**Name: Abdallah Galal Abdallah**

**Cleaning and Modeling step**

1. Add column for people working per hour (we find that some jobs salary per hour and other salary is annual salary and we want make all same annual salary)
2. Clean Salary estimate column: salaries with -1: 214 rows out(we look for -1 in estimate salary and we find that is 214 row contain -1 and we will drop this rows because core of analysis is salary)
3. Transform estimated Salary from to 17−17−24 Per Hour(Glassdoor est.)/ 53𝐾−53$−91K (Glassdoor est.)To 53K-91K (clean column remove text from it )
4. Split it to 2 columns (53K) and 91K: min and max
5. Multiply hourly by 12\*40\*4 (40 is the number of formal work hours in the weak) we want to calculate annual salary for jobs with hour salary and multiply annual \*1000
6. Get the average of salaries (min+max)/2
7. Replace -1 rate with 0 indicating (I think no rates by - )
8. Company name from name rate to name only
9. Job 741 is d misleading just drop it
10. The founded date has missing values represented by -1
11. The type of owner ship has missing values represented by unknown
12. The size has missing values represented by unknown
13. The industry has missing values represented by unknown
14. The Sector has missing values represented by unknown.
15. Classify Each Company size from A To G
16. Classify Each Company Revenue from Class 1 To Class 12
17. Calculate the age of the company with -1 representing the missing data
18. Count The number of Competitors for each company
19. Join The 2 data sets
20. Add new column count of skills [1 to 4 skill]
21. Add state column from location
22. Create unpivote table contain skills and count of it each skill
23. Now we have the dataset having the following columns:

[job title -job desc-rating-location -headquarters-size-founded-type of ownership-industry-sector-revenue-revenue class -size class -hourly-min-max-avg salary- company name-age-competitors-number competitors-same state-python-R-aws-excel-spark-job\_simp-state-skills count]

**Note: I used Dax to implement 4 measures**

1. **Num\_competitors to calculate number of competitors**

num\_comp = LEN(glassdoor\_jobs[Competitors])-LEN(SUBSTITUTE(glassdoor\_jobs[Competitors],",","")) + 1

1. **Python% to calculate the percentage of python needed on jobs**

Python% = (SUM(eda\_data[python\_yn])/COUNT(eda\_data[aws]))\*100

1. **Excel%to calculate** **the percentage of excel needed on jobs**

Excel% = FORMAT(round(((sum(glassdoor\_jobs[excel])/count(glassdoor\_jobs[aws]))\*100),0)/100,"#%")

1. **R% to calculate** **the percentage of excel needed on jobs**

R% = (sum(eda\_data[R\_yn])/count(eda\_data[aws]) )

**Note** : **I will see to the data set from 3 sides salary side ,job opportunities, Needed skills to see which the no of jobs and salaries of this jobs and which of them affect on the salary and I will show which skills needed**

**Insights**

* **What the total number of jobs ?**

**The total number of jobs is 741 job**



* **What the Max Annual Salary?**

**The max annual salary is 306K**



* **What the Min Annual Salary?**

**The Min Annual Salary is 15K**



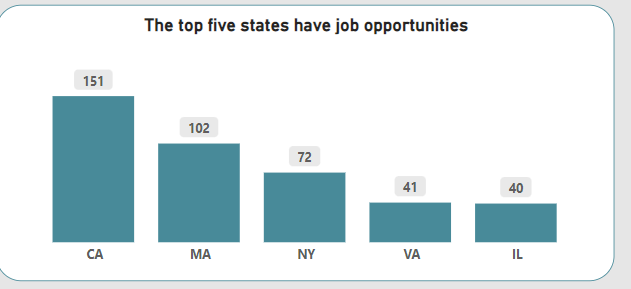
* **What the Avg Annual Salary?**

**The Avg annual salary is 101.35K**

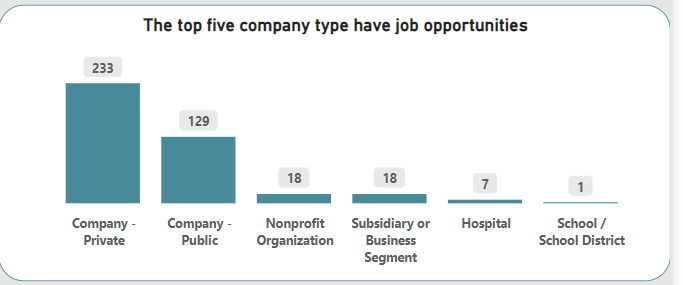


* **What the top 5 states have job opportunities ?**

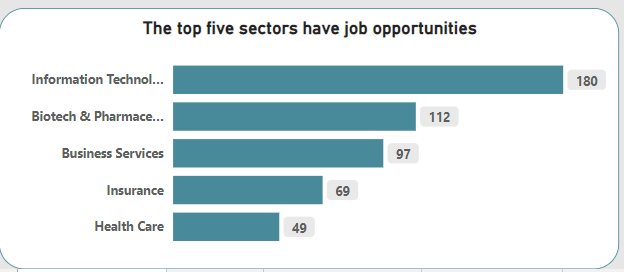
**[CA-MA-NY-VA-IL] ordered desc**



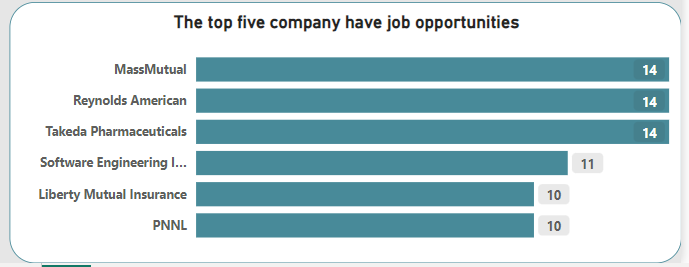
* **What the top 5 company type have** **job opportunities?**



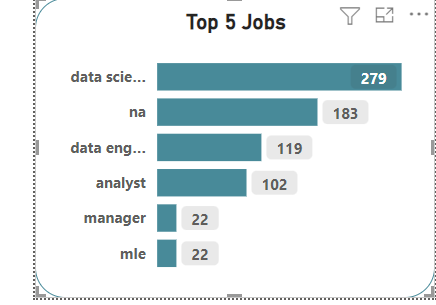
* **What the top 5 Sectors have** **job opportunities?**



* **What the top 5 Company have** **job opportunities?**

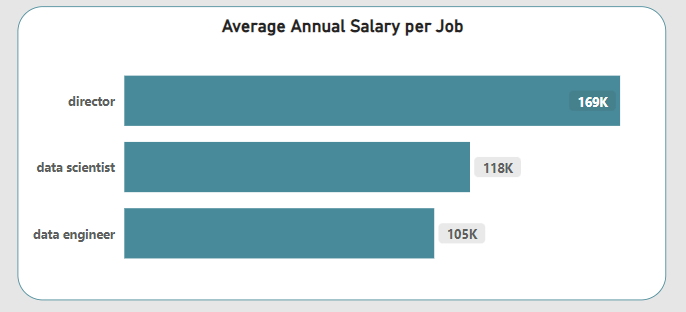


**Note : I was implement tooltip for each chart that show what the most job title needed**



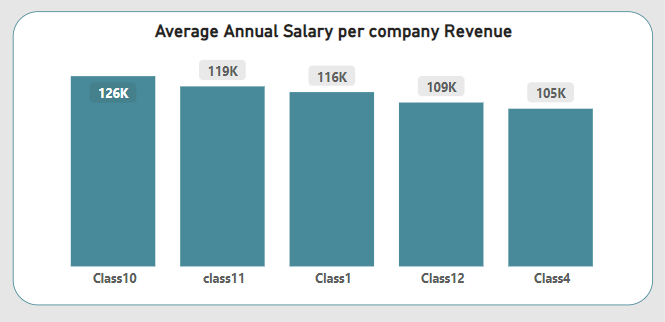
**Salaries Analysis**

* **What The Avg of Nannual Salary Per Each Job?**



* **What The Avg of Nannual Salary Per Each Revenue class?**

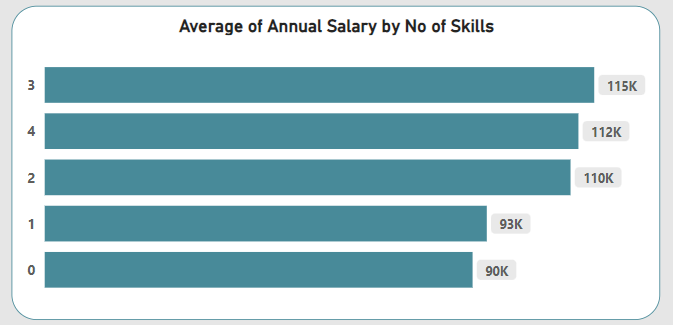
**Note clases assume class1 is the companies with high revenue +10billion and class 12 is the lowest revenue**



* **What The Avg of annual Salary Per Each Revenue class?**

**[0 skills skill not found in data set ,1 skill ,2skills ,3skills ,4skills]**

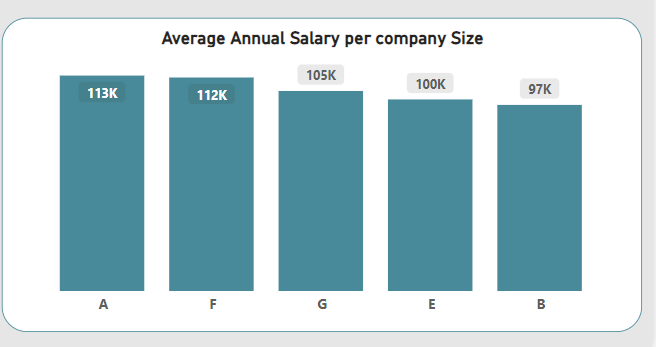
**The jobs with 3 skills achieve highest salaries**



**Note: I was implement drill through to show which 2 or 3 or 4 skills come together to now which skill affect on salary**

* **What The Avg of annual Salary Per Each Revenue class?**

**Note classes assume classA is the companies with high Emplyee number and class F is the lowest Employee Number**



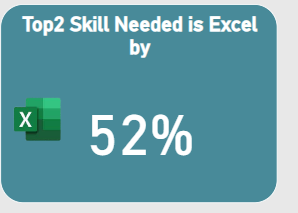
**Skils Analysis**

* **What the most skill needed ?**

**Is python needed in 53% of jobs**



* **Top 2 Skill is excel**



* **Low skill needed is R**



* **Skills ordered by need**

