Team: 404 Found
Weijia Bao, Xinyi Cheng,Jessica Feng, Keting Yang

Slide 1: Insights for employees
How we processed Data:
1. Filter the data with the condition of "Job age day" equals to 1, to avoid overcounting.
2. Extract top 5 Categories (Industry) and admin1(states)
3. Process the correlation map of industry, states, and education level.

Conclusion:
- Jobs on Indeed.com normally does not require high school degree. Top 5 states have significant more more job supply A universal finding over the five states are most people working in Healthcare/medical.
- A student interested in healthcare could go any of the five states for higher job opportunities. CA has largest supply of jobs, especially for IT. But not Food Industry.
○ Food: FL, TX ○ Health: NY, CA ○ Others: CA, FL○ Retail: NY, PA ○ IT: CA. TX

Slide 2 +3: Insights for employers
How we processed Data:
1. Filter the data with the condition of "Job age day" equals to 1, to avoid overcounting.
2. Exclude all samples with the average rating equals to 0(non-rated company)
3. Extract the clicks, average rating and description length columns
4. Normalized the data to the same scale by subtracting the mean and dividing the max.
5. Feed our data to a neural network with one hidden layer of three nodes.

Conclusion:
The formula we found for predicting the click number with average rating and description length is clicks = 0.09-0.005*average rating-0.04*description length. This indicates very weak linear correlation between the number of clicks and the average rating and description length.

With the scatter plot, we can see companies with 30-40 ratings tend to have the highest number of clicks, but so do the companies with extreme rating as 10 and 50. The reason that there's an decreasing trend for the rating between 40-50 might be the people who is rating the companies tend to be kind of "extreme" when they "like" the company and give 50 in general.

Smaller firms tend to post more jobs and also get more clicks. As the size of the company increase, they tend to post less jobs, but this trend cease when company size is larger than 1000+. We graphed total count clicks by company scale and job post, and the two graphs are similar. Thus we plot another graph of the ratio between them, which is clicks/(job*day). Averagely there are 75 clicks per job post per day, and this data is same for companies ranging from 0-49 employees to 1000+ employees. Therefore we conclude that the total number of clicks does not depend on size of the company.

Slide 4: Insights for Indeed
1. The employment distribution for each industry from Indeed is consistent with the one from U.S. Bureau of Labor. This implies that the posted jobs on Indeed.com are pretty diverse and cover all the popular industries.
2. We suggest Indeed.com to specify the industries on posted jobs from staffing firms. This will allow Indeed.com to have a better understanding of how the major industries of employment are distributed.
3. Indeed.com can provide our findings on rating and description length to employers when they post a job on Indeed.com.
4. Indeed.com can use that currently there are 75 clicks/job*day to promote sponsorship.