The University of Hong Kong

COMP7507
Visualisation and Visual Analytics
UCL Football Analytics Project Report
Group 5

Lee Wing Chun (3036212881)

A M M Mutasim Moyen (3036383226)

Md Tanjeeb Hossain (3036199560)

Asif Salam (3036383446)

Pan Jia Chen (3036383094)

Contents

Part I: Overall Work

INTRODUCTION

Football analytics has become an indispensable tool for understanding game dynamics, optimizing team strategies, and evaluating player performance. This project delves into the realm of UCL (UEFA Champions League) football data to uncover key trends, match-level insights, and individual player metrics. By leveraging data visualization and analytical techniques, we aim to provide a comprehensive overview that can inform coaches, analysts, and enthusiasts alike. This report outlines our approach, the tools utilized, the visualizations created, and the key findings derived from our analysis.

Project Components

This project visualises the UCL data in 7 tableau dashboards and a python visualisation. These items are consolidated and available online at cocolwc.github.io/COMP7507/.

HIGHLIGHTS

- Comprehensive Dashboard Suite: Developed a series of interactive Tableau dashboards covering league overviews, match-level details, defensive strategies, and individual player performance, offering multi-faceted insights into UCL football.
- Dynamic Visualization: Created a dynamic visualization using Python, presented as a video demonstration, to complement static dashboards and provide a more immersive understanding of game flow and player movement.
- Data-Driven Insights: Successfully identified and presented key performance indicators and strategic patterns, enabling a deeper understanding of UCL matches and player contributions.

VISUALISATION WITH EXPLANATIONS

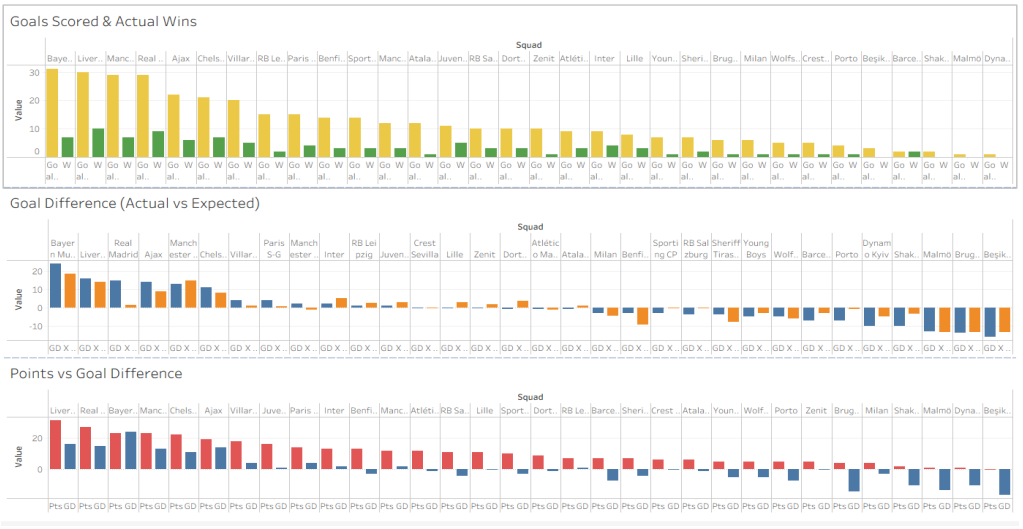
Our project developed a suite of Tableau dashboards designed to provide comprehensive insights into UCL football. Each dashboard serves a specific analytical purpose, offering interactive exploration of the data.

Dashboard 1a: UCL League Overview

These dashboards present league-wide statistics and trends for a given season. It allows users to gain a high-level understanding of team performance across the league, including metrics such as total goals scored, conceded, win/loss ratios, and league standings.

Dashboard 1a presents a streamlined league snapshot centred on two fundamental metrics: goal difference and total points. The top panel ranks squads by goal difference, highlighting how balanced offensive and defensive performances translate into overall dominance. Clubs such as Bayern Munich, Liverpool and Real Madrid exhibit substantial positive goal differences, while teams at the opposite end of the spectrum show values near zero or slightly negative. This simple layout allows viewers to see immediately which teams excelled on both ends of the pitch and which struggled to maintain parity.

The bottom panel ranks teams by total points, providing a complementary measure of success. Here again the heavyweights of European football occupy the upper tier, while mid-table and lower-ranked squads fall away. By juxtaposing these two bar charts, Dashboard 1a enables users to compare how goal differential aligns with points won. The interactivity built into Tableau—tooltips on hover and the ability to highlight individual bars—encourages exploration without overwhelming the viewer with excess detail.



Dashboard 1b: UCL League Overview

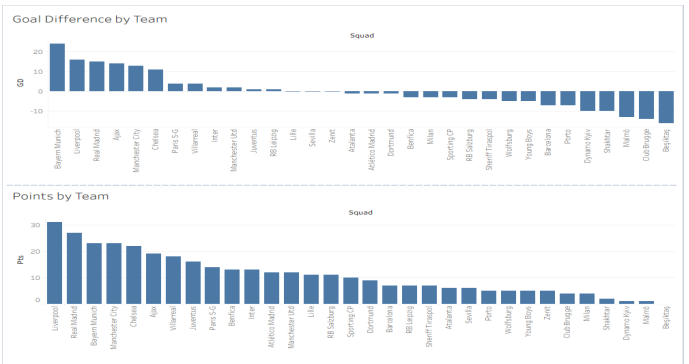
Dashboard 1b offers a richer, multi-layered examination of league performance by incorporating additional variables and comparative analyses:

Goals Scored & Actual Wins: This grouped bar chart displays each team's total goals scored alongside its number of matches won. The visual emphasizes that scoring prowess generally correlates with victories, yet differences emerge; some sides, for example, recorded high goal tallies but relatively fewer wins, indicating defensive frailties or inconsistent match outcomes.

Goal Difference (Actual vs Expected): The second chart compares actual goal difference (GD) to expected goal difference (xGD). Expected metrics are calculated from shot quality probabilities, so the side-by-side bars reveal which teams outperformed or underperformed their underlying statistics. Positive disparities suggest clinical finishing or strong defensive saves, while negative gaps imply that a team's results were worse than the quality of chances created or conceded.

Points vs Goal Difference: The final panel juxtaposes points earned with goal difference. A positive correlation appears: clubs with larger positive goal differences tend to accumulate more points. Yet outliers also emerge—teams whose points tally is disproportionately high or low relative to their goal difference—prompting further investigation into tactical style, match context or luck.

Together, these three visualisations provide a comprehensive, interactive overview of the league. They enable users to explore how attacking output, defensive solidity, and underlying performance metrics interact to produce results. Compared with Dashboard 1a, Version 1b delves deeper into the data by incorporating expected metrics and breaking down goals into wins versus total scoring. This added context helps analysts distinguish between sustainable success and results driven by variance or finishing streaks, making Dashboard 1b an essential complement to the more succinct overview provided by Dashboard 1a.

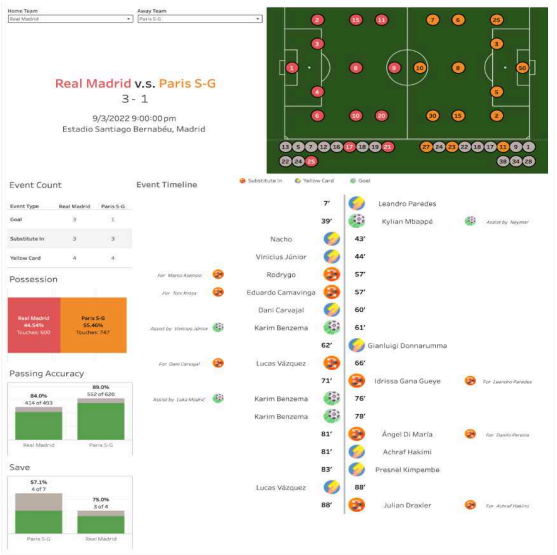


Dashboard 2: Match Level Analysis

Purpose

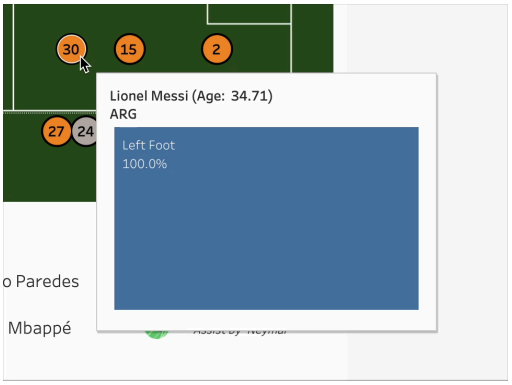
The dashboard is designed to offer a comprehensive overview of key match events and performance metrics. Focusing on individual matches, this dashboard displays key statistics, event timelines, and formation maps. Users can select specific matches to delve into detailed performance metrics, visualize player positions, and understand critical moments during the game. For example, coaches could use this dashboard to review the players performance in a selected match, and observe the outcome of the applied tactic, such as formation and player substitution.

Description

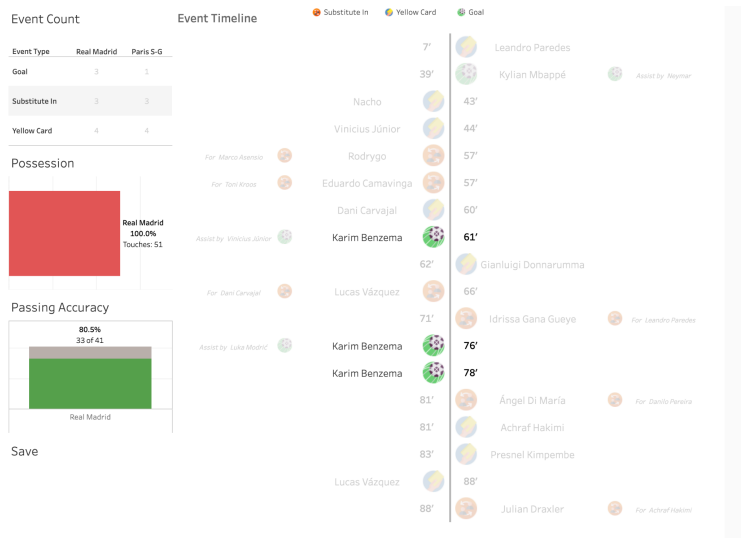


- **Match Summary and Team Filters:** At the top, the dashboard clearly displays the home and away teams, the final score, the date and time of the match, and the venue in a text box. The font colors used in this text box representing home team (red) and away team (orange) are aligned with the other graphs such as Possession and Tactical Formation. Users could choose the match by selecting the home team and away team in the 2 dropdowns above the summary text box.
- **Tactical Formation:** A prominent visual element is the football pitch diagram, showing the tactical formations and player positions for both teams with a scatter plot. Players are represented by numbered circles, providing a quick visual reference to their roles on the field.
    - Interaction:
        - Hover on the numbered circle could check player information such as name, age, nation and dominant body parts (e.g. right foot, left foot, head) used in

attack actions.



- Click on the numbered circle could:
  - Highlight the event that the selected player involved in the match
  - Filter the Possession and Passing Accuracy of the selected player
  - Note: The following screenshot show the example of clicking on the number 9 of red team - Real Madrid



- **Event Count and Timeline:** The "Event Count" section summarizes the count of significant events like "Goal," "Substitute In," and "Yellow Card" for each team. Next to this, the "Event Timeline" provides a chronological sequence of these events, noting the minute of occurrence and the player involved, often with an assist or the reason for the event (e.g., "Assist by Neymar"). This allows for a quick understanding of how the match unfolded.
- **Possession:** A stacked bar shows the possession statistics for both teams, including the percentage of possession and the total number of touches.
- **Passing Accuracy:** A bar chart visualizes the passing accuracy for both teams, displaying the percentage of successful passes out of total passes.
- **Save:** A bar chart presents goalkeeper save statistics, showing the percentage of saves made

out of total shots faced.

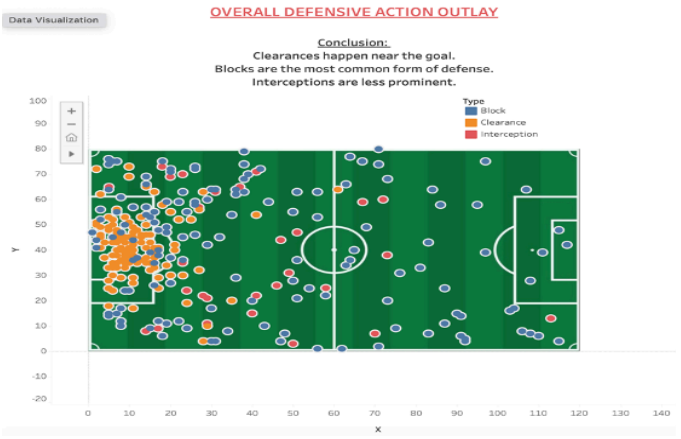
Dashboard 3: Defensive Strategies

This series of dashboards provides a detailed examination of defensive performance.

- **Dashboard 3a (General Overview):** Offers a broad look at defensive metrics across teams and players for 3 seasons - 2013/14, 2014/15, 2015/16. This section consists of the following two worksheets:

1. Defensive Action Heatmap

Its purpose is to visualize the spatial distribution of defensive actions across the pitch. We have used a density heatmap with x and y coordinates on a 1593x980 pixel football pitch background (X: 0-120, Y: 0-80).
The rationale behind this technique is that heatmaps effectively highlight high-density zones, aiding tactical analysis of defensive coverage.

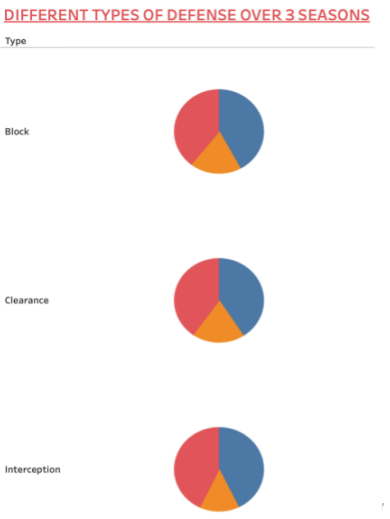


We can generate the following hypothesis from this dashboard: **Do concentration of a particular defensive activity near the goal lead to tangible results?**

The effectiveness behind this visualization is that it proves our hypothesis - The Defensive Action Heatmap showed concentrated activity near the goal especially clearances, suggesting teams with higher midfield defensive actions concede fewer goals. This has been backed up by the match results for the 3 seasons in question.

2. Defensive Trends Pie Charts

Its purpose is to analyze defensive action trends over matches for the 3 seasons in focus. We have used pie charts with *match_id* on columns, *action counts* i.e. blocks, clearance and interceptions on rows, colored by *season_id*. The blue color represents the 2013/14 season, the orange represents the 2014/15 season and the red represents the 2015/16 season.
The rationale behind this technique is that pie charts clearly depict temporal patterns, useful for tracking performance changes.



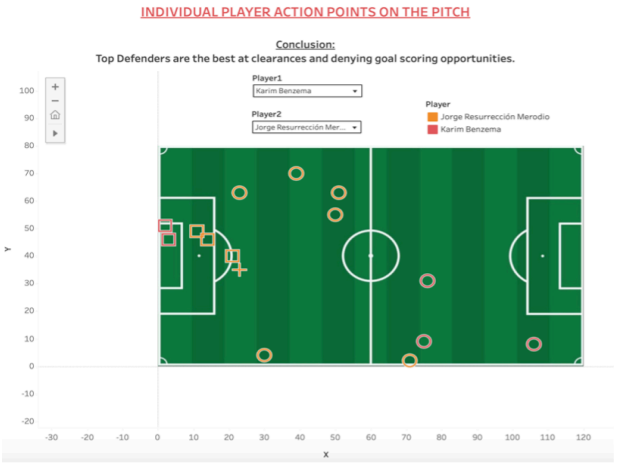
We can generate the following hypothesis from this dashboard: **Do defensive actions increase in high-stakes matches?**

The effectiveness behind this visualization is that it proves our hypothesis - The pie chart showed peaks in knockout stages, indicating that defensive actions increase in high-stakes knockout matches. This has been backed up by the match results for the 3 seasons in question.

- **Dashboard 3b (Player-Centric Insights):** Provides in-depth analysis of individual player defensive contributions, such as tackles, interceptions, and clearances. This section consists of the following two worksheets:

1. Player Comparison Action Points

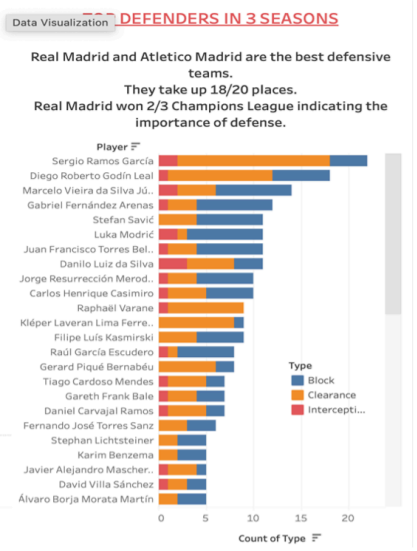
Its purpose is to map individual action points of any two players that participated in the 3 seasons. The x and y axis represent the pitch coordinates of a particular action. Each action point is differentiated by *Type* by using different shapes - square, circle, cross. The two players can be selected by using a parameter and the parameter is implemented by using a filter *SelectedPlayers*.



The rationale behind this technique is that each defensive action can be individually differentiated.

2. Top Defenders Bar Charts

This visualization ranks players by total defensive actions. The technique used is a stacked bar chart with players on rows, summed action counts *(Tackles Count, Interceptions Count, Clearances Count, Blocks Count)* on columns, colored by *Type*, with a *Top N Players* parameter for ranking.
The rationale behind this technique is that stacked bars show both total actions and type-specific contributions, ideal for identifying key performers.



We can generate the following hypothesis from this dashboard: **Are the top defenders part of title-winning teams or just individual matches ?**

The effectiveness behind this visualization is that it proves our hypothesis - The top defenders are part of Real Madrid and Atletico Madrid. They took up 18 out of top 20 positions and Real Madrid were the champions for 2 out of those 3 seasons.

- **Dashboard 3c (Individual Heatmaps):** Visualizes player movement and defensive zones, highlighting areas of high activity and strategic positioning.

The purpose of this dashboard is to compare the total defensive action zones of two selected players.
We are using density heatmaps with dynamic axis scaling, player on color, using *Player1 and Player2* parameters. The 3 different action zones are shown via line charts and differentiated by color.

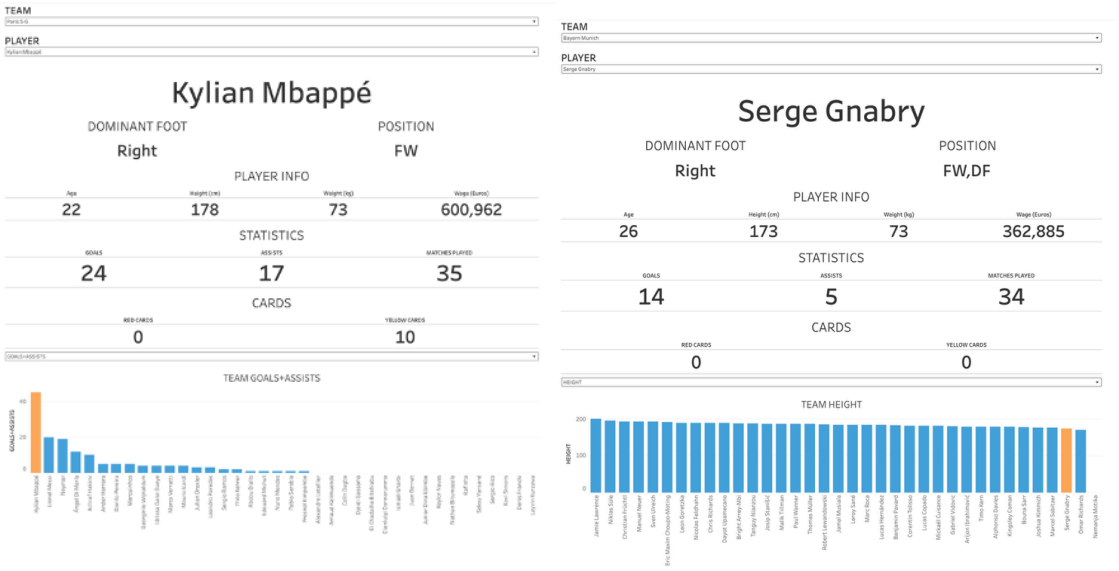
The rationale behind using this technique is to implement dynamic scaling which focuses on relevant areas on the pitch, enhancing player comparison for tactical decisions.

We can generate the following hypothesis from this dashboard: **Do players have specialized defensive roles based on position?**

The effectiveness behind this visualization is that it proves our hypothesis - Comparing heatmaps of two different players explicitly showcase that different positions of defenders on the field (Centre Back vs Left Back vs Right Back) have different roles i.e. some players are performing more blocks while some are doing more interceptions or clearances.

Dashboard 4: Player Statistics

Dashboard 4a: Player Level Dashboard

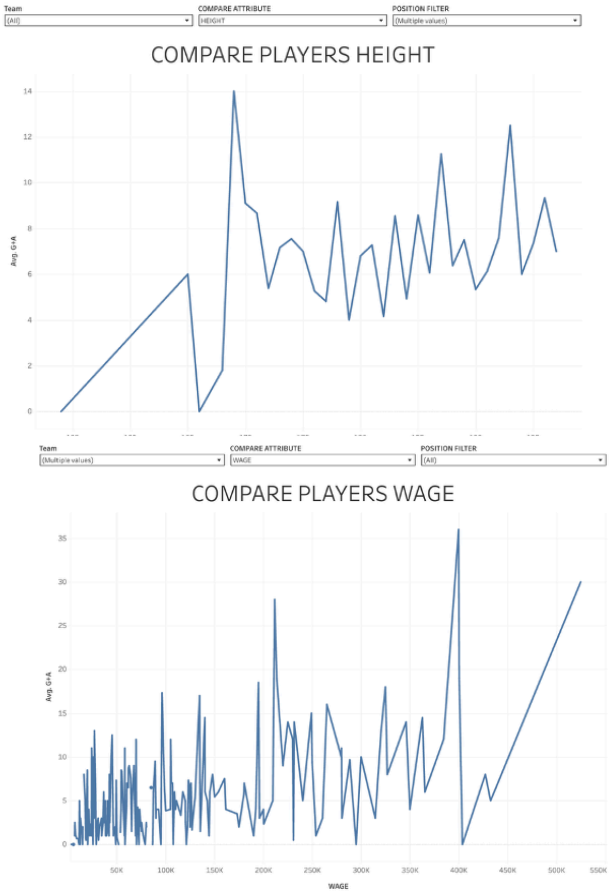


Dashboard 4a allows the user to view player level information and statistics. At the top of the dashboard, the user can select the team and player they wish to view. Then important player information like dominant foot and standard playing position is displayed. Further player bio information includes player age, height, weight, and wage. Below are important performance and match metrics like assists, goals, and the number of matches started. Finally, the player discipline is visualized by how aggressive the player is based on the red and yellow cards received. Under the player information is a bar chart to compare the player with others on the team. The selected player is highlighted in orange while the rest of the team remains blue. The comparison attribute can be selected from the dropdown menu at the top of the bar. The players can be compared based on goals, assists, and the rest of the player-specific attributes displayed above.

This dashboard gives the coach an overview of each player, which will allow them to make informed decisions on match lineup, substitutions, and formations. The dashboard will also highlight higher-performance members of the team, allowing the coach to utilize them more effectively. The coach can also look at how much each player is costing the team, aiding them in making decisions on squads for future tournaments. Using dashboard 4a along with the formation dashboard will support the coach in making intricate tactical decisions for upcoming matches.

In the above examples, the coach for PSG can view the player information for Mbappe and come to the conclusion that he is performing significantly higher than his teammates. The coach can then play him in his preferred forward position more often and give him more opportunities to score. In the second example, the coach for Bayern can see that Gnabry is one of the shorter players on the team. The coach can then design the formation and position him accordingly.

Dashboard 4b: Player Attributes Comparison



Dashboard 4b allows the coach to see an overview of how different player attributes impact player performance metrics. For this dashboard, the performance metrics utilized are the average sum of goals and assists of the players. On top the user can filter the players considered by the team they play in and their preferred position. The coach can also select the player attribute to compare, like height, weight, and wage.

This visualization will allow the coach to better understand and explain the factors affecting positive and negative player performance metrics. It will allow them to make informed decisions on player formations, positioning, and lineups for matches.

In the first example above, the coach can see that taller forward players are more likely to contribute to a goal, this can influence them to pick taller people as strikers for his lineup. The coach can also position taller players as forwards in his match formations. The coach can also see that player wages are not an indicator of abilities and can make judgements on future squads based on that.

Python Dynamic Visualization Video

In addition to the static dashboards, a dynamic visualization created with Python is featured in a video demonstration. This visualization complements the dashboards by offering a more fluid and engaging way to understand game dynamics.

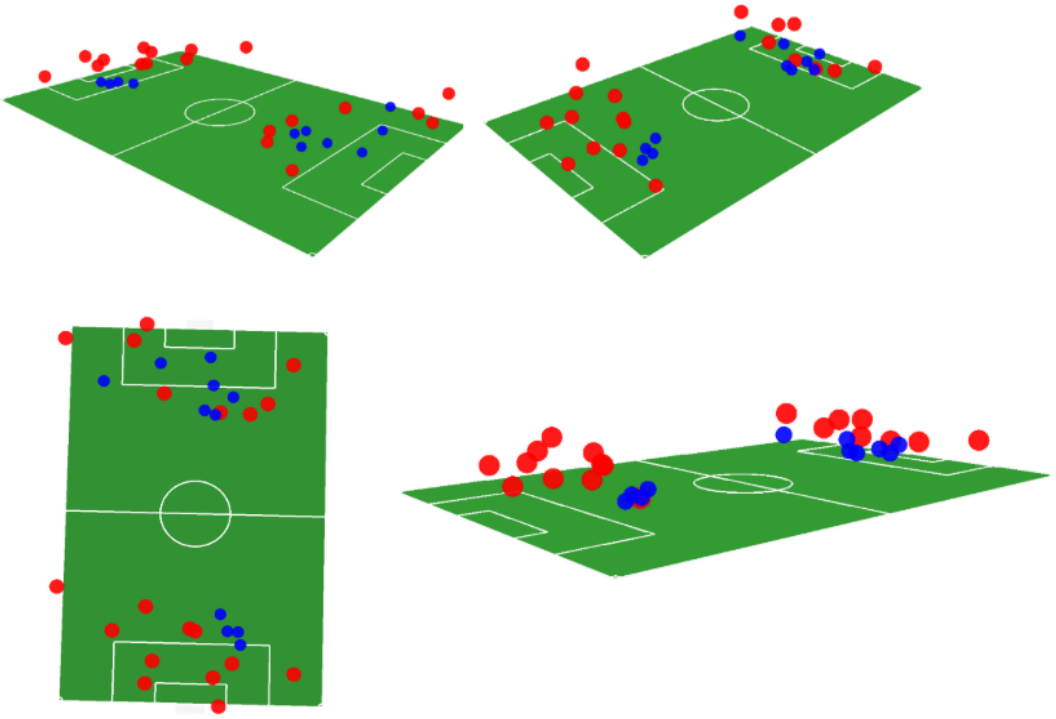
This code realizes the visualization of football shooting data through the construction of a 3D interactive scene, with the core technical paths as follows:

Scene Following

Based on the standard field parameters of the European Cup (length 105m, width 68m), the FootballField class is used to accurately draw 3D football field elements, including:

- A green field plane as the base layer;
- A system of white lines (sideline, center line, penalty area line, small penalty area line, center circle, corner arc);
- 3D models of the goals on both sides (with depth effect).

The 3D scene construction is realized using Plotly's Mesh3d (for drawing planes / three-dimensional graphics) and Scatter3d (for drawing lines / points).



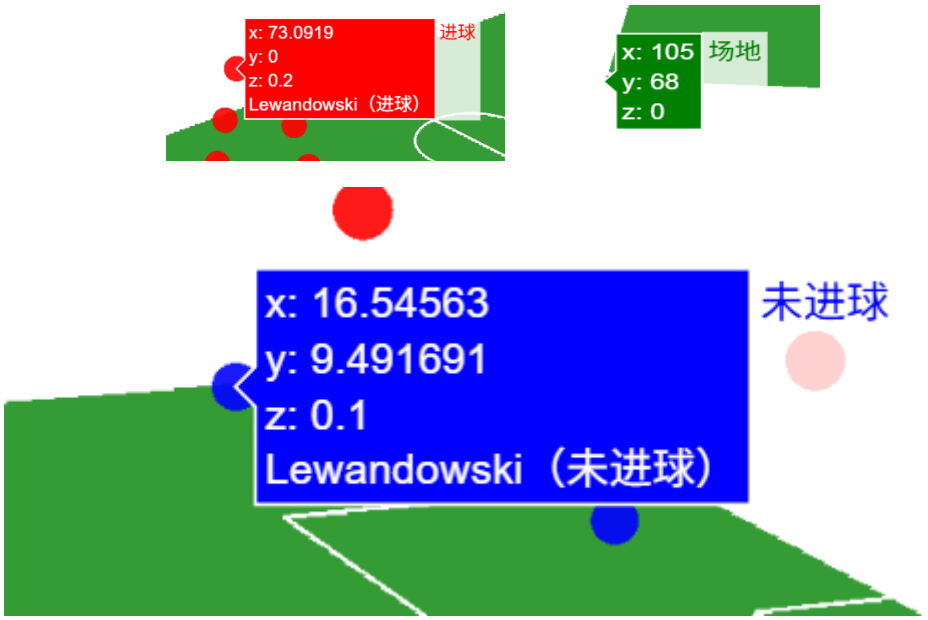
Data Mapping

The `add_shot_points` function maps shooting data to marker points in the 3D scene:

- Visual attributes are differentiated by "whether a goal is scored" (large red dots indicate goals, while small blue dots indicate missed shots);
- It supports data filtering by player name (e.g., focusing on "Lewandowski's" shots in the example); - Text information including the player's name and goal status is displayed when the mouse hovers over a marker.

Interactive Controls

It adopts Plotly's browser rendering mode, supporting scene rotation, zooming, and panning to facilitate observing shot distributions from multiple angles; The initial camera perspective is fixed (x=1.5, y=1.5, z=0.5) to balance the overall view of the scene and the visibility of details.



Purpose

The purposes of the visualization are as follows:

1. Intuitively present the spatial distribution of shots Map abstract shot coordinate data (position_x, position_y) onto a football field with real proportions, helping to understand the relationship between shot positions and field areas (such as the penalty area, corner area).

2. Comparative analysis of shooting effects By distinguishing goal and non-goal data through colors and sizes, it allows for quick identification of areas with high shot success rates, or the shooting efficiency characteristics of specific players (such as Lewandowski in the example).

3. Support targeted analysis Provide a player filtering function to facilitate focusing on the shooting habits of individual players (such as preferred areas, success rates) or comparing performance differences between different players.

DISCUSSION

DIFFERENT METHODS

Visualization Tools

Beyond the Tableau dashboards, a Python-based visualization was created to offer a dynamic, animated view of on-field events. This provided a complementary perspective to the static, aggregated views in Tableau, allowing for a more immersive understanding of game flow.

Data Sources and Extraction Techniques

The data for this project was sourced from multiple platforms, including:

Tableau Dashboard: Data Source & Data Preparation

Fbref.com is a comprehensive football statistics website. To efficiently gather the extensive match and player data, a custom Python web scraping script was developed. This automated web scraping approach allowed for the efficient collection of a large volume of granular football data, forming the foundation for our analytical dashboards and dynamic visualizations.

The scraping process involved:

Identifying Match URLs: The script first navigated to the main Champions League statistics page (https://fbref.com/en/comps/8/2021-2022/2021-2022-Champions-League-Stats) to extract URLs for

individual match reports.

Extracting Match Details: For each match URL, the script accessed the page and extracted various detailed information, including:

Match Info: Home and away team scores, expected goals (xG), team names, formations, venue, and match date/time. This data was then structured into a pandas DataFrame and saved as match_info.csv.

Event Data: A chronological list of in-game events such as goals, yellow cards, substitutions, and their corresponding players and timestamps. This was processed and saved as event.csv.

Formation Data: Player positions and numbers for both starting lineups and bench players, including their approximate coordinates on the pitch for visualization purposes. This was saved as formation.csv.

Player Statistics: Various tables containing detailed player statistics (e.g., Summary, Passing, Defensive Actions, Possession, Miscellaneous Stats, Goalkeeping, Shooting, Goal and Shot Creation). The script dynamically identified and extracted these tables, renaming columns for consistency and saving each as a separate CSV file (e.g., Liverpool Player Stats Table - Summary.csv).

Data Storage: All extracted data was organized into separate CSV files within dedicated directories for each match, ensuring a structured and accessible dataset for subsequent analysis in Tableau.

Defensive Dashboard: Data Source & Data Preparation

The defensive data was taken from StatsBomb which is hosted on GitHub. This link contains a JSON file *('competitions.json')* which contains season-wise and competition-wise data for multiple teams. Two Python scripts were written. The first one was used to extract season and competition IDs named *'checkseasons.py'*. The second Python script was written to extract relevant defensive data including season, players, teams, pitch coordinates and different types of actions. This used the IDs extracted using the first python script. This file was named *'get_defensive_data.py'*. This script extracted and converted the JSON file into a cleaned CSV file called *'defensive_actions_three_seasons.csv'*.

The link for the data source is:

'*https://github.com/statsbomb/open-data/blob/master/data/competitions.json'*

Python dashboard: Data Cleaning

Standardization of data format

Unify the format of the goal field: Standardize possible expressions in the original data such as "goal/no goal" into boolean values (TRUE/FALSE).

Standardize coordinate data: Ensure that position_x and position_y are values based on a unified

coordinate system (such as metric units based on the length/width of the football field), and correct possible format errors (such as coordinate values in string type, redundant symbols, etc.).

Deduplication and handling of abnormal values

Remove duplicate records: Delete duplicate shooting events caused by data crawling errors or duplicate entries (such as duplicate records of the same player at the same time and position).
Clean up abnormal coordinates: Eliminate coordinate values that exceed the reasonable range of the football field (for example, if the length of the field is 105 meters, a position_x of 150 meters is an abnormal value and needs to be corrected or deleted).

Handling of missing values

Complete key information: For records missing club or player_name, supplement and complete them by associating with team lineup tables or player IDs for matching.
Handle null values: If there are null values in position_x/position_y, it may be necessary to trace back the position according to the match video, or mark them as invalid data and delete them (to avoid affecting the visualization effect).

Data association and integration

Associate multi-table data: If player information and shooting events are separated in the original data (such as player tables and event tables stored independently), it is necessary to associate the two tables through key fields such as player IDs to supplement player_name and club information.
Unify data granularity: Ensure that each record corresponds to an independent shooting event, and avoid redundant records formed by merging multiple events.

DIFFICULTIES

- **Data Cleaning:** Splitting the location field into x and y coordinates in Python required careful debugging to handle null values, resolved by validating outputs in the script.
- **Top N Filter Issue:** The Top N Players parameter initially filtered by action counts rather than player ranks, fixed by correcting the RANK calculation and ensuring proper table calculation settings.
- **Performance:** Large datasets slowed Tableau's rendering; mitigated by filtering out null values and optimizing calculated fields.

- **Data Limitations:** The absence of off-ball movement data and match outcome details (e.g., goals conceded) limited hypothesis verification, requiring reliance on visual patterns and assumptions.

- **In-app Visualization Limitations:** Tableau is not natively designed for creating highly dynamic, fluid visualizations of player movement and on-field events. This limitation was addressed by developing a separate dynamic visualization using Python, which complemented the static dashboards but highlighted a gap in Tableau's capabilities for this specific type of analysis.

POSSIBLE EXTENSIONS

- **Video Integration:** Planned to link video clips of defensive actions to data points for richer analysis, but time constraints and lack of video data in the StatsBomb dataset prevented implementation.
- **Real-Time Data Updates:** Considered integrating ongoing season data for live analysis, but the static nature of the dataset and time limitations made this infeasible.
- **Match Outcome Analysis:** Wanted to correlate defensive actions with goals conceded to fully verify hypotheses, but the dataset lacked outcome data, limiting this exploration.
- **Advanced Machine Learning Applications:** The project could be expanded to use advanced machine learning techniques like deep learning for player tracking and match result prediction. This would require a different type of data (e.g., video or GPS tracking) and more computational resources than were available.
- **Correlation and Interaction between Dashboards:** Due to time constraints, the project team developed separate visualizations in Tableau and Python. A more comprehensive solution would have been to link all dashboards together in Tableau, enabling users to "drill down", from league level to match level, and finally player level, for a seamless analytical experience.

LIMITATIONS

- **Reliance on Tableau:** While a powerful tool for interactive dashboards, its primary strength is visualizing structured data. This necessitated a significant pre-processing and cleaning phase using a Python script to prepare the raw data.

- **Scope Limitations:** The analysis was focused exclusively on UCL data for a specific season, which may limit the generalizability of the findings to other leagues or time periods.
- **Data Granularity:** The reliance on event data from web scraping, rather than more detailed in-match tracking data, meant that certain aspects of player movement and team strategy could not be captured or analyzed.

CONCLUSION

In conclusion, the UCL Football Analytics project successfully developed a robust set of interactive dashboards and a dynamic visualization to provide comprehensive insights into football performance. We demonstrated the power of data analytics in uncovering key trends, analyzing match events, and evaluating player contributions. The project's outputs serve as a valuable resource for understanding the intricacies of UCL football and lay the groundwork for further advanced analytical endeavors.

Part II: Individual Contributions

ASIF SALAM (3036383446)

Details of Tasks Completed:

- Scraping all defensive data from StatsBomb open dataset available on GitHub.
- Wrote a Python script to convert the JSON file into a CSV file.
    - The file contains various competitions, seasons and statistics.
    - The script was tailor made to extract only the relevant defensive data - coordinates on the pitch, type of defensive action, season, team and players.
- Using the above dataset to prepare 5 worksheets in Tableau:
    - Worksheet 1 - Defensive Action Heatmap
    - Worksheet 2 - Defensive Trends Pie Charts
    - Worksheet 3 - Player Comparison Action Points
    - Worksheet 4 - Top Defenders Bar Charts
    - Worksheet 5 - Individual Heatmaps
- Using the above worksheets to prepare 3 dashboards in Tableau (the dashboard names are as written in Tableau):
    - Dashboard 1 - 'General DB'
    - Dashboard 2 - 'Player DB'
    - Dashboard 3 - 'Heatmap DB'

Results or Findings of the Tasks:

The finding of the tasks can be summarized as 'Conclusions' for each dashboard. ' Part I' contains the screenshots of these 3 dashboards (3a,3b,3c). The conclusions are given below:

- **'General DB'**
    - Clearances happen near the goal.
    - Blocks are the most common form of defense.
    - Interceptions are less prominent.
- **'Player DB'**
    - Top Defenders are the best at clearances and denying goal scoring opportunities.
    - Real Madrid and Atletico Madrid are the best defensive teams.
    - They take up 18/20 places.
    - Real Madrid won 2/3 Champions Leagues indicating the importance of defense.
- **'Heatmap DB'**
    - It indicates the range of activity that a defender goes through and the area where that activity happens.

Some visual decisions were taken based on the following:

- **Heatmaps vs. Scatter Plots:** Initially tested scatter plots for spatial data but chose heatmaps because they better visualize action density, crucial for identifying high-impact defensive zones.

- **Bar Charts vs. Pie Charts:** Considered pie charts for ranking defenders but selected stacked bar charts to clearly display both total actions and type-specific contributions.
- **Line Charts vs. Area Charts:** Explored area charts for trends but opted for line charts due to their clarity in comparing multiple seasons without visual clutter. These choices were driven by the need for intuitive, data-driven visuals that a coach could easily interpret to make tactical decisions, aligning with the rubric's emphasis on proper visualization design.

LEE WING CHUN (3036212881)

Details of Tasks Completed:

- **Web Development:** Developed the HTML project page, which serves as the central hub for showcasing the project's dashboards and video visualization.
- **Data Extraction & Preprocessing:** Wrote the core Python web scraping script to extract comprehensive match-level data from Fbref.com, including match info, events, formations, and player statistics. The data are outputted to CSVs and formed the primary dataset for the Tableau dashboard on Match Analysis.
- **Tableau Dashboard:** Designed and implemented Dashboard 2: Match Level Analysis, which provides a detailed view of individual match events and metrics.
- **Report Writing:** Contributed to the drafting of the LIMITATIONS and ANYTHING BUT HAVEN'T sections, outlining project constraints and potential future work.

Results or Findings:

- **Dashboard 2 - Match Level Analysis:** This dashboard enables users to conduct in-depth analysis of single matches, with key findings presented on the fly. The dashboard's design, which includes an event timeline and tactical formations, allows for a clear understanding of how key moments and player positioning influenced a match's outcome. The primary purpose of this dashboard is to serve as a comprehensive tool for detailed match information, providing a factual and data-rich overview rather than being designed to test a specific hypothesis.
- **Data Extraction:** The Python script successfully automated the collection of a large volume of structured football data, which was essential for the project's analytical foundation. The custom-built script was a necessary component of the project since there is no consolidated

online data source that included the specific formation data required for the pitch visualization in Dashboard 2.

A M M Mutasim Moyen (3036383226)

Details of tasks completed:

- Analysed various datasets available online to source in-depth data to form visualizations

- Wrote a Python program to web crawl and scrape data from Fbref.com. Obtained and pre-processed player-specific data, including player statistics, discipline, performance metrics, and general player bio information. Formed the comprehensive CSV dataset from scraped data for use in Tableau to generate dashboard 4a and 4b.

- Designed and implemented dashboard 4a and 4b, which includes player bio information, match statistics, and disciplines. Created a player comparison bar chart that highlights the current chosen player in comparison to the rest of the team. Generated a line chart to display how each attribute affects performance metrics that can be filtered with various drop-down menus.

- Aided in drafting the report and the overall vision of the visualizations.

Results or Findings:

- A Python program web crawled and scraped to obtain data from various web pages to make a consolidated dataset including all player data, as there were no readily available datasets. This allowed us to make comprehensive player visualizations that can provide further insights for the user.

- Dashboards 4a and 4b provide insights into player performance and the reasons for the performance. This can allow the user to better position players and plan formations. The user can also use the data to decide on future match line-ups and even plan for transfers in upcoming seasons.

PAN JIA CHEN (3036383094)

Details of tasks completed:

- **Data analysis and preparation & cleaning:** Searching for and researching datasets with 2D coordinate information. The research was conducted on authoritative football websites such as Hugginface, Kaggle, Statbomb, and ChampionLeagueF. Quality analysis was performed on approximately 15 different datasets, which were then cleaned into the target format. Additionally, research was carried out on visualization solutions in the Python environment.

- **Attendance & Presentation**:
  Responsible for demonstrating the effects of Python dynamic visualization during the presentation. Also, participated in all complete reports and meetings of this project, as well as the entire life cycle of the dashboard (which ran from the early part of July to August 2nd).

- **Pythono Dashboard:** Independently completed all tasks related to the "**Python dynamic video dashboard**" section, including data visualization, data preparation, data research, data cleaning, data merging, and research on visualization solutions.

- **Report Writing:** Completed all explanations about the Python dashboard in the report, finished writing the contribution section, completed the writing of the data preparation and data cleaning sections, and wrote the content table of contents.

Results or Findings:

Completed the "visualization of shooting positions of different players in the European Championship", and successfully deployed the visualization effect on HTML using plot3D combined with web deployment.

I found that 90% of attacks are concentrated within a distance of 10-45 meters from the goal, among which nearly 75% (actually 76.33%) of goals are concentrated in the range where the angle between the line connecting to the goal and the x-axis is between -60 degrees and 60 degrees, showing a fan-shaped distribution. Moreover, different players have different preferences for shooting positions. For example, Mbappé prefers to shoot at a relatively close distance from the goal, while Messi likes to shoot at medium distances between the midfield and the penalty area, and some defenders prefer to shoot from very far away.

HOSSAIN MD TANJEEB (3036199560)

Dashboard 1: UCL League Overview

Details of tasks completed:

- Conducted extensive exploratory analysis of the UCL 2021/22 dataset to identify key metrics for a league-level overview, including points, goals scored, goals conceded and goal difference.

- Built a comprehensive league-level dashboard from scratch in Tableau. This involved cleaning and reshaping the data, creating calculated fields, and designing three visualisations: a bar chart ranking teams by total points, a bar chart comparing goal difference across squads, and a scatter plot showing the relationship between goals scored and points. I sorted and formatted each chart with clear labels and descriptive tooltips to highlight top teams and ensure readability.

- Reviewed dashboards produced by teammates to ensure consistent metrics and design, contributing feedback on high-level analysis, defensive strategies and player-level dashboards, and helped decide which metrics to emphasise in the final presentation.

- Actively participated in group discussions to interpret the data, structure the narrative of the report and integrate individual dashboards into a cohesive website and report.

- Researched football analytics literature to support our analysis and summarised key papers for the literature review section, particularly on expected goals and performance indicators.

Results or Findings:

1. The ranking of teams by total points highlighted clear leaders. Bayern Munich and Liverpool topped the chart with the highest point totals, closely followed by Real Madrid, Ajax and Manchester City. The distribution showed a sharp drop-off after the top five, revealing a significant gap between title contenders and mid-table squads.

2. Comparing goal differences by team revealed a similar pattern: clubs with the most points also had the largest positive goal differences, underscoring the link between offensive firepower, defensive solidity and overall success. Most teams near the bottom of the chart hovered around zero or negative goal difference, reflecting inconsistent performances.

3. The scatter plot of goals scored versus points demonstrated a strong positive relationship: teams that scored more goals generally accumulated more points. Outliers were easy to spot: a few sides amassed respectable point totals despite modest goal tallies (indicating defensive efficiency), while others scored frequently but did not convert chances into wins, hinting at defensive vulnerabilities or game-management issues.

4. These findings align with football analytics research. Studies have shown that winning teams typically score over two goals per match and average more shots on target and better passing accuracy than drawing or losing teams. The concept of expected goals (xG) further contextualises these results: xG models assign a probability to each shot based on factors like distance and angle, providing a benchmark for how many goals a team should score. By comparing actual goal difference with expected goal difference, analysts can identify teams that over- or under-perform relative to the quality of chances created.