Data Coin

Influential Factors on Data Scientist Salaries as Reported on Glassdoor

Celia Sagastume, September 2021

Data Scientist

Job Search

Salary Range

Per Job Posting

Regression Model on Data Scientist Salaries as Reported on Glassdoor

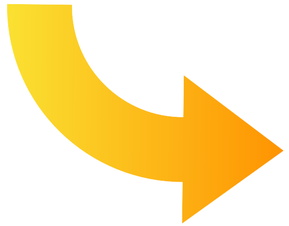
Model Notes

- 1,305 Samples
- 36 cities in the United States
- Scraped using Selenium

- Generated Features:
  - Age
  - Geolocation
  - Industry Map
- Binary Mapping:
  - Additional Pay
  - Internship
- Interaction term
  - West & Tech Industry
- Target Variable:
  - Salary

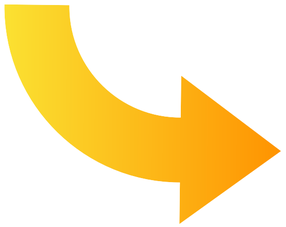
Glassdoor Salary Posting Search by City

- **Salary**
- **Salary Range**
  - Low, 25%, 75%, High
- **Additional Pay**
  - Cash Bonus
  - Stock Bonus
- **Total Pay**
- **Glassdoor Review Data**
  - Reviews, Jobs, Salaries, Interviews, Benefits, Confidence Score
- **Internship**
- **City**



Glassdoor Company Information by Result

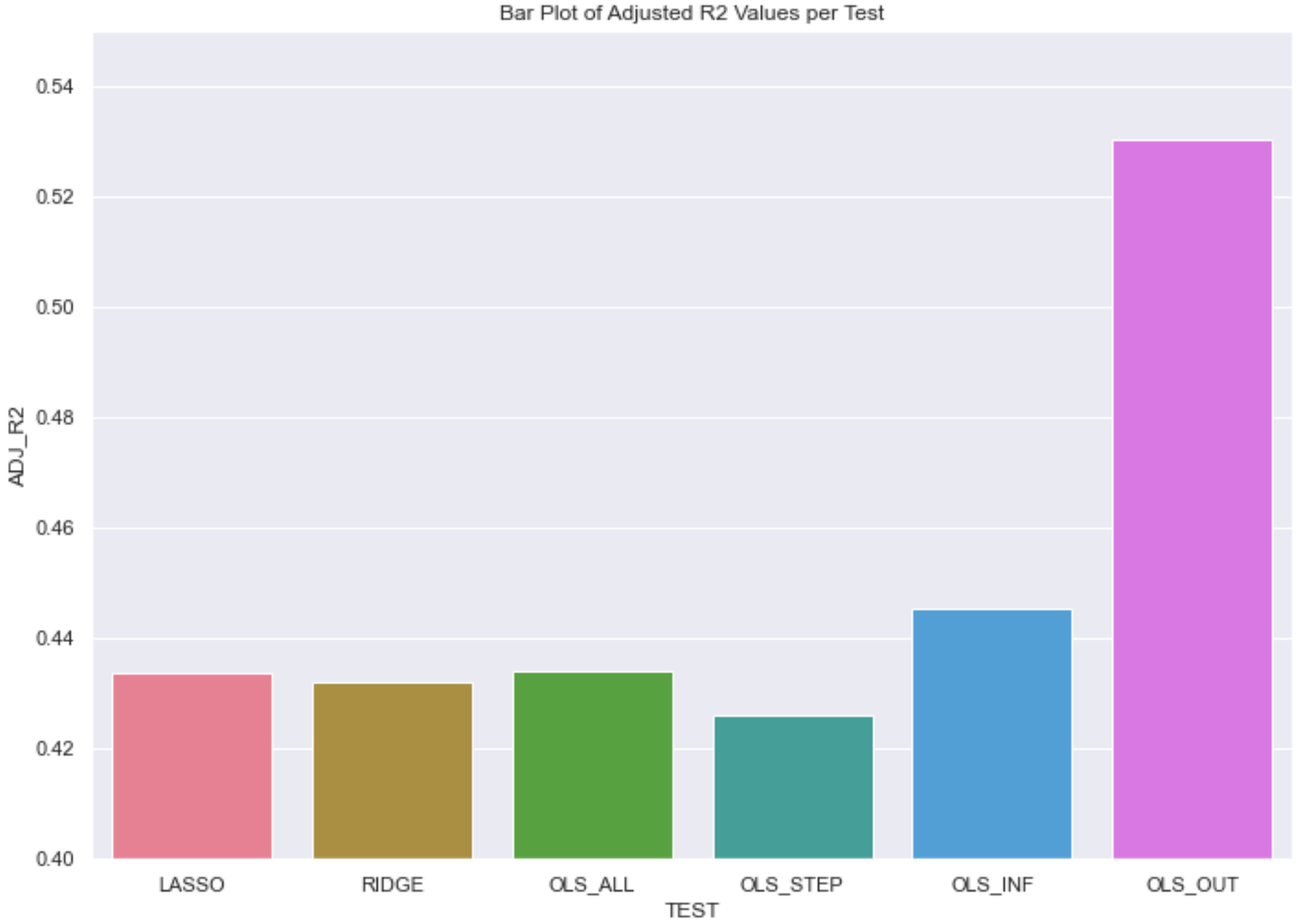
- **Company Type**
  - Public, Private, Government, Education, Subsidiary
- **Company Size**
  - Categorical
- **Industry**
- **Headquarter**
- **Year Founded**

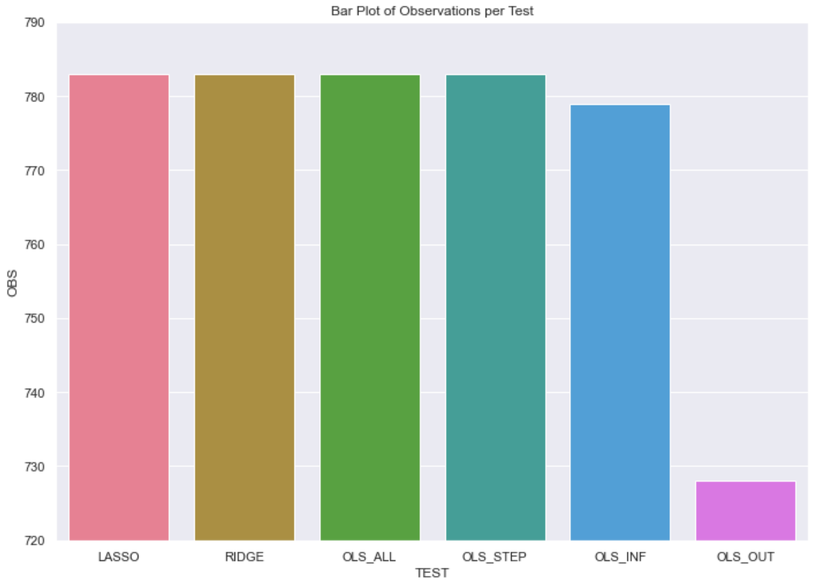


Numbeo Cost of Living Index by City

- **Cost of Living**
- **Purchasing Power**
- **Rent**
- **Rent + Cost of Living**

Model Performance



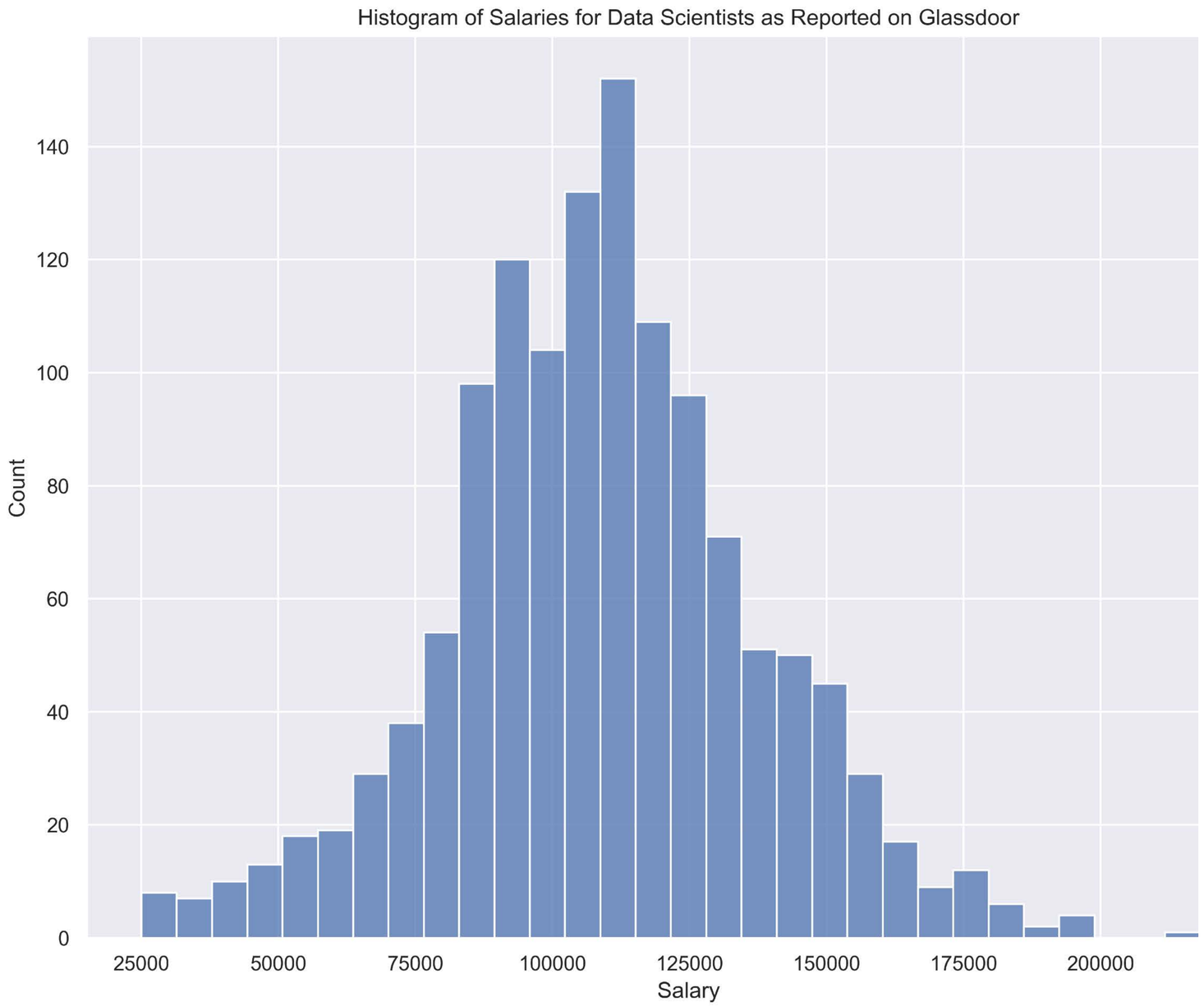


Performance Notes

- Best R^2 Value:
  - Ordinary Least Squares with Outliers Removed, *Adjusted R^2 = .530*
  - *Outliers Removed:* 55
- Second Best R^2 Value:
  - Lasso CV, *Adjusted R^2 = 433*
  - *No outliers removed*
  - *MAE = 16138.464503*
  - *RMSE = 20873.19*
- *Final Model: Lasso*
  - *MAE = 16,669*
  - *RMSE = 22,055*
  - *Adjusted R2 = .401*

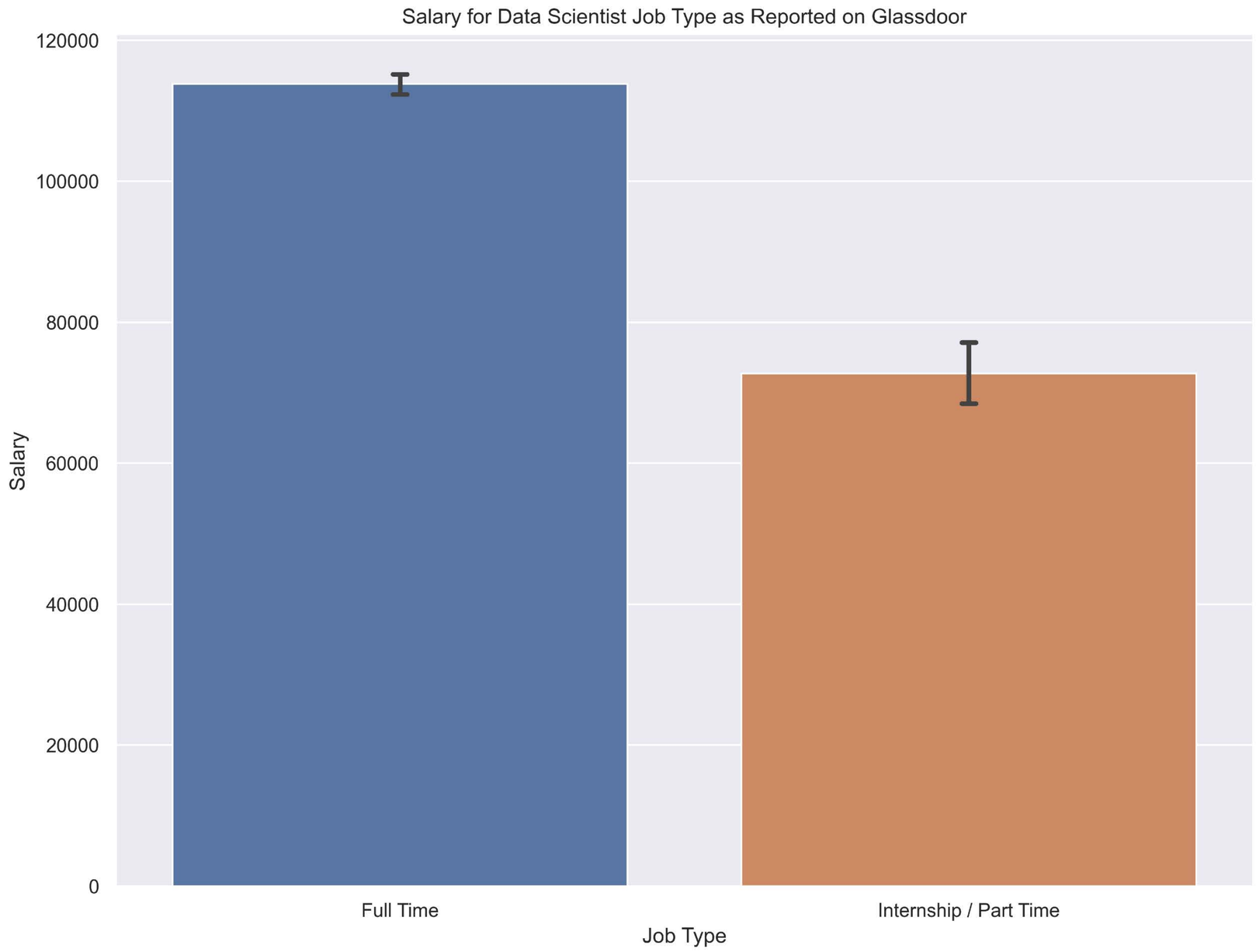
Data Scientist Salary

Mean: 109,038.829
Standard Dev: 28, 715



Internship

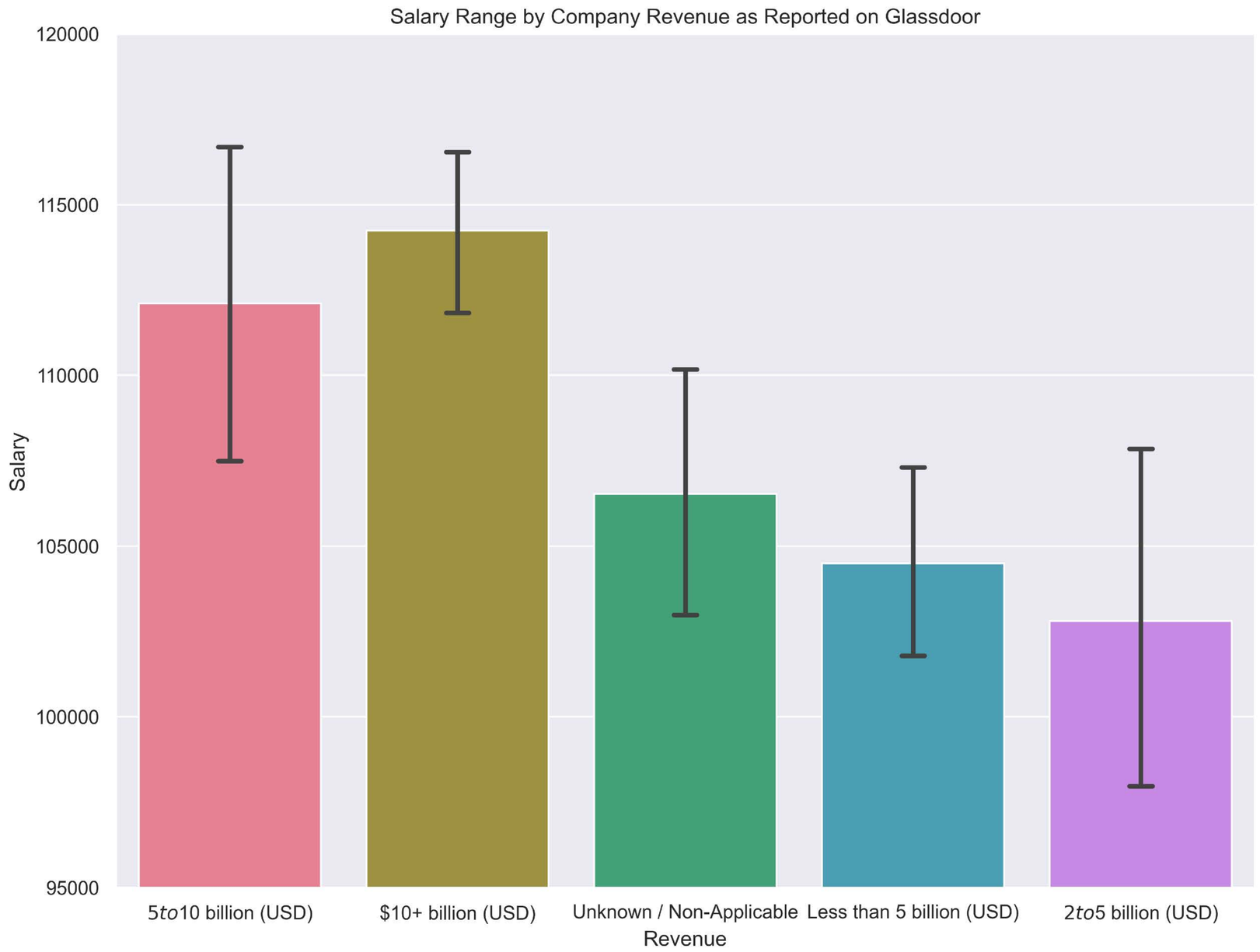
⬇37953.57



Revenue

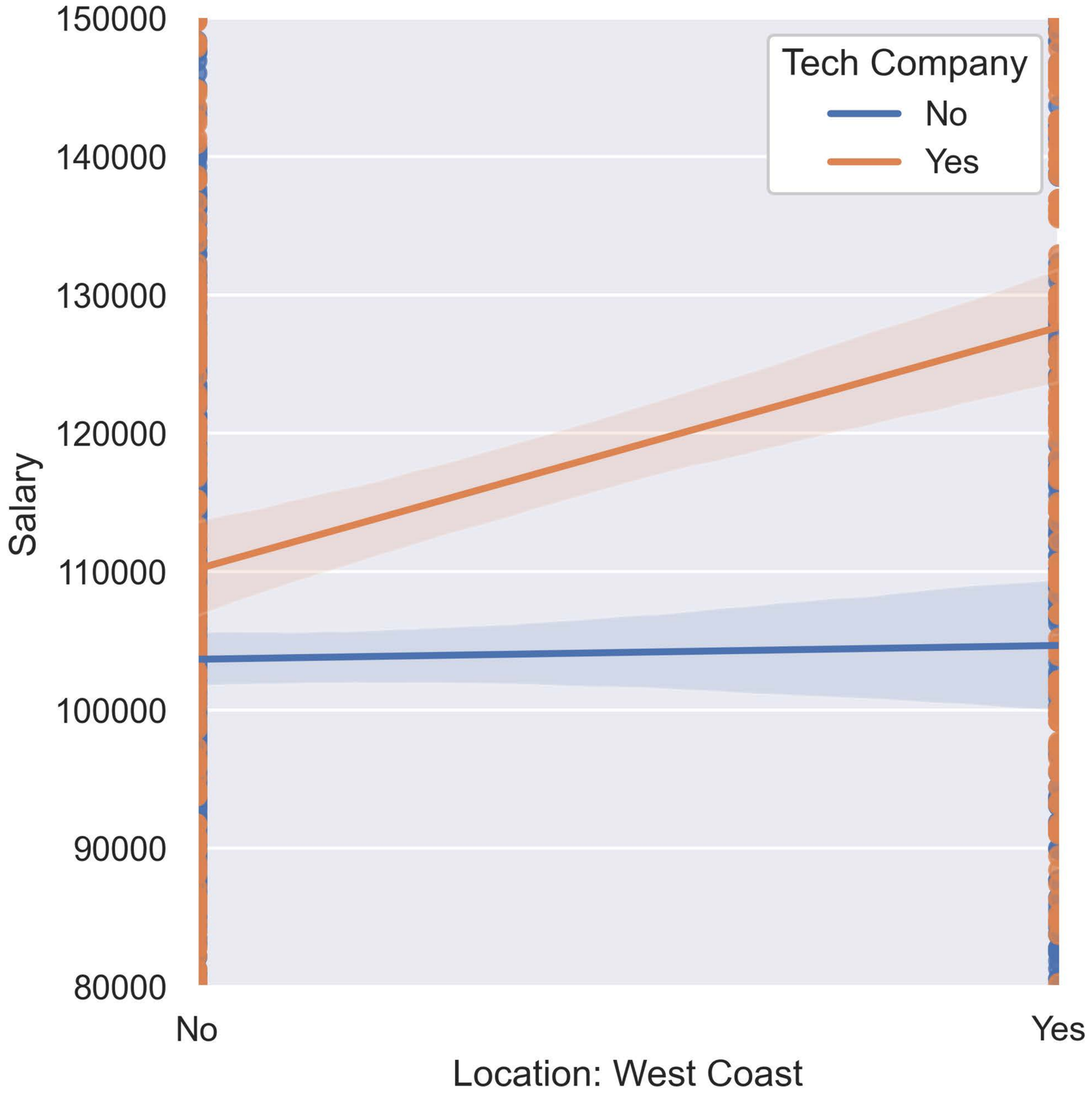
⬇ 15144.99

(2 to 5 Billion compared to 10 + Billion a year)



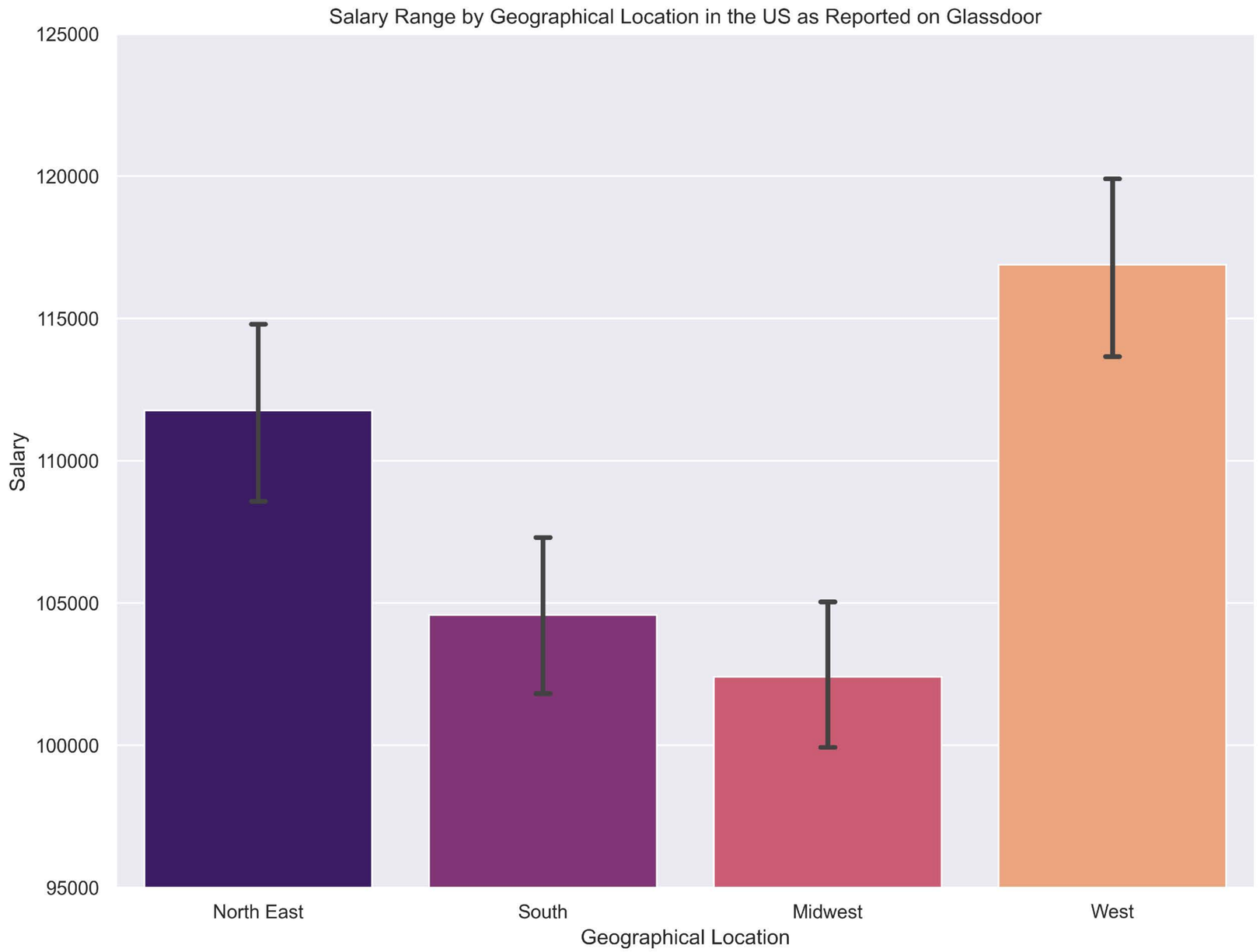
Interaction: West Coast & Technology

⬆ 11300.91



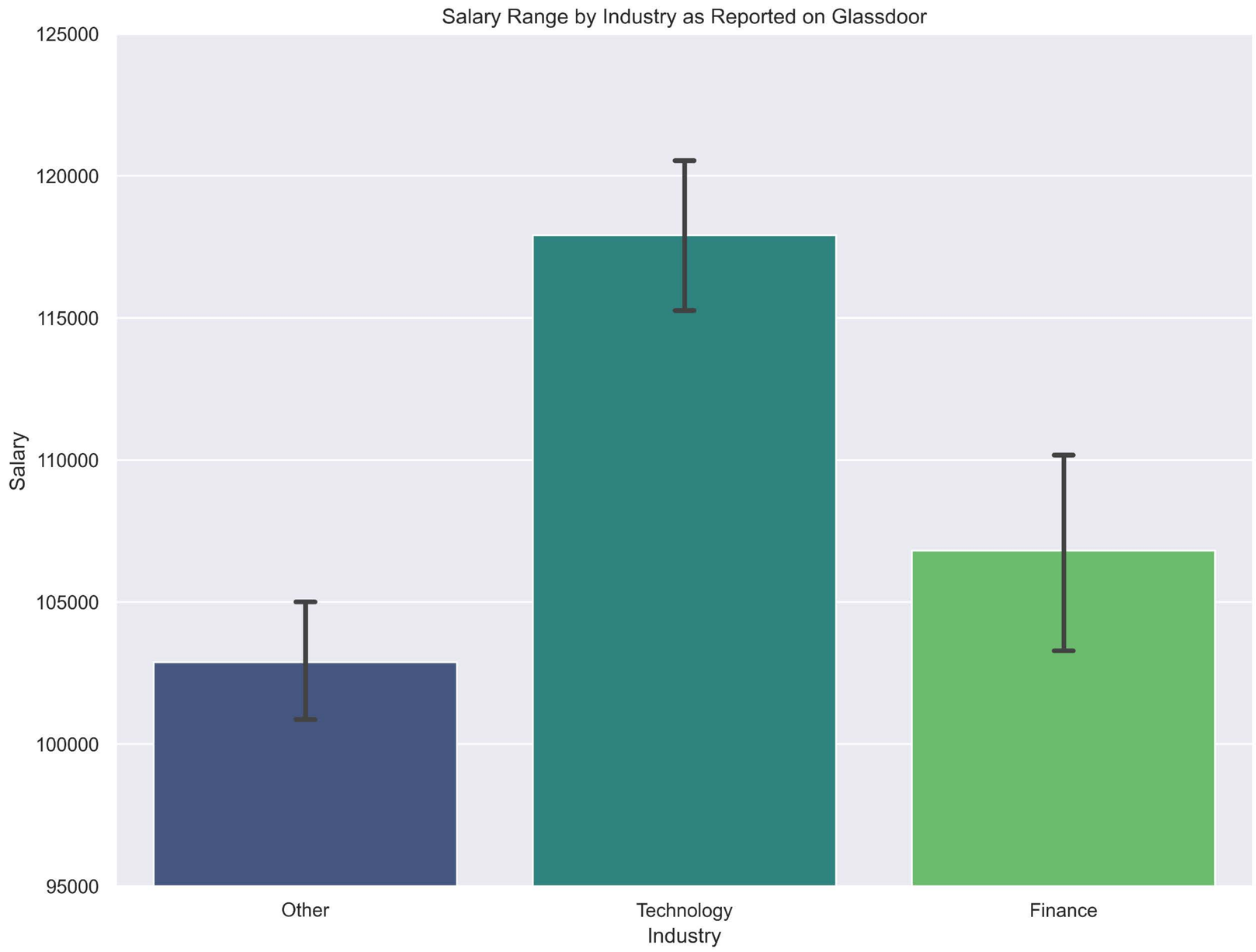
Interaction: West Coast

⬆ 11300.91



Interaction: Technology

⬆ 11300.91



Next Steps

- Interesting Findings:
  - Cost of Living Coefficient varied by data set
  - (Train: -28.14, Val: 368.99, Test: -75.92)
- Consider Headquarter of Company in conjunction with City of Job Posting
- Take into account Seniority level
- Language associated with Job Description
    - “Machine Learning” vs “Analyst”

Questions