

Are the Highest Paid Soccer Players and their Teams the Best?

Sophia Weidner

Bellevue University

DSC 540: Data Preparation

Professor Catie Williams

September 25th, 2022

Are the Highest Paid Soccer Players and their Teams the Best?

- API: <https://www.thesportsdb.com/api.php?ref=apilist.fun>
 - **Description:** This API can look up players, teams, leagues, league standings, etc across 300+ international and club soccer leagues.
 - **Criteria:** player name, team name, event name, list all leagues, list all countries, player former teams, player contracts, etc.

- CSV:

<https://www.kaggle.com/datasets/martj42/international-football-results-from-1872-to-2017>
 - **Description:** This dataset includes 50,000+ results of international soccer matches from the first official match in 1972 all the way up to 2019. International soccer matches are different from club matches; the players play on their country's team in this scenario. Some of these games are from the FIFA World Cup, whereas some are just friendly games. These results are just for **men's** soccer.
 - **Criteria:** date, home_team, away_team, home_score, away_score, tournament, city, country, neutral
 - Additional resources: <https://data.world/datasets/soccer>

- Website:

https://en.wikipedia.org/wiki/List_of_most_expensive_association_football_transfers

- **Description:** This website includes tables relating to the most expensive club player transfers ever, as well as the most expensive club player transfer per year, and player cumulative transfers. These transfers are all under UEFA's jurisdiction and the purchasing clubs are from England, France, Italy and Spain.
- **Criteria:** rank, player, from, to, position, fee, year

My goal with these datasets is to compare different aspects of soccer to see if the best players are the ones who are paid the most, and, in conjunction, if the teams that spend the most on their players are the best teams. The API that I found for this project gives me access to players and teams across several club teams as well as international teams, so I am able to see the success of players on teams who are paid to be there (club) versus who represent their country without a contract (international). The dataset I found has results from 1972 to 2019 of various international competition results for different tournaments; the FIFA World Cup is not the only place to pull data from. Lastly, the website I found is a Wikipedia article that contains data pertaining to the most expensive club player transfers and other fun numbers relating to player transfers.

The three sources work very closely together, and I believe they will show interesting trends amongst successful players, clubs, and international teams. The API and CSV both share criteria of country and events, so I am able to look up specific tournaments and the national teams that played in those events. The API and Website will work best together by searching for players; the website table is small and only shows one bit of information about a player, but the API has more detail. For example, if I look up a player, it shows all the teams they've played for,

their statistics on those teams, their overall statistics, and even the amount of awards that they've won.

Overall, between my three sources, I think I will be able to create solid comparisons between the most successful teams and players. Not only will I be able to look at data where club teams are able to purchase and loan players, but I will be able to study teams where players cannot be traded, and the talent of their team is based on pure luck/whether or not the "top" players are born in that country or not. I think one of my challenges may lay in the use of the API. The API has so much data stored in it, I will need to make sure that what I am looking up is concise and correctly matches what both the CSV file and Website tables contain. The CSV file and Website are both smaller snippets of time, whereas the API spans years and even generations of football. I am mostly interested in modern-day soccer, so anything from the year 2000 and on is my primary focus. I have also struggled in the past with cleaning my data sets, and the CSV file I've chosen is a bit clunky. I'm thinking I may want to remove some of the data, like the scores, and create some new columns that store data based on the total of wins a country has. That way, my CSV becomes a bit more "graph-able".