# ETL Project: Historical Military Data

By Tara and Marvin

# Extracting: Finding Data

- Found Historical Military Battle Data using Kaggle
- A previous user compiled a comprehensive data set with battle information - specifically, battle name, war name, location, actor names, winners, and weather conditions
- Each battle had a battle ID (ISQNO) that corresponded across csv files
- We used 3 csv files - battles.csv, battle_actors.csv, and weather.csv

# Transform: Cleaning Data

- Imported each csv file into a Jupyter Notebook
- Dropped unnecessary or duplicate rows
- Renamed all columns
- Reshaped battle_actors file from long to wide so data would be unique by ISQNO (battle ID)
    - Data originally contained one row per battle per actor
- Transformed NaN data to integers
- Changed ISQNO to a float so we could merge on this piece of common information
- Saved to a CSV that we could upload to postgres

# Load: Creating Database

- Set ISQNO as primary key and created table
- Every battle has at least 2 actors but can have up to 4 - thus, there are inevitably some null values
    - Tell sql to default to ‘null’ in the case that there are less than 4 actors in the table using the command DEFAULT ‘NULL’
- Manually import table

# Next Steps: Further Explorations

- Examine the relationship between weather conditions and battle wins
  - Did attackers win more often when the weather was good?
  - Are there certain weather conditions better for battles?
  - Weather conditions can elongate battles. Did this affect ultimate results of wars?
- Expand and research whether terrain can affect battle wins