# MGMT 59000 Visual Analytics, Fall 2023
# Data Visualization Project
# Part II. Final Product

## Topic: Leveraging LinkedIn Data for 2023 Job Market Insights

---


By

Group 2:

Akshita Sharma

Yukti Sanjay Jain

Soham Agarwal

Sanveed Adnan Qureshi

Varun Annapareddy

Abithaa Shree Venkatesh


PURDUE
UNIVERSITY

Mitchell E. Daniels, Jr.
School of Business

## INTRODUCTION AND BACKGROUND

As students eager to enter the job market, we selected this topic to better understand employment dynamics. Our immediate objective is to find a job, making this study crucial. By analyzing over 15,000 job postings on LinkedIn over a two-day period, we hoped to gain insights into roles and other pertinent features. The project's outcomes were expected to include comprehensive data visualizations, actionable insights for job seekers and employers, trends in job listings, and an understanding of the relationship between skills, pay, and job demand.

The project aimed to leverage the LinkedIn dataset to benefit job seekers and recruiters by:

**Enhancing Job Knowledge**: The dataset offers insights into available job types by detailing job titles, categories, required skills, and qualifications. This helps job seekers understand in-demand roles and industries.

**Streamlining Job Search:** Through visualizations, job seekers can pinpoint job-rich regions and set realistic salary expectations. The dataset's location and salary data can be visualized to show job distribution and salary ranges, respectively.

**Understanding the Market:** Visual representations of job postings by industry and region highlight job market trends. Data on hiring companies can also guide job seekers in researching potential employers.

The end goal was to create dashboards, especially for Krannert University students like us, to make data-driven decisions using interactive tools like Tableau.

## DATASETS

One dataset we used used the “job_postings.csv” dataset for this project. It has about 15,887 job listings posted on LinkedIn over the course of 2 days and 27 attributes of these job listings.

- We found the dataset on Kaggle[3]
- The data was created to get a snapshot of the current job market and explore the highest compensated titles, companies, and locations; and examine how industries and companies vary through their job offerings and benefits.
- Arsh Kon from Kaggle utilized LinkedIn's backend search API to collect the data. He published it under the license CC BY-SA 4.0.
- This data has been collected by analyzing job postings on LinkedIn over the course of two days.
- No important publications.
- The dataset has 15,887 rows (job listings) and 27 columns (variables). The variables that we are interested in for the project include: Title, Job roles, work type, formatted_work_type, location, views, sponsored, formatted_experience_level, salaries, application URL, listing dates
- The dataset includes city-level job listings across all US states.
- We aim to make graphs at the city/ state level.

The other dataset that we used for this project was “company_industries.csv.” It had about 15,880 job listings alongside the industry categories that these job listings belonged to.

- We found the dataset on Kaggle[3]
- The data was created to get a snapshot of the current job market, which industries have the largest number of job postings etc.
- Arsh Kon from Kaggle utilized LinkedIn's backend search API to collect the data. He published it under the license CC BY-SA 4.0.

- This data has been collected by analyzing job postings on LinkedIn over the course of two days.
- No important publications.
- The dataset has 15,880 rows (job listings) and 2 columns (variables). The variables that we are interested in for the project include: ‘company_id’ and ‘industry’
- The dataset includes industry and company level job listings across all US states.
- We aim to make graphs at the city/ state level.

**Limitations of the Dataset(s)**

1. **Temporal Limitation:** Given the job market's inherent dynamism, a two-day snapshot might not fully capture longer-term trends or potential seasonal fluctuations in job postings.

2. **Geographical Limitation:** While the dataset offers city-level job listings across all US states, specific regions or cities within these states with unique job market dynamics might not be detailed in the dataset.

3. **Industry Bias:** Given LinkedIn's prominence in certain industries, our dataset could exhibit unintentional biases towards these sectors, potentially sidelining others.

4. **Granularity:** The dataset presents average salaries without further segmentation into specific job roles or levels within each state or experience category, potentially concealing disparities within the same category.

**Understanding of Concepts**

The dataset and the subsequent visualizations touch upon several key concepts:

1. **Data Aggregation**: By calculating average salaries, job postings counts, and other metrics, we've employed aggregation techniques to distill and convey multifaceted data.

2. **Correlation vs Causation**: While correlations are evident (e.g., states with higher salaries might have more job postings), it's pivotal to understand that these don't necessarily denote causation.

3. **Visualization Best Practices**: Our visualizations adhere to industry standards, employing clear labels, legends, and scales to articulate the data narrative effectively.

**Effective Comparisons**

The data enables insightful comparisons across multiple dimensions:

1. **State-wise Analysis**: With data providing insights into job opportunities and average salaries across states, this inter-state comparison aids both job seekers and businesses.

2. **Experience-wise Salary Progression**: The data's portrayal of salary alterations based on experience levels facilitates a comparison of the compensation trajectory across diverse roles and industries.

3. **Sponsored vs Remote Allowed by Industry:** This comparison, sourced from `company_industries.csv` and `job_postings.csv`, facilitates an understanding of industries' preferences towards remote work and sponsorships.

4. **Job Count by Company Size**: This metric offers insights into the distribution of job opportunities across companies of varying sizes.

## DATA STORY

### Time Invested in the Project

1. **Data Exploration**: Prior to visualization, significant effort was directed towards comprehending the dataset's intricacies, pinpointing outliers, and discerning patterns.

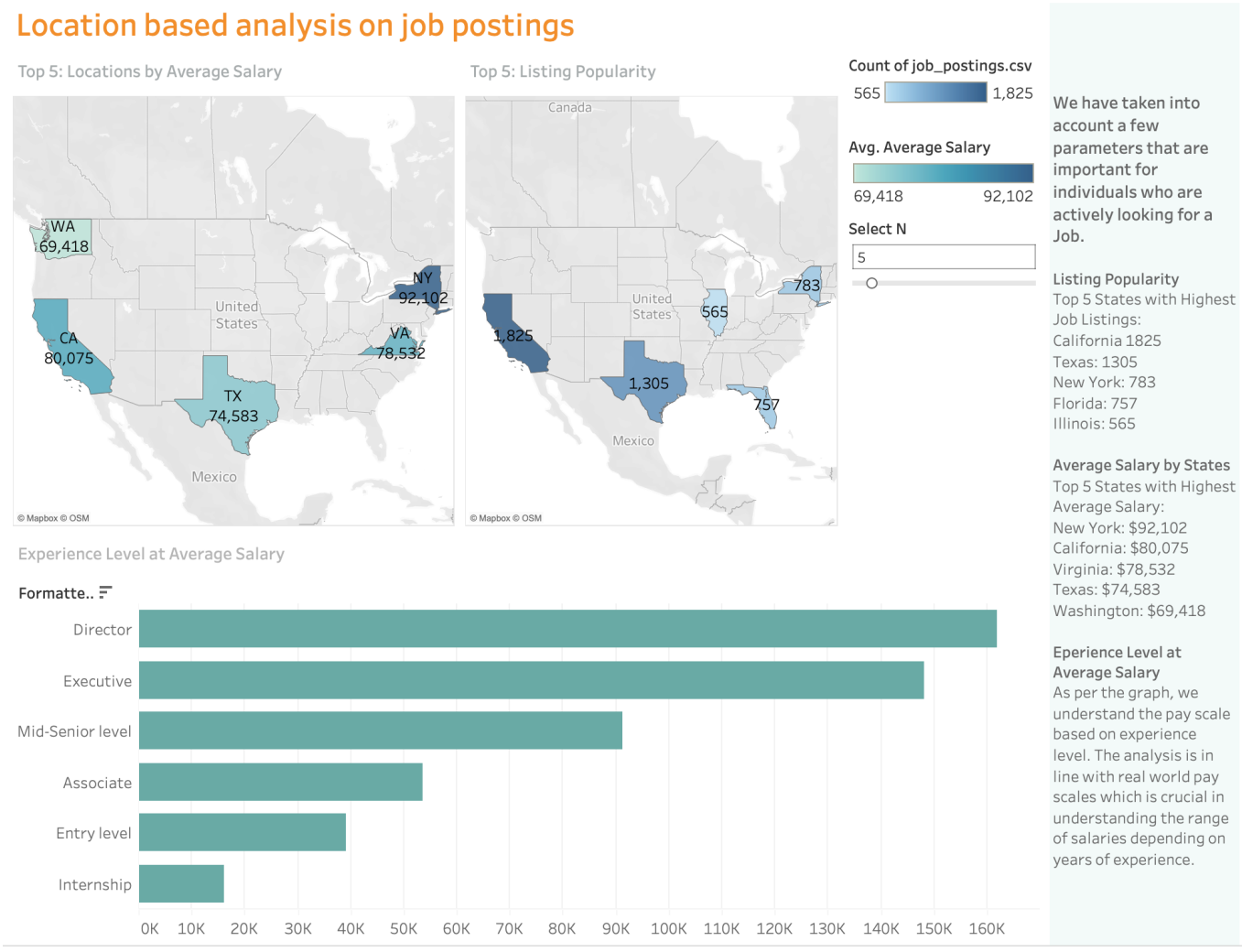
2. **Visualization Design**: Each chart was meticulously crafted to ensure clarity, appeal, and accurate data representation.

3. **Narrative Crafting**: The surrounding narrative was devised to offer context, insights, and actionable takeaways, catering to both job seekers and employers.

### Charts and Their Significance

Below are some charts that we created from our select datasets to enable job seekers, recruiters, and companies streamline their search, gain valuable insights, and make smarter decisions:

#### 1. Location Analysis and Pay Scale based on Experience Level



| Chart | Data Tables | Variables | Categories/ Units |
|---|---|---|---|
| Average Salary | job_postings.csv | Location vs Average Salary | Salary in dollars ($) |
| Listing Popularity | job_postings.csv | Location vs count of job postings | Absolute count of number of job postings |
| Experience Level | job_postings.csv | Formatted Experience Level vs Average Salary | Salary in dollars ($) |

**Location vs Average Salary: Economic Disparities Across Regions**

The chart delineates distinct economic disparities across the US. **New York**, with an average salary of $92,102, signifies the concentration of high-paying industries, notably finance and technology. However, a critical consideration is the elevated cost of living in New York, which might attenuate the net benefits of higher wages.

**California's** average salary of $80,075 is indicative of its robust tech sector, particularly in regions like Silicon Valley. Nonetheless, California's diverse economy, ranging from entertainment to agriculture, likely results in a wide salary spectrum.

The average salaries in **Virginia, Texas, and Washington** underscore the influence of regional industry strengths — tech industries in Washington, energy sectors in Texas, and defense contractors in Virginia, for instance.

**Location vs Count of Job Postings: Analyzing Market Vibrancy**

The pronounced number of job postings in **California (1,825)** suggests a vibrant job market. However, it's imperative to understand that high opportunity regions can also attract intense competition. California's multifaceted economy and the presence of numerous global corporations contribute to this heightened activity.

**Texas**, showcasing 1,305 listings, reflects its burgeoning reputation as an economic nexus. Factors such as favorable business policies, diverse industrial presence, and strategic location contribute to this growth. Yet, the potential for market saturation in certain sectors remains a consideration.

The listings in **New York, Florida, and Illinois** emphasize the significance of regional economic hubs, with each state having its industry-specific predominance, shaping the nature and volume of job postings.

**Experience Level vs Average Salary: The Compensation Trajectory**

At the entry level, an **intern** can anticipate a compensation of approximately $16,183. As experience accrues, the compensation trajectory becomes more favorable, culminating in director-level roles that command an average of $161,872. This progression, while seemingly linear, is influenced by variables such as industry specialization, skill demand, and continuous professional development.

These charts provide a **comprehensive view of the US job market**, influenced by diverse regional, industrial, and experiential factors. For job seekers and businesses alike, these insights are i**nstrumental in strategic decision-making and understanding the broader economic landscape**.

**2. Sponsorship and Work Type by Industry**

Work-type and number of sponsored jobs by industry

| Industry | Work .. | |
|---|---|---|
| Accounting | FULL_TIME | 32 |
| | CONTRACT | 1 |
| | PART_TIME | 0 |
| | INTERNSHIP | 0 |
| Airlines/Aviation | FULL_TIME | 4 |
| | TEMPORARY | 1 |
| Alternative Medicine | PART_TIME | 0 |
| | FULL_TIME | 0 |
| Animation | FULL_TIME | 0 |
| Apparel & Fashion | FULL_TIME | 25 |
| | PART_TIME | 10 |
| | TEMPORARY | 0 |
| | OTHER | 0 |
| | INTERNSHIP | 0 |
| Architecture & Planning | FULL_TIME | 12 |
| | INTERNSHIP | 0 |
| | CONTRACT | 0 |
| Automotive | FULL_TIME | 43 |
| | INTERNSHIP | 1 |
| | TEMPORARY | 0 |
| | PART_TIME | 0 |
| | CONTRACT | 0 |
| Aviation & Aerospace | FULL_TIME | 17 |
| Banking | FULL_TIME | 42 |
| | PART_TIME | 2 |
| Biotechnology | FULL_TIME | 30 |
| | CONTRACT | 8 |
| | INTERNSHIP | 1 |
| | TEMPORARY | 0 |
| | PART_TIME | 0 |
| Broadcast Media | FULL_TIME | 17 |
| | PART_TIME | 3 |
| | CONTRACT | 1 |
| | INTERNSHIP | 0 |
| Building Materials | FULL_TIME | 32 |

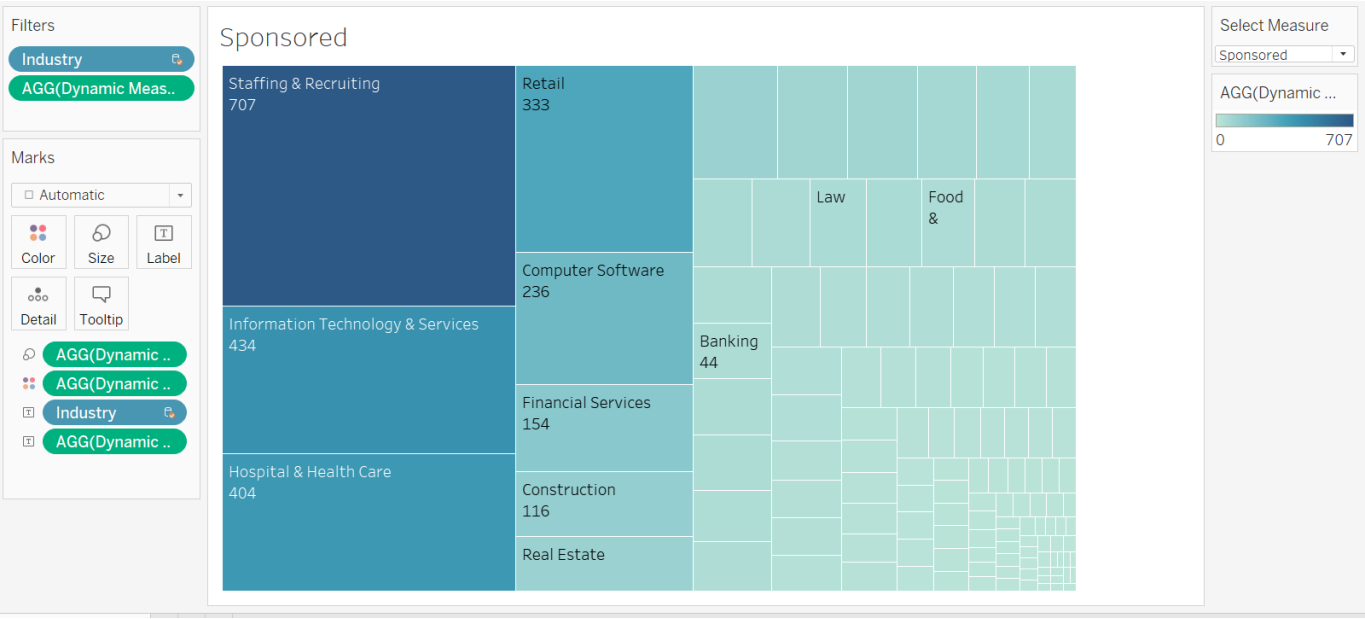
| Chart | Data Tables | Variables | Categories/ Units |
|---|---|---|---|
| Work-type and number of sponsored jobs by industry | company_industries.csv, job_postings.csv | Industry, Work-type, Sum of Sponsored | Count |

In this comprehensive Tableau crosstab analysis, we explore the distribution of sponsored jobs by work-type across a wide range of industries, shedding light on the varied employment landscapes. The striking commonality that emerges is the strong prevalence of Full-Time positions in the top five industries: Staffing and Recruiting, Hospital and Healthcare, Information Technology and Services, Computer Software, and Retail. This pervasive preference for Full-Time employment underscores its importance in the labor market.

However, a closer look reveals intriguing nuances in work-type distribution within each industry when it comes to sponsored jobs.. While Full-Time roles are prominent, Part-Time and Contract positions also occupy significant space, providing flexible options for both job seekers and employers. This diversity in work-type distribution speaks to the adaptability of different industries and caters to the varied needs of their workforces.

At the same time, Internship and Other work-types are found to be less common across the board, reflecting the traditional structures that remain prevalent in the job market. This analysis is invaluable for job seekers, employers, and policymakers, providing data-driven insights into labor market trends, recruitment strategies, and workforce planning within a broad spectrum of industries. Whether you're navigating the dynamic landscape of Information Technology or exploring opportunities in the ever-evolving Retail sector, this crosstab analysis offers a comprehensive basis for informed decision-making tailored to the specific characteristics of your chosen industry.

**3. Sponsorship vs. Remote Allowed by Industry: Top Industries offering Sponsorships/ Remote**



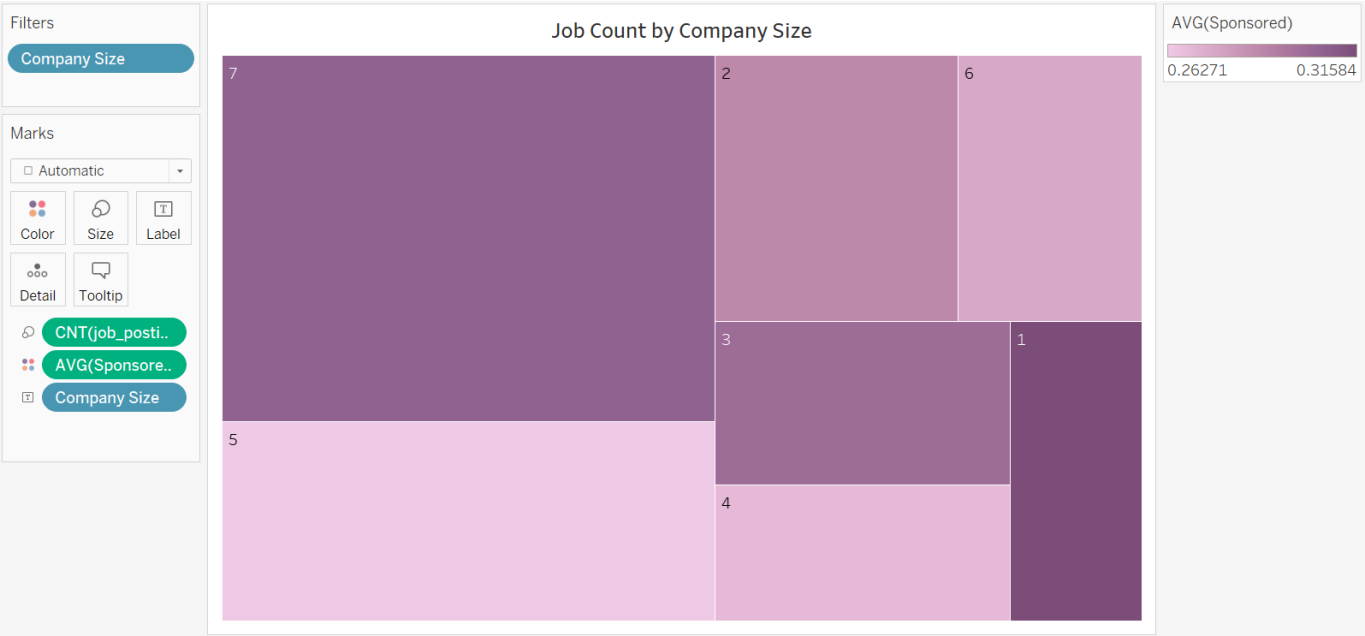

| Chart | Data Tables | Variables | Categories/ Units |
|---|---|---|---|
| Sponsorship vs Remote Allowed by Industry | company_industries.csv, job_postings.csv | Industry, Sum of Sponsored or Sum of Remote Allowed according to selection from Parameter 'Select Measure' | Count |

The **treemap** offers a quick insight into the top industries offering the most sponsorships or remote work. The variable adjacent to the 'Industry' can be changed according to one's preference from the parameter '**Select Measure**.' The sheet name and values change according to the selection. Hence by taking a quick look at the chart and viewing the size of the boxes and the numbers written within them, job seekers can immediately identify which industry offers the most sponsorships/ most number of remote jobs. For-example, from the chart it can be seen that "**Staffing & Recruiting**" offers the most sponsorships while "**Information Technology & Services**" allows the most number of remote jobs.

This information is quite valuable for job seekers as **international students/ workers require sponsorship** to work in the USA. Hence, they can use the insights offered by this chart to streamline / focus their job search. Similarly, many job seekers might be interested in cutting commuting and living costs by doing remote work, this chart

offers excellent insights to such job seekers as well. As with this project, the **aim was to facilitate the job seekers** direct/ focus their job search process, this chart enables them to achieve exactly that through clear and simple visuals.

### 4. Job Count by Company Size



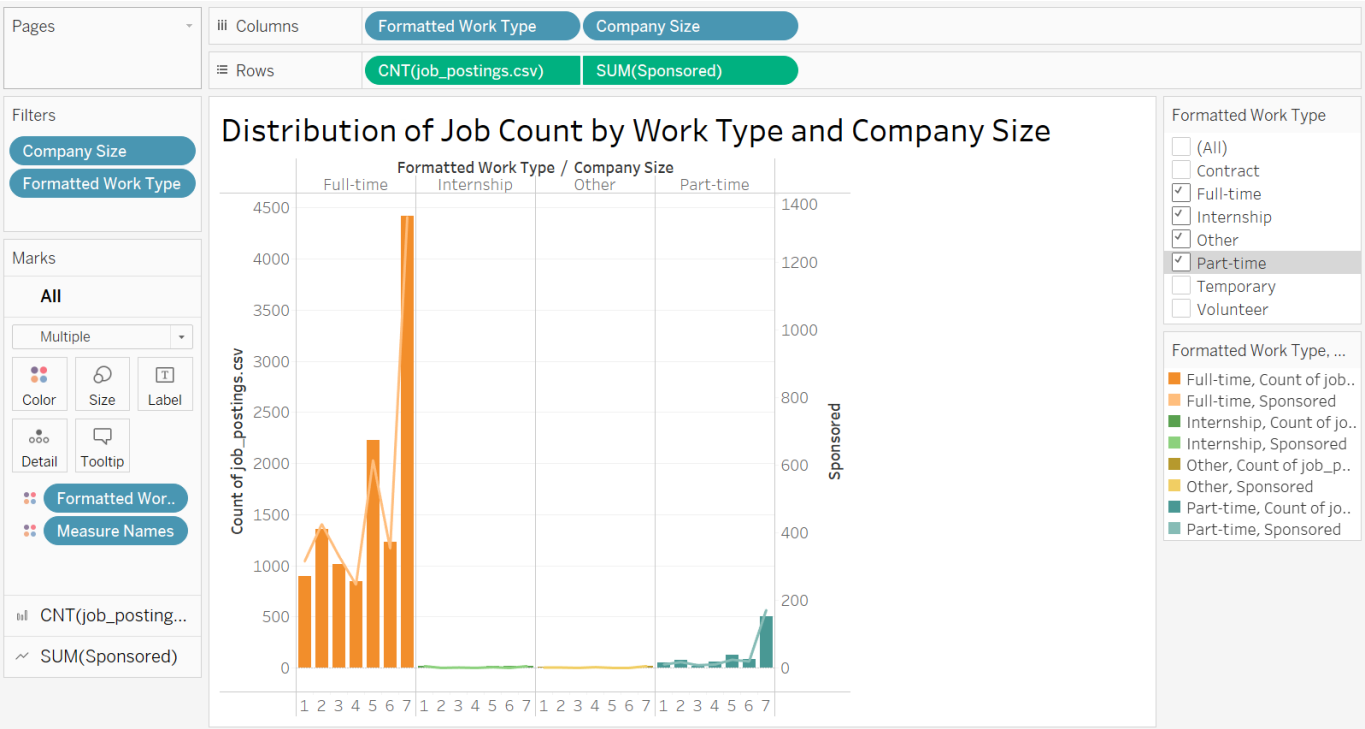

| Chart | Data Tables | Variables | Categories/ Units |
|---|---|---|---|
| Job Count by Company Size | company.xlsx, job_postings.csv | Sponsored, Company Size | Count of job_postings.csv, Average of sponsored |

This visual aid serves as a compass for job seekers, especially those seeking international positions:

- Quick Assessment: The chart's box size and color combination helps identify the abundance of job postings and sponsorship likelihood by company size.
- Example Highlight: Consider the smallest companies, which, despite having fewer job openings, boast a higher sponsorship rate.

This chart is invaluable for international job seekers who rely on sponsorship to work in the USA. It equips them with actionable insights to fine-tune their job search efforts effectively. Similar to this project's aim, it simplifies and clarifies the job search process, providing a roadmap for success through accessible visuals.

### 5. Distribution of Job Count by Work Type and Company Size




| Chart | Data Tables | Variables | Categories/ Units |
|---|---|---|---|
| Distribution of Job Count by Work Type and Company Size | company.xlsx, job_postings.csv | Formatted Work Type, Company Size | Count of job_postings.csv, Sum of sponsored |

Bar Chart: Number of Job Counts by Work Type and Company Size

The bar chart presents a comprehensive view of the distribution of job counts based on work type and company size. This visual representation offers insights into how different work types are distributed across various company sizes, shedding light on the employment landscape. The chart provides a clear snapshot of job opportunities across sectors and company sizes.

- Customizable Work Type Analysis: The horizontal axis categorizes job postings by work type. Using the provided checkbox filters, users can easily customize their view by selecting specific work types. This feature allows job seekers to tailor the chart to their specific interests, focusing on the work types that align with their career goals.
- Company Size Distribution: The vertical axis represents the number of job postings, making it easy to see the job counts for each combination of work type and company size. This information helps job seekers understand where the most opportunities are and which company size categories have a more significant demand for specific work types.

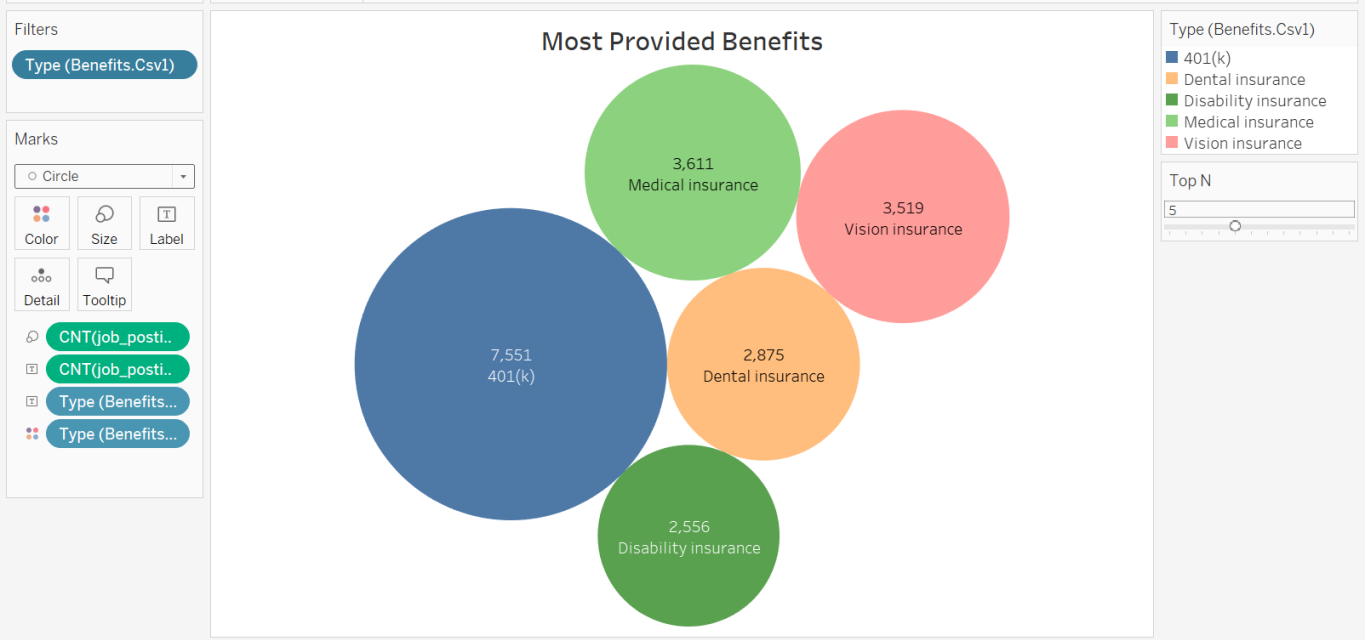
Dual-Axis Line Chart: Number of Sponsored Jobs by Formatted Work Type and Company Size (Dual Axis)

The dual-axis line chart complements the bar chart by illustrating the number of sponsored jobs in each combination of work type and company size. This dual-axis approach enables job seekers to assess not only the job count but also the sponsorship availability for their preferred work types and company sizes.

- Customized Sponsored Job Insights: The primary axis displays the number of sponsored job postings. Users can further refine their analysis by using the checkbox filters to focus on specific work types. This feature empowers job seekers to identify which work types and company sizes are more likely to provide sponsorship, aligning their job search with their individual requirements.
- Alignment with Job Counts: The dual-axis feature allows you to directly compare the number of sponsored jobs with the total job count in each category. Users can personalize this view by selecting specific work types via the checkbox filters, providing a comprehensive understanding of the job market tailored to their preferences.

Together, these visualizations are powerful tools for job seekers, allowing them to make data-driven decisions by customizing their analysis based on their chosen work types. The addition of checkbox filters enhances the user experience, streamlining the job search process and increasing the likelihood of finding suitable opportunities.

### 6. Most Provided Benefits




| Chart | Data Tables | Variables | Categories/ Units |
|---|---|---|---|
| Most Provided Benefits | benefits.csv1, job_postings.csv | Type | Count of job_postings.csv |

This dynamic circle chart visually represents job postings based on both the number of openings and the types of benefits offered. While the circle size reflects job counts, the color scheme distinguishes benefit types.

- Circle Size Significance: Larger circles indicate a higher number of job openings within specific sectors.
- Color-Coded Benefits: Each color represents a distinct benefit type. This feature allows users to quickly identify the perks associated with job postings.

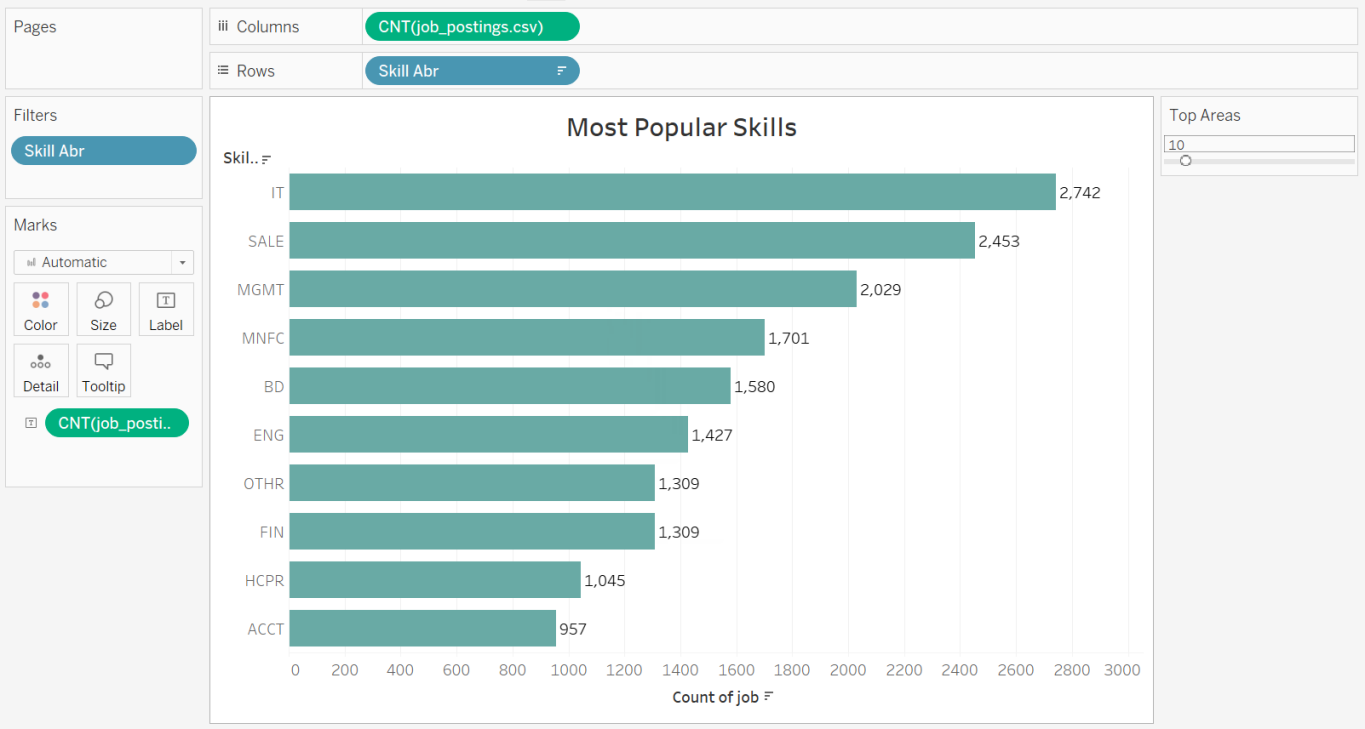
Key Insights:

- Diverse Benefits: The chart showcases a range of benefits, including "Medical Insurance," "401(k)," "Paid Maternity Leave," and more.

- Top N Benefit Filters: Users can filter the chart to focus on their top N preferred benefits. This functionality streamlines job searches based on specific benefit priorities.

For example, the number of job postings with 401(k) benefits are 7551. This simplifies the process of pinpointing employment opportunities in sectors that align with their preferred benefits, enhancing the overall job search experience.

### 7) Most Popular Skills




| Chart | Data Tables | Variables | Categories/ Units |
|---|---|---|---|
| Most Popular Skills | job_skills.csv, job_postings.csv | Skill Abr | Count of job_postings.csv |

This horizontal bar chart provides a clear and concise overview of job counts based on skills abbreviations. Each bar on the chart represents a specific skill, while the length of the bar corresponds to the number of job postings requiring that skill.

Key Insights:

- Skill Demand: The chart reveals the demand for various skills in the job market, making it easy to identify which skills are in high demand based on the length of the bars.
- Comparison: Job seekers can quickly compare the popularity of different skills, helping them understand which skills are most sought after by employers.
- Data-Driven Decisions: This chart empowers job seekers to make data-driven decisions about which skills to prioritize in their job search, enhancing their chances of finding relevant job opportunities.
- Tailored Job Search: By focusing on skills with higher job counts, job seekers can customize their job search to match the skills they possess or are interested in developing.

- Industry Insights: The chart can also provide insights into which industries or sectors have the greatest demand for specific skills.
- Filter by Top N Areas/Skills: Users have the flexibility to apply filters, narrowing down the chart to display their "Top N" skills. This feature streamlines the job search, ensuring a laser-focused approach based on specific skill priorities.

In summary, this horizontal bar chart simplifies the job search process by providing a visual representation of the demand for various skills in the job market. Job seekers can use this information to better target their job search efforts and increase their chances of finding job opportunities that match their skill set.

## SUMMARY AND CONCLUSIONS

This report outlines a comprehensive study on employment dynamics, utilizing datasets based on LinkedIn job postings. The project aimed to provide valuable insights for job seekers and employers, with a focus on understanding job roles, salary expectations, and market trends. The key datasets used included "job_postings.csv," job_skills.csv," "benefits.csv1," "company.xlsx" and "company_industries.csv." The study also acknowledged certain limitations, such as temporal and geographical constraints and potential industry bias. Key concepts explored in the project included data aggregation, correlation vs. causation, and visualization best practices.

The report presents several data visualizations, each with specific significance that help users understand regional economic disparities, gain insights into high-paying industries, cost-of-living considerations, and vibrancy and competitiveness of the job market. Other visualizations help users assess the most in-demand skills, job opportunities based on company size and sponsorship likelihood, distribution of job counts across different work types and company sizes, and the types of benefits offered. These visualizations can empower job seekers and recruiters to make data-driven decisions about which skills to prioritize in their job/candidate search, enhancing their chances of finding relevant talent/ job opportunities.

The project has successfully leveraged LinkedIn job posting data to provide actionable insights for job seekers, recruiters, and businesses. The visualizations presented in the report offer a deep understanding of employment dynamics, salary expectations, and industry trends. The report acknowledges limitations related to data temporality, geography, and potential industry bias, highlighting the need for careful interpretation of the findings. The data visualizations and analysis tools created in this project can be valuable resources for Krannert University students and job seekers in general, helping them make informed, data-driven decisions in their job search and career planning.

## REFERENCES

[1]https://www.wsj.com/articles/lets-face-it-linkedin-might-3 charts (be-the-best-social-network-right-now-11644402715
[2]https://smallbiztrends.com/2023/07/popular-social-media-sites.html?expand_article=1
[3]www.kaggle.com/datasets/arshkon/linkedin-job-postings.
[4]https://www.indeed.com/career-advice/resumes-cover-letters/skills-employers-look-for
[5]https://interstride.com/blog/how-to-answer-will-you-now-or-in-the-future-require-sponsorship-to-work-in-the-us/
[6]https://www.nerdwallet.com/article/loans/student-loans/average-salary-by-age

## CONTRIBUTIONS

| Author | Contributions |
|---|---|
| Akshita Sharma | 2 charts (Circle, Dual Axis Bar and line - both visualization and story), Introduction and Background |
| Yukti Sanjay Jain | 2 charts (Treemap, Horizontal Bar Chart - both visualization and story), Introduction and Background |
| Soham Agarwal | 1 chart, objectives, and goals, data story for chart and introduction to data story. |
| Sanveed Adnan Qureshi | 1 chart (Treemap - both visualization and story), datasets, summary and conclusions |
| Varun Annapareddy | 3 charts (2 Maps, 1 bar chart - visualisations with interactions), data description |
| Abithaa Shree Venkatesh | 1 chart (Crosstab, both visualization and story), formatting and editing, chart ideations |