Data Science Salary Analysis Report

Project Overview

This report presents an exploratory data analysis of a dataset containing employee salary information.

Objectives:

- Understand the overall salary distribution.
- Evaluate how salary varies by experience level, company size, and remote work ratio.
- Identify correlations among variables.
- Build a simple regression model to illustrate predictive relationships.

Dataset Summary

- **Observations:** 607
- **Features:** Work Year, Experience Level, Employment Type, Job Title, Salary, Company Size, Remote Ratio, Employee Residence
- **Time Period:** 2020–2024
- **Currency:** Salaries reported in USD

Data Cleaning Steps

- Verified column names and data types.
- Checked for missing values and found **no missing salary values.**
- **No duplicate records detected.**
- Encoded categorical experience levels for modeling.

Descriptive Statistics

Statistic	Salary (USD)
Mean	119,705.65
Median	112,500.00

Statistic	Salary (USD)
Minimum	28,500.00
Maximum	600,000.00
Standard Deviation	45,214.44

Key Visualizations

Salary Distribution

- Most salaries fall between **85,000** and **150,000 USD**.
- The distribution shows **right skew** due to high-end salaries.

Salary by Experience Level

- Clear trend of increasing salaries with experience.
- **Senior** and **Executive** roles earn significantly more.

Salary by Company Size

- Larger companies offer higher median salaries.

Salary by Remote Ratio

- Fully remote roles tend to pay more on average.

Correlation Matrix

- **Experience Level** shows strong positive correlation with salary.
- **Company Size** moderately correlates with higher salaries.
- **Remote Ratio** has a mild positive relationship with salary.

Regression Analysis

A simple linear regression was conducted to predict salary from experience level (encoded numerically):

- **Model:**
  Salary = $\beta_0 + \beta_1$ * (Experience Level)

- **Interpretation:**
  As experience level increases, predicted salary increases linearly.

Key Insights

- **Experience** is the most significant driver of salary.
- **Company size** and **remote work** also impact compensation.
- The salary distribution is **right-skewed**, with high outliers.

Recommendations

- Consider further modeling with multiple predictors (company size, remote ratio).
- Validate extreme salary values to ensure data integrity.
- Investigate potential factors behind salary disparities across roles and regions.