

Data Visualisation Project

Charles Suddens-Spiers - K21040272

ACADEMIC HONESTY INTEGRITY

Students at King's are part of an academic community that values trust, fairness and respect and actively encourages students to act with honesty and integrity. It is a College policy that students take responsibility for their work and comply with the university's standards and requirements. Online proctoring/invigilation will not be used for our online assessments. By submitting their answers students will be confirming that the work submitted is completely their own. Misconduct regulations remain in place during this period and students can familiarise themselves with the procedures on the College website. *I agree to abide by the expectations as to my conduct, as described in the academic honesty and integrity statement.*

1 PART 1: ANALYTICS

1.1 Exploratory Research Questions Proposed

Q1 - Analyse the development of teams' performances over time. Are there detectable trends in the performance of teams over the last decade? For this question, the focus will be on the Premier League dataset provided, examining data from 2014 to 2024. Subquestions for this section could be: which teams have consistently performed at the top level, and is there a particular pattern in performance fluctuations for specific clubs?

Q2 - How does kickoff time affect Premier League match outcomes, and are certain teams more successful at specific time slots? For this question, there will be a temporal focus, comparing teams' performances across all possible match times. Subquestions that may be assessed include: Do teams perform differently during various periods of the year? Is there a correlation between kickoff times or the number of games per month and team performance? Additionally, are certain teams scheduled more frequently for specific kickoff times than others?

Q3 - What is the relationship between financial investment and Premier League performance, and which clubs demonstrate the best and worst returns on investment over time? The focus of this question is on financial investment when it comes to the transfer of players in and out of clubs. Looking at the last decade, for teams that have consistently been part of the Premier League during this period, how much have they benefited from their transfer investments? This analysis will explore whether higher spending directly correlates with improved league positions, identify which clubs have achieved the most efficient success relative to their spending, and determine if any clubs show consistently poor return on their transfer market investments.

1.2 Data Types and Datasets

In order to answer these research questions, extensive research was conducted to find the most appropriate datasets that can be used. The specific datasets used for each research question will be described in the following sections, with links to all data sources provided in the Appendix.

1.2.1 Premier League Tables Dataset

The Premier League keeps a record of all the league tables since its establishment in 1992. This dataset is an official record and very

reliable, containing final standings, points earned, goals scored and conceded, and win/draw/loss records for every team across each season. The record of Premier League tables makes it easy to compare season to season performance and track long-term trends in team success. However, the tables do lack depth of understanding to really analyse performance to a high level. They provide end-of-season summaries instead of a detailed story.

1.2.2 FootyStats

FootyStats provides detailed Premier League statistics, with rich information on team and player performance metrics. This dataset includes valuable match-specific data such as kickoff times, attendance figures, referee assignments, shots on/off target, possession percentages, and various performance indicators. The temporal information about match scheduling is particularly useful for Q2, allowing analysis of how kickoff times affect team performance. Unlike the Premier League tables dataset, FootyStats offers granular match-by-match data, enabling deeper analysis of performance patterns throughout the season. This detailed information could be used to aid in answering all research questions by providing contextual factors beyond just final standings and goal difference. However, this dataset only provides information from 2007 to the present.

1.2.3 Transfermarkt

Transfermarkt is a comprehensive database of football transfers, player valuations, and market data essential for addressing Q3. This dataset contains detailed records of player transfers including fees, dates, selling/buying clubs, and market valuations at the time of transfer. The financial information spans multiple seasons, allowing for thorough analysis of club investment strategies and their correlation with Premier League performance over time. The data regarding player valuation could provide useful insight when investigating clubs' effectiveness of their investments. However, some potential issues and inaccuracies come with this dataset. It is reliant on the transfer fees reported being accurate; some fees may only be estimates, some may be undisclosed, and some may be more complicated and costly for clubs than just the one-time transfer fee payment may suggest. Additional complexities like add-ons, sell-on clauses, and performance bonuses further complicate the true financial picture of transfers. This is still the most appropriate dataset for player transfers due to the scale of the dataset as well as the level of detail provided.

1.3 Correlation

All of these datasets can be correlated by the common elements of club name and year/season. Each dataset contains these fields, allowing for direct comparison and integration between sources. The Premier League Tables dataset provides the foundational performance outcomes that the detailed match statistics from FootyStats and the financial information from Transfermarkt can explain. By linking these datasets through team names and seasons, the research questions can be explored. However, the time format and club naming are not consistent across all the datasets and some additional processing may need to be undertaken to ensure data from the desired time frames is being used correctly.

2 PART 2: DESIGN AND DISCUSSION

2.1 Question 1

The best team in the Premier League is the one that wins it. A team's overall performance is determined by its league position, which was the first metric that came to mind when considering team performance. To detect trends and enable comparisons, league positions over time can be effectively visualised as a line graph, as supported by Aigner et al. [1], who demonstrate that line graphs are particularly effective for revealing trends in sequential time-based data.

To compare multiple teams' performances simultaneously, each team is represented by a distinct line. Teams are differentiated by colours that correlate to their primary club colours, with team badges and names appearing at the end of each line. This creates a visually engaging presentation while allowing for clear differentiation between data points.

The y-axis is deliberately inverted, placing position 1 at the top of the scale and 20 at the bottom, mirroring the football table format. Since Premier League rules ensure no two teams can finish in the same position, there will never be positional overlap. When teams are relegated, their lines appear to move off the page between seasons, returning only if the team is promoted back to the Premier League.

Users can interact with this visualisation in several ways. Initially, no lines are displayed until the user activates the timeline controls. Users can play the animation to observe changes over time or step forward/backward through seasons. Additionally, clicking any point on the graph displays a popup with detailed performance statistics for that team in that specific season.

For this visualisation, inspiration was taken from the Premier League performance visualisation by Café Tactiques [6]. For my design, it is Figure A.1 in the Appendix.

2.2 Question 2

To answer this question, we need to investigate performance data at different intervals of time whilst comparing this against the other teams in the time frame. This would consist of both numerical and categorical data. In this instance, the different clubs would be categorical and the different kick-off times would be the numeric data. To measure the performance at these data points, we can look at the win percentage from each team during each kick-off time.

To create a visualisation of this performance, heat maps provide an ideal solution. According to Suematsu et al. [5], heat maps are particularly effective for "time-varying multi-variate data visualization unifying numeric and categorical variables," which perfectly describes our scenario of analysing team performance across different kick-off times. Their research demonstrates that heat maps excel at revealing patterns that might otherwise remain hidden in traditional statistical analyses.

The design uses colour hue and luminance to represent win percentages, with a colour scale from red (poor performance) through neutral grey to green (strong performance). This colour scheme creates an intuitive visual encoding that allows users to quickly identify which teams perform better or worse at specific kick-off times. We as humans associate red with bad and green with good so this was a natural choice. Users can easily spot temporal patterns such as teams that consistently perform well in evening matches or struggle with early weekend fixtures.

The interactive elements allow users to filter by season range and toggle between home and away performances. When clicking on a specific cell, detailed match statistics appear, showing the comprehensive breakdown of that team's performance at that particular kick-off time. There is even more detail that can be analysed, the visualisation can take into account the performances of teams during the first or second half only. This may prove an interesting analy-

sis of teams' performances in the first half compared to the second half.

This design draws inspiration from GitHub's contribution calendar visualisation [4], which similarly uses colour coded cells to represent activity intensity across different time periods, making it easy to identify patterns and outliers at a glance. My design for this visualisation can be found at Figure A.2 in the Appendix.

2.3 Question 3

One of the biggest factors when it comes to a team's performance is the players in the club. Part of this involves the movement of players in and out of the club. To answer question 3, we have to investigate the financial performance of a club when it comes to the transfer market and how this correlates to their performance in the Premier League.

A scatter plot was selected as the primary visualisation method as it excels at revealing correlations between two metrics. According to Elmqvist et al. [3], scatter plots are particularly effective for "multidimensional visual exploration" and allow users to "navigate through the multidimensional space" of complex datasets. This approach is ideal for examining the relationship between financial investment and sporting success.

Each team is represented by a point whose position encodes two primary metrics: total transfer spending (x-axis) and total points gained (y-axis). Teams are differentiated by colours matching their primary club colours and include their badges for immediate recognition. This encoding allows users to view the correlation between these two metrics for teams over the desired time period.

The interactive elements of this visualisation include: users can select a specific time frame using the season selectors. The user can select any of these scatter points and get a more detailed breakdown of the transfers made by the club during this time period, gaining an idea of the net spend, the names of players who have left and joined the club, and some more overall metrics of performance including points per million spent and the average league position.

The doughnut chart visualisation for financial breakdowns draws inspiration from EA's FC Career Mode interface [2], which presents club finances in a similar format. This familiar presentation helps users quickly comprehend the financial composition of each club's transfer activity, distinguishing between money spent and received. My design for this visualisation can be found at Figure A.3 in the Appendix.

REFERENCES

- [1] W. Aigner, S. Miksch, H. Schumann, and C. Tominski. *Crafting Visualizations of Time-Oriented Data*, pp. 83–127. Springer London, London, 2023. doi: 10.1007/978-1-4471-7527-8_4 2
- [2] Electronic Arts. *Fc 24 career mode*, 2024. Video game financial interface. 2
- [3] N. Elmqvist, P. Dragicevic, and J.-D. Fekete. Rolling the dice: Multidimensional visual exploration using scatterplot matrix navigation. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1539–1148, 2008. doi: 10.1109/TVCG.2008.153 2
- [4] GitHub. *GitHub contribution calendar visualization*, 2025. Accessed: 19/04/2025. 2
- [5] H. Suematsu, S. Yagi, T. Itoh, Y. Motohashi, K. Aoki, and S. Morinaga. A heatmap-based time-varying multi-variate data visualization unifying numeric and categorical variables. In *2014 18th International Conference on Information Visualisation*, pp. 84–87, 2014. doi: 10.1109/IV.2014.25 2
- [6] C. Tactiques. *2020–21 premier league review — graphs*. Café Tactiques Blog, June 2021. 2

Appendix A

Here are the datasets discussed in Part 1:

A.1 Visualisation Datasets

- **Premier League Tables Dataset** - Premier League tables from 1993 to 2024 - provided by module leader
- **FootyStats** - Detailed stats on the league tables over the years, player stats, match stats and more can be found [here](#).
- **Transfermarkt** - provides detailed information regarding the football transfer market, the [website](#) and [CSV](#)

A.2 Visualisation Datasets

- **pl-tables-1993-2024.csv** - Premier League tables from 1993 to 2024 - provided by module leader
- **transfers.csv** - Player transfer data - Found from [Kaggle](#)

A.3 Visualisations

A.3.1 Design 1

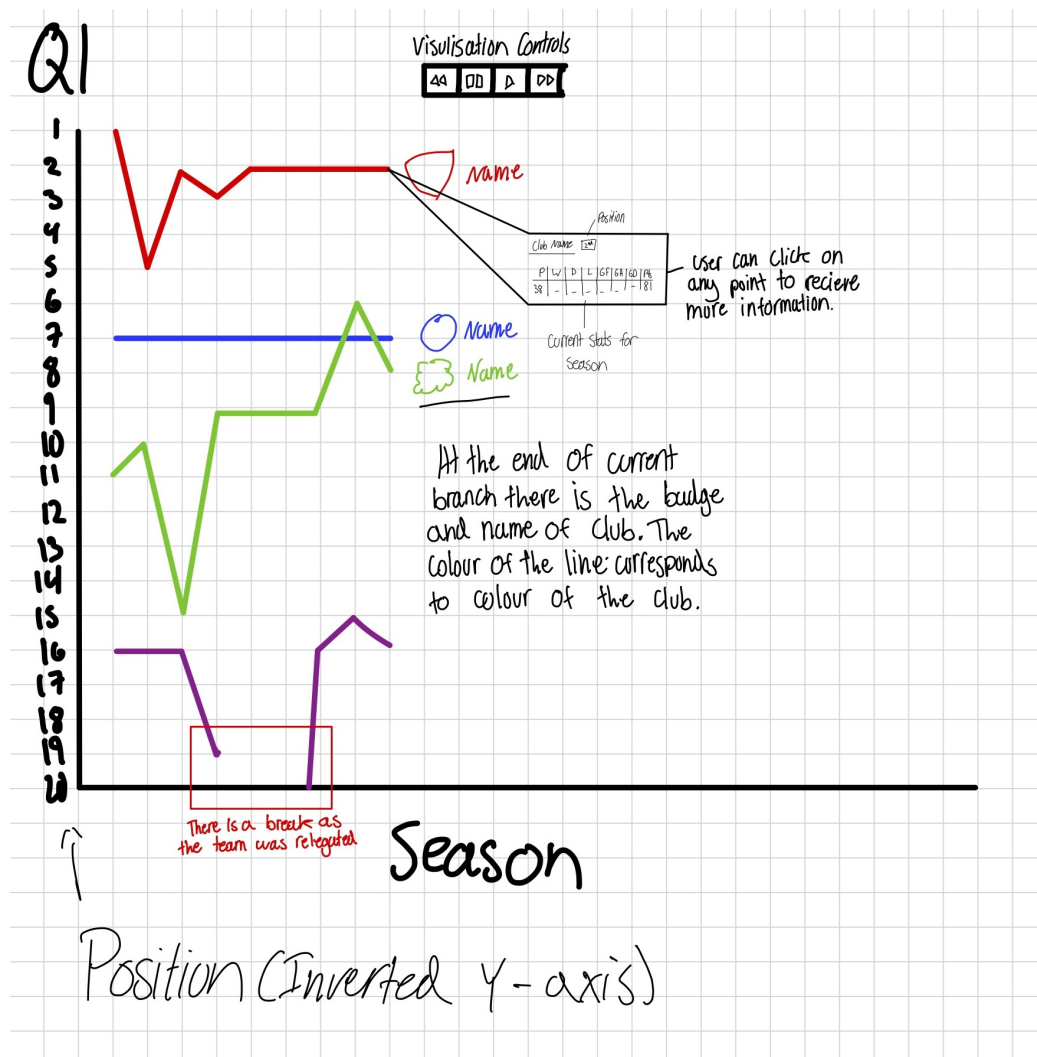


Figure A.1: Team Performance Trajectory Visualization showing position changes over seasons with interactive elements.

A.3.2 Design 2

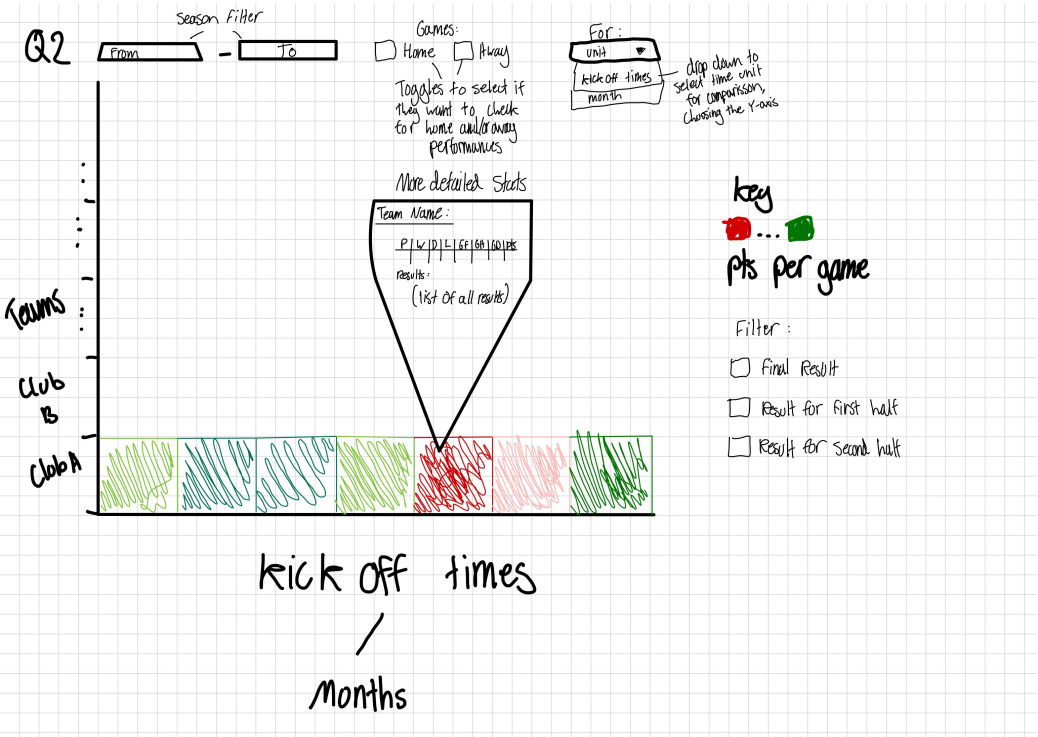


Figure A.2: Match Result Timeline Visualization showing performance patterns with kickoff times and filterable views.

A.3.3 Design 3

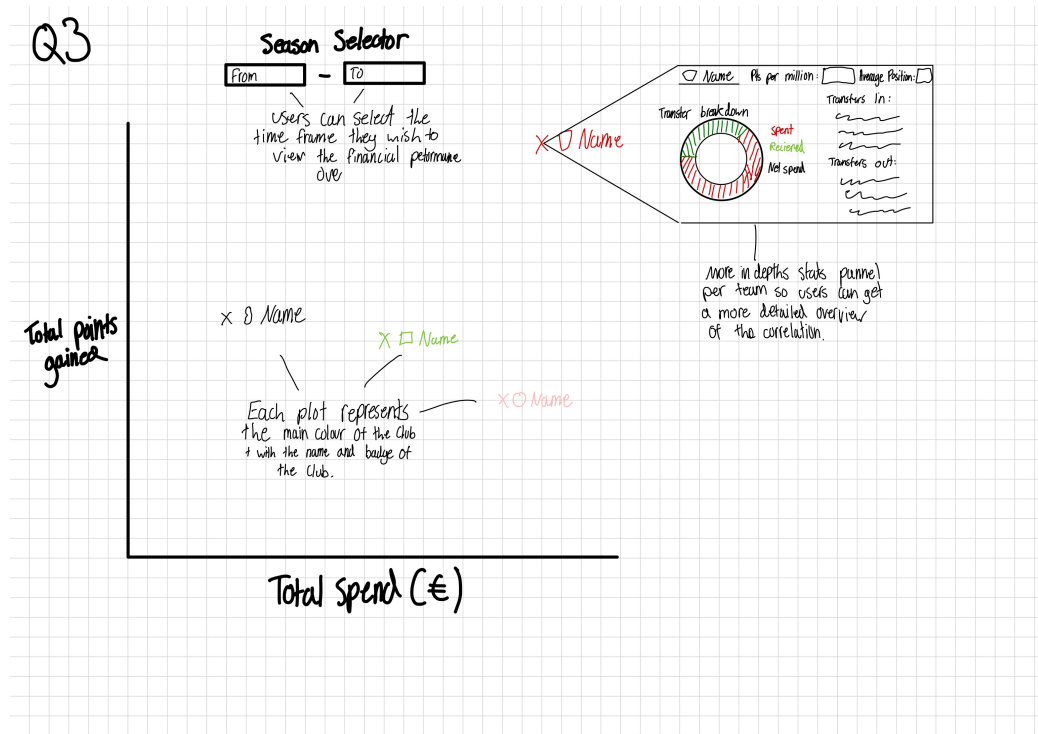


Figure A.3: Financial Performance Correlation Visualization comparing team spending against points gained.