

Predicting Data Science Salaries Based on Job Post Features on Indeed.com

Mai Tran

Metis

1/18/2022

Client – Indeed.com

- With the explosion of data and data science jobs, Indeed.com (the #1 job site on the internet) has tasked me to do predictive analysis of their job post features for various data science job roles:
 - Enhance user experience of job seekers including data science professionals
 - Keeping users informed and engaged on the site
 - Salary transparency
 - To avoid wage theft and exploitation
 - To research and explore new UX features

new

Data Analyst

Forrest Solutions 3.1 ★

New York, NY 10036(Midtown area)

 \$100,000 a year

 Temporary +1

➤ Easily apply  Urgently hiring

- Methodically collect **data** from various sources into standardized formats.
- Work with analytics team to define processes for **data** collection.

7 days ago

Design – Scraping Features

1. Salary (low end) - **target**
2. Company Rating (low < 3.5, high > 3.4, out of 5 stars)
3. Remote vs Non-Remote
4. Old vs New (Old = older than 29 days)
5. Non-Urgent vs Urgent
6. Non-EZ vs EZ (Easily Apply Button)
7. Job Title
8. Company
9. Location

Design – Process

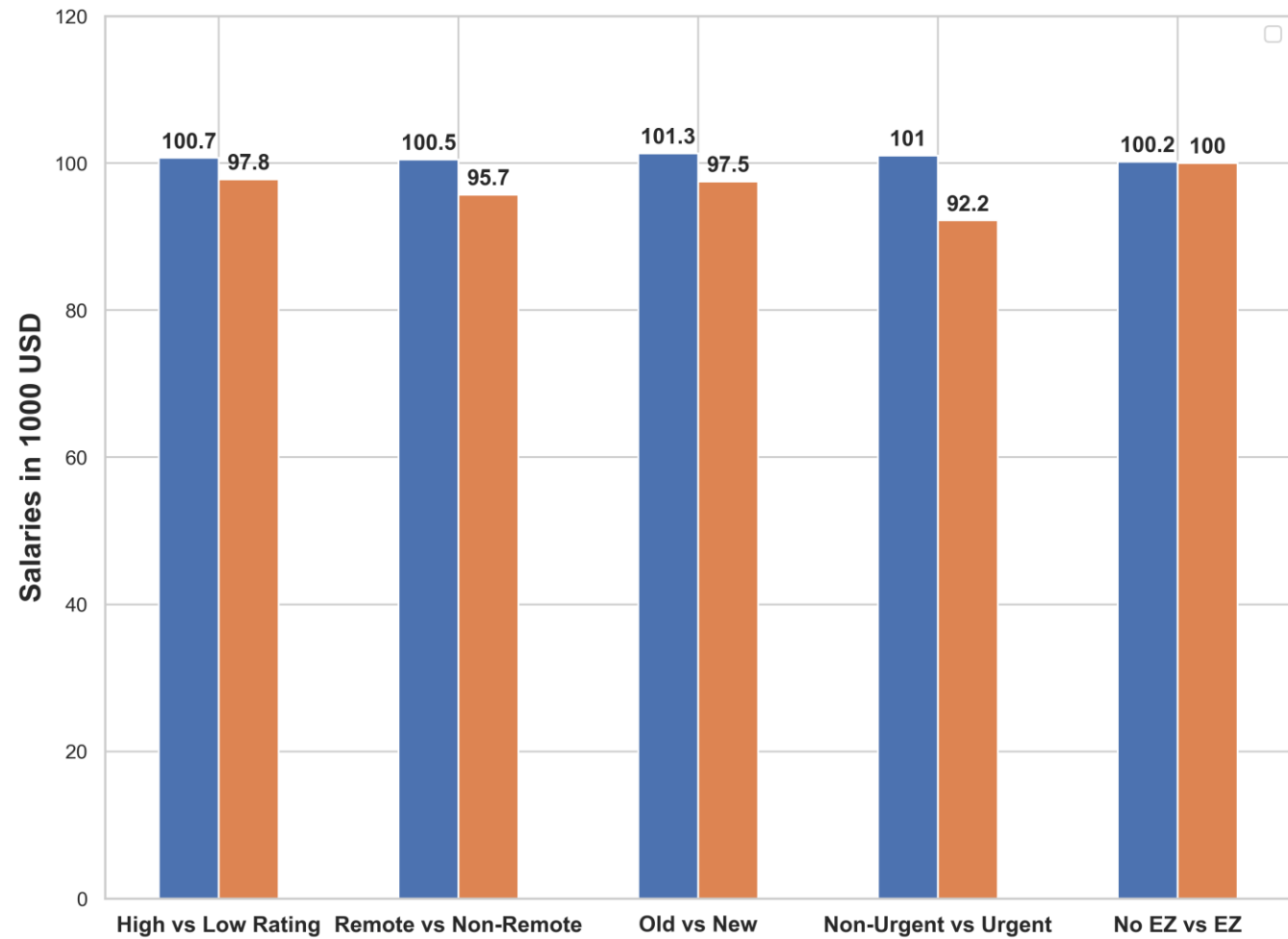
- Extracted features from four main Data Science Job Roles:
 1. Data Scientist (1025 jobs)
 2. Data Analyst (1053 jobs)
 3. Data Engineer (664 jobs)
 4. Machine Learning Engineer (1002 jobs)
- Performed predictive analysis on features using regression methods:
 1. Linear
 2. Lasso
 3. Ridge
 4. Elastic Net

EDA Results

- Features vary widely between job titles
- Only 30% of all job posts are younger than 30 days
- Machine Learning Engineer has highest median low-range pay, which is 107k
- Data Analyst has lowest median low-range pay, which is 68k, but pays more for Non-Remote positions
- Data Scientist Remote Jobs pay more than Data Scientist Non-Remote Jobs while all other roles pay more for Non-Remote jobs
- Data Engineer Non-Remote pays highest compared to other Non-Remote roles
- Data Analyst and ML Engineer pay more for urgent positions

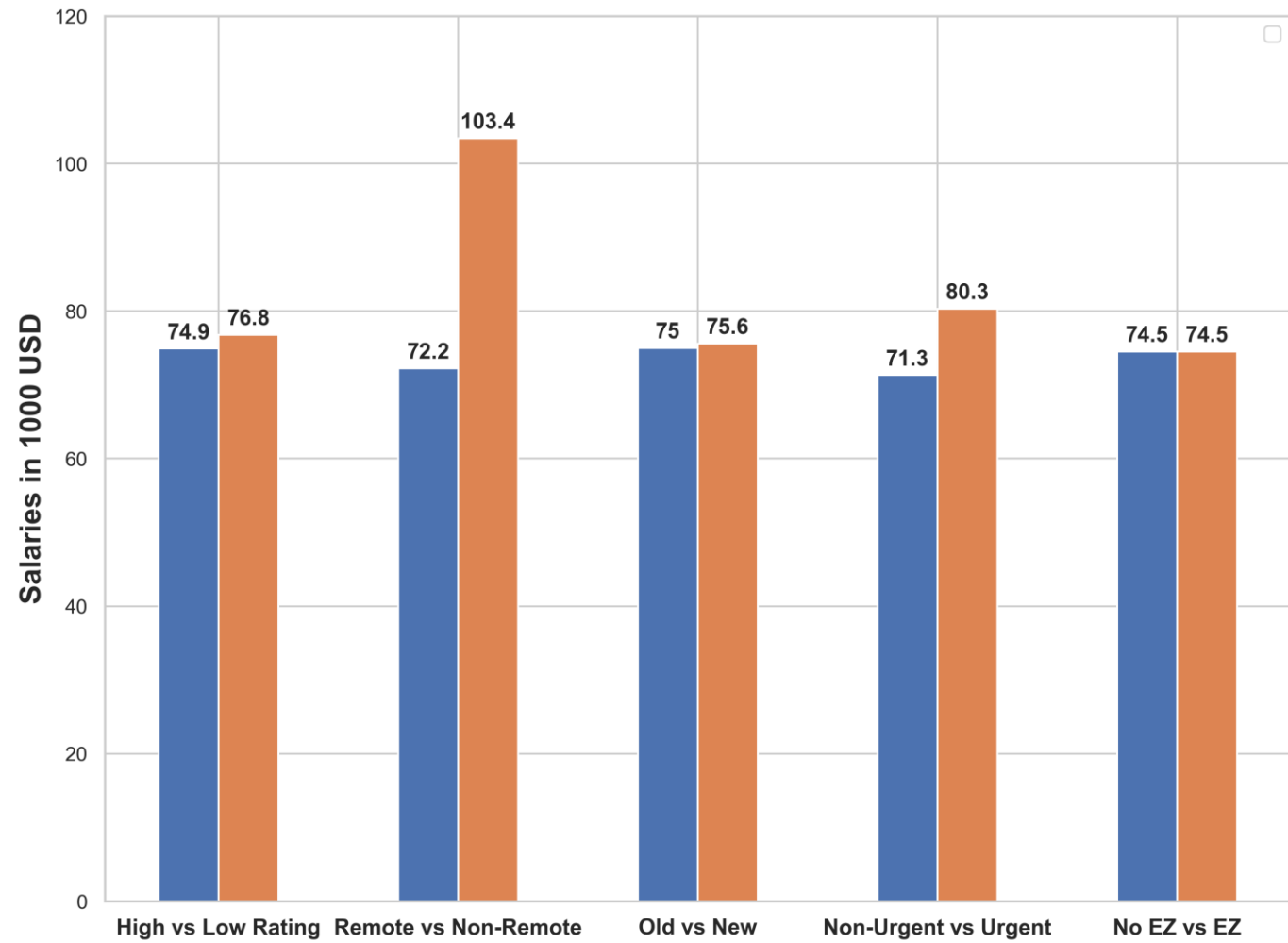
Data Scientist

Data Scientist Salary Differences Based on Job Post Features



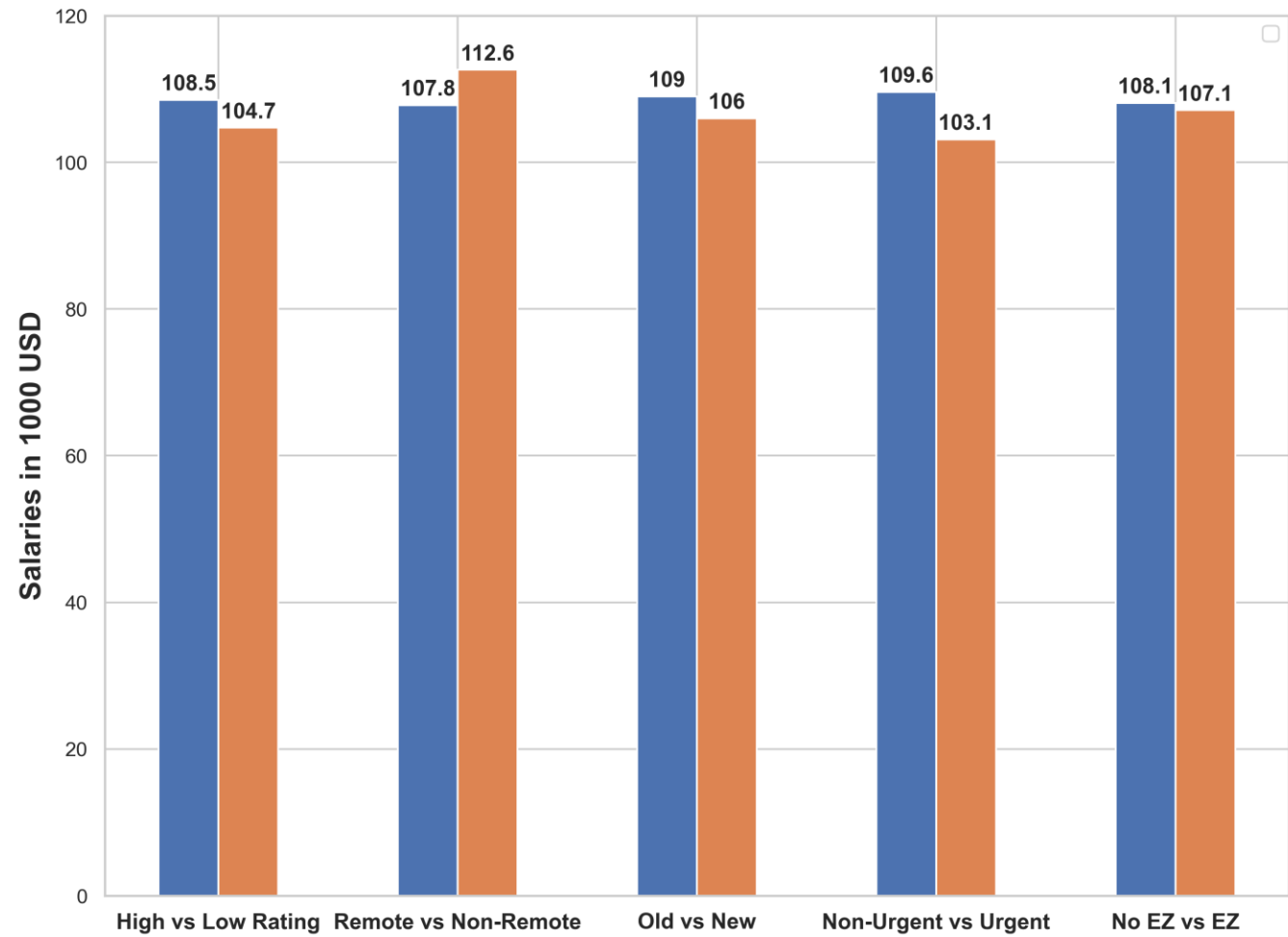
Data Analyst

Data Analyst Salary Differences Based on Job Post Features



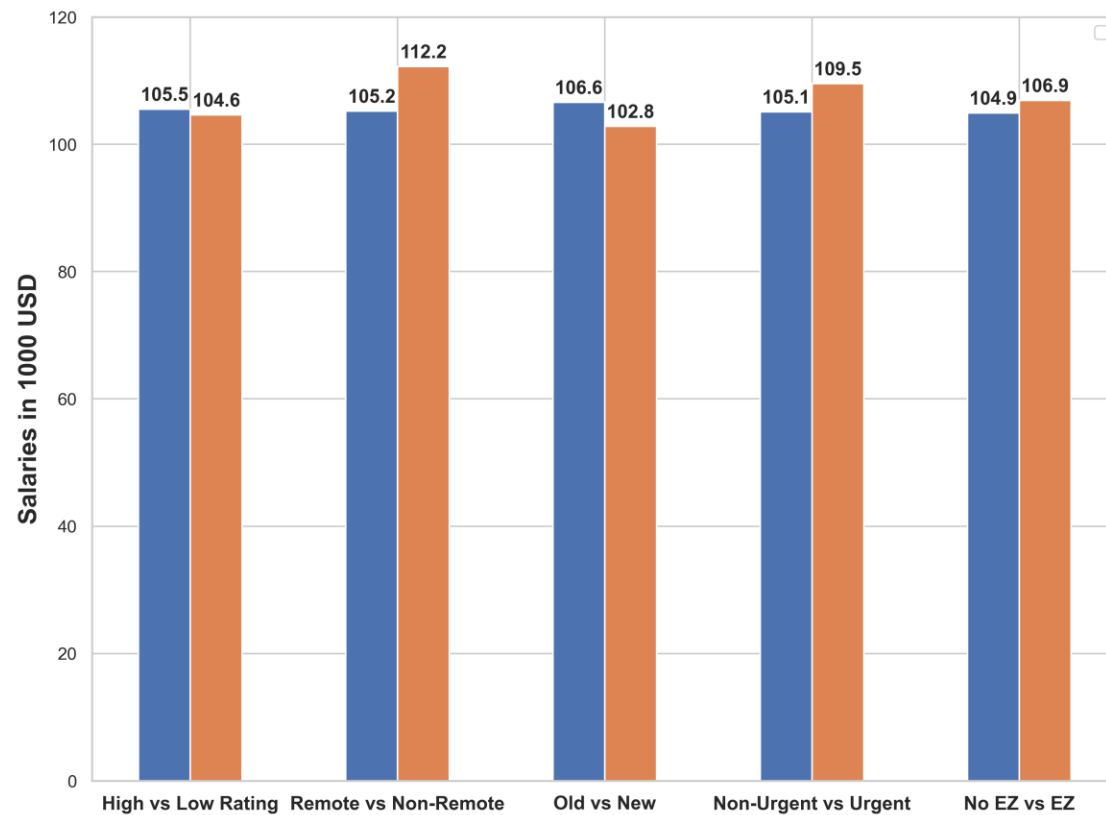
Data Engineer

Data Engineer Salary Differences Based on Job Post Features



Machine Learning Engineer

Machine Learning Engineer Salary Differences Based on Job Post Features

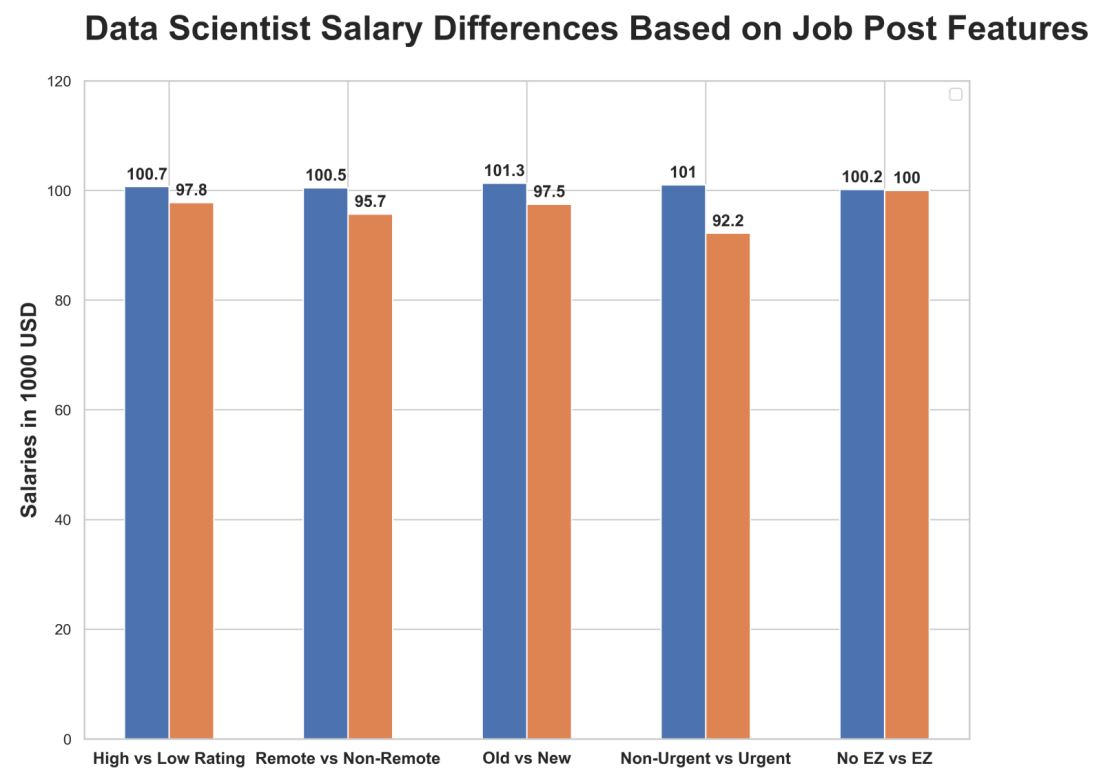
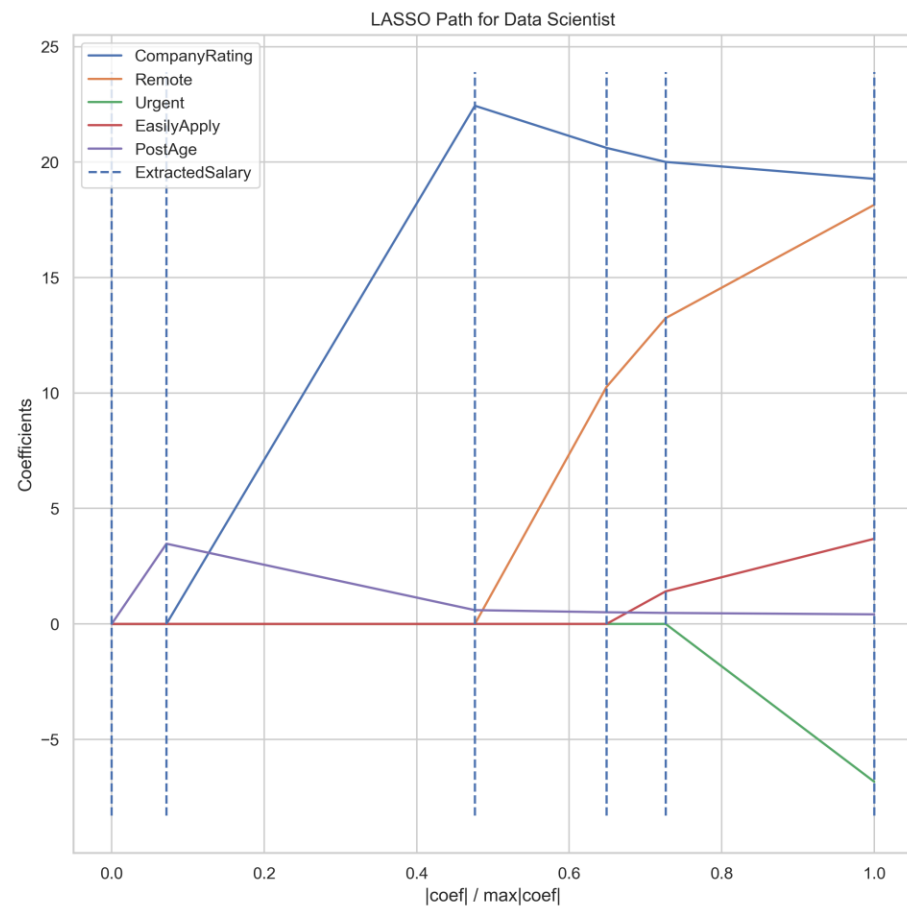


Predictive Analysis

- Data Scientist role
- No correlation at all – no best fit line
- Median Absolute Error (MAE) instead of (R^2)
- Used Cross-Validation with K-fold with 5 splits

Regression	Training (MAE)	Testing (MAE)
Linear	10.6	14.5
Lasso	10.9	12.0
Ridge	10.8	14.3
Elastic Net	10.8	12.0

Data Scientist LARS Path



Recommendations

- Each job title should have its own predictive model
- Recommend data job seekers to focus on high-paying urgent data roles that often have Easily Apply feature. Exs: Data Analyst and ML Engineer
- Use high-paying urgent roles with Easily Apply feature to gauge job demand
- Add "Urgently Hiring" and "Easily Apply" as filtering features in drop-down menu in search home page
- Years of experience as a job post feature since that feature has some predictability power for salary