Data Science Jobs Around the World

Background

As part of the Pacmann data analyst and business intelligence programme, I was curious to see how the different data science related roles are distributed across the globe. With the dataset I obtained from Kaggle, I built a dashboard to show the distribution of data science related roles across the globe, where do companies who hire these data science people come from and how the salary are broken down by various factors (location, employment type, seniority level, and working mode). Hopefully, this will provide me and similar individuals who are learning data science a better understanding of data science employment across the globe.

Objective

- Apply data preprocessing and exploratory data analysis to raw dataset
- Apply data visualizations skills to design user friendly dashboard
- Apply analytical thinking to determine analysis method and summarize meaningful insights

Dataset

Data Science Fields Salary Categorization Dataset contains a total of 607 data points taken from 2020-2022 with 9 columns :

#	Column Name	Description	Data Preprocessing
1	Working Year	The year the salary was paid (2020, 2021, 2022)	N.A
2	Designation	The role worked in during the year	N.A
3	Experience	The experience level in the job during the year. [EN - Entry level / Junior, MI - Mid level / Intermediate, SE - Senior level / Expert, EX - Executive level / Director]	Using VLOOKUP to map abbreviations with respective meaning. For example, EN is mapped to “Entry Level”.
4	Employment Status	The type of employment for the role. [PT - Part time, FT - Full time, CT - Contract, FL - Freelance]	Using VLOOKUP to map abbreviations with respective meanings. For example, FT is mapped to “Full-time”.

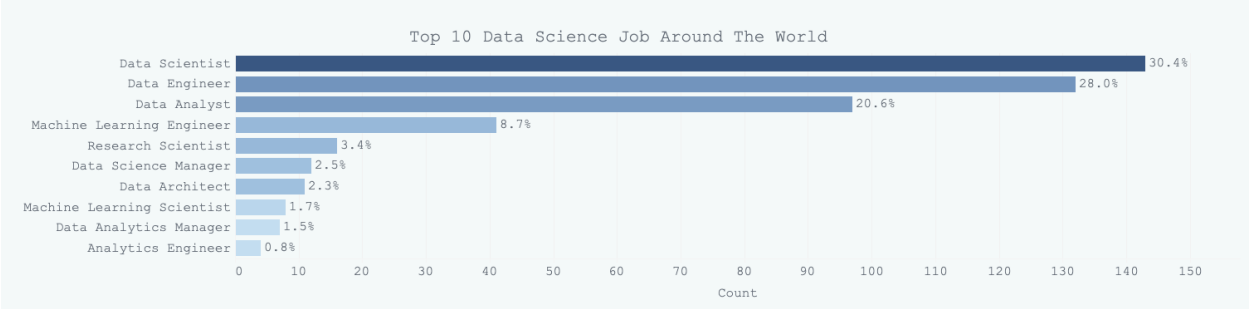
5	Salary In Rupees	The total gross salary amount paid.	Salary in Rupees was converted to USD.
6	Employee Location	Employee's primary country of residence during the work year as an ISO 3166 country code.	Using VLOOKUP to map abbreviations with respective meanings. For example, GB is mapped to “United Kingdom”.
7	Company Location	The country of the employer's main office or contracting branch.	Using VLOOKUP to map abbreviations with respective meanings. For example, GB is mapped to “United Kingdom”.
8	Company Size	The median number of people that worked for the company during the year. [S(small) - Less than 50 employees , M(medium) - 50 to 250 employees , L(large) - More than 250 employees]	Using VLOOKUP to map abbreviations with respective meanings. For example, GB is mapped to “S is mapped to Small”.
9	Remote Working Ratio	The overall amount of work done remotely. [0 - No Remote Work (less than 20%), 50 - Partially Remote, 100 - Fully Remote (more than 80%)]	Using VLOOKUP to map abbreviations with respective meanings. For example, 0 is mapped to “On-site”.

Insights

Note that the insights are merely deducted based on the dataset for the purpose of applying knowledge acquired through Pacmann course and do not in any way represent the population across the globe. We have no information of how the sample is collected and hence, the extent of bias and credibility of the dataset cannot be completely guaranteed.

1. Data Science Fields by Roles

From data analysis and visualization conducted, the **top 10 data science related jobs** around the world are as follows:

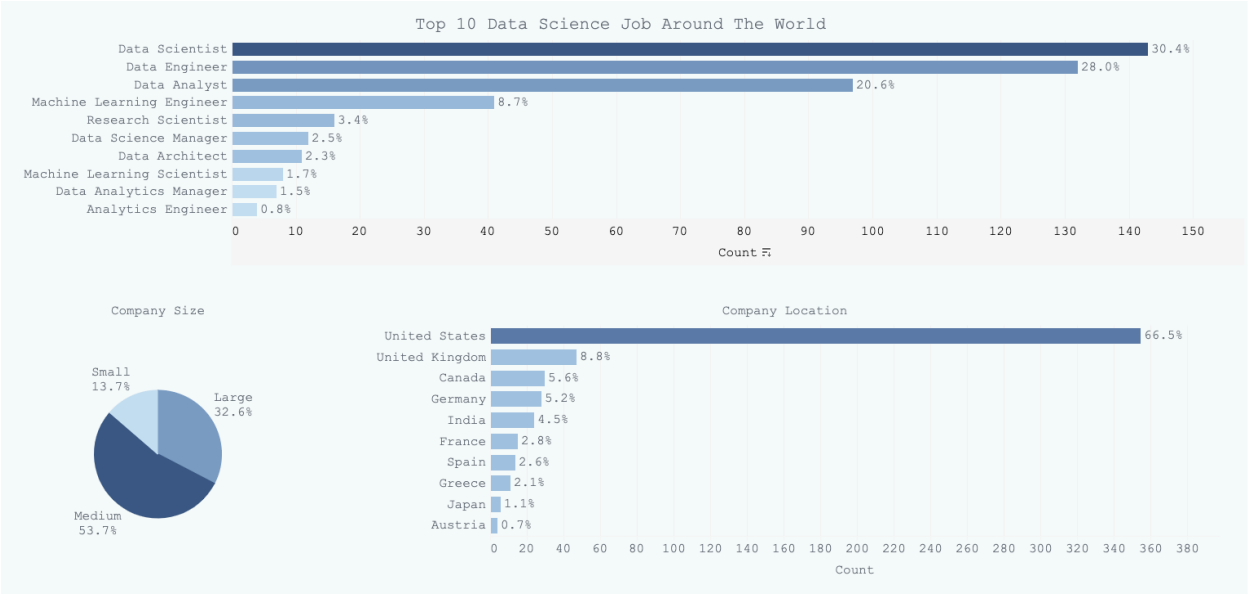


“Data scientists came first in the top 10 data science related jobs around the world, making up ⅓ of the sample data. Data engineer and analyst follow in the second and third place with 28% and 20.6% respectively. **”**

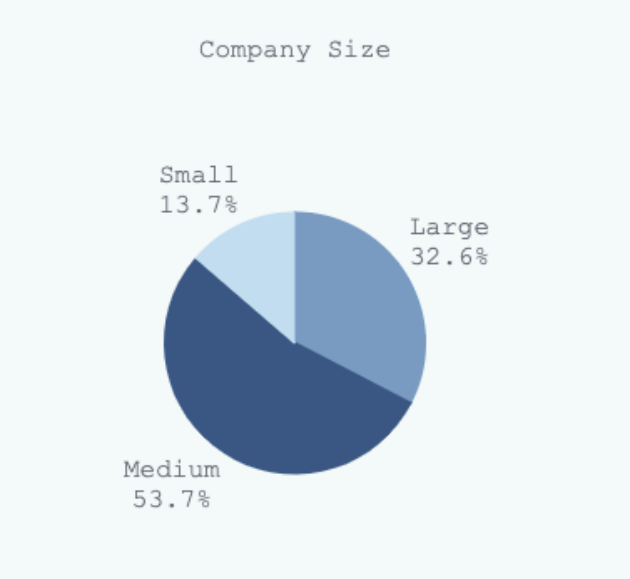
Below are deductions for these trends:
Disclaimer: that these deductions are not proven with data but serves as a guide for you to conduct further analysis.

- There could be **higher demand for data science, engineer and analyst roles** since there are more people in these roles. Individuals who are looking to transition into data science careers could look into these roles description and openings as there could be more job openings for such roles.
- On the flip side, after the top 3 roles, there is less than 10 percent sample taking up other data related roles. It could either be that the **demand is indeed low for these roles or there is unmet demand for such roles**.

2. Data Science Fields by Company Size & Location



Overview of Dashboard Interface - Companies

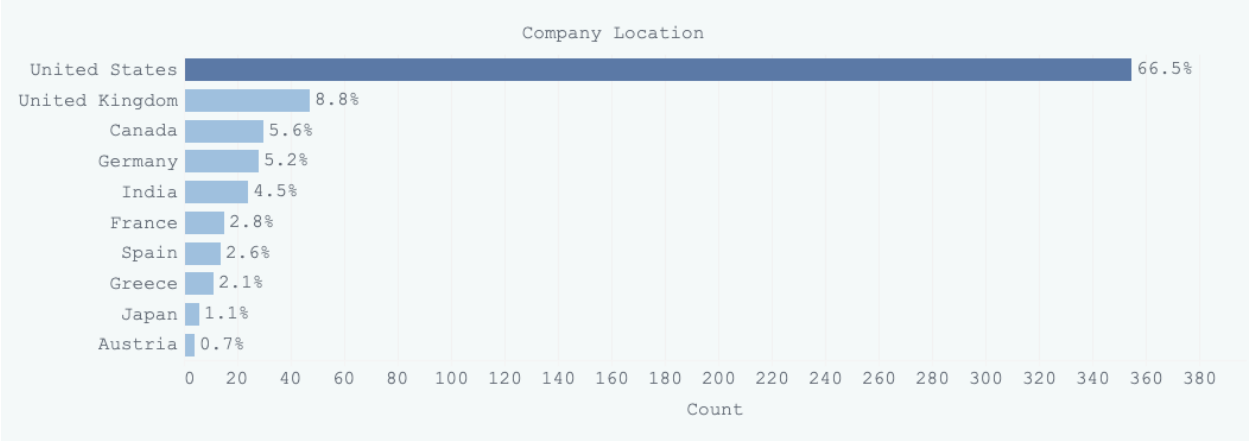


"**More than half (53.71%) of data science related employees work in medium-sized companies** (employees between 50-250). Large companies came in second place with 32.62% and small companies came last with 13.67%."

Below are deductions for these trends:

Disclaimer: that these deductions are not proven with data but serves as a guide for you to conduct further analysis.

- There could be **higher demand for data related roles in medium sized companies** as they tend to be at a **growing stage with hiring capabilities**. While large companies are usually more traditional and hence may not be data ready to hire so many data related roles and small companies are still focusing on PMF (Product Market Fit) and data is not yet a priority.
- **Individuals who are looking to go to data related roles can focus on medium-sized companies in their job search** as there could be more openings there.



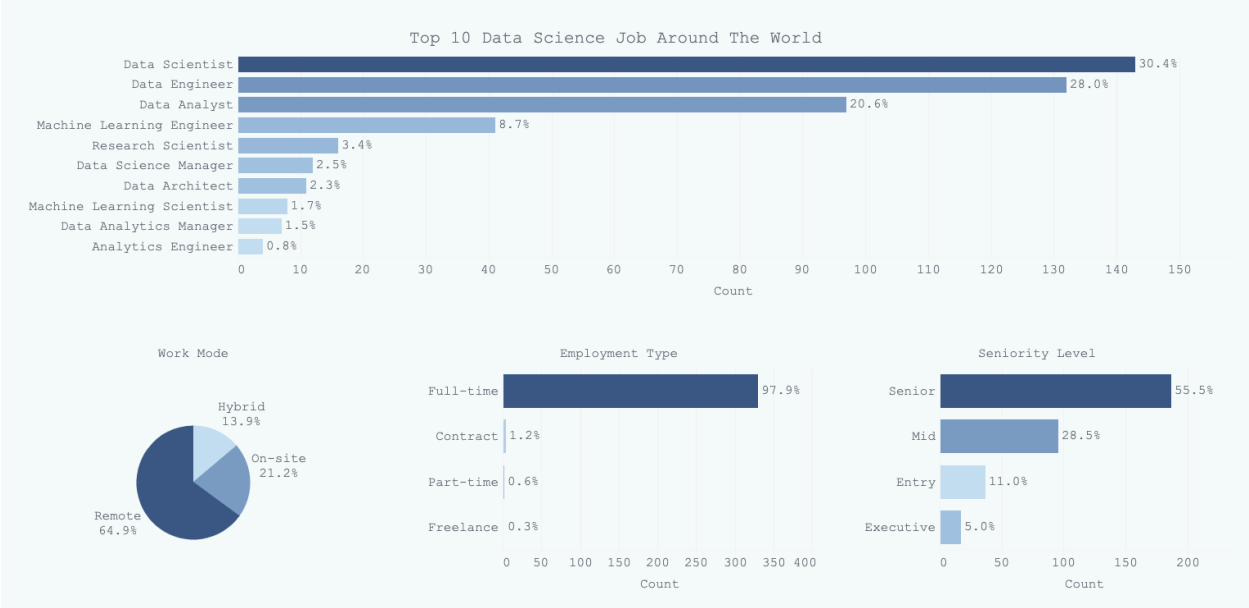
"**Companies from the United States came first** when it comes to the number of data science related hires in this dataset **with a whopping percentage of 66.5%** making up more than half of hiring companies around the globe. The rest of the top 10 countries make up less than 10% of the sample."

Below are deductions for these trends:

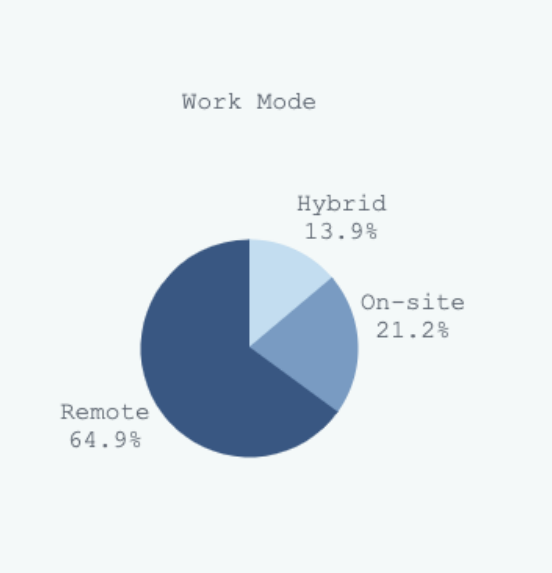
Disclaimer: that these deductions are not proven with data but serves as a guide for you to conduct further analysis.

- The United States is known to be more advanced in terms of technologies and hence, it is possible that the **US came first as they have advanced data capabilities to hire these roles**.
- The rest of the other countries in the top 10 also have advanced data capabilities but it could be possible that the sample is more biased towards US data science employees.
- Individuals who are looking for data related jobs can **explore these top countries to search for potential future employment**.

3. Data Science Fields by Working Mode, Employment Type, Seniority Level



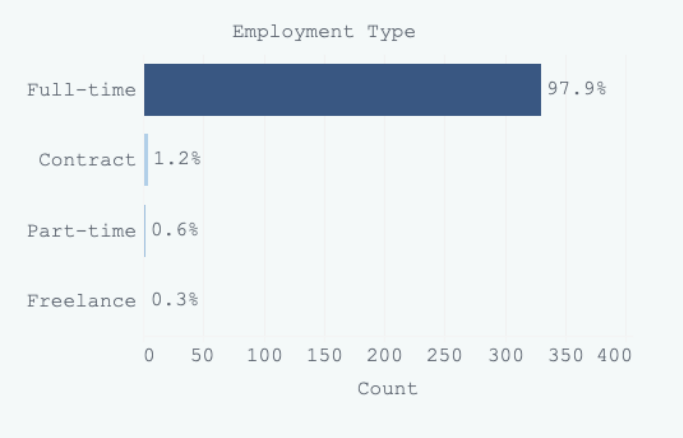
Overview of Dashboard Interface - Jobs



“**Remote working came in first with a percentage of 64.9% making up nearing 2/3 of working arrangement types for data science related roles**. On-site and hybrid make up less than 20% of the working arrangement.”

Below are deductions for these trends:
Disclaimer: that these deductions are not proven with data but serves as a guide for you to conduct further analysis.

- As the data is collected **during covid period (2020-2022),** it is expected that the **majority of data science employees are working remotely**. Additionally, **remote working arrangements are gaining popularity in post-covid workforce** as it reduces commuting time for employees and breaks down geographical barriers when it comes to closing in the talent gap.
- **Individuals who are looking to switch to data related roles can be rest assured that there are remote and hybrid working options** if they don’t prefer to travel that much. They can **focus on looking for remote or hybrid working opportunities.**

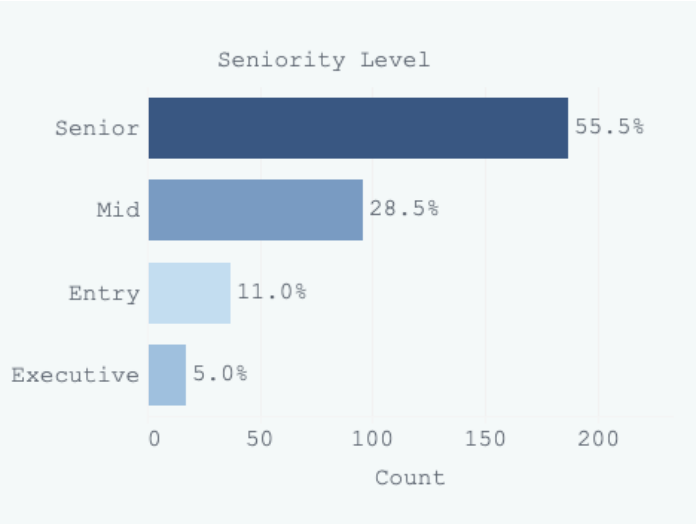


“**97.9% of data-related employees own full-time roles**. Less than 2% of roles are contract, part-time or freelance”

Below are deductions for these trends:
Disclaimer: that these deductions are not proven with data but serves as a guide for you to conduct further analysis.

- Sample could be biased towards full-time employees as there is so little percentage of other types of employment. **Although full-time employment makes up the most employment type, this could mean that other types of employment may indicate unmet demands and present opportunities**.

- With the emergence of freelancing platforms like Taskrabbit or Upwork in the US, one could potentially look for freelancing or part time opportunities for data-related work beyond full-time employment.

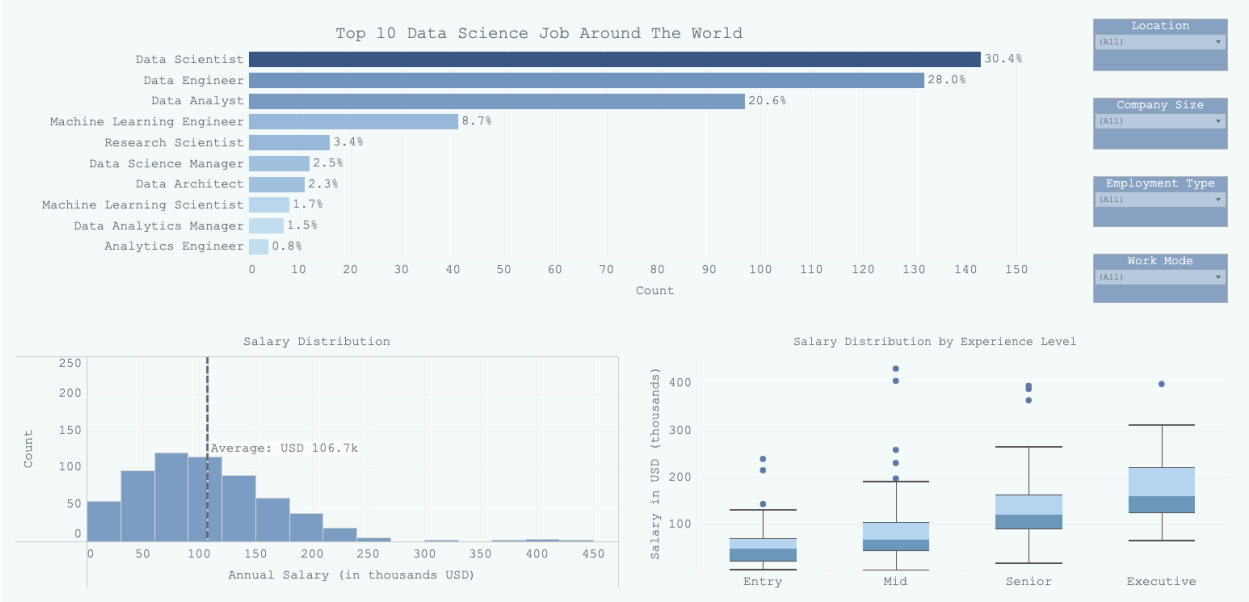


"**Senior data related employees make up more than half of the sample with a percentage of 55.5%.** Mid experience level came in second place, followed by entry and executive level."

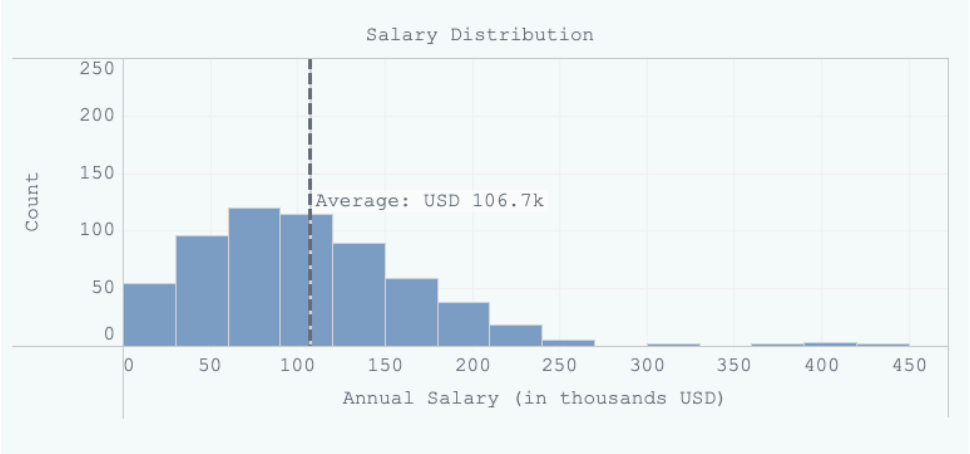
Below are deductions for these trends:
Disclaimer: that these deductions are not proven with data but serves as a guide for you to conduct further analysis.

- **Majority of samples in the data set are senior employees with substantial experience in data**.
- While executives only make up 5% and this could indicate there is difficulty in finding executives with extensive leadership and data knowledge.
- Individuals who are looking to switch to data related roles could look into roles with entry to mid experience level.

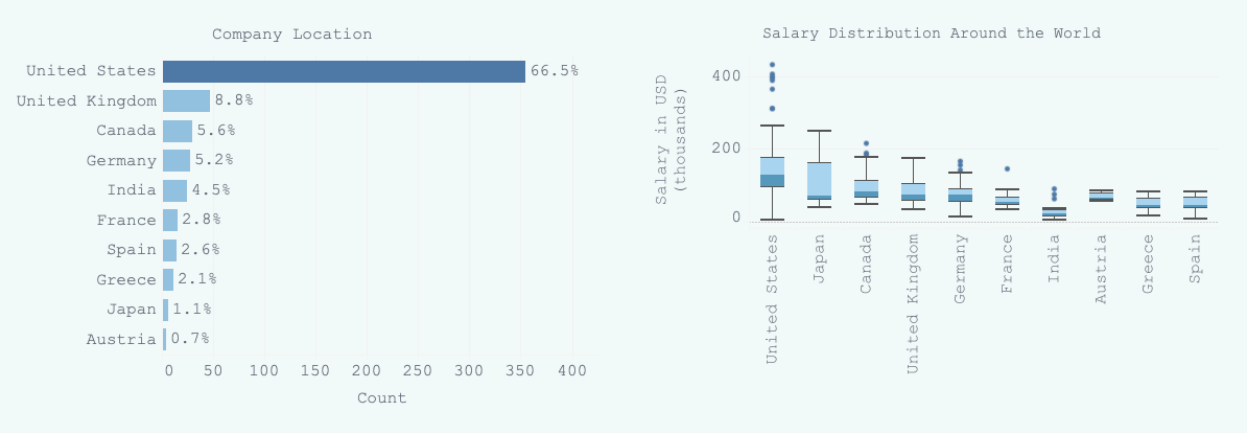
4. Data Science Fields by Salary



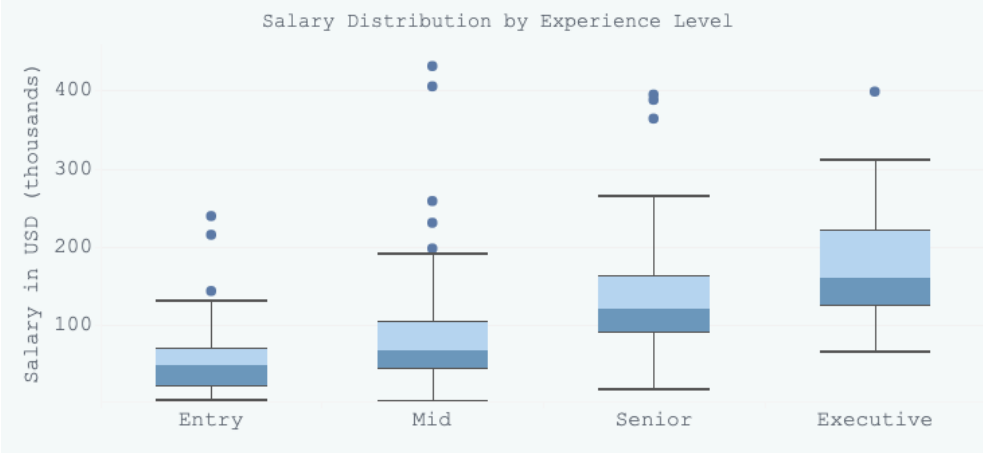
Overview of Dashboard Interface - Salary



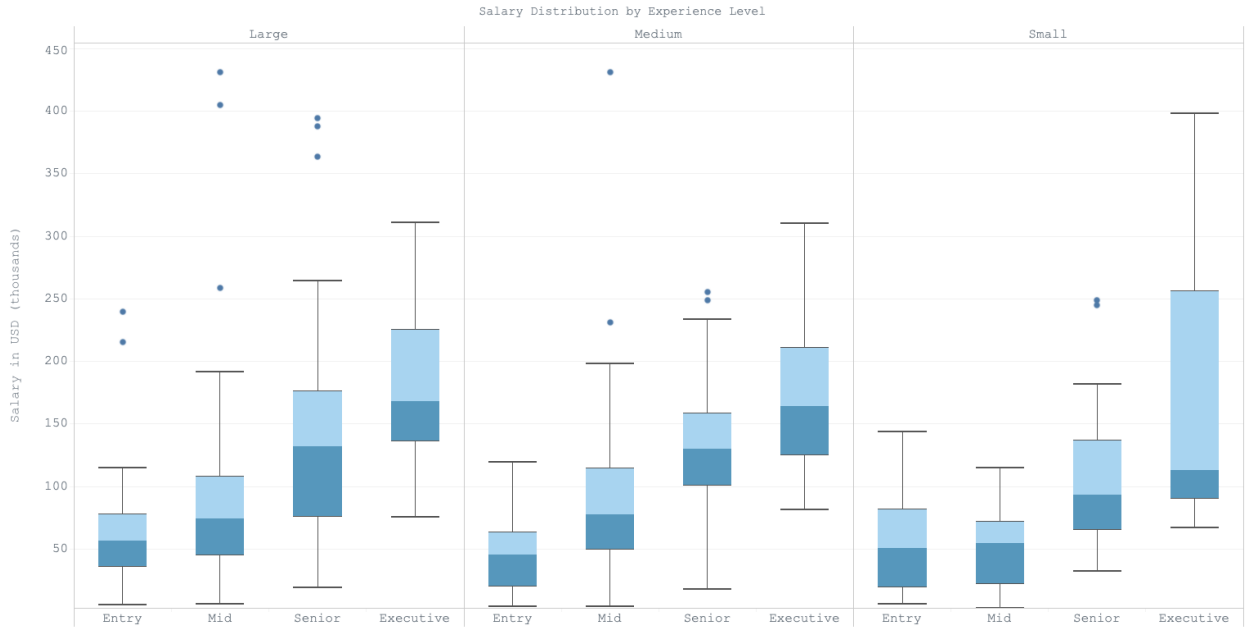
"**Average annual salary for data related roles is USD 106.7k .** With annual salaries going as low as <20k range to as high as 400-450k range."



“The United States pays the most salary for data related roles as compared to other countries in the top 10 locations of hiring companies. Although India is in the top 5 in terms of hiring numbers, it falls behind when it comes to compensation trailing in the last 5 with France, followed by Austria, Greece and Spain.”



“**With increasing experience, the salary range of data related roles increases with the highest salary increment from mid to senior level**. With the difference between entry level and mid level median around ~20k increase, mid level to senior level ~60k increase and senior to executive around ~40k increase. ”



“**Pay range varies by seniority level as well as company sizes.** The entry level pay range doesn't vary between company sizes. However, **mid level pay is the lowest range and executive have widest spread in small companies** ”

Below are deductions for these trends:

Disclaimer: that these deductions are not proven with data but serves as a guide for you to conduct further analysis.

- Data distribution is skewed and hence, the average may not be representative. Additionally with different seniority levels and job titles, the salary average could be different.
- There could be a lot of factors that influence the average and salary increase between each experience level like location, company size and employment type that can be further explored in more details for future improvements.
- Individuals who are looking to break into data science roles can use this findings as a guide for future research on salary based on their experience, location and hiring company sizes.

Tools

- Data preprocessing and EDA: Google Sheets
- Data visualization: Tableau
- Summarize insights: Google Docs

References

- Kaggle Dataset: link
- Tableau Documentation: link
- ISO 3166 Country Code mapping: link
- Rupees to USD conversion rate: link

Demonstration

- Tableau Dashboard: link