**ETL Project**

By: Todd Schanzlin, John Swierczynski, & Erin Lee

**Extraction:**

We used Kaggle to extract four different csv files. Three of our data sets are based on space launches/missions. The last csv file was a dataset of astronauts. The following files is the datasets we used from Kaggle.

* Astronauts.csv
* Space\_walks.csv
* Space\_missions.csv
* Global\_space\_launches.csv

**Transformation:**

After we extracted our csv files, we used pandas to read the files into DataFrames (Figure 1). From there we scanned the DataFrames to see what needed to be cleaned. To clean our data, we first dropped unwanted columns (Figure 2). In our datasets, we had different formats for dates (figure 3). We used the datetime function so that the date column in each DataFrame would be in the same format (Figure 4). We finished cleaning our data by renaming some columns.

Figure 1:



Figure 2:

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generatedFigure 3:

Figure 4:

A screenshot of a social media post

Description automatically generated

**Load:**

Our final step was to get our cleaned data into a Database. To do this we created an engine to connect to postgres. We used pandas\_to\_sql formula to transfer our DataFrames into a database.

**Summary:**