Executive Summary

This study investigated salaries of Data Science jobs posted on the Indeed.com, a popular job posting web site. Job postings from 16 metropolitan areas were analyzed to determine what factors most influence posted salaries. Certain words used in the title and descriptions of the postings predict higher and lower than average salary. Also, some metropolitan areas tend to have higher paying positions than others. It is recommended that our company consider this information in writing job postings, when determining competitive offering salaries, and when deciding where to locate positions.

Introduction

A 2012 Harvard Business Review Article listed Data Scientist as 'The Sexiest Job of the 21st Century' (https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century). For our company to stay competitive we must recruit and retain high quality Data Science talent. This study was undertaken to understand Data Science salaries, and factors that impact the salaries.

Methods

This research analyzed job listings from a 'Data Science' search on Indeed.com website. Listings in Raleigh and 15 other cities with a high concentration of technology jobs were pulled. The median salary was determined across all listings where an annual salary was available. Jobs were then categorized into being either above or below the median.

An analysis was done using multiple Data Science tools for natural language processing and classification to determine what factors could predict a listed salary above or below the median. The location of the job, words in the job title, and words in the job description were included in the analysis.

Findings

Location is one factor that helps predicts whether a job listing is above or below the median. Table 1 shows which cities tended to have salaries above, near, or below the median.

Table 1:

Above the median	***Near the median***		***Below the median***
San Francisco	Dallas	Denver	Pittsburgh
Philadelphia	Houston	Los Angeles	Portland
Austin	New York	Phoenix	Miami
Chicago	Raleigh	Seattle	Atlanta

However, location by itself is not a good predictor of the salary group.

Words in the title and/ or description, particularly when coupled with location, can predict whether a salary is above or below the median with 75% to 80% accuracy. Table 2 shows words that are predictive of salaries above or below the median.

Matthew Terribile July 2017

Table 2:

Above the median					
Principal Director Senior Head Lead	Leading Innovative Challenges	Quantitative Analytical Predictive Modeling Statistician	Machine Learning Data Scientist	Architect Programming Engineer	Client Field Sales
Below the median					
Staff Assistant Associate Specialist Assist	Analyst Analysis Qualitative Quality	Laboratory Environment Medical Clinical Bureau	Project Program Financial Research Development		

Discussion / Interpretation of Results

Words that indicate a leadership role or seniority, as would be expected, indicate a higher salary. Certain technical skill areas such as 'modeling', 'statistician', 'architect', and 'engineer' also indicate a greater salary, as do client facing skills ('client', 'field', 'sales'). Although words were most analyzed individually, having the term 'Data Scientist' in the title was found to lead to a higher salary. The terms 'machine' and 'learning' are believed to go together, that is, postings asking for 'machine learning' skills also get higher salaries.

Words that indicate an assisting role or analyst role, and words associated with the medical or clinical area indicate a lower than median salary. Additional terms found to go along with a lower salary were less expected, such as 'research', 'development', and 'financial'.

Although words in the job posting title and description are good indicators of whether the job salary is likely to above and below the median, many of these words can be used in multiple ways ('*Senior* Data Scientist' vs. *'Senior* Analyst'). Therefore, the context of the terms on the entire job posting must be considered.

Conclusion

The use of certain terms in job postings are predictive of the relative salary level vs. the median. The location of the position also has some, although less, predictive value.

Recommendations

These findings can help our company attract and retain the high quality data science talent we need. They should be referenced when:

- Writing job posting – use appropriate words for the type and level of skill recruiting.
- Deciding the posted pay level – the posted pay vs. the median salary should generally correlate with the terms used in the posting, along with the job location.
- Determining where to locate a job – certain locations may offer less expense talent. If these locations are strategic locations for our company, talent might be more economically recruited there.

Matthew Terribile July 2017