Progress Report

# Title

The Player Performance Lab

By: Daniel Tshiani

# Topic and Data

It looks like we will not be using API soccer, nor the other data sources that were listed in the initial project plan. I will stick will soley data from the newly released R package “itscalledsoccer”. I have also decided to limit the data to the 2024 season and I have decided that I will disaggreagate the app into different player tabs (offensive players, midfield players, defensive players, goalkeepers). The data has been collected, cleaned, and is ready for analysis. There are limitations to the data and there were similar limiations on all other data sources as well which were missing salary data as we begin to steer away from the MLS.

# Ethics

I am still not using PPI data and the documentation does not infer there is a malusage issue in the way I am building my app.

# Layout

I will disaggregate the player tab to Attacking players, Midfielders, Defensive players, and Goalkeepers. I only do the attacking players tab however, I think including the other tabs and briefly talking about them is important to empahise the idea that the model for player salary differs based on position. I will keep the other tab but probably put a pop up or something like that when click and pretend like a membership is needed for the other tabs. I will take out the teams tab though. I will also not give the option to analyze by league because a lot of data on salary is missing for leagues outside the MLS.

In the offensive tab, I will let users pick 2 variables to analyze with the recommendation of starting with goals and minutes played. I will also allow users to search based on player’s name and the search results will give the players current salary and the predictive players salary so users can see the difference and find players who are underpaid for their value. I am contemplating excluding players whos we are missing wage data on or setting their wage and assuming its zero, which represent a youth contract.

I decided not to overcomplicate the statistical model. In the paper Football players’ performance and market value, their model adjusted R-squared is about 37%. My adjusted R-squared is low also. I know this means the model is poor, but more than one semester needs to be dedicated to running various statistical tests to achieve the best model possible, and this focus is more for data science rather than stat.

# Outline

Update on demonstration outline: I have decided to create my own app and do it single-handedly. Because I am working independently with a shorter time-span to complete the project, I will aim for a 10 presentation rather than a 20 min presentation. The major risk to completion is working against time and lack of data for leagues outside the MLS. Mitigation measures is soley focusing on the attack and limiting the data to those who salary are available. The cut-off is may 6th, so everthing will be submitting at 11pm that day regarless of status of completion.