BUILDING A TEAM EFFICIENTLY

AN INSIGHT INTO THE TOO-MUCH TALENT EFFECT

DATA VISUALIZATION 2021 PROCESS BOOK

CADILLON Alexandre
SABAA Karim
SHABAAN Abed Alrahman

Contents

INTRODUCTION

A major survey conducted by FIFA in 2006 revealed that approximately 270 million people around the world are actively involved in the game. According to the leading sports analysis and statistics organization, Nielsen Sports, 46% of the world's population is interested in football.

So, the beautiful game trumps all others in terms of reach and popularity. It is therefore fascinating to think that there was a time when soccer did not exist at all and that it had to develop and grow like any other sport. Today, it has become an ultra-professionalized sport, with training centers from a very early age and in which being successful can change the life of a player, a city and even a country. Yet it still maintains the spirit of being a sport that is accessible to everyone and that can be shared with anyone.

With the amount of resources devoted to this game, the question naturally arises as to what extent it is an efficient sport. An analogy could be made with financial markets where developed economies are expected to be more refined and present less unpredictability.

Motivation

We came across a study on the relationship between talent and team performance. It found that people believe there is a linear and nearly monotonic relationship between talent and performance: participants expected that more talent improves performance and that this relationship never turns negative. Three studies revealed that the *too-much-talent effect* emerged when team members were interdependent (football and basketball) but not independent (baseball).

Through this representation, we want to illustrate this phenomenon in the context of European football, which is undoubtedly the most developed set of leagues both within football and among all team sports.

What is the relationship between the results of a team and the quality of its players? Is it possible to follow the progression of a team as it signs better players? Does this phenomenon have the same properties depending on how developed a league is? Our visualization tool will assist the user in answering these types of questions and help them to validate their hypotheses.

EXPLORATORY DATA ANALYSIS

Dataset

We decided to use a dataset from Kaggle as suggested. The dataset is called European Soccer Database. It contains data about teams, players and matches from 11 European leagues. There are +10,000 players and +25,000 matches from 2008 to 2016.

Match data contains team names, the date and the result. Additionally, team line ups, detailed match events (goal types, possession, corner, cross, fouls, cards, etc.) and multiple betting odds are available for most games.

Players and teams' attributes are sourced from EA Sports' FIFA video game series.

The dataset is fairly clean but there is some work to be done in the match table. When available, match events need to be parsed to identify meaningful stats like goals or cards and link them to particular players.

Initial EDA

Exploratory data analysis confirmed that the dataset contained much more information than the one we needed to test the hypothesis. We divided the exploration in several directions: games, teams, leagues, individual players.

The goal of our project guided us as to what meaningful connections to make between tables. The main challenge was to connect the player's attributes and the match tables so that the team talent could somehow be measured.

Additionally, we explored team matchups. This wasn't directly related to the *too-much-talent effect,* but it could lead to interesting visualization. Similarly, we also compared leagues between them, computing statistics such as goals scored or goal difference which could be a measure of a league's level of inequality.

DESIGN

Since the beginning, we were clear as to what our visualization should look like. The article that we based our project on referred to a common belief that a linear relationship existed between a team's success and its players' talent, so we knew we wanted the user to explore the relationship between those two metrics. The better suited kind of visualization for this is a scatter plot since linear relationships are easily identifiable.

We spent quite some time trying to find statistics that would measure talent and performance.

Performance

Match results can be seen from slightly different perspectives. We can think of it as a trinary variable (win, loss, draw), which is what we did at first, but the problem is that it is hard to extract overall performance from three variables. If the result was binary, it would have been easier.

This led to our second approach which was based on points achieved (using the modern rule of 3/1/0 points for win/draw/loss). From that we could measure the total number of points and normalize by the total maximal number of points.

We realized that leagues and seasons were hard to compare between them. If teams aren't facing each other, having more points wouldn't necessarily imply having better performance. Therefore we decided to group data by league and season so that we could always separate teams using those attributed, if needed.

Finally, we decided to include team's rank as another measure of a team's performance. This is a discrete variable as opposed to the point percentage and is more disconnected to the actual number of won/lost/drew games, but it also deserved to be studied since, in the end, that is what actual football is about. We noticed that not all leagues had the same size, so we decided to include a rank percentile variable as well, which in practice normalizes the rank by the number of teams in a league.

Talent

There are other attributes that could be used, more focused on certain facets of the game but we kept the overall rating as our main measure of a player's talent.

The match table contains information on team lineups. That data is not always complete, but it is what we decided to use to compute a team's talent. We thought of different approaches.

The first idea was to create a set of all the players that had participated during a season with a team. From there, we could compute the average overall rating. This option didn't weigh in the degree of participation from all the players. A well-known, constant and regular playing player had the same weight as an unknown young player from the club's training center that only played one game during the season, so we had to reassess how to take into account a player's regularity on the field.

For that, we decided to compute the team average overall rating in each match that all 11 players would be assigned the same weight at that point. Then, at a later stage we could compute an

average over every single game in a season. Even certain lineup info is missing in some games in a season, all the available games will be assigned equal weight. This approach was the better suited on for comparing teams between different leagues since we compute a mean of player rating per game so that the number of games played in a season is normalized.

Getting back to the *too-much-talent effect*

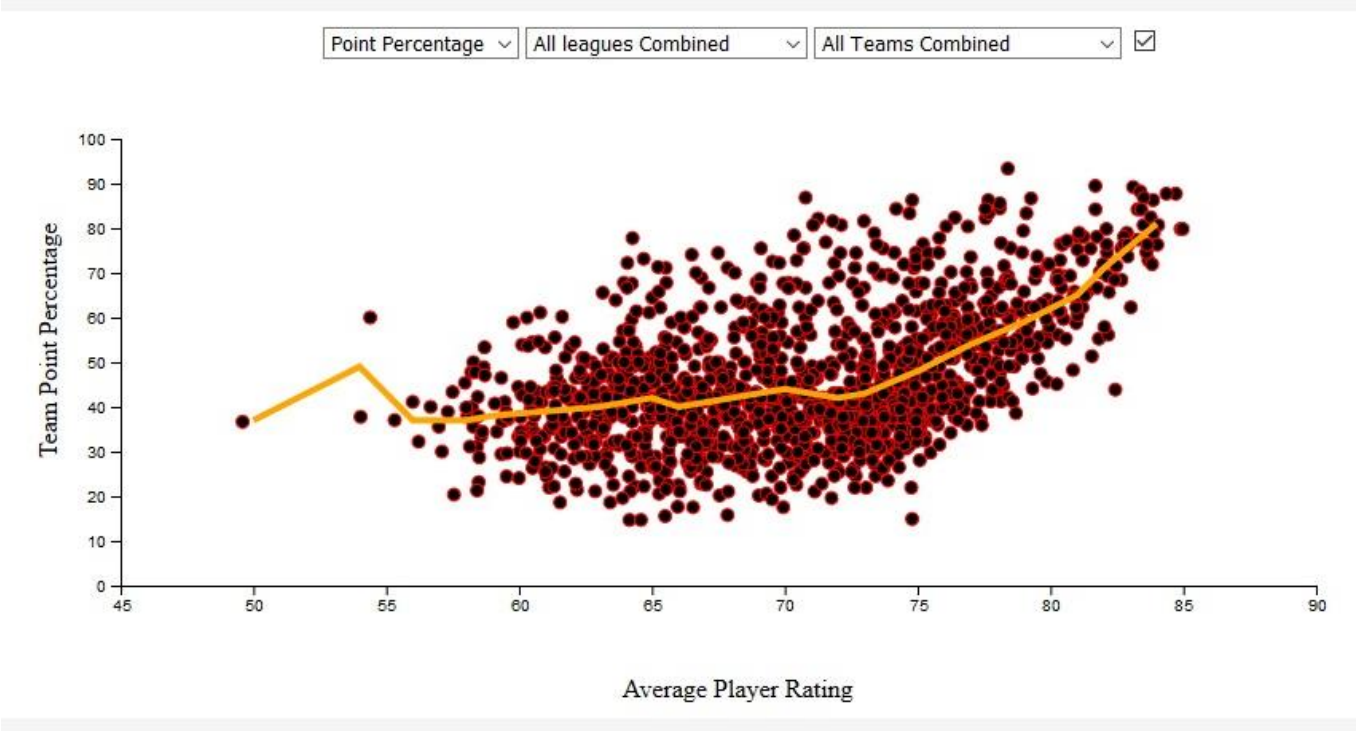
Finally we decided to guide the user in its conclusions, so we included a simple fitting curve so that the reader can make sense of the data ever if there are too many points clustered together..

We used a simple bin-averaging curve to represent the trend of the cloud of points.

Final visualization

We included some drop-down lists that allow the user to select what he wants to visualize, mainly leagues and teams. Due to the same number of seasons, the team scatter plot is lacking some points, especially if the team wasn't regularly playing in division 1.

We also included the possibility to hover over a certain point to get some additional stats on a season's team performance, mainly wins, losses and draws.



PEER ASSESSMENT

Abed and Alexandre shared most of the work in generating the data used and any eventual preprocessing. This part had to be revisited along the project since the original metric computed sometimes proved to be faulty in some aspects.

Abed and Karim worked together to do most of the work in building our website.