December 21, 2019

Team:

Daniel Mendoza

Chris Lachman

**ETL Project – Project Report**

**Etract:**

For our ETL project we looked data that delineates employment numbers by whether individuals went on to pursue graduate school. We used *data.world*  as our data source. Using datasets from FiveThirtyEight – “College Majors” (<https://data.world/fivethirtyeight/college-majors>), we pulled two data sets (“all-ages.csv” and “grad-students”) that looked into graduate student data: majors and employment data.

**Transform:**

Once the two data sets were extracted, we cleaned the data. For each data set, we created data frames. We cleaned our columns by extracting columns we did not need. Next, we merged both data sets and created .csv files that broke down our employment data by major category.

**Final:**

We loaded up our data into postgres via sqlalchemy. We chose sql because we were more comfortable.