job postings for the top 20 job positions in the chosen city. This allows users to identify the most in-demand roles within their desired location, aiding in job search and career planning efforts.

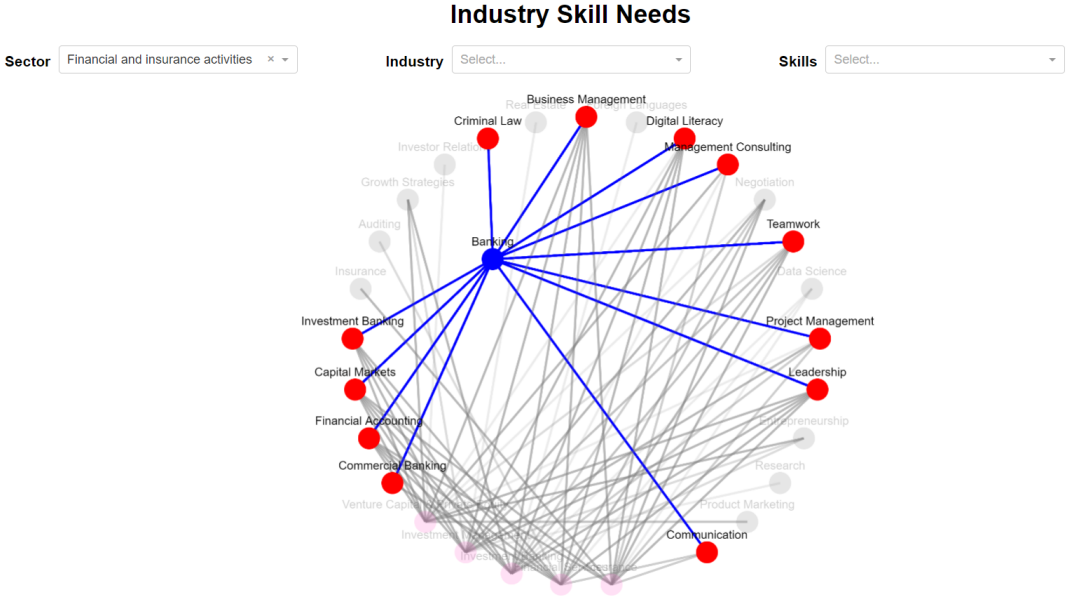
- **Industry growth rates in your city**
  This visualization allows users to explore the growth rates of various industries within a chosen sector and country. Users select the country and sector from dropdown menus, then view a line graph illustrating growth rates over the years. Each industry is represented by a different color, and users can toggle the visibility of specific industries for focused analysis.

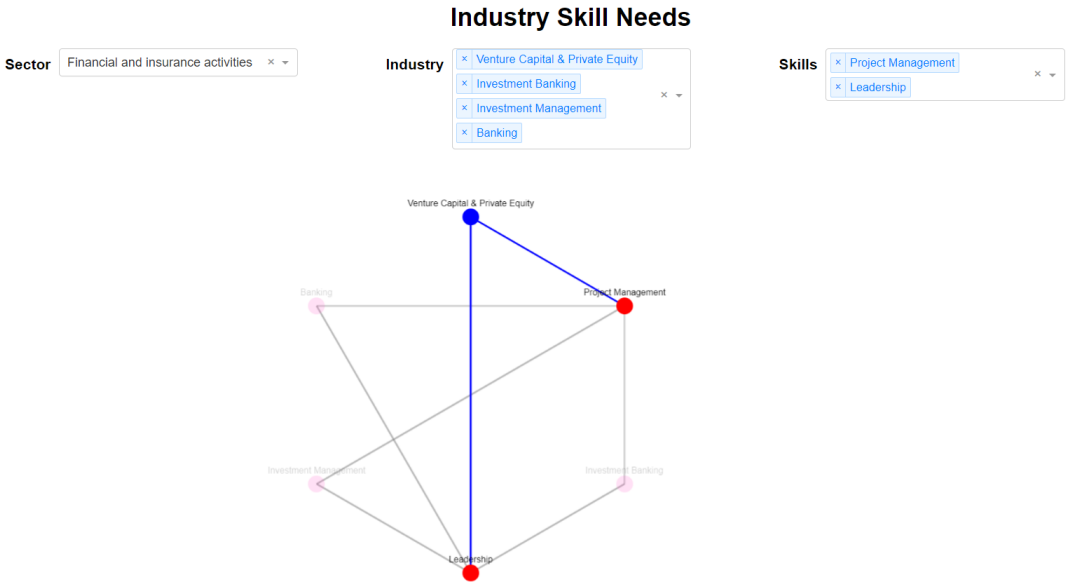


- **Industry skill needs for your job**

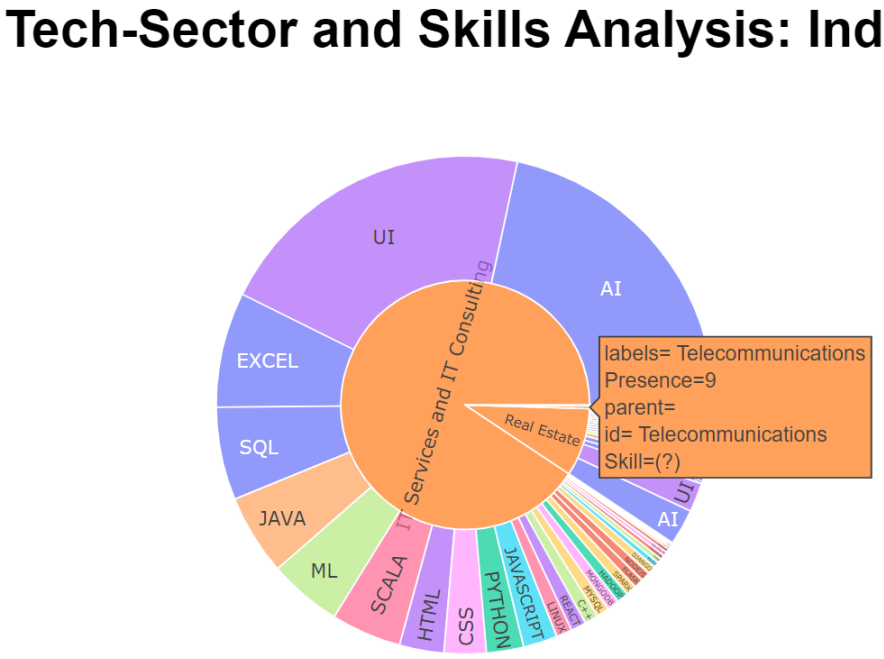
  This visualization provides insights into the skill requirements within a chosen sector and industry. Users can select the sector, industry, and specific skills of interest from three dropdown menus. The visualization then generates a network graph depicting the relationships between different skills required in the selected industry.

  Users can interact with the graph by clicking on individual nodes to view detailed information about each skill, including its relevance, demand, and associated job roles. This interactive feature allows users to explore the skill landscape within their chosen sector and industry, aiding in skill development and career planning efforts.

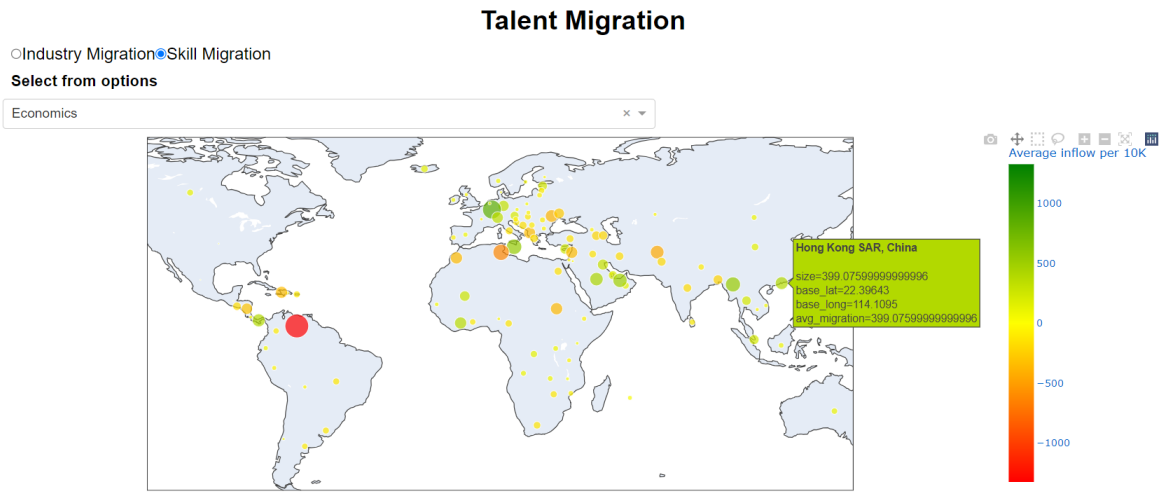




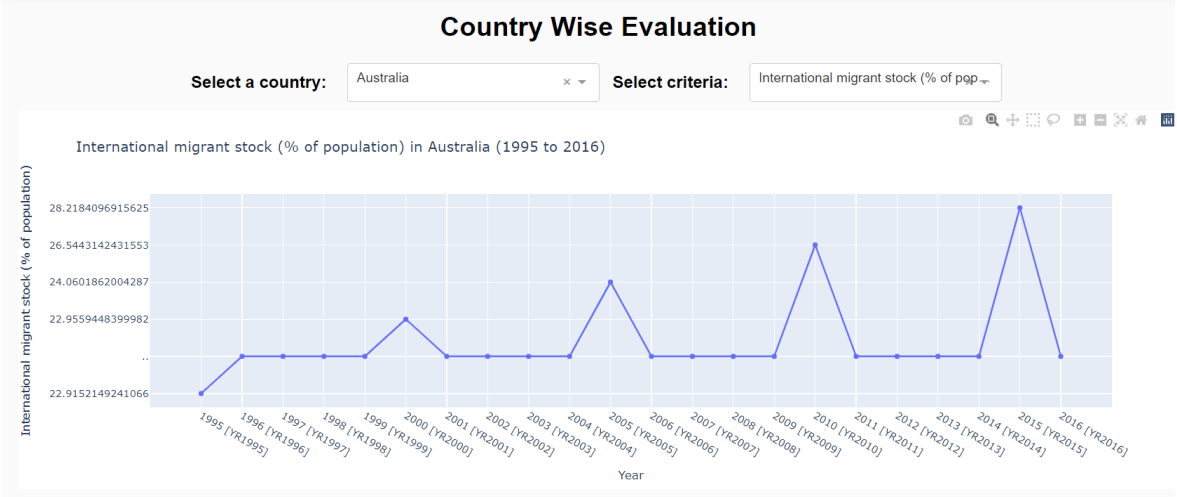
- **Tech-Sector and Skills Analysis in India**
  This visualization presents a pie chart depicting the distribution of skills within the tech sector in India. Each slice of the pie represents a specific skill, and the size of the slice corresponds to the relative prevalence of that skill within the sector. This visualization offers a quick overview of the most in-demand skills within India's tech industry, aiding professionals and policymakers in understanding skill trends and workforce needs



- **Talent migration from one part of world to other**
  Users can choose between two options: "Industry Migration" or "Skill Migration" using radio buttons. Upon selection, users can further specify their preference by selecting from a dropdown menu. The visualization then generates a Choropleth Map illustrating talent migration trends, providing users with a basic understanding of immigration patterns from one part of the world to another



- **Country wise evaluation**
  From dropdown menus, users can select a country and criteria (e.g., Unemployment with advanced education). The visualization generates a line chart illustrating the selected criteria for the chosen country over the years. This allows users to track trends and evaluate specific aspects of a country's performance or development over time.



5 Conclusion

In conclusion, navigating the complexities of the modern job market requires a strategic approach and access to comprehensive information. Our project aims to fulfill this need by harnessing the power of data visualization to enable individuals and businesses to make informed career decisions. By developing Career Compass, a data visualization platform, we aim to provide users with the tools and resources needed to confidently navigate the ever-evolving professional landscape.

By providing information from a variety of reputable sources such as LinkedIn, Ambition Box, and Kaggle, and using data preparation and visualization techniques, we have created a platform that provides valuable information about job opportunities, industry trends, and skill requirements. Users can explore the geographic distribution of jobs,

assess company profiles, and gain insights into industry growth and migration trends.

Our platform not only benefits job seekers by helping them find career opportunities that match their skills and aspirations but also provides valuable information to recruiters and companies looking for talent on sourcing, strategic planning, and market analysis. In addition, policymakers and researchers can use the platform to gain insights into labor market dynamics, economic trends, and global industry changes.

Furthermore, by accessing more authentic and reliable data sources, such as government databases and industry-specific reports, this dashboard can be further improved and become highly beneficial in providing accurate and up-to-date insights for users. Ultimately, Career Compass serves as a guide for both individuals and organizations, enabling them to make informed decisions, adapt to changing market conditions, and successfully navigate their professional journeys in an ever-evolving global economy.

6 Link to source code

To access our project source code visit our github repository at :
`https://github.com/jhaansireddy/CS661-Visual-Career-Guide`
Note: Since dataset (1) is too large (6.1 GB) we omitted to push it into our repository. Thus, it is mandatory that the dataset be downloaded from the source and the extracted files (3 csv files) be saved under *./Data/linkedin_jobs_and_skills* directory. Follow the instructions given in Readme to install requirements and run the project.

7 References

(1) `https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024`

(2) `https://www.kaggle.com/datasets/bhavikjikadara/top-worlds-companies`

(3) `https://www.kaggle.com/datasets/salehahmedrony/linkedin-digital-data`

(4) `https://www.kaggle.com/datasets/anurag629/jobs-data-world-wide`

(5) `https://www.kaggle.com/datasets/joebeachcapital/linkedin-jobs`