STAT 207 - Project 3

STAT 207: Data Science Exploration

Quinn Crockling & Gabe Price

UNIVERSITY OF
ILLINOIS
URBANA-CHAMPAIGN

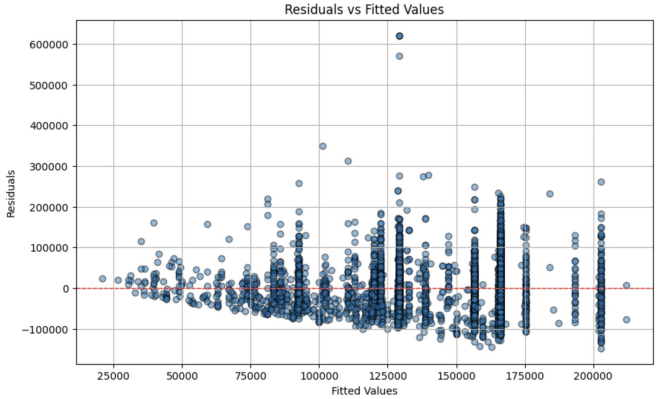
Background

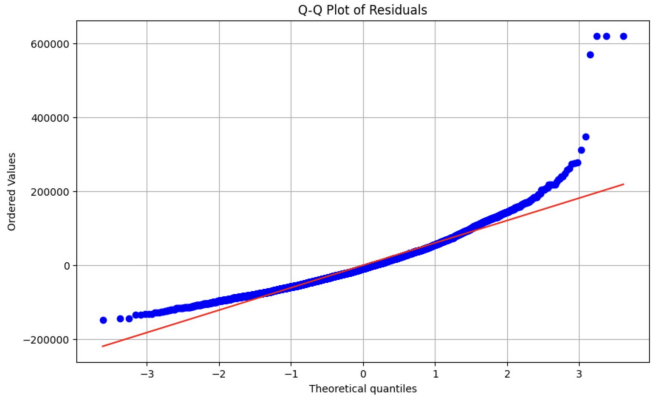
- **Dataset:** contains global data on data science roles, including job titles, expertise levels, salaries, company locations, and employee residences from 2024.
- **Linear Model:** explores salary trends using linear regression to identify factors such as expertise level, location, and year influencing compensation.
- **Logistic Model:** utilizes logistic regression to predict the likelihood of professionals residing in the U.S. based on variables like salary, expertise, and company location.

Linear Regression Analysis

Research Question: How do expertise levels, company size, and other factors influence the salary trends for data science professionals over time, and how can we predict future salaries across different expertise levels in the field?

- Higher salaries are linked to U.S. residence and advanced expertise, highlighting location and expertise as key salary factors.
- The low adjusted R^2 (0.205) suggests significant salary variability is influenced by unmeasured factors, requiring more robust models.

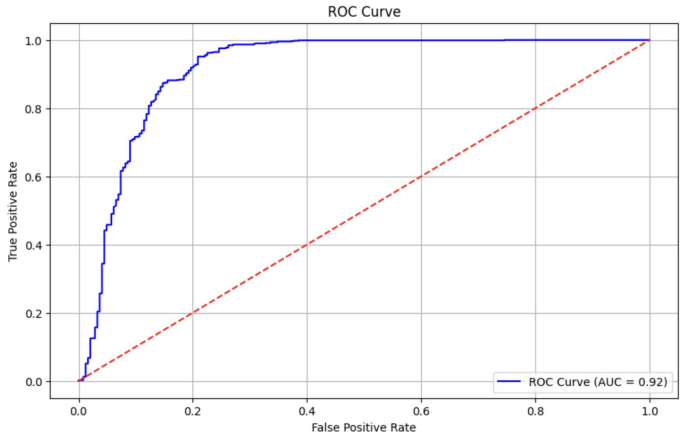




Logistic Regression Analysis

Research Question: What are the key factors influencing the likelihood of an employee's residence being in the United States for Data Science related jobs, and how accurately can these factors predict residence?

- Higher salaries and company location significantly predict U.S. residence for data science professionals.
- The model's strong AUC (0.92) highlights excellent classification performance, but some variability may still be unexplained.



Thank You

STAT 207: Data Science Exploration

Quinn Crockling & Gabe Price

UNIVERSITY OF
ILLINOIS
URBANA-CHAMPAIGN