Data science is a high-demand, high-compensation industry that many people want to enter. To do this, more and more people are taking data science boot camps such as the Data Incubator to develop the technical skills and professional network to find their first data scientist position. However, critical questions are not satisfactorily answered by the marketing materials of these companies. Are there data science positions in my area? How long did it take graduates to find a job? What type of compensation were those positions? The risks of expensive classes, large time commitments, and vague job placement details make many potential students wary of enrolling in boot camps.

Additionally, the incoming skill level of students vary widely. The best boot camp for a former software engineer with years of programming experience and a degree in mathematics is quite different than a business analyst whose primary analysis tool is Windows Excel. Improper matching of the student’s needs with the education can lead to great dissatisfaction, especially when some boot camps cost tens of thousands of dollars to attend for a review of concepts the student understands well.

I want to create an app that answer some of these questions about job placement for potential boot camp students. By using the LinkedIn API to scrap information about former students of the most popular boot camps, I can construct a tool to find people with similar backgrounds to the applicant and display

a) the times to first positions,

b) companies that hired them (including geographic and industry distribution), and

c) estimated salaries (by scraping Glassdoor salaries for the labeled positions where applicable and industry averages in the locality for the others).

An initial analysis of ten of the most popular of data science boot camps using LinkedIn shows that approximately 26,000 profiles exist for alumni of these boot camps. Many of these positions teach both software engineering and data science skills, so I limited the search to profiles with position titles including the term “data” (see Figure 1). Heavy emphasis is now being placed on online visibility to find new positions, so this data set likely represents a good sampling of people who were trying to find positions. This curation focus will also improve the quality of the dataset, allowing variables such as school ranking and location, major, number of graduate degrees, number of previous positions, and years of work experience to be extracted. By finding the nearest neighbors to the potential applicant in career space, the experiences of the most similar candidates can inform the user of what outcomes look like for various boot camps.

This preliminary search also showed that some data science bootcamps hire a large number of their own graduates (see Figure 2). This data set does not incorporate every former student of these companies and there are numerous reasons that students may go to work for their former instructors, but the understandable concern for job placement rates among boot camps makes this information interesting and important for boot camp shoppers.

An outline of the project development is presented here:

1. Construct a script to gather data from the LinkedIn API about boot camp alumni   
   (currently under construction)
2. Explore the data set to see the data quality of each variable and perform feature selection
3. Create scripts to gather supplementary data from additional sources such as Glassdoor and US News & World Report to create salary estimates and school rankings
4. Combine all the information into a model finding the most similar profiles from appropriate dimensional weighting of a combination of numerical and categorical variables
5. Build an application which takes the user’s LinkedIn Profile or resume, then displays the *x* most similar profiles, including links to allow people to message them on LinkedIn in case they have questions