PROJECT PART 3

TOPIC – LINKEDIN JOB POSTING DATASET

TEAM-03

Team Members

SAI POOJA PRAVALLIKA APPIKATLA (11659092)

SOWJANYA TEJAVATH (11656749)

VENKATA NAGA GOHITH SAI V (11609635)

ARPANA JADAV (11564523)

PRADEEP KUMAR YAMJALA (11716180)

VINEETH KUMAR (11706849)

Introduction and background:
We choose to research LinkedIn job postings because we think they are very important in the current job market. The job market is changing quickly, especially after the pandemic. Both job seekers and employers are having a hard time dealing with this changing job market. We think that data from job postings can help us understand the skills that are in demand, how the job market is changing, and what new job opportunities are available. We were inspired by studies like " Online Job Postings: Analyzing, Visualizing, and Predicting" and "Text Mining and Visualization of Job Postings for Business Intelligence". These studies showed us how important it is to mine online job postings. We hope that our research will help job seekers and employers find each other more easily. We also hope that it will help us understand the job market better.

Datasets:
From our project we will be learning how to visualize the job dataset which we collected from some online resources. We framed different visualization questions to visualize our data in a story format which includes different charts. We think it will help others to get to know about how the job market is changing day by day, and also helps to know what skills companies are looking for in a candidate making it easier for the candidate to develop new skills according to the market demand. Along with that it also helps candidates to know about the location of companies offering jobs with specific skill sets. Once the candidate is recruited in order to clarify the dilemma of taking the offer or not, our visualization helps them to do some background check with the benefits offered by the company.

- We found the dataset for our project that is of 'LinkedIn Job Postings' through an online source on Kaggle which includes the important eight aspects required for both the employer who's providing a job and employee who are currently recruited as well as the candidates who are seeking to become an employee of the company.
- The dataset which we have taken for our project is from an open source available on Kaggle and so it does not have any particular publications associated with it.
- The dataset we used in our project consists of 8 different variables including job postings, companies name, industries in which the companies are established, varied specialties, number of employees, different benefits, industries of job and skills required for those jobs.
- The dataset which we have chosen for our project of LinkedIn job postings has covered the whole United States of America along with other countries. Yes, we will be able to make the graphs at the city as well as the state level as in the dataset it has given details regarding various cities as well as their states as for example Austin Tx, Los Angeles California and so on. Even we have the data of many countries and so with its help we can make the graphs on national and global levels as well.
- The dataset which we have utilized for our project is on the basis of a survey and it is for the job postings and thus there is no particular date, time or year provided in it but we have the pay period in which it is given that whether the employee in a particular company has provided the pay to him/her on the yearly , quarterly or monthly basis.

Data Stories:
Narration of Data Visualizations used in our project in a story format:

To begin with , the reason behind using different kinds of visualizations in our project is to show the wide range of possible graphs/charts using which we can make the lengthy data into attractive, user-friendly visualizations which provides clarity as well as simplicity for understanding information which is in general a bit cumbersome. The first visualization which we used in our project is 'Industry and Company Data' by using a simple tabular format to highlight the Industry as well as Company ID, along with its names and the palace where it is located describing its number of employees working in it. We have also added the filter for the Industry and Name so that the user on the basis of his/her need can utilize it to its optimum level. The tabular format has been taken into it as we have all the attributes placed on the rows which are all discrete values and thus it gives the best overview of the data. Now comes the second visualization of 'Number of Job Postings for Different Industries and Companies' and for this we have used bar graph highlighting the Name and Posting Domain both as discrete field in the rows and job posting in rows having continuous field by using Cnt as an aggregate function so that as a user if a person needs to know where to apply for a particular company they can directly use the filter of Posting Domain and Name provided in our visualization. The Bar chart concept taught in the class by professor under the Week 2 and 4 Modules have helped us to make our visualization in such a manner projecting the data to be interactive in the simplistic way and as we had attributes under both rows as well as columns we were able to choose bar graph in these. Third, is the visualization of 'Types of Worktype with Applies and Sponsor' by using Tree Map as learnt in the Week 4 Module in Basic Visualization ppt provided by the instructor as with the help of it we can display the complex amount of data in a small region having higher efficiency to understand. In this, we have considered using the Color, Size and Label Marks to make our visualization differentiate the full time, contractual and hourly worktype among the applies and sponsors. The size helped us to mold our data on a scale of high to low and the label used to highlight the numbers so that the reader can easily grab the information without having much work to do on his/her side. As an applicant it helps them to know what kind of job the company provides and whether he/she should apply to it on the basis of their requirements. Then comes the fourth visualization of 'Companies that are having Employee/Applies more than they used to' where we have taken the Name in the column and sum of applies and sum of employee count in rows, while having the filter on the Name and Employee count sum aggregate function for the ease of a user. For this visualization we have used the combination chart consisting the line and bar chart for comparison where the bars in blue color depicts the details of applies whereas the lines in orange color depicts the employee count while labeling the highest as well as the lowest values providing data to be compared across two categories over a period of time. The combination chart is very useful when we need to describe a large amount of data in the simplest way which was also learnt by us during Week 4 Module in Basic Charts ppt and implemented in our Lab Assignments which enhanced our knowledge to use the method in our project as well. The fifth visualization used in our project is 'Companies that are having maximum and minimum salary' for which again we have used the Bar chart but in a clustered manner where one bar chart highlights the companies having maximum salary and the other bar chart highlights the companies having minimum salary as we have gone into deeper concept of using bar chart from Week 2 Module to Week 4 Module and implementing the ways in our assignment it boosted our confidence to use it in the project as a result of which we came upon the idea of visualizing max and min of salary providing companies. Also, in this visualization we have used a dark background so that the color used to mention the data in bars can seem to be more efficient that is the purple and red color highlighting along with the labels of the amount of salary which helps

the user to know the expected salary are being met by which companies so that they can apply in those moreover it helps the companies as well to know about how much salary range the competitors provides to their employees. Sixth, visualization is about the 'Companies that offer remote work opportunities' in which we have used the aggregate function of Sum on the size and angle mark to show the total of companies offering remote working opportunities and the color mark is used for the Name attribute to differentiate the companies on the basis of their color. As we have learnt in the Week 4 Module under Basic Charts ppt we know that choosing Pie Chart is not very welcoming in data visualization under tableau but for visualizing the data which do not have much of the complex values pie chart seems to be appropriate and thus for our project in the sixth visualization we added pie chart to describe when we can use the same. The last second visualization we used in our project is of 'Company Benefits' where we have kept type and name as discrete values under columns and count aggregate function as a continuous field on the benefits attribute under rows. These visualization will be helpful for both the job seeker as well as giver as the employees now-a-days are not just acquiring job to have the salary but they are also interested into the perks provided by the companies and the employers can look and develop their strategies in a different manner so that the benefits provided by them can attract more applicants and retain the current employees. The last and eighth visualization is of 'States with maximum job postings' for which we have selected a geographical presentation of map. The longitude as well as the latitude has been set up on its own depicting the map of USA in which the State attribute have been used under Color and Label Mark which denotes the names of all the states of USA in different colors along with their name and the number of job posting are also labeled so that by looking at the visualization we can directly know the exact numbers and can see which state advertises the job of their specific companies through which we can find the maximum , minimum and even range of job offerings. The usage of geographical presentation has been taught to us in the Week 2 Module under Basic Chart ppt where we have experimented different types of maps such as density, symbol and filled maps and it has been also used during our Lab Assignments. Overall, the Graph Selection Matrix provided to us under Week 3 Module has been very helpful to learn and understand when and why we should use the specific graph for the visualizations. Thus, the whole story in one way or the other seems to be helpful for the employers , employees and applicants in this competitive world market as it provides ease, concise and proper description along with the filter option to adjust the data visualization according to their needs.

Summary and Conclusions:

In conclusion, we want users to get a better understanding of our "Linked In Job Postings Data set" by viewing our Visualizations. From the Visualization user can understand different aspects how job posting changing these days according to the location(states), domains(medical, technical, etc), types of work(full time,part time,etc), job posting application domain, applicants to posting ratio, Max and Min salary, what companies are allowing remote jobs.

REFERENCES:

1.(*LinkedIn Job Postings Dataset*, 2023)
https://www.kaggle.com/datasets/rajatraj0502/linkedin-job-2023
2.(., 2022)https://www.kaggle.com/datasets/dilshaansandhu/international-jobs-dataset
3.(*International Jobs Dataset*, n.d.)https://www.kaggle.com/datasets/thedevastator/global-it-jobs-analysis
4.(*Data Science Job Postings (Indeed USA)*, n.d.)https://www.kaggle.com/datasets/yusufolonade/data-science-job-postings-indeed-usa
5.(*LinkedIn_job_data*, n.d.)https://www.kaggle.com/datasets/shashankshukla123123/linkedin-job-data
6. (kaggle)https://www.kaggle.com/datasets/killbot/linkedin

CONTRIBUTION:

Topic Selection : Sowjanya Tejavath, Sai Pooja Pravallika A

Dataset: Sai Pooja Pravallika A, Sowjanya Tejavath

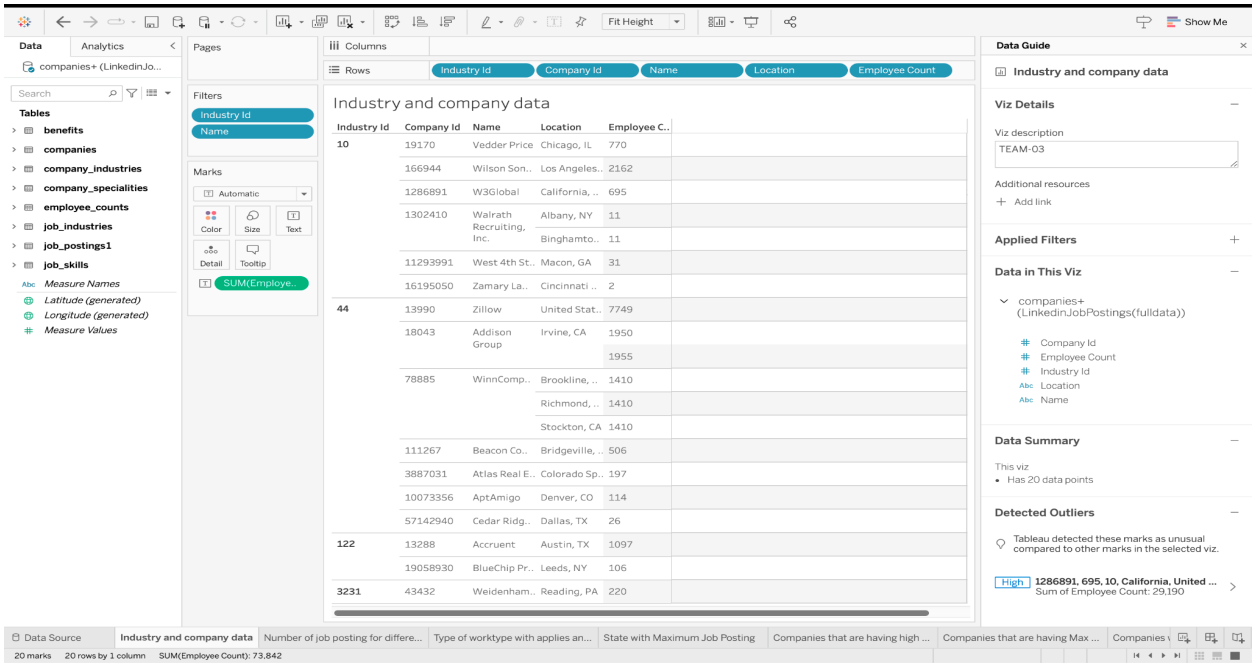
References: Gohith, Pradeep

Dataset Sorting: Pradeep, Arpana,Vineeth

Visualization plan: All team members(we all together brainstormed the ideas)

Execution in Tableau: All team members(each person is working with different charts)

1.INDUSTRY AND COMPANY DATA(Identify relationships between industry Id and their respective companies ?)



Motive and Variables used in the Visualization:

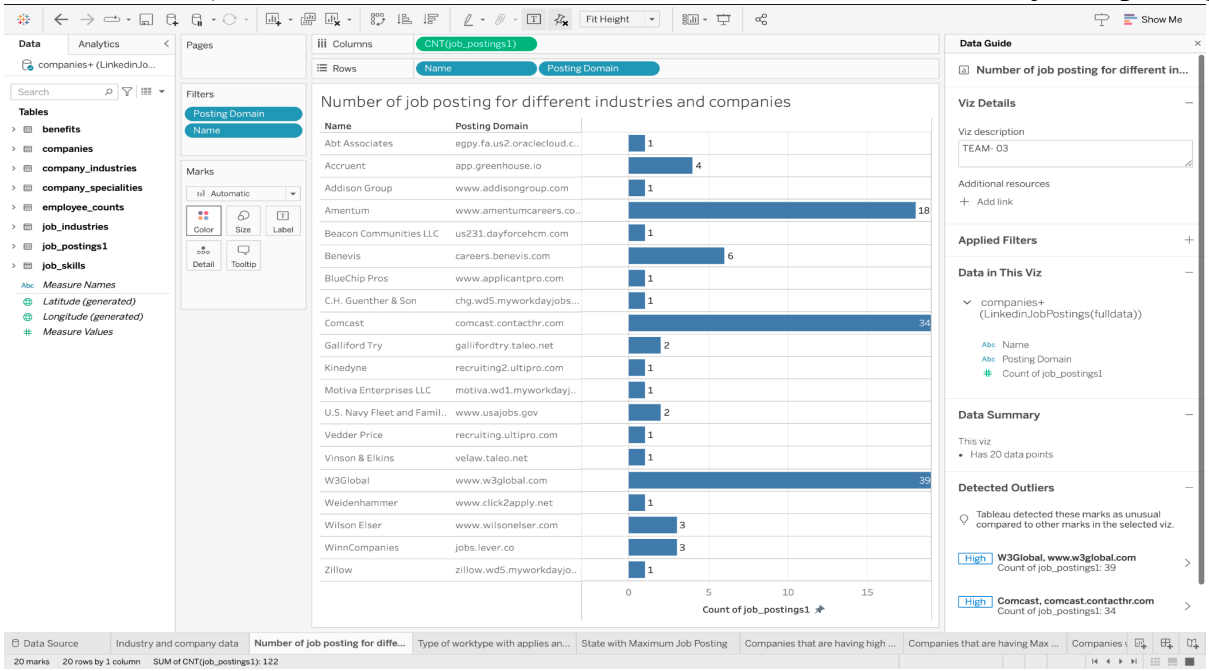
The main motive of the above visualization is to provide users with better understanding about how the “Industry and Company Data” works. From the visualization we can get information about Industry Id, Company Id, Name , Location and their corresponding employee count. Key Variables for the visualization are Industry Id and Name of the Employee, they help users to filter out as per the requirement of the data.

Principles in Visualization:

Principles used in the above visualization are 1.Clarity: The visualizations are clear and easy to understand. The data is labeled appropriately,and the axes are well-scaled. 2. Accuracy: The data is accurate and reliable. The data sources are properly cited, and the dashboard includes a data summary. This Table chart "Show Me the Numbers: Designing Tables and Graphs to Enlighten" by Stephen Few.

Data source for visualization: We have taken these attributes from the data set ‘ Linked In dataset’ provided through Kaggle.

2.NUMBER OF JOB POSTINGS FOR DIFFERENT INDUSTRIES AND COMPANIES:(What are the states that have maximum job posting?)



Motive and Variables used in the Visualization:

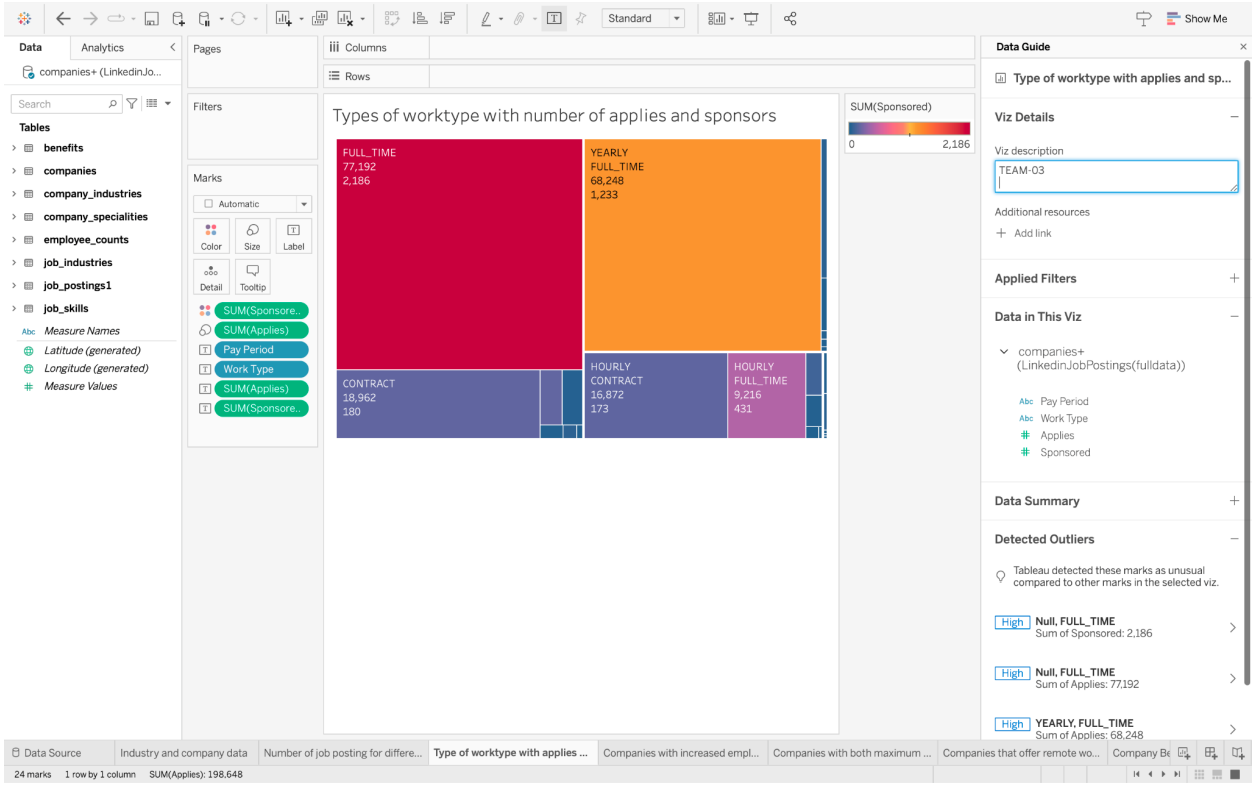
The main motive of the above visualization is to provide users with better understanding about "Number of Job Postings for Different Industries and Companies.We want people to understand the information about what companies (IBM, Infosys, Chase, etc) were offering what kind of industry sectors(accounting, marketing, data science,etc) with their number of openings to that particular position with the posting name. Key variables used in the visualization are Posting Domain and Name.

Principles in Visualization:

1.Relevance: The data is relevant to the topic of the dashboard, which is the number of job postings for different industries and companies. 2.Interactivity: The dashboard is interactive, which allows users to explore the data in more detail. For example, users can filter the data by industry, company, or location. Bar Chart "Information Dashboard Design: Displaying Data for At-a-Glance Monitoring" by Stephen Few.

Data source for visualization: We have taken these attributes from the data set ' Linked In dataset' provided through Kaggle.

3. TYPES OF WORKTYPE WITH APPLIES AND SPONSORS(what are the number of job posting domains for different industries and companies?)

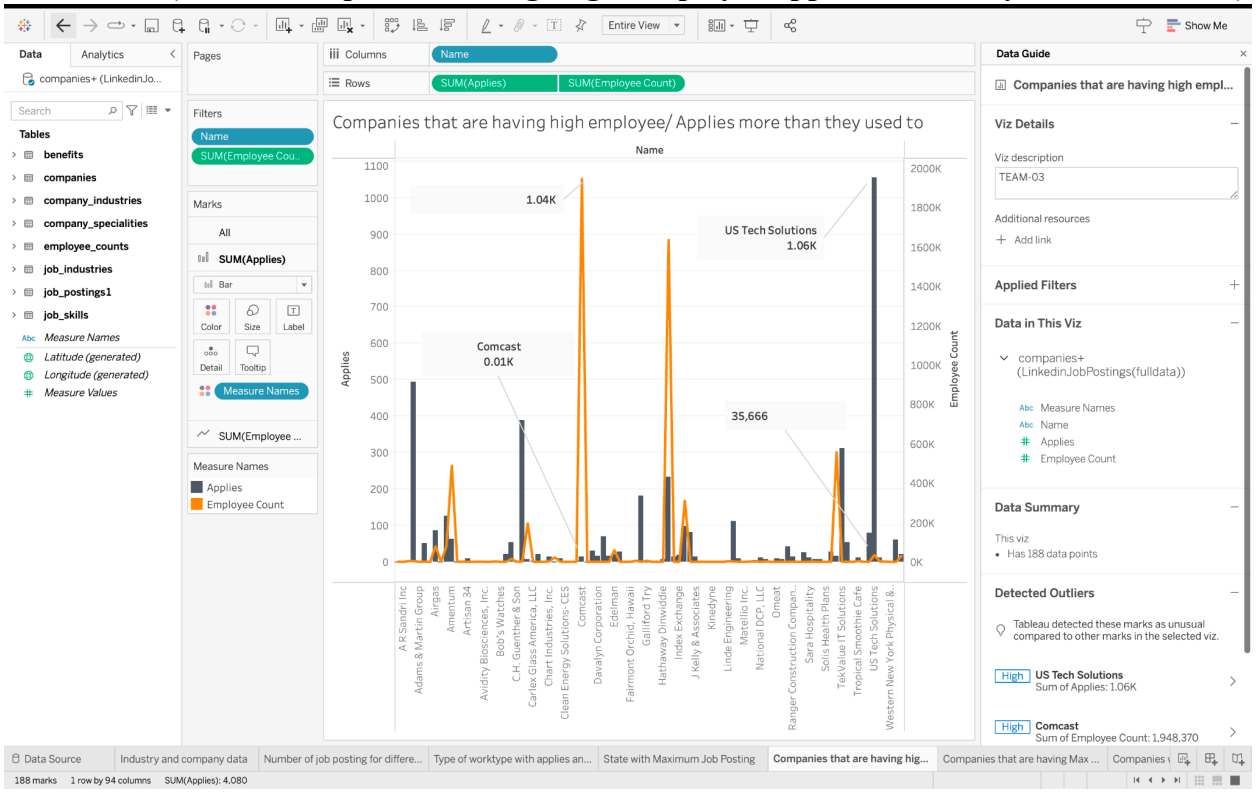


Motive and Goals of visualization: The motive for the above chart is to let people know there are different types of work such as full time, hourly contract, etc. And also there are specific applications for those work types associated with different numbers of sponsors.From the chart we can say that there are more full_time opportunities and less volunteers. Key Variables used for the visualization are Work Type and their payment period.

Principles in Visualization:1.Clarity: The visualizations are clear and easy to understand. The data is labeled appropriately,and the axes are well-scaled. The titles and subtitles are informative and help to explain the data.2.Interactivity: The dashboard is interactive, which allows users to explore the data in more detail.For example, users can filter the data by type of work, number of applies, or number of sponsors. utilizing a TreeMap chart."Data Points: Visualization That Means Something" by Nathan Yau.

Data source for visualization: We have taken these attributes from the data set ‘ Linked In dataset’ provided through Kaggle.

4.COMPANIES THAT ARE HAVING EMPLOYEE/APPLIES MORE THAN THEY USED TO (Which companies having high employee/applies that they used to have?)



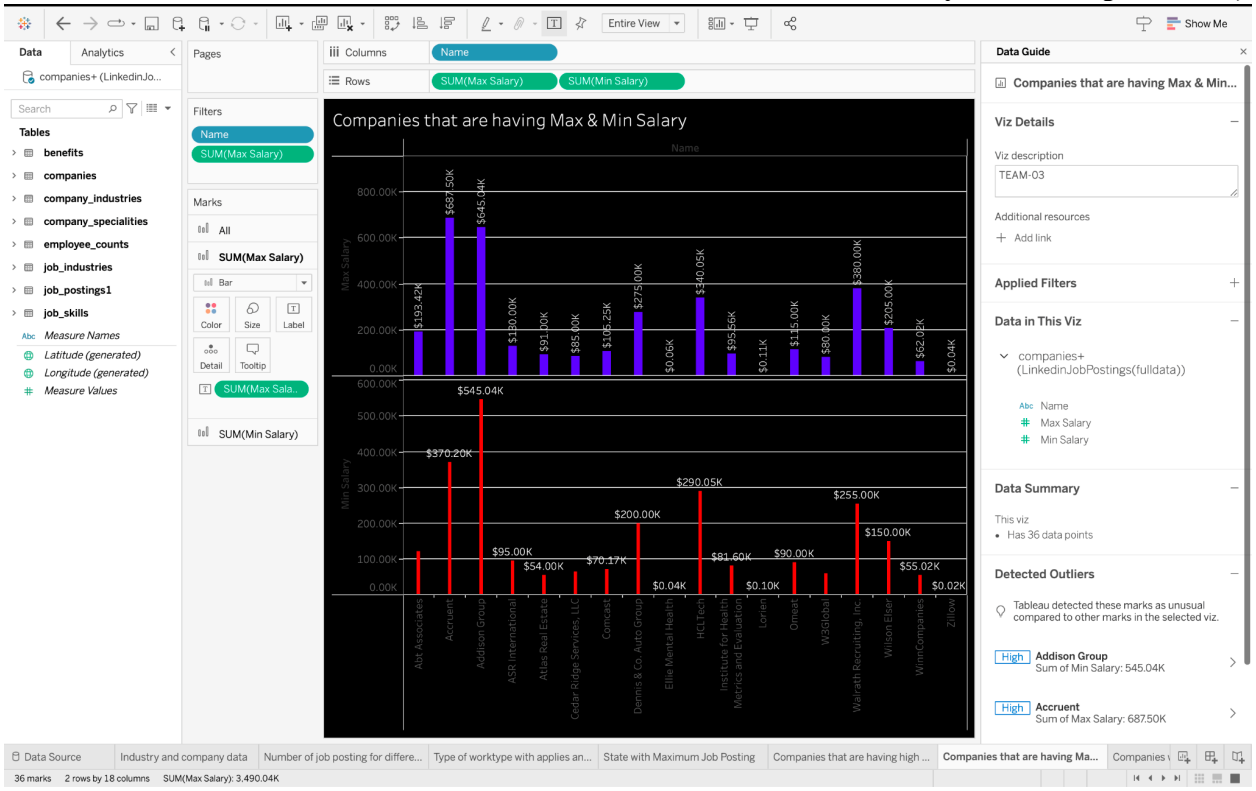
Motive and Goals of visualization: The main motive for the above visualization is to get a better understanding about the company's employee count to the ratio of Number of applications to each company. For better understanding and to differentiate we used the combination of bar and line chart. Key variables used for the Visualization are Applicants and Employee count with respect to names.

Principles in Visualization:

Clarity: The visualizations are clear and easy to understand. The data is labeled appropriately, and the axes are well-scaled. The titles and subtitles are informative and help to explain the data. Accuracy: The data is accurate and reliable. The data sources are properly cited, and the dashboard includes a data summary. Line Chart & Bar Chart"The Visual Display of Quantitative Information" by Edward Tufte.

Data source for visualization: We have taken these attributes from the data set ‘ Linked In dataset’ provided through Kaggle.

5.COMPANIES THAT ARE HAVING MAXIMUM AND MINIMUM SALARY(Provide a visualization which shows about the max and min salary of companies ?)

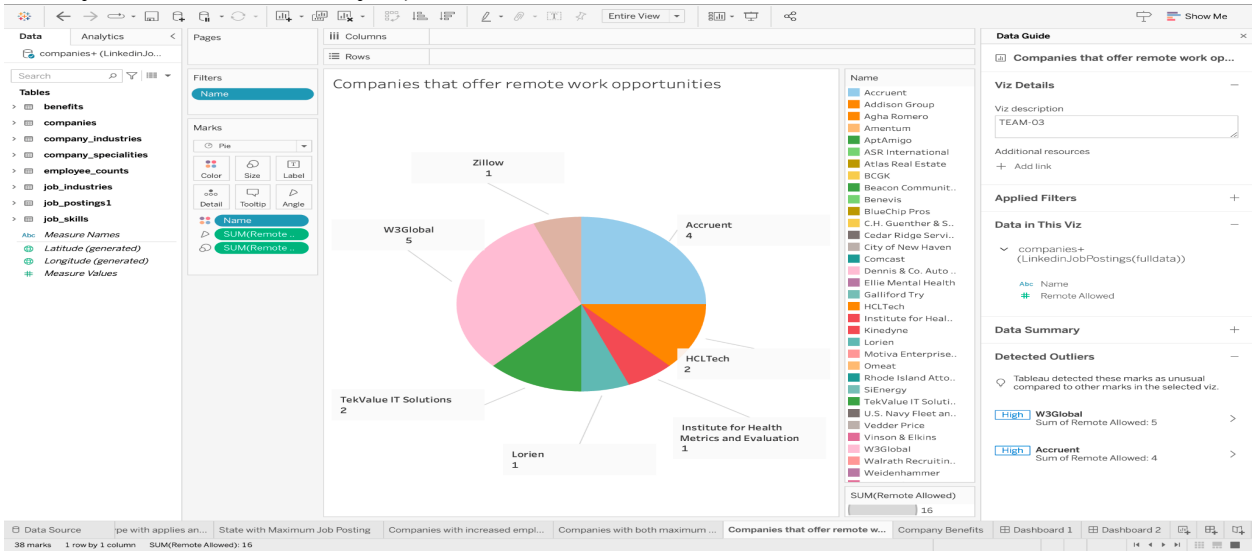


Motive and Goals of visualization: The main motive for the above visualization is to get a better understanding about the average salary range for job postings at major companies. The salary range is displayed along the y-axis, with the company on the x-axis. Key Variables used for the above visualization are maximum and minimum salary with respect to each company name. The chart gives a high-level view of compensation at major companies hiring on LinkedIn.

Principles in Visualization:Clarity: The visualizations are clear and easy to understand. The data is labeled appropriately, and the axes are well-scaled. The titles and subtitles are informative and help to explain the data. Accuracy: The data is accurate and reliable. The data sources are properly cited, and the dashboard includes a data summary. Line Chart "The Visual Display of Quantitative Information" by Edward Tufte.

Data source for visualization: We have taken these attributes from the data set ‘ Linked In dataset’ provided through Kaggle.

6.COMPANIES THAT OFFERS REMOTE WORK OPPORTUNITIES(which companies allow you to work remotely ?)



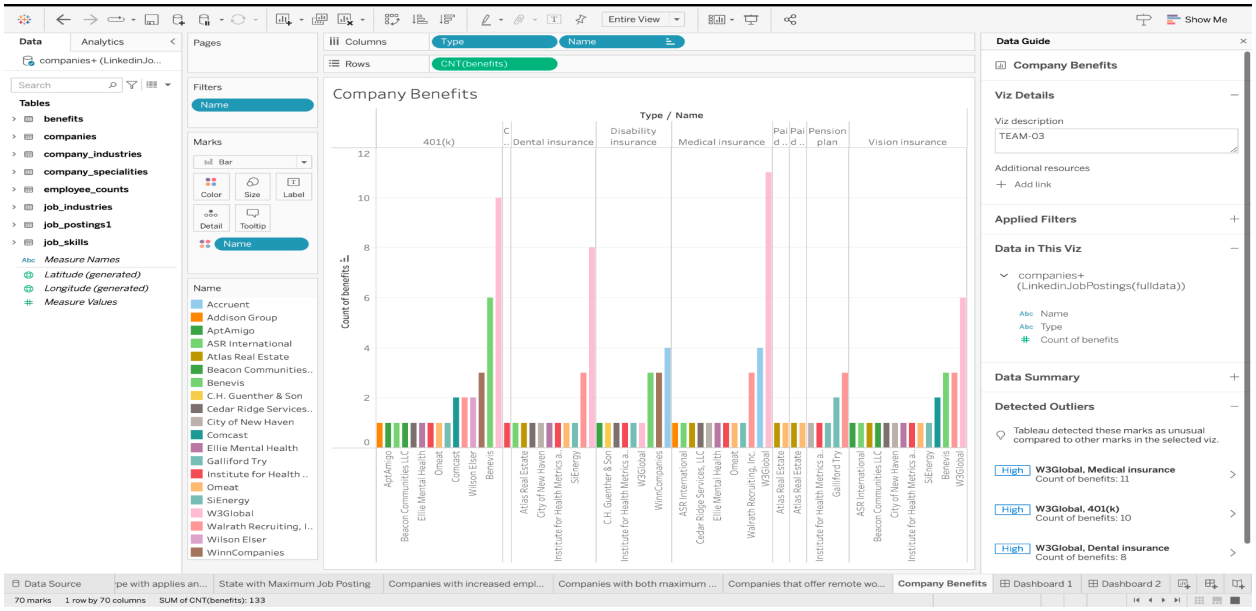
Motive and Goals of visualization:The main motive for the above visualization is to get a better understanding about the remote work tag indicating the job posting explicitly states the role can be done remotely or virtually. Visualization displays the percentage of job postings from each company that are listed as offering remote work. Key Variables used for the visualization are Remote Work allowed.

Principles in Visualization:

1.Relevance: The data is relevant to the topic of the dashboard, which is the number of job postings for different industries and companies. 2.Interactivity: The dashboard is interactive, which allows users to explore the data in more detail.For example, users can filter the data by industry, company, or location. Pie Charts "Storytelling with Data" by Cole Nussbaumer Knaflic.

Data source for visualization: We have taken these attributes from the data set ‘ Linked In dataset’ provided through Kaggle.

7.COMPANY BENEFITS(Show the visualization for the company benefits ?)



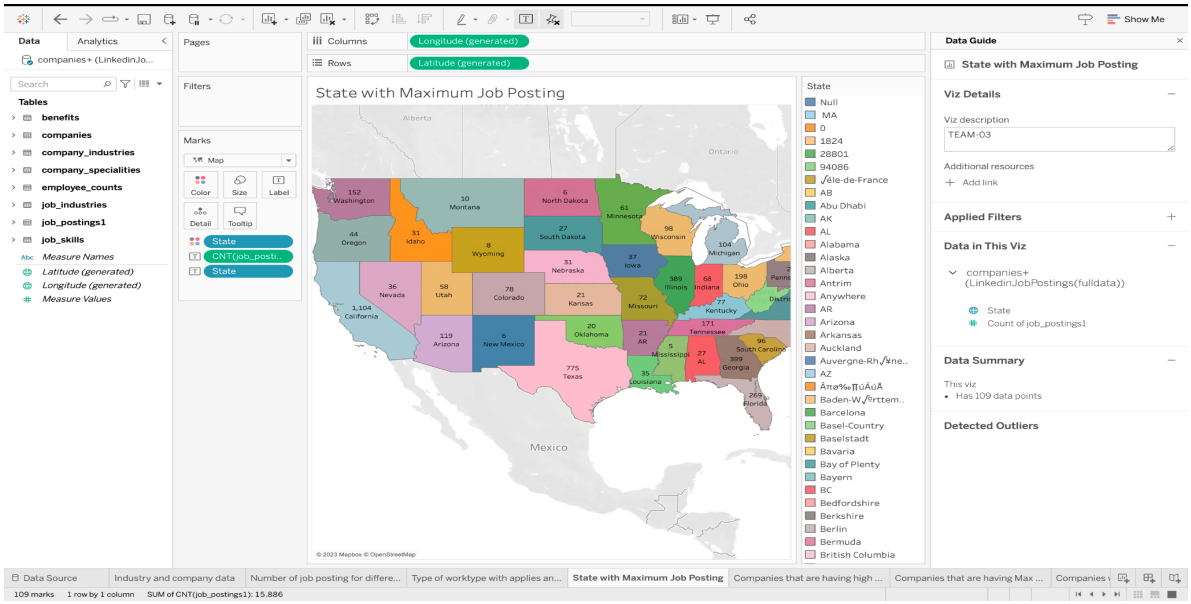
Motive and Goals of visualization: The motive for the above visualization is to understand the companies and their benefits to the employees. From the Visualization we can see the count of benefits offered in each job type(Medical,vision,etc) by every other company. Key Variables used in the chart are Benefits, Type and name of the benefits and company.

Principles in Visualization:

The visualization uses a variety of data visualization principles, including simplicity, clarity, accuracy to the better understanding of the visualization. Line Chart "The Visual Display of Quantitative Information" by Edward Tufte.

Data source for visualization: We have taken these attributes from the data set ‘ Linked In dataset’ provided through Kaggle.

8.STATES WITH MAXIMUM JOB POSTINGS(What are the states that having maximum job posting form a question?)



Motive and Goals of visualization: The motive for the above visualization is to understand the job postings in each state. From these we can say that the most job postings were in Texas and least job postings were from North Dakota. Key Variables are state, job postings from attributes job_industries, job_postings.

Principles in Visualization:

1.Colour: Colour is used to differentiate each state for clarity of the visualization. Accuracy: The data is accurate and reliable. The data source is properly cited, and the visualization includes a data summary. Map chart "Making Maps Third Edition: A Visual Guide to Map Design for GIS" by John Krygier and Denis Wood.

Data source for visualization: We have taken these attributes from the data set ‘ Linked In dataset’ provided through Kaggle.