Pokemon ETL Project

By: Angel Milla, Francis Marquez, Adam Burstyn, & Sabrina Karam

In the Extract element we used two different sources, a CSV from Kaggle.com [<https://www.kaggle.com/abcsds/pokemon>] and a JSON from a public API [<https://rapidapi.com/Chewett/api/pokemon-go1?endpoint=apiendpoint_e0cea6ae-58e1-4b27-85ed-756803df433e>]. These sources were used to find the names, types, and stats from each of the main series Pokemon featured both in regular Pokemon and Pokemon GO.

In the Transform element, we used both Pandas and Python to help clean up the data. In the CSV file, we did the following: used .group\_by function to get the types of different Pokemon, added a none category for Pokemon who didn’t have a secondary type, renamed the column names so that the tables would match in the database, and renamed the index to ‘type\_ID’. In the JSON file, we did the following: dropped the columns that were not necessary, in the Pokemon GO we dropped all types of forms except Normal Type, created a function to both split the column types into main and sub as well as checking if any new types had been added, merged the tables, dropped multiple copies of the pokemon “Arceus’ from the final table, and checked the see how the new data was comparable in the two DataFrames.

In the Load element, we took the data and put in into pgAdmin and created a database. We decided to use relational databases because we are able to create tables by merging and discovered interesting information from these merges that would not have occurred if we didn’t use relational databases. The next page has the two tables created from using pgAdmin and the merging of the databases.

 