

# Process book - LinkedIn Job Postings 2023

## 1. Introduction

This project aims to develop a comprehensive website utilizing the LinkedIn Job Postings - 2023 dataset, providing insightful and user-friendly visualizations of the job market landscape. By leveraging data from LinkedIn job postings in 2023, the platform is designed to offer intuitive visualizations that help users navigate the job market effectively. The goal is to empower individuals with data-driven insights, enabling informed decisions regarding job searches, career paths, and recruitment strategies. By analyzing various attributes such as job titles, industries, locations, salaries, skills, and company profiles, our visualization will provide valuable information for job seekers, employers, and researchers exploring employment trends and patterns.

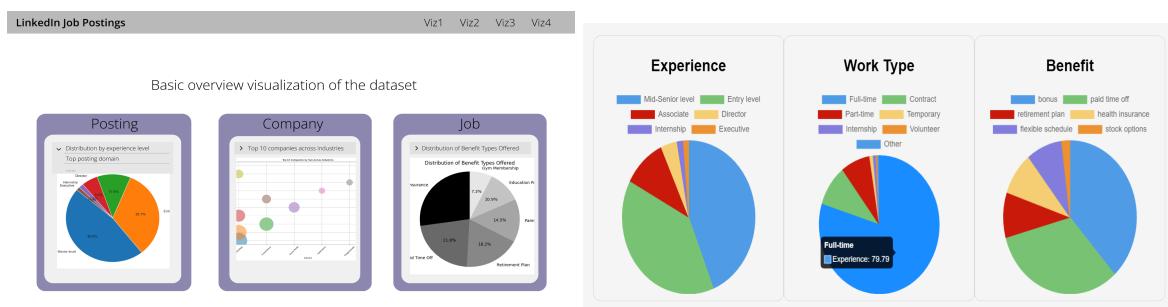
## 2. Visualization

### Overview

In this section, we offer a comprehensive overview of the data encompassing job postings, companies, and job details sourced from the LinkedIn Job Postings - 2023 dataset. We employ pie charts to delineate the distribution of various job positions within our dataset, ranging from entry-level roles to executive positions. Notably, the data indicates a prevalence of mid-level and entry-level positions across the board.

Furthermore, our analysis extends to the nature of work types prevalent across the job postings. Utilizing pie charts, we illustrate the distribution of work types such as full-time, part-time, internships, and volunteer positions. It emerges that full-time roles dominate the landscape, with internships representing a smaller fraction of the overall dataset.

Additionally, we delve into the benefits associated with each job posting. By extracting and recalculating the prevalence of different benefits across the dataset, we gain insights into the common benefits offered within the job market. This comprehensive analysis sheds light on the prevailing trends and practices within the employment landscape captured by the LinkedIn Job Postings - 2023 dataset.



In the left the initial sketch and right the final version

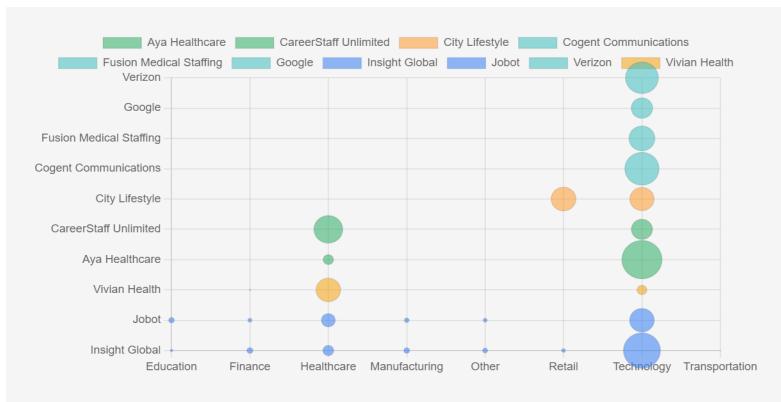
## **Industry Leaders**

Delving deeper into the insights gleaned from the LinkedIn job postings dataset, we aim to provide readers with a comprehensive understanding of the industry leaders and their distribution across various sectors. Utilizing data visualization bubble charts, we present the top company leaders and their prevalence within distinct industries.

Employing a bubble chart, we showcase the dispersion of these industry leaders and the volume of job postings attributed to each company in the year 2023 on LinkedIn. By varying the size of bubbles corresponding to each company, we illustrate the magnitude of job postings they generated within specific industries. This visualization aids readers in discerning the primary areas of focus for these influential companies.

Upon closer examination, it becomes evident that the majority of these prominent industry leaders primarily operate within the technology sector. Notably, Aya Healthcare emerges as a dominant force within the technology industry, consistently leading in terms of job postings. This observation underscores the significant hiring activity within the technology sector, with Aya Healthcare spearheading recruitment efforts.

The utilization of a bubble chart facilitates clear visualization and interpretation of these insights. By providing readers with visual cues regarding the distribution and prominence of industry leaders across various sectors, our analysis offers valuable insights into the prevailing trends within the job market landscape. This strategic approach enhances reader comprehension and facilitates informed decision-making regarding career pursuits and industry exploration.

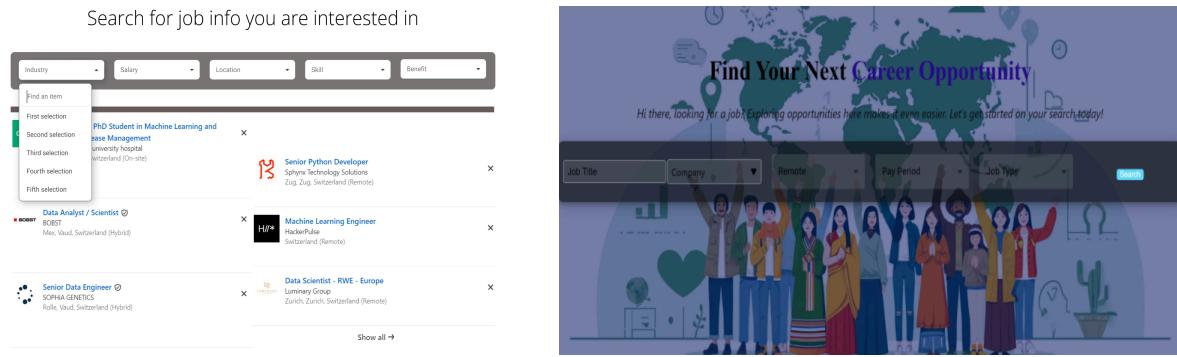


## **Job Searching**

With the LinkedIn job posting dataset at our disposal, navigating through various job positions and company listings manually was a necessity. To streamline this process, we developed a Searchable Job Posting Table aimed at enhancing the job-seeking experience for users. This interactive table enables users to effortlessly filter and search for job opportunities based on multiple criteria, including job title, company name, remote work availability, pay period, and job type.

The platform features a user-friendly interface coupled with straightforward backend functionality, ensuring an efficient and intuitive exploration of the dataset. This standalone website serves as a complement to the original dataset, accessible via the search button on the main platform. Inspired by alternative approaches observed on various websites, we opted to redefine this website significantly.

This redesigned platform offers readers an enhanced and visualized experience, simplifying access to the wealth of data available from LinkedIn for the year 2023. By providing a seamless and interactive interface, users can delve into the dataset with ease, facilitating efficient data exploration and analysis.



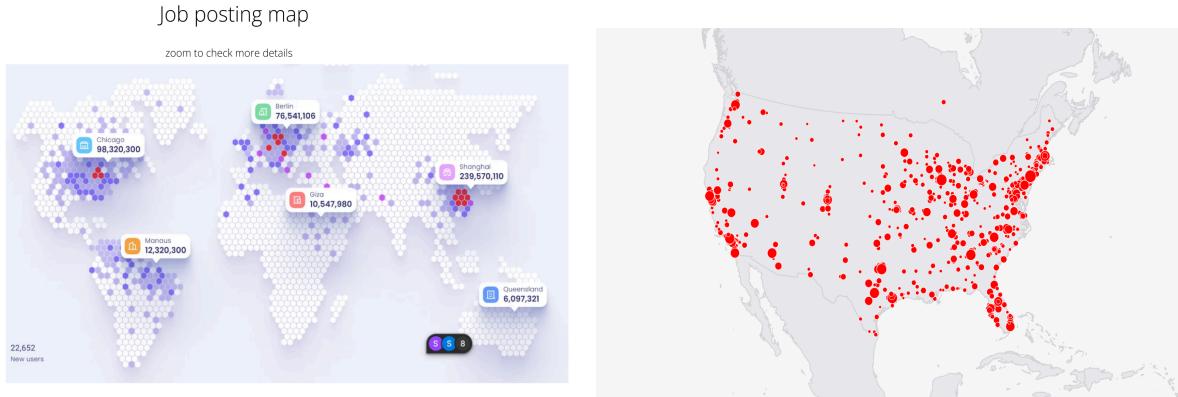
In the left the initial sketch and right the final version

### Interactive Map

Utilizing location data and additional information, we crafted an interactive map designed to offer deeper insights into the distribution of job postings across North America. Each point on the map corresponds to a specific geographical location, representing the concentration of job postings for a particular job title within that area. By deploying an information window triggered by cursor interaction with each circle, users can access additional details regarding the job postings associated with that location.

Upon exploration, it becomes apparent that the majority of job postings within our dataset are clustered in the United States. This visualization effectively highlights the geographical hotspots for various job opportunities, providing readers with a clearer understanding of where particular job roles are in high demand compared to others.

The utilization of this interactive map offers readers a more nuanced perspective on the distribution of job postings, allowing for a more informed analysis of regional job market trends. By seamlessly integrating geographical data with job posting information, this visualization enhances user comprehension and facilitates data-driven decision-making regarding job search strategies and career pursuits.



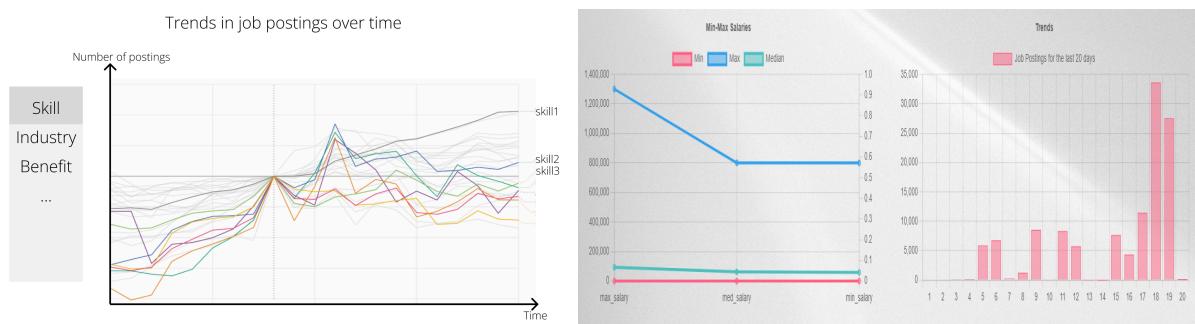
In the left the initial sketch and right the final version

## Statistics

Regarding salary visualization, we meticulously presented the distribution of maximum, median, and minimum salaries across different job postings. This approach aids in providing a nuanced understanding of the salary landscape within the dataset, allowing users to discern potential salary ranges associated with various job opportunities. By depicting these metrics graphically, we facilitate easier interpretation and comparison of salary information, thereby empowering users to make informed decisions regarding job prospects.

In exploring trends, our objective was to identify significant patterns over time, as outlined in milestone 2. However, due to the dataset's limitation to the year 2023, our analysis primarily focuses on this temporal window. Subsequently, we encountered challenges in accurately visualizing trends for the entire year of 2023, prompting us to narrow our focus to the most recent 20 days of available data.

Despite this limitation, our analysis revealed intriguing trends, notably an observed increase in job postings during weekends. This trend suggests a potential correlation between job posting activity and the timing of the week, offering valuable insights into recruitment patterns and hiring behaviors.

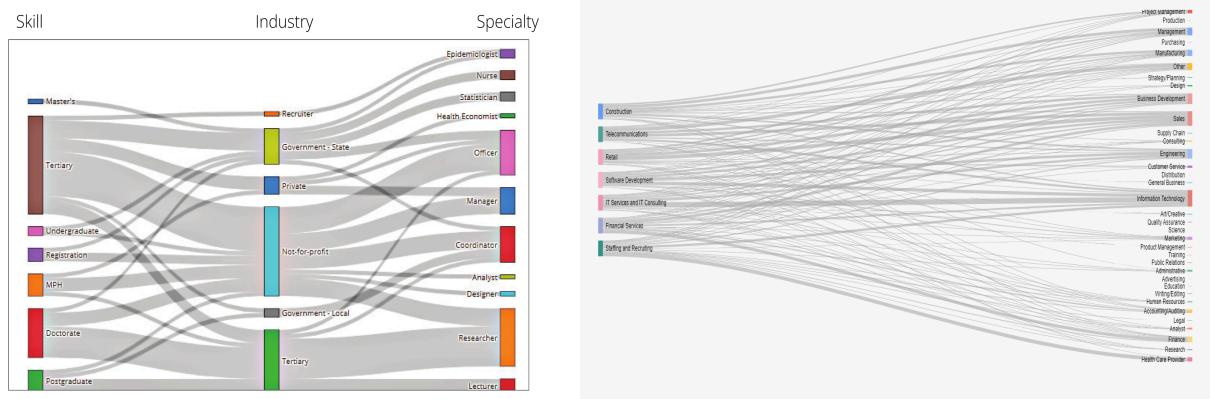


In the left the initial sketch and right the final version

## Sankey

We've utilized a Sankey diagram to visually represent the intricate relationships between various skills, industries, and specialties within the job market. In this diagram, the width of the connecting bands signifies the volume of transitions between different categories. For instance, a wide band flowing from the skill "Research" to the industry "Information Technology & Services," and subsequently converging on the specialty "Supply Chain," indicates a substantial number of individuals possessing analytical skills being employed within the supply chain sector of the Information Technology & Services industry. Moreover, this interactive visualization offers users the opportunity to hover over the flow lines, revealing additional insights. Users can access information such as the percentage of different skills required for a specific industry, the skill requirements for particular positions, and their respective proportions.

Links btw attributes - Sankey Diagram



In the left the initial sketch and right the final version

## 3. Path to obtain the final result

### 3.1 Data processing and analysis

To analyze and visualize the LinkedIn Job Postings data:

- Start by importing the data and inspecting its structure
- Handle missing values by removing or imputing them, and ensure all columns have correct data types.
- Remove duplicates to maintain data integrity. Standardize categorical data values for consistency and consider normalizing numerical data. Identify and handle outliers appropriately.
- Create useful new features, such as extracting date components or creating new categorical features.
- Aggregate and group data to provide insights, such as by job title, industry, or location.
- Utilize NLP (Natural Language Processing) techniques to extract benefits from the job descriptions for the benefits data.
- Extract longitude and latitude information from the state or country names to facilitate visualization on maps.

### 3.2 Data visualization

After preprocessing the data and ensuring its readiness, our primary objective was to closely replicate the original searches. To achieve this, we opted for React to leverage its extensive library ecosystem,

facilitating smoother implementation and enhancing overall functionality. By harnessing React's capabilities, we aimed to create a robust and user-friendly search experience that aligns closely with our requirements.

### 3.3 Web construction

#### **Job search platform:**

The web page consists of several key components implemented in both the frontend and back-end. The front end, developed with React, includes interactive elements such as text inputs, dropdown menus, and buttons that allow users to specify their job search parameters. The backend, developed using Flask, handles search queries and retrieves matching job listings from a well-cleaned database.

#### **Frontend Implementation for the main website:**

**React Components:** The application is built using React components such as App, FilterComponent, and App: The main component that serves as the entry point for the application, orchestrating the layout and navigation.

**PieChart:** Displays the distribution of job postings across different categories, such as job types or experience levels.

**BarChart:** Visualizes the frequency of job postings or other metrics across different categories, providing a clear comparison.

**BubbleChart:** Illustrates the relative size and importance of various industry players or job postings using bubble sizes.

**SalariesChart:** Showcases salary distributions, comparing median, minimum, and maximum salaries across different roles or industries.

**MapBubbleChart:** Presents a geographical distribution of job postings, with bubbles representing locations and sizes indicating the frequency of postings.

**SankeyDiagram:** Demonstrates the flow of skills to industries and specialties, with the width of the bands representing the volume of transitions.

**CustomDropdown:** These components manage the state and user interactions.

**State Management:** The useState and useEffect hooks are used to manage component states and side effects. For instance, useState stores search parameters, results, and UI states, while useEffect is used to initialize animations and handle updates to the selected job details.

**Styling:** CSS is used extensively to style the components, ensuring a visually appealing and responsive design.

#### **Backend Implementation:**

**Flask API:** The Flask application serves as the backend, providing a /search endpoint that handles GET requests. It uses the Pandas library to filter job listings based on the query parameters received from the front-end.

**CORS:** Cross-Origin Resource Sharing (CORS) is enabled to allow the front-end to communicate with the back-end hosted on a different domain or port.

## **4. Challenges and Design Decisions**

### **4.1 Job Searching**

Developing this application from scratch involved several challenges. Ensuring the dataset was clean and properly formatted required extensive preprocessing to handle inconsistencies and missing values. Establishing seamless frontend-backend communication was difficult, particularly with CORS issues and correct HTTP request handling. Efficient state management was essential for tracking user inputs, search parameters, and results. Designing a user-friendly and visually appealing interface required balancing aesthetics with functionality, ensuring responsiveness across devices. Setting up a robust build and deployment pipeline for the React application was also challenging. Due to time constraints, we created a screen recording to demonstrate the application's functionality. The recording showcased job search features, including filtering jobs, viewing search results, and detailed job information. This served as a proof of concept, demonstrating correct functionality and performance in a controlled environment.

### **4.2 Trend**

What we envisioned is a line chart where the number of postings fluctuates over time with certain trends. However, the data quality is not as good as expected, the main problem is the posting time, which is the dominant variable in our trend graphs. As there is only job posting data in 2023, the limitation is we cannot get a trend over years. And by processing data, we found that the majority of job postings were posted concentrating in two time periods in August and November 2023. So in the line chart we got, there are two significant peaks in these two periods, while only sporadic distribution at other times. So in the end we decided to put only the number of job postings per day for the lastest 20 days.

### **4.3 Sankey**

By processing the “industry - skill - speciality” data, we found that there are hundreds of industries and skills, and 21341 specialties, which is much more than the number we expected. So we chose the top 10 companies with the highest number of jobs offered, and the corresponding skills required for them as data for the Sankey diagram. For the specialties, as there are too many unique values, even for one particular industry, for example, there are more than 900 unique specialties in the “Advertising” industry, and it is hard to categorize them as there are no common or regular patterns of the specialties. It is unlikely to show all the specialties as nodes in the Sankey graph. Also, we tried to create a network graph with this “industry - speciality” data, but there are still too many nodes and the graph would be very messy. Therefore, we just keep the “industry - skill” links in the Sankey diagram.

## **5. Expansion of the Sketches and Changes**

Overview, interactive map, and job searching sections are well implemented basically as sketches, without much change.

- **Trends**

What we envisioned in the sketch is a line chart where the number of postings fluctuates over time with certain trends. For example, ideally, there might be a trend that numbers of job postings of remote allowed jobs during corona might be significantly higher than previous years, and decrease in recent years after

corona. And there might be increasing requirements in certain industries like “IT Services and IT Consulting” and “Retail”.

However, as we described in 4.2, the challenge is that the data is not as expected, the majority of the posting time concentrate on two periods, so we did not get the ideal distribution of job posting count or a certain trend in the visualization. In the end we decided to put only the number of job postings per day for the lastest 20 days, and one additional visualization of the approximate distribution of salary.

- **Sankey**

In the sketches, there are supposed to be 3 columns and there are links between industries and skills required for the job, and links between industries and companies’ specialties. But as we mentioned in 4.3, according to the processed “industry - skill - speciality” data, there are too many unique values in these 3 attributes. So we kept the top 10 companies with the highest number of jobs offered, and the corresponding skills required for them as data for the Sankey diagram.

## **6. Peer assessment**

Lluka and Yunlong:

- Main Website and Data Visualizations: Lluka and Yunlong worked on the main website and were responsible for the data visualizations. They implemented the React components and handled the deployment of the website. They also managed the data preprocessing required for the visualizations. They contributed equally for the process book.

Renqing:

-Job Search Platform: Renqing worked on developing the job search platform separately. This involved creating a user interface that allows users to search for job postings and integrating it with the main website. Limited contribution for the process book and screencast.