ETL Final Report

**Extract:**

(your original data sources and how the data was formatted (CSV, JSON, MySQL, etc)

We began the ETL project by getting CSV data from Kaggle. We decided to analyze Space X launch activities. We were able to get 3 datasets that we wanted to extract, transform, and load. We began by using Jupyter notebook. We used pandas as well as sqlalchemy. We created a data frame after uploading the data to read the data and all of the columns.

**Transform:**

(what data cleaning or transformation was required)

The column names in the three sets of data did not exactly match. We needed to transform all of the data sets with the same names for the column titles so that we would be able to combine the different rows into the same column. After we renamed and synchronized the column names in each file, we were able to run each file separately in Jupyter; the “data\_min\_vers.csv”, “databasespacex.csv”, and “Updated\_spacex\_2006\_to\_present.csv”. This is displayed in the uploaded Jupyter file notebook “data\_proj2.ipynb”

**Load:**

(the final database, tables/collections, and why this was chosen)

We used MySQL workbench to do a Union code to combine the result from multiple SELECT statements into a single result set. With the Union code we were able to take into account the value retrieved by all of the SELECT statements even if the data types of the corresponding columns did not match. Our final csv file, “sqloutput\_1.csv”, combines the datasets and shows how the Union SQL code transformed and added the rows to the columns and combined the different CSV files into one.

Please upload the report to Github and submit a link to Bootcampspot