

ELT Project Report: Super Bowl & US Presidential Election Results

Participants: Tiffany McGowan, Corey Perkins, Mallory Kuba

DATA SETS:

- Super Bowl finals from years 1967 to 2020 from Kaggle:
<https://www.kaggle.com/timoboz/superbowl-history-1967-2020>
- US presidential election results from years 1797 - 2009:
<https://gist.github.com/namuol/2657233> from Lou Acresti's Github.

EXTRACT: (1 hour)

- Download Kaggle dataset and GitHub dataset as CSVs.
- Validate both datasets.
- *Challenge:* Load CSVs into Postgres using PG Admin's import tool.
- CSVs will be imported to pandas using `from_csv` with pandas.

TRANSFORM: (1.5 hours)

- Once imported, verify data integrity and create aggregates as necessary.
- Drop irrelevant columns in each dataset.
- Rename appropriate columns for readability.
 - Extract the year from (full) date columns. Create new column for "year". Complete for both data sets.
- Once imported, join the tables based on year. Note: A US presidential election is every four years, while a Super Bowl is every year. US Presidential elections began in 1797 and Super Bowl elections began in 1967.

LOAD: (1 hour)

- While extracting/loading, read through each dataset and create a Postgres table to load into.
- Extract appropriate columns for each dataset—only include relevant columns in analysis.
- Write the portion of the python script that loads the data into the Postgres database.

FURTHER EXPLORATION:

- Do election results have an impact on super bowl winners/losers? (and vice versa)