## Project Proposal

Combine Superbowl information into one database, including: game, halftime show, and advertisement data from 1967 to 2020.

## Data Cleanup & Analysis

\* Sources of Data:

* Super Bowl History: <https://www.kaggle.com/timoboz/superbowl-history-1967-2020>
* Super Bowl Game & Halftime Data: <https://www.kaggle.com/amithasanshuvo/tv-halftime-shows-and-the-big-game-dataset?select=tv.csv>
* Super Bowl Ads: <https://www.kaggle.com/prondeau/superbowlads>

\* The type of transformation needed for this data:

Cleaning – standardizing column names

Joining – Primary Key: Super Bowl year

\* The type of final production database to load the data into:

Relational

\* The final tables or collections that will be used in the production database.

* halftime\_musicians
* super\_bowls
* tv
* name\_mvp
* superbowl-ads

Steps Required to Reproduce ETL Process:

1. Download CSV files from Kaggle
2. Load into a dataframe using Pandas in Jupyter
3. Create database Postgres connection

## Project Report

\* \*\*E\*\*xtract: your original data sources and how the data was formatted (CSV, JSON, pgAdmin 4, etc).

Data Sources:

* Super Bowl History (CSV): <https://www.kaggle.com/timoboz/superbowl-history-1967-2020>
* Super Bowl Game & Halftime Data(CSV): <https://www.kaggle.com/amithasanshuvo/tv-halftime-shows-and-the-big-game-dataset?select=tv.csv>
* Super Bowl Ads(CSV): <https://www.kaggle.com/prondeau/superbowlads>

Format:

* Download CSV files from Kaggle
* Load into a dataframe using Pandas in Jupyter
* Create database Postgres connection

\* \*\*T\*\*ransform: what data cleaning or transformation was required.

Limit and rename columns for data standardization; identify primary keys

\* \*\*L\*\*oad:

Final database: Super\_Bowl\_Analysis

Tables/collections: ERD

* halftime\_musicians
* super\_bowls
* tv
* name\_mvp
* superbowl-ads

Why this was chosen: Relevant/timely as Superbowl on 2/7

GitHub: Super Bowl Analysis