Dataset Description:

Naukri.com is an Indian Job Portal that holds a huge database of the jobs, resumes and recruitment consultants. This website was considered as a common platform for the jobseekers and hiring people to come together. A dataset was obtained by pre-crawling this website and it is a subset of the bigger dataset which has got more than 9.4 million job listings created by extracting from the website. The resultant dataset has got 22K records of job descriptions. This dataset is the recently updated dataset that conveys latest information about the availability of jobs. This dataset has never been visualized which sets the foundation for the creativity in visualizing the data. The tool that is used for visualizing the dataset is D3js which is a java script library used for creating the interactive and dynamic visualization.

The dataset consists of the following features:

* Company
* Education
* Experience
* Industry
* Job Description
* Job Id
* Job Location Address
* Job Title
* Number of Positions
* Pay Rate
* Post Date
* Site Name
* Skills

Business Scenario:

**Claims:** The ministry of Labor and Employment in India releases the monthly status report of the number of jobs created by them in different sectors in order to tackle the unemployment issue. But sometimes this report doesn’t match with the reality and this takes many months to come into notice of others. In order to approve/disapprove the claims given by the ministry, I have created a visualization that gives detailed information about the number of jobs available in all the sectors over the major cities in India. If the results from the visualization matches with the report then the claims are approved or else, it can be disapproved.

Data Preparation:

The dataset consists of many columns like Job Id, company, experience etc. which are not suitable for the case study that has been proposed. So, a subset was obtained from the original by selecting only the columns that are mentioned below:

* Job Location
* Industry
* Skills
* Total Number of jobs

Once the features are selected, there are certain records where a certain job is posted for more than one location and this is represented as “Chennai, Delhi, Hyderabad, Bengaluru”.

In this case this particular record is duplicated into four records but with different job location details in order to avoid the loss of information. The names of the job locations where not standard and they were many different names for the same city that would result in the repetition

of information while visualizing it. For E.g. Chennai and Madras are new and old name for that city so they are standardized only to new name. The records are selected in such a way that the all the four features are present and the if the number of jobs field was empty then it is replaced by one in order to be suitable for calculation.

Visual Tasks:

The dataset was visualized in the form of a tree so that it can be easy to drill down each level according to the need. The visualized tree has got 4 levels nested in it. They are

* First Level speaks about cities
* Second Level speaks about Industries
* Third Level speaks about Skills
* Fourth Level speaks about Total number of Jobs

The visualization was made interactive by giving the drill down flexibility to the user by clicking on every node so that they can avoid looking into the details which are not required. Some animation effects like collapsing the sublevels and expanding the size of the last levels was given so that the users attention gets increased on that sublevel or last level. The visualization falls under the discover, search and derive. In discover task we have stated the hypothesis which is the claim made by the ministry and this is verified true or not from the visualization. In search task the user picks the aspect which they are interested and proceed further. In derive task the variable total number of jobs is derived from the number of jobs variable.

Visual Encoding:

All the nodes were indicated using circle with a label(text) describing about the node and the paths represented in the form of lines were used to indicates the connection between the nodes. All the levels are differentiated with increase in size and also the brightness of the color with respect to the depth of the levels. The last level which indicates the total number of jobs is presented in the form of an expanding circle with count embedded within it and the size of this circle is large which compared to other levels. The path which they traversed through the tree has been highlighted using green color which makes the user easier to traverse back and forth since the visualization is pretty huge and this helps in differentiating from the other paths.