

A decorative graphic on the left side of the slide consisting of two overlapping parallelograms. The front one is blue and the back one is a light green color. They are positioned diagonally, with the blue one in front of the green one.

StrikerRank

Fisher Marks and Thomas Orozco



Project Overview

- Expected Goals (xG): probability shot will go in based on various factors (%)
 - Real Goals (rG): actual goals (0 or 1)
 - $rG - xG$: one useful way of measuring how good a striker is
-
- Market Value: estimated transfer fee a team would have to pay a player's team to acquire the player



Data Sources

- FBRef: a good source for raw soccer stats
- TransferMarkt: a good source for player market values

FBRef Scraper: <https://github.com/fisherm123/soccerdata>

TransferMarkt Scraper:

<https://github.com/fisherm123/transfermarkt-scraper>



Tech Stack

- Scrapers
 - Python
- Database
 - PostgreSQL
 - pgAdmin4 (easy for importing large csv files)
- Web App
 - Flask
 - Psycopg2 (connects to PostgreSQL servers)
 - HTML/CSS/JavaScript

GitHub: <https://github.com/fisherm123/strikerrank>

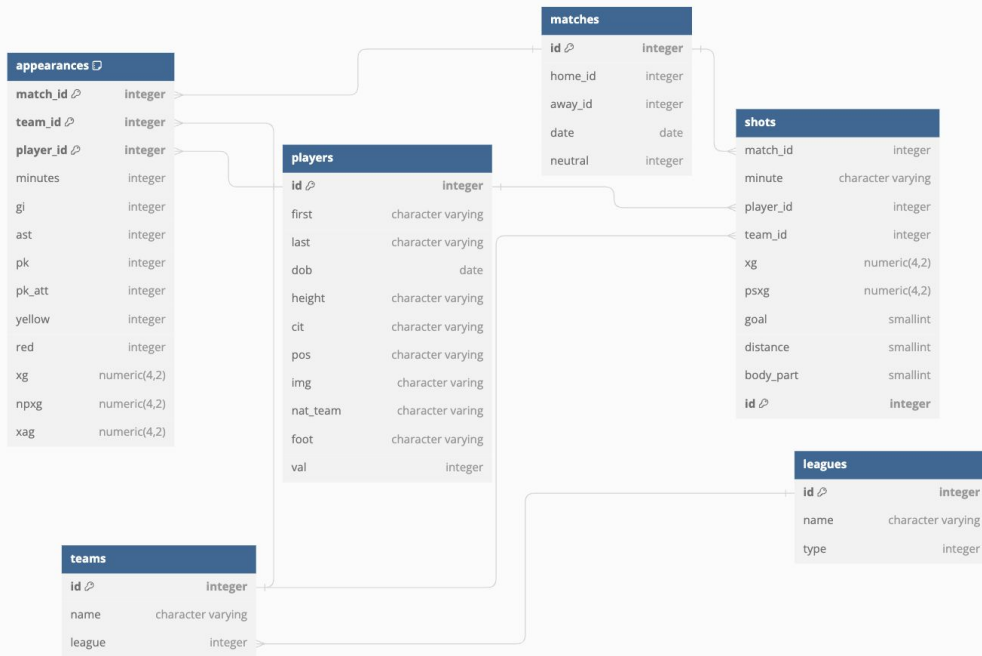
Database Design



Schema

Key Design Choices:

- No match score
 - Smaller files
 - Longer compute
 - Player-focused analysis
- No team attribute for players
 - Transfers mid-season
- Player image is saved as a link





Main Features

- Ranked list of players by our key metric alongside market value
- Dynamic filter (options adjust automatically if more data is scraped)
- Increase query size
- Player-specific info menu



Good Decisions

- Scraping our own data
 - After initial time investment can get whatever data we needed
 - Expansion of project as simple as rerunning scraper
- Multi-tier database design



Bad Decisions

- Scraping our own data
 - High initial time investment
- Proper Data Cleaning
 - Frequently failed quality checks
 - Harder to find problem than solve it from beginning