October 17, 2019

Data Visualization and Analysis

ETL Project Summary

Jennifer Lawless, Jie Bai, Mingming Chen

Data Science Job Market in Austin

For our ETL project, we wanted to look at the job market for data scientists, specifically, the different companies, salaries, and skills desired. We found some datasets on both Kaggle.com and Data.World, then we wanted to look at currently available jobs on indeed.com and the most common skills required.

We cleaned the data by removing unnecessary columns, such as reviews, and filtered by the location, so that we only looked at Austin. To look at the companies with the most jobs, we counted the unique companies, with the University of Texas at Austin, being the employer with the most data science jobs currently available.

We also wanted to look at only jobs with the title of Data Scientist, so that we could look at the most desired skills, from all data sets. In this section, we removed all columns except for the job title, skill, and company and found that for data scientists in general, SQL, Python, SAP, Machine learning, R, and SAS are the most frequently desired skills.

We loaded the data into postgres and set up tables to view the data and to be able to query the dataset.

We had some issues loading the data into postgres and searching for data.