The Cutting Edge DataFest Write-Up

Main Goal:

The main goal of this analysis was to understand the discrepancy between observed search trends and job posting data from Indeed. Specifically, we found that many job postings originated in the Northern regions of the country, per capita, compared to the vast amount of searches that originated in the South.

(Failed) Attempts at Analysis:

- We originally had hoped to find differences and similarities between the different job postings using the K-Means clustering algorithm
  - We found this to not produce specific results due to the fact that many of the positions titles were specific to their industry
- We also used an LSTM Network to create a prediction forecast of total employment numbers in the US
  - This was not used as it was not relevant to our final presentation

Data Used:

We used the given data from indeed.com to get the city and state of the listings, the date they were posted, and the industry the posted job was it. Externally, we used data from the US Bureau of Labor Statistics, the 2010 US Census, and Google Trends.

Data Analysis:

- Determined Indeed's job listing throughout regions of the US
- Normalized to average the population throughout each state to show job listings per capita
- States with the most searches for "Indeed" is observed to have lowest job listings per capita
- Employment data from Bureau of Labour Statistics was used to make predictions about the industries the job searchers were coming from
- Trends for those industries were extrapolated to find potential increases in user base for Indeed

Strategies:

Industries that are falling could result in the unemployment of millions. We could sponsor companies and organizations that provide industry switching education programs. These programs will efficiently allow individuals to experience new opportunities quickly new positions towards employment.