Sarah Johnson, Yongsen Mercier, Paolo Roque

Team Chimchar: ETL Report

**What is our Project?**

We wanted to find data to help in our quest to find the strongest Pokémon battle team. In order to do this, we needed to find out the different stats of the first three generations of Pokémon in the video games. Hopefully using this information we can dominate in the competitive Pokémon scene.

**Extract:**

The original data source came from a CSV file in Kaggle titled The Complete Pokémon Dataset and an API called the PokéAPI. We used API calls to get the different Pokémon types to add to the CSV.

**Transform:**

We cleaned the CSV by removing unnecessary columns such as Japanese name, egg statistics since we are using full grown Pokémon, and the unorganized type information. By rearranging the columns to set the index to the Pokédex number and putting the name in the first column, the dataframe became a lot easier to use. We then added the type information from the API to the dataframe before saving it as a csv and moving it to MongoDb.

**Load:**

We chose MongoDb to display the data because certain Pokémon have multiple types and it makes more sense to have an unstructured document for each Pokémon because not all categories may apply to them. The database was called pokemon\_database with the collection being named pokemon\_collection.