**Team Members:** Michael Cruz, Yafreisi Gonzales de Basilio, Collins Mokaya, Aracely Vivanco

**Background:**

[EA Sports FIFA](https://www.ea.com/games/fifa/fifa-20) is the best-selling sports video game franchise in the world. It’s recognized globally for it’s realistic soccer gameplay.

In FIFA, Player Attributes determine the quality and the features of a player's technical skills, behaviours and performance on the pitch. The combination of attributes help mimic player’s real world action and provide FIFA with it’s realistic gameplay. Each player has a collection of attributes rated from 0 to 99 each. The average of a player's attributes sums up their OVR (Overall) Rating.

Player Attributes are rated by EA’s team of coaches, professional scouts and season ticket holders from around the world.

There are over 35 player attributes which range from: Acceleration, Pace, Composure, Strength, Dribbling, etc..

**Player Attributes**

QUALITY RATING RANGE

Excellent 90 to 99

Very Good 80 to 89

Good 70 to 79

Fair 50 to 69

Poor 40 to 49

Very Poor 0 to 39

**Abstract:**

Can FIFA’s attribute data, help determine a player’s success in all different phases of the game?

Could a combination of FIFA’s Player Attributes help predict player performance in various leagues and tournaments around the world?

**Data Sources:**

1. Complete list of FIFA 20 Player attributes can be found on Kaggle

<https://www.kaggle.com/stefanoleone992/fifa-20-complete-player-dataset#players_20.csv>

1. For the purposes of this project, actual player performance was scraped from [https://fbref.com](https://fbref.com/en/), a child site of SportsReference.com, focused exclusively on Soccer.

**Extract**

* Uploaded Kaggle Fifa dataset (CSV) using pandas into two dataframes.
* Only focused on Premier league players from [https://fbref.com](https://fbref.com/en/)
  + Webscraped all premier league team URLS from the first table by using BeautifulSoup and requests from

<https://fbref.com/en/comps/9/Premier-League-Stats>

* + Iterated through all Premier league team URLs to get players performance data from the first table by using panda’s .read\_html

**Transform**

* Fifa dataset
  + Split into two dataframes
  + Renamed and dropped unnecessary columns.
* [https://fbref.com](https://fbref.com/en/)
  + Renamed and dropped unnecessary columns.
  + Needed to create an ID to be able to match back to fifa set, decided to use the sofifa\_id from out fifa dataset
  + Used long\_name from both dataset and merged them together, however we had some NAs
  + Used fuzz from the fuzzywuzzy package to fuzzy match all the NAs to the sofifa\_ID
    - We were able to match 88% of the NAs to sofifa\_ids
  + Assigned sofifa\_ids to the fbref dataframe

**Load**

* Loaded our three tables (players, player\_skill and player\_perf) into postgress using sqlalchemy