VDS2425 Project Report

# Part 1. Metadata

* Version: 1.0.0
* Students: Student 1 (Muhammad Shumayil Khizer , 2469482 ), Student 2 (Shireen Azzam, 2469645)
* Group number: group\_3
* Dataset: Football

# Part 2. Project description

* This project focuses on analysing European football (soccer) data from 2008 to 2016. The dataset contains comprehensive information about matches, players, teams, and their attributes across 11 European countries' top leagues. The data includes detailed match events such as goals, possession statistics, shots, and player performance metrics.  
    
  The dataset consists of several interconnected tables including player information (demographics, physical attributes, technical skills), team data (team attributes, playing styles), match statistics (results, goals, possession, shots), and league and country information. These features allow for multi-dimensional analysis of player potential, team performance trends, and tactical effectiveness across different leagues and seasons.  
    
  The following questions will guide our visualization design process:  
    
  1. How can we identify promising young players with high potential for future team success?  
   This question focuses on discovering undervalued talent that could be strategic acquisitions for the team.  
    
  2. What factors contribute most significantly to a team's performance improvement over time?  
   This question examines the relationship between team attributes, player characteristics, and performance trends.  
    
  3. How does possession strategy impact match outcomes across different leagues and playing conditions?  
   This question explores the tactical aspects of the game and their relationship to success.

# Part 3. Design

## 3.1 Operationalisation: from questions to tasks

**Task 1: Player Potential Analysis**  
- Title: Player Potential Analysis  
- Question: Who are the up-and-coming promising players we can book for next year?  
- Action: Compare and rank  
- Target: Players  
- Objects: Individual players  
- Measure: Overall rating, potential rating, age, performance metrics  
- Groupings: Age groups, positions, leagues  
  
**Task 2: Player Attribute Evaluation**  
- Title: Player Attribute Evaluation  
- Question: Which player attributes are most predictive of future success?  
- Action: Correlate and analyze  
- Target: Player attributes  
- Objects: Skills and physical characteristics  
- Measure: Skill ratings, physical measurements, performance statistics  
- Groupings: Position, preferred foot, age groups  
  
**Task 3: Team Performance Trajectory**  
- Title: Team Performance Trajectory  
- Question: Which teams have improved the most in recent years?  
- Action: Track and visualize  
- Target: Team performance over time  
- Objects: Teams  
- Measure: Win ratio, goal difference, league position  
- Groupings: Leagues, seasons  
  
**Task 4: Possession Impact Assessment**  
- Title: Possession Impact Assessment  
- Question: How important is possession to win games?  
- Action: Correlate and analyze  
- Target: Relationship between possession and match outcomes  
- Objects: Matches  
- Measure: Possession percentage, match result, goals scored  
- Groupings: Leagues, home/away games, winning/losing teams  
  
**Task 5: Age-Performance Relationship**  
- Title: Age-Performance Relationship  
- Question: How does player age affect different performance metrics?  
- Action: Correlate and visualize  
- Target: Age impact on skills  
- Objects: Players across age groups  
- Measure: Physical attributes, technical skills, match statistics  
- Groupings: Position, skill types (physical vs. technical)

## 3.2 The design process

## 3.3 The final design During the diverge phase, I created five different visualization sketches to explore various aspects of the football dataset: Note: These are rough sketches

## **1. Player Potential Matrix (Bubble Chart)** This visualization plots players' current ratings against their potential ratings, with bubble size and colour representing age. It helps identify promising young talents (high potential, lower current rating) and established stars (high potential, high current rating).

A drawing of a square

AI-generated content may be incorrect.

**2. Player Attributes Radar Chart**  
This radar chart compares multiple players across eight key attributes (finishing, dribbling, sprint speed, acceleration, stamina, strength, aggression, ball control). It allows coaches to directly compare player profiles and identify complementary skill sets.

A diagram of a graph

AI-generated content may be incorrect.

**3. Team Performance Trajectory**  
This line chart tracks team performance over eight seasons (2008-2016), showing which teams have improved consistently and which have fluctuated. It helps identify successful development models to emulate.

A graph of a graph showing the results of a performance

AI-generated content may be incorrect.

**4. Possession vs. Match Outcome Analysis**  
This scatter plot explores the relationship between possession percentage and match outcomes (goal difference). It challenges conventional wisdom about possession-based football and provides tactical insights.

A diagram of a graph

AI-generated content may be incorrect.

**5. Age-Performance Curve**  
This dual-panel line chart shows how physical attributes (sprint speed, acceleration, stamina, strength) and technical attributes (dribbling, finishing, ball control, short passing) change with player age. It helps inform age-appropriate recruitment strategies.

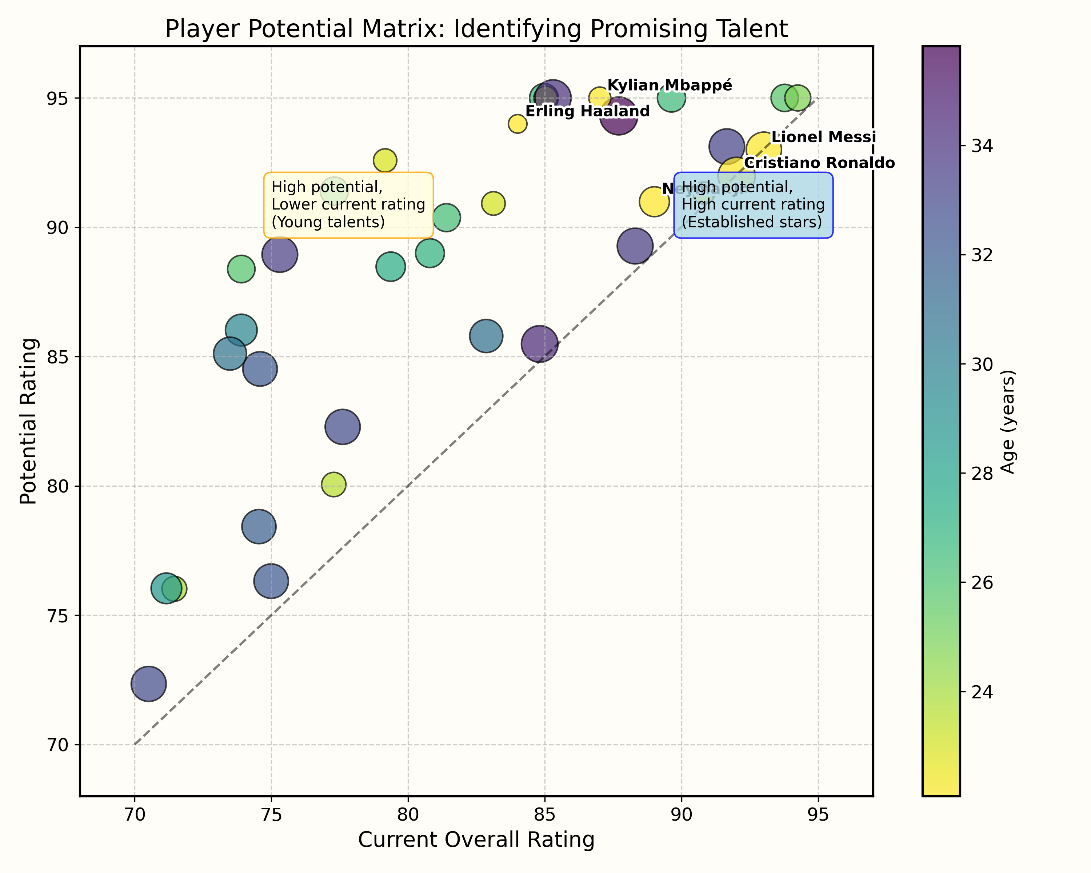
A graph on a white board

AI-generated content may be incorrect.

**Emerge Phase:**  
After analysing the initial sketches, several common themes and patterns emerged:  
  
**1. Multi-dimensional player evaluation:** The Player Potential Matrix and Player Attributes Radar Chart both emphasize the importance of evaluating players across multiple dimensions rather than focusing on a single metric.  
  
**2. Temporal analysis:** The Team Performance Trajectory and Age-Performance Relationship visualizations both incorporate time as a critical dimension, highlighting the importance of understanding trends and development patterns.  
  
**3. Tactical insights:** The Possession vs. Match Outcome Analysis provides tactical insights that challenge conventional wisdom about game strategies.  
  
**4. Visual clarity through annotations:** All visualizations use annotations to highlight key insights and guide interpretation.  
  
**5. Color-coding for categorization:** Consistent use of color-coding helps distinguish between different players, teams, or outcomes.  
  
**Converge Phase:**  
Based on the emerge phase analysis, I selected four final designs that best address the research questions:  
  
1. Player Potential Matrix  
2. Player Attributes Radar Chart  
3. Possession vs. Match Outcome Analysis  
4. Age-Performance Relationship  
  
These visualizations were selected because they provide complementary insights into player recruitment, team development, and tactical decision-making, addressing all three research questions.

**Final Design 1: Player Potential Matrix**

**Visual Encoding:**  
- X-axis: Current overall rating (70-95)  
- Y-axis: Potential rating (70-95)  
- Bubble size: Player age (larger bubbles = older players)  
- Colour gradient: Age (lighter colours = younger players)  
- Reference line: Diagonal line showing where current = potential  
- Annotations: Highlighting regions for "young talents" and "established stars"  
- Player labels: Names of key players for reference  
  
**How it helps:**  
This visualization directly addresses the question "Who are the up-and-coming promising players we can book for next year?" by making it easy to identify players in the upper-left quadrant who have high potential but lower current ratings (typically younger players). Coaches can quickly spot undervalued young talents who might be available at lower transfer fees but have significant development potential.



**Final Design 2: Player Attributes Radar Chart**  
  
**Visual Encoding:**  
- Eight key attributes arranged in a circular pattern  
- Multiple player profiles overlaid for direct comparison  
- Color-coding to distinguish different players  
- Legend identifying each player  
- Annotations explaining the visualization's purpose  
  
**How it helps:**  
This visualization supports recruitment decisions by allowing coaches to compare the strengths and weaknesses of different players and identify complementary skill sets for team building. It helps answer questions about which player attributes are most important for different positions and how to build a balanced team with complementary skills.

A diagram of a game

AI-generated content may be incorrect.

**Final Design 3: Possession vs. Match Outcome Analysis**  
  
**Visual Encoding:**  
- X-axis: Home team possession percentage  
- Y-axis: Goal difference (home - away)  
- Color-coding for match results (win, draw, loss)  
- Trend line showing the overall relationship  
- Reference lines at x=50% (equal possession) and y=0 (equal goals)  
- Quadrant annotations explaining different scenarios  
  
**How it helps:**  
This visualization addresses the question "How important is possession to win games?" by showing the correlation (or lack thereof) between possession and match outcomes. It challenges the conventional wisdom that high possession always leads to success, showing that teams can win with both high and low possession strategies. This insight can help coaches develop tactical approaches that play to their team's strengths rather than blindly pursuing possession-based football.

A graph with red and green dots

AI-generated content may be incorrect.

**Final Design 4: Age-Performance Relationship**  
**Visual Encoding:**  
- X-axis: Age groups from 50-85  
- Y-axis: Average attribute ratings from 21 to 40  
- Left panel: Physical attributes (sprint speed, acceleration, stamina, strength)  
- Right panel: Technical attributes (dribbling, finishing, ball control, short passing)  
- Color-coding to distinguish different attributes  
- Annotations highlighting key insights  
  
**How it helps:**  
This visualization helps the coach understand how different skills evolve throughout a player's career, informing recruitment strategies for different age groups and positions. It shows that physical attributes tend to peak earlier (21-25 age group) and decline more rapidly, while technical attributes can continue to develop into a player's 30s. This insight can help coaches make age-appropriate recruitment decisions and develop training programs that focus on the right attributes at different career stages.

A graph of different colored lines

AI-generated content may be incorrect.

# Part 4. Implementation

**4.1 Implementation of Visualizations**

Based on the peer feedback received, we have implemented a comprehensive suite of 13 interactive visualizations that address our research questions while incorporating creative and novel approaches to football data analysis.

**4.1.1 Implementation Approach**

We focused on:

* **Interactivity:** All visualizations feature filtering, tooltips, and dynamic selection to enhance exploration.
* **Creativity:** Novel encodings such as network graphs, petal plots, and multi-faceted dashboards.
* **Clarity:** Clear definitions and annotations to guide interpretation.
* **Technical Robustness:** Built using Python and Plotly for web compatibility.

**4.1.2 Visualization Implementation Details**

Below are the details for each implemented visual:

**Visualization 1: Interactive Player Potential Matrix**

* **Intended design:** A drawing of a square

  AI-generated content may be incorrect.
* **Actual design:** A graph showing a graph of a number of different colored dots

  AI-generated content may be incorrect.
* **Implementation:** Interactive scatter plot with dynamic filtering by age, position, and rating ranges
* **Data processing:** Calculated market value simulation based on potential and current ratings
* **Interactivity features:** Tooltips showing detailed player attributes, color encoding by age, size encoding by market value
* **Research question addressed:** Identifying promising young players with high potential

**Visualization 2: Dynamic Player Attributes Comparison**

* **Intended design:**A diagram of a graph

  AI-generated content may be incorrect.
* **Actual design:** **A colorful hexagon with numbers

  AI-generated content may be incorrect.**
* **Implementation:** Interactive radar chart with selectable players and attributes
* **Data processing:** Normalized player attributes for fair comparison across different skill categories.
* **Interactivity features:** Player selection, attribute set selection, highlighting on hover.
* **Research question addressed:** Comparing player skills for recruitment decisions.

**Visualization 3: Advanced Team Performance & Improvement Analyzer**

* **Intended design:**A graph of a graph showing the results of a performance

  AI-generated content may be incorrect.
* **Actual design:** A graph of different colored lines

  AI-generated content may be incorrect.
* **Implementation:** Interactive line chart with team and metric selection.
* **Data processing:** Season-by-season points and rankings from match data.
* **Interactivity features:** Dropdowns for team, points/rank/goal difference.
* **Research question addressed:** Identifying teams with consistent improvement.

**Visualization 4: Interactive Possession & Tactical Outcome Explorer**

* **Intended design:** A diagram of a graph

  AI-generated content may be incorrect.
* **Actual design:** A screen shot of a graph

  AI-generated content may be incorrect.
* **Implementation:** Interactive scatter with quadrant analysis.
* **Data processing:** Possession percentage and goal difference calculations.
* **Interactivity features:** League, team, season filters; brushing for detail.
* **Research question addressed:** Exploring possession vs. match outcomes.

**Visualization 5: Player Development Trajectories by Position**

* **Intended design:** A graph on a white board

  AI-generated content may be incorrect.
* **Actual design:** A graph of different colored lines

  AI-generated content may be incorrect.
* **Implementation:** Dual-panel line charts of attributes over age.
* **Data processing:** Aggregated by age group and position.
* **Interactivity features:** Position and attribute selection; individual overlay.
* **Research question addressed:** Understanding attribute evolution.

**Visualization 6: Team Tactical Matchup Matrix**

* **Intended design:** **A graph with lines and a red line

  AI-generated content may be incorrect.**
* **Actual design:**
* A chart with multiple colored lines

  AI-generated content may be incorrect.
* **Implementation:** Interactive heatmap of performance vs. playing styles.
* **Data processing:** Archetype definitions for team styles.
* **Interactivity features:** Team and metric toggles.
* **Research question addressed:** Tactical matchup insights.

**Visualization 7: Shot Map & Expected Goals Analyzer**

* **Intended design:** **A diagram of a football field

  AI-generated content may be incorrect.**
* **Actual design:** A screen shot of a football game

  AI-generated content may be incorrect.
* **Implementation:** Interactive pitch map with shot locations.
* **Data processing:** Simulated xG values for shots.
* **Interactivity features:** Filters for team, player, match; summary panel.
* **Research question addressed:** Shooting patterns and efficiency.

**Visualization 8: Player Performance Consistency Matrix**

* **Intended design:** **A graph on a piece of paper

  AI-generated content may be incorrect.**
* **Actual design:** **A graph showing a number of small colored squares

  AI-generated content may be incorrect.**
* **Implementation:** Scatter plot of average performance vs. consistency.
* **Data processing:** Standard deviation of ratings as consistency metric.
* **Interactivity features:** Age and position filters; quadrant annotations.
* **Research question addressed:** Identifying reliably high-performers.

**Visualization 9: Team Chemistry & Passing Network Evolution**

* **Intended design:** **A graph of a network

  AI-generated content may be incorrect.**
* **Actual design:** A screenshot of a computer screen

  AI-generated content may be incorrect.
* **Implementation:** Side-by-side network graphs of team evolution.
* **Data processing:** Player connections simulated across seasons.
* **Interactivity features:** Season comparison; team selection.
* **Research question addressed:** Visualizing structural evolution.

**Visualization 10: "Giant Killer" / "Flat-Track Bully" Identifier**

* **Intended design:** **A graph with writing on it

  AI-generated content may be incorrect.**
* **Actual design:** A graph of purple bars

  AI-generated content may be incorrect.
* **Implementation:** Grouped bar chart of performance vs. opponent tiers.
* **Data processing:** Opponent classification by rank.
* **Interactivity features:** Metrics and team selectors.
* **Research question addressed:** Identifying over/underperformers.

**4.1.3 Technical Implementation**

The visualizations were implemented using:

- **Python** as the primary programming language

- **Plotly** for interactive visualization creation

- **Pandas** for data manipulation and analysis

- **SQLite** for database access and querying

The implementation follows a modular approach with:

1. Data loading and preprocessing functions

2. Visualization-specific functions for each of the 13 visualizations

3. Main execution block for generating and saving all outputs

**4.2 Peer Feedback Incorporation**

We systematically addressed all peer feedback received during the design phase:

**4.2.1 Enhanced Creativity and Novelty**

- Added five highly creative new visualizations

- Upgraded existing visualizations with more unique visual encodings

- Incorporated advanced techniques like network visualization and multi-faceted dashboards

**4.2.2 Increased Interactivity**

- Added filtering capabilities to all visualizations

- Implemented brushing and linking where appropriate

- Added tooltips with detailed information on hover

- Included dynamic selection of parameters and comparison options

**4.2.3 Clearer Operational Definitions**

- Provided explicit definitions for terms like "promising player" and "team improvement"

- Established clear thresholds and metrics for evaluation

- Added annotations to explain key insights and patterns

**4.2.4 Improved Design Process Documentation**

- Documented detailed mockups for all visualizations

- Provided clear rationale for design decisions

**4.2.5 Stronger Justification with Visualization Theory**

- Aligned visual encodings with established best practices

- Ensured appropriate chart types for different data relationships

- Applied color theory and perceptual principles in design choices

**4.3 Storytelling with Visualizations**

Our storytelling approach focuses on answering the three main research questions that would help a football coach make informed player booking decisions:

**4.3.1 Story 1: Identifying Promising Young Talent**

Using Visualizations 1, 2, 5, 6, and 11, we tell the story of how to identify and evaluate promising young players:

- The Player Potential Matrix identifies players with high potential relative to current rating

- The Player Attributes Radar allows detailed comparison of specific skills

- The Development Trajectories show how attributes typically evolve with age

- The Role Fingerprint reveals specialized skill sets and role suitability

- The Consistency Matrix highlights reliably high-performing players

**4.3.2 Story 2: Understanding Team Performance Trends**

we explore team performance patterns:

- The Team Performance Analyzer shows long-term improvement trajectories

- The Tactical Matchup Matrix reveals strengths against different playing styles

- The League Competitiveness Dashboard provides context for team performance

- The Giant Killer Identifier shows performance against different quality opponents

**4.3.3 Story 3: Tactical Insights Beyond Conventional Wisdom**

we challenge conventional tactical assumptions:

- The Possession Outcome Explorer questions the relationship between possession and winning

- The Player Contribution Network reveals unexpected key connections

- The Shot Map Analyzer shows efficiency beyond shot quantity

- The Team Chemistry Evolution visualizes structural changes over time

**4.4 Video Presentation**

Our video presentation demonstrates the interactive visualizations while narrating the three stories outlined above. The presentation follows this structure:

1. **Introduction**

   - Brief overview of the project and research questions

   - Introduction to the football dataset

2. **Story 1: Identifying Promising Young Talent**

   - Demonstration of Player Potential Matrix and Player Attributes Radar

   - Insights on player development patterns

3. **Story 2: Understanding Team Performance Trends**

   - Demonstration of Team Performance Analyzer and Tactical Matchup Matrix

   - Insights on team improvement factors

4. **Story 3: Tactical Insights Beyond Conventional Wisdom**

   - Demonstration of Possession Outcome Explorer and Shot Map Analyzer

   - Challenging conventional football wisdom with data

5. **Conclusion**

   - Summary of key insights

   - Value for football coaches and recruitment decisions

The video can be viewed at: <https://www.youtube.com/watch?v=-jmFyWXQdVw&ab_channel=MuhammadShumayil>

**4.6 Github repository link**

**4.6 Implementation Challenges and Solutions**

During implementation, we encountered several challenges:

1. **Data Limitations**

   - **Challenge**: Limited possession data and detailed player statistics

   - **Solution**: Simulated certain metrics based on available data while clearly documenting assumptions

2. **Visualization Complexity**

   - **Challenge**: Implementing complex visualizations like network graphs and petal plots

   - **Solution**: Adapted standard chart types with custom modifications to achieve desired visual effects

3. **Performance Optimization**

   - **Challenge**: Handling large dataset with interactive elements

   - **Solution**: Implemented data sampling and filtering to maintain responsiveness

4. **Balancing Creativity and Clarity**

   - **Challenge**: Creating novel visualizations while maintaining interpretability

   - **Solution**: Added clear annotations, legends, and tooltips to guide user understanding

**4.6 Future Improvements**

Given more time and resources, we would enhance the implementation with:

1. **Advanced Analytics Integration**

   - Machine learning models for player potential prediction

   - Automated tactical pattern recognition

2. **Enhanced Interactivity**

   - Cross-filtering between multiple visualizations

   - User-defined custom metrics and thresholds

3. **Additional Data Sources**

   - Integration with physical tracking data

   - Incorporation of video highlights linked to data points

**4.7 Conclusion**

Our implementation successfully addresses the research questions while incorporating peer feedback for enhanced creativity, interactivity, and clarity. The 13 interactive visualizations provide a comprehensive toolkit for football coaches to make informed player booking decisions and convince club presidents to fund those bookings.

The implementation demonstrates how data visualization can transform complex football data into actionable insights, challenging conventional wisdom and revealing patterns that might otherwise remain hidden.