

# Data Salaries Analysis

---

Team #4:

Javier Alarcon Velazco

Eugenio A. Castillo Elizondo

Oscar Tlacuilo

Priscila Elieth Matus Díaz

# Introduction

This dataset provides comprehensive insights into data science job salaries spanning from 2020 to 2024. It encompasses various dimensions of employment, including experience levels, job titles, employment types, and company characteristics.

## Objective:

- Benchmarking salaries across various parameters.
- Uncovering relationships between salary and factors like experience, job type, and company size.
- Empowering career decisions through data-driven insights."

# Questions Formulation

01

## Highest Salaries

Which job titles have the highest salaries?

03

## Salary Increase

Was there a salary increase between 2023 and this year?

02

## US Salary

Compare salaries between the United States and the rest of the countries in the dataset.

04

## Company Size

Is there a relationship between company size and country of origin?

# Questions Formulation

05

## Company Size

Does the company size influence the salary?

07

## Experience Level

What is the distribution of experience levels among data science professionals?

06

## Remote Work

How much does remote job impact salary?

08

## Companies Sizes

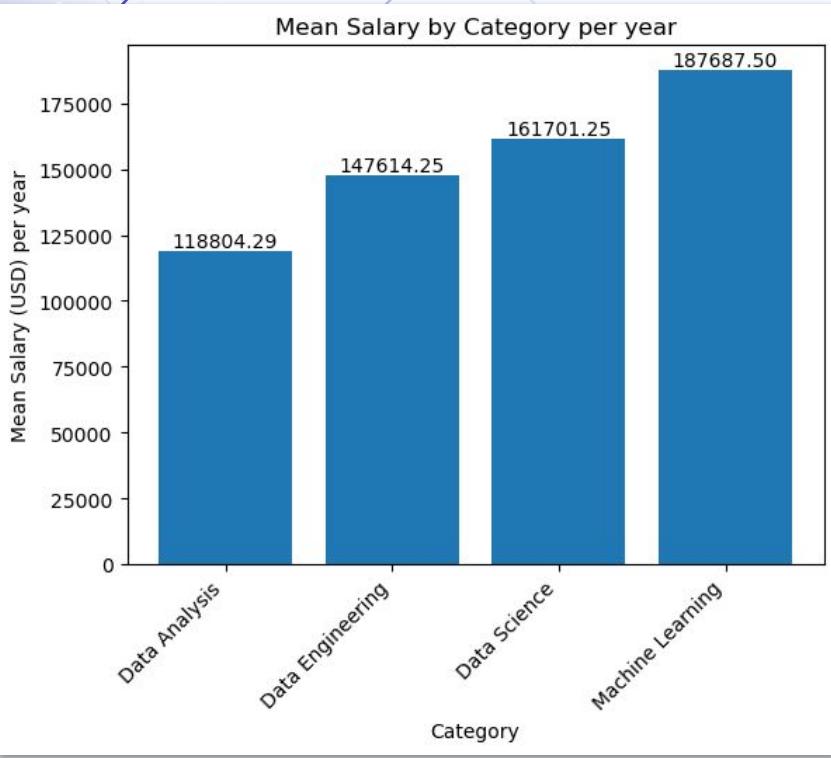
Depending on the size of the company, what do companies tend to look for more?

# 01

Which job titles have  
the highest salaries?

---





- Mean Salaries by Category
- Analysis of Standard Deviation
- Inclusion of International Data

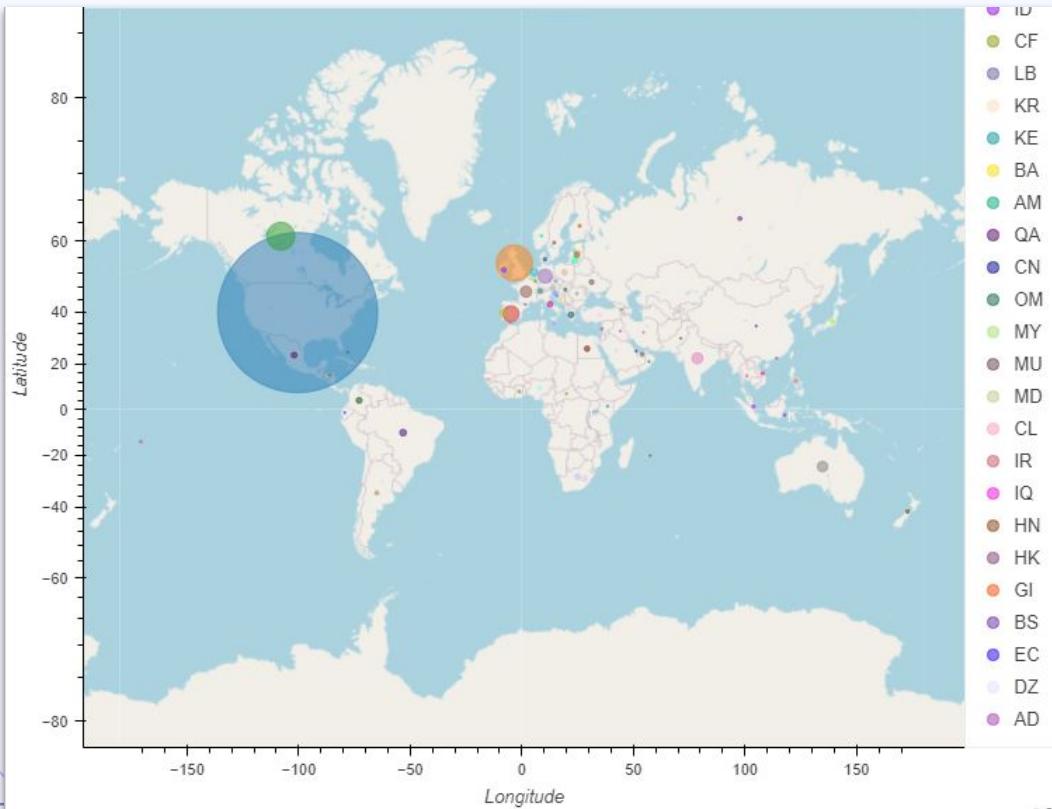
	count	mean	std	min	25%	50%	75%	max
category								
data_analysis	3326.0	118804.292544	56628.107362	15000.0	81050.0	110000.0	146962.75	774000.0
data_engineering	3685.0	147614.251289	57945.045153	17025.0	107000.0	140000.0	180000.00	750000.0
data_science	4653.0	161701.251666	69154.395573	16000.0	120000.0	155000.0	200000.00	750000.0
machine_learning	1892.0	187687.502643	70861.536009	15966.0	142200.0	182000.0	226600.00	800000.0

# 02

**Compare salaries between the United States and the rest of the countries in the dataset.**

---

# World Map of Job entries



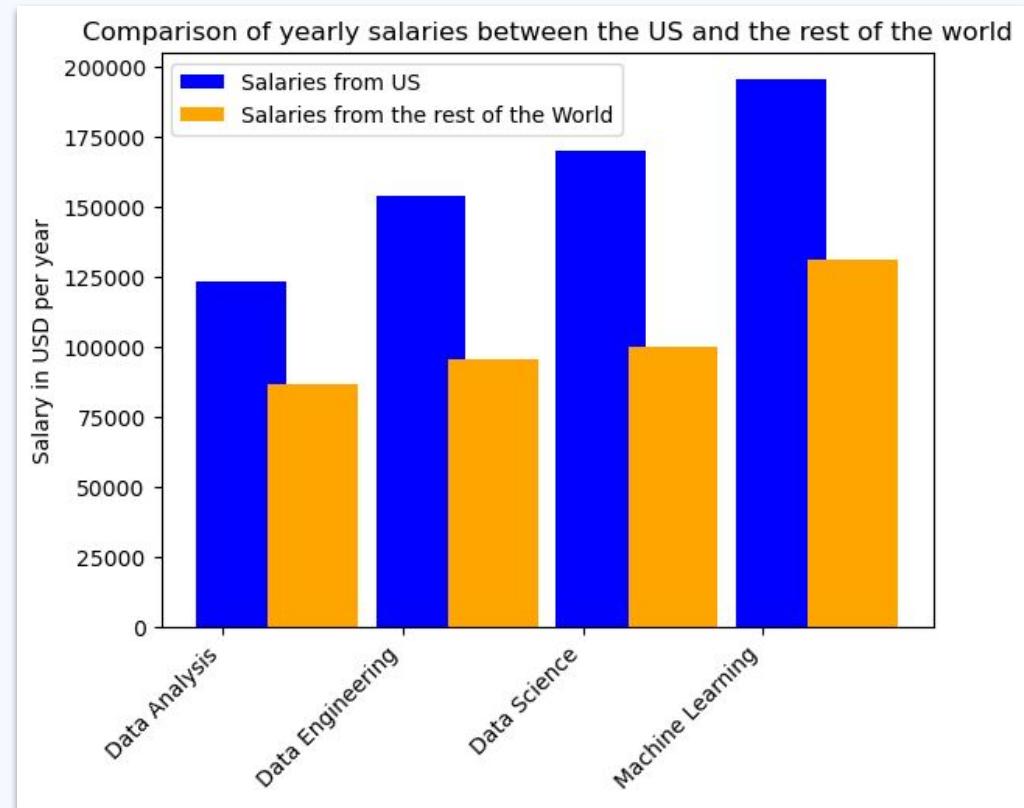
- Employing nominatim.openstreetmap API
- Highlights concentration in United States
- Provides insight into global distribution

# Yearly salaries US vs the rest of the world

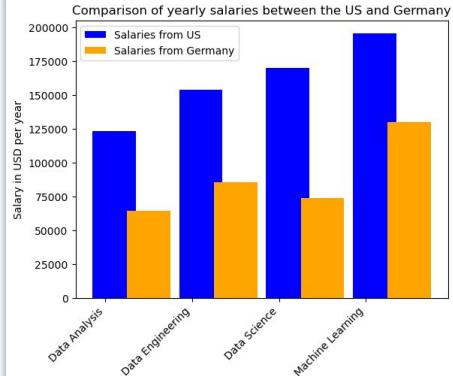
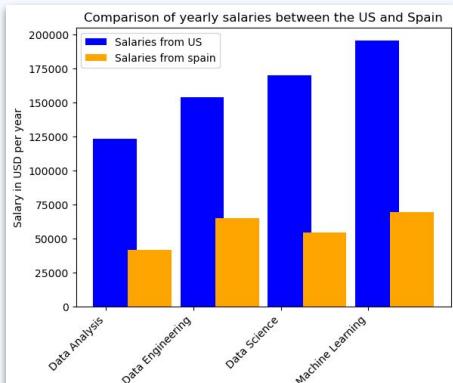
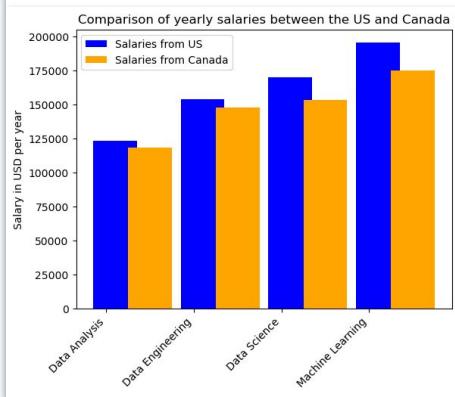
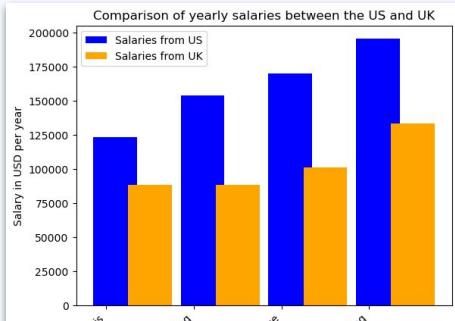
```
count    12975.000000
mean     157410.126474
std      66151.750528
min      20000.000000
25%     111600.000000
50%     148000.000000
75%     191475.000000
max      750000.000000
```



```
count    1863.000000
mean     97393.819646
std      65608.455145
min      15000.000000
25%     51676.000000
50%     79981.000000
75%     123040.000000
max      800000.000000
```



# Yearly salaries Top 4 countries entries count

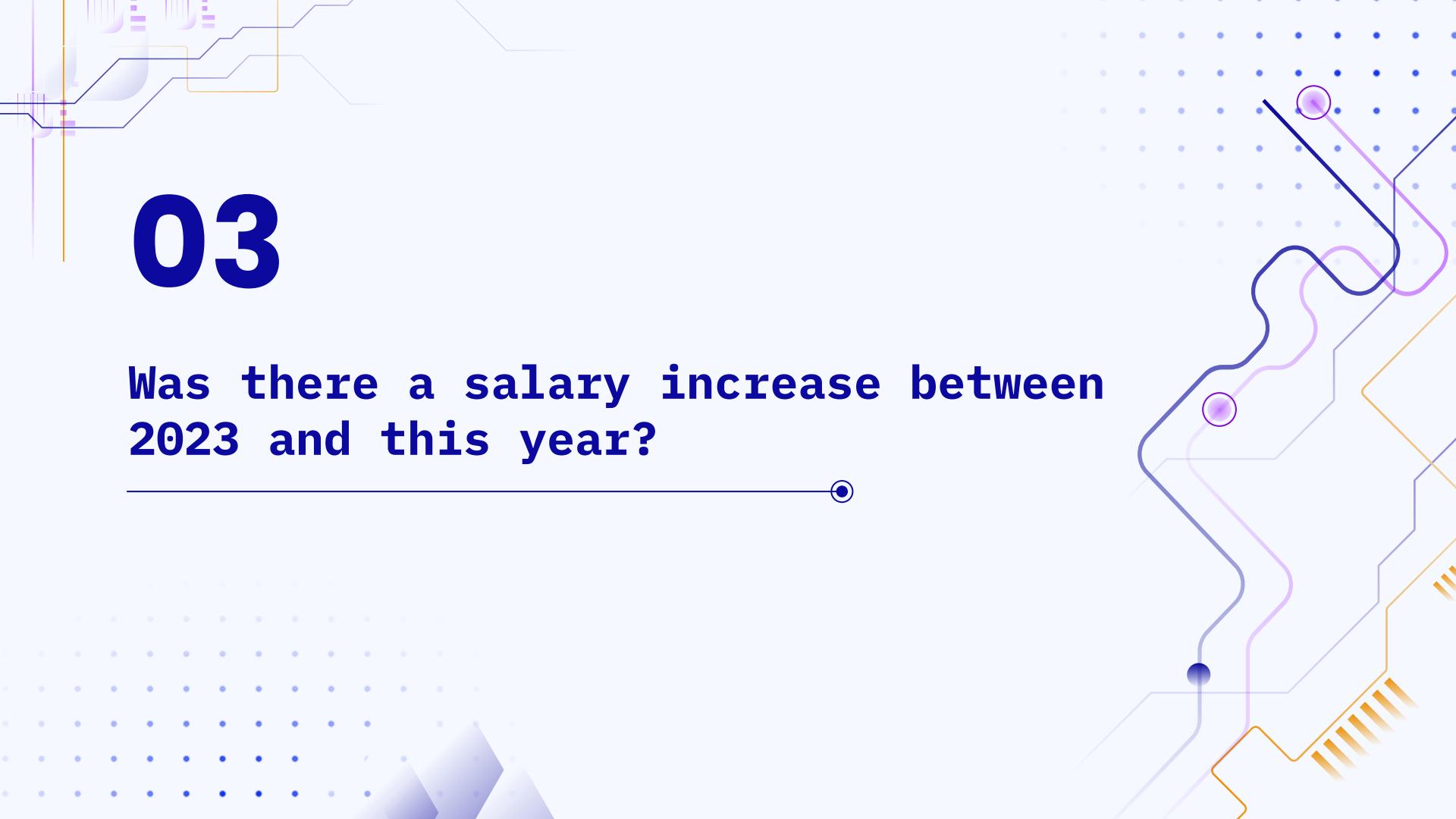


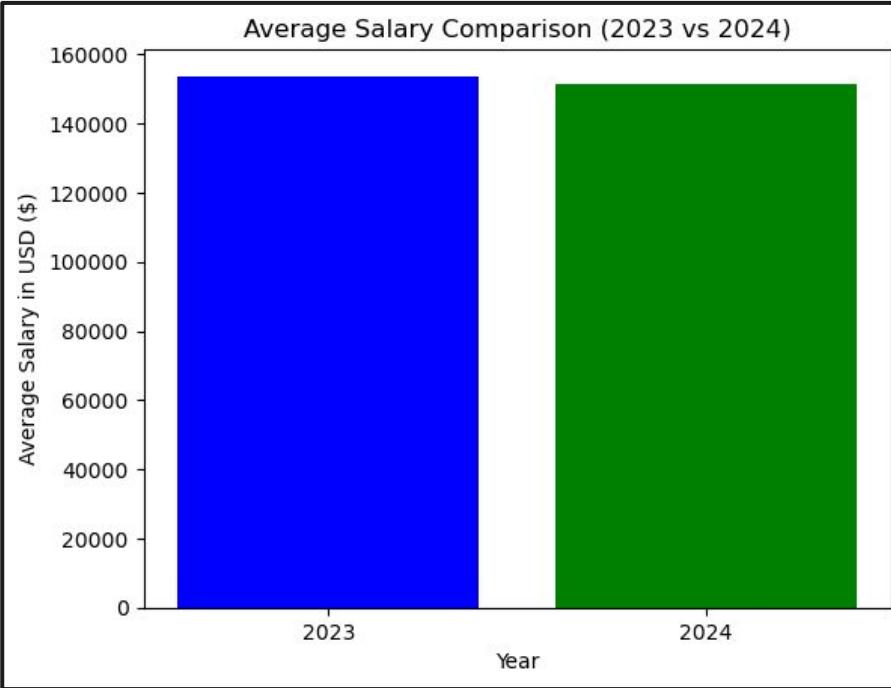
Location	Count
UK	655
Canada	392
Spain	127
Germany	98

# 03

Was there a salary increase between  
2023 and this year?

---





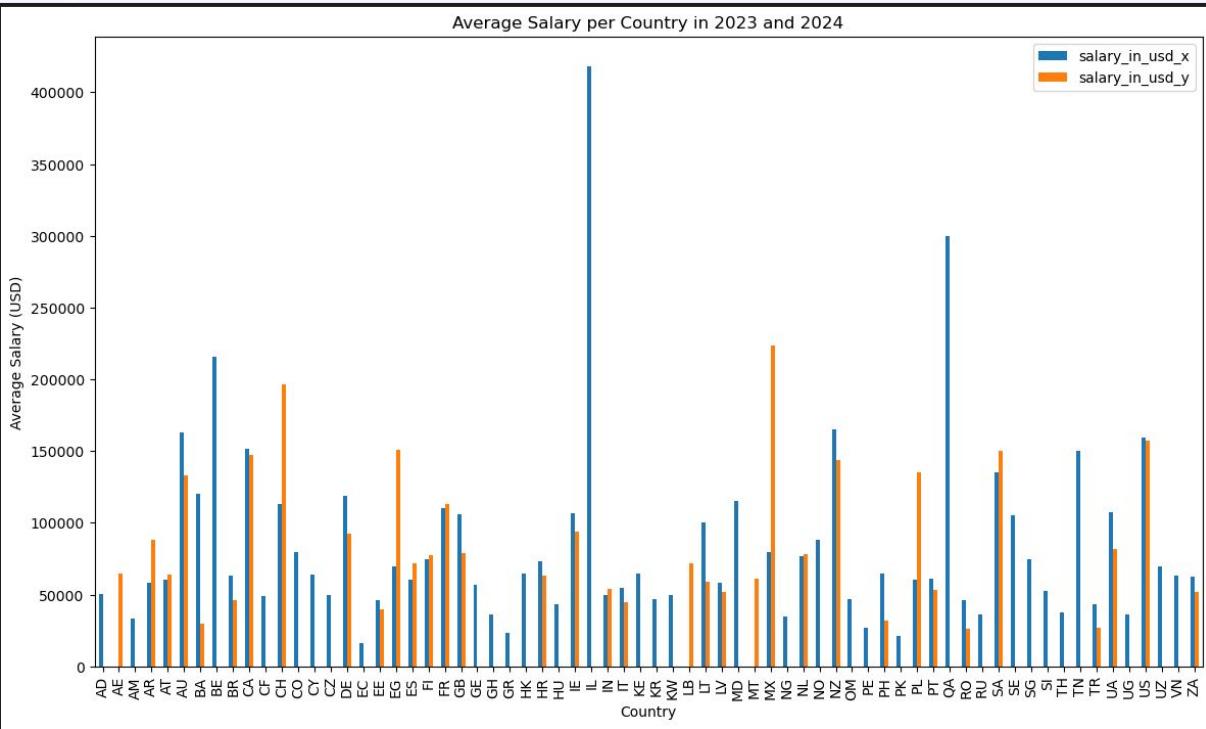
work_year	salary_in_usd	
10	2023	60805
11	2023	60207
16	2023	48644
19	2023	43779
21	2023	38915

work_year	salary_in_usd	
20	2024	39615
33	2024	27611
66	2024	800000
68	2024	774000
69	2024	750000

**Average 2023 Salary = \$153,732.66**  
**Average 2024 Salary = \$151,510.09**

## 2023 Salaries

	employee_residence	salary_in_usd
0	AD	50745.000000
1	AM	33500.000000
2	AR	58333.333333
3	AT	60462.000000
4	AU	163288.823529



## 2024 Salaries

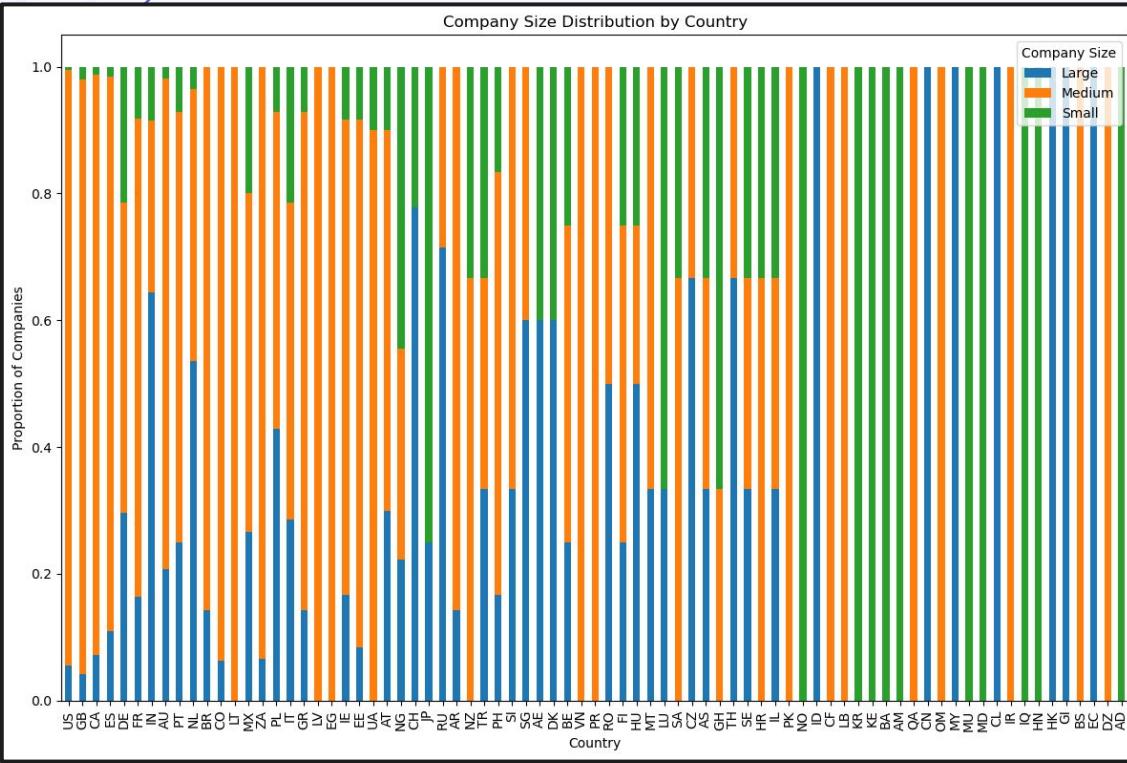
	employee_residence	salary_in_usd
0	AE	65000.000000
1	AR	88500.000000
2	AT	64166.000000
3	AU	133389.565217
4	BA	30000.000000

# 04

**Is there a relationship between  
company size and country of origin?**

---





company_size	L	M	S	total_companies	Large	Medium	Small
company_location							
US	723	12185	67	12975	0.055723	0.939114	0.005164
GB	27	615	13	655	0.041221	0.938931	0.019847
CA	28	359	5	392	0.071429	0.915816	0.012755
ES	14	111	2	127	0.110236	0.874016	0.015748
DE	29	48	21	98	0.295918	0.489796	0.214286
...	...	...	...	...	...	...	...
GI	1	0	0	1	1.000000	0.000000	0.000000
BS	0	1	0	1	0.000000	1.000000	0.000000
EC	1	0	0	1	1.000000	0.000000	0.000000
DZ	0	1	0	1	0.000000	1.000000	0.000000
AD	0	0	1	1	0.000000	0.000000	1.000000

# 05

Does the company size influence the salary?

---

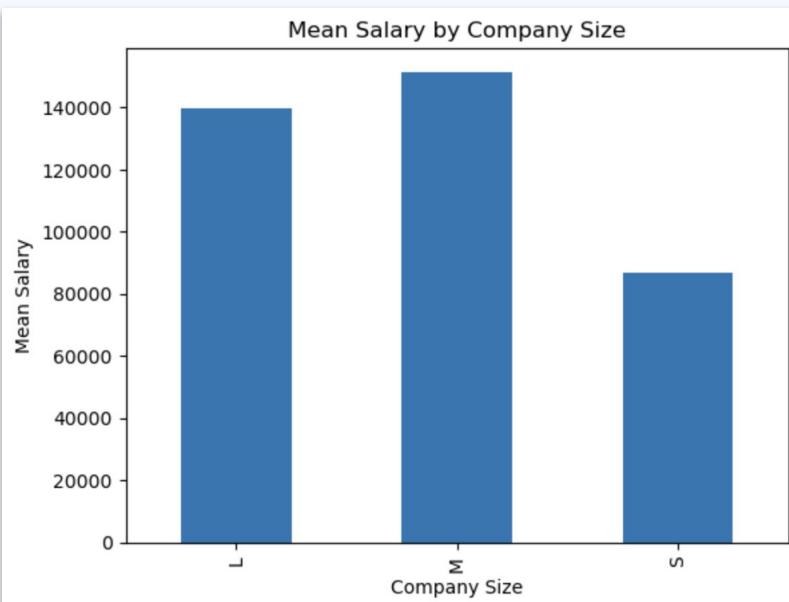


# Company Size

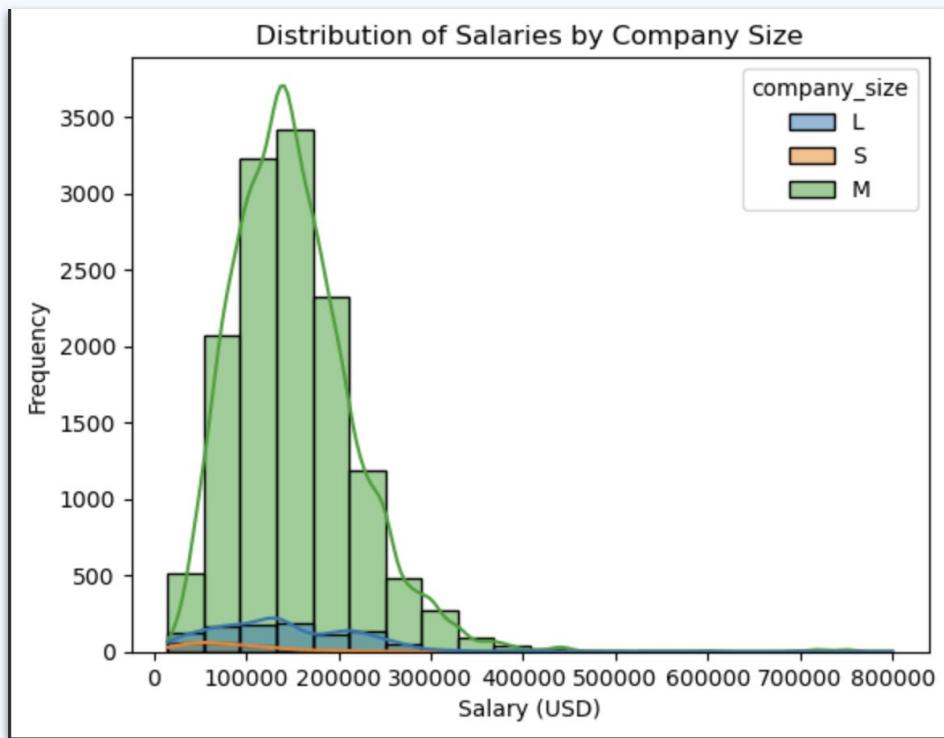
Statistics of Salary by Company Size:

Company Size	Count	Mean	Std	Min	25%	50%	75%	
L	983	139602.46	74183.85	15000.0	79833.0	136000.0	200000.0	423000.0
M	13674	151450.54	68295.97	15000.0	105000.0	143150.0	186000.0	800000.0
S	181	86614.57	58501.97	15809.0	50000.0	70179.0	114047.0	416000.0

- Medium size companies are paying better salaries



# Company Size



- Histogram of 20 bins

# 06

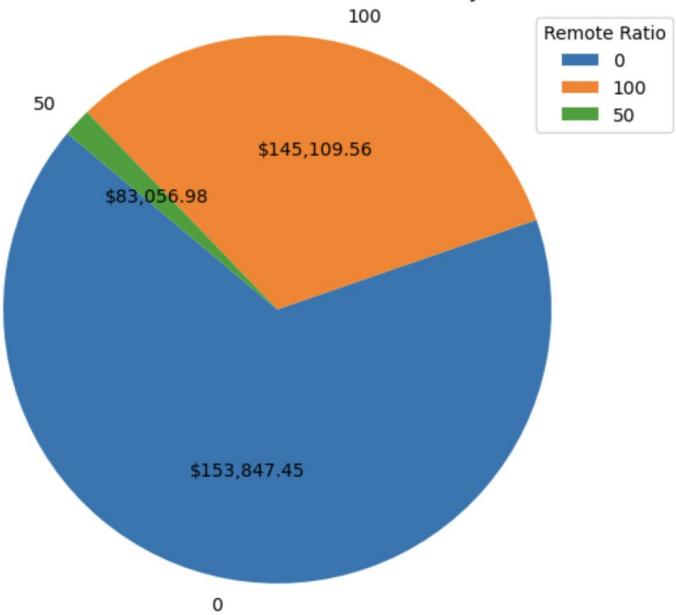
## How much does remote jobs impact salary

---

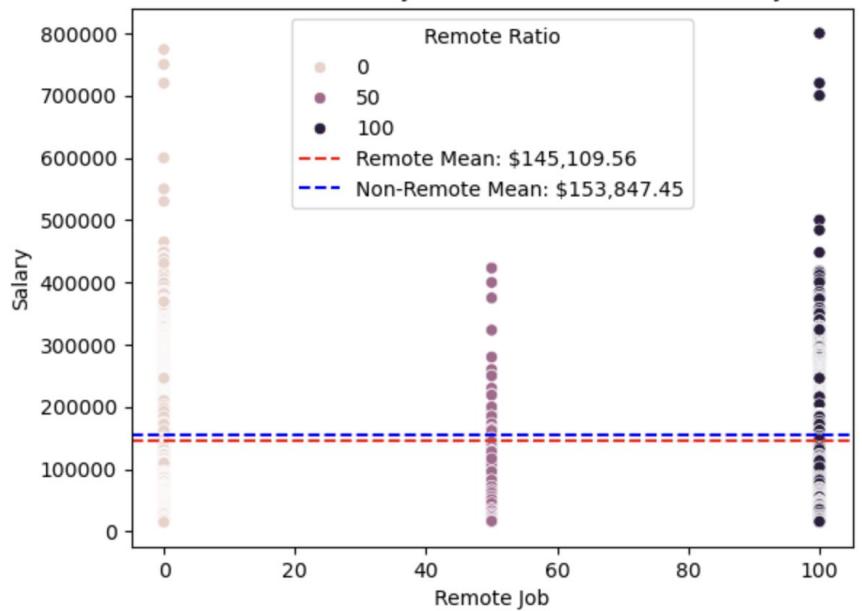


# Remote Job

Distribution of Remote vs. Non-Remote Jobs

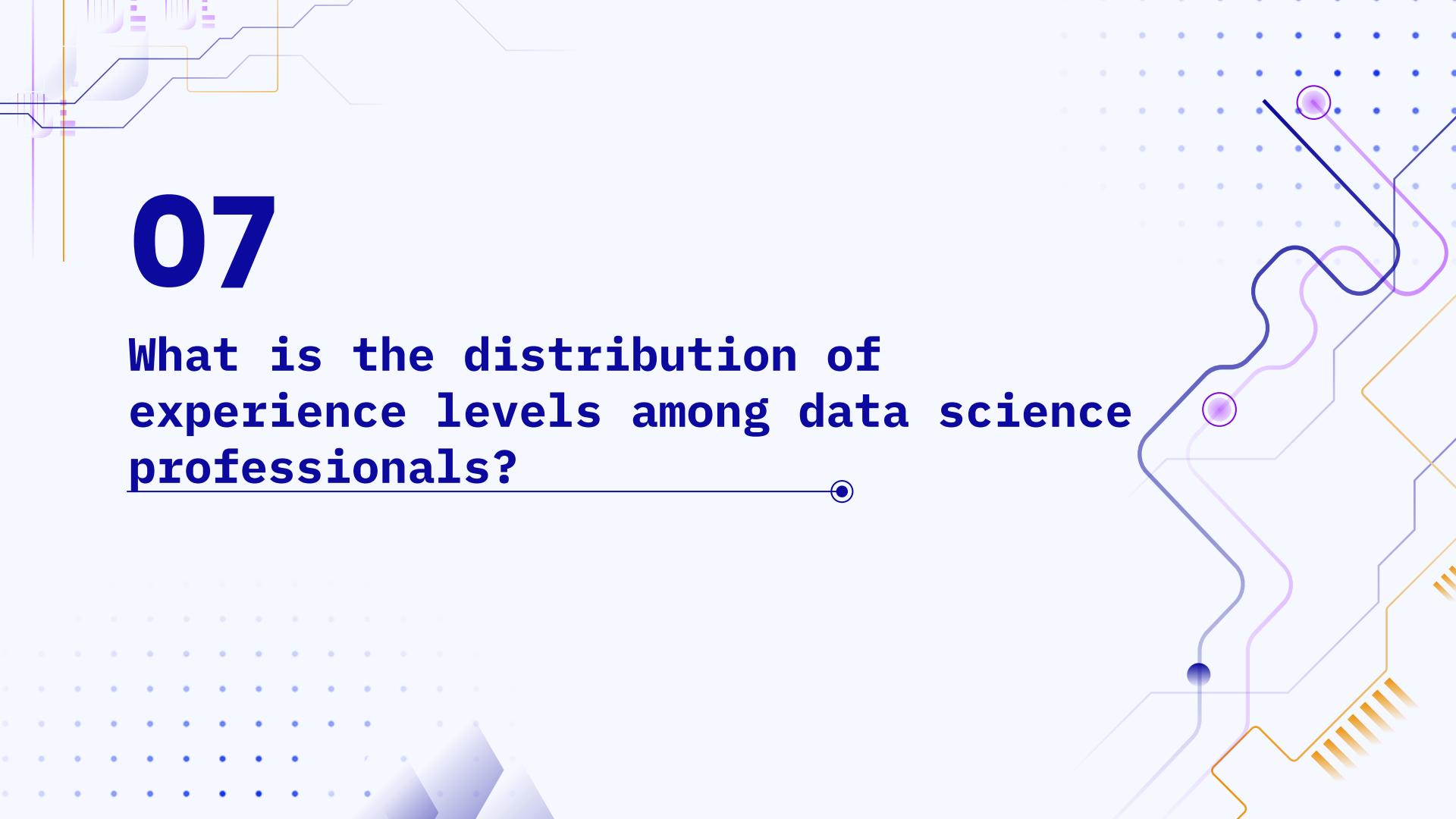


Scatter Plot of Salary for Remote and Non-Remote Jobs

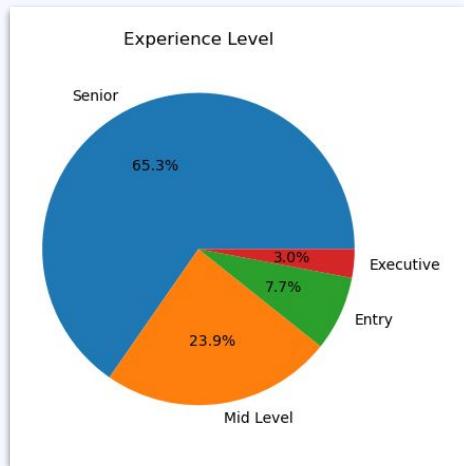


# 07

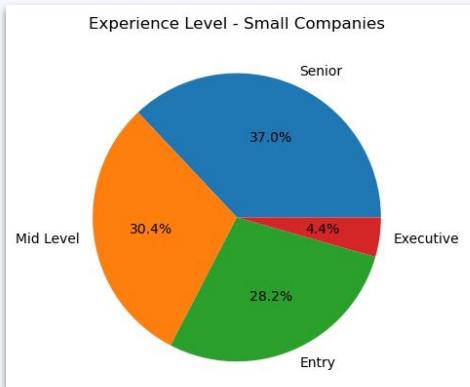
**What is the distribution of experience levels among data science professionals?**



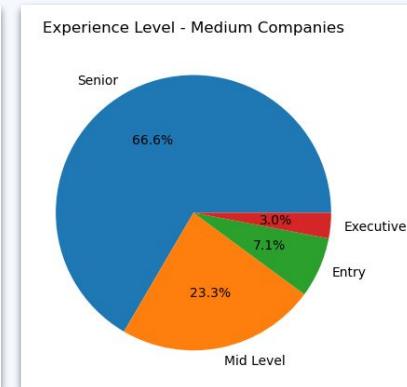
# Experience Level by Companies Size



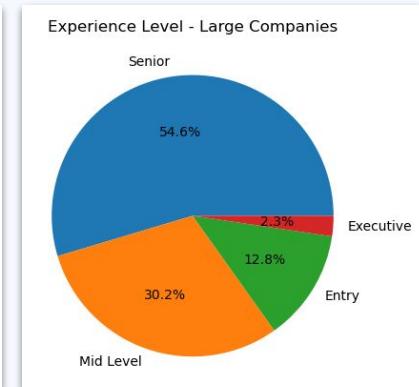
GENERAL: **SENIOR**  
(2020-2024)



SMALL COMPANIES: **SENIOR – MID LEVEL**



MEDIUM COMPANIES: **SENIOR**



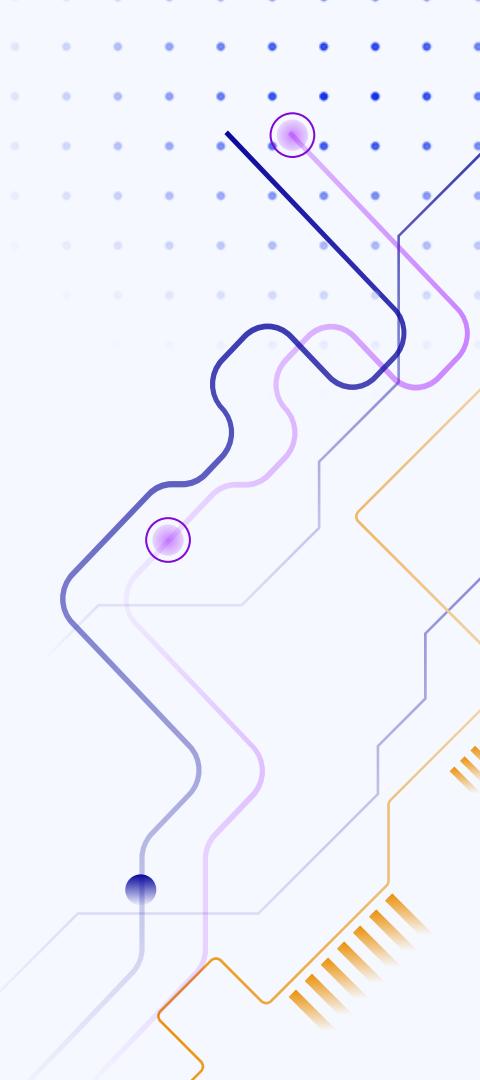
LARGE COMPANIES: **SENIOR**

	count	unique	top	freq
company_size				
L	983	4	SE	537
M	13674	4	SE	9104
S	181	4	MI	67

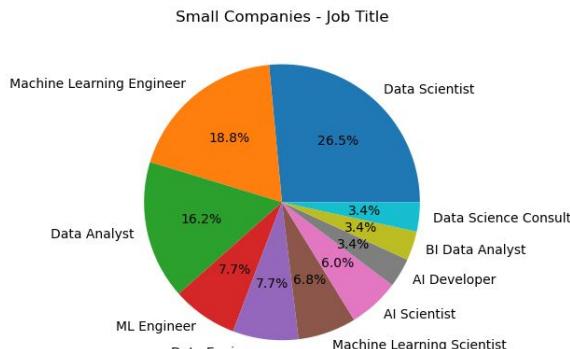
# 08

**Depending on the size of the company, what do companies tend to look for more?**

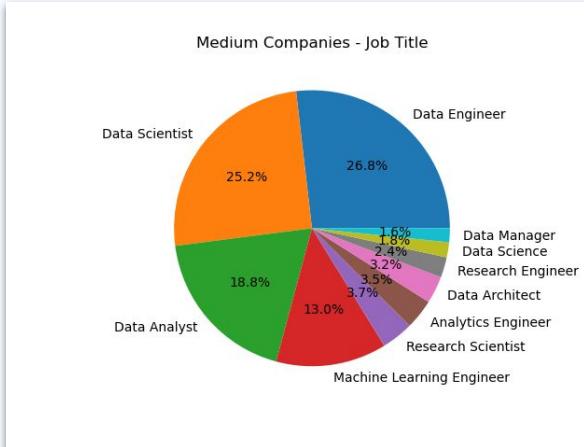
---



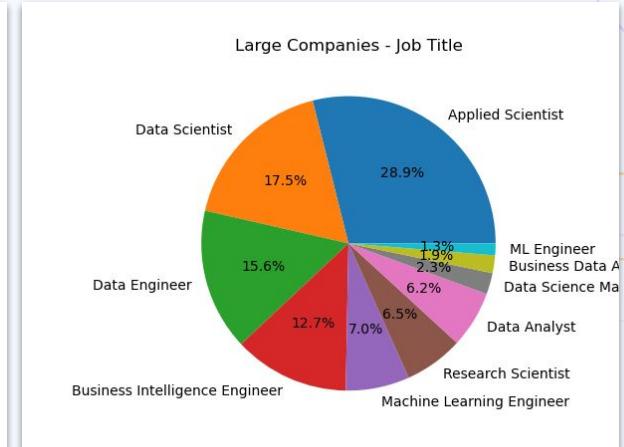
# Job Title by Companies Size



Small Companies: **Data Scientist**



Medium Companies: **Data Engineer**



Large Companies: **Applied Scientist**

	count	unique	top	freq
company_size				
L	983	95	Applied Scientist	223
M	13674	126	Data Engineer	3033
S	181	50	Data Scientist	31

# Conclusions

---

# Conclusions

## **1. Which job titles have the highest salaries?**

Machine learning roles lead with highest average salaries, while data analysis and engineering are comparable. Categorizing by job frequency provides insight into broader trends. Data science and machine learning show greater salary variability, reflecting diverse industry roles.

## **2. Compare salaries between the United States and the rest of the countries in the dataset.**

In comparing US salaries to those globally, despite the initial disparity in dataset representation, deeper analysis revealed lower wages in European countries, with Canadian salaries closer to US levels, emphasizing the nuanced influence of geographical location and cost of living on global wage disparities.

# Conclusions

## **3. Was there a salary increase between 2023 and this year?**

There was no increase between 2023 and 2024 salary wise, in fact it was the other way around, as 2024 experimented a 1.44% decrease in salary.

On the other hand, diving deeper into each country, we found that 17.65% of the countries in the dataset had an increase in salary, 27.94% had a decrease in salary and 54.41% did not have a point of comparison.

## **4. Is there a relationship between company size and country of origin?**

We determined that countries that have significantly more business activity have more “Medium” established companies as they are the most popular. Further down the list of countries in the dataset we found that countries with less business activity tend to have more “Small” companies.



# Conclusions

## **5. Does the company size affects the salary?**

We can say it doesn't. Because Medium size companies offer better salaries than Larger ones.

## **6. How much does remote or in-site jobs impact salary?**

There is no significant difference in-between. The difference comes for 50/50 jobs

# Conclusions

## **7. What is the distribution of experience levels among data science professionals?**

The results indicate that there is a large growth of work for those with senior experience from small to large companies, however in small companies, there is a slight margin between MID and senior level that we can quantify the state of the differences, they are not statistically significant.

## **8. Depending on the size of the company, what do companies tend to look for more?**

The companies have a tendency in the area of data regardless of size, the only variance is the area in which they perform and manage this data, we could observe that the same job titles are conserved but they move according to the size of the company.

# Next steps/limits

**Expanding the project with more questions and exploring relations between columns. Consider examining the correlation between the cost of living and salaries for comparison. Additionally, developing a dashboard or web app for user-friendly data representation.**

## Analysis limitations:

- Data entry may lack precision.
- Clarify the percentage breakdown between gross salary and annual compensation in money versus stock.
- The dataset is heavily biased towards USA salaries, potentially limiting comparability with other countries.

# Thanks !

## Team #4:

Javier Alarcon Velazco

Eugenio A. Castillo Elizondo

Oscar Tlacuilo

Priscila Elieth Matus Díaz