Exploratory Analysis of Data job salaries around the world

Exploratory Data Analysis (EDA)

At some point, certain categories may have few records, which could affect the reliability of statements based on the data. However, for the sake of studies related to data exploration, visualization, and comparison, I will proceed with the analysis even in these cases.

This is a portfolio project, and although the data is real, I have no control over how it was initially collected.

The objective of this stage will be to explore salaries by careers, understand which countries have the best and worst salaries, and if there is any relationship that can be highlighted solely through comparative data observation and data visualization.

About the dataset analyzed

The dataset was retrieved from Kaggle. This dataset has the goal to show what is being paid globally for jobs related to Data.

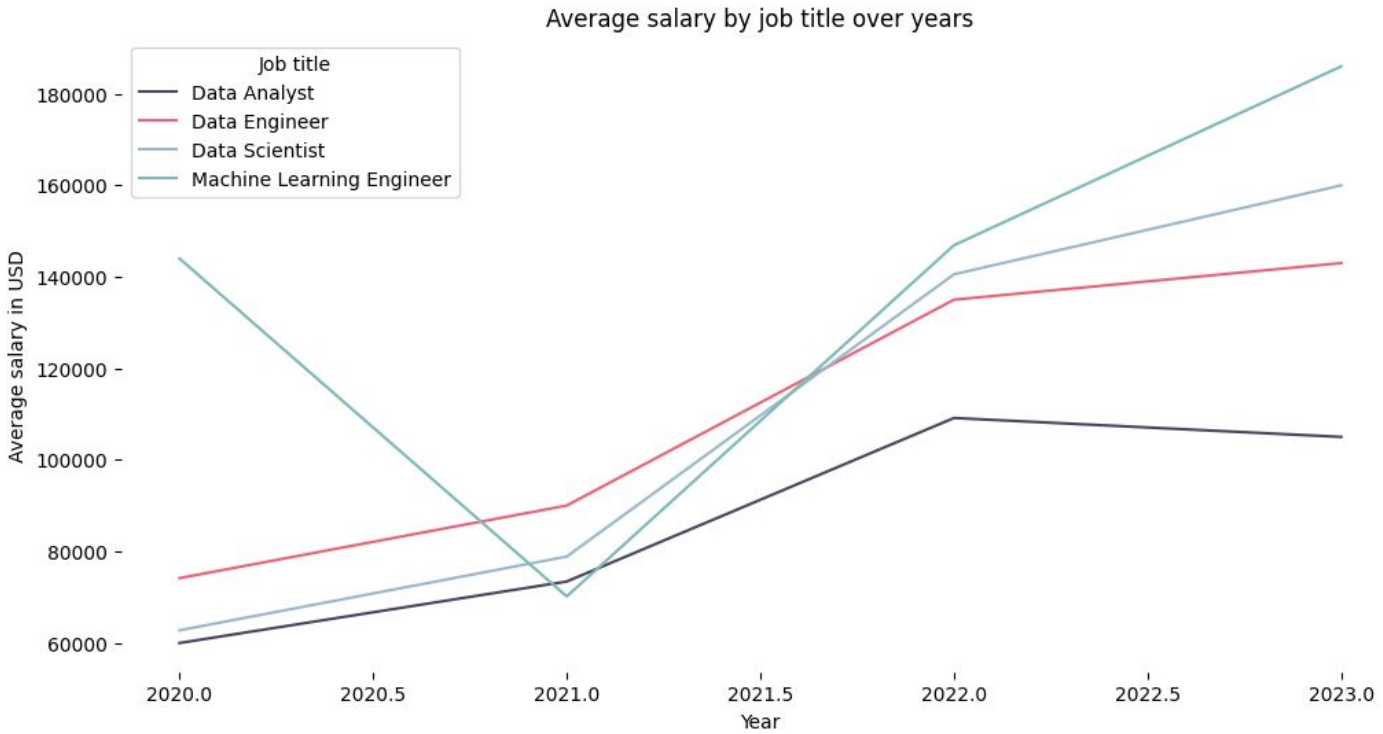
After data maintenance, the dataset contains the followed fields:

- **Work year:** The year the salary was paid.
- **Experience level:** The experience level in the job during the year
- **Employment type:** The type of employment for the role
- **Job title:** The role worked in during the year
- **Salary:** The total gross salary amount paid
- **Salary currency:** The currency of the salary paid as an ISO 4217 currency code.
- **Salary in USD:** The salary in USD
- **Employee residence:** Employee's primary country of residence in during the work year
- **Remote ratio:** The overall amount of work done remotely
- **Work model:** If In-person, Hybrid or Remote
- **Company location:** The country of the employer's main office
- **Company size:** If Large, Medium or Small

The dataset after modification contains data from 2020 to 2023.

1. Salaries by Career Over Time

Salaries by Career Over Time



Machine Learning Engineer salary decrease

The number of Machine Learning Engineer’s records by year worked:

- **2020:** 4
- **2021:** 18
- **2022:** 108
- **2023:** 940

In 2020, there were only four salaries registered, and three of them were from US companies, which offer the highest salaries in USD. However, in 2021, there were 18 salaries registered, with the majority from other countries that pay less in USD for the same position.

The drop in Machine Learning Engineer salaries from 2020 to 2021 isn't indicative of a market trend; it's merely a result of data bias.

Conclusions

Salaries have shown significant growth from 2020 to 2023. However, this doesn't accurately reflect the growth rate, as there were fewer records from 2020 to 2021 compared to 2022 and 2023.

The **Data Analyst** and **Data Engineering** careers exhibited salary stagnation from 2022 to 2023. This could be attributed to several factors, such as the accelerated growth in these fields during the pandemic, followed by a transition to a phase of equilibrium.

The continuous growth, even from 2022 to 2022 in **Data Science** and **Machine Learning** careers may be related to the emergence of artificial intelligence.

2. The countries with the highest salaries in 2023

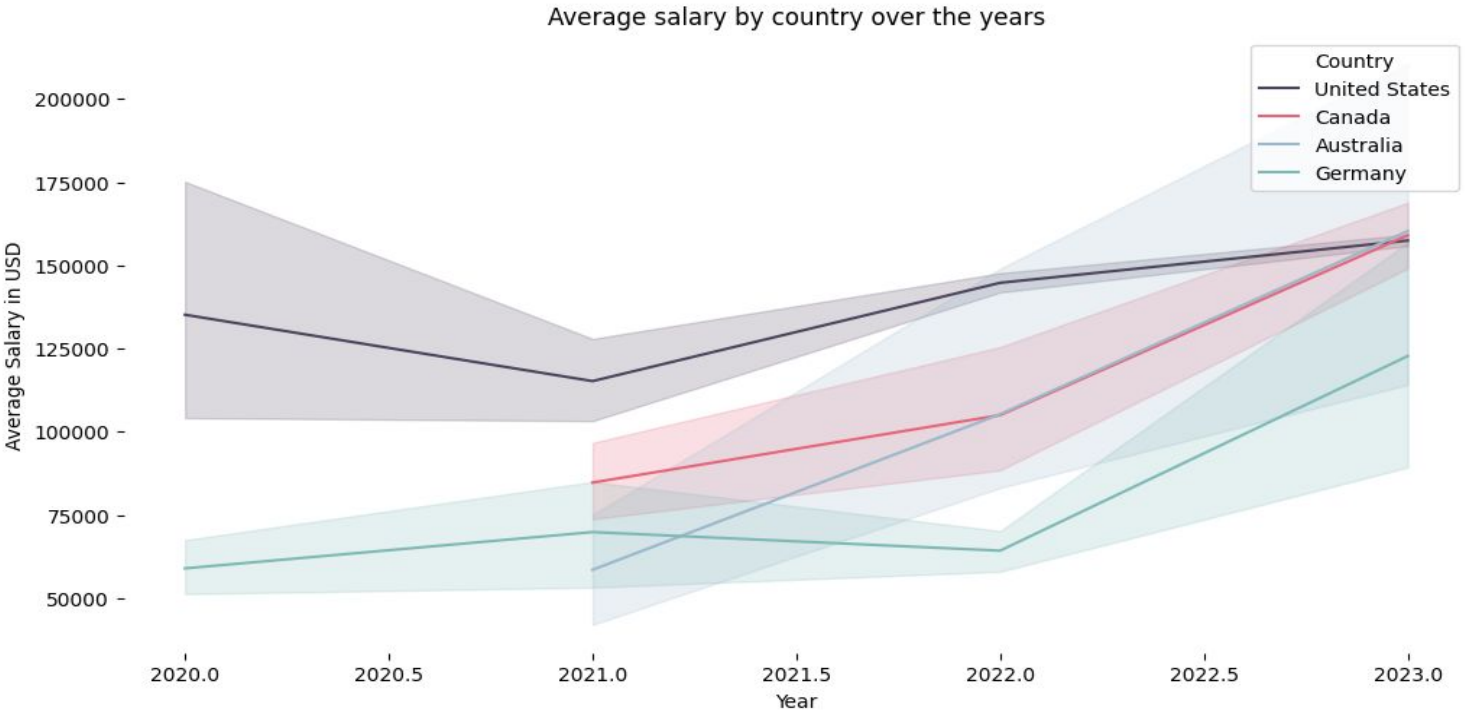
The four top-ranking countries in terms of salary

The **majority of the recorded salaries are from the US**, which could introduce bias into the analysis. However, for study purposes, I disregarded this factor and proceeded with the analysis anyway. Considering the median number of records per country is 3, I decided to **analyze only countries with more than 10 salaries registered**.

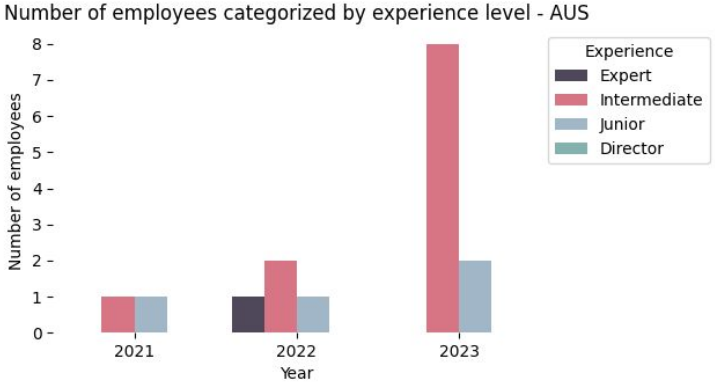
The four top-ranking countries in terms of salary:

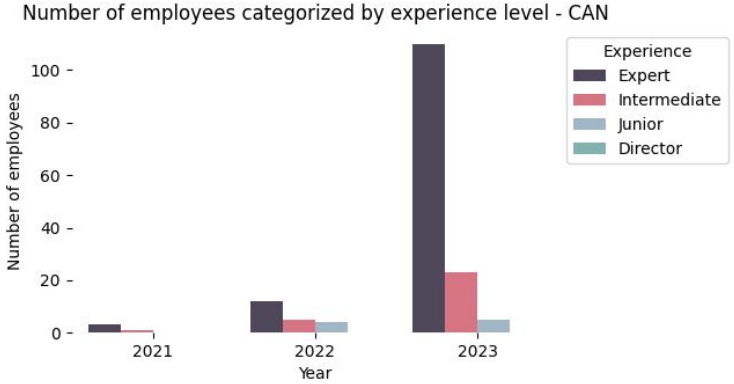
- **Australia:** $160325.00
- **Canada:** $159083.00
- **United States:** $157481.00
- **Germany:** $122766.00

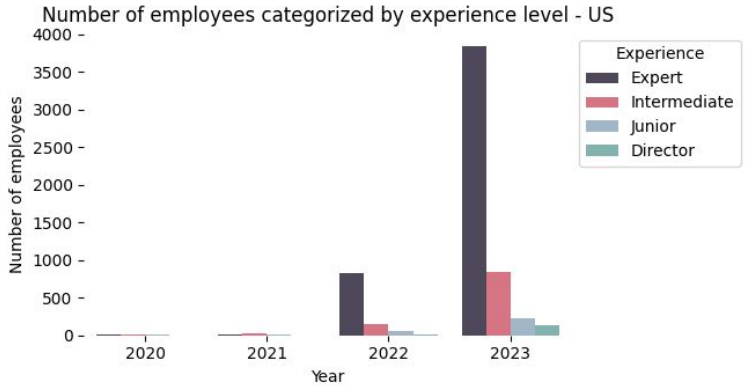
Salary behavior over the years

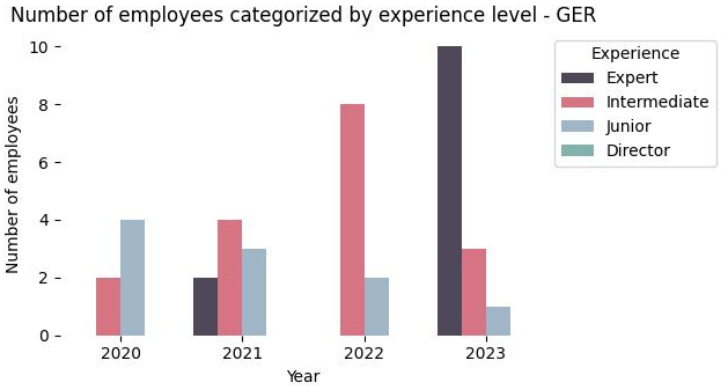


Number of employees by work experience

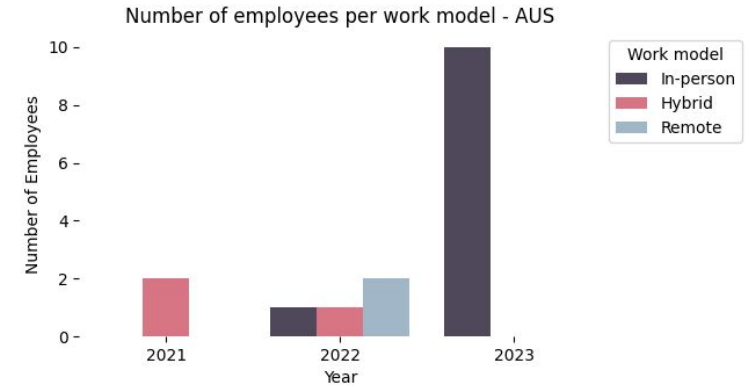


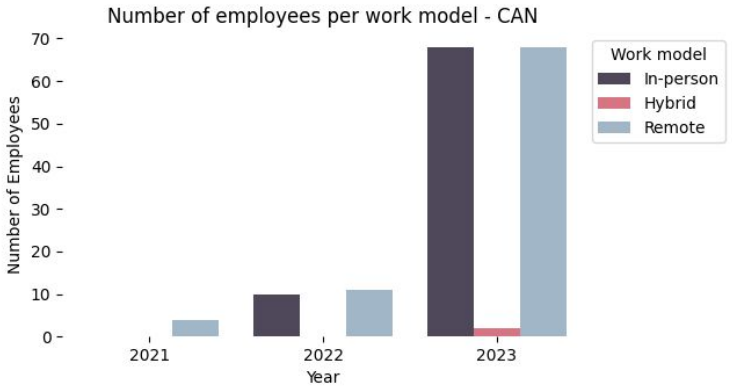


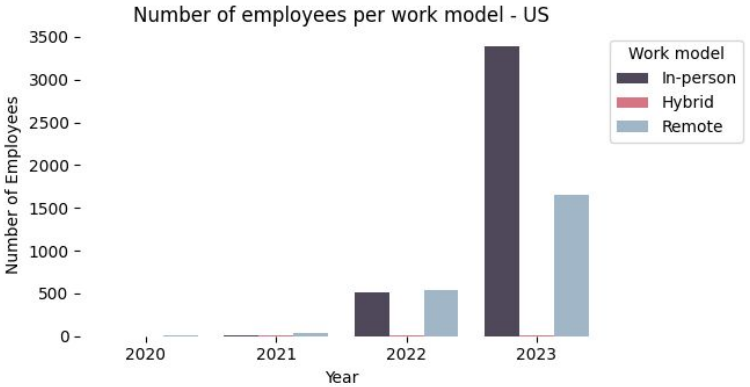


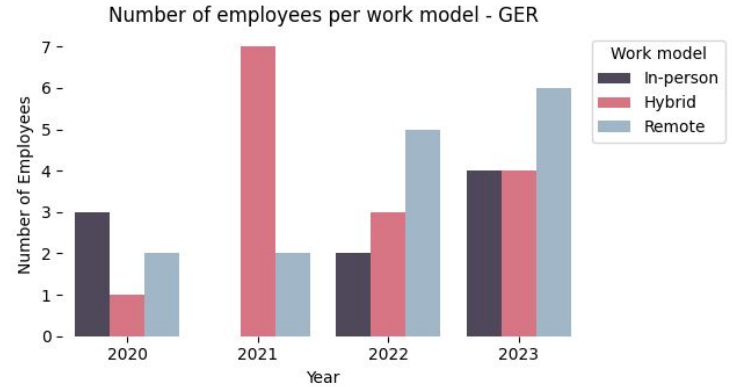


Number of employees by work model









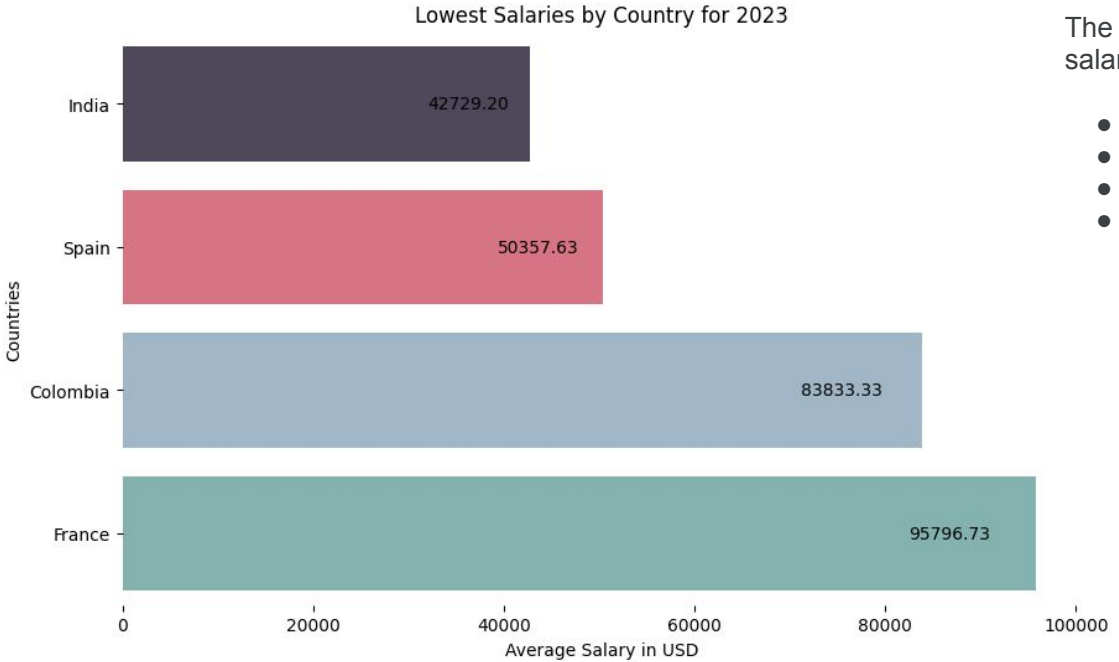
Conclusions

1. Over the year, the mean salaries in all of these countries have grown.
2. With the exception of Australia, all countries show an increase in the number of records for expert-level employees. This suggests that the growth in salaries may be partly due to an increase in senior employees, rather than a general rise in the career field itself (although there is growth regardless).
3. Australia exhibited a significant number of intermediate-level records, however, It has a low number of total records.
4. In the United States, Canada and Australia, the majority of recorded contracts are In-person. In contrast, Germany show a more balanced distribution between contract types, and all of them have high salaries.
5. The high increase in Germany salary from 2021 to 2023 may be due to the increase in number of records of Expert employees.

In general, much of the salary growth is attributed to the increase in records of employees in high-level positions, and not necessarily just the growth of the field itself. Contract models do not show significant influence on salaries.

3. The countries with the lowest salaries in 2023

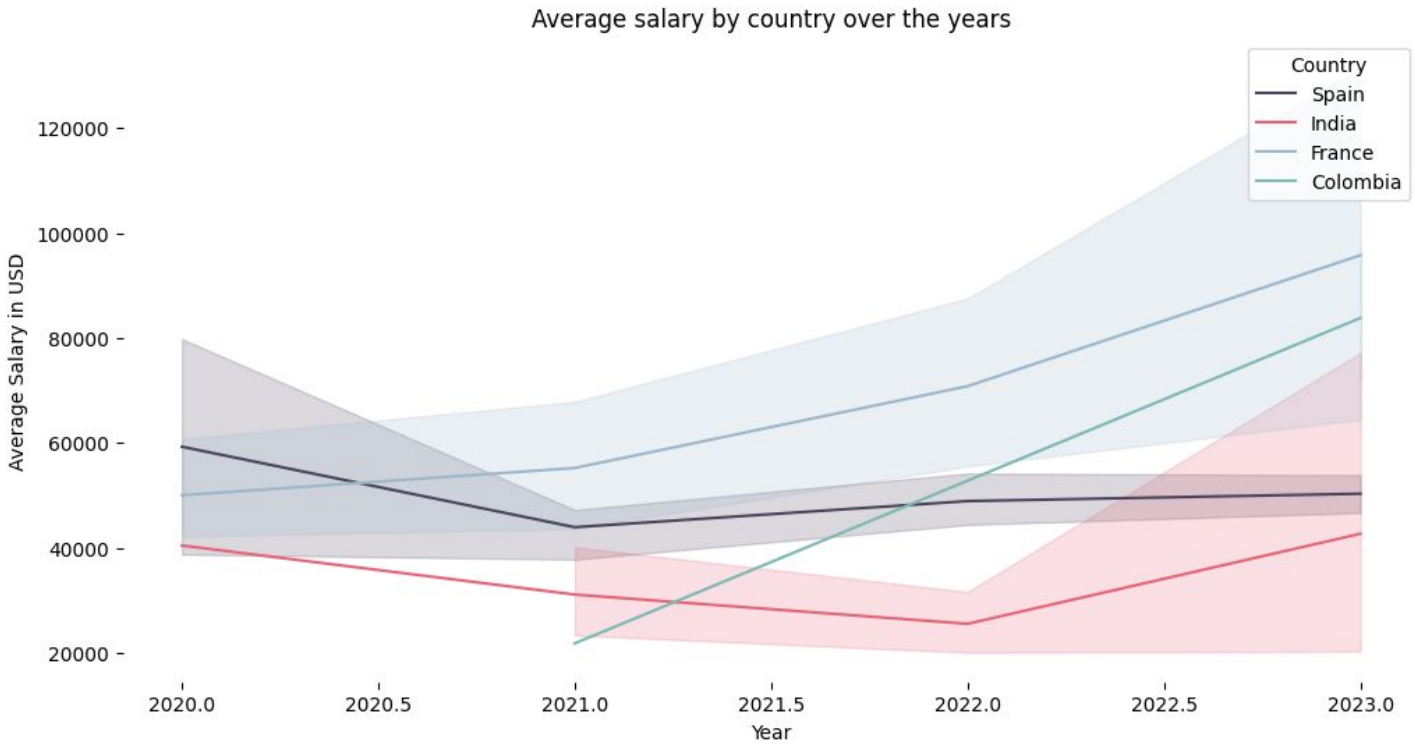
The four lowest-ranking countries in terms of salary



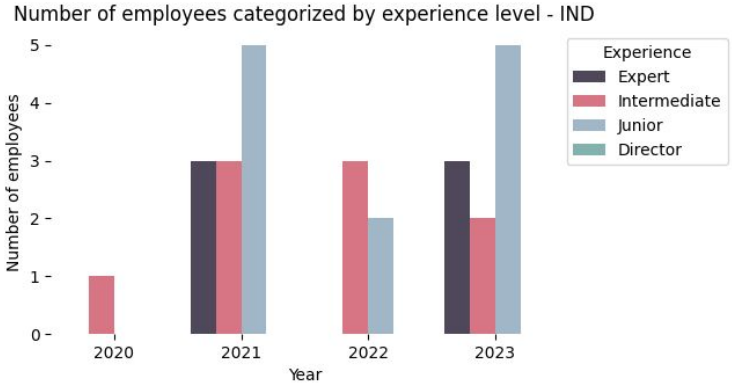
The four lowest-ranking countries in terms of salary are:

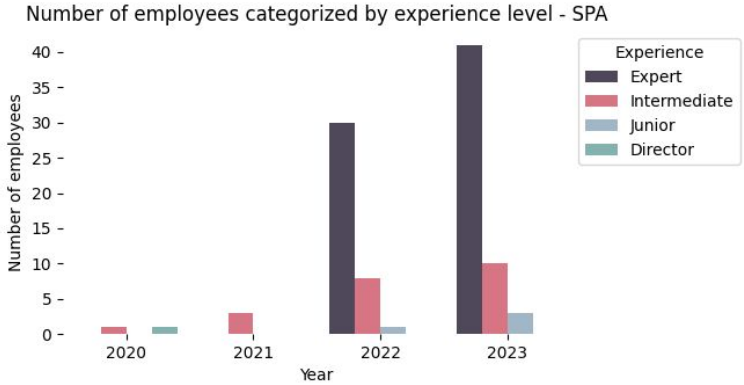
- **India:** $42729.00
- **Spain:** $50357.00
- **Colombia:** $83833.00
- **France:** $95796.00

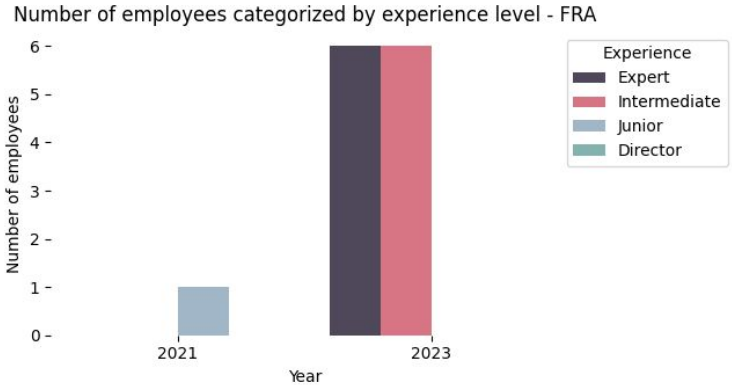
Salary behavior over the years

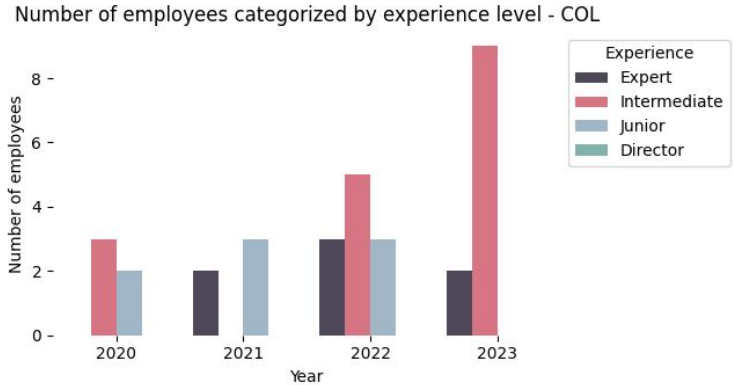


Number of employees by work experience

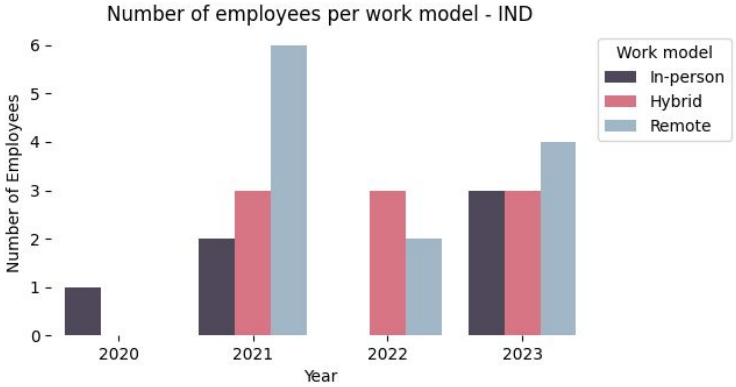


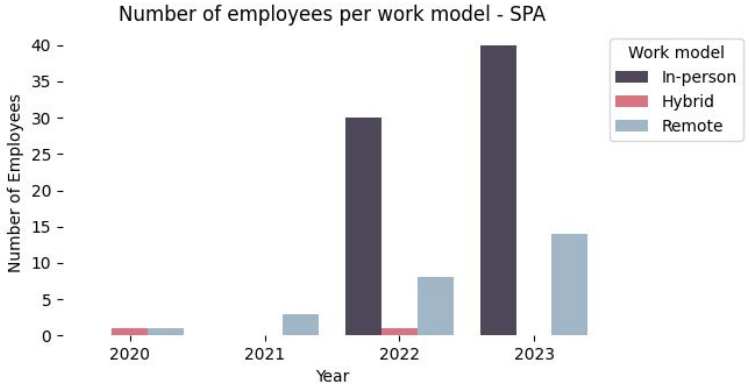


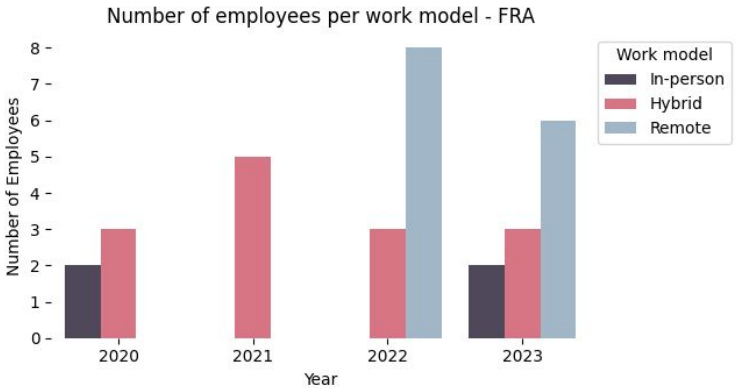


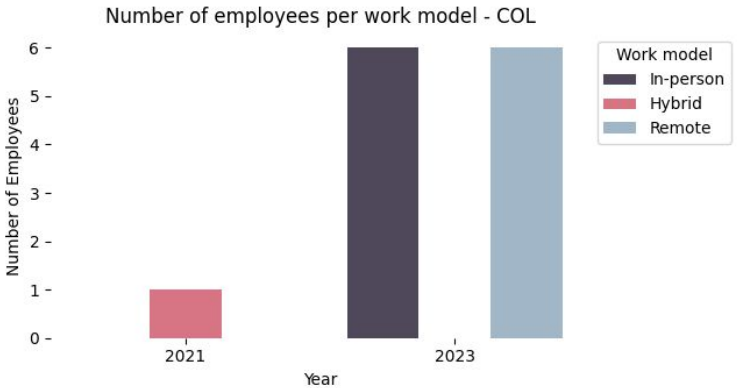


Number of employees by work model









Conclusions

1. Despite their low salary base, these countries have seen salary growth each year, which could be attributed to overall career growth around the globe.
2. The limited number of records for these countries may introduce bias into the analysis.
3. A higher frequency of junior-level positions in India may explain the lower salaries there. Of course, these salaries also reflect the country's reality, including the cost of living, culture, politics, etc.
4. For countries like Colombia (COL) and France (FRA), a substantial number of intermediate-level records could contribute to a lower average salary when compared with other countries.
5. Spain, despite its lower salaries, has a considerable number of expert-level positions. The low salaries in USD may be due to other factors such as a lower minimum wage and a lower cost of living compared to other countries.

Countries with lower salaries do not share many common factors regarding the number of employees by experience level or specific work model, and they still exhibit growth throughout the year. In other words, aside from the low number of records, some reasons may be unique to each country, such as technology policies, prioritization of other sectors, low minimum wages, or even a lower cost of living.

4. Average salary by company size

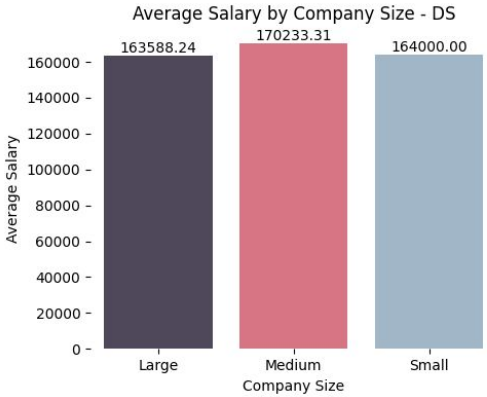
Average salary by company size

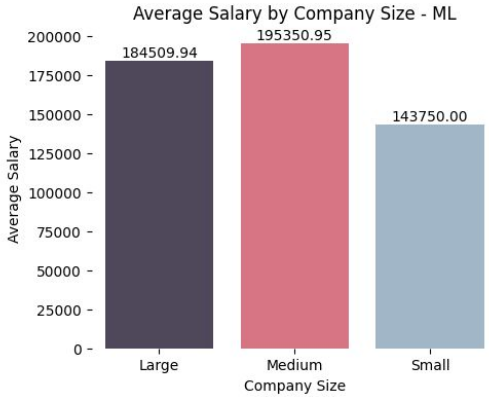
To analyze this aspect, I selected a specific country and year to minimize data bias:

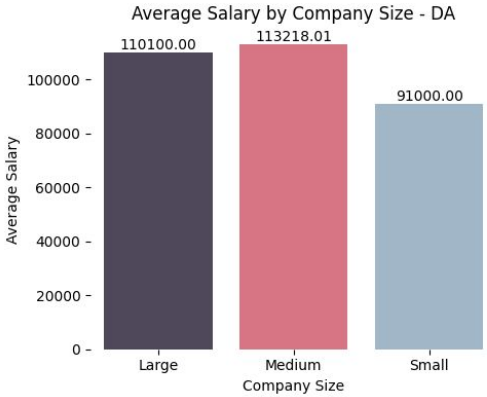
- Country: United States
- Year: 2023

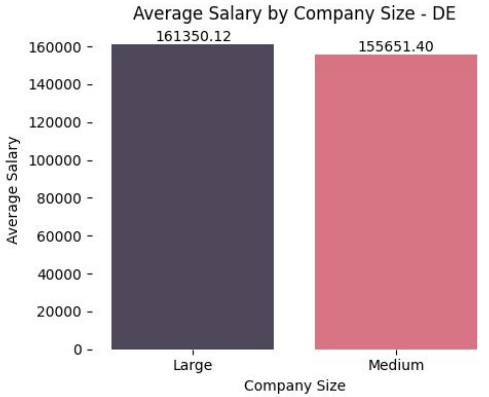
These selections are based on the number of records available.

Average salary by company size









Here we can observe that for most careers, the trend of higher salaries is seen in medium-sized companies, followed by large companies, while smaller salaries are found in small companies. However, it can be noticed that the monthly variation can be very low.

5. Next steps

Possible next steps

All topics can be further explored by examining the data in a more granular fashion with greater statistical rigor. Assuming the United States as a standard country for analysis due to the large number of records across all careers over the years, we can conduct more robust analyses, such as predicting salary growth patterns for each field in the coming years and forecasting work model trends for the future.

When looking at countries with the highest and lowest salaries, it would be important to delve into the data in a more granular manner. This involves understanding which areas are contributing to significant salary variations, and which ones are driving salaries up or down. This approach would provide more precision in determining whether the reasons are career-related trends or external factors affecting companies, such as political decisions, for instance.