Daniel Kang

Data Analytics Bootcamp

ETL Project

6 March, 2019

Extract:

* CSV file was downloaded from Kaggle.com
* CSV files were downloaded from U.S. Department of Education database

Transform:

Kaggle file

* Dataset was basically clean and not much formatting was done
* Reset the index to a column

U.S. Dept of Edu files

* Had to manually clean and re-aggregate the CSV files
* Had to format the values in the cells so that MySQL would be able to read the data
* Multiple tables were stored in a single CSV file so I had to re-aggregate the unique tables into their own CSV files

Load:

* The data was loaded into MySQL from Jupyter
* The chosen data, when utilized, should be able to help answer what indicators could best predict future student success, whether it be gender, financial background, or whether they enrolled in test prep courses.