Salary Base Dataset

Tak Sum Li

Objective

The objective of this dataset is to explore the relationship between education level, years of experience, and salary among employees in a company.

Dataset Metadata

- The number of entries: 6703
- Feature:
    - Age
        - Range: 21-62 years old
    - Gender
        - Range: Male, Female, Other
    - Education
        - Range: High School, Bachelor's degree, Master's degree, PhD
    - Job Title (Categorical)
    - Year of Experience
        - Range: 0-34 years
    - Annual salary:
        - Range: 350 - 250k US dollar

Summary

In data manipulation, the average salaries were identified based on education level. In addition, gathering information on years of experience and corresponding salaries.
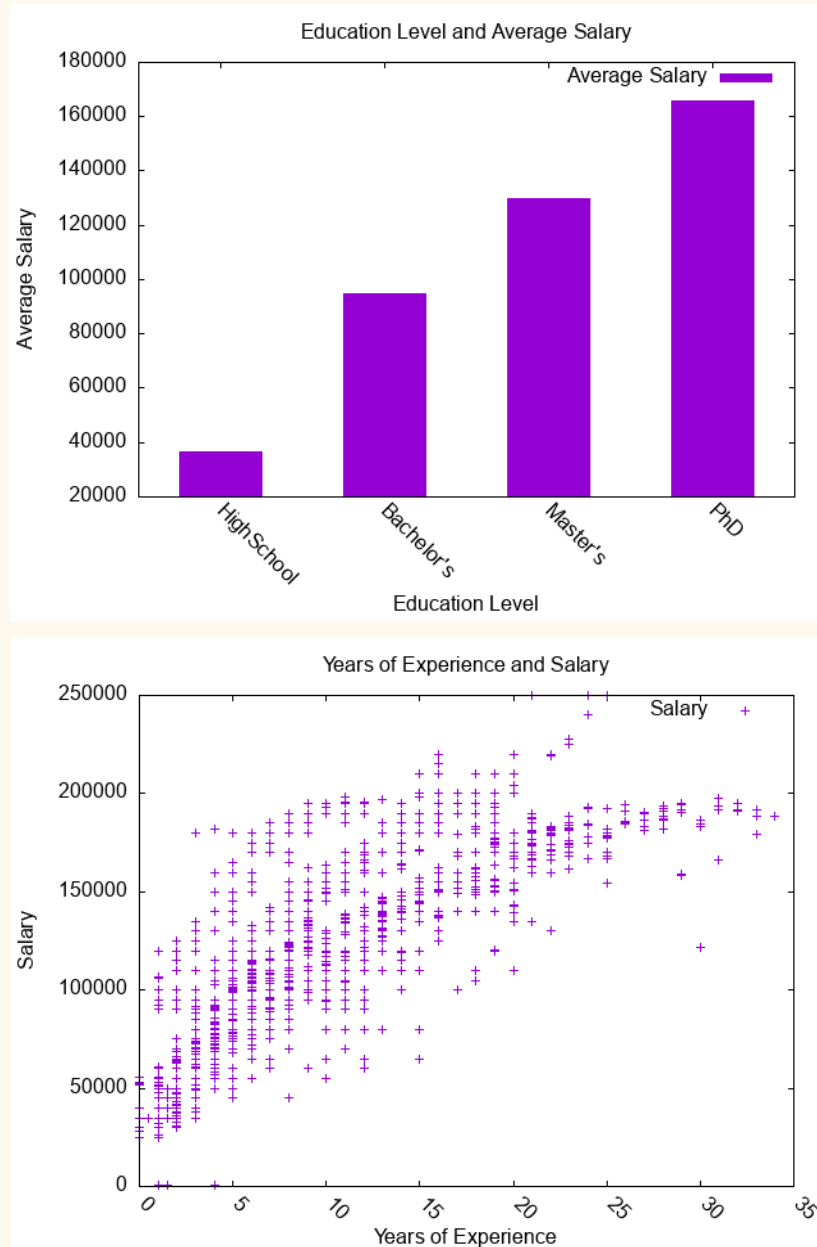
Methodology

1. Cleaning up the original data set using **sed** to improve the readability
2. Collecting lines with education, salary, experience using **awk** and **sort** then output to a csv file for plotting
3. Meantime, using an **awk** script to compute the the average salaries
4. Using **gnuplot** to plot the correlation between the average salaries and education level in histogram
5. Using **gnuplot** to plot the correlation between the years of experience and the salary in scatter plot

Findings

Analyzing the dataset and associated plots revealed a significant correlation between both education level and years of experience with salary.

Plots

Reference

https://www.kaggle.com/datasets/sinhasatwik/salary-base-data/data