



Data Science Salaries

Michael Costa
DSI-5

The background features several abstract network graphs. These graphs consist of nodes (represented by small circles) connected by lines (edges). The nodes and edges are colored in various shades including blue, teal, purple, red, and orange, creating a vibrant, interconnected pattern across the slide.

Objectives & Approach

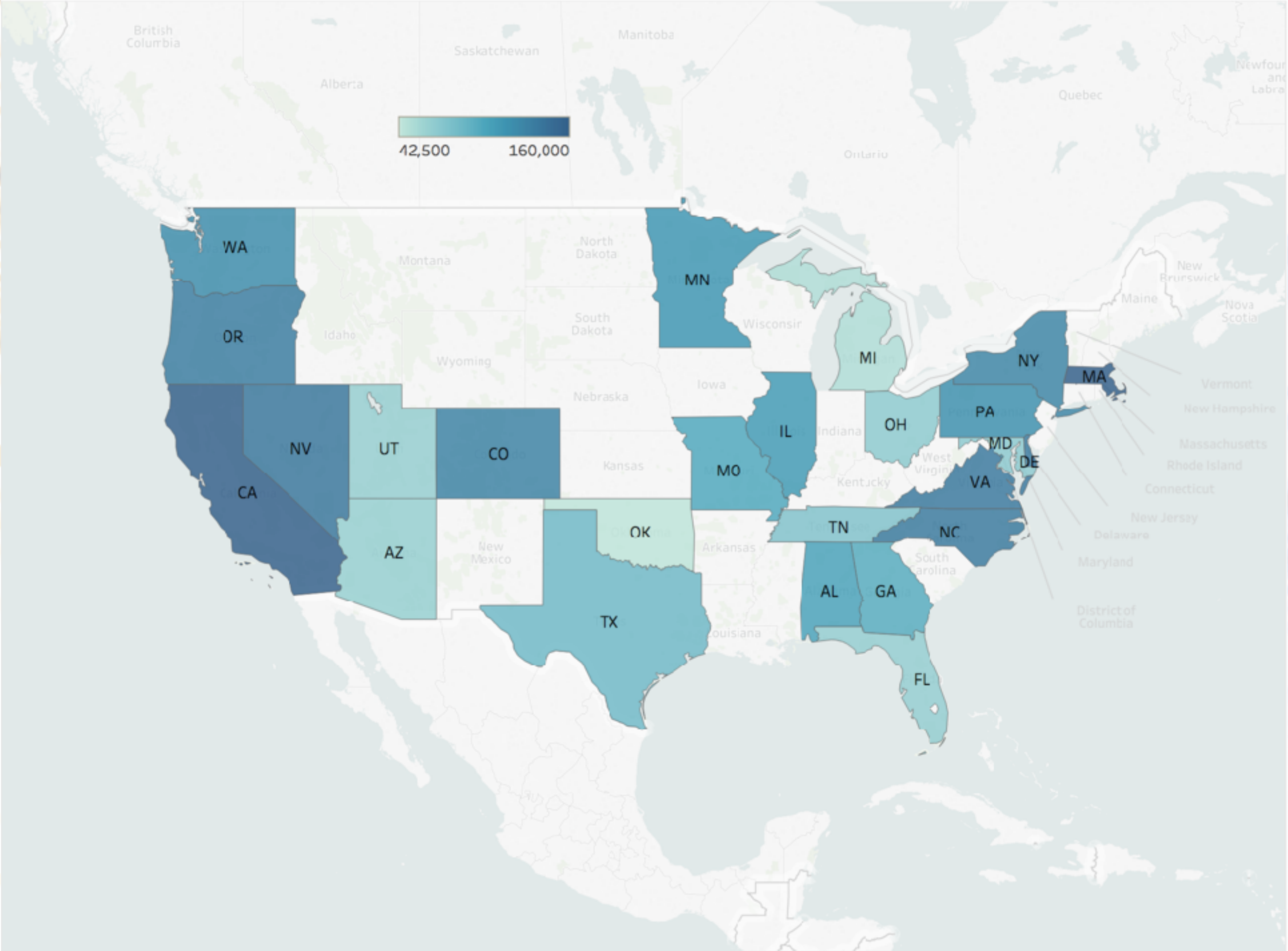
Use data on current job opening to leverage for future hires.

Gather statistics on salaries of data scientists by scrapping [indeed.com](https://www.indeed.com).

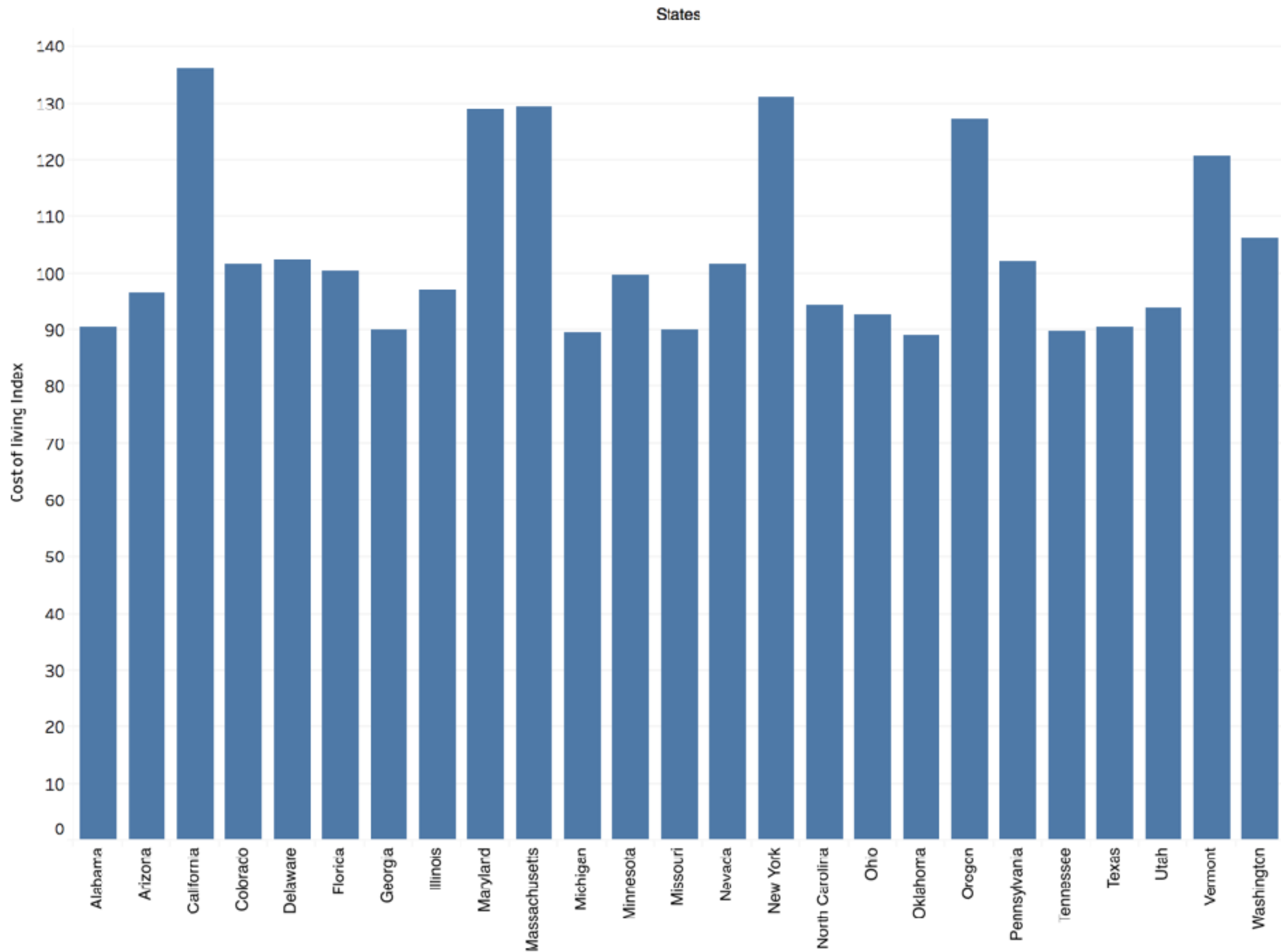
Model to determining location and keywords in job titles.

Collect and model using job descriptions for HR purposes.

State Median Data Science Salary



Cost of living index






Modeling Results

The national median was 125k .

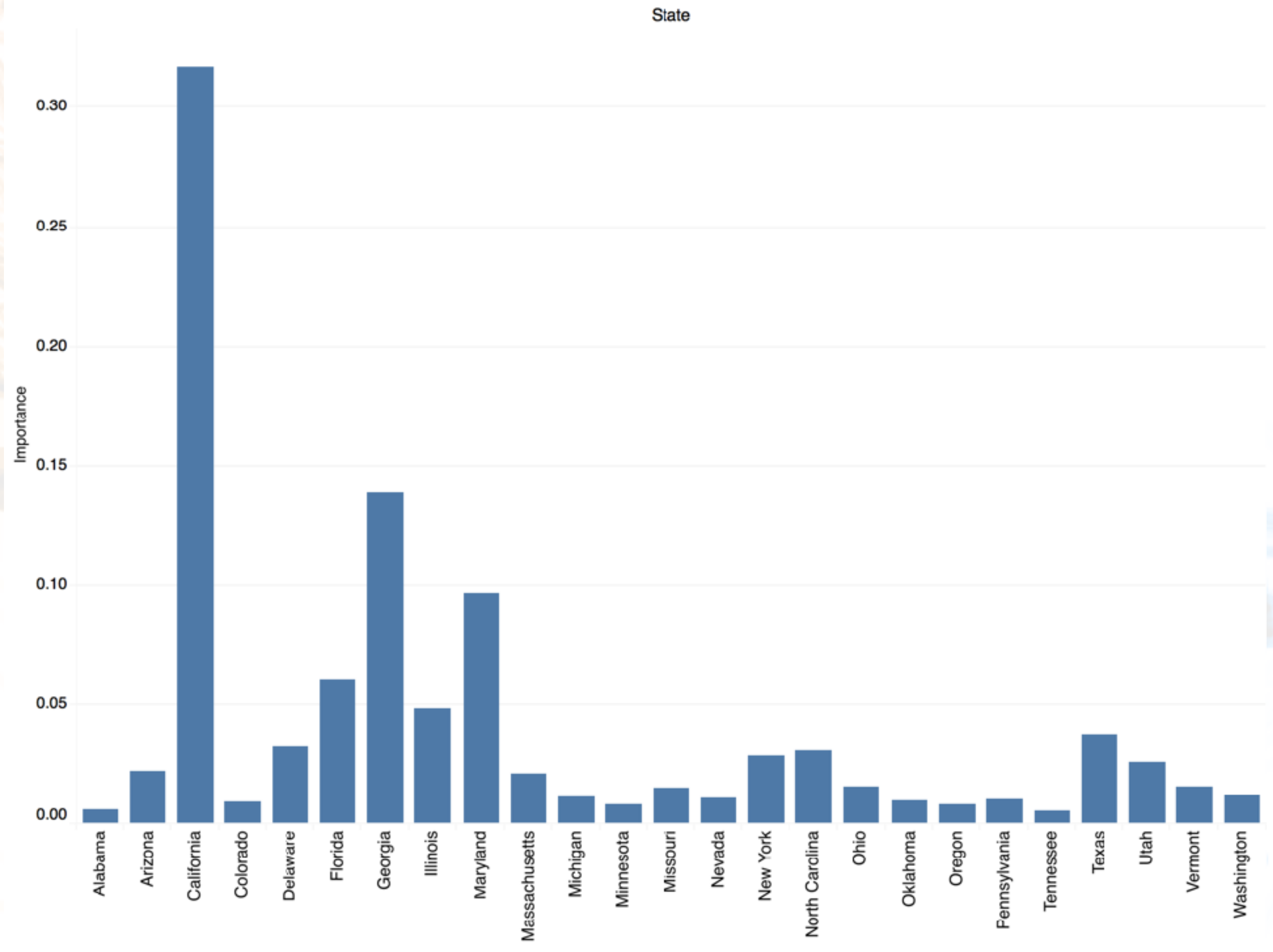
A combination of location and language data yields our best results.

DC areas high cost of living needs to be considered.

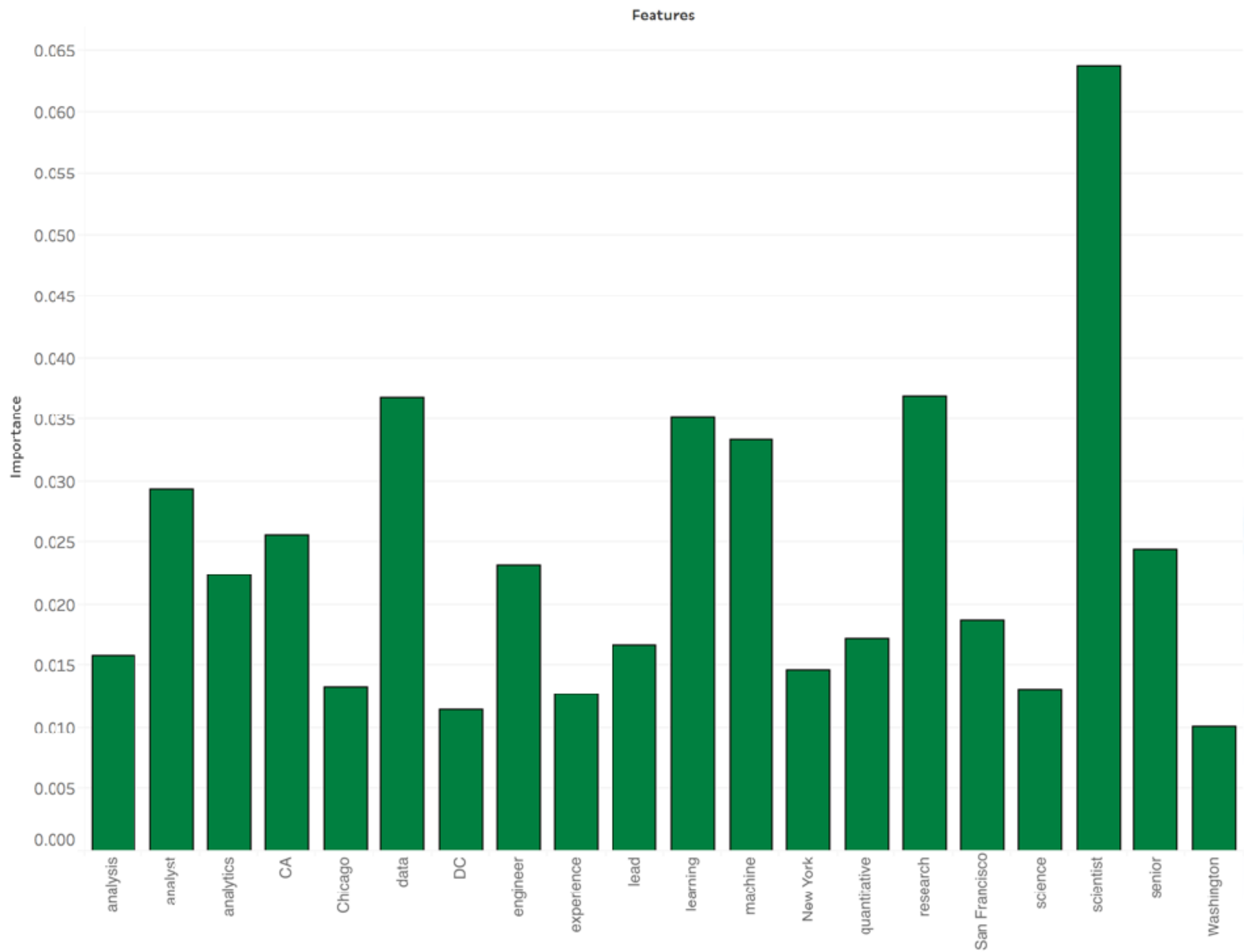


indeed.com did not make this easy. And the lack of salaries presented challenges.

Model Importance by state



Important features





Conclusions & next steps

Machine Learning seems stand out.

Location seemed to be the biggest indicator.

A lot of potential for additional research.

