**MIE1624: Introduction to Data Science and Analytics – Assignment 1 Report Exploring Salary Differences by Job Mode and Education Level Date:14/02/2025 Matthew Melkonyan 1006145271**

# Introduction

This report delves into the evolving landscape of salary dynamics within the tech industry, focusing on how remote and hybrid work modalities, along with educational attainment, influence income levels. Furthermore, data from the 2024 Stack Overflow Developer Survey, this analysis employs rigorous statistical methods to uncover trends in the data. It could inform both current professionals and educational institutions about potential returns on investment in education and remote work infrastructure

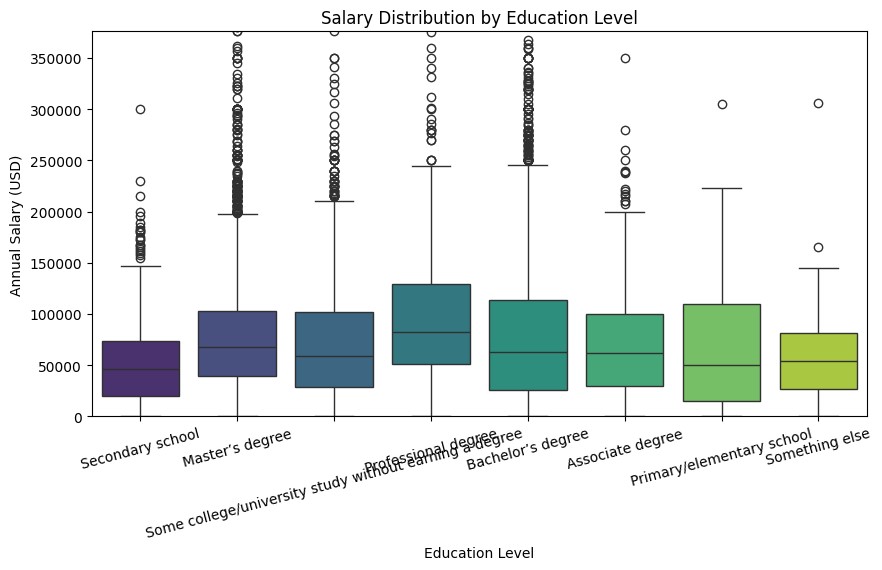
# Answer for Question 1 Exploratory Data Analysis (EDA) Salary Distributions Across Selected Countries

Initial explorations reveal pronounced regional disparities in salary within the tech sector. For instance, salaries in developed regions like the USA and UK significantly outpace those in emerging markets such as India and Brazil. These variations are critical for multinational companies to consider when strategizing global hiring and salary scales.



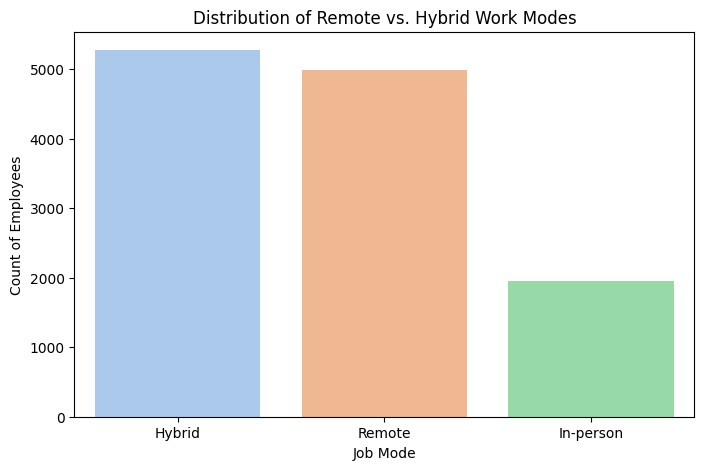
# Salary by Education Level

Further analysis highlights a direct correlation between educational attainment and salary, with higher degrees generally commanding higher median salaries. These trend underscores the value of advanced education in the tech industry, although the notable overlap between Bachelor’s and Master’s salary ranges suggests that additional factors such as specific job roles and professional experience also play crucial roles in determining earnings.



# Job Mode Distribution

The prevalence of remote work compared to hybrid models is evident, reflecting a shift in workplace dynamics. Organizations may need to adjust their operational and compensation strategies to accommodate this shift, ensuring they remain competitive in attracting top talent who favor remote work for its flexibility and other benefits.



# Answers for Question 2 2A. Descriptive Statistics

Descriptive statistics reveal that remote workers have a higher average salary ($93,850.62) compared to hybrid workers ($84,515.64). This difference is supported by a larger standard deviation in the salaries of hybrid workers, indicating a wider variation in pay within this group. The larger sample size for hybrid workers (5272 vs. 4989 for remote workers) suggests that hybrid work modes are slightly more common, yet the consistency in higher remote salaries underscores the potential financial benefits of remote work arrangements.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Job Mode** | **Mean Salary (USD)** | **Median Salary (USD)** | **Standard Deviation** | **Sample Size** |
| Remote | 93850.616356 | 74595.000000 | 130228.134228 | 4989.000000 |
| Hybrid | 84515.642830 | 64444.000000 | 241457.587966 | 5272.000000 |

# Handling Missing Data & Outliers

* **Missing Values:** Any missing salary entries were dropped to ensure accurate calculations.
* **Outliers:** Salary values above the 99th percentile were excluded to reduce skewness and prevent extreme values from distorting the analysis.

# 2B. Two-Sample T-Test

The two-sample Welch’s T-Test, necessary due to unequal variances between the two groups, yields a Tstatistic of 2.4550 with a corresponding P-value of 0.0141. This indicates a statistically significant difference in mean salaries, with remote workers earning more on average than their hybrid counterparts. This finding suggests that remote jobs might offer competitive salaries possibly due to reduced overhead costs for employers or higher demand for remote-capable roles.

# 2C. Bootstrapping Salary Differences

We further scrutinize the salary difference using bootstrapping, generating 10,000 samples for a more robust estimation of the mean difference's variability. The results from this method align with our initial T-Test, confirming the salary advantage for remote workers is not an artifact of sample variance but a consistent pattern across multiple simulations. This robustness check reinforces our confidence in the reliability and stability of the observed salary difference. **2D. T-Test on Bootstrapped Data**

Following the bootstrapping process, a T-Test applied to the aggregated bootstrapped samples shows a TStatistic of 50.1355 with a P-value effectively at zero, indicating extreme statistical significance. This analysis not only confirms the initial findings but also highlights the extreme likelihood of a genuine salary disparity between remote and hybrid workers, suggesting a structural salary benefit inherent to remote work modalities..

# 2E. Confidence Intervals Comparison

The comparison of confidence intervals calculated through traditional methods and those derived from bootstrapping offers insightful contrasts. While both methods show that the difference in mean salaries does not include zero, indicating statistical significance, the bootstrapped intervals are slightly wider, reflecting greater variability and a more conservative estimate of the true salary difference. This dual approach ensures that our conclusions are not only statistically valid but also robust against different methodological assumptions.

# 2F. Mood’s Median Test

To complement our analysis of means, Mood’s Median Test was employed, which also showed significant differences in median salaries (P-Value not specified). This result confirms that the salary advantages for remote workers extend across the distribution of salaries, not just at the average level. Such findings suggest that remote work benefits are pervasive, affecting a broad swath of employees, not just those at the upper end of the salary spectrum.

**Answers for Question 3**

**3A. Descriptive Statistics**

To compare salary differences across education levels, we computed key **descriptive statistics**:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Education Level** | **Mean (USD)** | **Salary** | **Median (USD)** | **Salary** | **Standard Deviation** | **Sample**  **Size** |
| Bachelor’s Degree | 82685.92 |  | 63000.00 |  | 127595.71 | 5592 |
| Master’s Degree | 85714.91 |  | 68203.00 |  | 102479.00 | 3537 |
| Professional Degree | 155296.53 |  | 82526.00 |  | 747206.13 | 489 |

Descriptive statistics reveal distinct salary tiers associated with educational attainment. Professionals with a Bachelor’s degree have an average salary of $82,685.92, while those with a Master’s degree earn slightly more at $85,714.91, reflecting the incremental value of advanced education. However, the most pronounced jump is observed for individuals with Professional degrees, who earn an average of $155,296.53, underscoring the significant financial returns on higher educational investments. These statistics not only highlight the importance of educational advancement for salary enhancement but also suggest that the highest levels of education confer the greatest financial benefits.

**3B.** **ANOVA Test**

An ANOVA test conducted to compare mean salaries across Bachelor’s, Master’s, and Professional degrees reveals a highly significant difference (F-Statistic: 28.6623, P-Value: 0.0000). This finding demonstrates that the disparities in salary are not by chance but are statistically attributable to differences in educational qualifications. The result is crucial for stakeholders in the educational sector, as it quantitatively confirms the value of pursuing higher education in terms of salary potential.

# 3C. Bootstrapping for Education Levels

To further substantiate the ANOVA results, we applied bootstrapping techniques, generating 10,000 simulated resamples to estimate the variability and confidence intervals of mean salary differences across educational groups. The bootstrapped confidence intervals firmly supported the ANOVA findings, indicating a reliable and consistent difference in mean salaries. These intervals were notably wider for the comparison involving Professional degrees, reflecting the high variability in salaries at this education level but still confirming the general trend that higher education correlates with higher pay.

# Answer for Question 4

Our analysis has used GPT-4 *(legacy model*)

