Sports Analysis of The English Premier League

CSE3020 – DATA VISUALISATION

PROJECT BASED COMPONENT REPORT

By

Aayush Singh

20BCE0634

Atharva Parkale

20BCE0480

Biggyat Kumar Pandey

20BCE2763

School of Computer Science and Engineering

May 2021

DECLARATION

I hereby declare that the report entitled "Sports Analysis of The English Premier League" submitted by me, for the CSE3020 DATA VISUALISATION (EPJ) to VIT is a record of bonafide work carried out by me under the supervision of Dr. S.VENGADESWARAN .

I further declare that the work reported in this report has not been submitted and will not be submitted, either in part or in full, for any other courses in this institute or any other institute or university.

Place   : Vellore

Date   : 10.12.2021

Signature of the Candidate

CONTENTS

Abstract

Background

This project lies in the domain of sports analytics. Sports analytics are a collection of relevant, historical, statistics that can provide a competitive advantage to a team or individual. Through the collection and analysis of this data, sports analytics inform players, coaches, and other staff in order to facilitate decision-making both during and prior to sporting events. The sports industry uses sports analysis to increase revenue, improve player performance and a team's quality of play, prevent injury, and for many more enhancements.

Since this project focuses on the analysis of a football league, statistics like the number of shots, shots on target, and ball possession, can offer a guide to how two teams have performed. If we then improve the granularity of the stats and include selected performance indicators, we can get an even more detailed and insightful report.

Problem Statement

This project is centered around the sports performance analysis of one of the most popular football leagues in the world, the English premier league (2020/2021 season). The project comprises various tasks ranging from data collection all the way to the delivery of feedback to the user through the dashboard and aims to improve performance for the players and coaches.

Procedure

Data Abstraction

The majority of the data has been collected from the official premier league website. A few other datasets have been collected from data aggregating websites like Kaggle. This data was then inserted in the appropriate datasets and used to implement various graphs, charts and plots to effectively and efficiently allow the user to summarize, compare and present the statistical information.

Task Abstraction

Various Interactive plots, charts and graphs like Bar plots, Pie Charts, Radar Charts, and Scatter Plots have been used for the appropriate statistics in order to effectively convey the information provided in the datasets.

Further, Correlation of various statistics has been calculated and represented in order to exhibit the relationship between the statistics. The interactive nature of the graphs allows the user to search, zoom and see the exact values. The relevant data is also displayed in a table alongside the graphs

Design and Dashboard Implementation

The dashboard has been implemented using flexdashboard and it is designed such that it is graphically distinctive as well as functionally efficient. All the statistics are displayed in the sidebar that leads to a page containing the graphs and tables relevant to that statistic. The graphs, plots and charts have been displayed in a tabset column to effectively use the available space.

Conclusion

The dashboard successfully exhibits the statistics in a visually brilliant and comprehensive manner using interactive graphs, plots and charts. The insight provided by the dashboard can be used by the coaching stuff of a team to monitor improvements in game plans, team strategies and player performance. Players can also use the information provided to enhance development, game understanding and decision making by enabling players to view and analyze their own performance, and identify their strengths and weaknesses. In addition to this, casual viewers would also greatly benefit from this dashboard as it would aid their understanding of the game and allow them to compare and summarize comprehensive details and relevant statistics of the entire season at a glance.

INTRODUCTION TO THE PROJECT

OBJECTIVE

The objective of this project is to design a dashboard that represents statistical data from the 2020/2021 season of the English premier league in an accurate, efficient, and elegant way. The dashboard must also display relationships between various analytics in order to convey prompt and clear insights to the user. This data must be represented in a concise and intuitive visual manner using diverse plots, charts, and graphs that can provide relevant information at a glance.

Since this project focuses on the analysis of a football league, statistics like the number of shots, shots on target, and ball possession, can offer a guide to how two teams have performed. If we then improve the granularity of the stats and include selected performance indicators, we can get an even more detailed and insightful report.

Through the collection and analysis of this data, the sports analytics dashboard can empower players, coaches, and other staff with the knowledge required to facilitate decision making both during and prior to sporting events. The sports industry uses sports analysis to increase revenue, improve player performance and a team's quality of play, prevent injury, and for many more enhancements. Further use cases have been elaborated in the problem statement.

PROBLEM STATEMENT

This project is centered around the sports performance analysis of one of the most popular football leagues in the world, the English premier league (2020/2021 season). The project comprises various tasks ranging from data collection all the way to the delivery of feedback to the user through the dashboard and aims to improve performance for the players and coaches. The dashboard to be created must accurately and efficiently display relevant statistics which can then be used by various users. The use cases for this dashboard can be divided into three categories- For coaches, for players, and for sports enthusiasts. These use cases are further elaborated in the given table:

For Coaches	For Players	For Sports enthusiasts
Having a record of both team and player performance is invaluable when monitoring improvements in game-plans, team strategies and player performance. Analysis also facilitates an objective viewpoint that can help in decision making, developing targets and training methods	Enhances development, game understanding and decision making by enabling players to view and analyse their own performance, and identify their strengths and weaknesses.	Allows for a comprehensive understanding of the sport and highlights relevant insights through visualization. For example, Having statistics like possession, shot accuracy, passes etc. can help determine which team performed better.

FUNCTIONAL REQUIREMENTS

We aim to create a dashboard which represents the relevant statistical data visually in chart, plots, and graphs and displays immediate and accurate feedback of the team and player performance by allowing the user to visualize and compare the given data. Thus in order to provide a comprehensive understanding for the dashboard must exhibit statistics like:

	Player statistics
Offense	Number of total shots, shots on target, goals, Goal to shot percentage, Assists, Passes, Pass accuracy, Touches, Offsides
Defence	Tackles, Red cards/Yellow cards, Fouls, Corners, Interceptions

Team Statistics
Matches played,Possession %, Cumulative player statistics (shots, goals, passes etc.), Points, League position, Wins/losses/draws ,Home wins/losses/draws, Away wins/losses/draws

DATA ABSTRACTION

Data abstraction is a principle of data modeling theory that emphasizes the clear separation between the external interface of objects and internal data handling and manipulation. In many programming languages, interfaces (or abstract classes) provide abstraction, and their concrete implementations form the implementation.

In simple words it is the reduction of a particular body of data to a simplified representation of the whole.

Abstraction, in general, is the process of taking away or removing characteristics from something in order to reduce it to a set of essential characteristics.

The picture represents the data frame which is the leaderboard for the all time top 10 goal scorers of the English Premier League

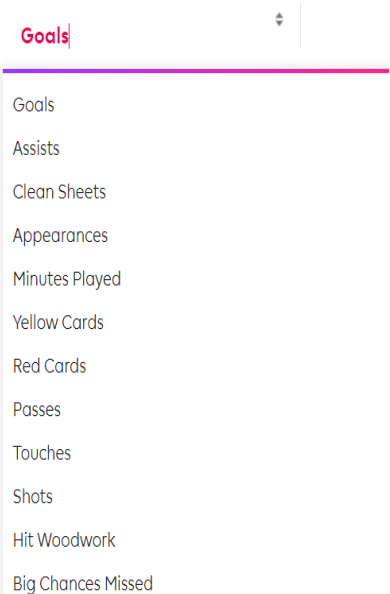It is in a table format with namely five attributes and they are:

Rank : for the rank of the player

Player name : for the name of the player
Club(current) :for current club of the player (if playing)
Nationality : for the nationality of the player
Stats : for the actual number

	Rank	Player	Club	Nationality	Stat
Goals	1.	Alan Shearer	-	England	260
Assists	2.	Wayne Rooney	Derby County	England	208
Clean Sheets	3.	Andrew Cole	-	England	187
Appearances	4.	Sergio Agüero	FC Barcelona	Argentina	184
Minutes Played	5.	Frank Lampard	-	England	177
Yellow Cards	6.	Thierry Henry	-	France	175
Red Cards	7.	Harry Kane	Tottenham Hotspur	England	167
Passes	8.	Robbie Fowler	-	England	163
Touches	9.	Jermain Defoe	Rangers	England	162
Shots	10.	Michael Owen	-	England	150
Hit Woodwork					
Big Chances Missed					

Here each attribute has around 10 elements from which it is derived that the data set used is a 5 x 10 discrete table with different levels of measurement of the data that are:
- Rank : interval
- Player name : nominal
- Club : nominal
- Nationality : nominal
- Stats : ratio

All the other attributes are nominal like the player name, club, nationality because they are characterized in names whereas for rank is interval as it is characterized in a way we can tell the exact difference between two values and for stats its ratio as we can tell the exact difference and it can be zero for different cases

The picture represents the data frame which is the leaderboard for the 10 clubs with the most yellow cards in the previous edition of the English Premier League

It is in a table format with namely five attributes and they are:

Rank : for the rank of the clubs

club name : for the name of the club

games played:total number of games played

goals against: the goals the have conceded for the entire season

yellow cards : the number of yellow cards issued to the club

Rank	Club Name	Games Played	Goal Against	Yellow Card
1	Sheffield United	38	63	73
2	Aston Villa	38	46	70
3	Fulham	38	53	67
4	Newcastle United	38	62	65
5	Manchester United	38	44	64
6	Leicester City	38	50	61
7	Leeds United	38	54	61
8	Everton	38	48	59
9	Tottenham Hotspur	38	45	57

Here each attribute has around 10 elements from which it is derived that the data set used is a 5 x 10 discrete table with different levels of measurement of the data that are:

- Rank : interval
- club name : nominal
- games played: ratio
- goals against: ratio
- yellow cards : ratio

mostly of the attributes are ratio like the games played,goals against, yellow cards because we can tell the exact difference and it can be zero for different cases whereas for rank which is interval as it is characterized in a way we can tell the exact difference between two values

The picture represents the data frame which is the leaderboard for the 10 clubs with the most shots and goals in the previous edition of the English Premier League

It is in a table format with namely four attributes and they are:

Rank : for the rank of the clubs

club name : for the name of the club

shots:total number of shots on and off target

goals : the goals scored

Rank	Club	Goals	Shots
1	Manchester City	83	599
2	Manchester United	73	526
3	Leicester City	68	485
4	Liverpool	68	608
5	Chelsea	58	556
6	West Ham United	62	468
7	Arsenal	55	459
8	Leeds United	62	522
9	Tottenham Hotspur	68	443
10	Everton	47	400

Here each attribute has around 10 elements from which it is derived that the data set used is a 4 x 10 discrete table with different levels of measurement of the data that are:

- Rank : interval
- club name : nominal
- shots: ratio
- goals : ratio

most of the attributes are ratio like the goals, shots because we can tell the exact difference and it can be zero for different cases whereas for rank which is interval as it is characterized in a way we can tell the exact difference between two values

The picture represents the data frame which is the leaderboard for the 10 players with the most passes in the previous edition of the english Premier League

It is in a table format with namely five attributes and they are:

Rank : for the rank of the clubs

player name : for the name of the club

club:the club of the player

nationality: players nationality

stas : the number of passes

	Rank	Player	Club	Nationility	Stat
1	Rank	Player	Club	Nationility	Stat
2	1	João Cancelo	Manchester City	Portugal	964
3	2	Virgil van Dijk	Liverpool	England	924
4	3	Lewis Dunk	Brighton and Hove Albion	England	872
5	4	Rodri	Manchester City	Spain	855
6	5	Pierre-Emile Højbjerg	Tottenham Hotspur	Denmark	843
7	6	Marc Guéhi	Crystal Palace	England	821
8	7	Joachim Andersen	Crystal Palace	Denmark	815
9	8	Rúben Dias	Manchester City	Portugal	793
10	9	Liam Cooper	Leeds United	Scotland	761
11	10	Çaglar Söyüncü	Leicester City	Turkey	760

Here each attribute has around 10 elements from which it is derived that the data set used is a 5 x 10 discrete table with different levels of measurement of the data that are:

- Rank : interval
- player name : nominal
- club:nominal
- nationality: nominal
- stats : ratio

most of the attributes are nominal like the nationality, club ,player name as they are characterized by names whereas for stats it is rank because we can tell the exact difference and it can be zero for different cases whereas for rank which is interval as it is characterized in a way we can tell the exact difference between two values

These are just few types of data frame that we abstracted in the similar manner we have analyzed various data sets for this project with various levels of measurement and table
similarly the data sets used at different level for various plots were data from csv files and obtained through web scraping the following url:

FOR GOALS SHOTS AND TOUCHES
https://github.com/AtharvaParkale/ZER01/tree/master/Aayush%20code
FOR PASSES ASSISTS AND FOULS
https://github.com/AtharvaParkale/ZER01
FOR RED CARDS AND YELLOW CARDS

https://sports.ndtv.com/english-premier-league/stats/most-yellow-cards-player-statsdetail/2020-21

https://sports.ndtv.com/english-premier-league/stats/most-yellow-cards-team-statsdetail/2020-21

https://sports.ndtv.com/english-premier-league/stats/most-red-cards-player-statsdetail/2020-21

https://sports.ndtv.com/english-premier-league/stats/most-red-cards-player-statsdetail/2020-21"
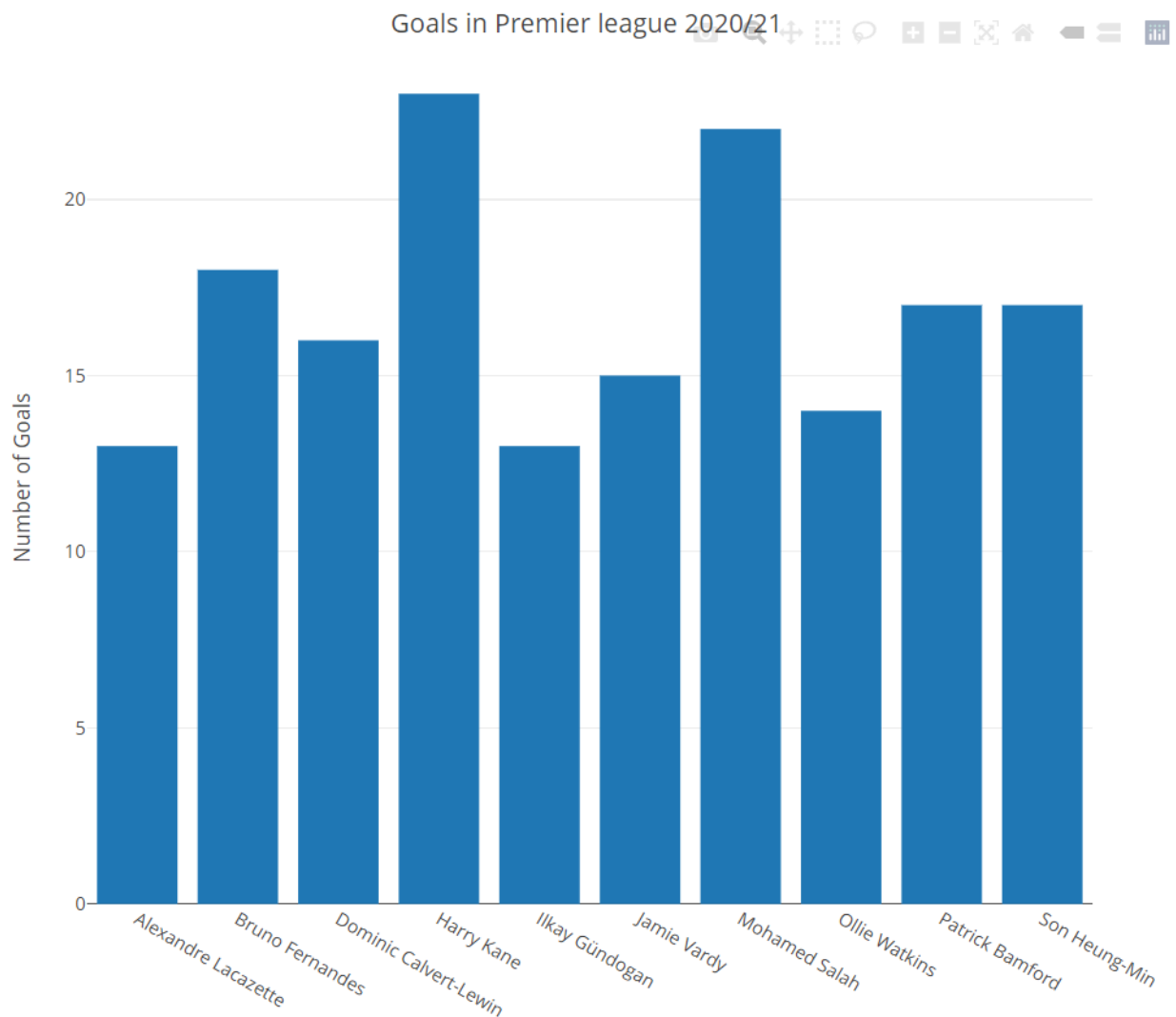
TASK ABSTRACTION

Statistic: Goals

Barplot representing the number of goals by Player

Actions performed:

1. Discover ,present and Enjoy the data in an engaging and easy to understand visualization model

2. Identify, Summarize and Compare the goals by the top ten players in the season

3. Search for the number of goals scored by a player

Target: Distribution of the number of goals by the top 10 players

The data is arranged by aligning it against a common scale for easy comparison and visualization
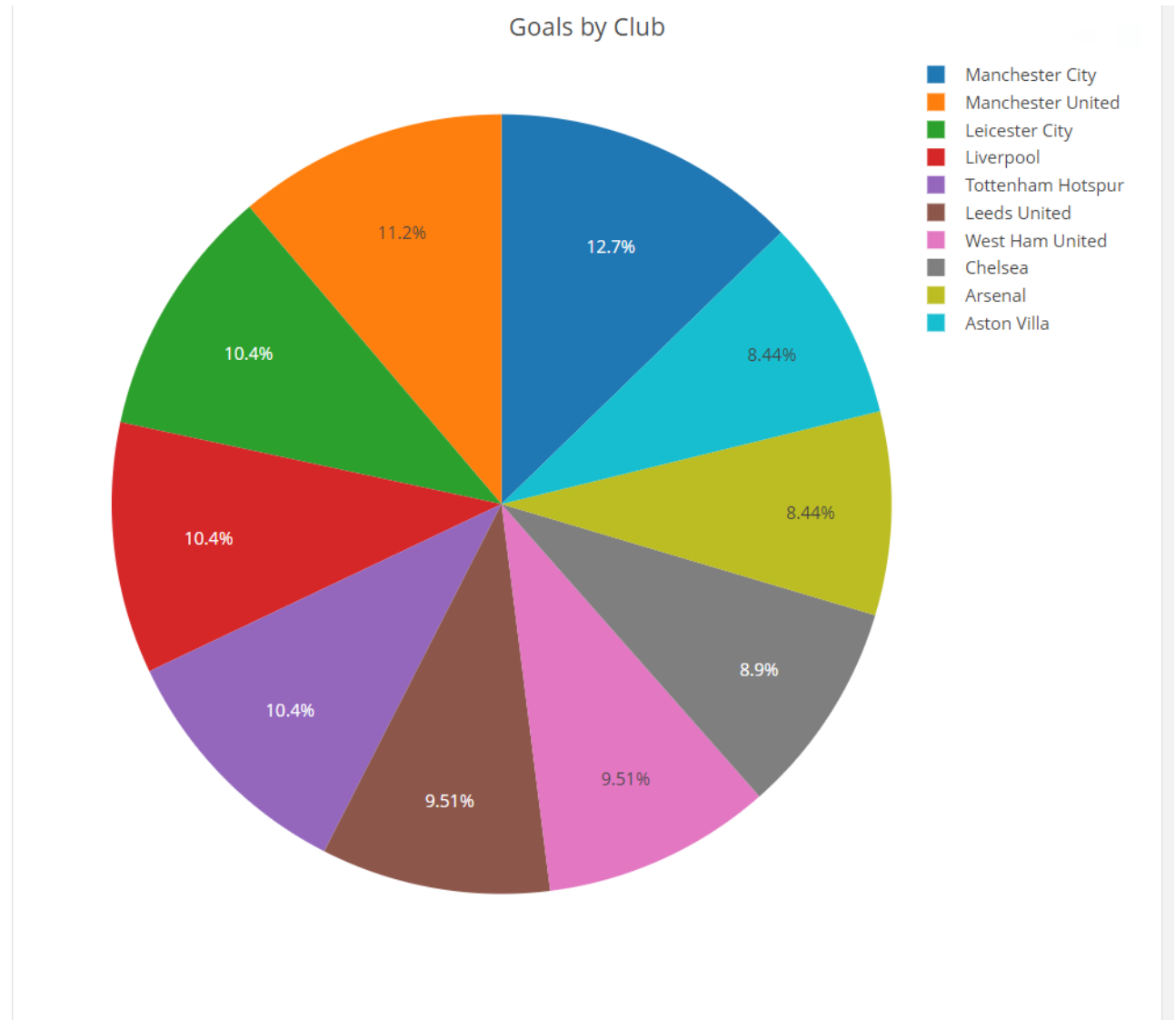
Statistic: Goals

Pie chart representing the number of goals by Club

Actions performed:

1. Discover , present and Enjoy the data in an engaging and easy to understand visualization model

2. Identify, Summarize and Compare the goals scored by the top ten clubs in the season

Target: Distribution of the number of goals by the top 10 clubs

The categorical data i.e team names have been mapped using different colours

Goals by Club

Statistic: Goals Conceded per game vs Clean sheets by defender

Scatter plot representing the number of goals conceded per game, clean sheets, and tackles by a defender.
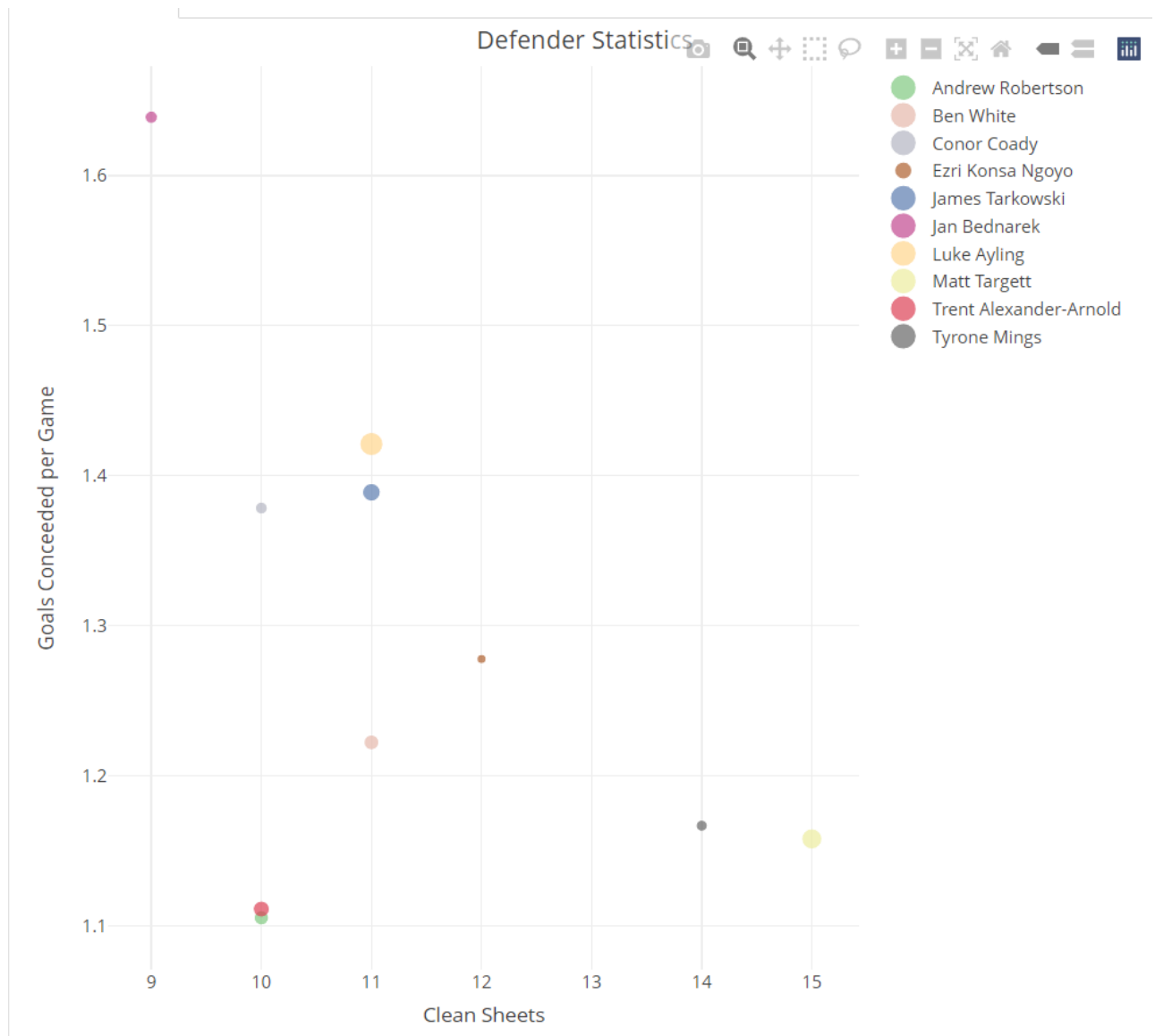
Here the dataset was first narrowed down based on the position such that only statistics of the defenders with the most appearances are considered. Goals conceded per game is derived data, derived from the number of goals conceded and the number of games.

The key defence statistics are goals conceded per game, clean sheets and number of tackles. The relationship between these 3 is represented in the scatterplot where goals conceded per game is the x axis, Clean sheets is the y axis and number of tackles is the size of the point. Each point is color coded to represent the categorical variable defender name.

Actions performed:

1. Discover ,present and Enjoy the data in an engaging and easy to understand visualization model

2. Identify, Summarize and Compare the goals conceded per game, cleans sheets and tackles by the top defenders in the season

3. Search for the number of goals conceded per game, clean sheets and tackles by a defender. Derive other data like the relationship between clean sheets and tackles.

Target: This visualization allows us to identify Trends, Outliers, Similarity and Features . The quantitative values are arranged separately in the scatterplot. The channels size and color are used to map further quantitative and categorical values respectively.
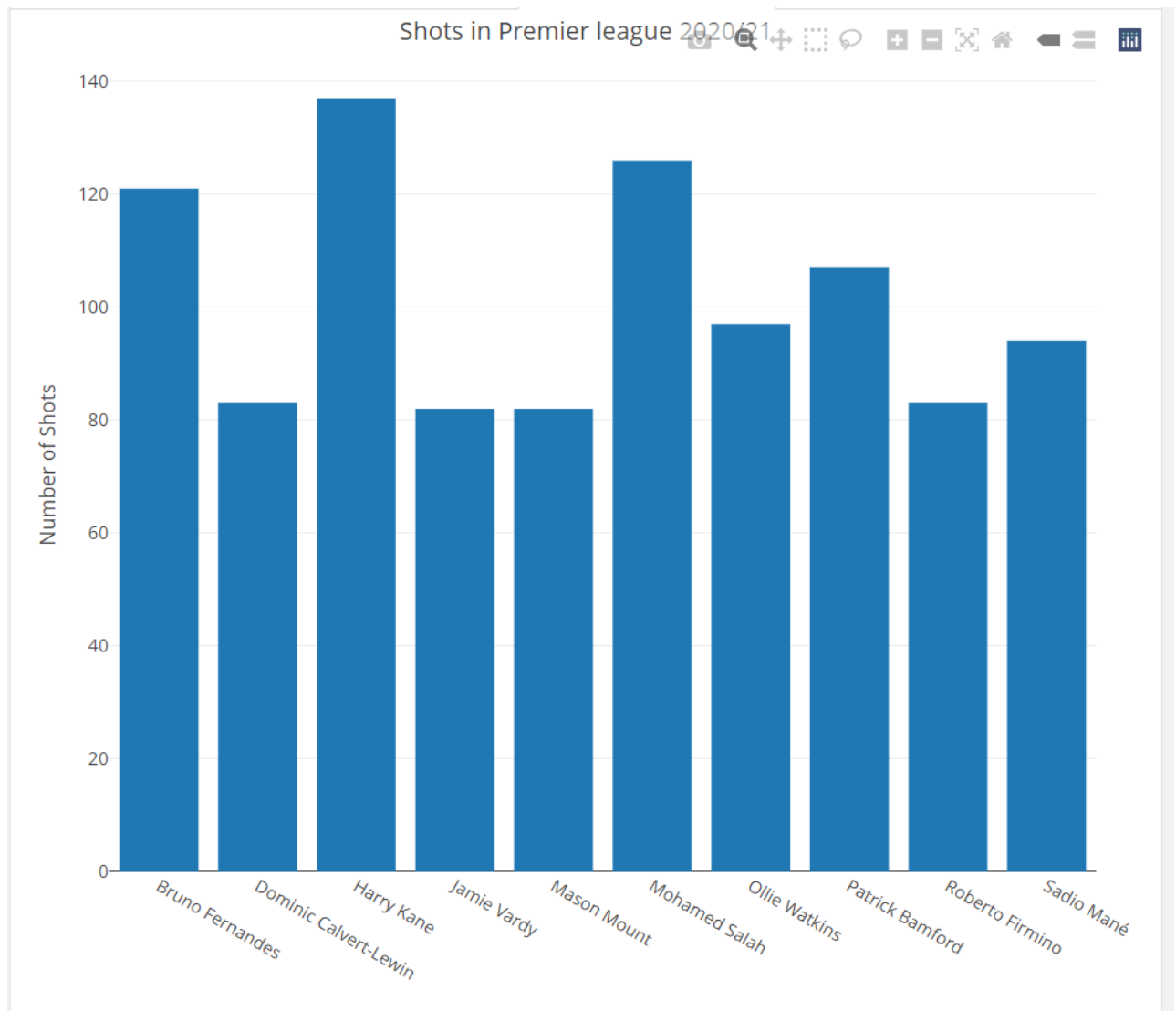
Statistic: Shots

Barplot representing the number of Shots by Player

Actions performed:

1. Discover ,present and Enjoy the data in an engaging and easy to understand visualization model

2. Identify, Summarize and Compare the shots by the top ten players in the season

3. Search for the number of Shots by a player

Target: Distribution of the number of Shots by the top 10 players

The data is arranged by aligning it against a common scale for easy comparison and visualization
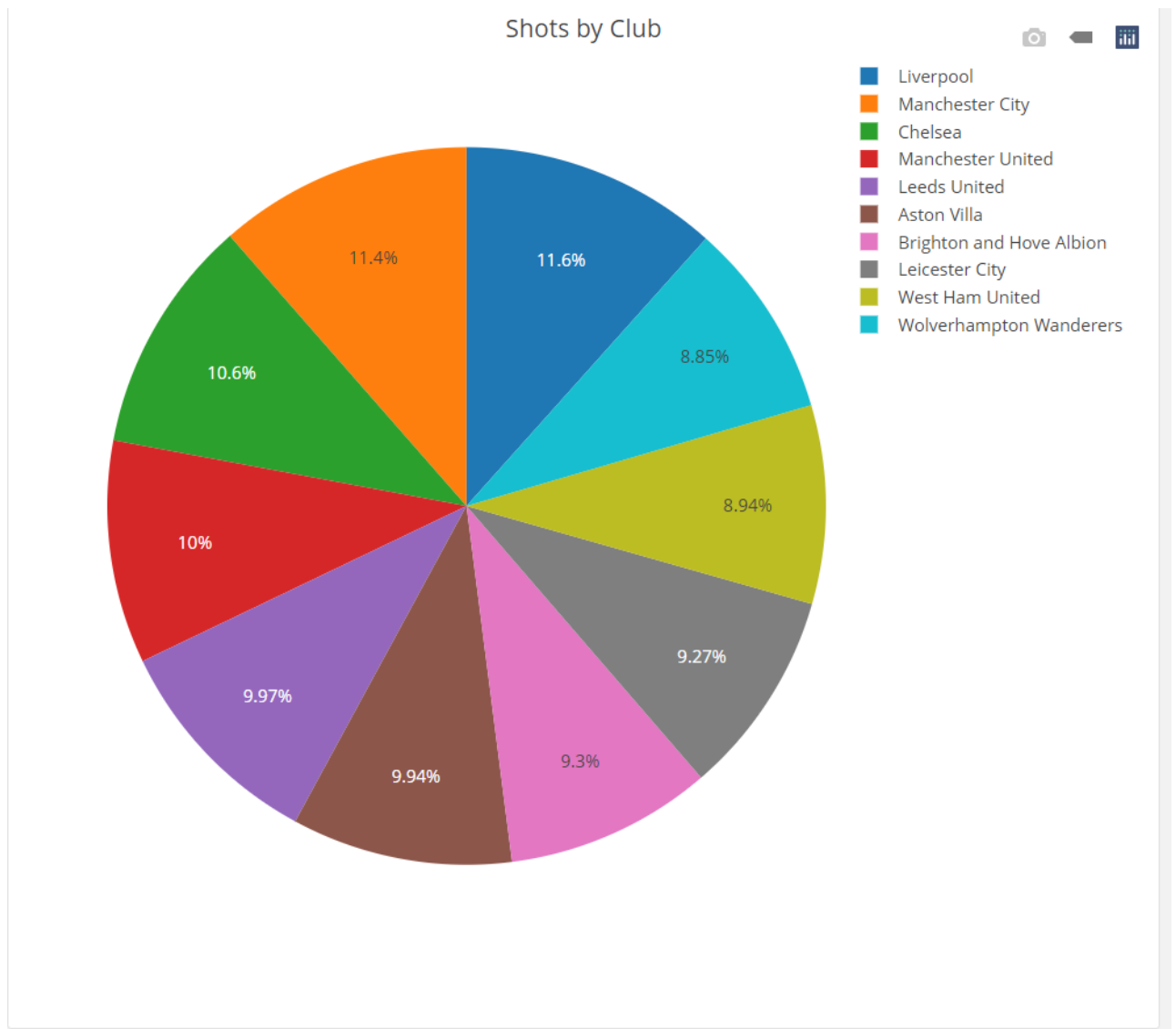
Statistic: Shots

Pie chart representing the number of shots by Club

Actions performed:

1. Discover ,present and Enjoy the data in an engaging and easy to understand visualization model

2. Identify, Summarize and Compare the shots by the top ten clubs in the season

Shots by Club

Legend:
- Liverpool
- Manchester City
- Chelsea
- Manchester United
- Leeds United
- Aston Villa
- Brighton and Hove Albion
- Leicester City
- West Ham United
- Wolverhampton Wanderers

Values shown: 11.6%, 11.4%, 10.6%, 10%, 9.97%, 9.94%, 9.3%, 9.27%, 8.94%, 8.85%

Statistic: Correlation between shots and wins

Pearson correlation between shots and goals

Allows us to know if more shots leads to more goals. Pearson correlation is performed between the statistics and a scatterplot is plotted. A regression line allows the user to better visually understand the correlation. The statistics used here are the cumulated club statistics of the top 10 clubs with the most goals in the season.

This visualization allows us to identify Trends, Outliers, and Features in the graph as well as depicting the correlation, comparison, and presentation of the data. We can see a positive correlation here telling us more shots leads to more goals. This can help in decision making and match prediction

```
[1] "Correlation: 0.65"
```

Relation between number of Shots and Goals
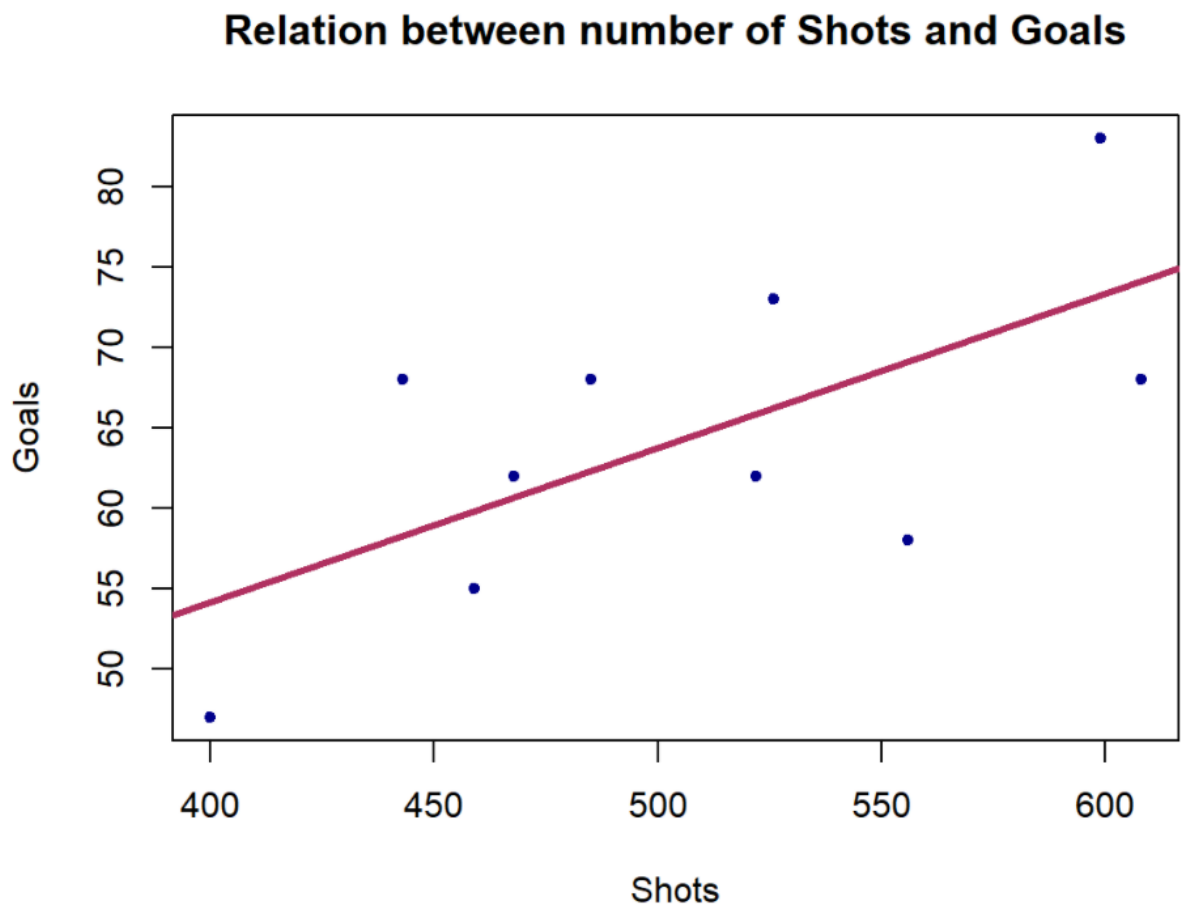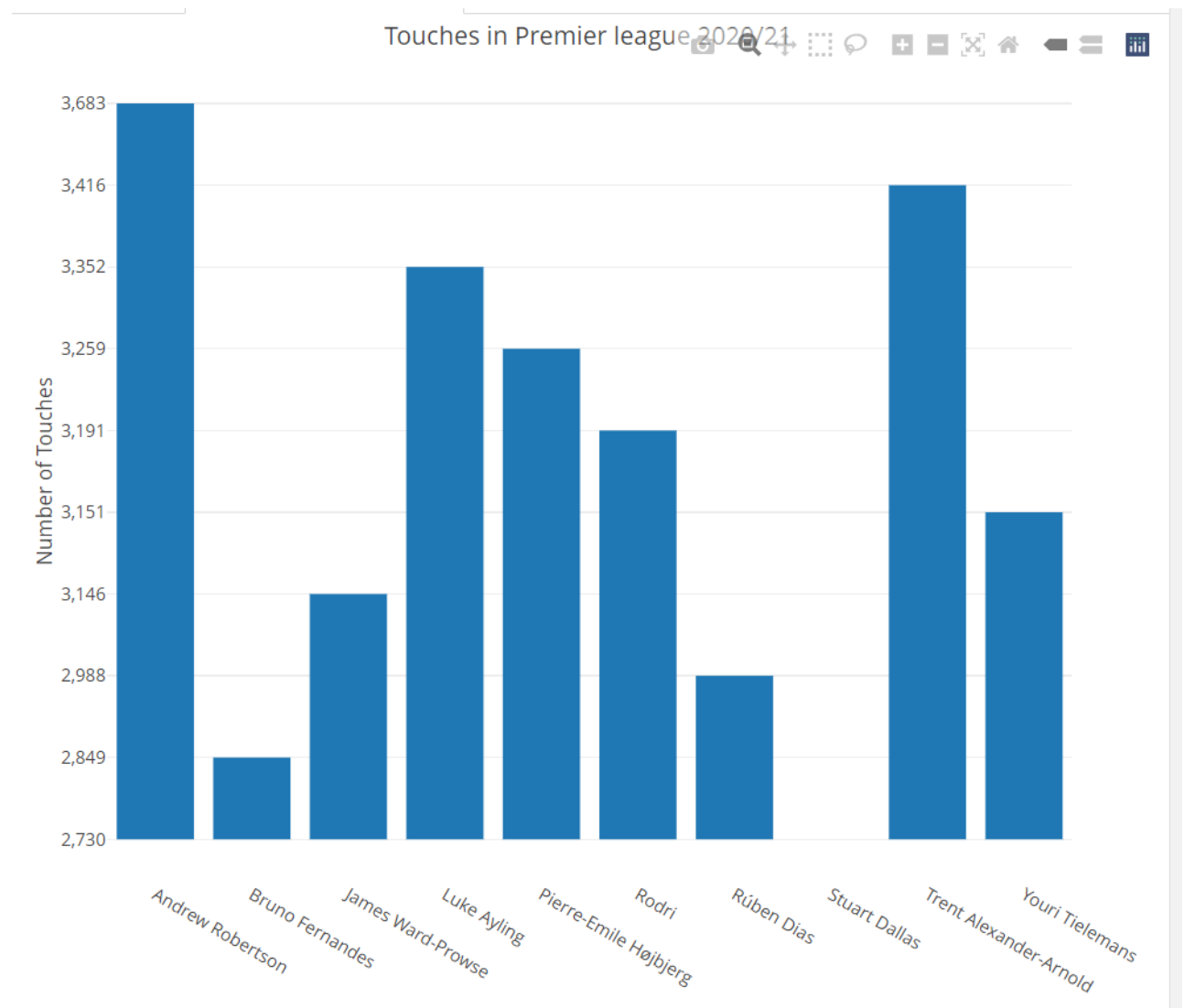

Statistic: Touches

Barplot representing the number of touches by Player

Actions performed:

1. Discover ,present and Enjoy the data in an engaging and easy to understand visualization model

2. Identify, Summarize and Compare the touches by the top ten players in the season

3. Search for the number of touches by a player

Target: Distribution of the number of touches by the top 10 players

The data is arranged by aligning it against a common scale for easy comparison and visualization

Touches in Premier league 2020/21

Statistic: Touches

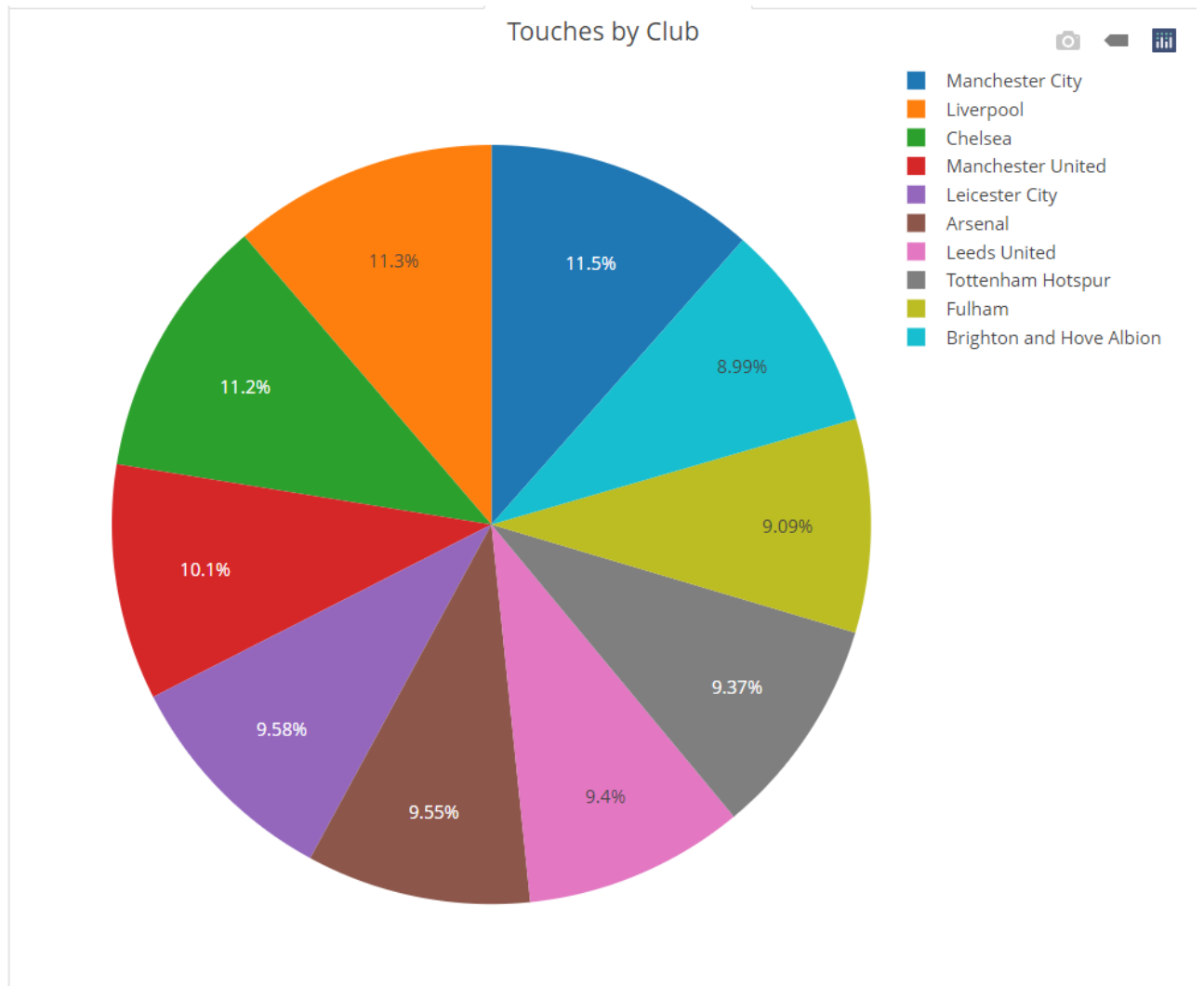**Pie Chart representing the number of touches by Club**

Actions performed:

1. Discover ,present and Enjoy the data in an engaging and easy to understand visualization model

2. Identify, Summarize and Compare the touches by the top ten clubs in the season

Target: Distribution of the number of touches by the top 10 clubs

The categorical data i.e team names has been mapped using different colours

Touches by Club

Legend:
- Manchester City
- Liverpool
- Chelsea
- Manchester United
- Leicester City
- Arsenal
- Leeds United
- Tottenham Hotspur
- Fulham
- Brighton and Hove Albion

Values shown: 11.5%, 11.3%, 11.2%, 10.1%, 9.58%, 9.55%, 9.4%, 9.37%, 9.09%, 8.99%

Statistic: Player Statistics

Radar Chart depicting the Goals per match, Passes per match, Shooting accuracy and Appearances of a player.

The radar chart considers 5 individual statistics that allow us to holistically judge a player's performance. The radar chart considers a maximum value of each statistic as the maximum value in the dataset. The individual statistics of any player is then compared with the maximum values to obtain the radar chat.

The statistics of multiple players can be displayed together for easy comparison. This chart can also help identify the strengths and weaknesses of a player or a team.

Actions performed:

1. Discover ,present and Enjoy the data in an engaging and easy to understand visualization model

2. Identify, Summarize and Compare the statistics of all player

3. Search for the statistics of a player.

Target: Analyse the features of player statistics using spatial data.

The data is arranged by aligning it against a common spatial arrangement for easy comparison and visualization.

Player Stats for Top 3 Goal Scorers
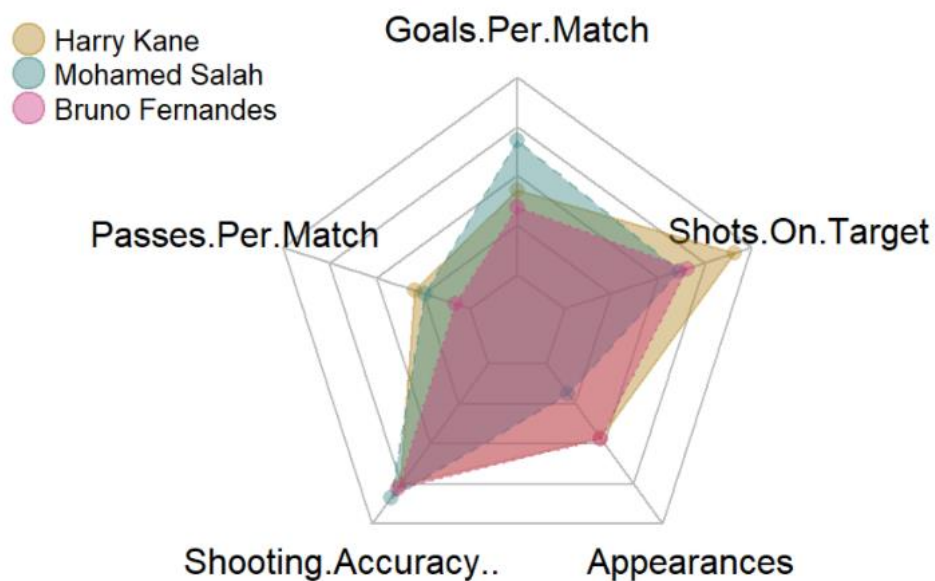
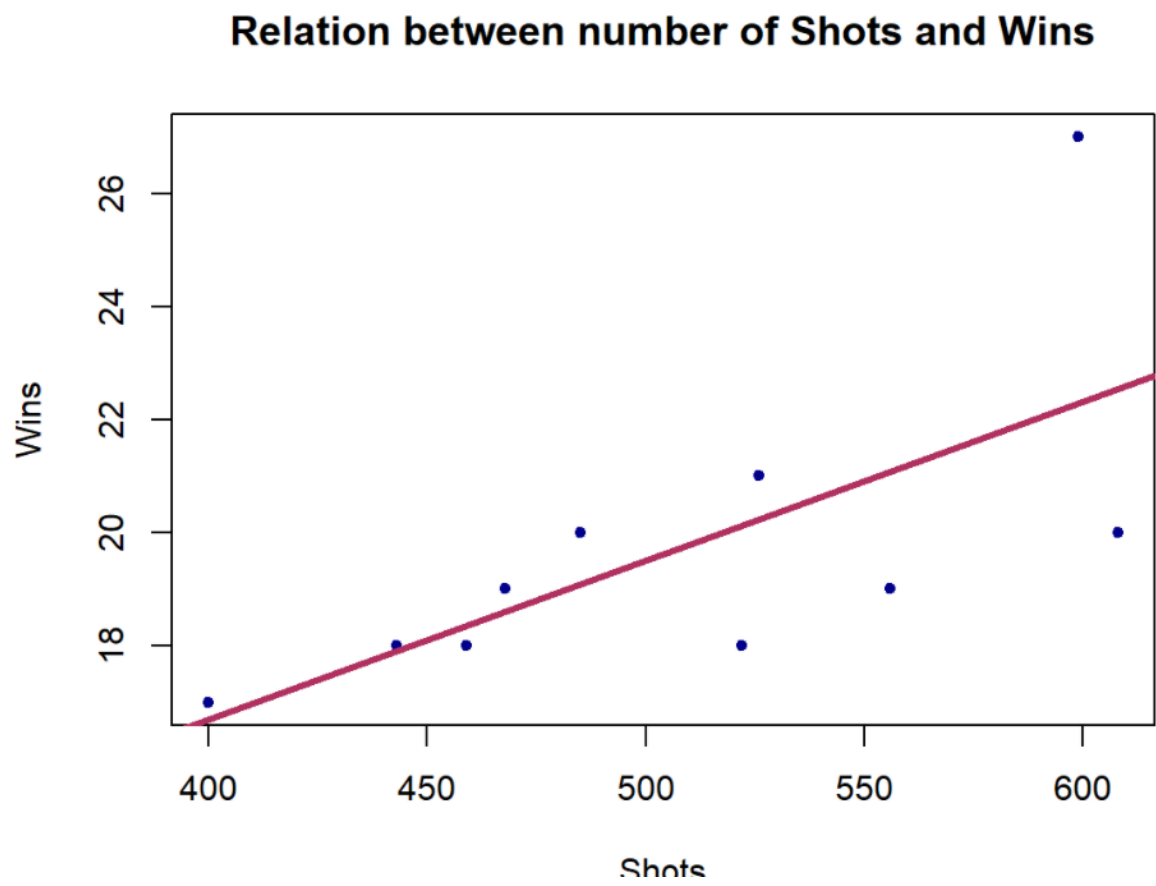**Statistic: Correlation with wins**

Pearson correlation between wins and:

Shots

```
[1] "Correlation: 0.67"
```
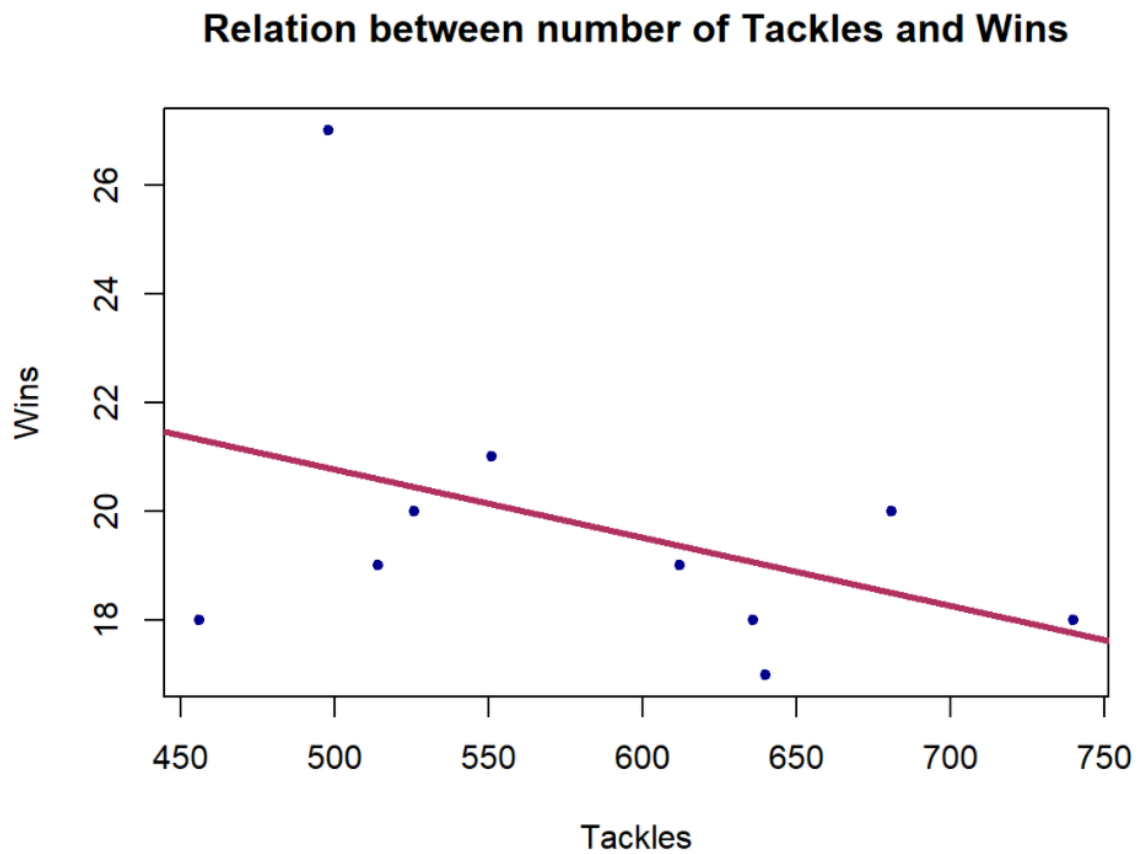

Relation between number of Shots and Wins

Touches

[1] "Correlation: 0.62"

Relation between number of Touches and Wins
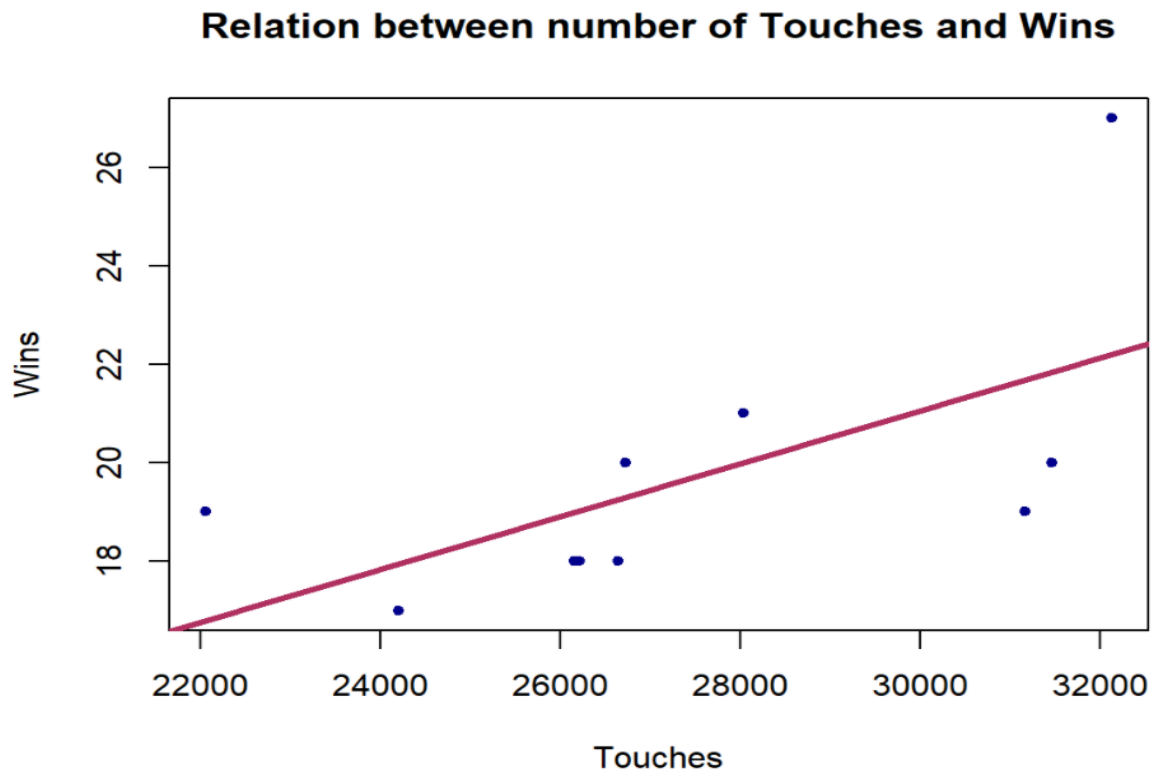
Tackles

[1] "Correlation: -0.4"

Relation between number of Tackles and Wins

Allows us to know which statistic is most important when the goal is to win a match. Pearson correlation is performed between the statistics and a scatterplot is plotted. A regression line allows the user to better visually understand the correlation.

The statistics used here are the cumulated club statistics of the top 10 clubs with the most wins in the season.

The output from these graphs tell us that there is a moderate positive correlation between number of shots and touches with the wins. However, the number of tackles has a moderate negative correlation with wins. This information can help in decision making as well as score prediction.

These visualizations allow us to identify Trends, Outliers, and Features in the graph as well as depicting the correlation, comparison, and presentation of the data.

Statistic: Passes

Barplot representing the number of passes by Player

Actions performed:

1. Discover ,present and Enjoy the data in an engaging and easy to understand visualization model

2. Identify, Summarize and Compare the goals by the top ten players in the season

3. Search for the number of passes conceded by a player

Target: Distribution of the number of passes by the top 10 players

The data is arranged by aligning it against a common scale for easy comparison and visualization
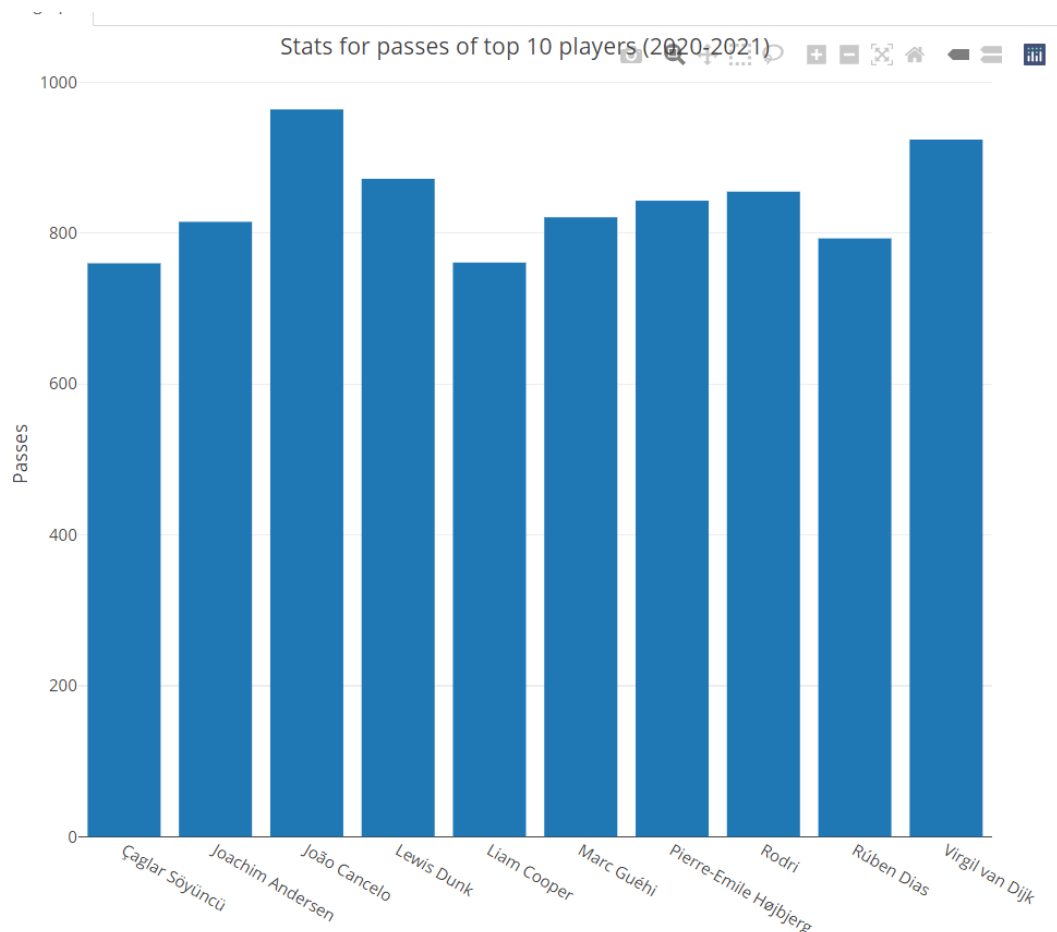
.

Statistic: Passes

Pie chart representing the number of passes by Club

Actions performed:

1. Discover , present and Enjoy the data in an engaging and easy to understand visualization model

2. Identify, Summarize and Compare the goals scored by the top ten clubs in the season

Target: Distribution of the number of passes by the top 10 clubs

The categorical data i.e team names have been mapped using different colours

Stats for passes of top 10 teams (2020-2021)

Statistic: Assists

Barplot representing the number of assists by Player

Actions performed:

1. Discover ,present and Enjoy the data in an engaging and easy to understand visualization model

2. Identify, Summarize and Compare the goals by the top ten players in the season

3. Search for the number of assists by a player

Target: Distribution of the number of assists by the top 10 players

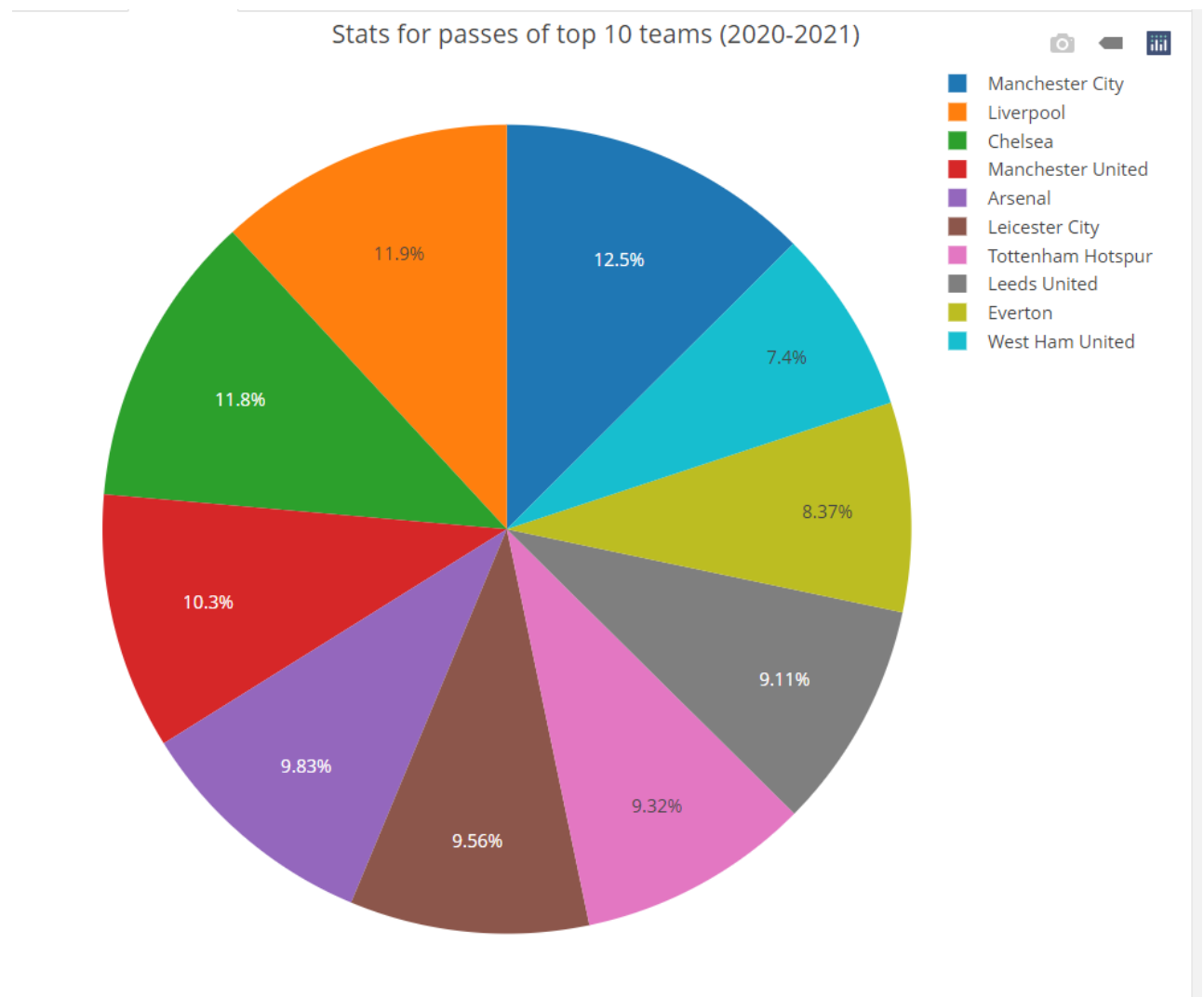The data is arranged by aligning it against a common scale for easy comparison and visualization

Barplot representing the number of assists by Player

Statistic: Assists

Scatter plot representing the number of goals conceded by a team, assists, and wins by top 10 team.

Here on the x axis we have the number of assists and on the y axis we have the number of goals. Goals are dependent on the number of assists. With the help of plot we can make a conclusion that the number of goals increases as the number of assists increase.

The size of each dot represents the number of wins by that particular team.

Actions performed:

1. Discover ,present and Enjoy the data in an engaging and easy to understand visualization model

2. Identify, Summarize and Compare the goals conceded per game, cleans sheets and tackles by the top defenders in the season

3. Search for the number of goals conceded , assists and wins by a team.

 Target: This visualization allows us to identify Trends, Outliers, Similarity and Features . The quantitative values are arranged separately in the scatterplot. The channels size and color are used to map further quantitative and categorical values respectively.
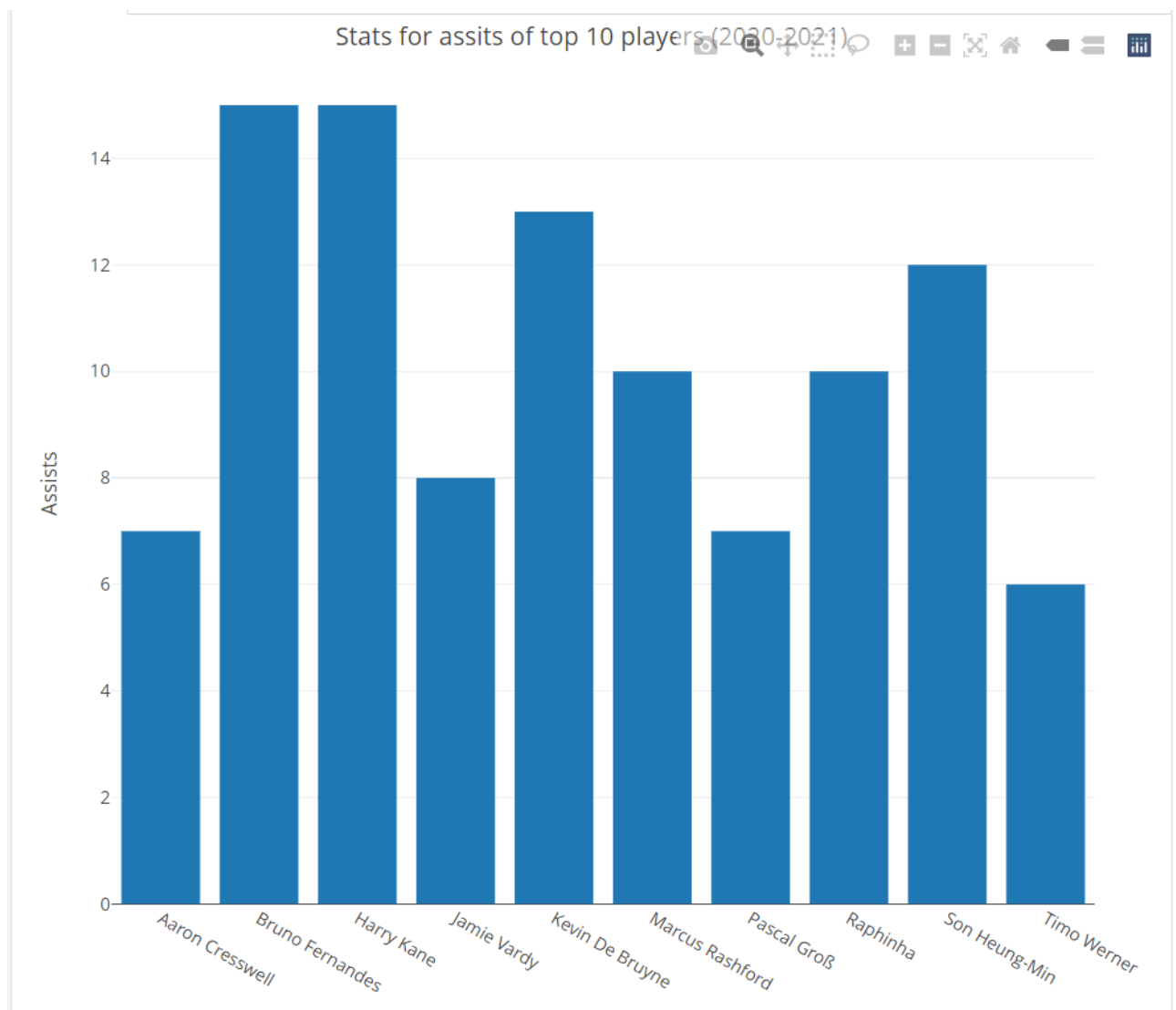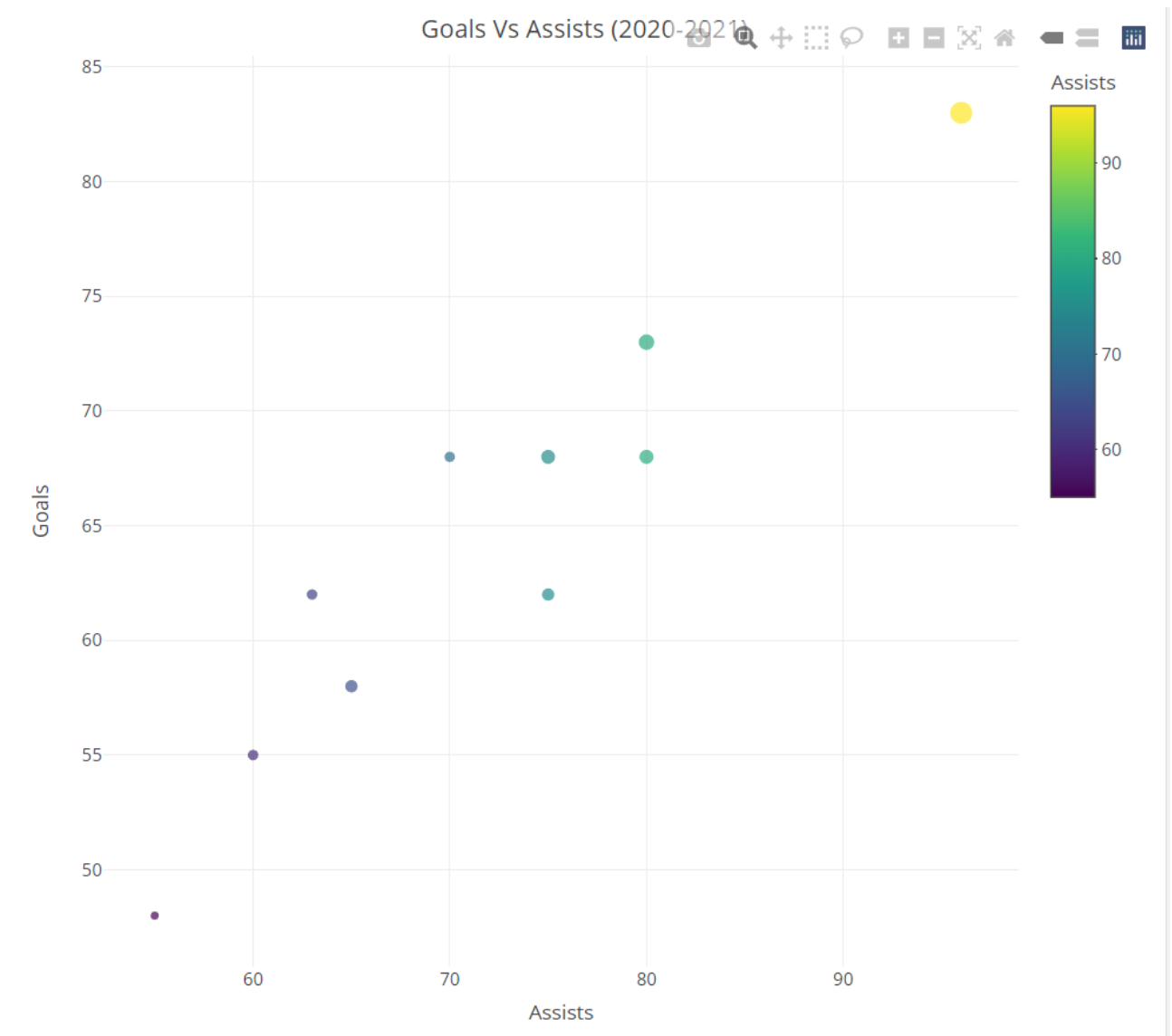
Statistic: Fouls

Barplot representing the number of fouls by Player

Actions performed:

1. Discover ,present and Enjoy the data in an engaging and easy to understand visualization model

2. Identify, Summarize and Compare the goals by the top ten players in the season

3. Search for the number of fouls conceded by a player

Target: Distribution of the number of fouls by the top 10 players

The data is arranged by aligning it against a common scale for easy comparison and visualization

Stats for fouls of top 10 players (2020-2021)
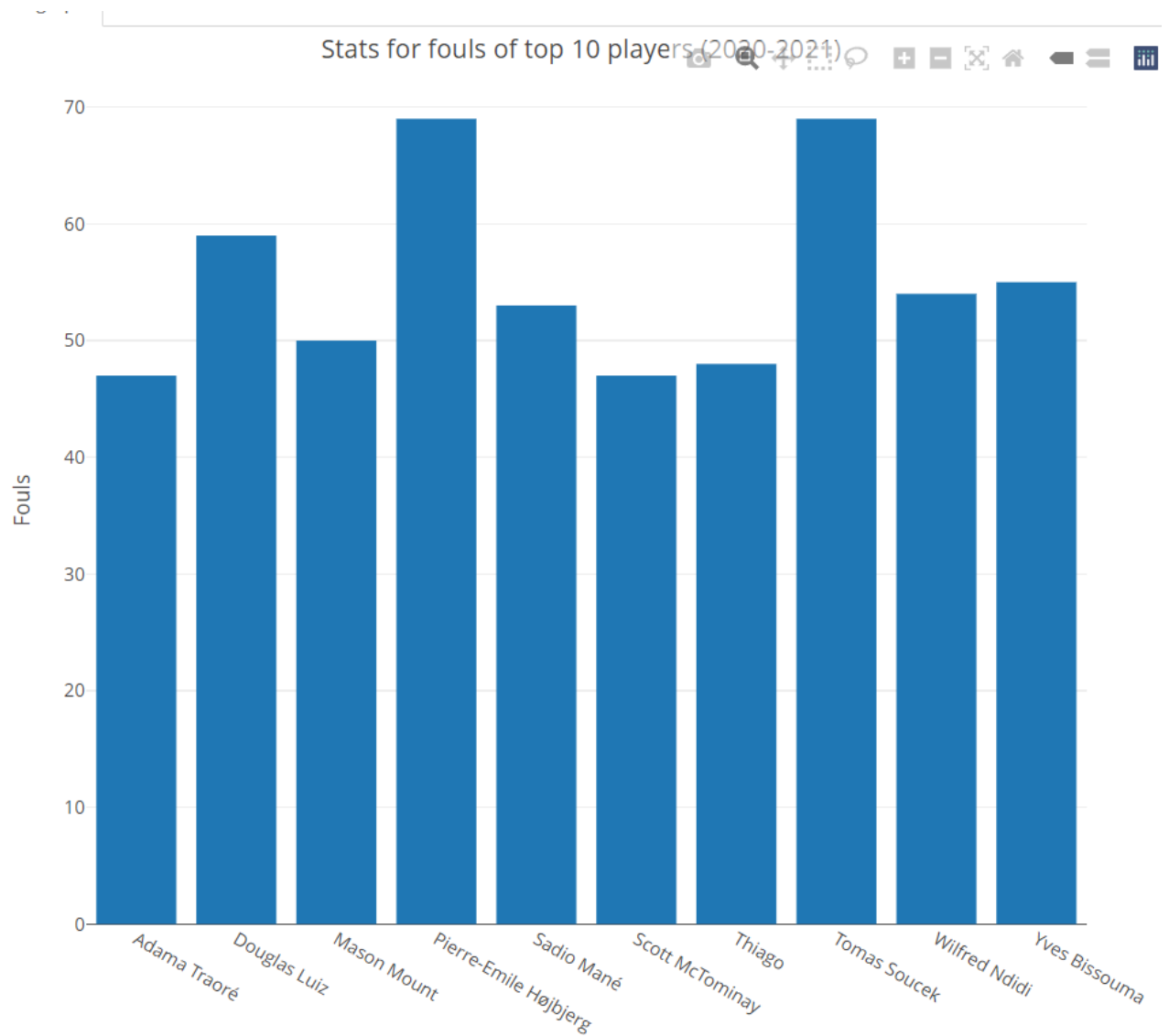
Statistic: Fouls

Grouped Bar Plot representing the number of red and yellow cards by a Team

Actions performed:

1. Discover ,present and Enjoy the data in an engaging and easy to understand visualization model

2. Identify, Summarize and Compare the goals by the top ten players in the season

3. Search for the number of red and yellow cards for a team.

Target: Distribution of the number of red and yellow cards by the top 10 teams

The data is arranged by aligning it against a common scale for easy comparison and visualization

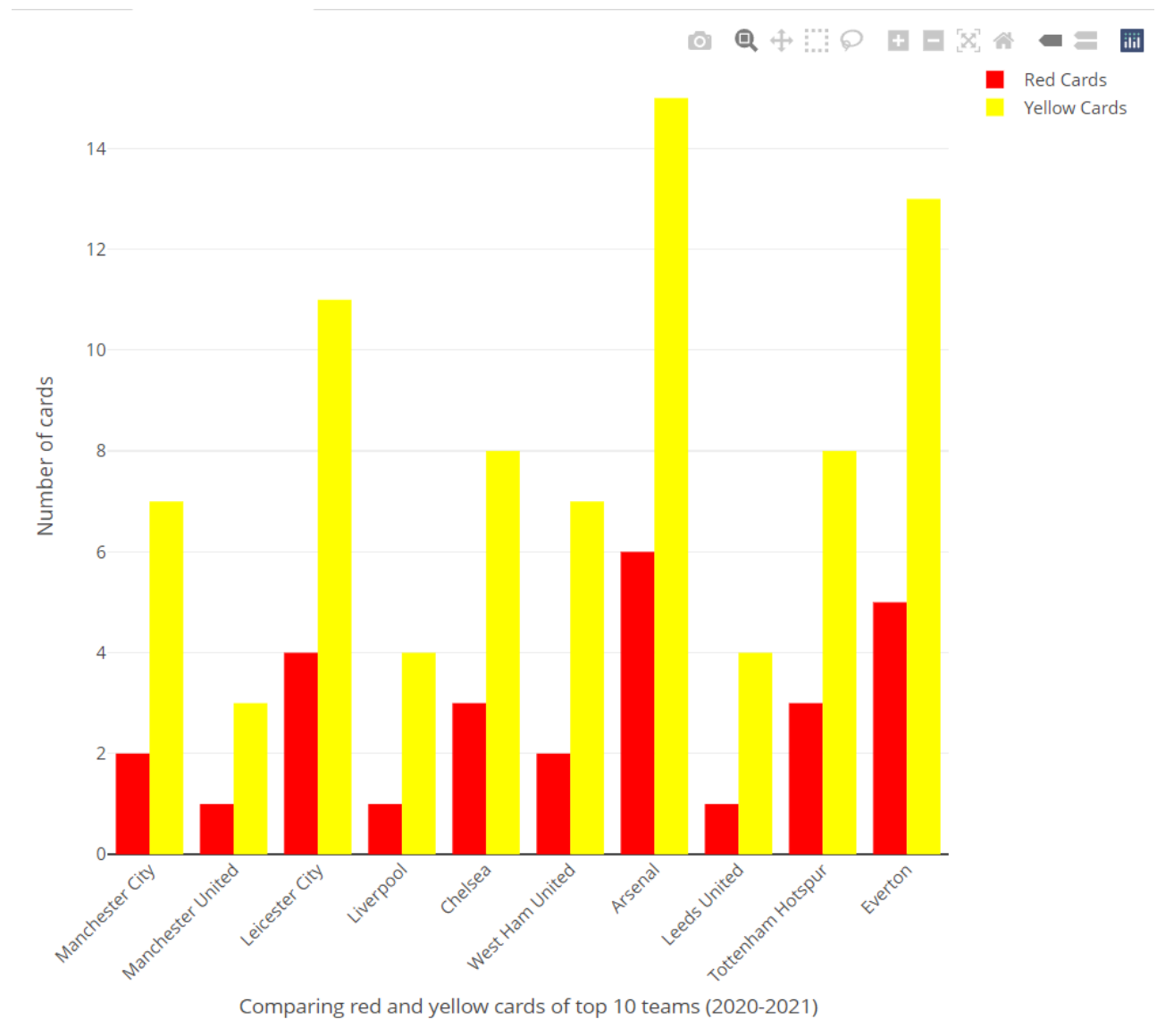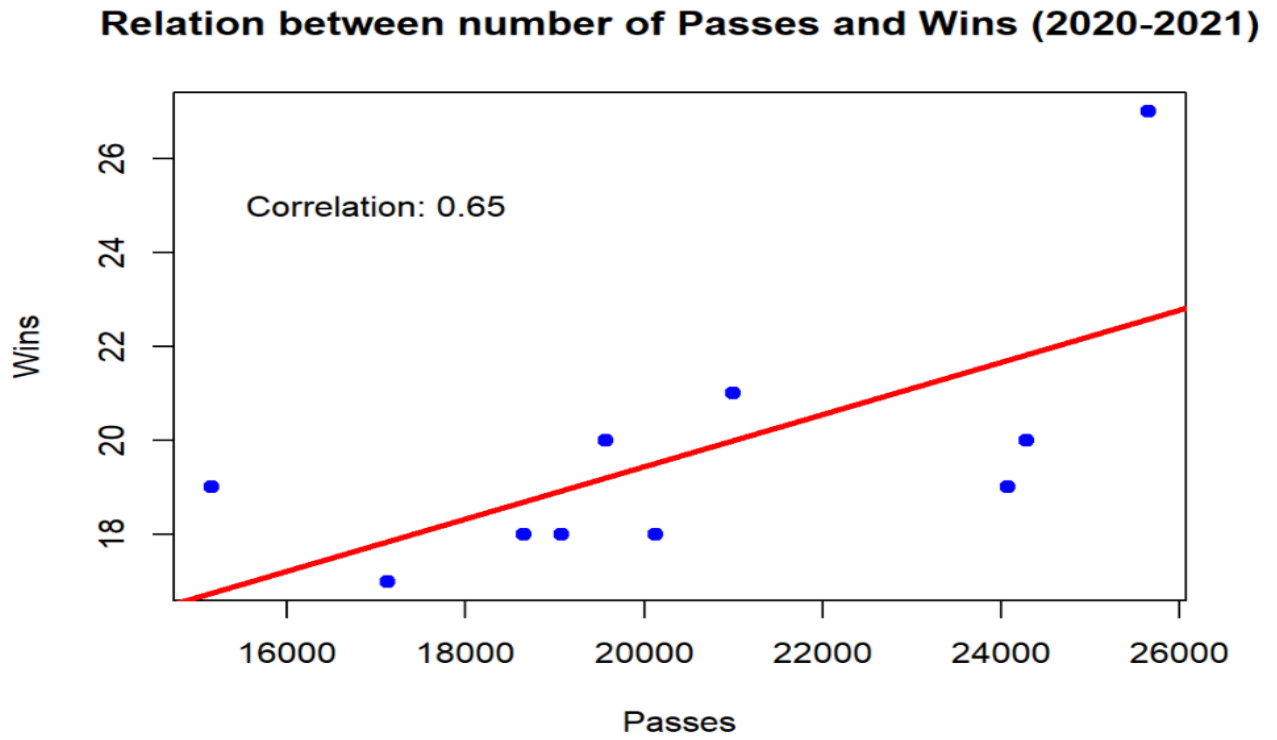Comparing red and yellow cards of top 10 teams (2020-2021)

Statistic: Correlation with wins

Pearson correlation between wins and passes:

Relation between number of Passes and Wins (2020-2021)

Correlation: 0.65

Wins

Passes

Pearson correlation between assists and wins:

Relation between number of Passes and Wins (2020-2021)
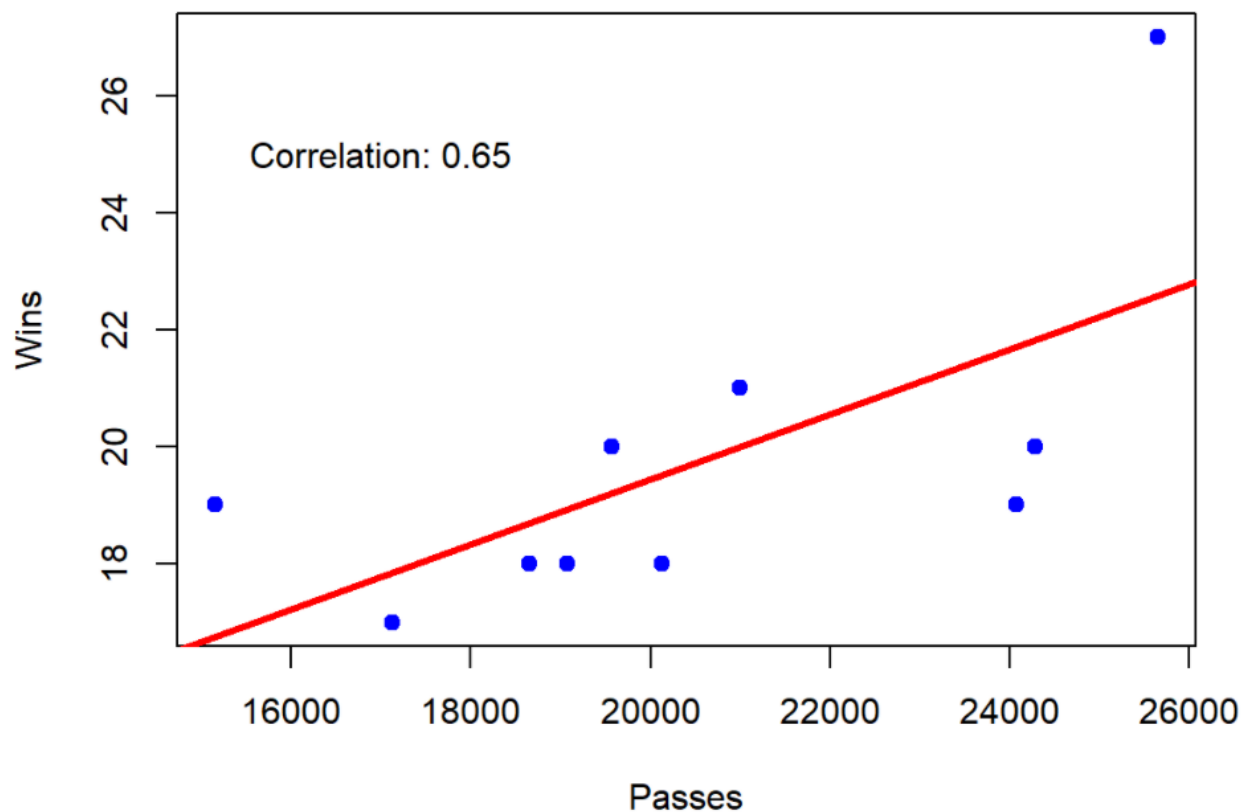
Correlation: 0.65

Allows us to know which statistic is most important when the goal is to win a match. Pearson correlation is performed between the statistics and a scatterplot is plotted. A regression line allows the user to better visually understand the correlation.

The statistics used here are the cumulated club statistics of the top 10 clubs with the most wins in the season.

The output from these graphs tell us that there is a moderate positive correlation between number of passes and assists with the wins. This information can help in decision making as well as score prediction.

These visualizations allow us to identify Trends, Outliers, and Features in the graph as well as depicting the correlation, comparison, and presentation of the data.

Statistic: Yellow cards

Barplot representing the top 10 players with most yellow cards (2020/2021)

Actions performed:

1. Discover , present and Enjoy the data in an engaging and easy to understand visualization model

2. Identify, Summarize and Compare the yellow cards issued to the top ten players in the season

Target: Analysis of the most yellow cards issued to the players

The data is arranged by aligning it against a common scale for easy comparison and visualization and for further clarity we have the reference we have the legend

TOP 10 PLAYERS WITH MOST YELLOW CARDS

Legend:
- John McGinn Aston Villa
- Harry Maguire Manchester United
- Conor Gallagher West Bromwich Albion
- Kalvin Phillips Leeds United
- Pierre-Emile Hojbjerg Tottenham Hotspur
- Mason Holgate Everton
- Hector Bellerin Arsenal
- Isaac Hayden Newcastle United
- Luke Shaw Manchester United
- John Lundstram Sheffield United

yellow card numbers
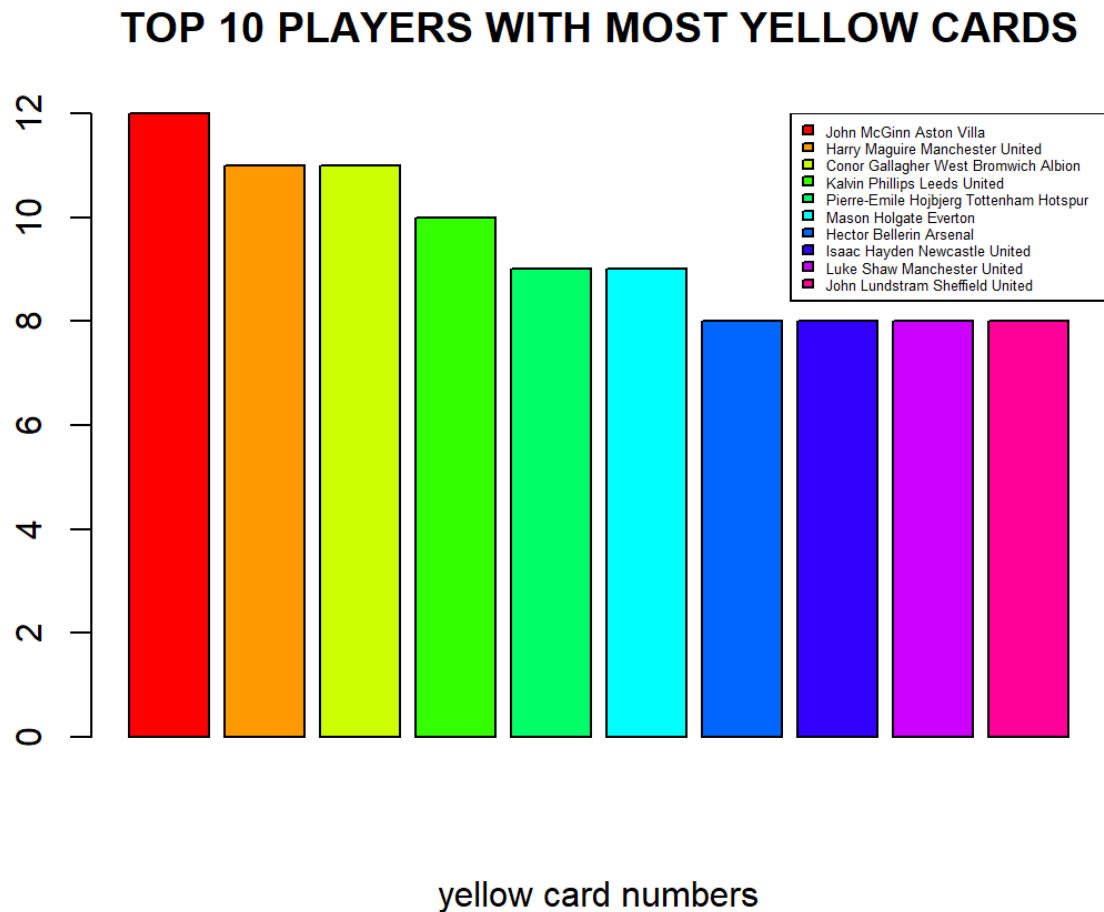
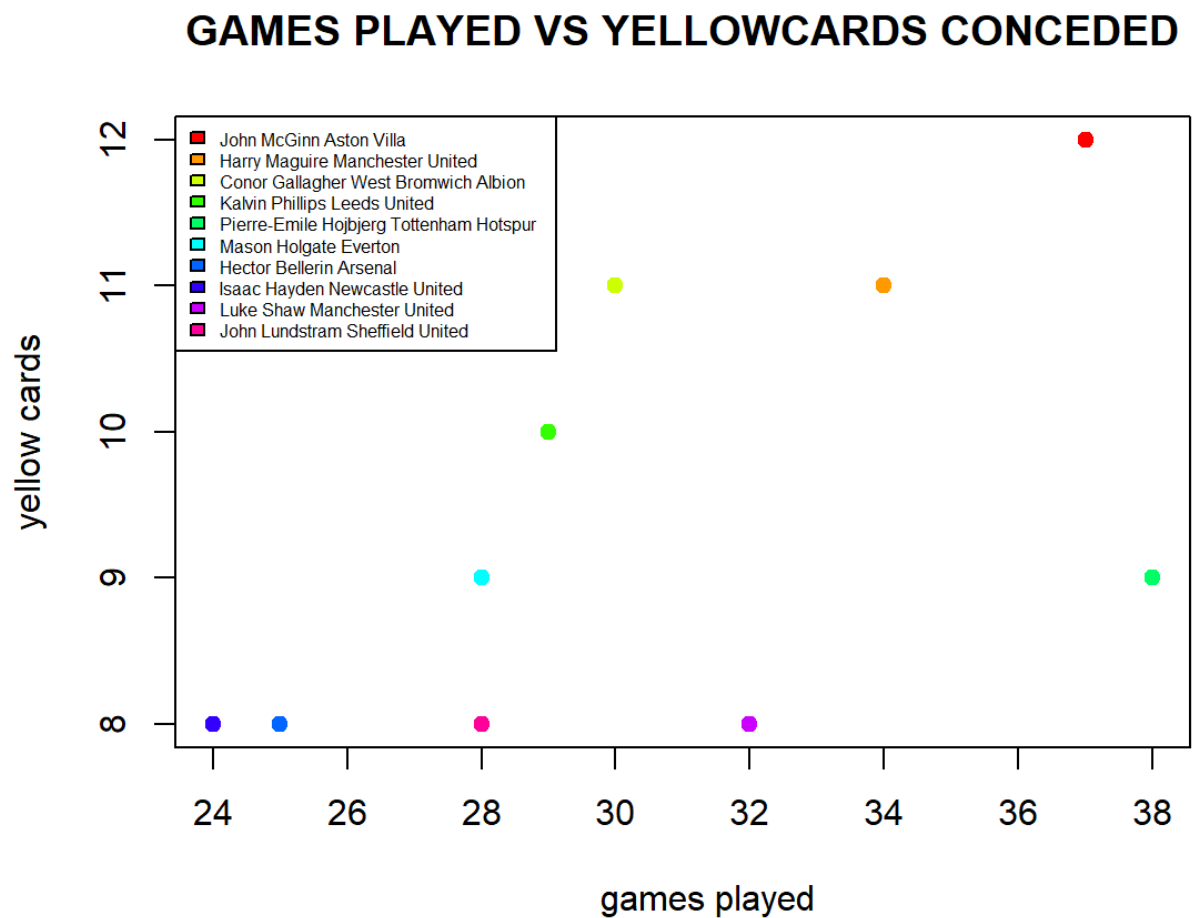**Statistic: Games played against yellow cards conceded**

plot representing the yellow cards issued against the games played

Actions performed:

1. Discover , present and Enjoy the data in an engaging and easy to understand visualization model

2. Identify, Summarize and Compare the yellow cards issued to the top ten players in the season

Target: Analysis of the frequency of yellow cards issued to the players on the basis of their appearance in the game

The data is arranged by plotting the number of yellow cards conceded against the total number of games played by the players for easy comparison and visualization and for further reference we have the legend

GAMES PLAYED VS YELLOWCARDS CONCEDED

Legend:
- John McGinn Aston Villa
- Harry Maguire Manchester United
- Conor Gallagher West Bromwich Albion
- Kalvin Phillips Leeds United
- Pierre-Emile Hojbjerg Tottenham Hotspur
- Mason Holgate Everton
- Hector Bellerin Arsenal
- Isaac Hayden Newcastle United
- Luke Shaw Manchester United
- John Lundstram Sheffield United

Statistic: player ranking as per yellow card average per game

plot representing the ranking for top 10 players on the basis of average yellow cards issued against them per game

Actions performed:

1. Discover , present and Enjoy the data in an engaging and easy to understand visualization model

2. Identify, Summarize and Compare the yellow cards issued to the top ten players in the season

Target: Analysis of the power ranking of yellow card average for the players per game

The data is arranged by plotting the number of average yellow cards per game conceded against the ranking of the players for easy comparison and visualization and for the reference we have the legend

PLAYER RANKING VS PER GAME YELLOWCARD AVERAGE

Legend:
- John McGinn Aston Villa
- Harry Maguire Manchester United
- Conor Gallagher West Bromwich Albion
- Kalvin Phillips Leeds United
- Pierre-Emile Hojbjerg Tottenham Hotspur
- Mason Holgate Everton
- Hector Bellerin Arsenal
- Isaac Hayden Newcastle United
- Luke Shaw Manchester United
- John Lundstram Sheffield United

yellow card average per game

Statistic: top 10 clubs with the most yellow cards for 2020/2021

pie plot representing the top 10 clubs with the most yellow cards for 2020/2021
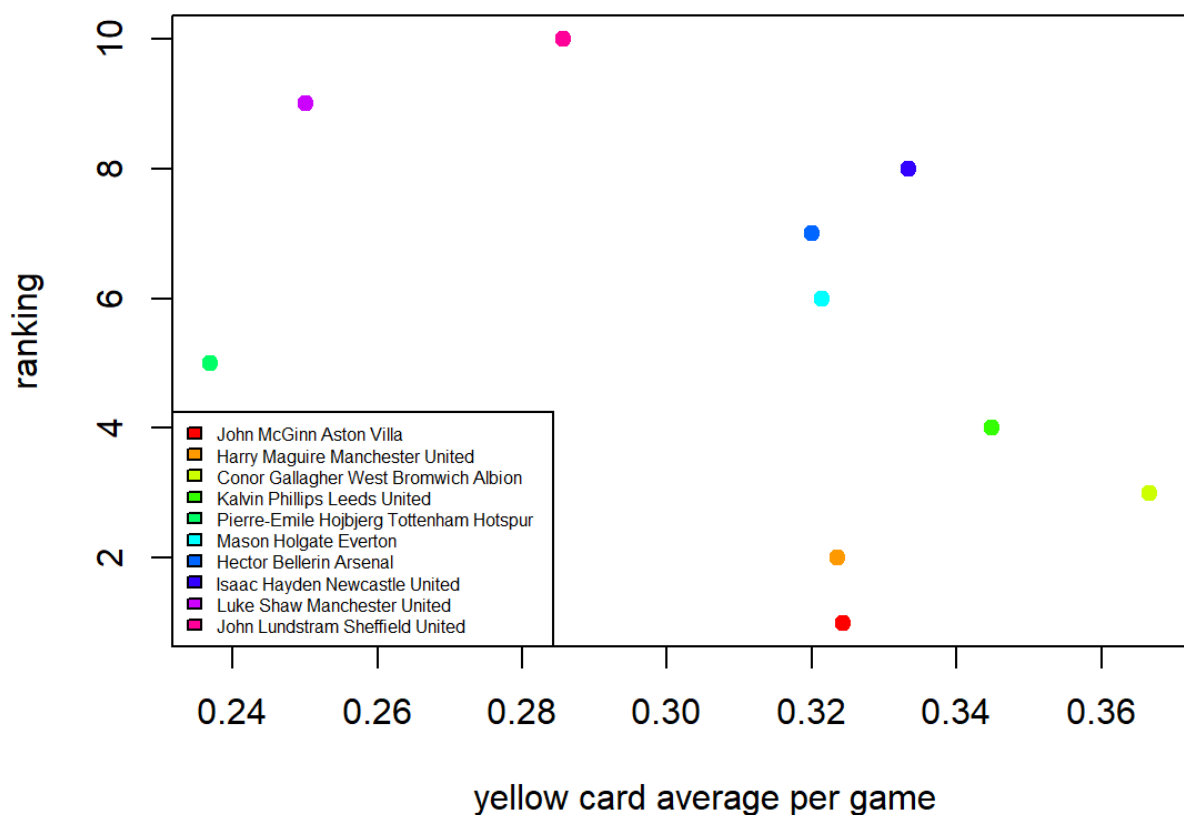
Actions performed:

1. Discover , present and Enjoy the data in an engaging and easy to understand visualization model

2. Identify, Summarize and Compare the yellow cards issued to the top ten players in the season

Target:Distribution of the top 10 clubs issued with most yellow cards

The categorical data i.e team names have been mapped using different colours for easy comparison and visualization and for the reference we have the legend

YELLOW CARDS 20/21(clubs)

Statistic: Games played against red cards conceded

plot representing the top 10 players with the most red cards against games played
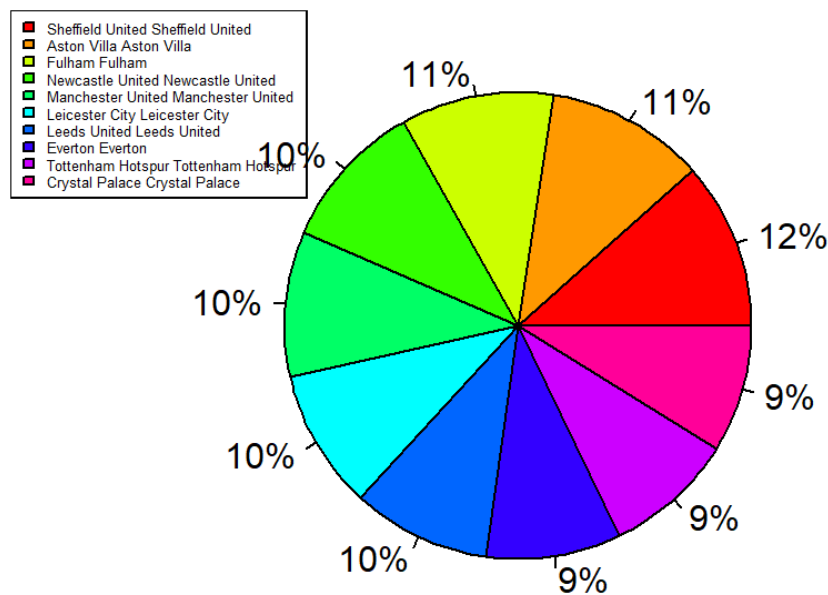
Actions performed:

1. Discover , present and Enjoy the data in an engaging and easy to understand visualization model

2. Identify, Summarize and Compare the 10 players with the most red cards issued in the following season

Target:Analysis of the power ranking of red card average for the players per game The data is arranged by plotting the number of average yellow cards per game conceded against the ranking of the players for easy comparison and visualization and for the reference we have the legend

GAMES PLAYED VS RED CARDS CONCEDED

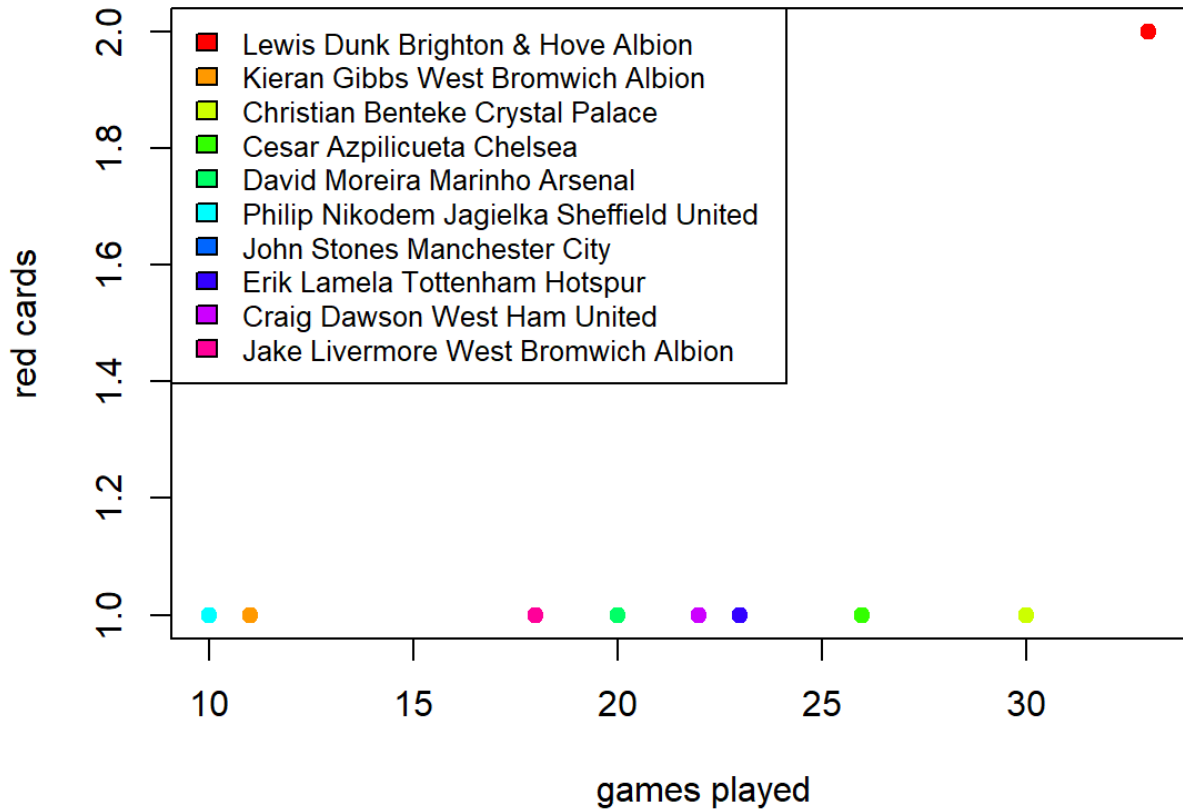**Statistic: player ranking as per red card average per game**

plot representing the top 10 players with the most red cards against games played

Actions performed:

1. Discover , present and Enjoy the data in an engaging and easy to understand visualization model

2. Identify, Summarize and Compare the 10 players with the most red cards issued in the following season

Target:Analysis of the power ranking of red card average for the players per game The data is arranged by plotting the number of average yellow cards per game conceded against the ranking of the players for easy comparison and visualization and for the reference we have the legend

PLAYER RANKING VS PER GAME REDCARD AVERAGE

Legend:
- Lewis Dunk Brighton & Hove Albion
- Kieran Gibbs West Bromwich Albion
- Christian Benteke Crystal Palace
- Cesar Azpilicueta Chelsea
- David Moreira Marinho Arsenal
- Philip Nikodem Jagielka Sheffield United
- John Stones Manchester City
- Erik Lamela Tottenham Hotspur
- Craig Dawson West Ham United
- Jake Livermore West Bromwich Albion

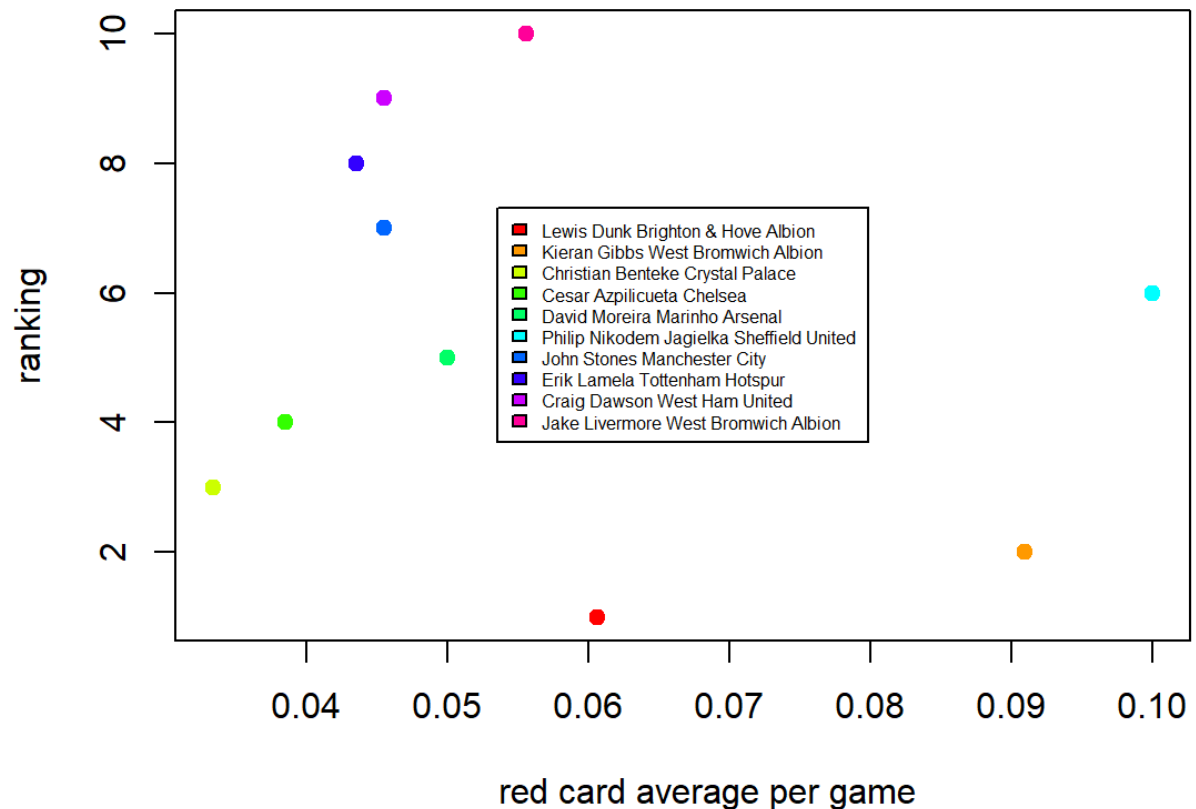**Statistic: top 10 clubs with the most red cards for 2020/2021**

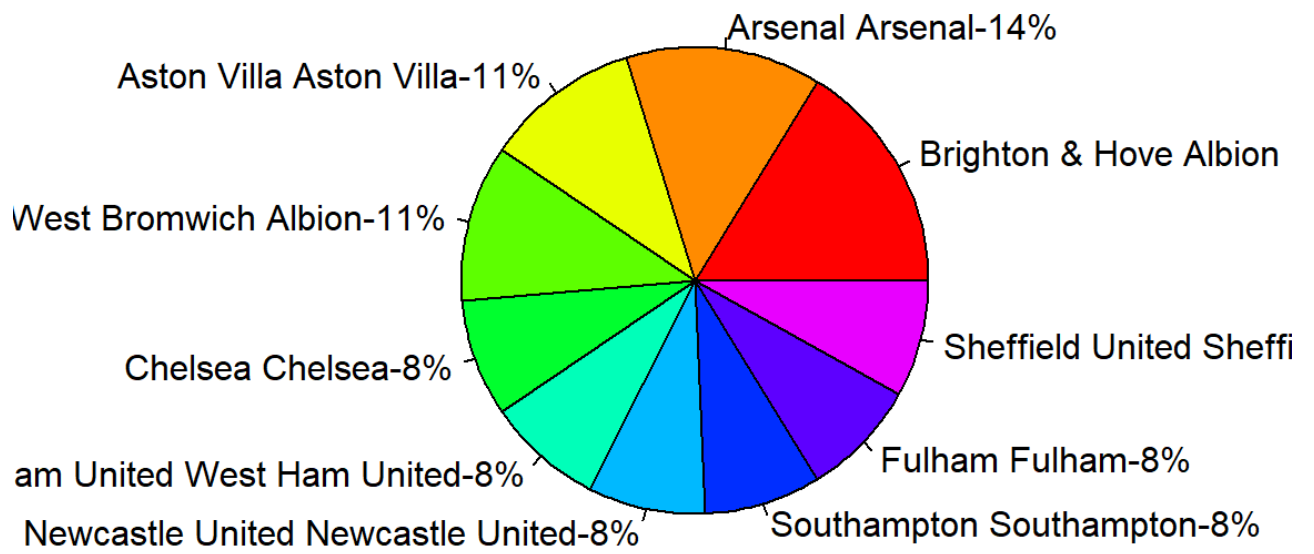**pie plot representing the top 10 clubs with the most red cards for 2020/2021**

Actions performed:

1. Discover , present and Enjoy the data in an engaging and easy to understand visualization model

2. Identify, Summarize and Compare the yellow cards issued to the top ten players in the season

Target:Distribution of the top 10 clubs issued with most red cards the following season

The categorical data i.e team names have been mapped using different colours for easy comparison and visualization and for the reference we have the legend

RED CARDS 20/21(club)

Arsenal Arsenal-14%

Brighton & Hove Albion

Sheffield United Sheffi

Fulham Fulham-8%

Southampton Southampton-8%

Newcastle United Newcastle United-8%

am United West Ham United-8%

Chelsea Chelsea-8%

West Bromwich Albion-11%

Aston Villa Aston Villa-11%

DESIGN OF PROPOSED SYSTEM :

Link:

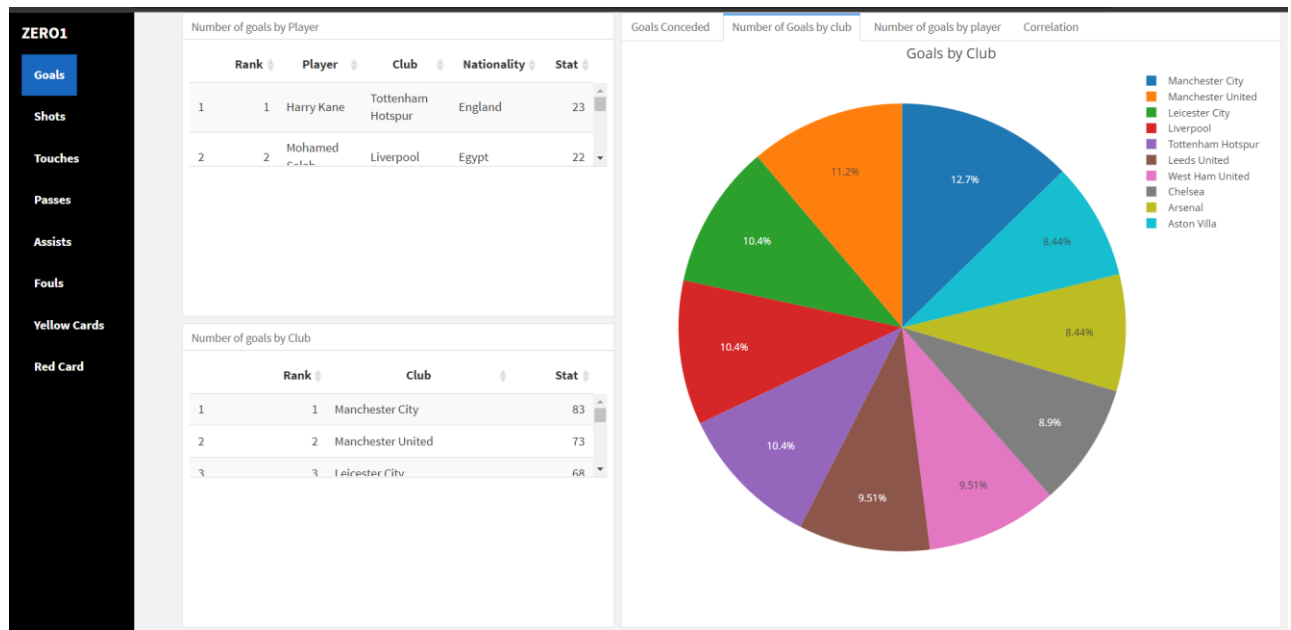https://www.figma.com/file/LWd79jU9Nlm1sRaloDj73m/ZeR01?node-id=2%3A2

DASHBOARD IMPLEMENTATION

The dashboard has been implemented using flexdashboard and it is designed such that it is graphically distinctive as well as functionally efficient. All the statistics are displayed in the sidebar that leads to a page containing the graphs and tables relevant to that statistic. The graphs, plots and charts have been displayed in a tabset column to effectively use the available space.


CONCLUSION

The dashboard successfully exhibits the statistics in a visually brilliant and comprehensive manner using interactive graphs, plots and charts. The insight provided by the dashboard can be used by the coaching staff of a team to monitor improvements in game-plans, team strategies and player performance. The visual representation of the data also facilitates an objective viewpoint that can help in decision making as well as developing targets and training methods. Furthermore, having records of team and player performance as well as the relationship between the various statistics is invaluable in determining future performances.

The dashboard can also be of great use to individual players as the insight provided can help enhance development, game understanding and decision making by enabling players to view and analyze their own performance, and identify their strengths and weaknesses. In addition to their own performance, they can also compare the performances of rival players and teams and use this information to train accordingly.

For a casual viewer or a sports enthusiast, the graphs, plots and charts demonstrated in the dashboard allows for a comprehensive understanding of the sport and highlights relevant insights through visualization. It also makes it easier for users to compare player and team performance without having to analyze a bulk of data.

APPENDIX

SCREEN SHOTS
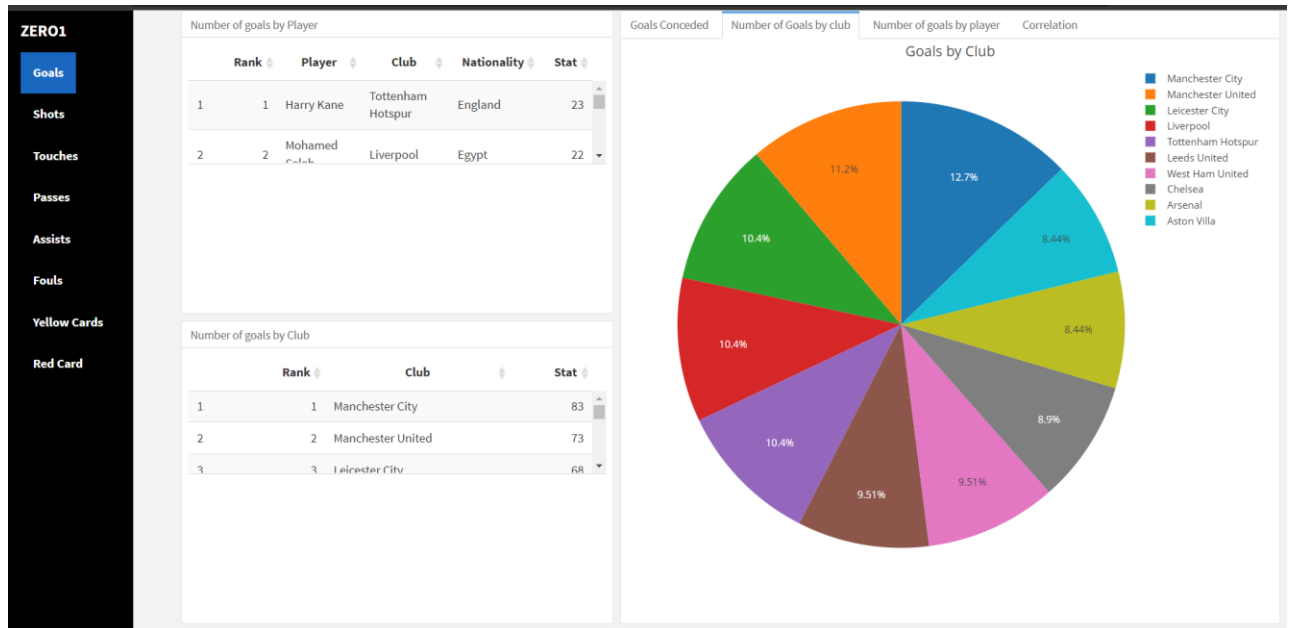
SAMPLE CODING


```
1    ---
2    title: "ZERO1"
3    output:
4      flexdashboard::flex_dashboard:
5        orientation: column
6    runtime: shiny
7    ---
8    <style>
9    .navbar {
10     margin: 0;
11     padding: 0;
12     height: 100%;
13     display: block;
14     position: fixed;
15     width: 150px;
16     background:black;
17     font-weight:700;
18   }
19
20   body {
21           margin-left: 200px;
22           padding-top:0px
23   }
24   </style>
25
26
27 ```{r setup, include=FALSE}
28   library(flexdashboard)
29
30 ```
31        Goals
32 ===============================
33   Column {data-width=400}
34 -------------------------------------------------------------------
35 ### Number of goals by Player
36 ```{r}
37   library(tidyverse)
38   g<-read.csv("C:/Users/athar/Desktop/ZERO1/Aayush code/premierleaguegoals.csv")
39   library(DT)
40   datatable(g,options = list(lengthChange = FALSE,searching=FALSE,paging=FALSE,info=FALSE))
41 ```
42 ### Number of goals by Club
```

```r
34 ----------------------------------------------------------------
35 ### Number of goals by Player
36 ```{r}
37 library(tidyverse)
38 g<-read.csv("C:/Users/athar/Desktop/ZER01/Aayush code/premierleaguegoals.csv")
39 library(DT)
40 datatable(g,options = list(lengthChange = FALSE,searching=FALSE,paging=FALSE,info=FALSE))
41 ```
42 ### Number of goals by Club
43 ```{r}
44 library(tidyverse)
45 #w<-read.csv("C:\\Users\\aayus\\Documents\\clubgoals.csv")
46 w<-read.csv("C:/Users/athar/Desktop/ZER01/Aayush code/clubgoals.csv")
47 library(DT)
48 datatable(w,options = list(lengthChange = FALSE,searching=FALSE,paging=FALSE,info=FALSE))
49 ```
50
51 Column {.tabset .tabset-fade data-width=600}
52 ----------------------------------------------------------------
53
54 ### Goals Conceded
55
56 ```{r}
57
58 #goalsconceded
59 df<-read.csv("C:/Users/athar/Desktop/ZER01/Aayush code/pl_20-21.csv")
60 library(dplyr)
61 #defenders<-subset(df,Position=="Defender")
62 #defenders<-defenders%>%
63  # arrange(desc(Appearances))
64 #defenders<- head(defenders,10)
65 #defenders["cpg"]<-defenders$Goals.Conceded/defenders$Appearances
66 #library(RColorBrewer)
67 #coul <- brewer.pal(length(defenders$Name), "Paired")
68 #par(mar=c(5,8,2,3),xpd=TRUE)
69 #plot(defenders$Clean.sheets,defenders$cpg,ylab='Goals conceded per game',cex=defenders$Tackles/70, main ='Goals conceded p
   Sheets",col=coul,pch=20)
70 #legend("bottomleft",defenders$Name, cex = 0.4, fill =coul,inset=c(-0.40,0))
71
72 defenders<-subset(df,Position=="Defender")
73 defenders<-defenders%>%
74   arrange(desc(Appearances))
```
915:1   pie plot

Console
```

```r
472
473 # Assists
474 Column {data-width=400}
475 ----------------------------------------------------------------
476
477 ### Data for individual player
478
479 ```{r}
480
481 assists <-read.csv("assists1.csv")
482 library(DT)
483
484
485 #datatable(assists)
486 datatable(assists,options = list(lengthChange = FALSE,searching=FALSE,paging=FALSE,info=FALSE))
487
488 ```
489 ### Data for top 10 teams
490
491 ```{r}
492 w <-read.csv("assists_wins.csv")
493 library(DT)
494
495 #datatable(w)
496 datatable(w,options = list(lengthChange = FALSE,searching=FALSE,paging=FALSE,info=FALSE))
497
498 ```
499
500 Column {.tabset .tabset-fade data-width=600}
501 ----------------------------------------------------------------
502
503 ### Bar graph
504
505 ```{r}
506
507 #assists <-read.csv("assists1.csv")
508
509
510 #midpts<-barplot(assists$Stat,ylab="Total assists",col=rgb(0.8,0.1,0.1,0.6),main="Stats for Assists",cex.names = 0.8,las = 2,ylim = c(0, 20))
511
```
915:1   pie plot

Console

Code Link : <u>https://github.com/AtharvaParkale/ZER01</u>