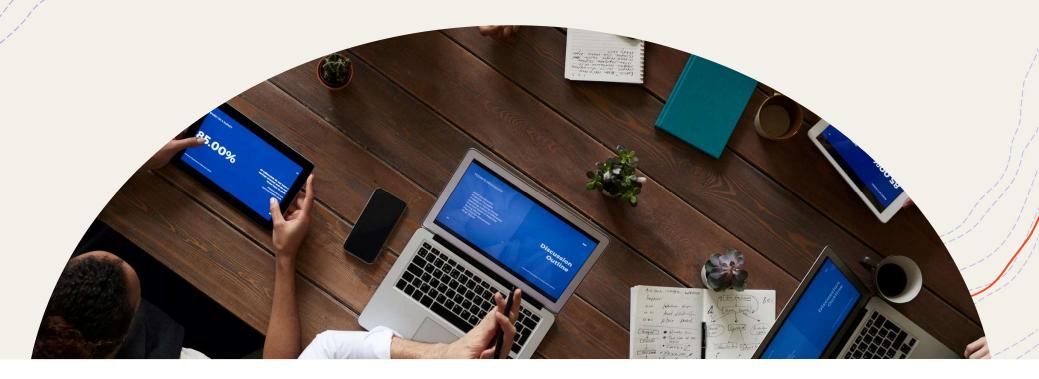# Want to get paid?

A deep dive into Data Science job postings on Glassdoor.com

By Jason Palumbo

An Introduction
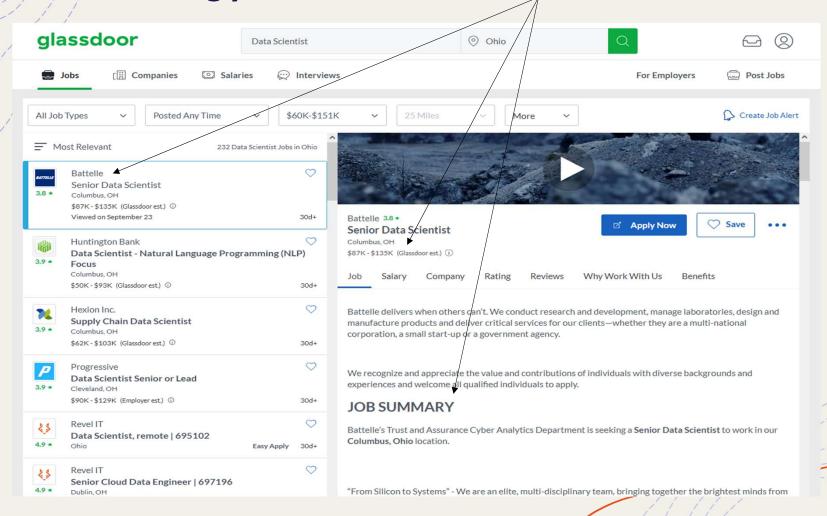
What features of a job listing can best predict the salary of a paid work position within the field of Data Science?

Methodology

Feature Rich Data

Data Transformations And Cleaning

"natural language processing"

"machine learning

"artificial intelligence"

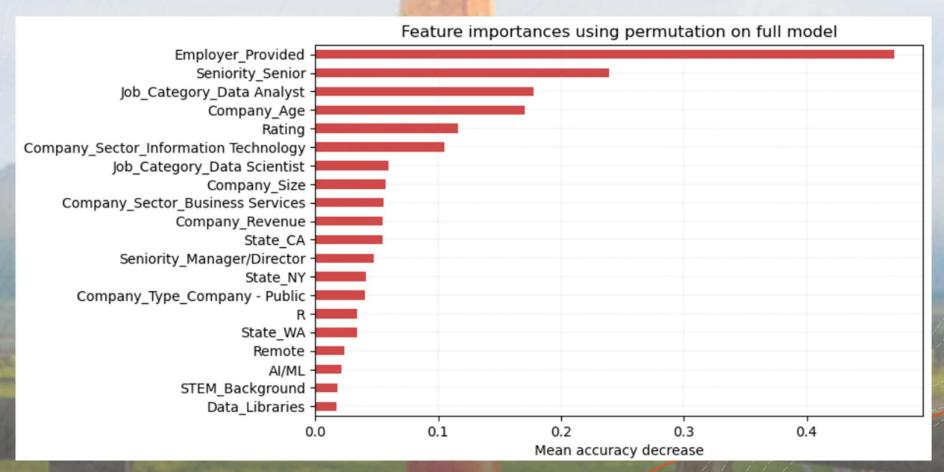"deep learning"

"neural networks"

"ai/ml"

"ai"

"ml"

"nlp"

→ "ML/AI/NLP"

Results

Model Table

	R^2	μ RMSE	σ RMSE
Linear Regression	.51	$19k	$3k
Lasso Regression	.51	$19k	$3k
Ridge Regression	.51	$19k	$3k
Random Forest	.94	$16.7k	$3.8k

On average the Random Forest model can correctly predict a salary within 15% of the projected salary median of $103k

Conclusion

Feature importances using permutation on full model

IMPACT

For a practitioner who is financially motivated within the data science community, he/she should concentrate on attaining a role at a younger, tech and data focused company, and aspire to enter a senior non-Data Analyst role.

FUTURE WORK

- Independent sources of data
- Normalize the data based on metrics like population
- Use GDP per capita to compare salaries across states or regions.
- Explore modeling with XGBoost.

+Jason Palumbo
+GitHub: @giasonep