

# Data Science Salaries Analysis

Dinan Sooriyaarachchi

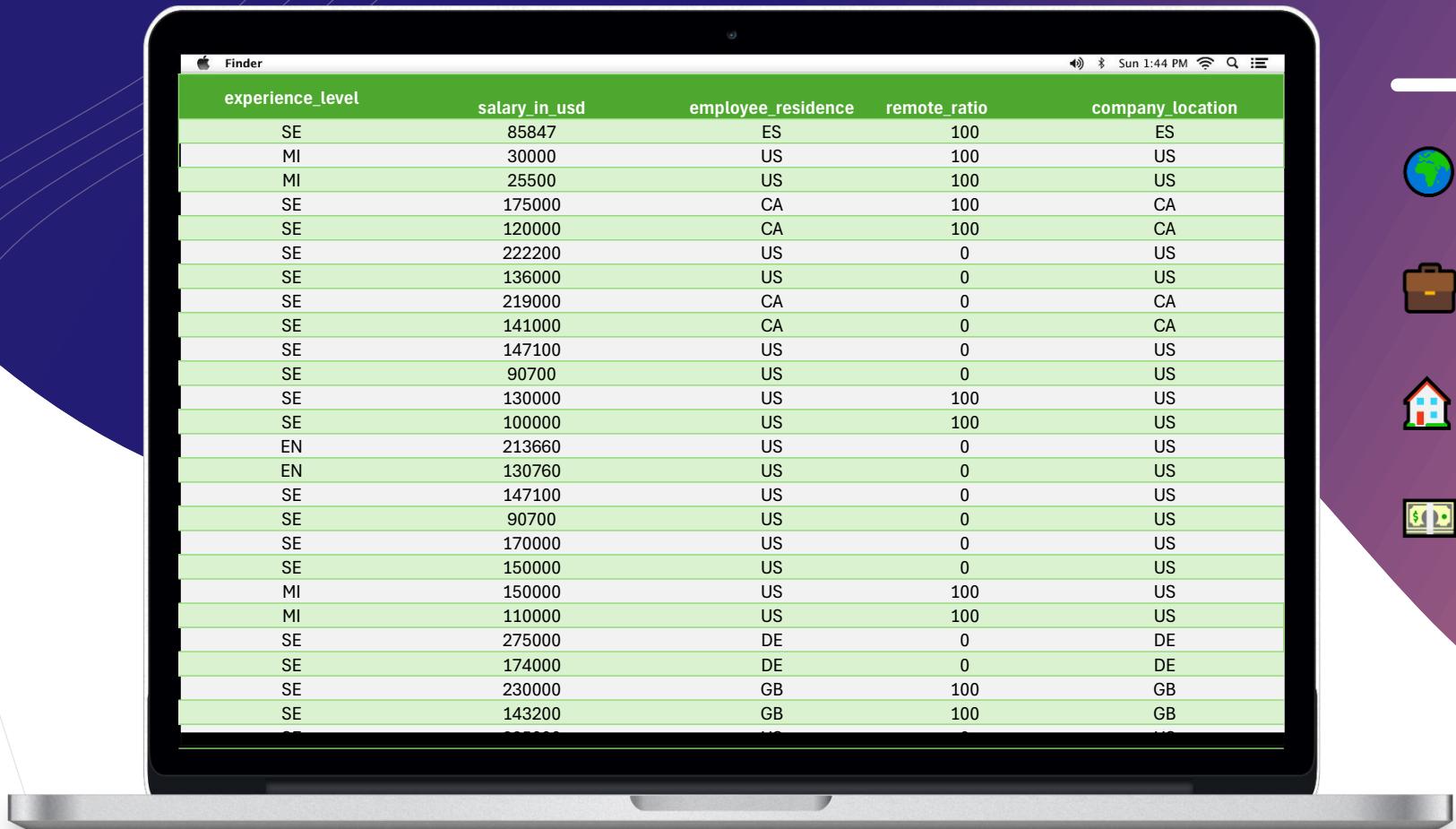
# THE PROBLEM

---

**How do experience level, work type (remote vs. on-site), and geographical location collectively influence the salaries of data science professionals?**

- The data science field is rapidly expanding with compensation becoming more complex due to factors like remote work, experience, and location.
  - This analysis helps uncover the patterns to better navigate career decisions and optimize salary offerings
  - Understanding the relationship between these factors and salary is critical for both data professionals and employers to make informed decisions
-

# Dataset Overview



A laptop screen showing a dataset in a Finder window. The window title is 'Finder' and the path is 'Macintosh HD /'. The dataset is presented as a table with the following columns: experience\_level, salary\_in\_usd, employee\_residence, remote\_ratio, and company\_location. The data consists of approximately 30 rows, with values such as SE, 85847, ES, 100, ES; MI, 30000, US, 100, US; and various other entries for different experience levels, salaries, residences, and locations.

experience_level	salary_in_usd	employee_residence	remote_ratio	company_location
SE	85847	ES	100	ES
MI	30000	US	100	US
MI	25500	US	100	US
SE	175000	CA	100	CA
SE	120000	CA	100	CA
SE	222200	US	0	US
SE	136000	US	0	US
SE	219000	CA	0	CA
SE	141000	CA	0	CA
SE	147100	US	0	US
SE	90700	US	0	US
SE	130000	US	100	US
SE	100000	US	100	US
EN	213660	US	0	US
EN	130760	US	0	US
SE	147100	US	0	US
SE	90700	US	0	US
SE	170000	US	0	US
SE	150000	US	0	US
MI	150000	US	100	US
MI	110000	US	100	US
SE	275000	DE	0	DE
SE	174000	DE	0	DE
SE	230000	GB	100	GB
SE	143200	GB	100	GB

 Location

 Experience Level

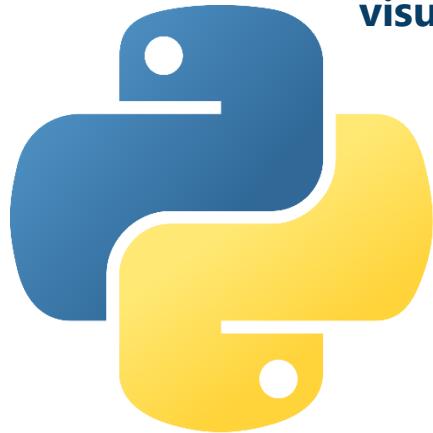
 Remote Work Ratio

 Salary in USD

# TOOLS AND METHODS USED FOR DATA ANALYSIS

## Python

Used for the primary tools for data-cleaning, in-depth analysis, and data visualization

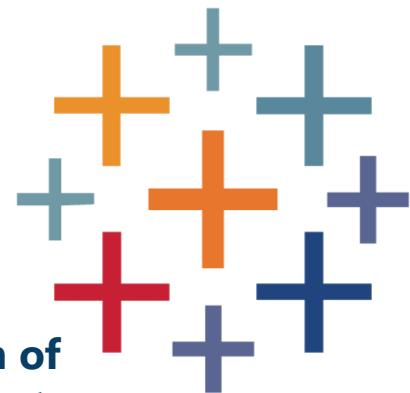


### Libraries:

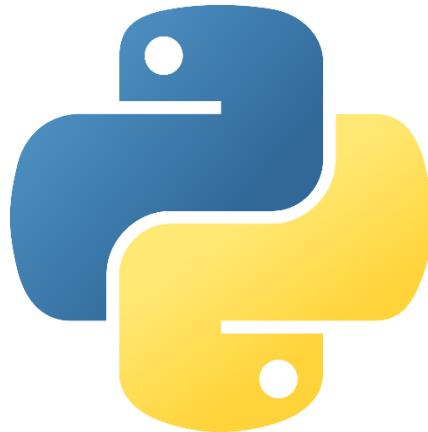
- Pandas
- Seaborn
- Matplotlib

## Tableau

Used to create interactive visualizations, allowing for better exploration of the dataset.



# Data Cleaning Code in Python



- **Checked for missing values:**
  - Ensured there were no missing entries in any of the columns.
- **Standardized column names:**
  - Converted all column names to lowercase and replaced spaces with underscores for consistency.
- **Ensured consistency in categorical columns:**
  - Converted 'experience\_level' uppercase for uniformity.
- **Removed duplicates:**
  - Checked for and removed any duplicate rows in the dataset.

## Checking for missing values

```
[ ] print("Missing values in each column:\n", salaries.isnull().sum())
```

```
Missing values in each column:  
work_year          0  
experience_level   0  
employment_type    0  
job_title          0  
salary              0  
salary_currency    0  
salary_in_usd      0  
employee_residence 0  
remote_ratio        0  
company_location   0  
company_size        0  
dtype: int64
```

## Make all the column names to a consistent lowercase and replace with underscores

```
[ ] salaries.columns = salaries.columns.str.lower().str.replace(' ', '_')
```

## Ensure consistency in categorical columns

Convert 'experience\_level' and 'remote\_ratio' to uppercase for consistency

```
[ ] salaries['experience_level'] = salaries['experience_level'].str.upper()  
salaries['remote_ratio'] = salaries['remote_ratio'].astype(str)
```

## Remove duplicate (if any)

```
[ ] salaries = salaries.drop_duplicates()
```

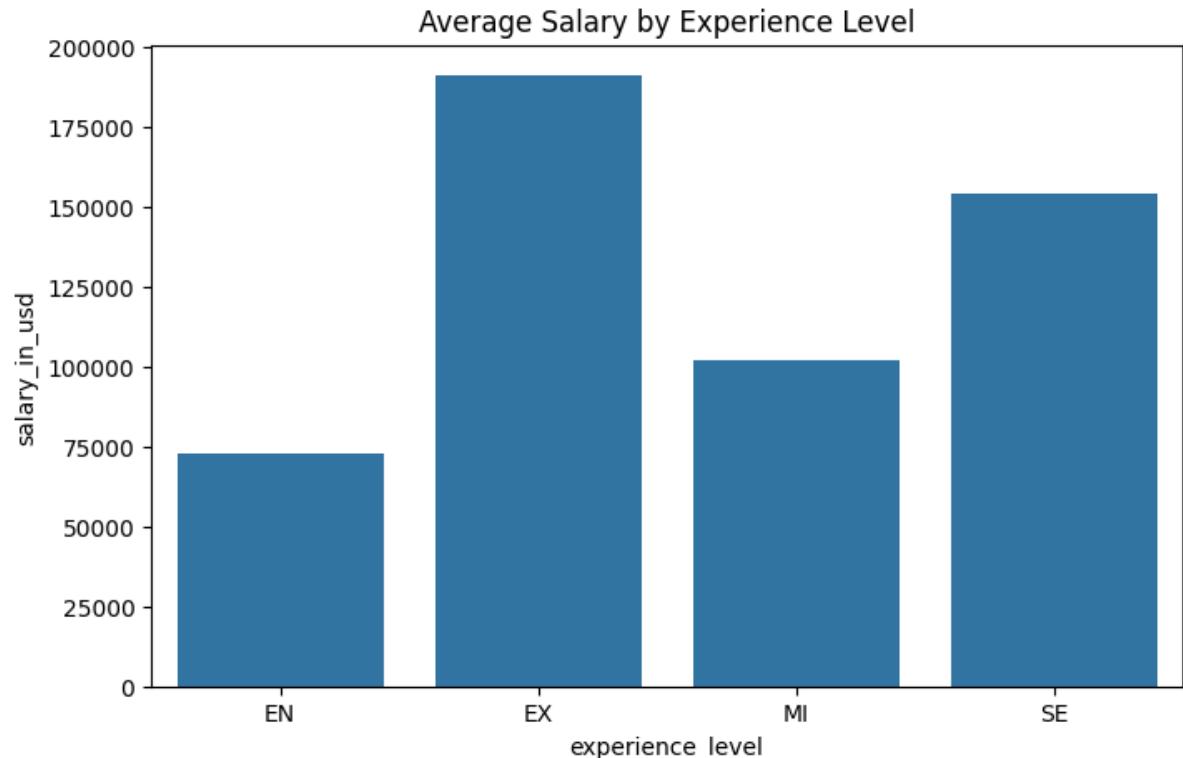
## Preview Cleaned Data Set

```
[ ] print(salaries.head())
```

	work_year	experience_level	employment_type	job_title	salary	salary_currency	salary_in_usd	employee_residence	remote_ratio
0	2023	SE	FT	Principal Data Scientist	80000	EUR	85847	ES	100
1	2023	MI	CT	ML Engineer	30000	USD	30000	US	100
2	2023	MI	CT	ML Engineer	25500	USD	25500	US	100
3	2023	SE	FT	Data Scientist	175000	USD	175000	CA	100
4	2023	SE	FT	Data Scientist					

# Salary by Experience Level

(EN)  
(EX)  
(MI)  
(SE)



- Entry-level professionals (EN) earn significantly less compared to executive-level (EX) professionals, who command the highest salaries.
- Companies are willing to pay a premium for experienced professionals, which might reflect the value of expertise and leadership in data science role

# Salary by Experience Level

```
[2]: import matplotlib.pyplot as plt  
import seaborn as sns
```

- ▼ Analyzing experience level impact on salary

```
[ ] experience_salary = salaries.groupby('experience_level')['salary_in_usd'].mean().reset_index()
```

- ▼ Plotting experience level vs salary

```
[ ] plt.figure(figsize=(8, 5))  
sns.barplot(data=experience_salary, x='experience_level', y='salary_in_usd')  
plt.title('Average Salary by Experience Level')  
plt.show()
```

Code Snippet

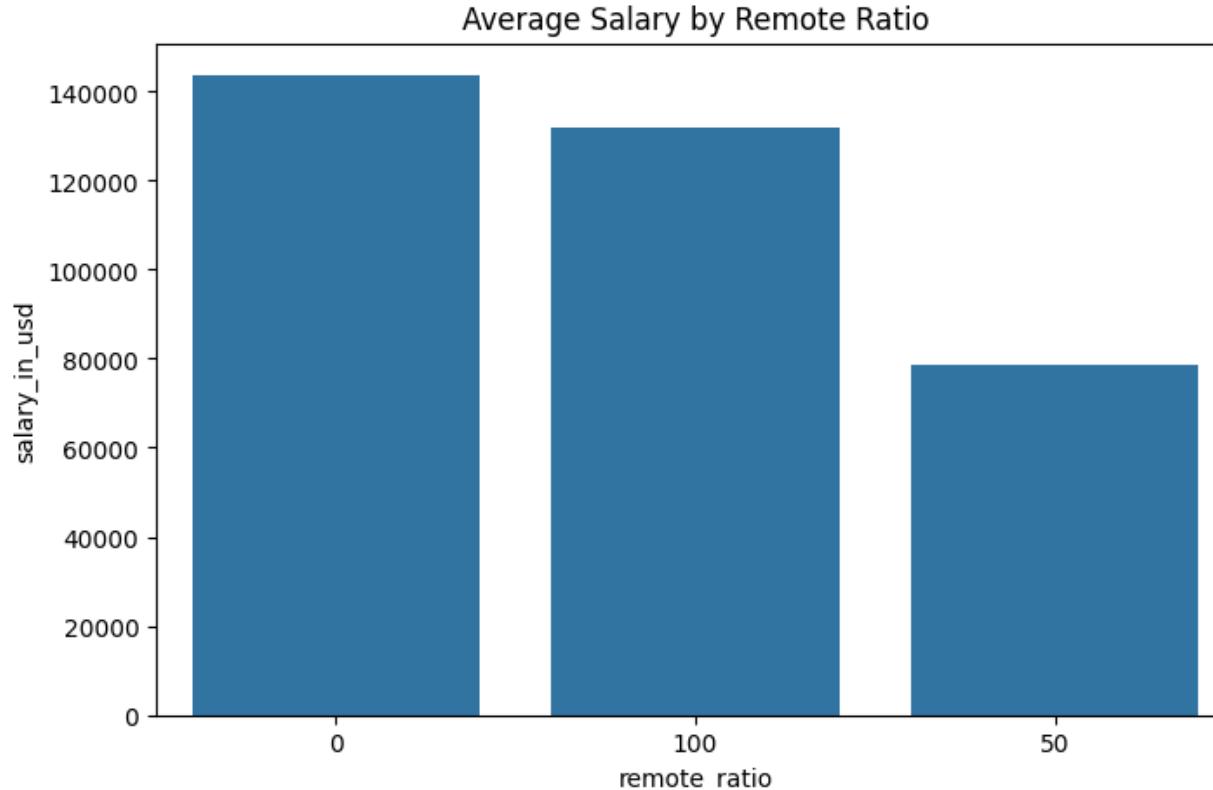
# Remote Work Impact on Salary

## Series Explanation

Fully on-site (0%) and fully remote (100%) roles have higher average salaries than hybrid roles (50%).

## Insight

Remote work flexibility or on-site roles may require companies to offer higher pay to attract talent.



## Additional Observation

Hybrid roles (50%) tend to offer lower salaries, reflecting less flexibility or incentive

# Remote Work Impact on Salary

## Code Snippet

### ✓ Analyzing Remote work impact on Salary

```
[ ] remote_salary = salaries.groupby('remote_ratio')['salary_in_usd'].mean().reset_index()
```

### ✓ Plotting remote ratio vs salary



```
plt.figure(figsize=(8, 5))
sns.barplot(data=remote_salary, x='remote_ratio', y='salary_in_usd')
plt.title('Average Salary by Remote Ratio')
plt.show()
```

- Groups data by remote ratio and calculates average salary.
- Plots a bar chart to show salary differences for on-site, hybrid, and remote roles
- Fully on-site (0%) and fully remote (100%) roles have higher salaries than hybrid (50%).
- Companies may offer more to attract fully remote or on-site workers.

# Geographical Impact on Salary

Gbn b



Insight :

- Location plays a crucial role in determining salary, countries with strong tech industry's, higher demand, and well-developed economies offer significantly higher pay.
- Companies in lower-paying regions and countries, may offer additional incentives (such as remote work options) to attract top talent.

# Geographical Impact on Salary ; Code Snippet

## ▼ Analyzing geographical impact on salary

```
[ ] top_locations = salaries['company_location'].value_counts().nlargest(10).index  
geo_salary = salaries[salaries['company_location'].isin(top_locations)]  
geo_salary_avg = geo_salary.groupby('company_location')['salary_in_usd'].mean().reset_index()
```

## ▼ Plotting company location vs salary

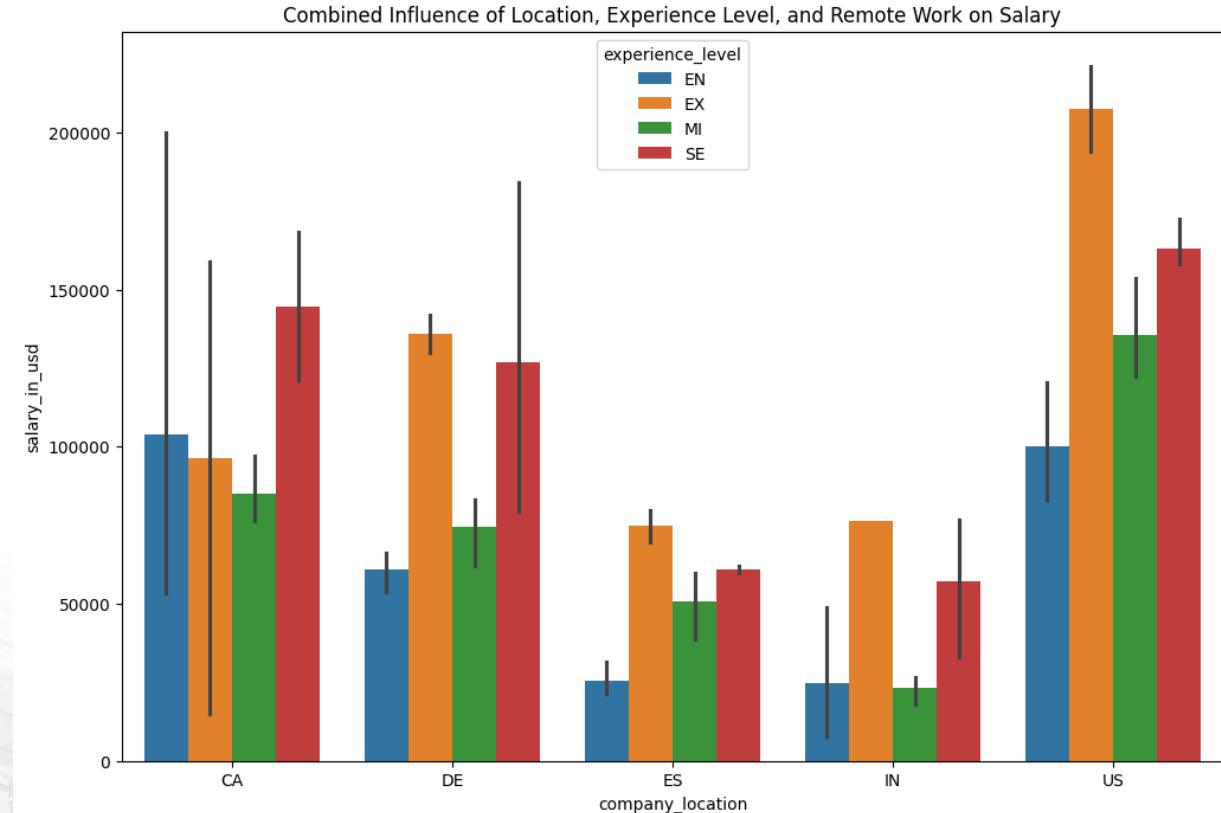
```
[ ] plt.figure(figsize=(12, 6))  
sns.barplot(data=geo_salary_avg, x='company_location', y='salary_in_usd')  
plt.xticks(rotation=45)  
plt.title('Average Salary by Top 10 Company Locations')  
plt.show()
```

### Insight :

- Location plays a crucial role in determining salary, countries with strong tech industry's, higher demand, and well-developed economies offer significantly higher pay.
- Companies in lower-paying regions and countries, may offer additional incentives (such as remote work options) to attract top talent.

## Combined Influence of Location, Experience Level, and Remote Work on Salary

- This analysis highlights that the collective influence of experience level and location on salaries is significant.
- Executive roles consistently offer the highest salaries across all locations, especially in countries like the US and Canada.
- This suggests that data science professionals seeking higher compensation should aim for senior or executive roles in high-paying regions like the US.



Overall, this combined chart shows that regardless of the location, experience level plays a crucial role in determining salaries.

The salary gap between entry-level and executive positions is most pronounced in the US, indicating a greater emphasis on experience and expertise.



Top 5 locations: CA (Canada), DE (Germany), ES (Spain), IN (India), and US (United States).



The y-axis shows average salary in USD, with significant variation across locations.



Bars are color-coded by experience level—EN (Entry), MI (Mid), SE (Senior), and EX (Executive). Executive roles in the US and Canada earn the most, while India follows a similar salary hierarchy but at lower rates.

# Combined Influence of Location, Experience Level, and Remote Work on Salary

## Code Snippet

- Combined Analysis: Experience Level, Remote Work, and Location

- Combined analysis: Average salary by experience level and remote ratio for top 5 locations

```
[ ] combined_salaries = salaries.groupby(['company_location', 'experience_level', 'remote_ratio'])['salary_in_usd'].mean().reset_index()
```

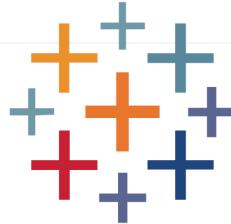
- Filter for top 5 locations for a cleaner visualization

```
[ ] top_combined_locations = combined_salaries['company_location'].value_counts().nlargest(5).index  
combined_salaries = combined_salaries[combined_salaries['company_location'].isin(top_combined_locations)]
```

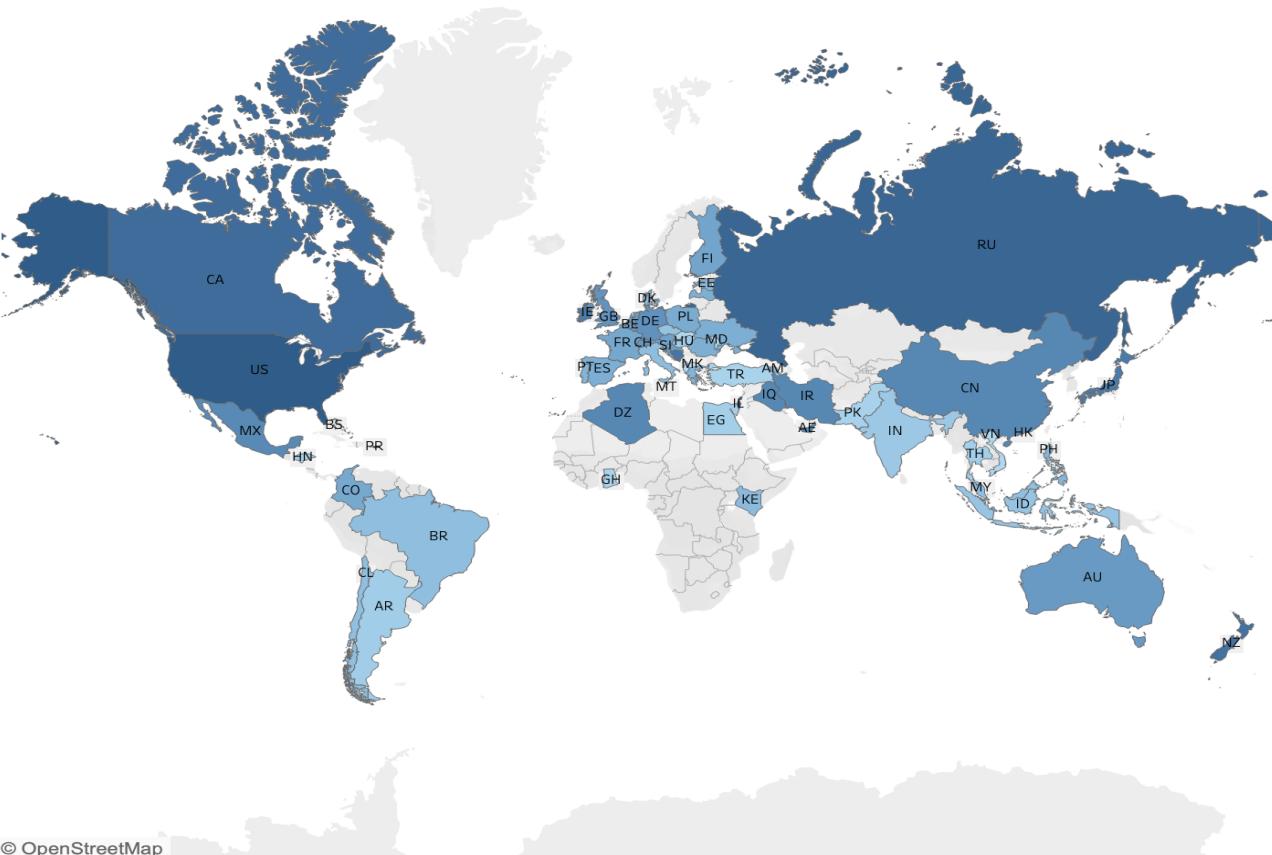
```
[ ] plt.figure(figsize=(12, 8))  
sns.barplot(data=combined_salaries, x='company_location', y='salary_in_usd', hue='experience_level')  
plt.title('Combined Influence of Location, Experience Level, and Remote Work on Salary')  
plt.show()
```

## Choropleth Map - Salary by Location

# Tableau



# Choropleth Map Salary by Location



Map based on Longitude (generated) and Latitude (generated). Color shows average of Salary In Usd. The marks are labeled by Company Location. Details are shown for Company Location. The data is filtered on Experience Level and Remote Ratio. The Experience Level filter keeps EN, EX, MI and SE. The Remote Ratio filter ranges from 0 to 100.

## Insights

- Visualizes the average salary of data science professionals across regions.
- Darker shades = higher salaries.
- Lighter shades = lower average salaries.

## Key Observations

- The US, Canada, and Germany show the highest salaries.
- Brazil and India have lower salaries due to lower living costs.
- Significant variation across continents, with developed countries offering more competitive compensation.

# Conclusion

## Key Findings

### Remote Work and Salary:

- Fully remote and fully on-site roles tend to offer higher salaries than hybrid roles.
- Remote work flexibility is increasingly becoming a factor in salary competitiveness.

### Experience Level and Salary:

- Executive-level roles command the highest salaries, especially in developed countries.
- Entry-level roles in the US and other tech hubs have higher starting salaries than those in developing regions.

### Geographical Differences:

- The US, Canada, and Germany offer the most competitive salaries for data science professionals.
- Countries like India and Brazil show lower salary averages, reflecting local economic factors.

## Recommendations:

### For Companies:

Offer flexible remote options: This can help attract talent from lower-paying regions and make salaries more competitive globally.

### Invest in senior roles:

High pay for experienced professionals ensures leadership in innovation and productivity.

### For Professionals:

Consider remote roles: Remote work can offer competitive salaries, even in regions with lower average compensation.

### Seek experience in tech hubs:

Gaining experience in high-salary regions can open opportunities for higher compensation later.

# THANK YOU

