For this milestone, we collected, cleaned, and formatted data from several sources. A custom scrapping tool was built in Ruby using Nokogiri for this milestone. The first source was QuakerNet which is the internal Penn tool to connect with alumni. This data was pulled from over a dozen industry searches and data was scrapped from the HTML source.

The second source was PennLink jobs posted. Over 1000 jobs were scrapped and using the same custom scrapping tool and relevant ones were added to the database. The third source was an open data CSV of potential employers. In all of these datasets, but especially the ones scrapped from the web, the majority of time was spent on cleaning and trimming the data; i.e. some job names contained the application due date in the position title and was removed.

DB Connection

password: tahmidisabitch

The schema we used is the default one after connecting; i.e. do not have to run the command 'alter schema'

connection string:

```
sqlplus
'cis450project@(DESCRIPTION=(ADDRESS=(PROTOCOL=TCP)(HOST=cis450project.creb8qtnn
bvb.us-west-2.rds.amazonaws.com)(PORT=1521))(CONNECT_DATA=(SID=DBPROJ)))';
```

SQL Files are in Github Repo