Step 2. ETL Setup

Initiates **logging** info to keep track of whole ETL pipeline process and capture errors

Extraction: The objective of the extraction function is to take data from different csv files and return two data frames.

- Each file path connects to an individual csv file
- Logs successful data extraction
- Returns two separate data frames

Transforming: Transforming the data will:

- Clean: remove missing values, uniform character casing, and renaming column names to avoid duplicates
- Filter: filters rows based on certain conditions, merge data on a common column
- Structure: Create a calculated column to that shows an aggregation of new info from the merged data
- Output a properly structured table for analysis

Loading: This function will store the transformed data into MySQL, which is best for structured data like what our group is using

- Connect to sqlit3 and load the data into a sql database
- Since we are uploading our transformed data into Google Cloud, we must consider what database is best to store our data. BigQuery, which we used in previous assignments, is best for analytical workloads in large datasets, cohesive with our transformed data. It also has high scalability for adjusting our queries and integration.

Run and Execute ETL