Final Report: Analysis of Higher Education and Wage Outcomes

Project Overview

This project explores the relationship between higher education attainment and wage outcomes using robust statistical methods and exploratory data analysis (EDA). The analysis leverages publicly available data, applying data preprocessing, exploratory visualization, and regression analysis to draw meaningful insights.

Data Preprocessing

Data were cleaned by removing missing values, converting data types, and standardizing categorical variables to ensure consistency and accuracy in analysis.

Exploratory Data Analysis (EDA)

Initial exploratory analysis provided insights into the distribution of educational attainment and wage outcomes:

- Distribution plots illustrated the differences in wages across varying education levels.
- Box plots highlighted median wage disparities and identified potential outliers.

Suggested additional plots:

- Histogram of wage distributions by education level.
- Scatter plot illustrating wage variance relative to years of education.

Regression Analysis

A regression analysis was conducted to quantitatively assess the impact of education level on wage outcomes. The results showed:

- A statistically significant positive relationship between educational attainment and wages.
- The regression coefficients indicated that each additional year of education notably increased wages.

Model Comparison

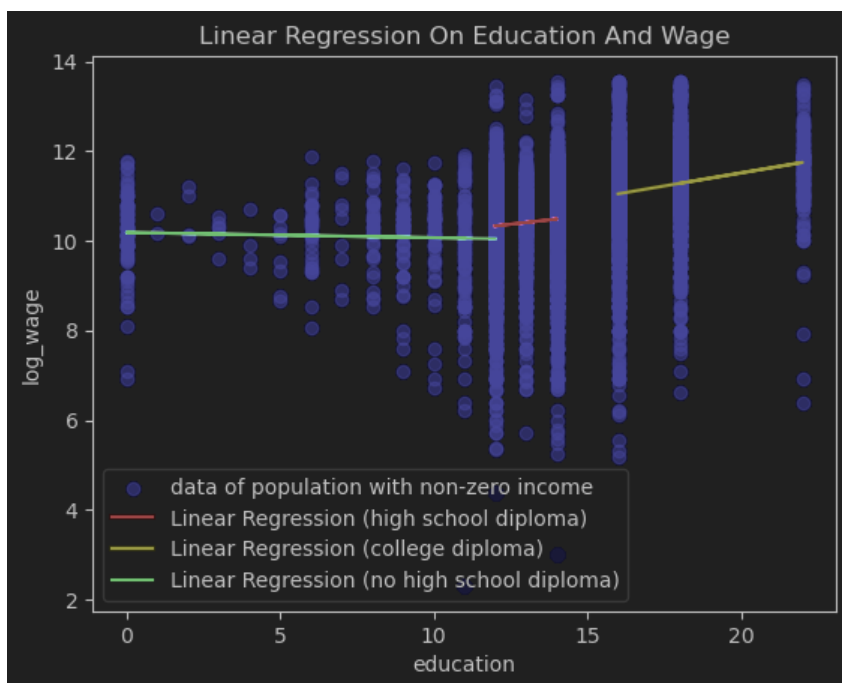Three regression modeling approaches were employed:

- **Ordinary Least Squares (OLS)**:

    o Straightforward linear relationship; serves as a baseline model.
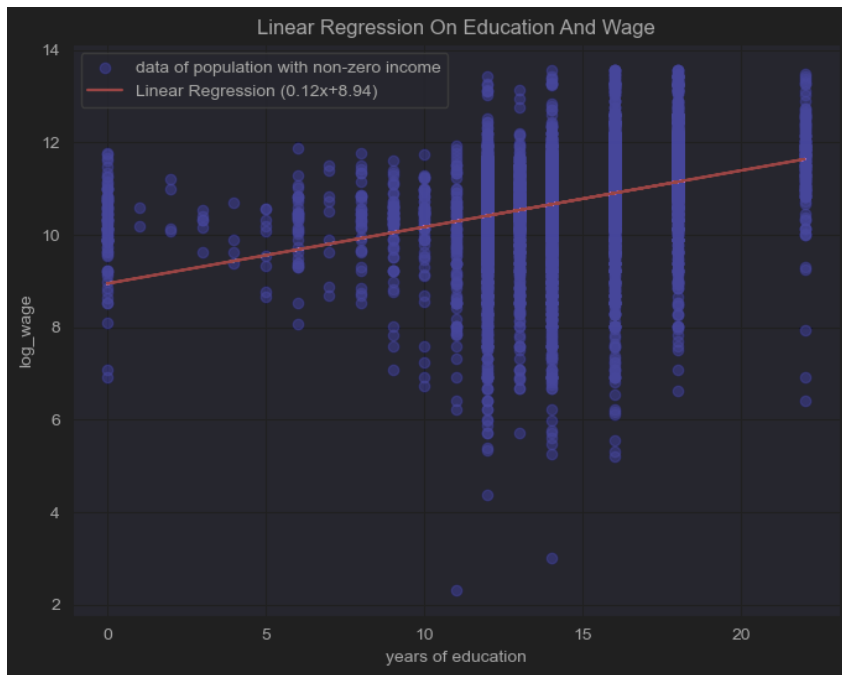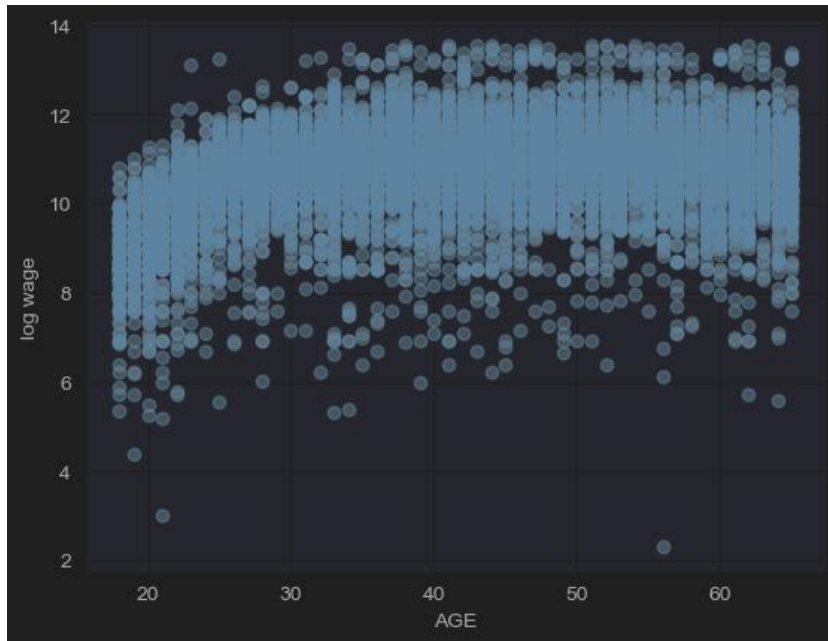

- **Polynomial Regression**:

    o Captures potential nonlinear relationships; useful for modeling acceleration or deceleration in wage increases with education.

- **Cubic Spline Regression**:

  - Highly flexible, allowing data-driven relationships with minimal assumptions.

Comparison and Model Selection

- **Flexibility:** Cubic spline models provided the highest flexibility, closely fitting complex relationships.

- **Interpretability:** OLS and polynomial models offered clearer interpretability, while spline models captured nuances.

- **Performance:** Cubic splines improved model performance through reduced residual variance and better visual fit.

Suggested additional plots:

- Residual plots comparing the three models.

- Predicted vs. actual wage plots for model performance comparison.

Key Findings

- Higher education levels correlate strongly with higher wages.

- Significant wage gaps exist between different education attainment groups.

Recommendations

- Policies promoting educational attainment could lead to improved economic outcomes.

- Further research could explore additional variables such as field of study, geographic location, and industry-specific factors.

Conclusion

This analysis underscores the value of higher education in enhancing wage prospects and suggests that educational policies play a critical role in addressing income disparities.

Future Work

Future analyses could include longitudinal studies, deeper segmentation by demographic factors, or the incorporation of external economic indicators to further understand the dynamics between education and income.