Indeed Salary Analysis for Data Science

Matthew Terribile

Data Scientist

July 2017

Summary

- Objective is to understand what factors affect salaries for Data Science job postings

- Analyzed job postings on Indeed.com for 16 cities
  - Median salary of $95,000
  - Categorized postings into 'above' or 'below' the median salary

- Salary category (above or below median) can be predicted with about 75% accuracy based on
  - Job location
  - Use of certain words in the Title and Description
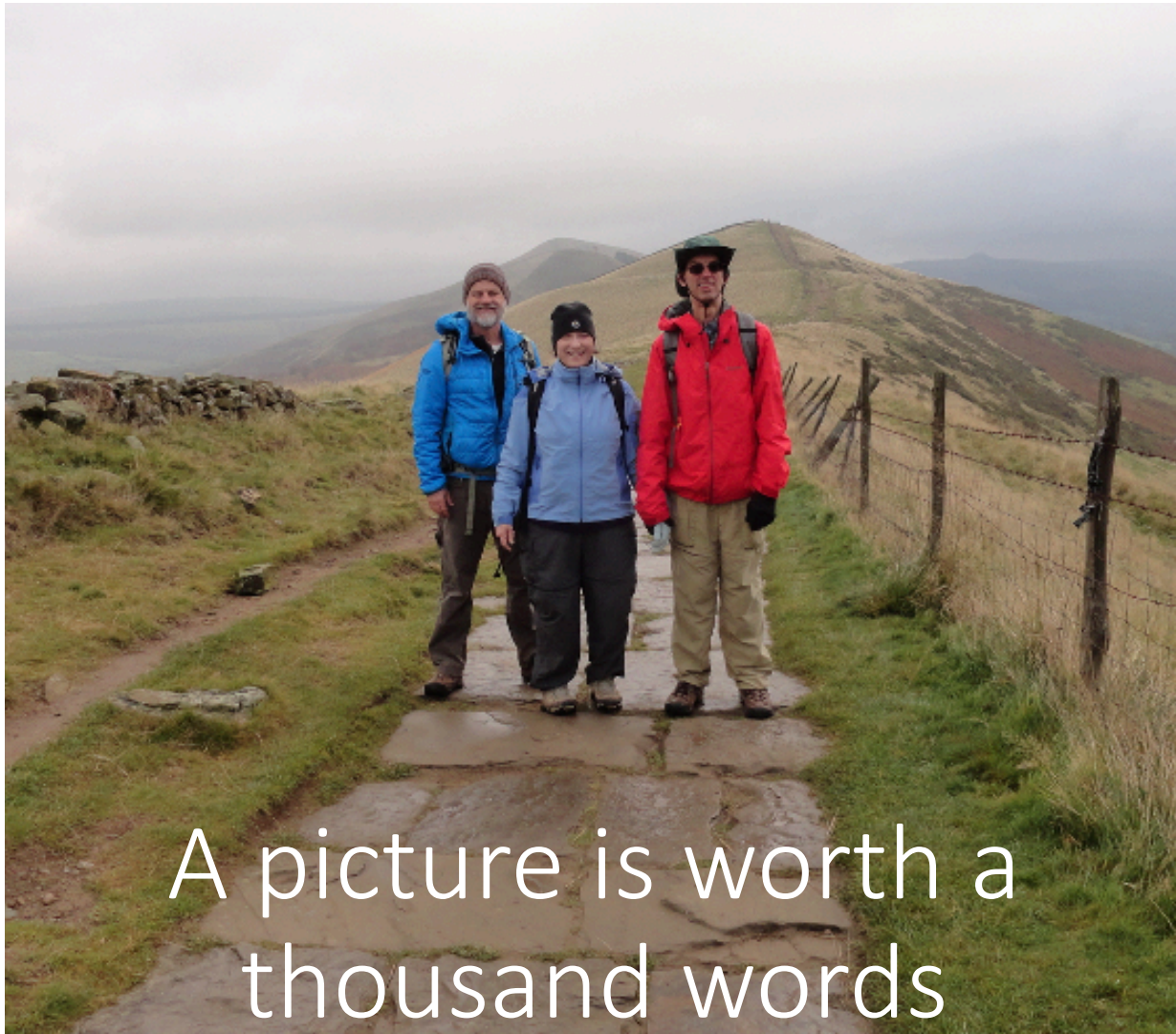
A picture is worth a thousand words
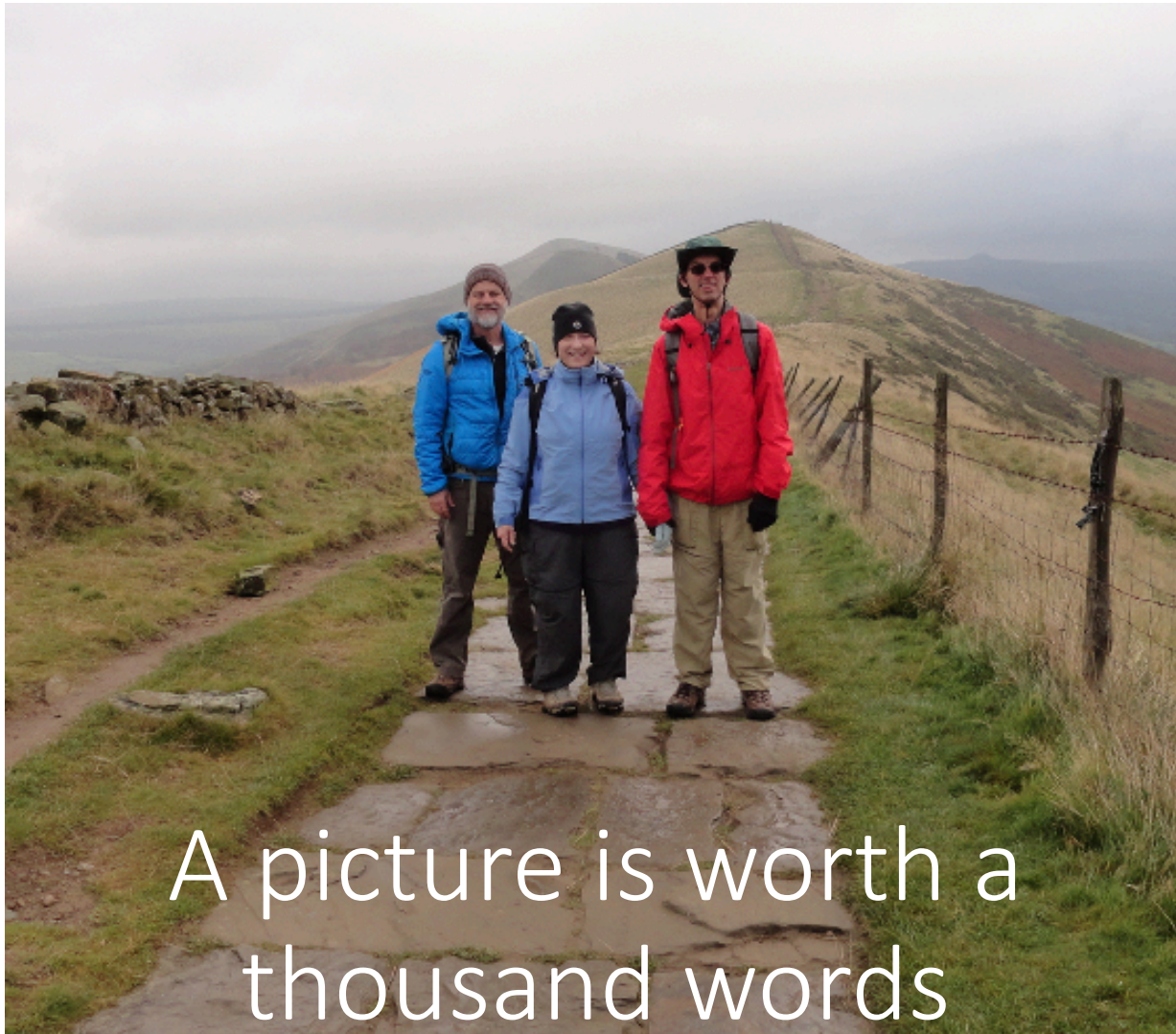
A picture is worth a thousand words

A word can be worth thousands of dollars

Agenda

- Procedures used

- Findings
  - City jobs listed in
  - Words in the title
  - Words in the descriptions

- Caveats and Conclusion

Procedure

- Selected 16 US metro areas including Raleigh

- 'Scraped' over 28k Data Science job listings from those cities

  - Approximately 300 had annual salaries

- Determined the median salary (½ greater, ½ less) to be $95k

- Used Data Science techniques to isolate factors that predict whether a job will be listed with a salary above or below the median

  - Natural language processing

  - Classification algorithms

Location Findings

Some cities are likely to have salaries above the median:

Cities vs. Median Salary
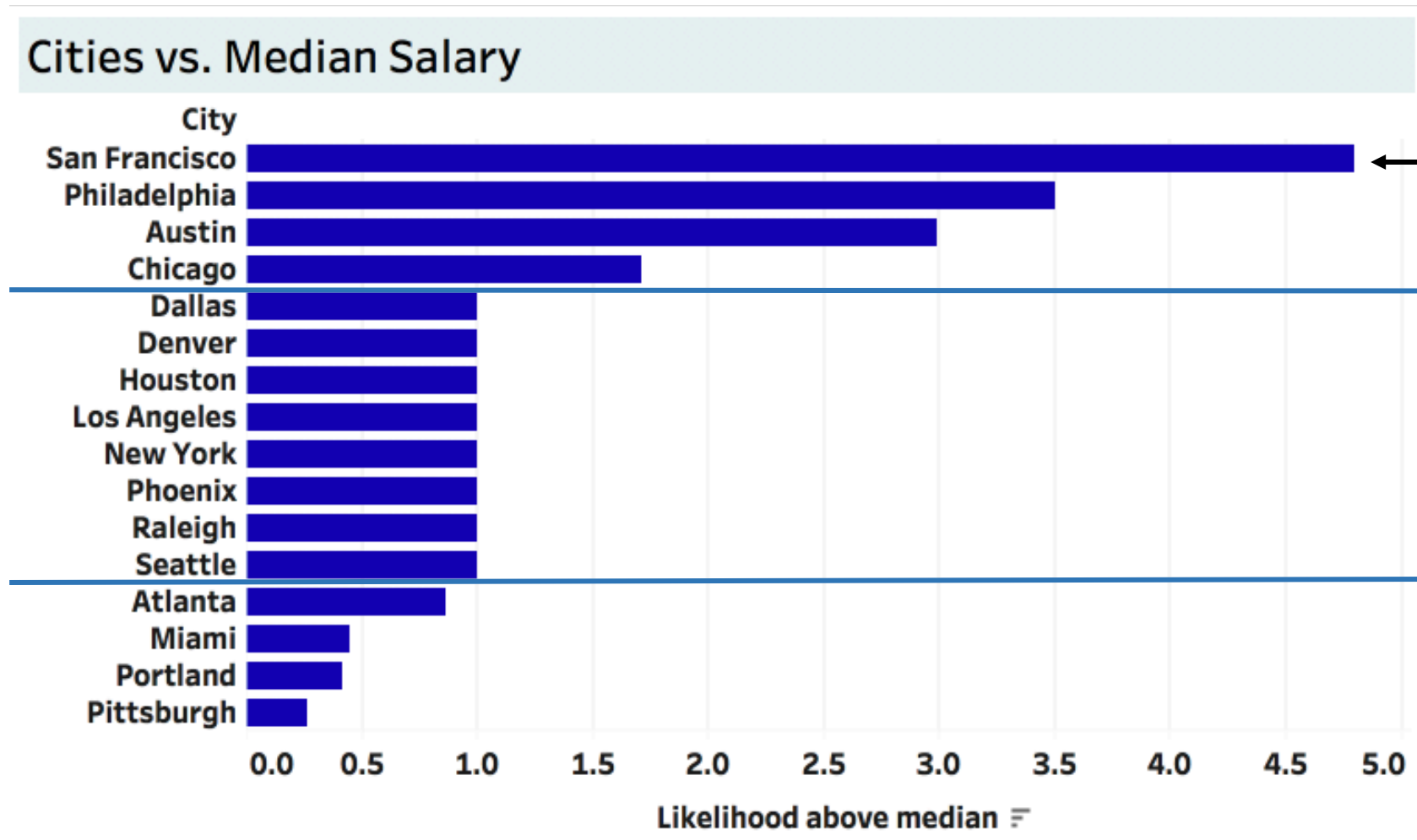
San Francisco postings are 4.8 times as likely to have a salary above the median than postings in
Dallas, Denver, Houston, Los Angeles,
New York, Phoenix, Raleigh, or Seattle

Title Words for Higher Salaries

These words in the title <u>increase</u> the likelihood that salary is above the median:

Title Terms vs. Median Salary
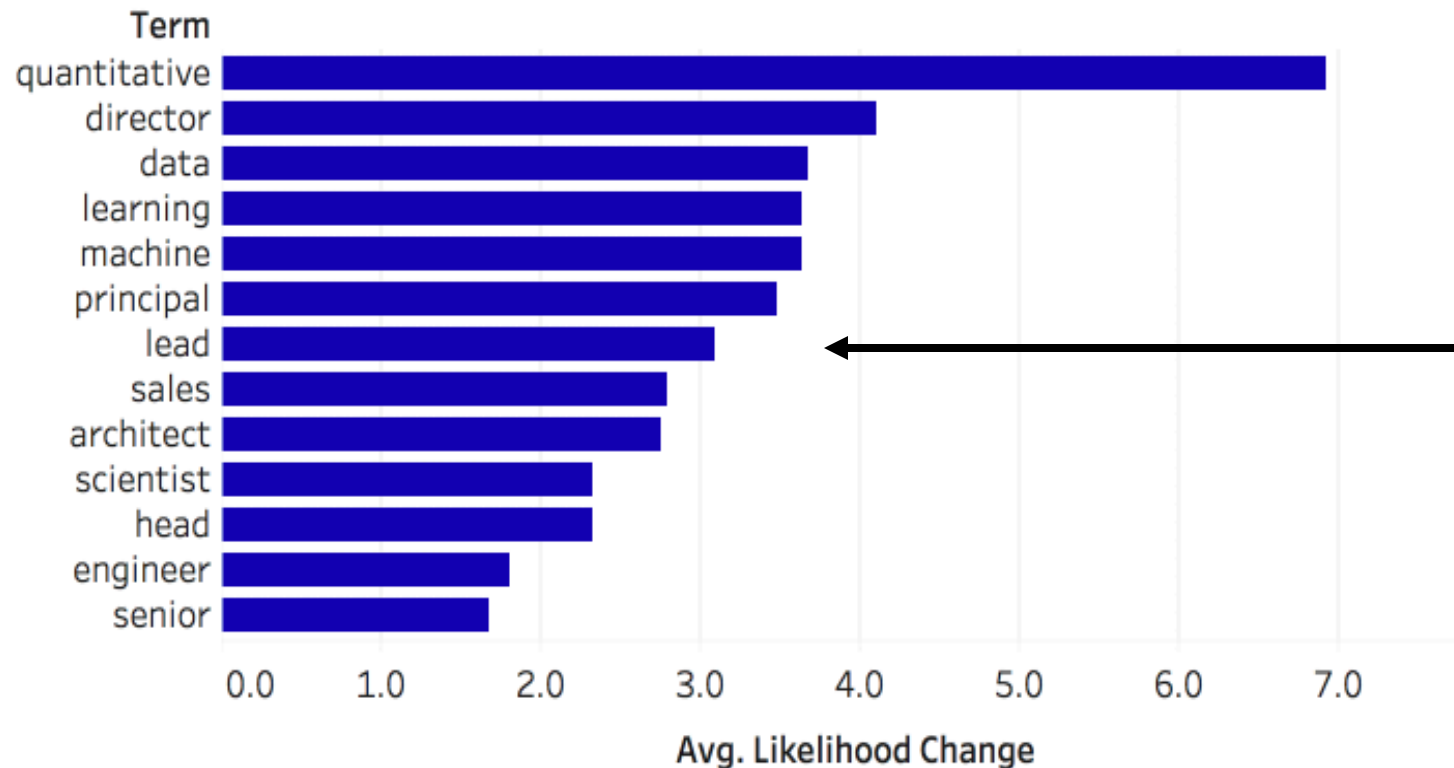
Listing with the word 'lead' in the title is about 3 times as likely to be ABOVE the median salary than those without

Title Words for Lower Salaries

These words in the title <u>decrease</u> the likelihood that salary is above the median:

Title Terms vs. Median Salary

Term
- clinical
- analyst
- medical
- environmental
- laboratory
- research
- bureau
- specialist
- associate
- assistant

Avg. Likelihood Change

Listing with these words in the title are about 2 to 2.5 times as likely to be BELOW the median salary than those without

Description Words for Higher Salaries

These words in the description <u>increase</u> the likelihood that salary is above the median:

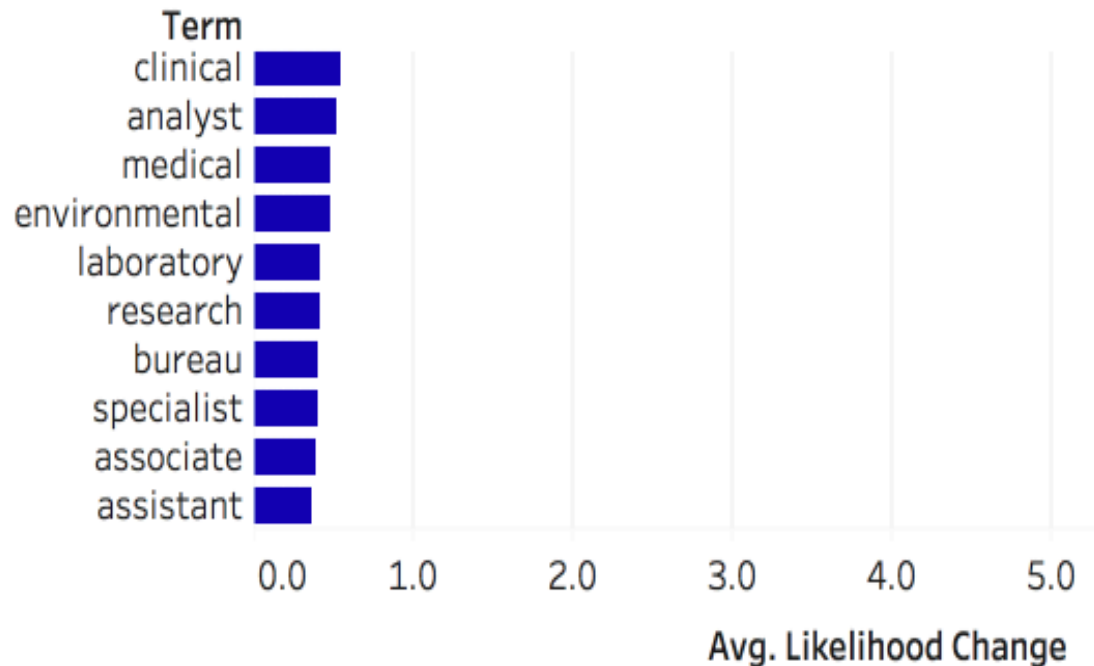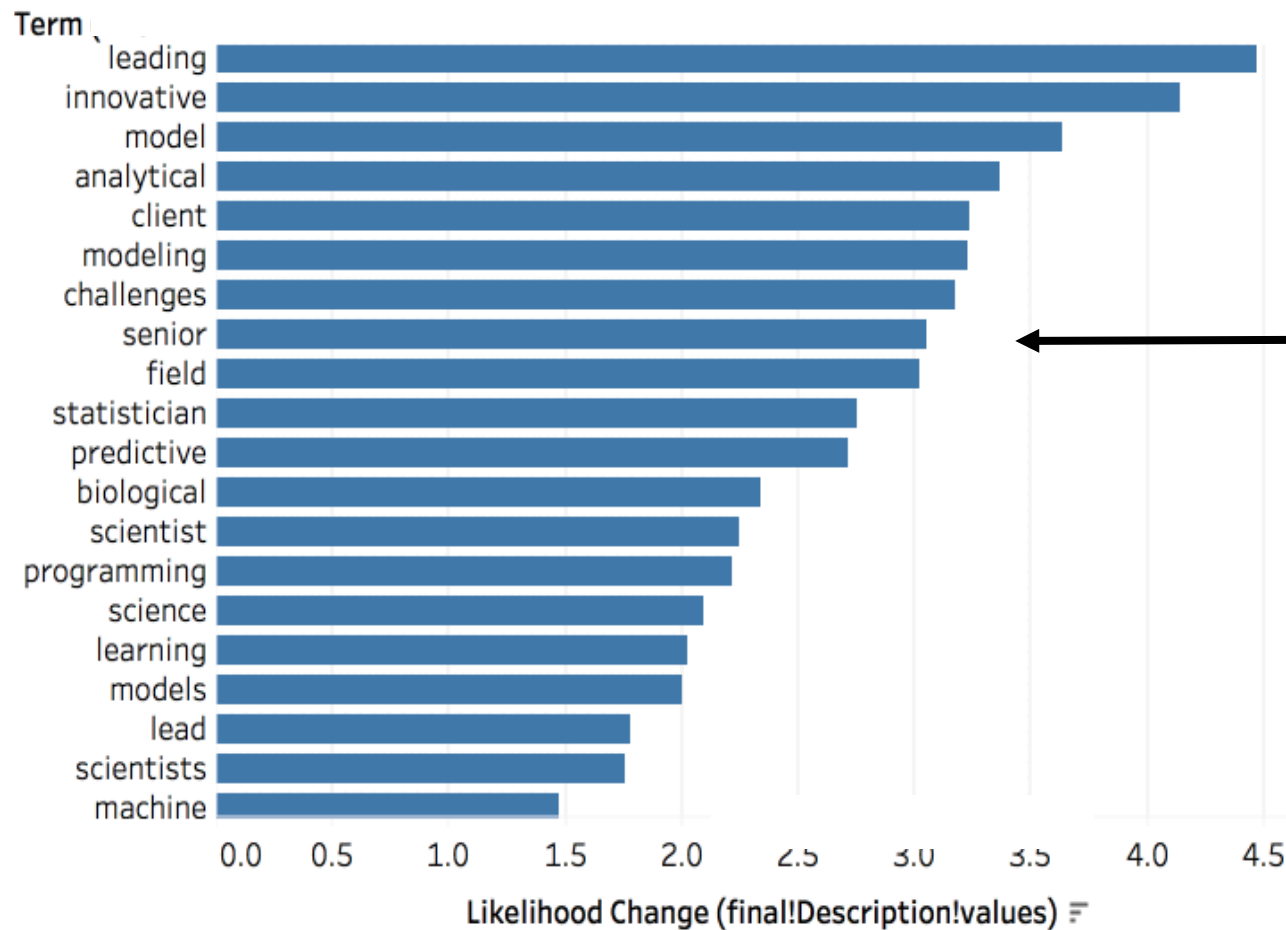## Description Terms vs. Median Salary

Term

Term	Likelihood Change
leading	4.5
innovative	4.1
model	3.6
analytical	3.4
client	3.3
modeling	3.2
challenges	3.2
senior	3.0
field	3.0
statistician	2.7
predictive	2.7
biological	2.3
scientist	2.2
programming	2.2
science	2.1
learning	2.0
models	2.0
lead	1.8
scientists	1.7
machine	1.5

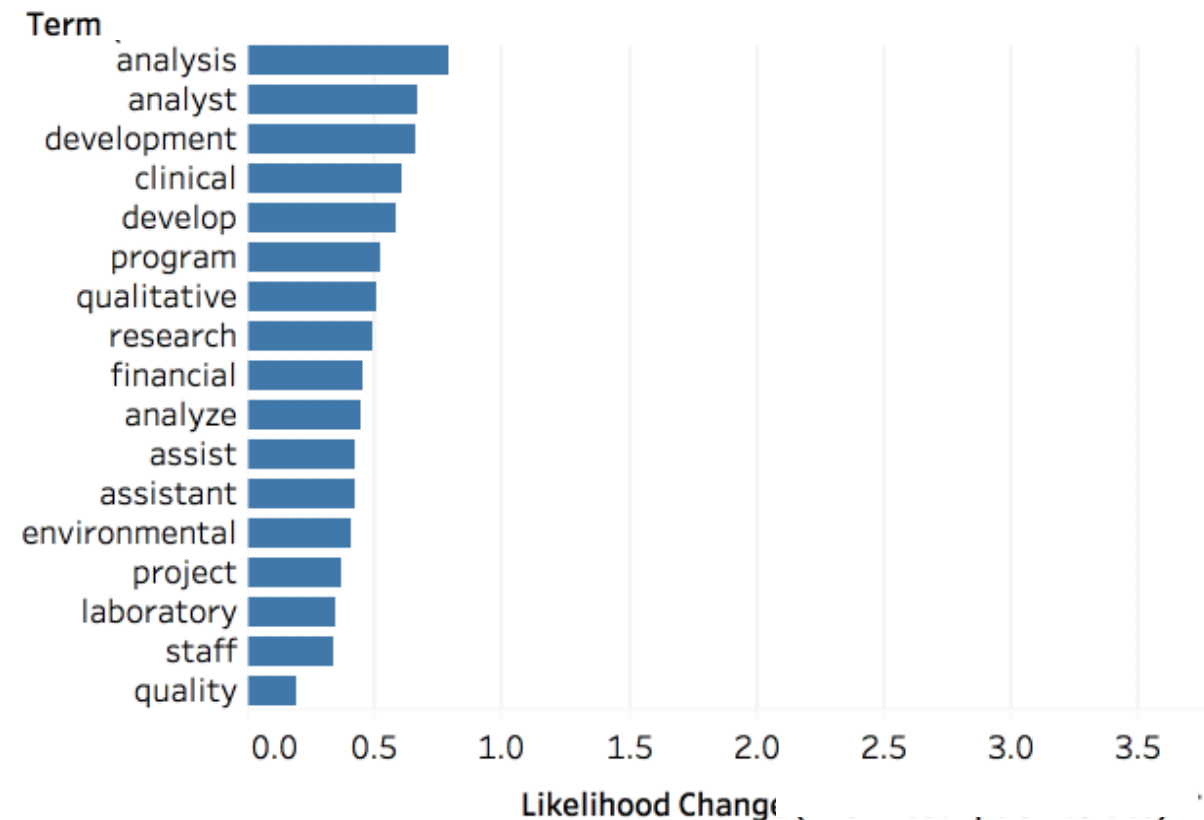Likelihood Change (final!Description!values)

Listing with the word 'senior' in the description is about 3 times as likely to be ABOVE the median salary than those without

Description Words for Lower Salaries

These words in the title <u>decrease</u> the likelihood that salary is above the median:

Description Terms vs. Median Salary

Listing with these words in the description are likely to be BELOW the median salary

Term	Likelihood Change
analysis	
analyst	
development	
clinical	
develop	
program	
qualitative	
research	
financial	
analyze	
assist	
assistant	
environmental	
project	
laboratory	
staff	
quality	

0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5

Summary of terms

- These (types of) terms tend to indicate a salary above the median

Principal	Leading	Quantitative	Machine	Architect	Client
Director	Innovative	Analytical	Learning	Programming	Field
Senior	Challenges	Predictive	Data	Engineer	Sales
Head		Modeling	Scientist		
Lead		Statistician			

- These (types of) terms tend to indicate a salary below the median

Staff	Analyst	Laboratory	Project
Assistant	Analysis	Environment	Program
Associate	Qualitative	Medical	Financial
Specialist	Quality	Clinical	Research
Assist		Bureau	Development

Caveats

- Location by itself is not a good predictor of salary.

- Adding a word to the title or description does <u>not</u> guarantee a higher salary

  - The context of the word needs to be considered
    'Senior' Data Scientist vs. 'Senior' Analyst

  - Models are only 75% - 80% accurate

- Although the Indeed search was done on the term 'Data Science' not all of the jobs were primarily data science

Results and Recommendations

- Whether salary is above or below the median ($95k) can be predicted with 75 – 80% accuracy by
  - Terms in title and description (primarily)
  - Location (partially)

- This information can be useful for
  - Writing the job posting
  - Deciding the posted pay level
  - Determining where to locate a job