DATA SCIENCE JOB INSIGHTS: INDUSTRY TRENDS AND ANALYSIS

INTERACTIVE IMFORMATION VISVALIZATION

FEBRUARY 2024

PREPARED BY:
DENZIL DSOUZA

UNIVERSITY OF WASHINGTON
Information School

ABOUT THE DATA

Dataset Link: Kaggle

THE DATASET ENCOMPASSES GLOBAL INFORMATION, DETAILING COUNTRIES AND THE ARRAY OF DATA SCIENCE JOB OPPORTUNITIES THEY PRESENT.

Demographics:

1. work_year: The year in which the data was recorded.
2. job_title: The specific title of the job role, like 'Data Scientist', 'Data Engineer', or 'Data Analyst'.
3. job_category: A classification of the job role into broader categories for easier analysis.
4. salary_currency: The currency in which the salary is paid, such as USD, EUR, etc.
5. salary: The annual gross salary of the role in the local currency.
6. salary_in_usd: The annual gross salary converted to United States Dollars (USD).
7. employee_residence: The country of residence of the employee.
8. experience_level: Classifies the professional experience level of the employee.
9. employment_type: Specifies the type of employment, such as 'Full-time', 'Part-time', 'Contract', etc.
10. work_setting: The work setting or environment, like 'Remote', 'In-person', or 'Hybrid'.
11. company_location: The country where the company is located.
12. company_size: The size of the employer company, often categorized into small (S), medium (M), and large (L) sizes.

DATA QUALITY AND SANITY CHECKS

Jupyter Notebook would be used for performing the Sanity checks.

We have **9355 records** in the dataset.

Step 1: Analysis of missing Values.

We do not see any missing values in the column.

```
Nulls in each column
Column              Count
work_year               0
job_title               0
job_category            0
salary_currency         0
salary                  0
salary_in_usd           0
employee_residence      0
experience_level        0
employment_type         0
work_setting            0
company_location        0
company_size            0
```

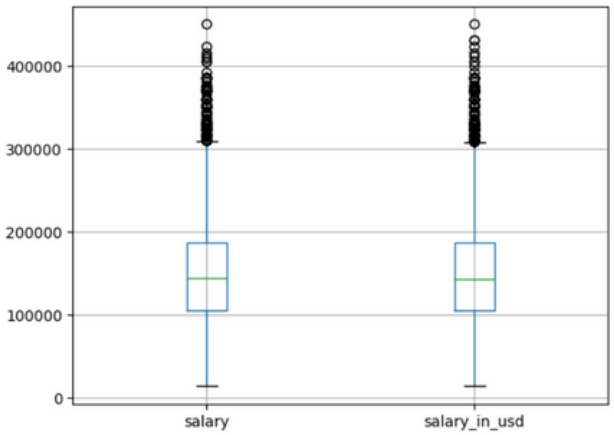
Characteristic/Salary	Salary	Salary in USD
Count	9355	9355
Mean	1409928	150300
Standard Deviation	63609	63177
Minimum	14000	15000
25th Quadratile	105200	105700
50th Quaratile	143860	143000
75th Quadratile	187000	186723
Maximum	450000	450000

Step 2: Data distribution of the salaries.

The salary variables would be the key to analyze the trends over the years and the disparity across various job categories in the data science domain.

Insights:

1. The mean and the median are relatively closer indicating that the data is symmetrically distributed.
2. The minimum and maximum values are $14,000 and $450,000 indicating a high variability in the data.
3. The high standard distribution of $63,608 would indicate a good variability in the data.
4. The maximum outliers are above the 75th percentile.
5. The IQR values are between between $105,700 to $186,723 for Salary in USD.



QUESTIONS FOR EDA

Let's embark on a comprehensive exploration of this dataset, unraveling its intricate patterns and insights.

Salary Distribution Across Job Categories

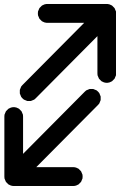
What is the average salary for data science jobs in the country that offers the highest number of job opportunities in the field?

Job Category Analysis by Numbers

How many job openings are currently available for the top 10 job titles in the country with the highest concentration of data science positions?

Geographical Analysis of Salaries

How has the average salary evolved in the top countries with the highest volume of data science job opportunities?



Notable Salary Shifts in Job Categories

From 2020 to 2023, which job category experienced the most substantial increase in average salary, and conversely, which category saw the most significant decrease in average salary?

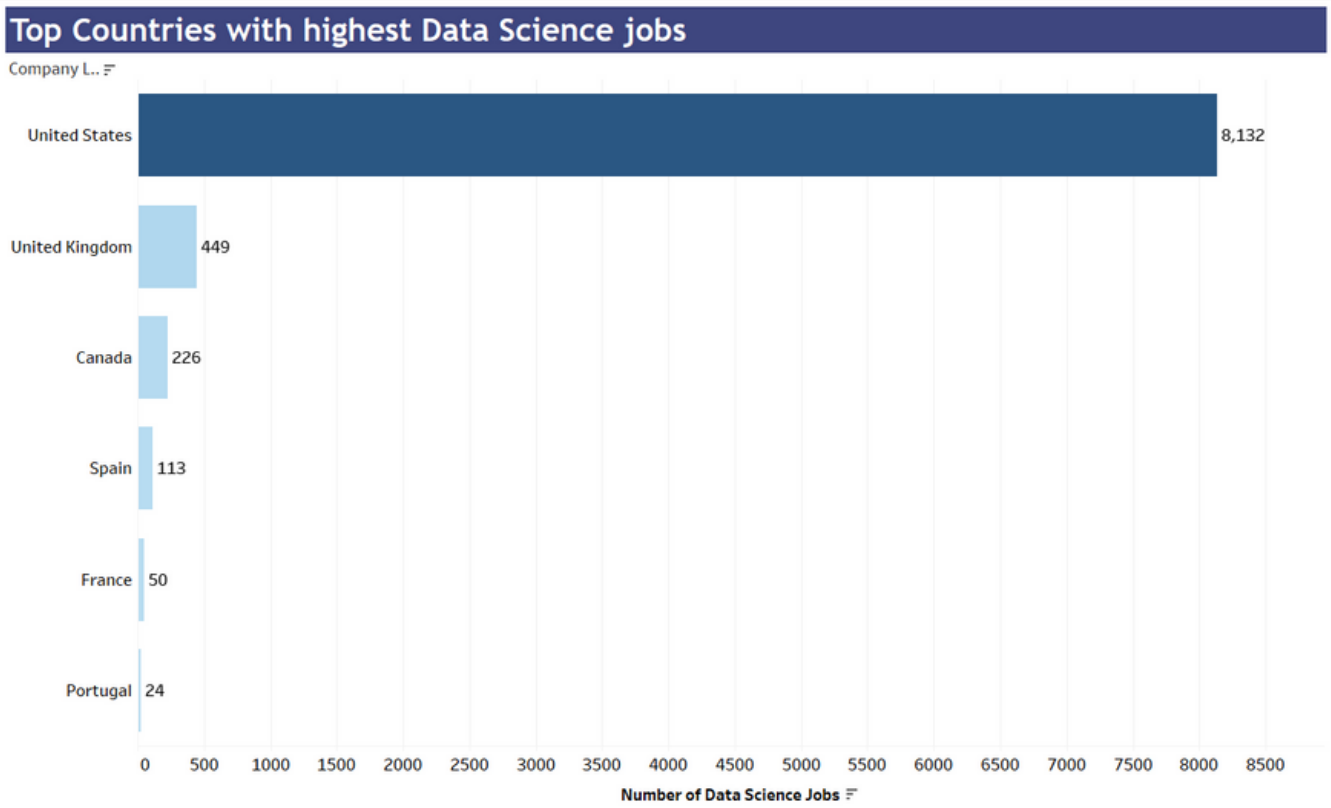
Salary Analysis Across Key Countries and Work Settings

Are there significant variations in average salaries across different work settings for data science roles globally?

SALARY DISTRIBUTION ACROSS JOB CATEGORIES

WHAT ARE THE AVERAGE SALARIES FOR DATA SCIENCE JOBS IN THE COUNTRY THAT OFFERS THE HIGHEST NUMBER OF JOB OPPORTUNITIES IN THE FIELD?

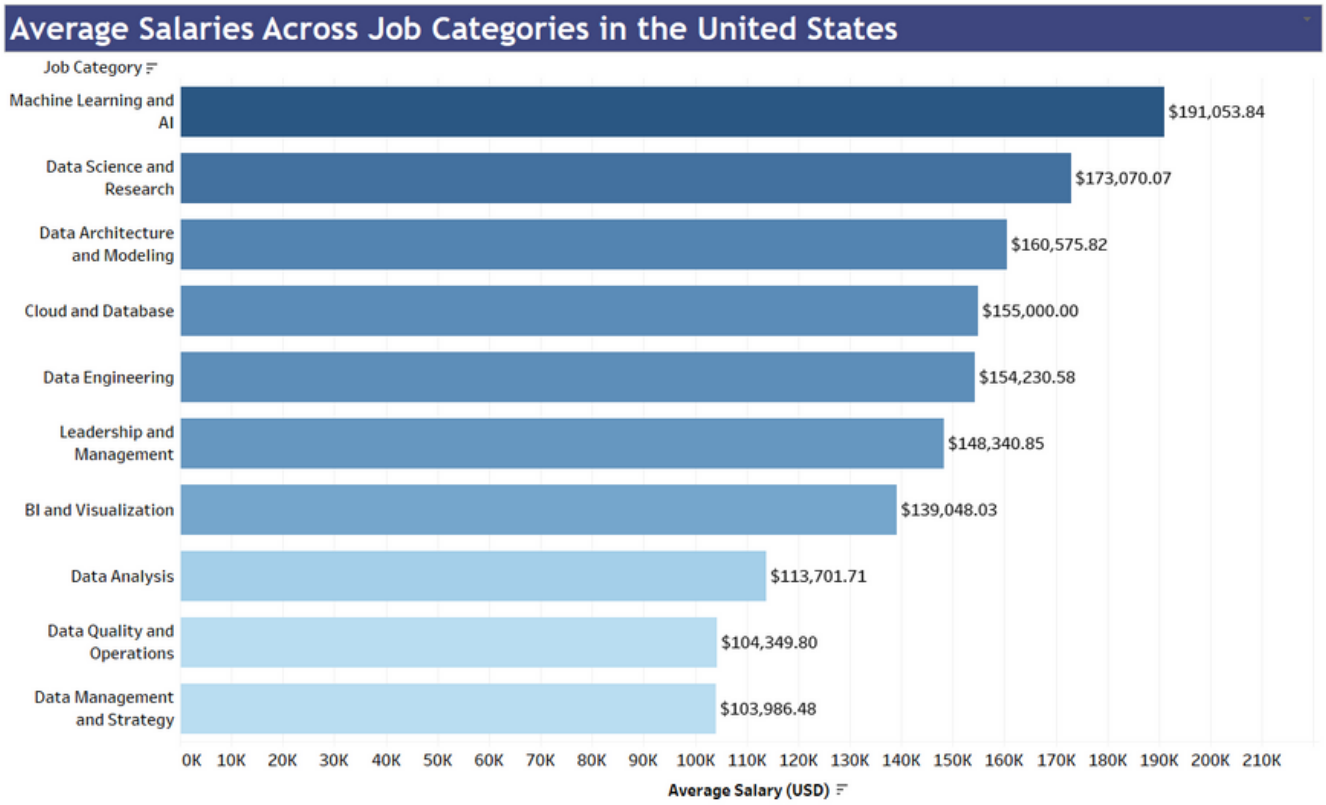
- We first analyze the country with the highest number of data science jobs. Our top destination is the **United States** with .**8,132** jobs.
- We do have a **limitation** due to the large amount of data from the United States compared to other countries.
- The dataset doesn't consider other demographic factors like **population** which would affect the number of jobs.
- An article from Statista (Taylor, 2023) supports this claims while one blog from Careerera (Careerera, 2021) mentions countries like **Australia**, **Germany** and **Israel** that are not a part of this dataset



SALARY DISTRIBUTION ACROSS JOB CATEGORIES

WHAT ARE THE AVERAGE SALARIES FOR DATA SCIENCE JOBS IN THE COUNTRY THAT OFFERS THE HIGHEST NUMBER OF JOB OPPORTUNITIES IN THE FIELD?

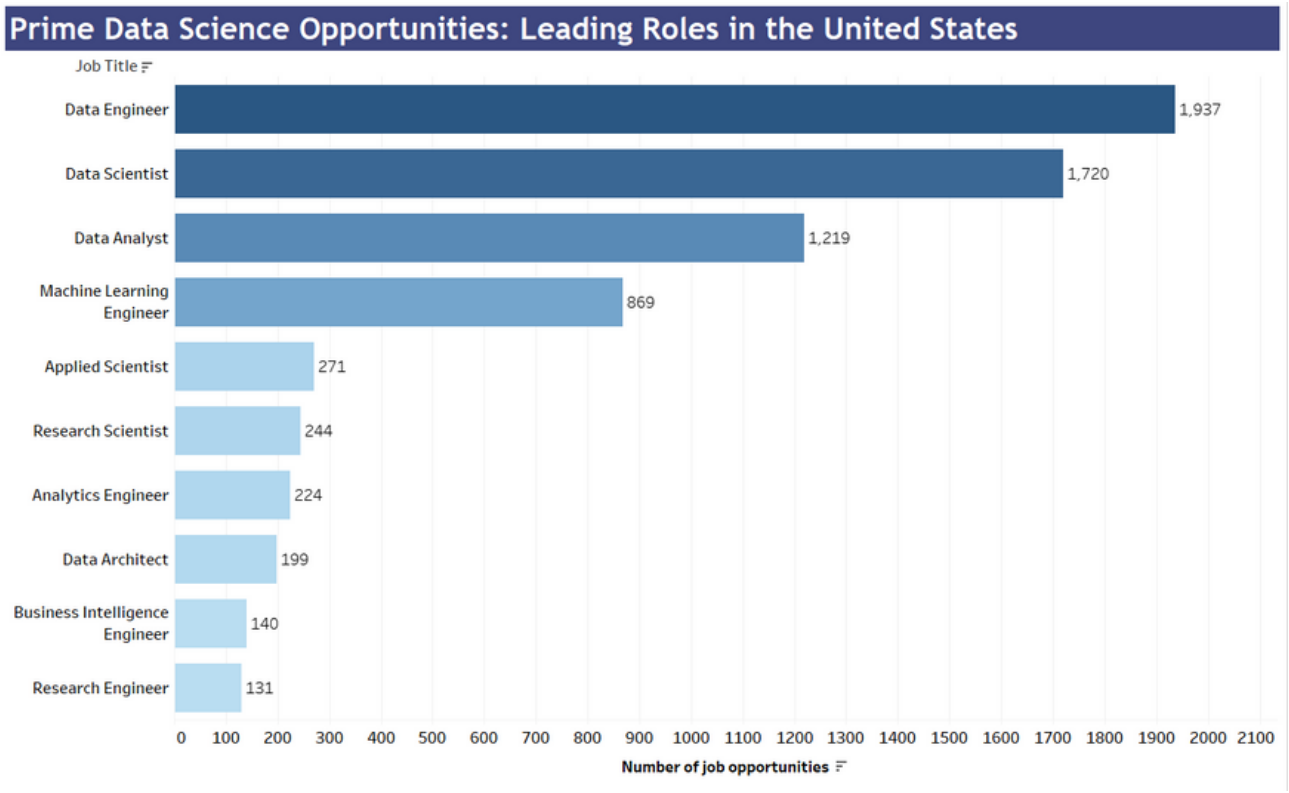
- In our next visualization, we will analyze the job categories for the highest performer i.e. the **United States**.
- Our winner is **Machine Learning and AI** with an average salary of **$191,053.84.**
- **Data Management and Strategy** has the lowest average of **$103,986.48**
- Given the high skill requirements inherent in Machine Learning and AI roles, salaries in these fields typically reflect this expertise (Nkolakowski, 2023).
- Data Management and Strategy usually run of predefined frameworks making it simpler and easier to implement causing the lower average.



JOB CATEGORY ANALYSIS BY NUMBERS

HOW MANY JOB OPENINGS ARE CURRENTLY AVAILABLE FOR THE TOP 10 JOB TITLES IN THE COUNTRY WITH THE HIGHEST CONCENTRATION OF DATA SCIENCE POSITIONS?

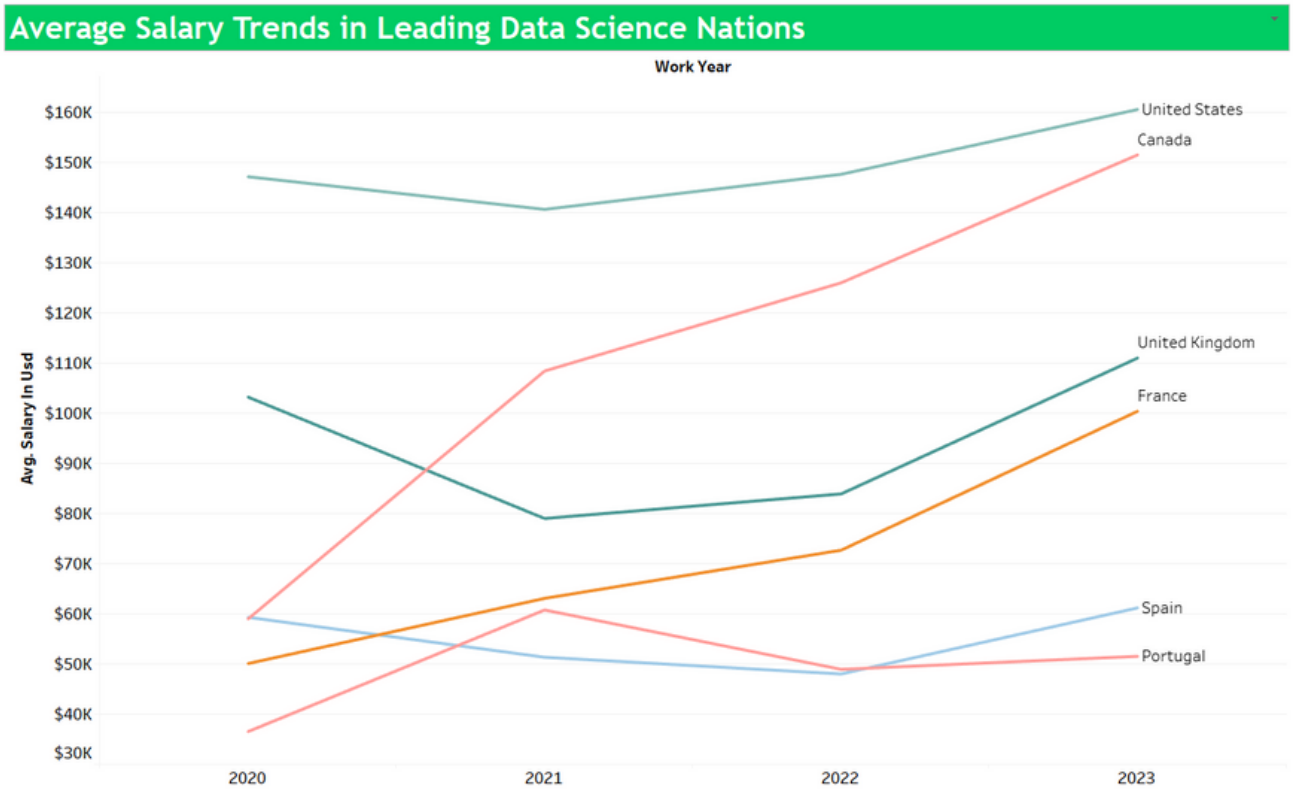
- We will be analyzing jobs across the **United States**, considering the wide variety of job offerings and the **richness of data** available from US-based positions.
- We can observe that there is a high number of jobs available for **Data Engineers** at **1937** and the least number of jobs available for **Research Engineers**, with just **131** positions
- There has been an increase in the number of **Data Engineers** over the years (Paige, 2020) with a **50%** year over year growth rate.
- As per Gartner **80%** of the Data Science projects fail and its up to the data engineers to find the core functionality of the data before it could be used for analysis.



GEOGRAPHICAL ANALYSIS OF SALARIES

HOW HAS THE AVERAGE SALARY EVOLVED IN THE TOP COUNTRIES WITH THE HIGHEST VOLUME OF DATA SCIENCE JOB OPPORTUNITIES?

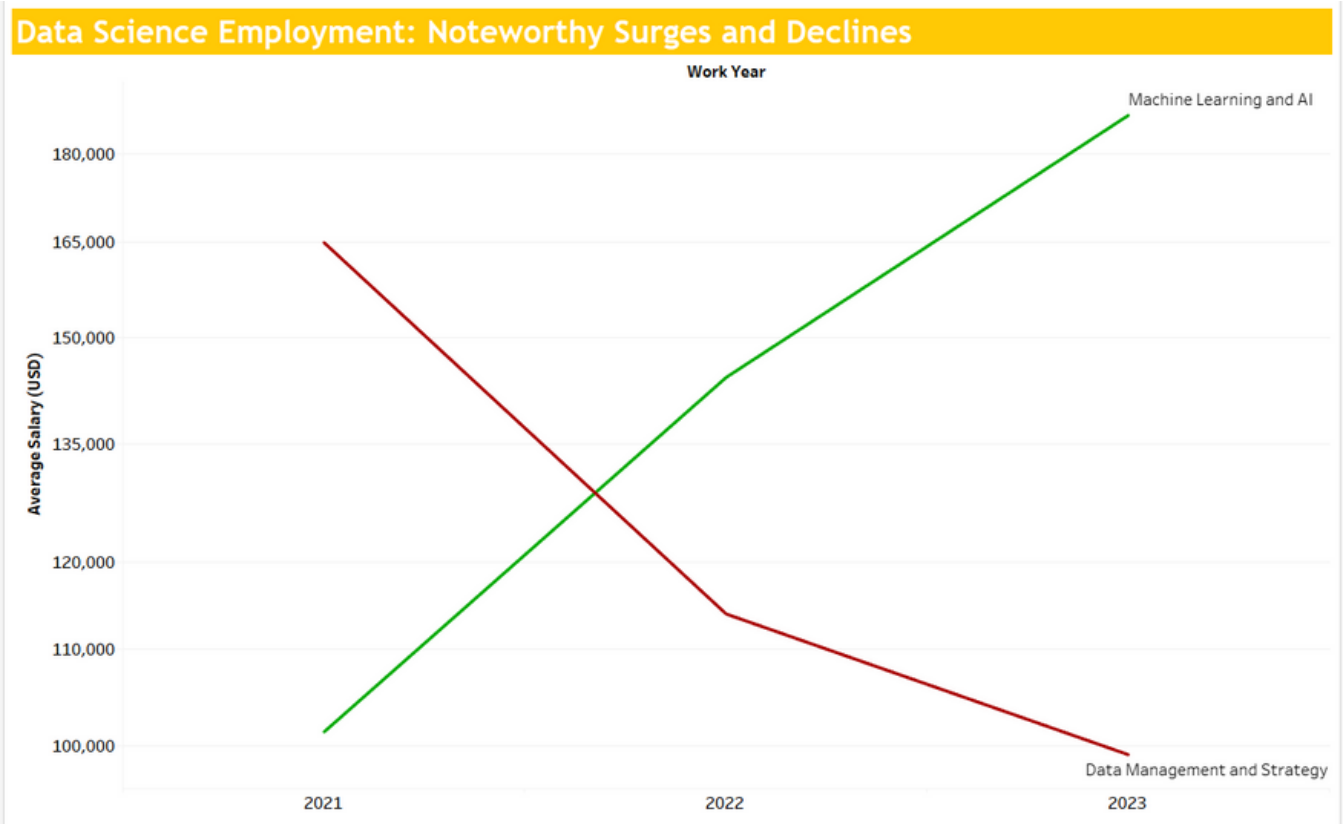
- Salaries for top performers have experienced an **upward trend** since 2020.
- However, there were **declines** in salaries in **2021**, likely stemming from the **repercussions** of the **pandemic** in 2020, which strained organizations' budgets for project funding. (Robinson, 2023)
- European countries typically exhibit greater resilience in managing crises compared to the United States.
- As a result, three out of the four top European countries did not experience significant crises in 2021, leading to stable data science salaries. (ceps. (n.d.).[6])



NOTABLE SALARY SHIFTS IN JOB CATEGORIES

FROM 2020 TO 2023, WHICH JOB CATEGORY EXPERIENCED THE MOST SUBSTANTIAL INCREASE IN AVERAGE SALARY, AND CONVERSELY, WHICH CATEGORY SAW THE MOST SIGNIFICANT DECREASE IN AVERAGE SALARY?

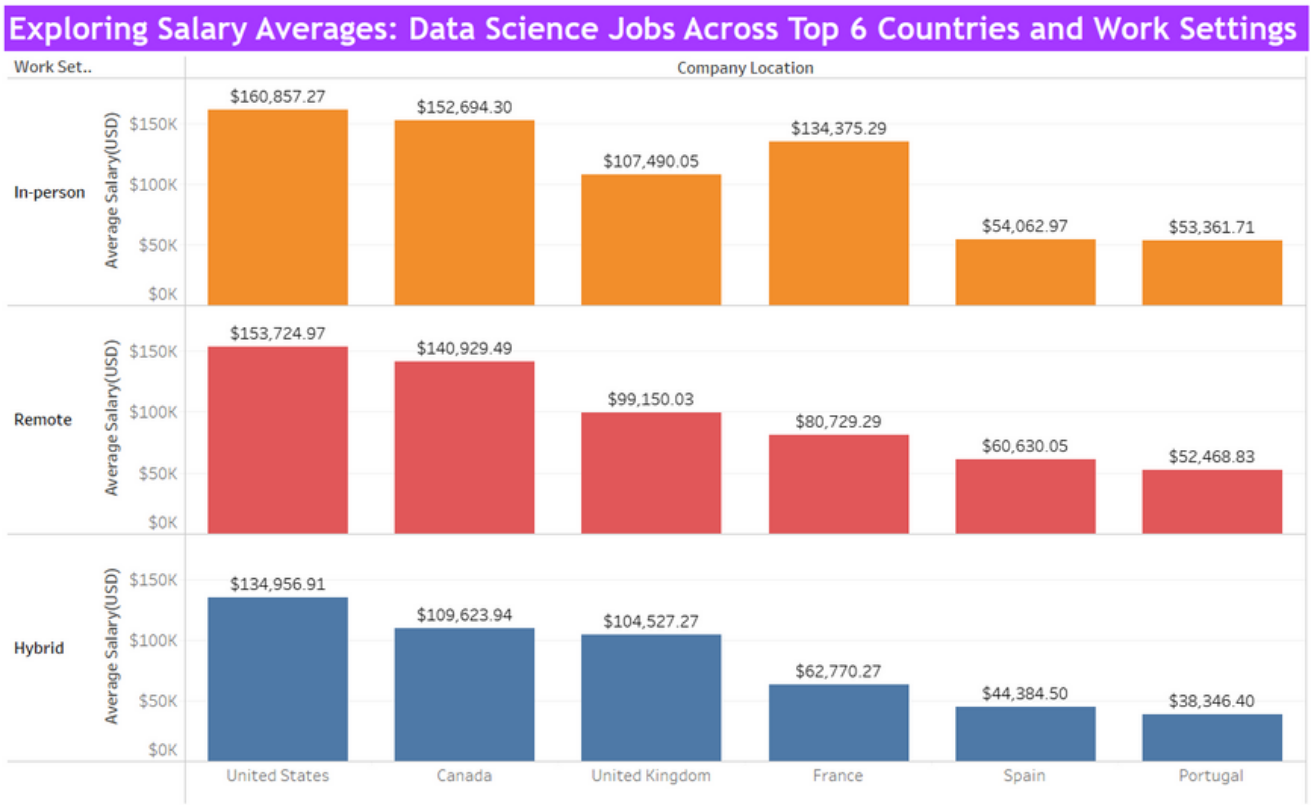
- In recent years, salaries in **Machine Learning and AI** roles have experienced a remarkable surge, reflecting the growing demand for expertise in these fields.
- Conversely, salaries for **Data Management and Strategy** positions have faced significant declines.
- Previously noted, the rising demand for **advanced skills** and **continuous learning** has been the driving force behind the salary escalation in Machine Learning roles.
- Conversely, the **automation** of data management tasks by cloud providers such as AWS and Azure has contributed to a reduction in salaries for these positions. (McCullough, 2023)



SALARY ANALYSIS ACROSS KEY COUNTRIES AND WORK SETTINGS

ARE THERE SIGNIFICANT VARIATIONS IN AVERAGE SALARIES ACROSS DIFFERENT WORK SETTINGS FOR DATA SCIENCE ROLES GLOBALLY?

- As observed, **in-person** positions tend to command the highest salaries compared to hybrid and remote roles.
- A noteworthy finding from this visualization is that remote positions are now being compensated more favorably than hybrid roles, signaling a shift in the emerging workplace dynamics.
- An article from Empire Resume (Gold, 2023) highlights a significant **23% surge** in wages for remote workers compared to their in-person counterparts, particularly among white-collar professionals.
- Nonetheless, there's a contrasting trend emerging as organizations are contemplating wage reductions for remote workers, citing cost savings from reduced overheads such as office space and utilities.



REFERENCES

1. Taylor, P. (2023, November 29). Big Data & Analytics market share by country 2021. Statista. https://www.statista.com/statistics/1258046/worldwide-big-data-business-analytics-market-share-by-country/#:~:text=Big%20data%20and%20business%20analytics%20market%20share%20worldwide%202021%2C%20by%20country&text=In%202021%2C%20the%20United%20States,around%205%20percent%20market%20share.

2. Careerera. (2021, June 22). Which country is best for Data Science Jobs?. Careerera.com. https://www.careerera.com/blog/which-country-is-best-for-data-science-jobs

3. Nkolakowski. (2023, August 18). Which artificial intelligence (A.I.) jobs pay the most?. Dice Insights. https://www.dice.com/career-advice/which-artificial-intelligence-a.i.-jobs-pay-the-most

4. Paige, J. (2020, June 11). Why Data Engineers are more in demand than data scientists. Smith Hanley Associates. https://www.smithhanley.com/2020/06/11/data-engineers-more-in-demand-than-data-scientists/

5. Robinson, D. (2023, September 4). Power Bi Dashboard data science salary analysis. Medium. https://medium.com/@thatsweeetfriend/power-bi-dashboard-data-science-salary-analysis-7fa0a57911fb#:~:text=Well%2C%202020%20was%20a%20tricky,went%20on%20a%20positive%20streak.

6. Why European companies might be better equipped to cope with the ... - ceps. (n.d.). https://www.ceps.eu/why-european-companies-might-be-better-equipped-to-cope-with-the-crisis-than-their-us-counterparts/

7. McCullough, D. (2023, December 7). 10 ways to improve data management with automation. Pure Storage Blog. https://blog.purestorage.com/perspectives/10-ways-to-improve-data-management-with-automation/

8. Gold, M. (2023, September 7). Do remote jobs pay less money?. Empire Resume. https://empireresume.com/do-remote-jobs-pay-less-money/

GOT ANY QUESTIONS?

E-mail me at **denzild@uw.edu**

Visualizations: Tableau Public Sanity Checks: Jupyter Notebook