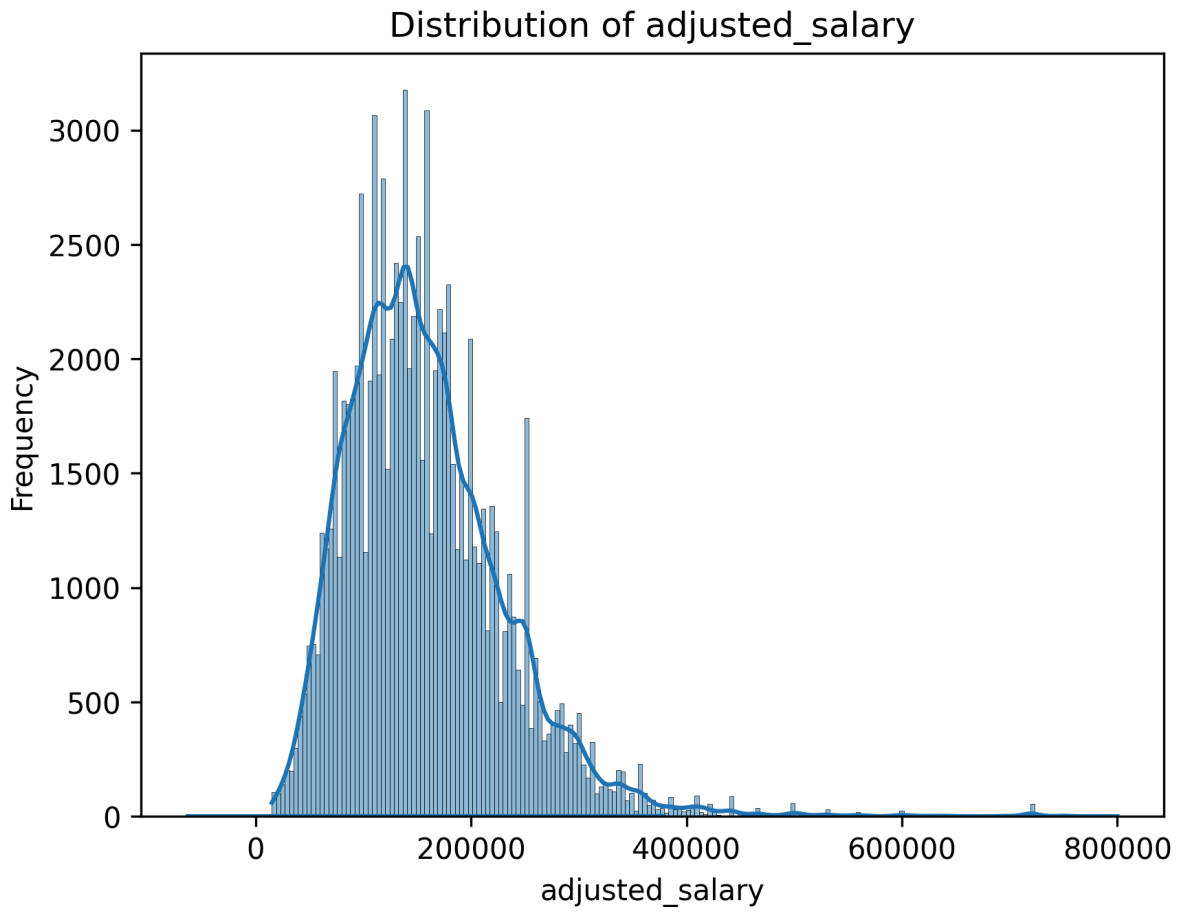
Summary of EDA Findings

1. Distribution and Central Tendency of Adjusted Salaries



The adjusted salary variable shows a strongly **right skewed** distribution. Salary values are not evenly distributed around the mean, with a clear concentration at lower values and a smaller number of very high salaries.

Summary statistics indicate the following:

- The mean adjusted salary is approximately 157,745 USD.
- The standard deviation is approximately 73,659 USD, indicating a high level of variability across salaries.

Most observations fall within a relatively concentrated range:

- The majority of salaries are between 100,000 and 180,000 USD, which represents the highest density region of the distribution.

Beyond this central range:

- Salary values extend significantly higher, with extreme observations reaching approximately 700,000 to 800,000 USD.
- These values form a long upper tail in the distribution.

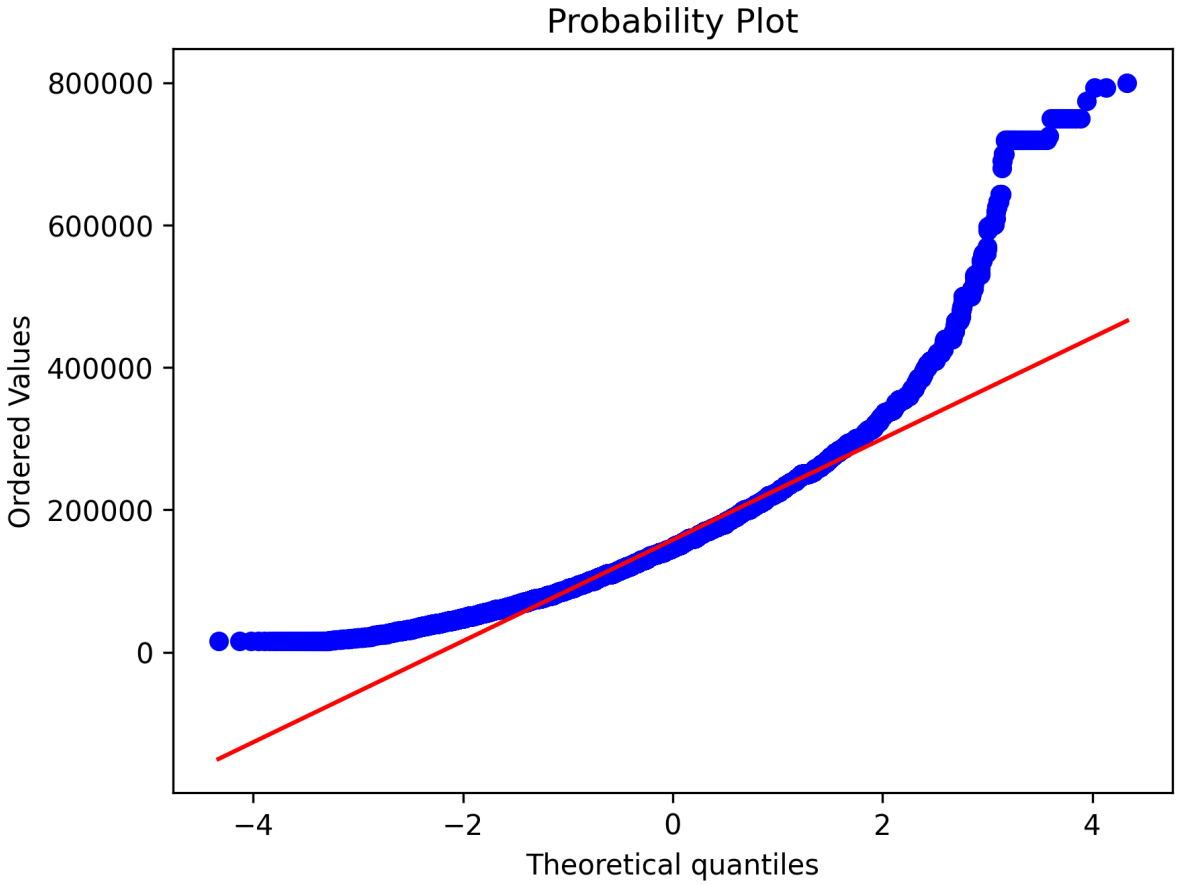
Measures of distribution shape further support these observations:

- A skewness value of approximately 1.34 indicates a pronounced right skew.
- A kurtosis value of approximately 4.75 indicates heavier tails than those expected under a normal distribution.

These characteristics show that high salary values are consistently present in the data rather than occurring as isolated anomalies.

Because of this structure, the mean salary is influenced by extreme values and does not fully reflect a typical salary outcome. Median based measures and percentile ranges provide a more stable representation of central salary behavior.

2. Normality and Distribution Shape



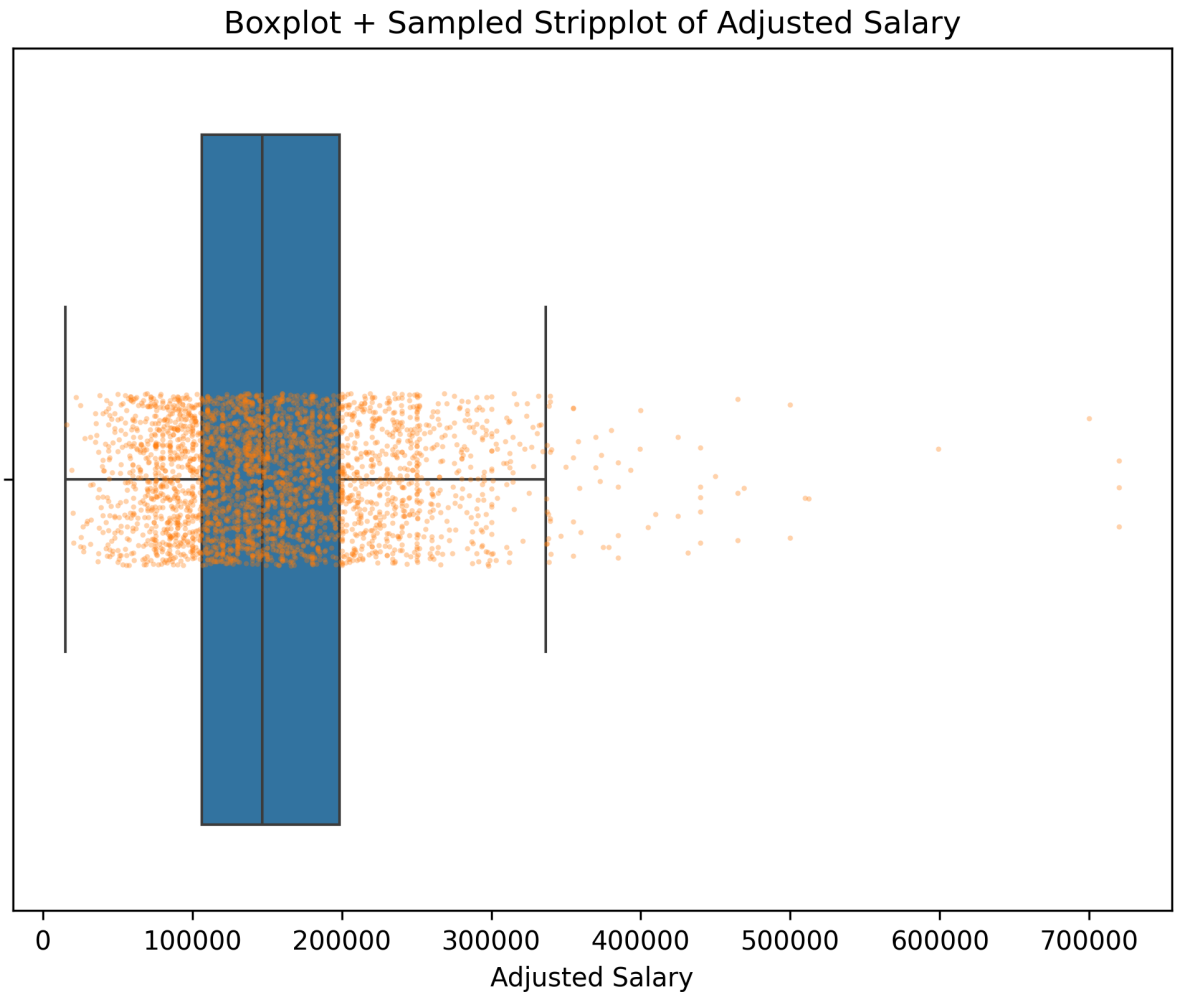
A normality assessment using a probability QQ plot shows clear deviations from the theoretical normal line.

Observations near the center of the distribution show moderate alignment with normal expectations, indicating that the middle portion of the data behaves reasonably close to a normal pattern.

At the lower end of the distribution, salary values compress above the theoretical line. At the upper end, salary values diverge sharply upward from the line.

This pattern indicates that the salary data does not satisfy normality assumptions, particularly in the upper quantiles. Analytical methods that rely on normality would therefore require data transformation or the use of robust statistical alternatives to reduce the influence of extreme values.

3. Variability and Presence of High Value Outliers



The boxplot and sampled stripplot provide additional insight into salary variability.

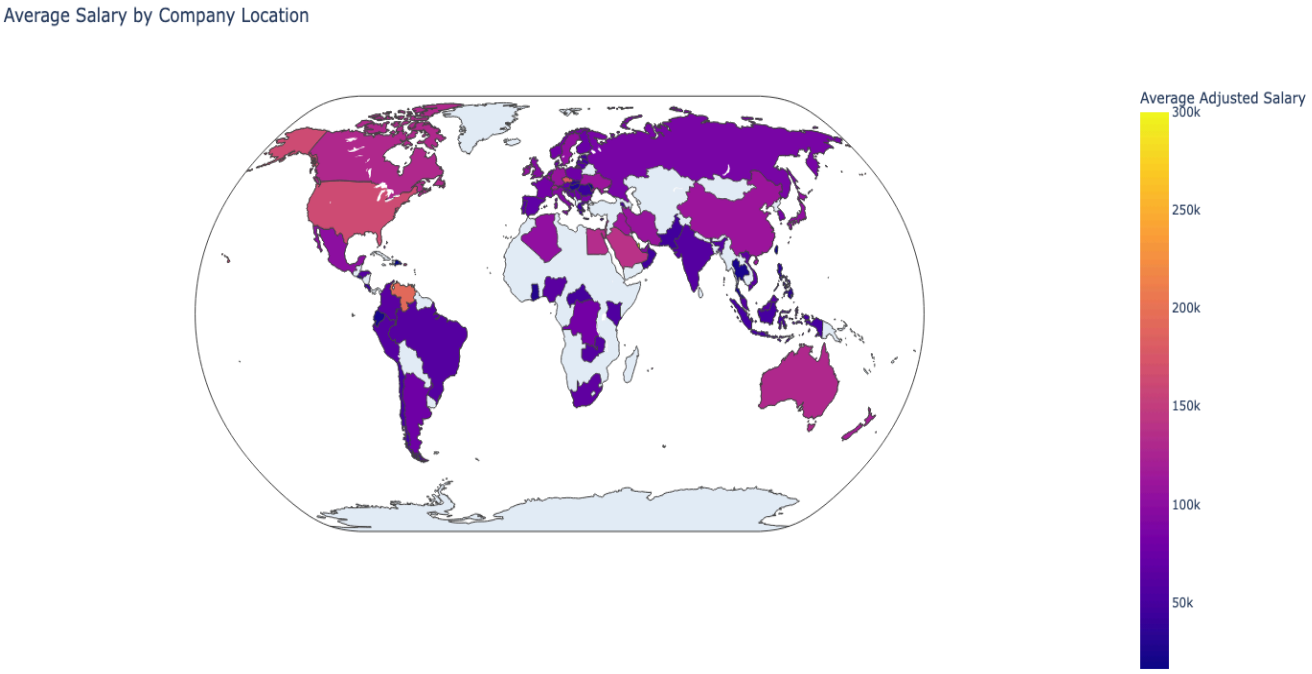
Key observations include:

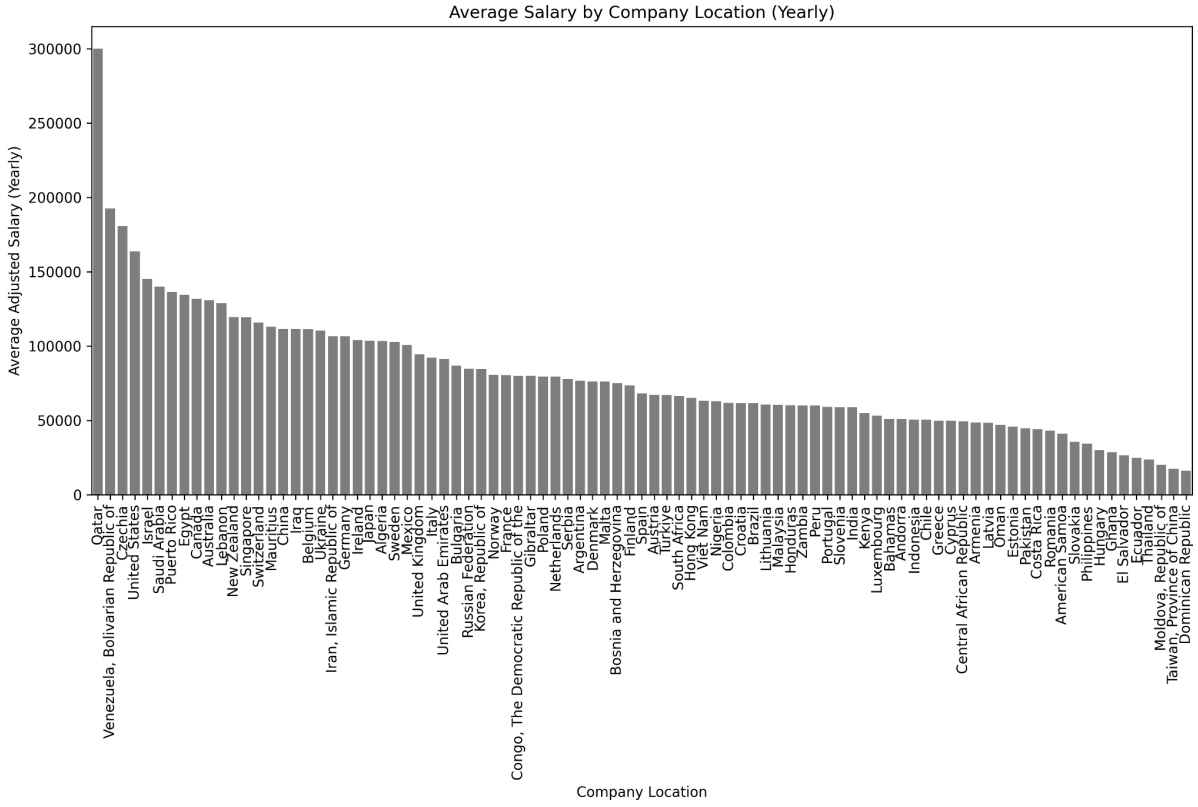
- The interquartile range captures a relatively compact band of salaries centered near the median.
- A large number of observations fall above the upper whisker, confirming the presence of many high salary outliers.

These outliers are not concentrated at a single value. Instead, they are spread across a wide range above approximately 300,000 USD, indicating the existence of multiple high compensation contexts rather than a single extreme anomaly.

This structure suggests that the salary distribution consists of a stable core of typical roles combined with a smaller but meaningful group of exceptionally high paying positions.

4. Salary Differences by Company Location





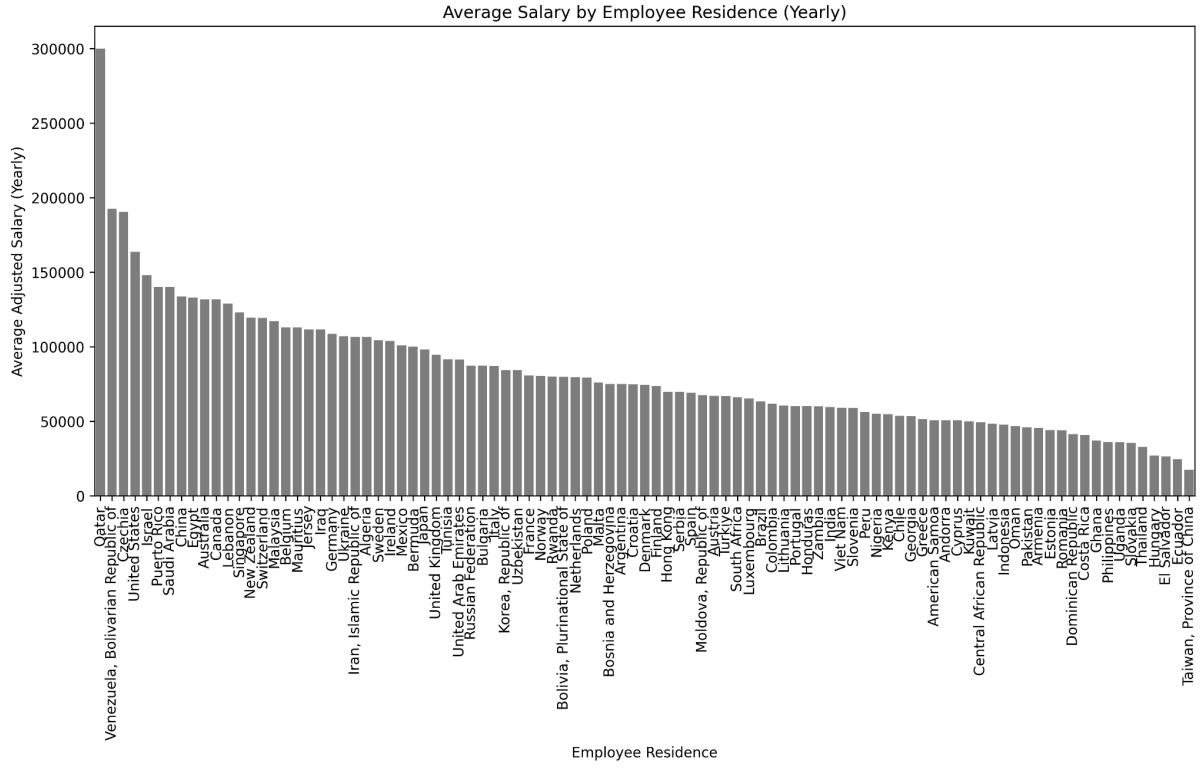
Average adjusted salaries vary substantially across company locations, indicating clear geographic stratification.

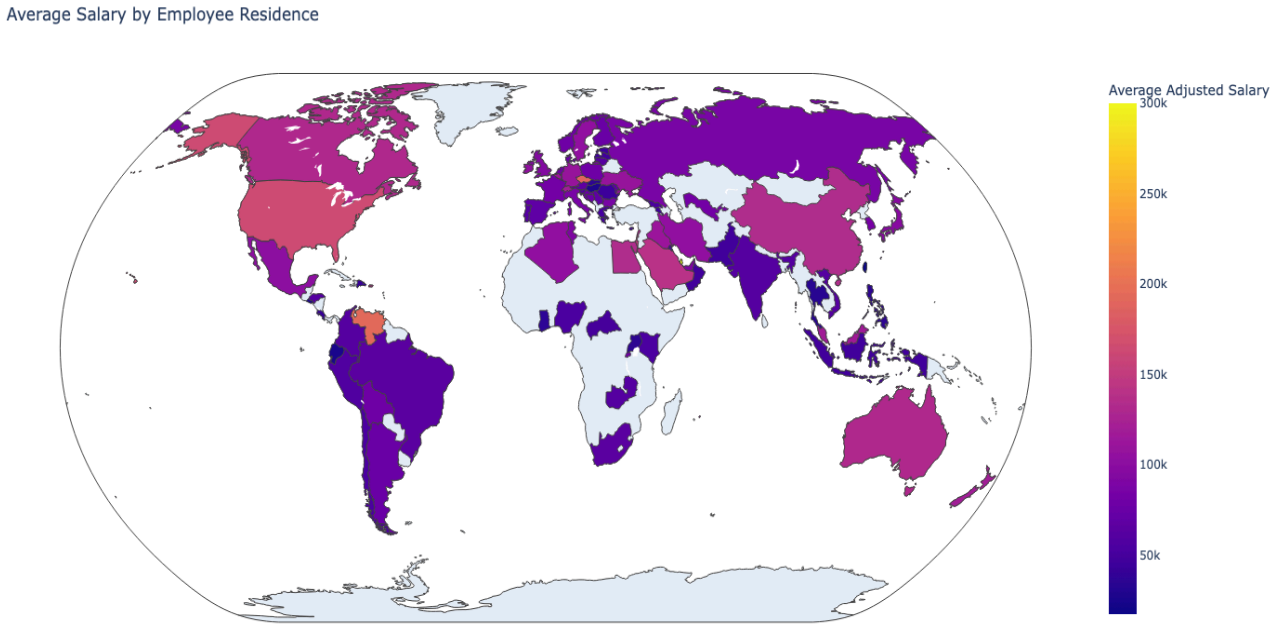
The observed patterns include:

- A small number of countries with average salaries approaching or exceeding 200,000 USD, placing them at the upper extreme of the distribution.
- A large group of countries with average salaries between 70,000 and 130,000 USD.
- Several countries at the lower end of the distribution with average salaries below 60,000 USD.

The overall distribution of country level averages is uneven and right skewed, reflecting the shape of the underlying salary distribution. Countries with fewer observations tend to exhibit greater volatility in average salary values, indicating that sample size influences the magnitude of observed extremes.

5. Salary Differences by Employee Residence





Grouping salaries by employee residence reveals patterns similar to those observed for company location, with important differences.

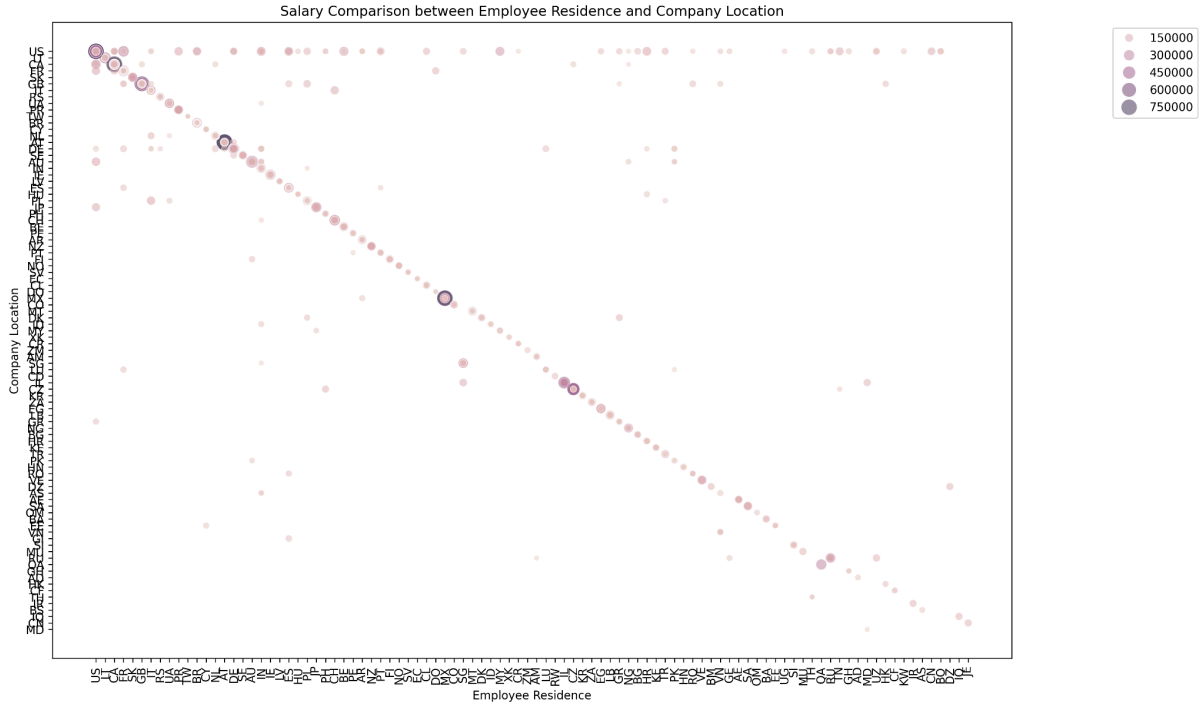
Key observations include:

- Average salaries by residence span a wide range, from below 50,000 USD to well above 200,000 USD.
- Higher average salaries are concentrated among a limited number of countries.
- Most countries fall within a broad middle band of average salary values.

Countries with lower average salaries are more numerous than those with high averages, contributing to the overall skewness observed in the salary distribution.

These patterns indicate that employee residence is strongly associated with compensation levels, although residence alone does not fully determine salary outcomes.

6. Relationship Between Employee Residence and Company Location



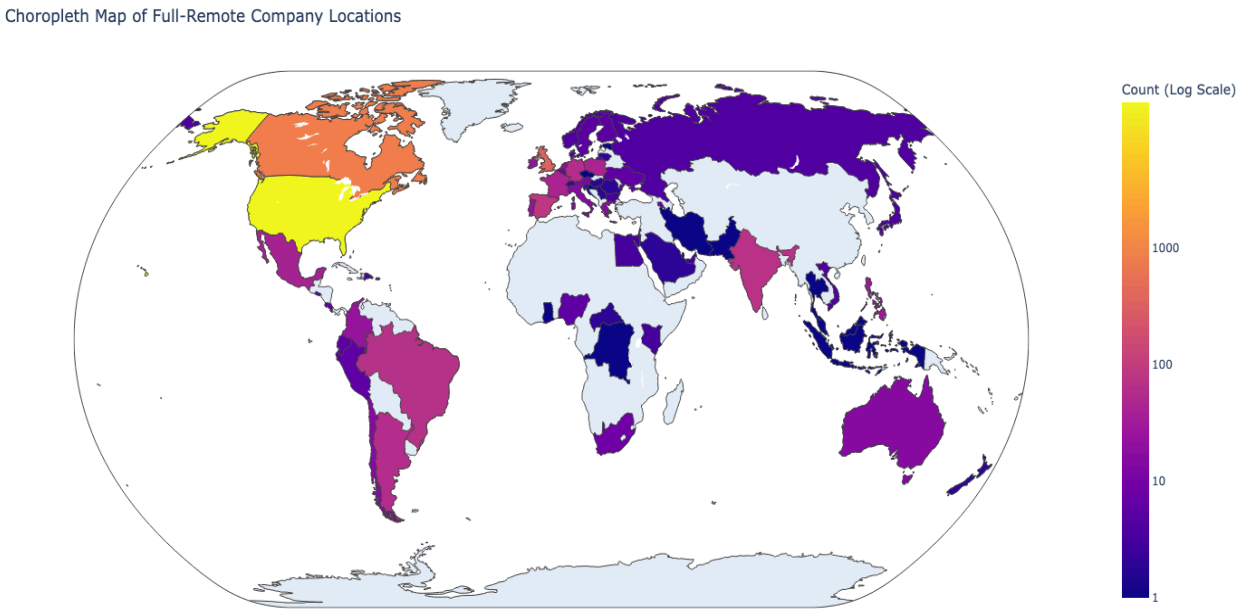
A direct comparison between employee residence and company location reveals two dominant patterns.

Diagonal concentration
Many observations lie along a diagonal pattern, indicating that employees often work for companies located in the same country as their residence.

Off diagonal dispersion
A substantial number of observations fall outside the diagonal pattern, representing cross border employment relationships.

Higher salaries appear both along the diagonal and among select off diagonal cases. This indicates that while domestic employment relationships dominate numerically, cross border arrangements can be associated with higher compensation, though this is not universally the case.

7. Full Remote Company Location Distribution



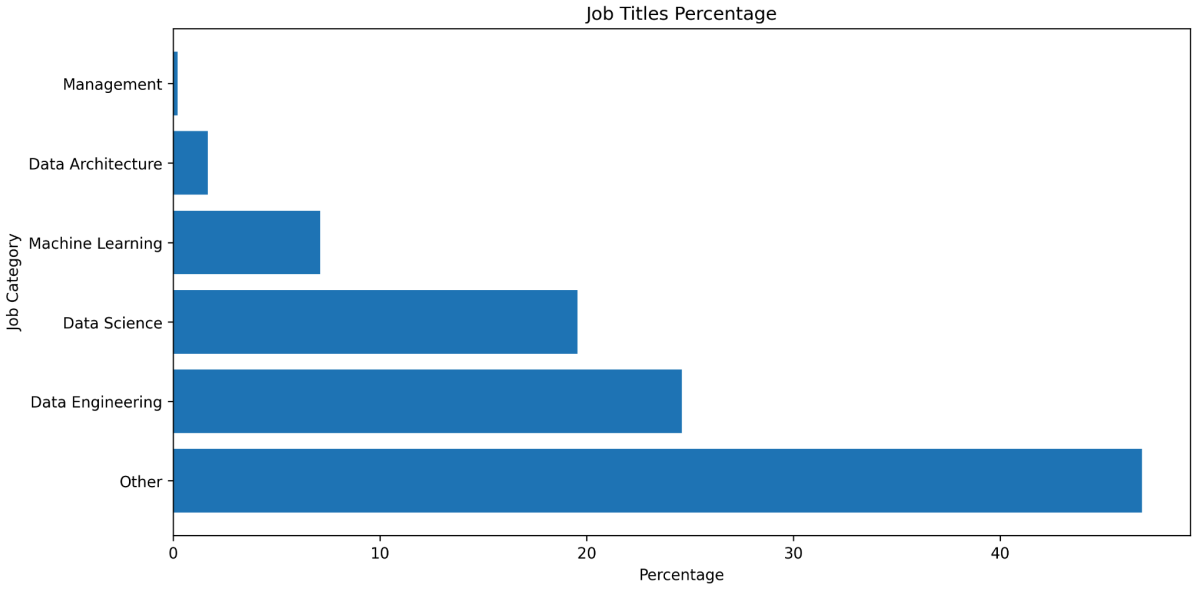
The choropleth visualization of full remote company locations shows a highly concentrated geographic distribution.

Key observations include:

- A small number of countries account for the majority of full remote roles.
- Most countries appear with very low counts, often only a small number of observations.
- The log transformed color scale highlights concentration rather than uniform global representation.

This uneven distribution indicates that analyses involving remote roles are disproportionately influenced by a limited set of regions and should be interpreted with this concentration in mind.

8. Job Category Composition



The distribution of job categories reveals a clear imbalance across role types.

JOB CATEGORY	APPROXIMATE SHARE
Other	Approximately 45 to 47 percent
Data Engineering	Approximately 25 percent
Data Science	Approximately 19 to 20 percent
Machine Learning	Approximately 7 to 8 percent
Data Architecture	Approximately 2 to 3 percent
Management	Less than 1 percent

The dominance of the Other category suggests that many roles do not map cleanly to narrowly defined job titles. This aggregation likely obscures meaningful variation in both responsibilities and compensation within that category.

This analysis of data science job salaries demonstrates that compensation levels vary substantially based on factors such as company location and employee residence. Clear differences emerge across geographic regions, and these differences contribute meaningfully to overall salary dispersion observed in the data.

By examining salary distributions, geographic patterns, and role composition, the analysis provides a structured understanding of how location and employment context influence compensation outcomes. These insights support more informed evaluation of salary expectations, compensation planning, and workforce distribution decisions.