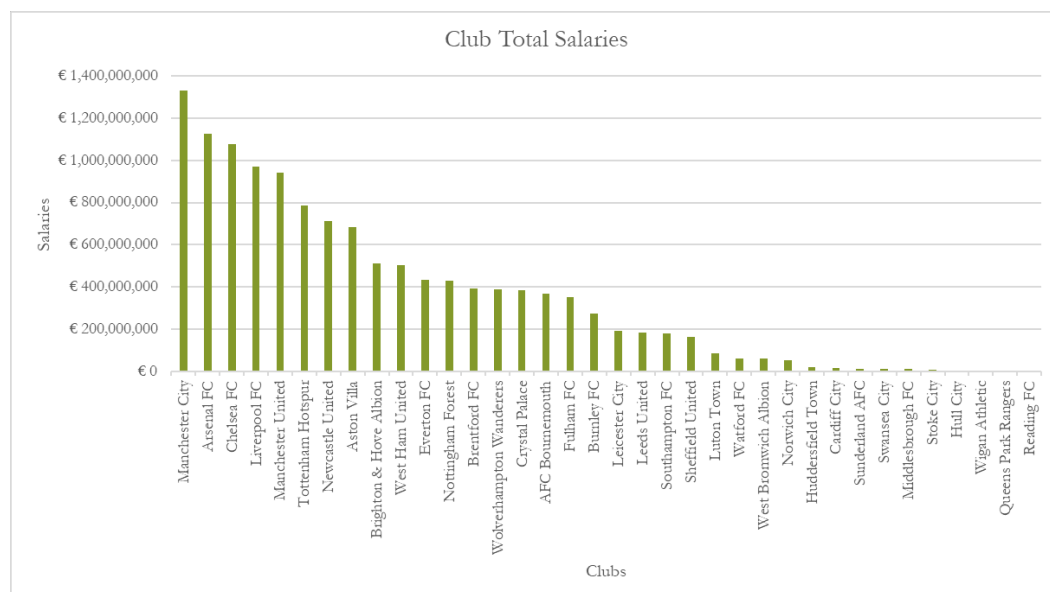
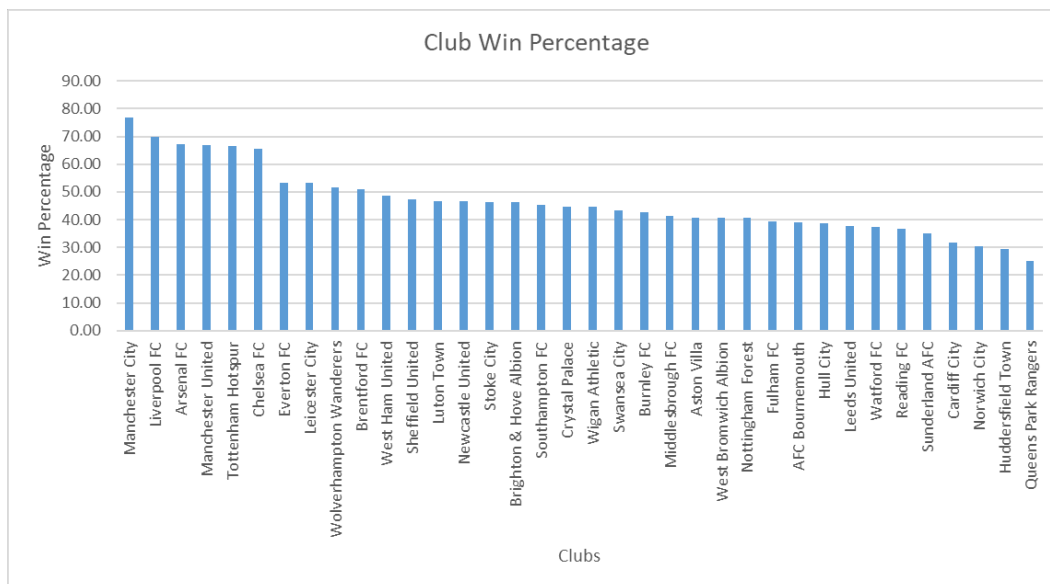


Road to Winning in English Premier League Clubs
By Amado Uyehara

Executive Summary:

The English Premier League has undergone revolutionary changes over the past few decades. These changes have had profound effects on the clubs with changes in their ownership and their recent success on the pitch. On August 4, 2008, Sheikh Mansour's Abu Dhabi United Group purchased the majority of Manchester City Football Club, which would revolutionize the EPL for the foreseeable future. On the final day of the 2011 season, Manchester City FC played against Queens Park Rangers and defeated them 3-2 to win their first English Premier League Title in club history. Under new wealthy ownership, Man City was able to pay a higher amount of salaries to acquire the talent to win a title. This trend of owners overspending has led to the creation of top club giants in the EPL resulting in the higher the salaries a club is willing to pay the higher the win percentage. In this paper, I will be exploring how English Premier League teams climb the success ladder, our research looks at the player's salaries and how that results in a club's success.



Major Findings:

- The English Premier League is the highest-level soccer division in England and one of the most competitive leagues in the world
- The top 3 teams with the highest win percentage are Manchester City, Liverpool FC, and Arsenal FC
- The top 3 teams with the highest salary payroll are Manchester City, Arsenal FC, and Chelsea FC
- The EPL has 94 high-earning players with English Citizenship, 25 with Brazilian Citizenship, 21 with French Citizenship, and 17 with Dutch, Portuguese, and Spanish Citizenship
- The top 3 teams with the most amount of players with salaries above €10,000,000 are Chelsea FC with 27, Manchester United with 25, and Liverpool FC with 24
- Teams with lower win percentages are usually relegated from the league, so there are more than 20 teams as the data is across multiple years
- The top 3 teams with the players they paid the most for are Manchester City, Manchester United, and Liverpool FC
- The teams with above a 50% win percentage are Arsenal FC, Brentford FC, Chelsea FC, Everton FC, Leicester City, Liverpool FC, Manchester City, Manchester United, Tottenham Hotspur, and Wolverhampton Wanderers

Main Conclusions:

There is a significant correlation between the English Premier League Clubs' win percentage and total salaries. The initial hypothesis was that the more a team spends on acquiring top-tier talent, the more they win. Using the example of Manchester City is showing a small sample size. Many other teams have been doing this for much longer and also teams who are just joining the trend. Chelsea FC, Manchester United, Liverpool FC, and Arsenal FC have all been more successful and are among the top giants of the EPL. Manchester City has now surpassed them and has become the best team in the EPL.

With some of my findings, some lower-paying teams find success. For smaller clubs aiming high in the EPL standings, adopting a "Moneyball" approach involves making strategic and cost-effective player decisions based on data and analytics. By identifying undervalued talents and utilizing advanced statistics, these clubs can punch above their financial weight and compete with more resourceful opponents. This also comes with the help of a competent coach/manager who has something to prove. Since this is a player-only dataset, coaches aren't included so there wasn't any data to analyze. Among the teams that found more success than the bigger spending clubs were Brentford FC, Everton FC, Leicester City, and Wolverhampton Wanderers. Of all these teams only one has recently won an EPL title, Leicester City. Since winning that title, many of the players left the club to seek higher-paying contracts. Since Leicester City is a smaller club, they didn't have enough money to retain their talent from their Cinderella run. In the 2022-2023 season, Leicester City was relegated and currently is playing in

the English Football League Championship which is the second-highest division in England. To maintain success in the English Premier League, the hypothesis was correct the more money you spend the more games you will win.

Backstory:

The English Premier League (EPL) was established in 1992, the league emerged as a breakaway from the Football League, driven by a desire to enhance commercial viability and global appeal. Over the years, the EPL has transformed into a football powerhouse, captivating audiences worldwide with its high-level competition and star-studded lineups. The league's format, featuring 20 clubs, has become a sporting spectacle, attracting the best talents from around the globe. From the intense rivalries of Manchester United and Arsenal to the Cinderella stories of underdog teams, the EPL's history is rich and filled with drama. As it continues to be a dynamic force in the sport, the English Premier League's journey reflects the ever-changing nature of the game.

The day that Manchester City was bought by Sheikh Mansour's Abu Dhabi United Group changed the way the EPL has operated ever since. Many large clubs have seen the success of having extremely rich owners, such as Paris Saint-Germain (PSG). They were already the top team in their respective league Ligue 1, but they wanted to seek more and win a UEFA Champions League Trophy. As for the EPL, many teams have switched owners and tactics to get to the top of the league. Clubs such as Chelsea FC and Newcastle FC have been bought out by very wealthy owners who are willing to pour their money into player salaries to win a title. Other major clubs haven't been bought out by new owners, however, they are facing a large amount of pressure from fans to spend more money on top-tier talent. This would evenly disperse the talent rather than 2-3 teams getting all of it. Manchester United has been facing growing pressure amongst its fans to sell the team to someone who will spend large amounts of money for talent like Manchester City, Arsenal FC, and Chelsea FC.

Today, Manchester City FC finished their extremely successful 2022-2023 season winning the treble, which means they won the three biggest trophies an English team could win. They had won the English Premier League, FA Cup, and UEFA Champions League. Their immense success comes with the fact that they acquired Erling Haaland for a record fee of \$114 million. He was everything they needed to finally win the treble. Their salary had increased much higher than any other teams in the EPL, but they only did it knowing they would win.

Research Process:

Define:

- What variables are important to clubs winning more games?

Collect:

- Where can I find data that both displays player salaries and their club's success?
- Are there any data quality issues?

Organize:

- How will the data properly be prepared for analysis?

Visualize:

- Creating a dashboard to see the variables affecting the clubs in real-time.

Analyze:

- Histogram Charts to appropriately show the distribution of the data.

Define:

My data warehouse will demonstrate that the English Premier Clubs that spend more money on player salaries, see more success in the league. The data is from a transfermarkt database. This database houses all of the transfer data for every league in Europe. It includes a lot of information from multiple international competitions and the exact values that players make and that clubs pay. I got the CSV file datasets from a Kaggle competition [Football Data from Transfermarkt](#). There are nine datasets of which I used only 4. Most of the other files are used to connect the leagues and use international competition stats. I did not use the other 5 as they weren't important in what I wanted to analyze. Looking more into the facttransfermarkt data warehouse and the variables I chose to analyze and what they represent:

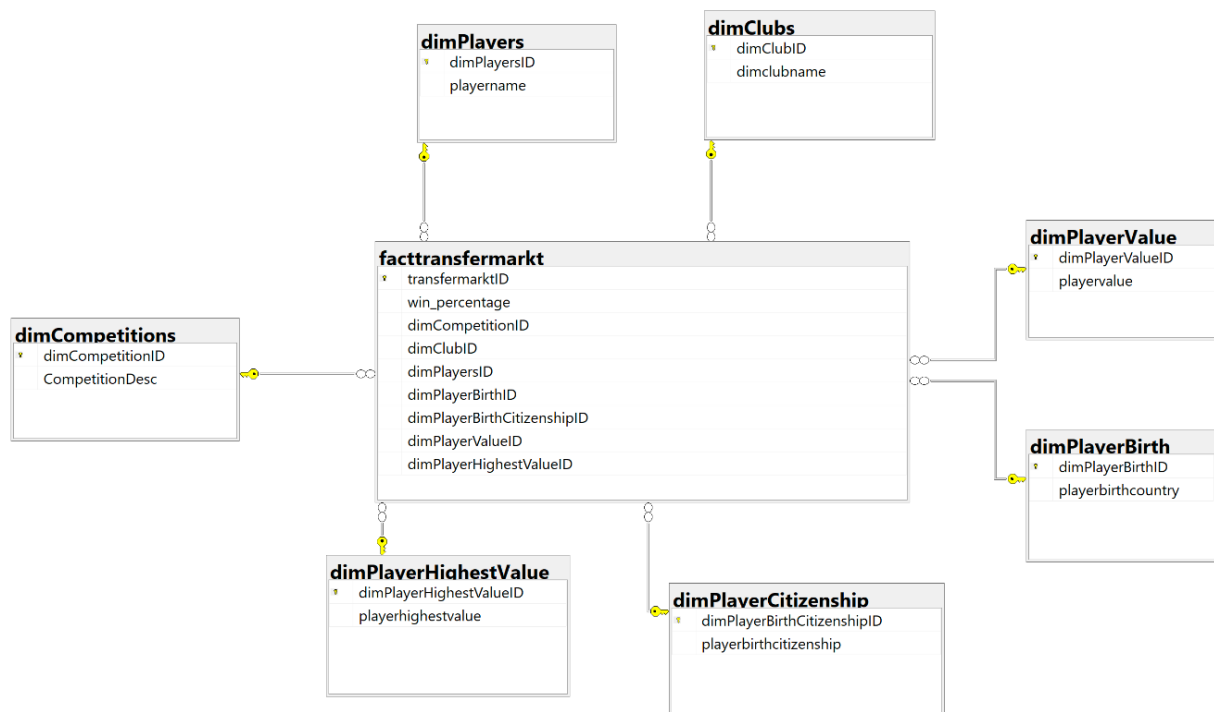
- transfermarktID: Unique identifier for each row in the dataset primary key of the table
- win_percentage: float related to the team that each player is on and how many games they win out of the total
- dimCompetitionID: 1 - premier league
- dimClubID: unique identifier for each team in the data warehouse (24 teams)
- dimPlayerID: unique identifier for each player in the data warehouse (357 players)
- dimPlayerBirthID: unique identifier for which country a player was born in (56 countries)
- dimPlayerBirthCitizenshipID: unique identifier for which country is a citizen of (52 countries)
- dimPlayerValueID: unique identifier for which contract value a player is (35 values)
- dimPlayerHighestValueID: unique identifier for the highest value a player has had on their contract (40 values)

Collect:

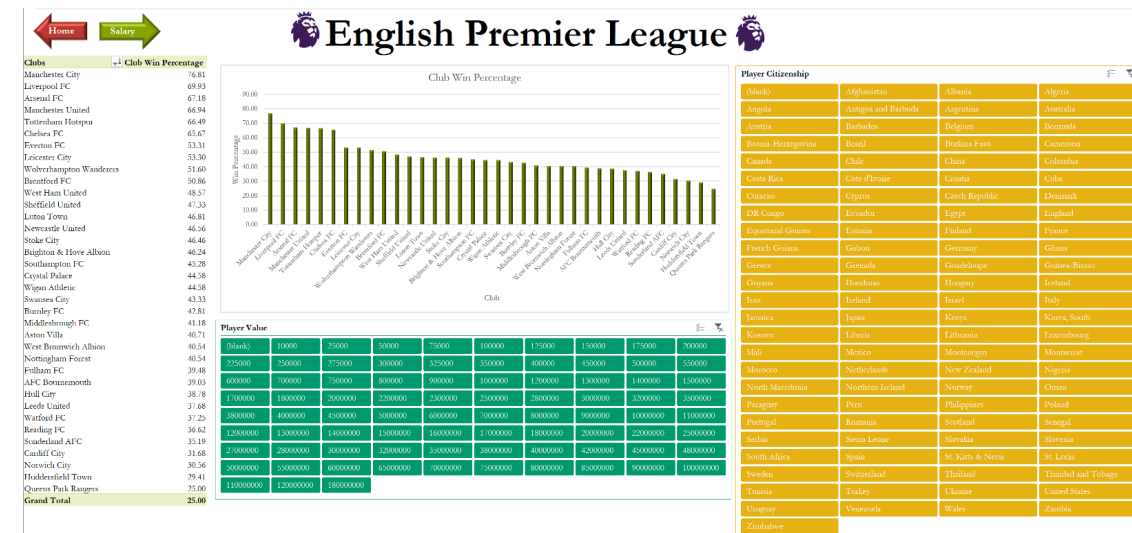
There was a lot of data within the nine files from the Kaggle competition, but I needed to narrow down the necessary files to four to get the data I was looking for. The four files I used were Competitions, Games, Clubs, and Players. Since there were four different files, I needed 4 different staging tables. Loading those staging tables in I noticed there were no stats on the team's overall win percentage. I calculated the win percentage of each team by using data from the Games staging table. After doing so, I would perform data quality checks. The first one was to see if there were any null values. There were many null values, but they belonged mostly to the market value, highest market value, and player birth columns. I decided to keep these null values in and deal with them later after compiling a transfermarktstaging table. Since players may defect from their birth countries or are born in another country than their parent's nationality, there are a lot of reasons why their birth country would be null. As for salaries, many players haven't had a chance to earn a professional contract, so those would show up as null. Any other fields where there were null values were dropped from the database. Descriptive Statistics for win percentage and player value can be found in the Excel sheet attached.

Organize:

To achieve the dataset I wanted to analyze I needed to create eight dimension tables for competition, player name, player country of birth, player country of citizenship, club, player value, and the player's highest value. I used their descriptions to load the specific data needed for each dimension table and cast their primary keys as the unique identifier for the distinct values before inserting them into the facttransfermarkt table. The win percentage was directly inserted from the tranfermarktStaging table. The entity-relationship diagram of all dimension tables and the fact table looks like this:



To make a more interactive visualization other than just a simply unadjustable histogram plot chart, I used Microsoft Excel to make an interactive dashboard. This dashboard uses pivot charts, pivot tables, and slicers to help understand and filter the data. For the Win Percentage dashboard, the user can navigate the player's current value and the country of citizenship. This allows the user to see which teams have specific salaries, while also being able to see which country they are from and compare that to the club's win percentage. The Player Salary dashboard allows the data to be filtered through specific clubs, player citizenship, and the win percentage per club. This allows the data to explain which clubs have a larger international squad, the amount of money they needed to spend for it, and the win percentage along with it.



Analyze:

The analysis of the data warehouse supports the hypothesis that increased expenditure on player salaries by English Premier League clubs is positively associated with higher levels of success in the league. The examination of financial and performance metrics consistently reveals that those clubs using greater financial resources for player salaries tend to achieve better standings within the league. Since the English Premier League is based in England, every team must have players of English citizenship, the other big-spending teams tend to spend more on other international talent ranging from other European and South American countries. During the data analysis, it was detectable that the bigger clubs with higher spending budgets have more international players than English players, which could be a reason why their payroll is so high and they can win more often than others. Although not every year the team with the highest payroll is the winner of the EPL, some teams are the underdog winners. The analysis found in the data is conclusive that the richest teams usually win more, but more data and a deeper understanding of that data is necessary to be able to see the correlation of "Moneyball" teams and their Cinderella runs (ex. Leicester City). The findings in the data analysis support the hypothesis, indicating a noticeable connection between heightened spending on player salaries and enhanced performance in the English Premier League.