**ETL Project Proposal** [**cwru-cle-data-pt-07-2020-u-c**](https://cw.bootcampcontent.com/CW-Coding-Bootcamp/cwru-cle-data-pt-07-2020-u-c/tree/master)

Group 3 ETL Members: Paul Han, Jon Zagorski, Tana Zimmer, Kimberly Gore

**Does higher home attendance lead to a better record for the season? Do we think a higher record results in higher home attendance?**

We are proposing scraping the PFR website for team attendance and also scraping the ESPN website for win/loss data for each team. We are hoping to be able to develop a flask app on AWS for holding all the data.

To do this we will:

1. Extract the Data

Scrape the ESPN and PFR websites for the data. Jon has already tried a pandas pd.table\_read of PFR to show it is possible. Scrape for 1999 to 2019 seasons.

1. Transform the Data

To transform the data we will need to clean the data and drop any n/a values and organize the data so that it can be imported into postgres in a concise manner.

1. Load the Data

To load the data, we will be building a Flask app to host the database and give it a splash page hosting some predictions on the current NFL season.

**Source examples:**

<https://www.pro-football-reference.com/years/2019/attendance.htm>

<https://www.espn.com/nfl/standings/_/season/2019>

We know this is an ambitious project for 1 week but we think that we can make a functional page to hold all of the data and our predictions.