

Dataset information

The dataset used in this report was downloaded from [Kaggle](#) and is intended for educational purposes only. It contains data on Age, Salary, Gender, and Experience. Based on this dataset, we formulated the following questions to guide our analysis.

Key Insights

- 📊 **Age Distribution:** Most survey respondents are between **24 and 34 years old**.
- 🎓 **Education Level:** Nearly **50% hold a Bachelor's degree**, while **less than 10%** have High School as their highest qualification.
- 📁 **Top 5 Job Titles by Count:**
 - 1 Software Engineer
 - 2 Data Scientist
 - 3 Software Engineering Director
 - 4 Data Analyst
 - 5 Senior Project Engineer
- 💰 **Top 5 Job Titles by Average Salary** (*with more than 20 entries*):
 - 1 Director of Data Science
 - 2 Marketing Director
 - 3 Software Engineering Manager
 - 4 Senior Project Engineer
 - 5 Data Scientist

Introduction

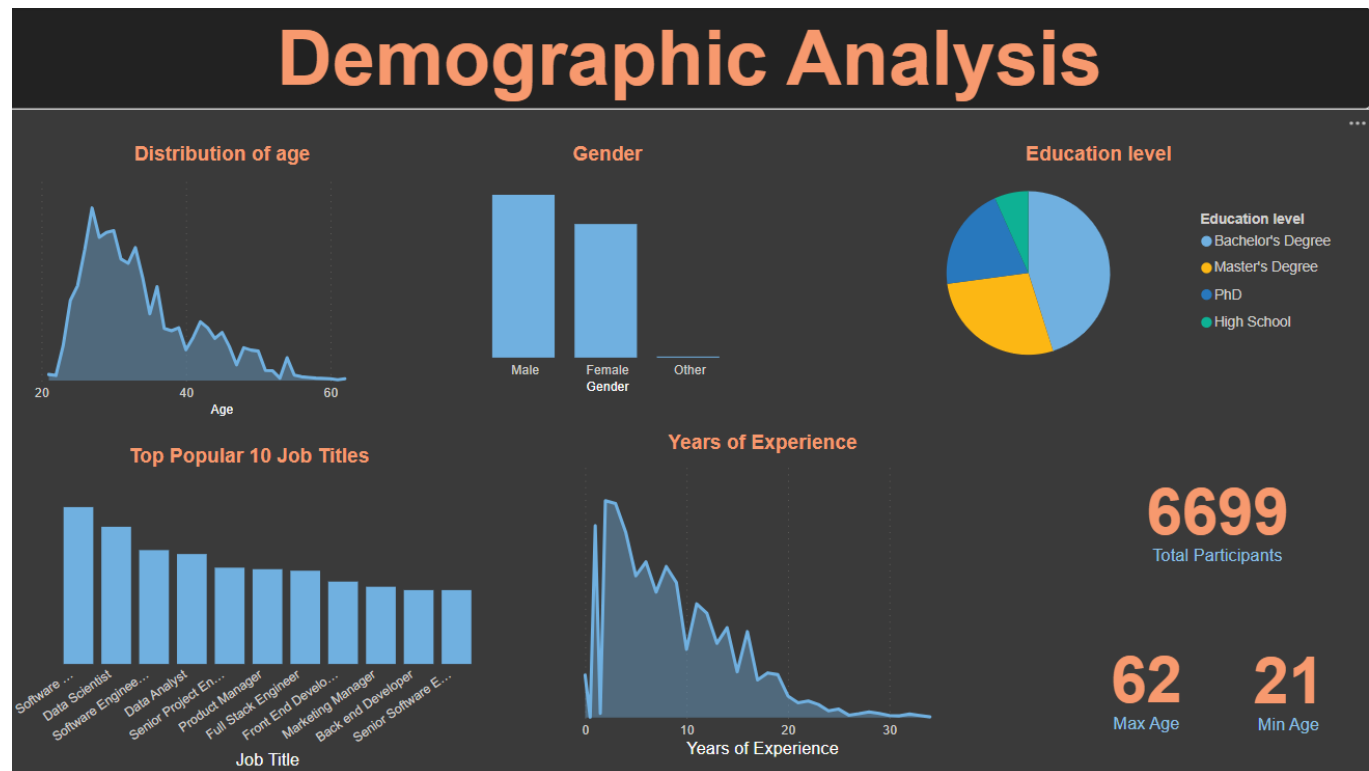
Using this dataset, we aim to conduct comprehensive data analysis to uncover actionable insights and trends. By leveraging the available features, we can address critical questions such as:

- Does a gender pay gap exist within the dataset?
- How significant is the impact of education level on earning potential?
- Can we build a predictive model to estimate salary based on key variables?

This analysis will enable us to derive meaningful conclusions and support data-driven decision-making.

Analysis

Demographics



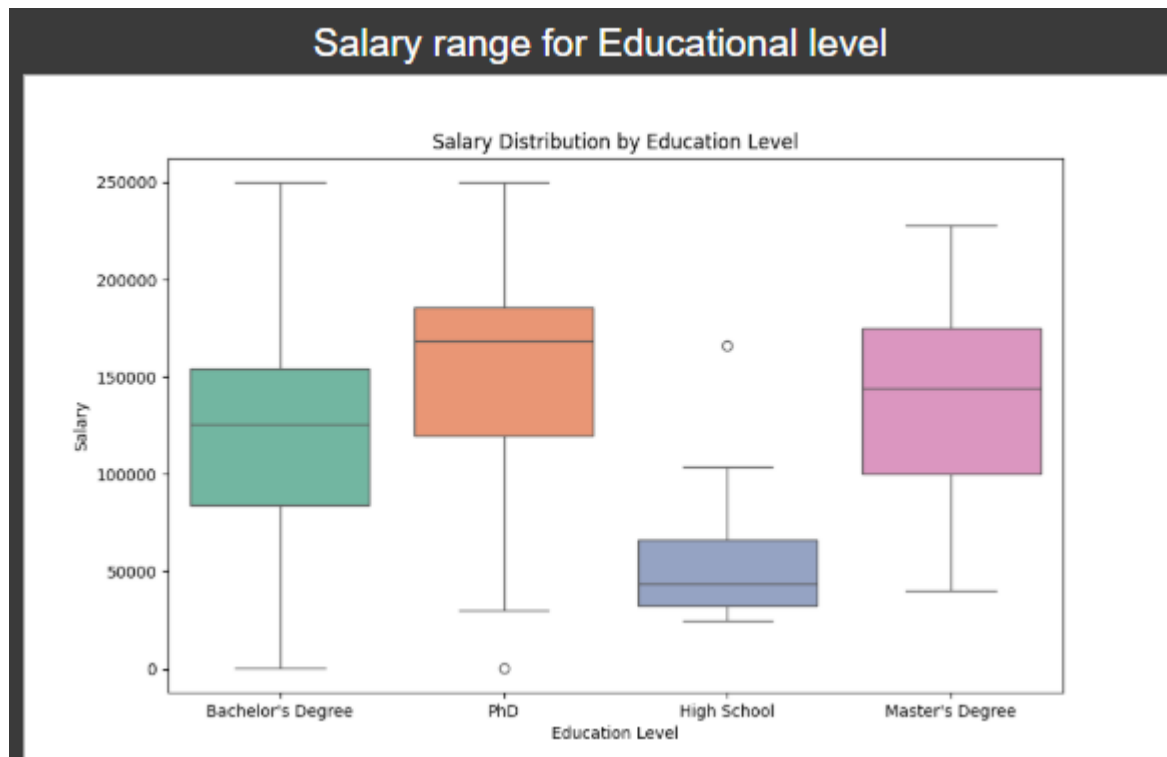
The survey included **6,699 participants**, with the majority holding tech-related job titles. However, it also captured responses from individuals in other common roles, such as accountants and operations professionals, though these groups were in the minority.

In terms of demographics:

- The age range of participants spans from a **minimum of 21 years** to a **maximum of 62 years**.
- The largest age group represented was **24–34 years old**.
- Gender distribution included **3,672 men**, **3,013 women**, and **14 individuals categorized as "other."**
- Educationally, **7% of participants did not hold a degree**, while the remaining **93% had completed at least an undergraduate degree**.

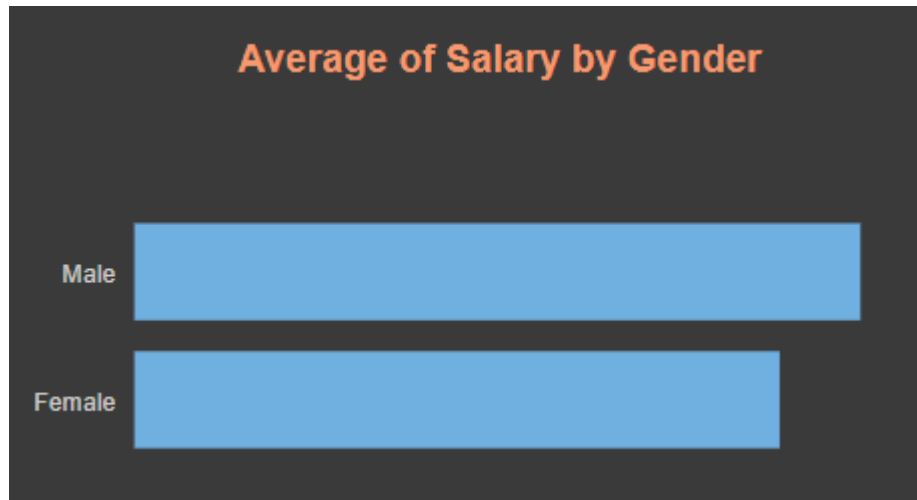
Salary

By Education

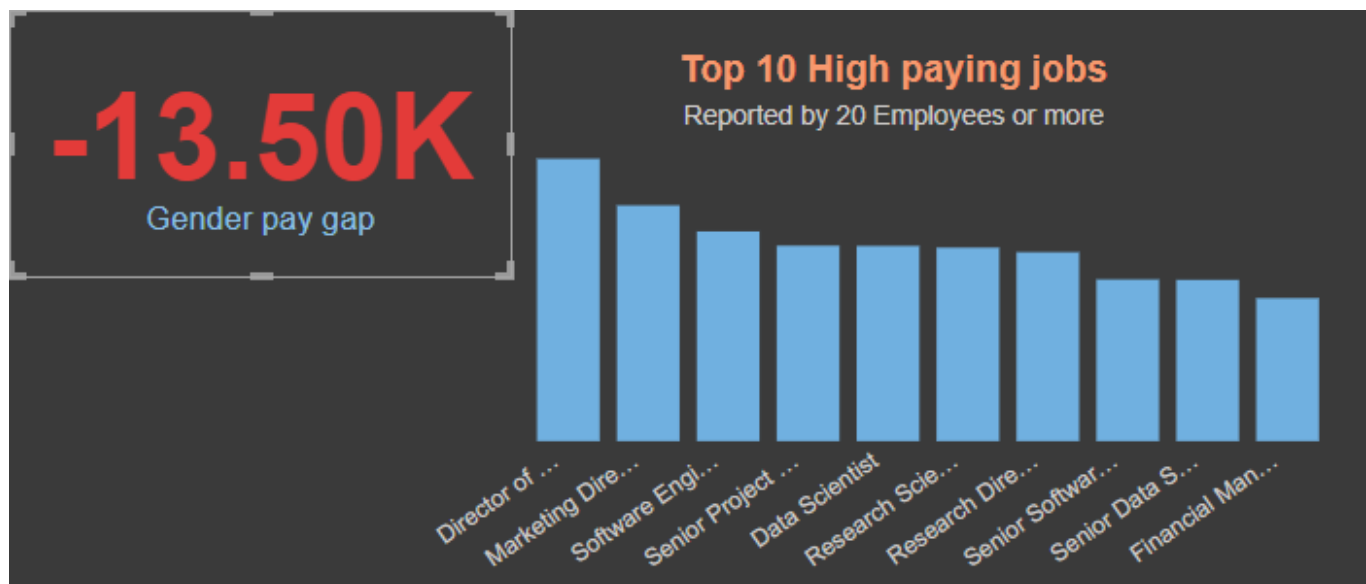


Across all job roles, individuals with at least a degree consistently earned significantly higher salaries compared to those with only a high school education. Notably, **Ph.D. holders** had the **highest median salary**, followed closely by those with a **master's degree**, highlighting the strong correlation between advanced education and earning potential.

By Gender



When comparing average salaries between men and women across all job titles, **women earn \$13,000 less on average**. This trend persists across most roles, with women consistently earning less than their male counterparts.

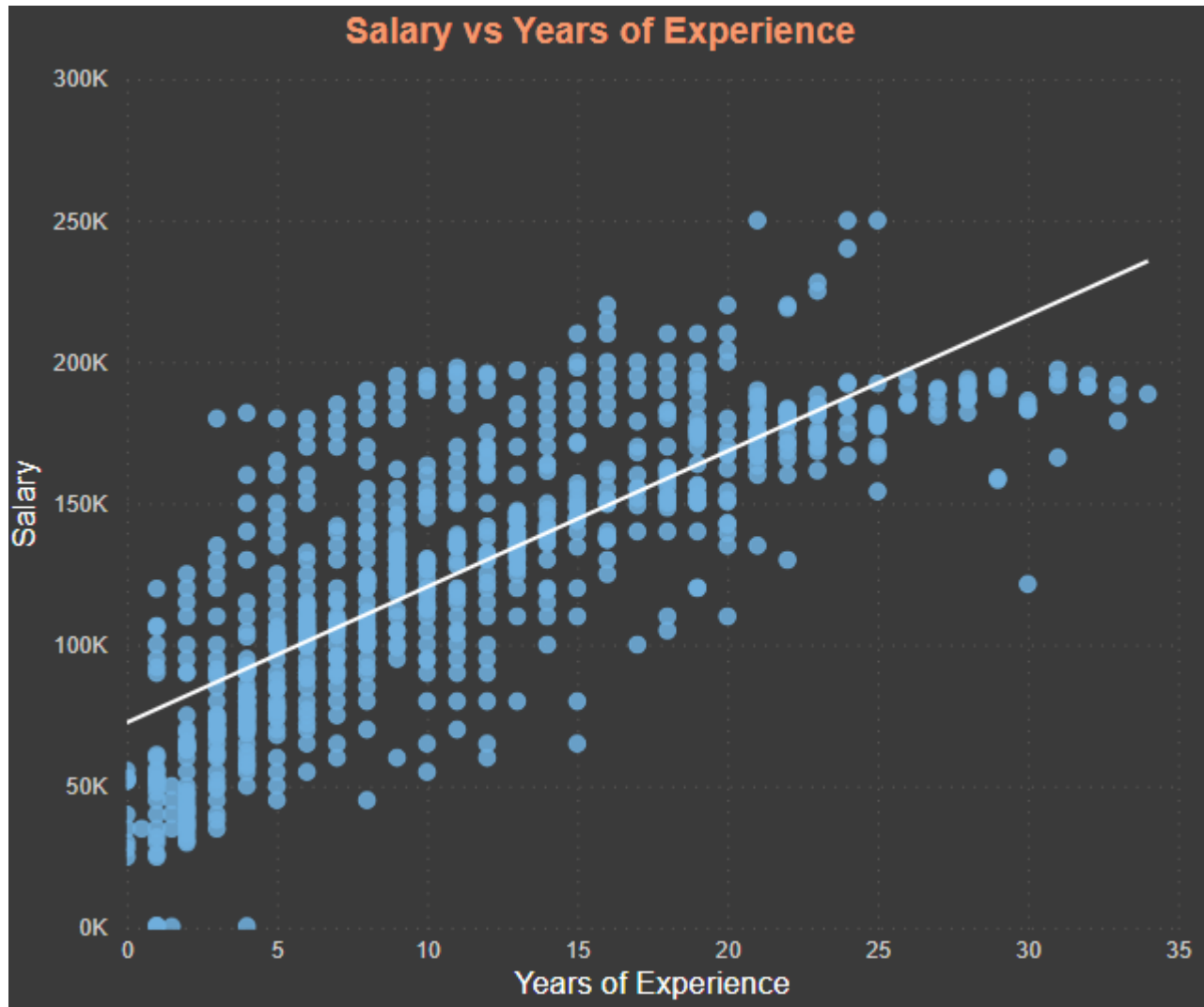


In the **top 10 highest-paying jobs**, the gender pay gap is particularly pronounced:

- **3 out of 10 jobs** show a salary difference of more than \$20,000 in favor of men.
- **3 out of 10 jobs** have women earning slightly more on average than men.

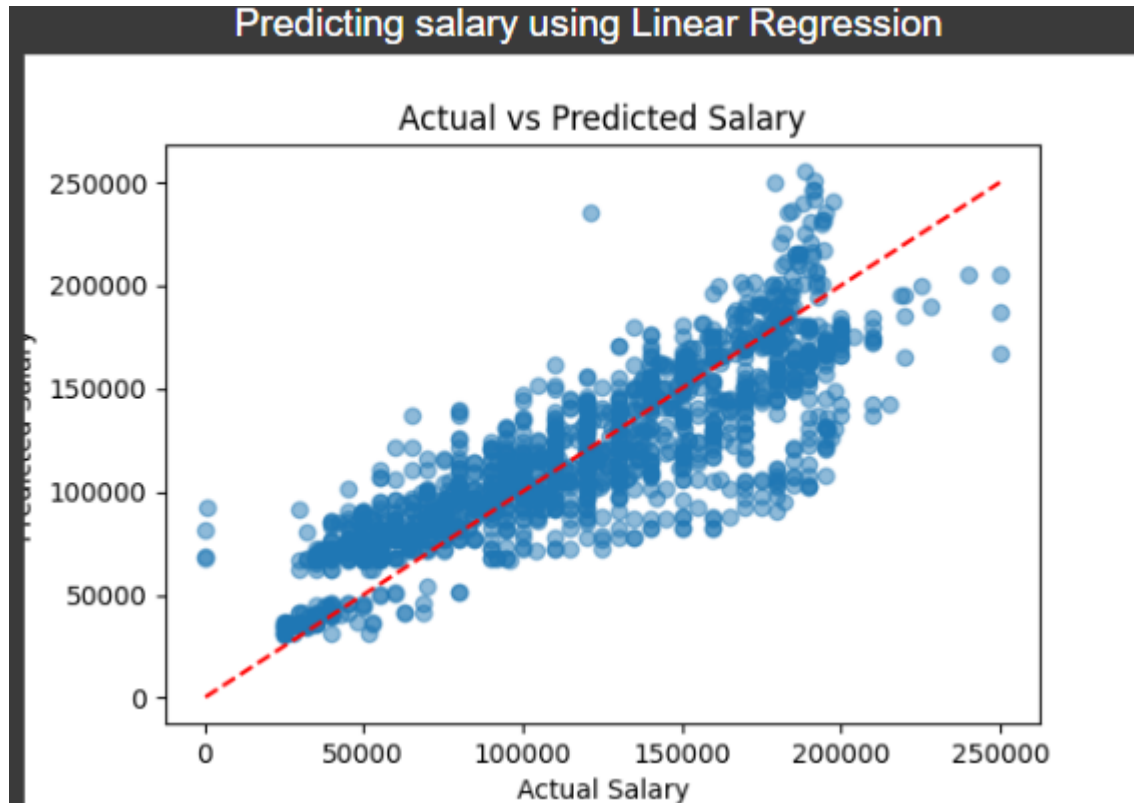
These findings underscore the existence of a **significant gender pay gap**, where women are likely to earn less even with the same education level and job title. Addressing this disparity remains a critical challenge in achieving workplace equity.

By Experience



The scatter plot displays a moderately strong positive correlation between years of experience and salary. The distribution of data points suggests a linear trend, where increased experience is generally associated with higher earnings. While some dispersion exists, the overall pattern indicates a clear upward movement from lower-left to upper-right, reinforcing the expectation that salaries increase with professional experience.

Prediction



Additionally, I conducted a **simple linear regression** on this dataset to evaluate whether the available features could effectively predict an individual's salary. The model achieved an **accuracy of 70%**, which is a strong result given the dataset's simplicity. While the accuracy could potentially improve by **eliminating outliers**, the current performance is still impressive and demonstrates that these features are sufficient for making reliable salary predictions. This suggests that even with basic data, we can build a practical and effective predictive model.

Conclusion

I began this project with the goal of uncovering insights and trends related to salary distribution across different genders and the factors influencing earnings. The dataset revealed that the majority of the workforce falls within the **24–30 age group**, and most participants have completed at least an undergraduate degree. Given the tech-dominated nature of the roles, salaries tend to be relatively high, as these positions often require advanced education and specialized skills.

Further analysis highlighted a persistent **gender pay gap**, with women consistently earning less than men, even when controlling for factors like education and job title.

Despite the dataset's simplicity, the machine learning model achieved **strong accuracy**, demonstrating its effectiveness in extracting meaningful patterns. This underscores the potential of such models to tackle both complex challenges and everyday tasks efficiently.

Related documents

Rough Notes: [Salary Analysis](#)

Performed Actions: [Salary Analysis - EDA](#)