Analysis of Data Scientist Salaries

Jingfei Cai
July 28, 2017

Agenda

- Objectives

- Data Collection

- Data Analysis
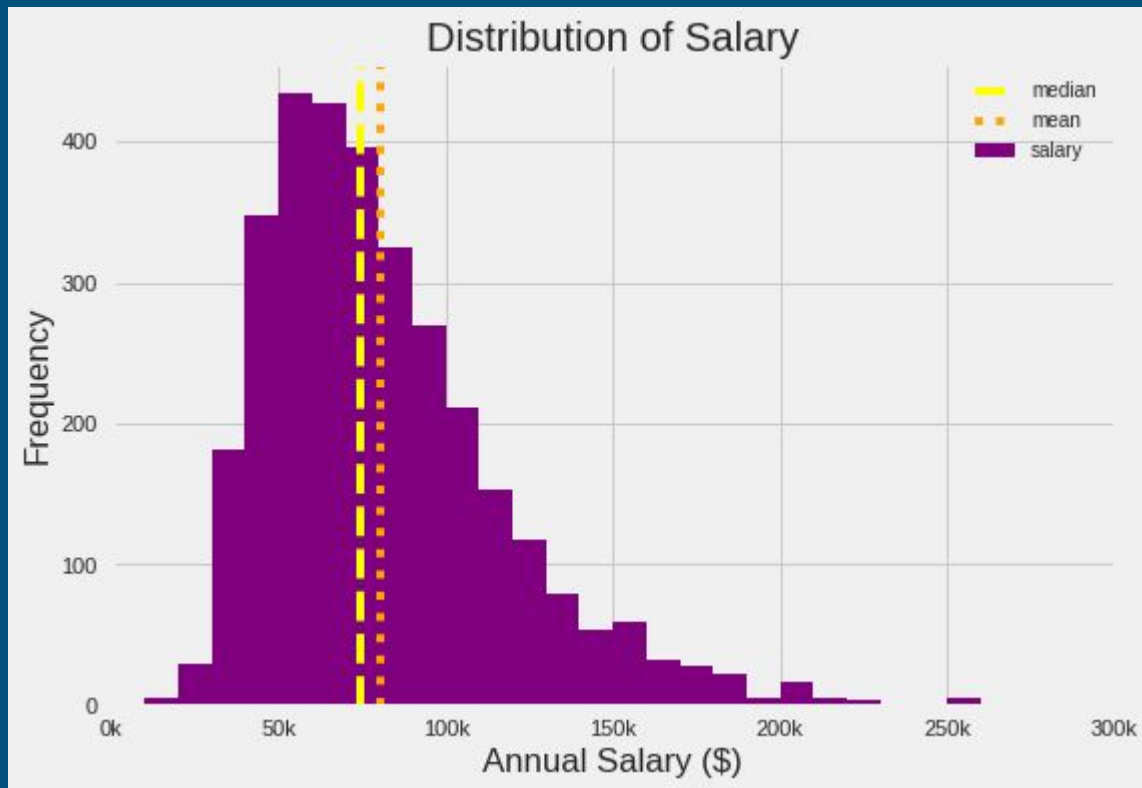
- Summary

Objectives

- Predict salary levels (above/below median) for data science jobs

- Identify factors that impact salaries

  - Location? Job title? Job description?

  - Keywords?

Data Collection

- Source:

  Indeed.com

- Query keywords:

  data scientist, city name (~90 search terms, within 100 miles radius)

- Information collected:

  company, location, job title, job description snippet, salary

- Data actually used in the models:

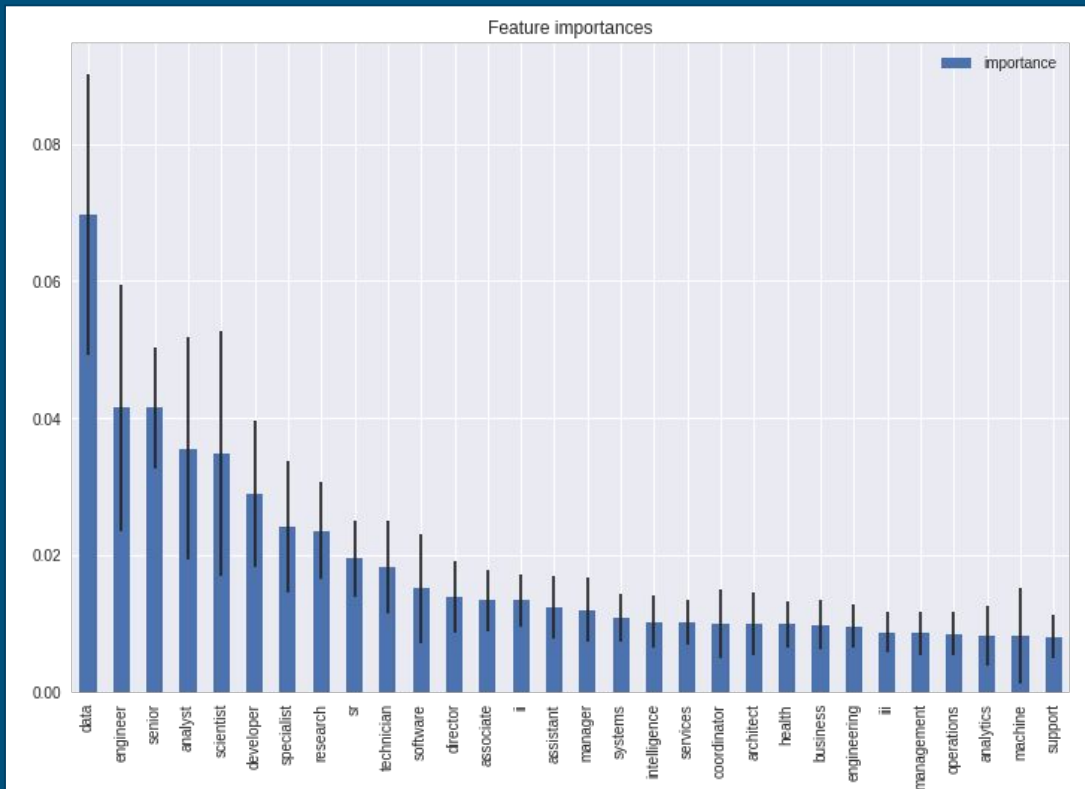 Job listings with annual salary information

Data Analysis

- ~ 3,200 job listings from 48 US states
- Median salary: $74,669
- Best accuracy: 73% ± 6%
- Most impactful factor: Job title

Important Keywords in Job Titles

Important Keywords in Job Titles

Feature importances

Keyword	Impact	Keyword	Impact
data	positive	software	positive
engineer	positive	director	positive
senior	positive	associate	negative
analyst	negative	ii	negative
scientist	positive	assistant	negative
developer	positive	manager	positive
specialist	negative	systems	positive
research	negative	intelligence	positive
sr	positive	services	negative
technician	negative	coordinator	negative

Limitations / Next Steps

- Moderate accuracy

- Need to accumulate more data over a longer period of time

- Collecting the full job description (not just the snippet) may improve prediction accuracy and provide more insight on which skills and keywords are of great significance

- Consider collecting data using other methods such as surveys

Summary

- Job title is the most impactful factor in predicting salary levels

- Important keywords identified (with positive or negative effect on salary)

- Limited accuracy

Any Questions?