Final Presentation : Visualization and Analysis

Sungha Kang
Jeongin Choi
Yue Tong Chang

Original / Cleaned Dataset :
https://www.kaggle.com/datasets/iamsouravbanerjee/software-professional-salaries-2022?select=Sa

Intro: Data set and Background

The data set we are going to analyze is 'Software Professional Salary 2022' from Kaggle. It shows us employees who work in the IT sector in India and how much they earn for their job. It provides us with detailed roles for software jobs, in what territory they are specialized in and its employee status. With this data set, we wanted to analyze job markets regarding software and IT industries.

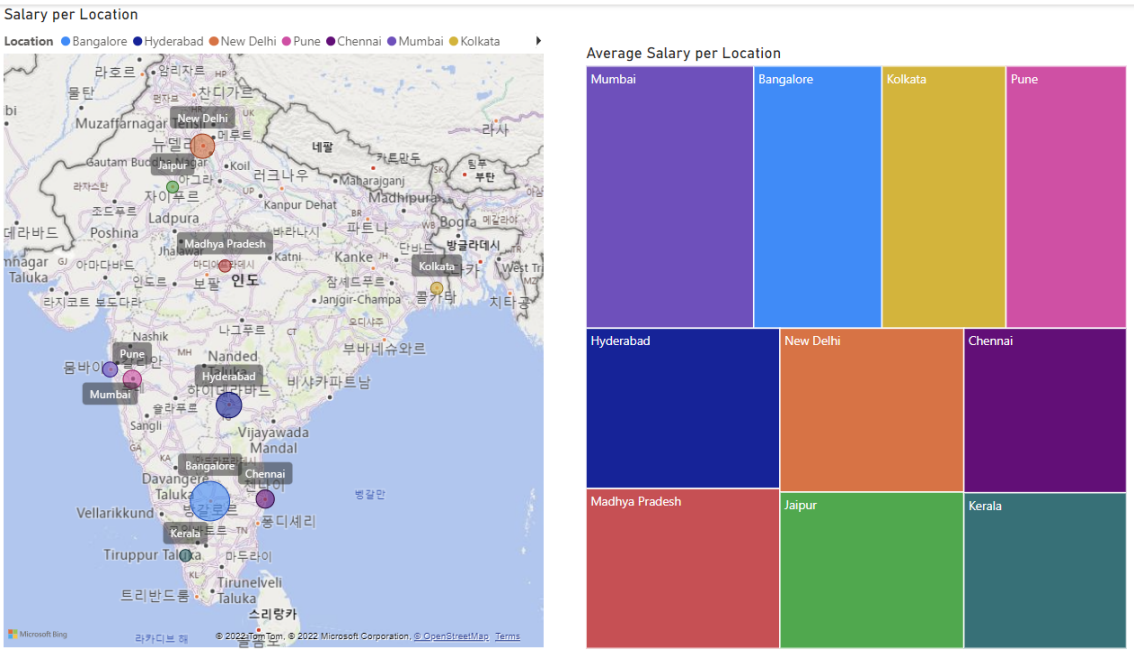
Now, let's look into the data. Our data is actually not very complicated. It has 7 columns and about 22,000 rows which is enough for statistical analysis and to get the generalized conclusion. Columns consist of rating which reflects employees' satisfaction with their employer, company name, company location and salary. It also contains job title and job role which is a little confusing. Job title is actually very similar to employment status plus job roles.

With this data, we wanted to focus on three things.
First, we will look into the relationship between company plus location and salary. We will find out differences between Indian regions.
Second, we will analyze salary differences among differing job roles and employment status.
Finally, we will look into salary differences regarding their job role.

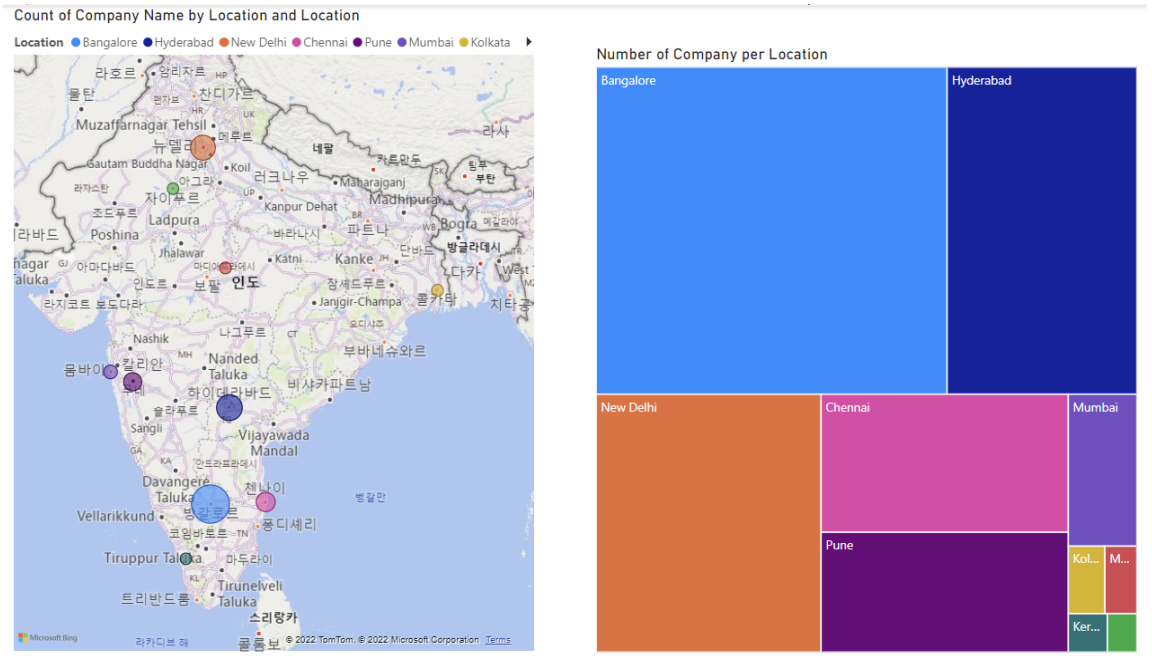
Rating	Company Name	Job Title	Salary	Salaries R	Location	Employment Status	Job Roles
3.8	Sasken	Android Developer	400000	3	Bangalore	Full Time	Android
4.5	Advanced Millennium T	Android Developer	400000	3	Bangalore	Full Time	Android
4	Unacademy	Android Developer	1000000	3	Bangalore	Full Time	Android
3.8	SnapBizz Cloudtech	Android Developer	300000	3	Bangalore	Full Time	Android

1. The Correlation Between Company, Location and Salary

Interactive Visualization :
https://app.powerbi.com/groups/me/reports/b58c82ce-a59e-45ec-b460-384a69c446ae/ReportSection64c188ab93b785ed8b50



Above map shows the region and its average salary. Mumbai is the one with the highest salary(961,180INR, US$ 12,037) while Kerala is the one with the lowest(553,577INR, US$ 6,932). This gap is huge considering the most common salary in India is 546,509 INR. We can say that there is a considerable gap between salaries in IT jobs according to the region.

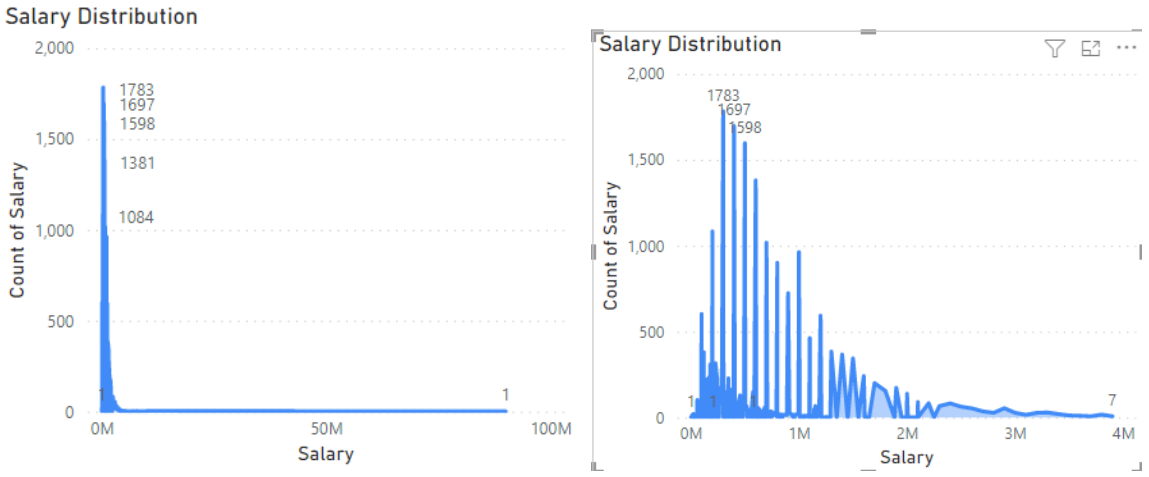


Second map shows the number of the companies that are located in the region. Actually, there is more eminent difference regarding the number of the companies than the average salary. You can see that Bangalore is home for so many companies while Kolkata, Madhya Pradesh, Kerala and Jaipur have less companies.

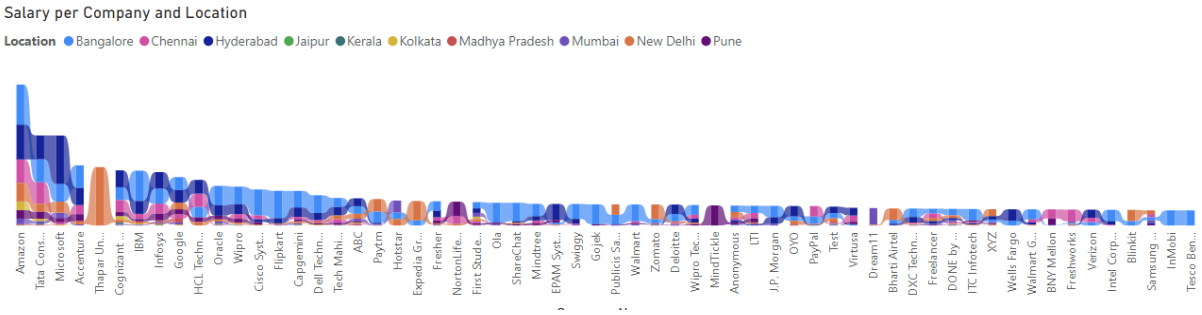
Mumbai was first to have a high salary while it hosts less companies. Kolkata has the same tendency with Mumbai. We can say that those two regions have good quality jobs. Bangalore was first to have a number of companies and second to have high average salary so we can say that if you are a job seeker who has a moderate skill set and hopes for decent pay, Bangalore is the best place for you to get a chance. Otherwise, if you have an extraordinary skill set and are looking forward to an affluent salary, you'd be better off going to Mumbai.

```
     Rating       Company.Name         Job.Title            Salary          Salaries.Reported
 Min.   :1.000   Length:22770       Length:22770       Min.   :    2112   Min.   :  1.000
 1st Qu.:3.700   Class :character   Class :character   1st Qu.:  300000   1st Qu.:  1.000
 Median :3.900   Mode  :character   Mode  :character   Median :  500000   Median :  1.000
 Mean   :3.918                                         Mean   :  695387   Mean   :  1.856
 3rd Qu.:4.200                                         3rd Qu.:  900000   3rd Qu.:  1.000
 Max.   :5.000                                         Max.   :90000000   Max.   :361.000
   Location         Employment.Status    Job.Roles
 Length:22770       Length:22770       Length:22770
 Class :character   Class :character   Class :character
 Mode  :character   Mode  :character   Mode  :character
```

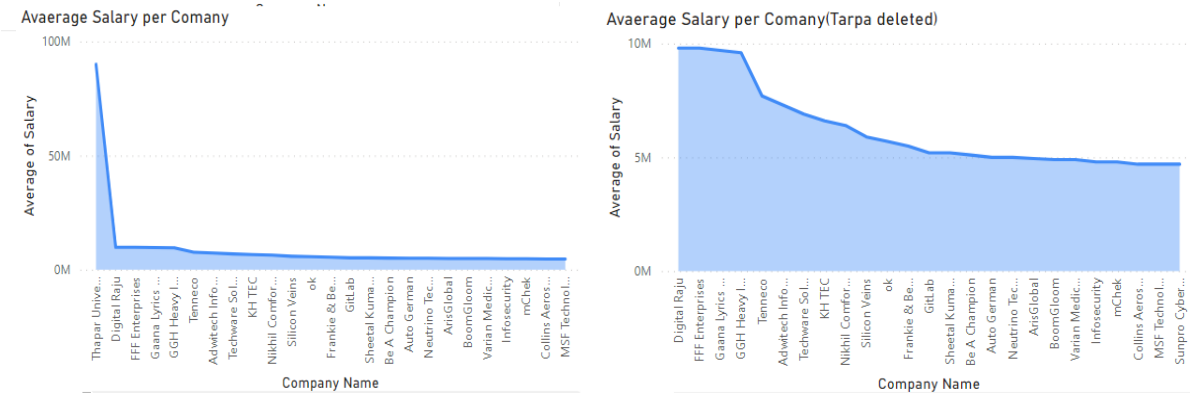
The lowest salary(Inuit, Bangalore) was $2,112 while the maximum salary(Thapar University, New Delhi) reached $90,000,000. The distribution of the salary is on the bellow graph.



For there are too many values that are reported only once in high salary range, I cut off salary range over 4 million to make graph more clear. In the graph on the right side, we see salary is mostly distributed between zero to one million.The mode of the salary was 300,000IND and median was 500,000IND. The Mean of salary was about 690,000IND. We can infer that high-salary made the average salary go up.



There are many small businesses that hire only one employee while big companies such as Amazon, Tata Consultancy Services, Microsoft, Cognizant Technology Solutions, Infosys, Google, HCL Technician, Accenture, IBM and so on hire a number of employees. Global companies hire employees from all around India and its ratio is marked with colors. From the chart, we can easily find out that Amazon is hiring about 80 employees from 8 regions while Thapar University hires its employees only from New Delhi. Blue which stands out for Bangalore is easily seen in most of the companies since Bangalore is no.1 home for so many companies in India.

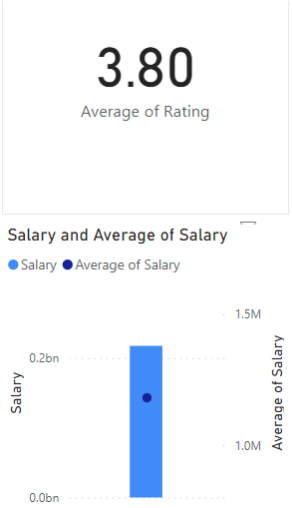


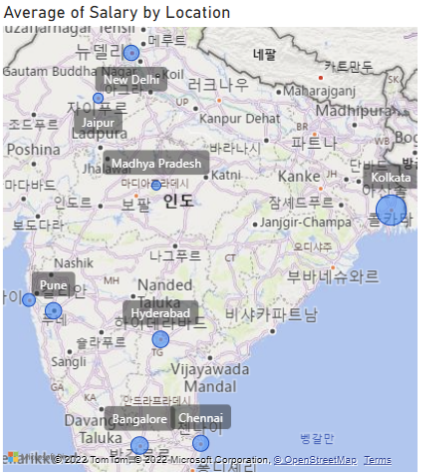
Above graph shows how much an employee earned from the company. Tarpa University gave so much of the salary to its employees that it made the line very flat elsewhere. So I deleted Tarpa University and there we see which company pays well for its employees. Here, we see companies with high average salaries.
(graph)
However, we should be careful because it is rather one high salary sata rather than the average of salaries. Because there are so many companies with the salary record reported only once, I filtered it to exceed ten for salary reported.

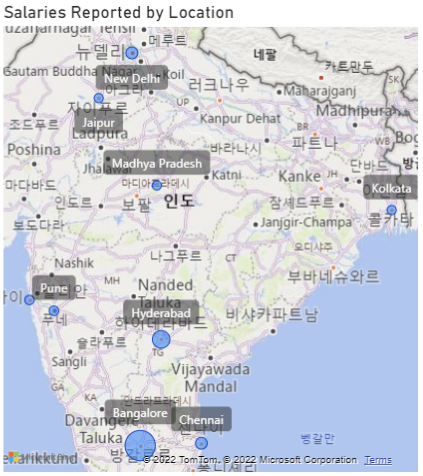
Now it's turn to focus on how much global companies are paying salaries in India. First I picked the top 5 companies. I arranged the companies by their average salaries. Next, I filtered them removing those with reported salaries under 30 to focus on big companies with many employees. Amazon, Microsoft, Accenture, IBM, and Google are the top 5 companies to be analyzed.

It will cover Ratings, total expenditure on work cost(bar graph) and average salaries(dot on the bar graph). Also, the map shows the regional average and the number of the employees there. It is ordered by total expenditure on work cost.

Amazon

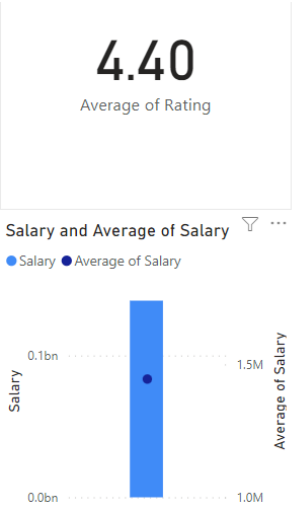


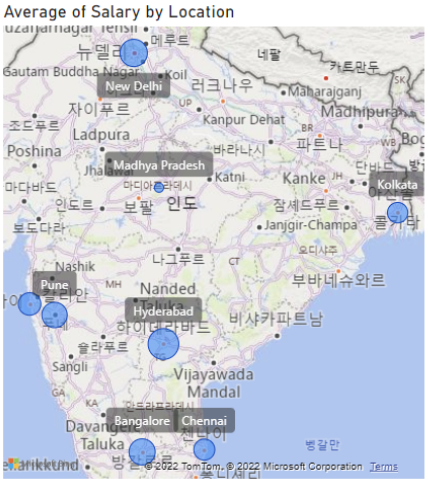


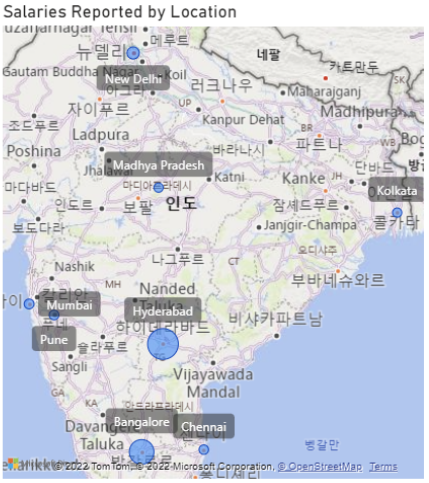


Amazon was top for expansion for its employees. However, ratings which show employees' satisfaction for the company were relatively low. Its overall expansion for labor cost was over 0.2 billion IND. Its average salary was the highest reaching over 1.4 million IND. However, rating was the lowest among 5 companies. Its average salary is high in Kolkata whil number of employees is small.

Microsoft

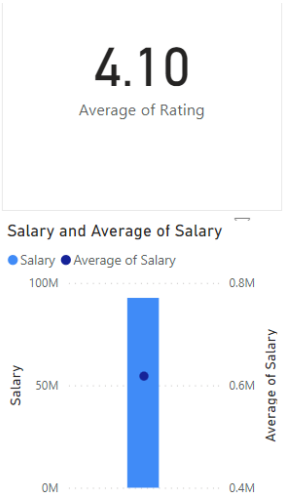


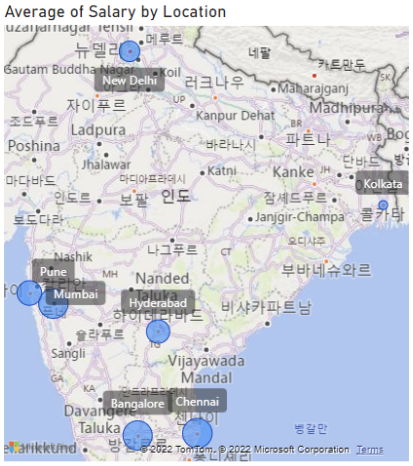


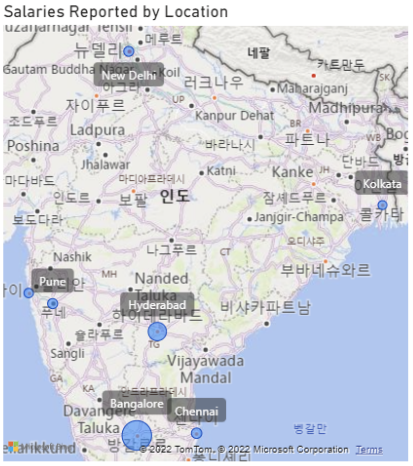


Microsoft spent over 0.13 billion IND and average salary is about 1.4 million IND. Its average salary was high in New Delhi.

Accenture

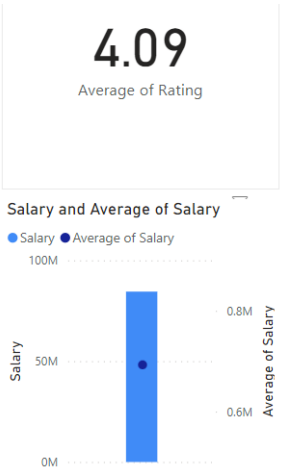


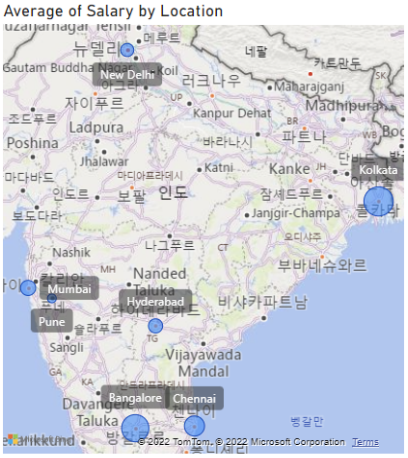


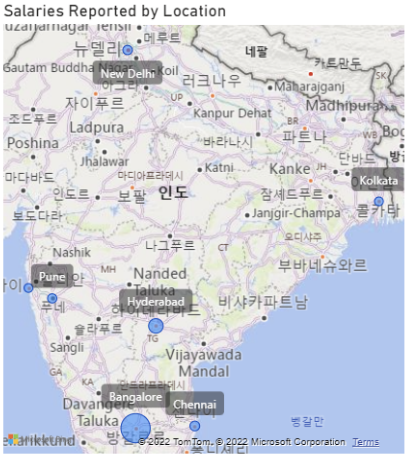


Accenture shows third-most highest expenditure on total work cost. Its salaries are relatively uniform all over the region.

IBM

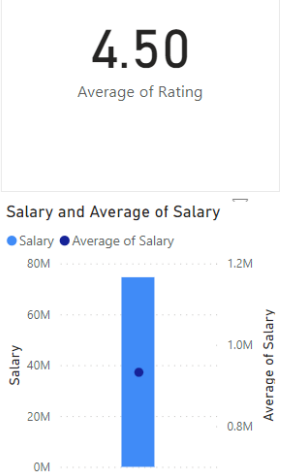


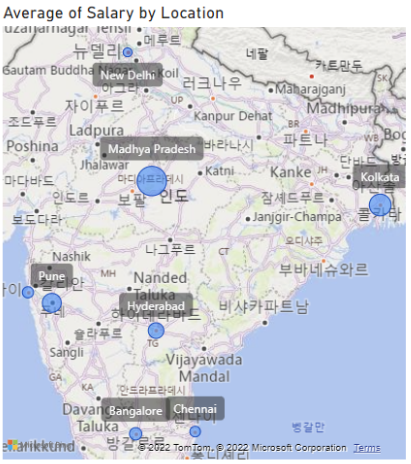


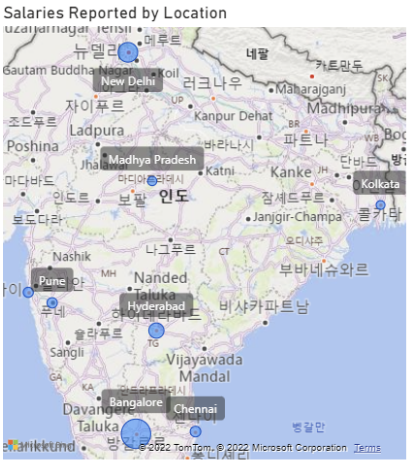


IBM displays perceivable differences according to the region. Its expenditure is high in Kolkata and low in Hyderabad.

Google







Google's workforce in Madhya Pradesh earns the most.

Company	**Total expenditure on work cost**	**Average Salary**	**Std(all employees)**	**Region with high avg salary(IND)**	**Region with a number of employees(number)**
Amazon	0.21billion	1,181,222	900036	Kolkata(27,080,000)	Bangalore(1545)
Microsoft	0.13billion	1,444,881	915670	Hyderabad(15,911,48)	Hyderabad(328)
Accenture	92million	618,120	891137	Chennai(700,400)	Bangalore(538)
IBM	84million	692,803	892541	Kolkata(825,000)	Bangalore(324)
Google	74million	932,500	893978	Madhya Pradesh(1,925,333)	Bangalore(114)

Bangalore has so many job opportunities there while its average salary is also high(Second)

1. Amazon

Location	Average of Salary
Kolkata	2708000.00
Bangalore	1289500.00
Hyderabad	1191466.67
Pune	1179636.36
Chennai	1133056.00
New Delhi	1056888.89
Mumbai	781818.18
Jaipur	468000.00
Madhya Pradesh	363800.00

Location	Salaries Reported
Bangalore	1545
Hyderabad	616
New Delhi	207
Chennai	196
Pune	34
Mumbai	26
Kolkata	6
Madhya Pradesh	4
Jaipur	2

2. Microsoft

Location	Average of Salary
Hyderabad	1591148.94
New Delhi	1420720.00
Bangalore	1359800.00
Pune	1318400.00
Mumbai	1174857.14
Chennai	1098666.67
Kolkata	1068000.00
Madhya Pradesh	500000.00

Location	Salaries Reported
Hyderabad	328
Bangalore	235
New Delhi	46
Mumbai	10
Pune	7
Chennai	5
Kolkata	1
Madhya Pradesh	1

3. Accenture

Location	Average of Salary
Chennai	700400.00
Pune	698933.33
Bangalore	692800.00
Mumbai	566666.67
Hyderabad	533520.00
New Delhi	460000.00
Kolkata	150000.00

Location	Salaries Reported
Bangalore	538
Hyderabad	244
Chennai	40
Pune	30
New Delhi	27
Mumbai	14
Kolkata	1

4. IBM

Location	Average of Salary
Kolkata	825000.00
Bangalore	792745.76
Chennai	679333.33
Mumbai	600000.00
Hyderabad	584117.65
New Delhi	566666.67
Pune	498285.71

Location	Salaries Reported
Bangalore	324
Hyderabad	92
Chennai	17
New Delhi	11
Pune	9
Kolkata	2
Mumbai	2

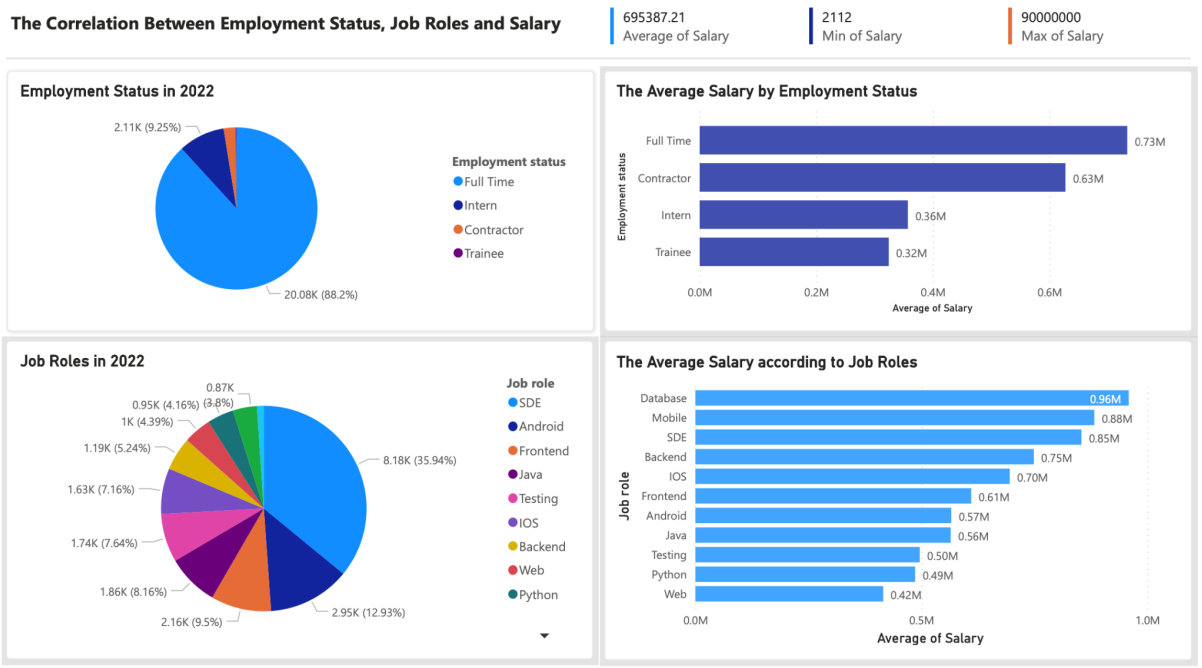
5. Google

Location	Average of Salary
Madhya Pradesh	1925333.33
Kolkata	1450000.00
Pune	1276000.00
Hyderabad	1069777.78
Bangalore	868749.91
Mumbai	800000.00
Chennai	762666.67
New Delhi	651048.00

Location	Salaries Reported
Bangalore	114
New Delhi	60
Hyderabad	36
Chennai	11
Pune	9
Mumbai	8
Madhya Pradesh	3
Kolkata	1

2. The Correlation Between Employment Status, Job Roles and Salary

Interactive Visualization :
https://app.powerbi.com/links/j0t0pA08MS?ctid=1a034549-971a-474d-bd7f-fd0b38f0060b&pbi_source=linkShare&bookmarkGuid=f4e2a38e-2a41-4914-9344-51247c542193



*If you go to the link and press each variable(ex. full time), it shows how much that variable occupies in the other chars. The analysis below is based on it.

The highest percentage (88.2%) of the four employment statuses are full-time workers, and the highest percentage (35.94%) of the job roles are SDE(Software Development Engineer). The average wage for the entire sample is $695,387.21, with a minimum of $2,112, and a maximum of $90,000,000.

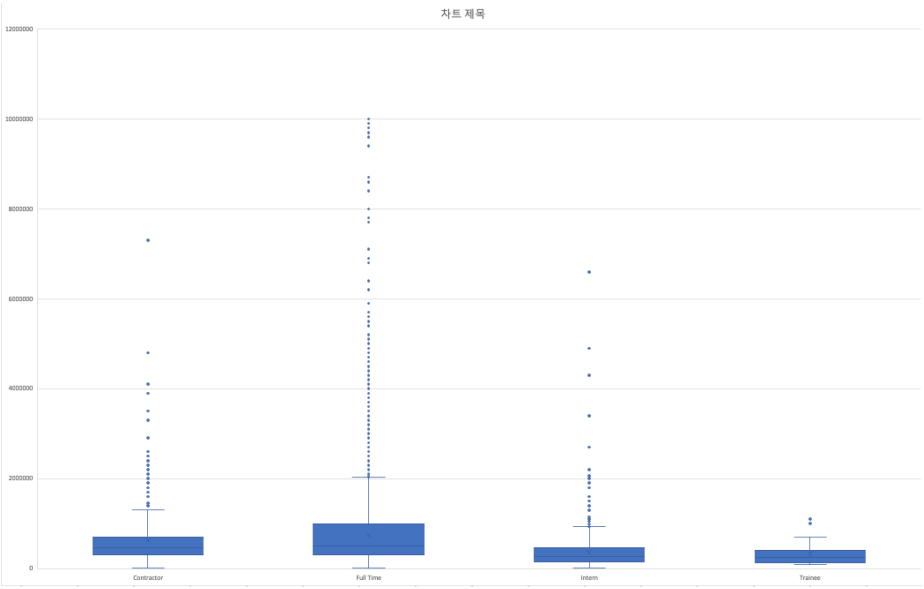
Other job roles are shown to earn similar to or less than the average salary, but in the case of SDE, they tend to earn higher than the average wage. However, although there are very few jobs that perform Database and Mobile tasks, the highest wages are shown (than SDE). These two jobs are expected to be professionals who require high education or require difficult work performance.

Another thing to look at is that contractors account for a much smaller percentage than the intern, but since they are in the upper class in company, they make much higher profits. In particular, contractors in the Backend sector earn the highest wages. In the case of Interns, it is confirmed that they earn similar wages regardless of job roles, but the SDE and Testing interns are making relatively high profits.

The two charts on the left used pie charts to represent the proportion and size of all samples efficiently, and the other charts on the right used bar charts to compare salary differences through height/length.

Box plot

It's the box plot for employment status - salary. It's hard to read the distribution with a box plot because there are outliers that are completely out of range, so I think it's more effective to visualize it with a bar chart.



It's also similar if you remove the outlier(maximum value; 90,000,000).

Now, let's consider the implications of making those visualizations interactive. It might seem that making content and data interactive is more about keeping up with trends, but actually adding interactivity to your visualizations is profoundly powerful. Interactive visualizations enable the following benefits:

1. **Multiple questions:**

Interactivity allows us to pose multiple questions per visualization – allowing us to switch axes or to add confabulating factors, which could be very useful for survey data, and allowing viewers to break down responses to a specific question by gender, age or perhaps occupation.

2. **Focus on detail:**

Interactivity allows users to zoom into a visualization – physically selecting an area of interest and blowing up that area of the chart. Hover information is also incredibly useful – if you're built a choropleth that visualizes the relative populations of different countries, allowing users to hover over a country to get the exact population value is much more useful than simple providing a color gradient legend.

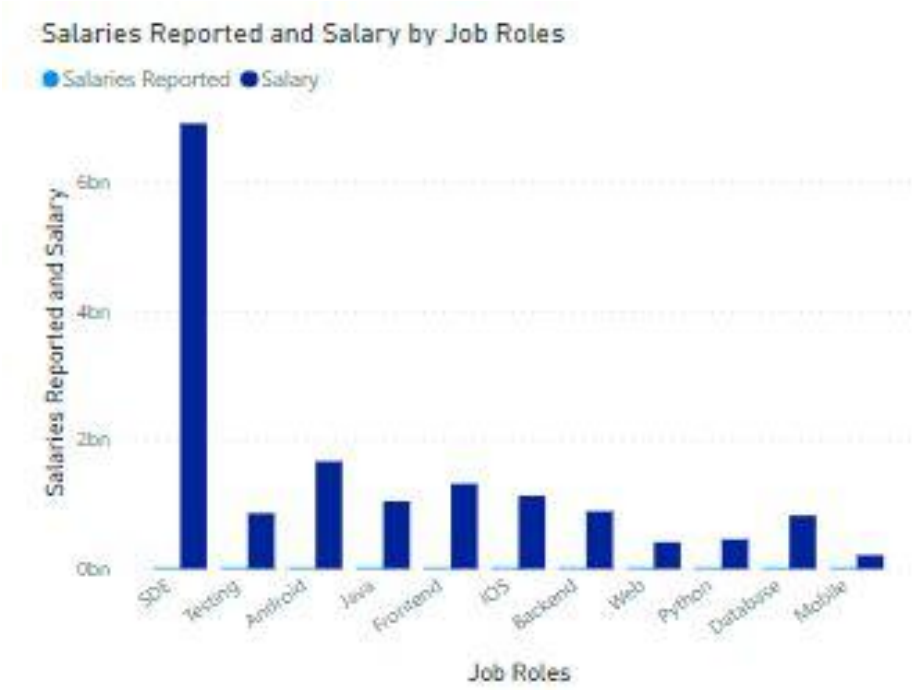
3. **User experience:**

Interactivity is particularly powerful when you allow users to select points or series in a chart and for a summary of the relevant data to appear. This allows us to build clean looking visualizations that are actually very rich and provide a tool for viewers to explore your data as little or as much as they're interested.

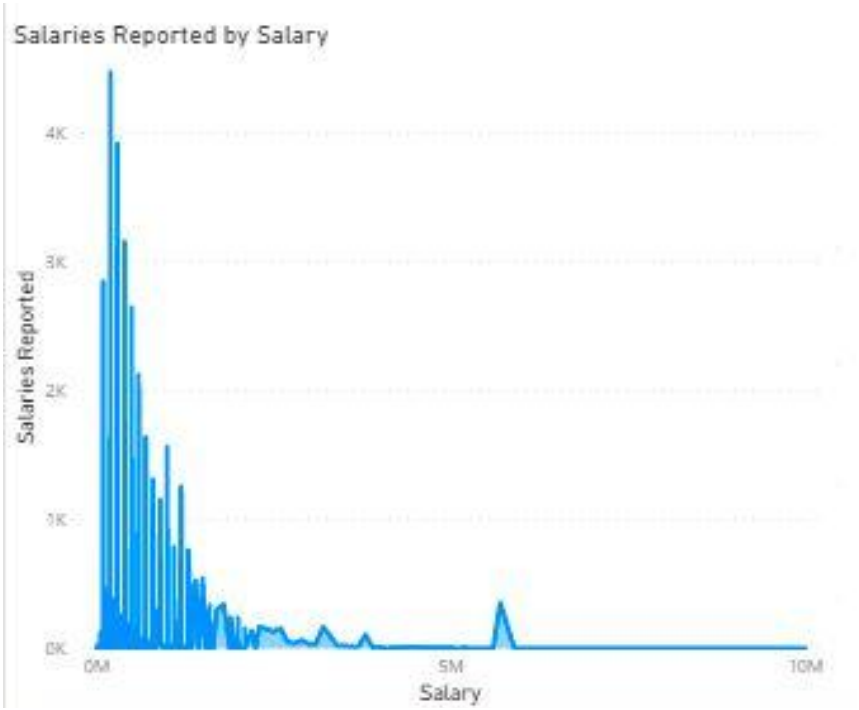
3. The Correlation Between Job Role, Salary reported and Salary

Interative Visualization:
https://app.powerbi.com/groups/me/reports/65ce01af-a9de-4a17-bdef-7f02b4c02449/ReportSection542dd31cd257294fa99b

In the graph of salaries reported and salary by job roles, we can see that the salary reported (in light blue) does not depend on the amount of salaries or the job roles, since it is spread evenly among the 11 job roles. In contrast we can see that there is one particular job role that get paid almost 4x higher than the other job roles in terms of salary

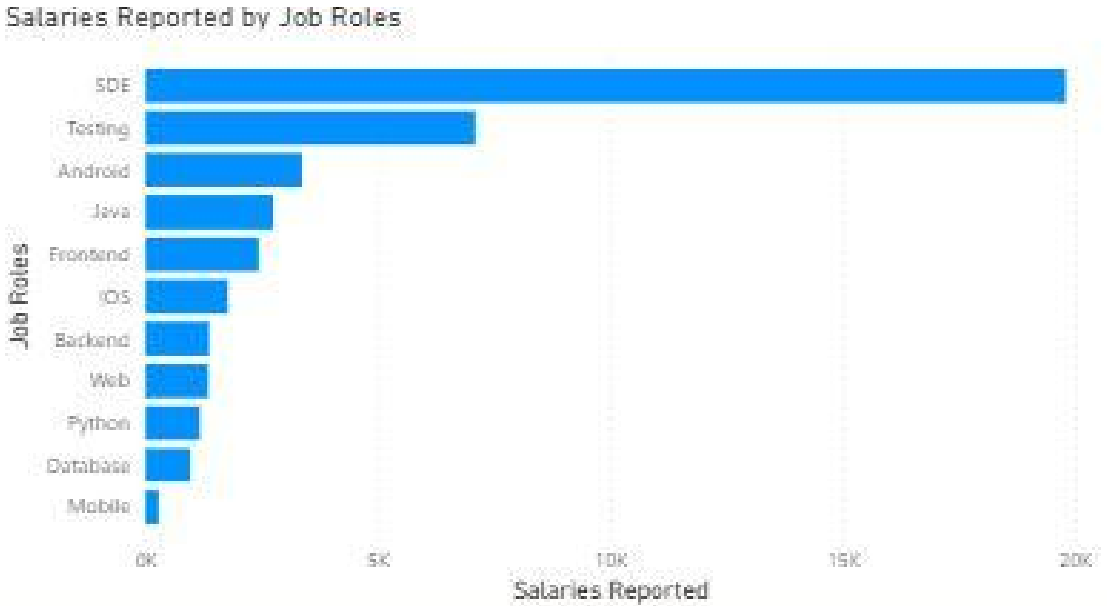


In this graph the salaries reported by salary, it depicts that more people who has less income are more likely to report their salary. The graph has been scaled a bit so that we can clearly see that this is a right skewed graph.



In the bar chart below, We can see that SDE has the most salary reported, which we were not able to see in the first graph. In the first graph, we can acknowledge that SDE has a relatively high salary among the other job roles. Similarly in salary reported, people are more likely to report their salary in SDE. We can say that workers working for SDE may be more prone to report their salaries compare to other workers because they have higher a salary.

By knowing the salary reported we are able to ge the total sample size of the data. Sample size would influence our precision of our result and the power of our conclusion. With 42K samples of IT salaries in India. From Google, the population of India is about 1.38 billion, and comparing to 42K, we can say that the sample size is relatively insignificant. Thus, even though this data set could be useful to discover some general patterns in the IT job field. The result that we get from this data set is not powerful enough.



Conclusion

Based on these data, The significance of analysis and visualization could provide some general information for the public, such as job picking or starting a company. These information can also sometimes provide information for companies, such as the amount of compensation. The legal compliance of a company may also be a concern for people, thus they can use these data to obtain useful information in order to benefit themselves. For the direction of future research a bigger sample data would be preferred and we can even compare these data with other countries such as the US or Australia to improve the system.