

A4: Project Proposal

Tharun Sikhinam

- **Motivation/problem statement:**

The purpose of this project is to perform an Exploratory Data Analysis on the Transfer Market of the Big Five European Soccer Leagues. This analysis is particularly interesting with the large sums of foreign investment and the increase in the number of sports agents in recent years. With each clubs bringing in vast amounts of money through TV rights and other sources, it would be interesting to know how the clubs have put this money to use. A detailed analysis of the economics of the football transfer market can help player scouts and clubs make informed decisions.

Furthermore, combining the transfer market data with player ratings will help us understand what factors affect the value of a player in the transfer market. It would be interesting to identify player trends and make future predictions if possible.

- **Data used:**

1) The data used for this project is obtained from the following Github Repository.

<https://github.com/ewenme/transfers>

The author of the GitHub repo scraped the data from <https://www.transfermarkt.us/> and details the process in this [article](#)

The scraping was done in accordance with the Terms of Use in this link:

<https://www.transfermarkt.co.uk/intern/anb>

After reading the Terms of Use, it clearly states that

4.8. With the transmission of Content, Transfermarkt grants you the irrevocable, temporally and spatially unlimited and transferable right to reproduce, distribute, publish, issue, make publicly accessible, modify, translate and store the Content. This includes the right to edit, design, adapt to the file formats required for use, and to change and / or improve the presentation quality.

This dataset consists of all transfers in the European football market from 1990-2019. The data consists of a player name, club-out, club-in along with the transfer value grouped by different leagues. The dataset also consists of free transfers and loans as well. Most of the data is updated and contributed by the users of the website, and there might be few factual inaccuracies in the transfer figures stated.

2) The other dataset we will be using as part of this analysis is the FIFA player rating dataset obtained from this kaggle link <https://www.kaggle.com/karangadiya/fifa19>

Although the Kaggle link says the data is licensed under [CC BY-NC-SA 4.0](#), the data is actually obtained by scraping sofifa.com. I wasn't able to find any terms of use or a license agreement associated with the website. Until the license/terms of use are clearly understood

I don't plan to use the data. I have contacted the website to inquire more about this issue and am waiting to hear back from them.

This dataset holds important information about the player such as Player Country, position, player attributes such as Sprint speed, defense, attack, shooting etc., These attributes can be combined with the transfer market dataset to understand the types of players that fetch the maximum value.

- **Unknowns and dependencies:**

I haven't obtained the player ratings dataset that is licensed to be used for academic purposes yet. I would have to search for alternative sources of data or not perform the second part of the analysis as part of this project.

I might potentially encounter another problem when trying to join the two datasets. Since each player doesn't have a unique identifier, the joins should be done on first name, last name basis which might not be the best way to join the two datasets.