Cairo University

Faculty of Engineering
Cairo University

Data Science Project

Supervised by: Eng. Mohammad Abdallah

Team 8

Presented by:

Name	Sec.	B.N.
Abdallah mahmoud	1	41
Gaser Ashraf	1	23
Ahmed Mahmoud Hafez	1	11
Omar Abdelfatah	2	5

Spring 2023

Team Contribution

Abdallah mahmoud	**2 Predictive** **1 Descriptive**
Gaser Ashraf	**1 Inferential** **1 exploratory** **1 Descriptive**
Ahmed M.Hafez	**2 exploratory** **1 Inferential**
Omar Abdelfatah	**—**

Gaser Questions:

Question 1:

How does the number of touches a player has correlate with their defensive contributions (such as tackles, interceptions, or blocks)?

1- Define the problem:

The problem is to investigate the relationship between the number of touches a player has and their defensive contributions. The goal is to determine if there is any correlation between these variables and understand how they influence each other.

2- Data collection:

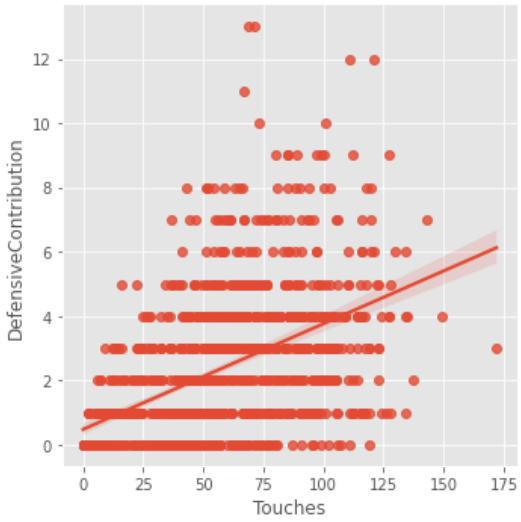
We write selenium script using python to crawl data of liverpool matches last two seasons 2020/2021 - 2021/2022 in premier league from https://fbref.com/en/

3- Data cleaning and preprocessing:

We removed any rows that had missing data, and we calculated the Defensive Contribution as the number of tackles + the number of interceptions + the number of Blocks.

4- Exploratory data analysis (EDA):

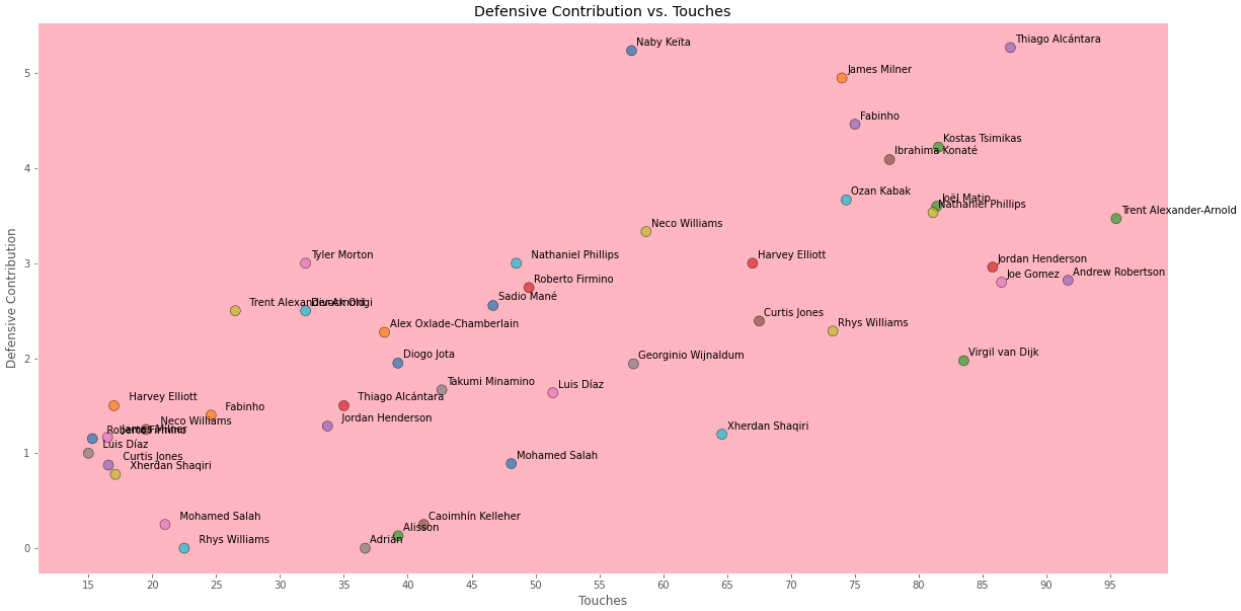
We draw a scatter plot to visualize the relationship between two variables.



5- Data modeling and analysis:

we used a statistical techniques to analyze the relationship between the number of touches and defensive contributions we calculate the correlation between number of touches and defensive contributions and it’s 0.47

6- Result Interpretation:



7- Communicating results:

Player Performance Evaluation:

- By understanding the correlation between the number of touches and defensive contributions, the club can assess player performance more comprehensively. They can identify players who excel in defensive contributions based on their touches and utilize this information for player evaluations, contract negotiations, and squad selection.

Tactical Decision-Making & Team Strategy:

- The analysis highlights the importance of players with a high number of touches in terms of defensive contributions. Coaches and managers can utilize this information to devise tactical strategies that maximize defensive effectiveness. For example, they might focus on involving players with more touches in defensive situations or positional setups that allow these players to contribute more defensively.

Player Development:

- The analysis can also guide player development programs. The club can emphasize skill training and tactical awareness for players who have a high number of touches but a comparatively lower defensive contribution. By addressing this gap, the players can improve their defensive skills and become more well-rounded assets to the team.

Question 2:

What is the performance of Liverpool's players based on their position?

1- Define the problem:

The problem is to assess the performance of Liverpool's players based on their positions. We want to determine which positions have been the most effective and identify any patterns or trends in player performance.

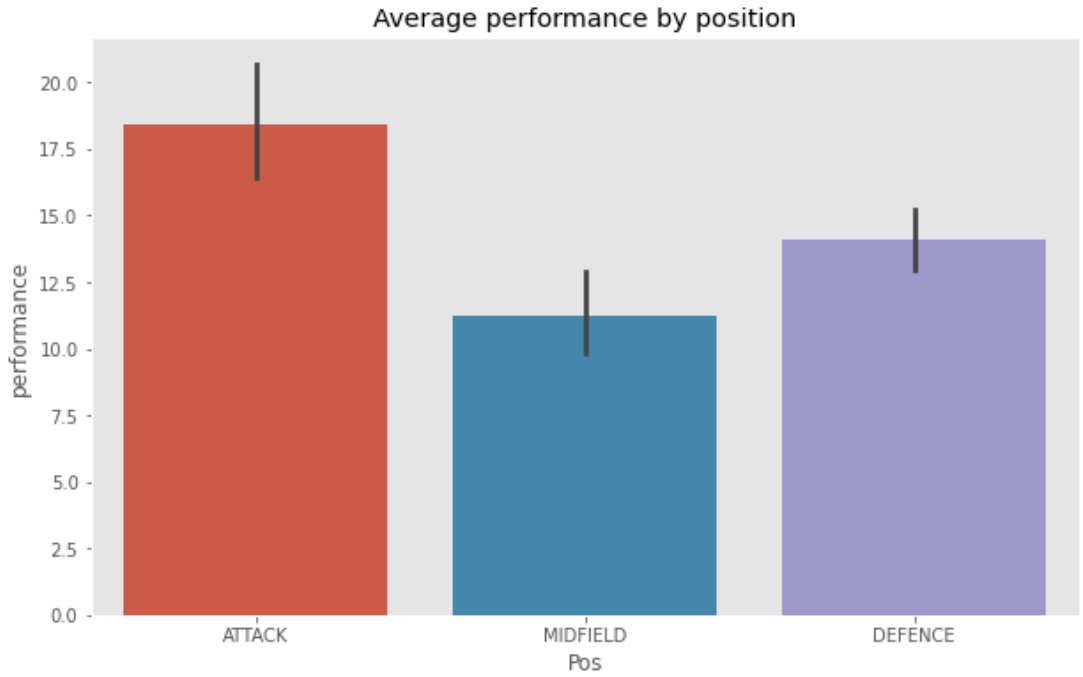
2- Data collection:

We write selenium script using python to crawl data of liverpool matches last two seasons 2020/2021 - 2021/2022 in premier league from https://fbref.com/en/

3- Data cleaning and preprocessing:

We removed any rows that had missing data, Split the positions into three categories: attack, midfield, and defense. You can assign each player to one of these categories based on their primary position. Calculate the performance for each player using a weighted formula based on their position. The weighted formula will assign different weights to different performance metrics based on their relevance to the specific position. For example, if goals and assists are more critical for attackers, you can assign higher weights to those metrics for players in the attack category.

4- Exploratory data analysis (EDA):



5- Communicating results:

Player recruitment and transfers:

- If the analysis reveals that the attack region has been performing well while the defense and midfield regions have lower performance, it can guide decision-making in player recruitment and transfers. The club could focus on strengthening the defense and midfield positions by targeting players who have demonstrated strong performance in those areas. This analysis helps align the recruitment strategy with the identified areas of improvement.

Question 3:

Does age affect the number of minutes played?

1- Define the problem:

The problem is to determine if there is a relationship between a player's age and the number of minutes they play in matches.

2- Data collection:

We write selenium script using python to crawl data of liverpool matches last two seasons 2020/2021 - 2021/2022 in premier league from https://fbref.com/en/

3- Data cleaning and preprocessing:

We removed any rows that had missing data, Split the Age into four groups :

first group: 17-20

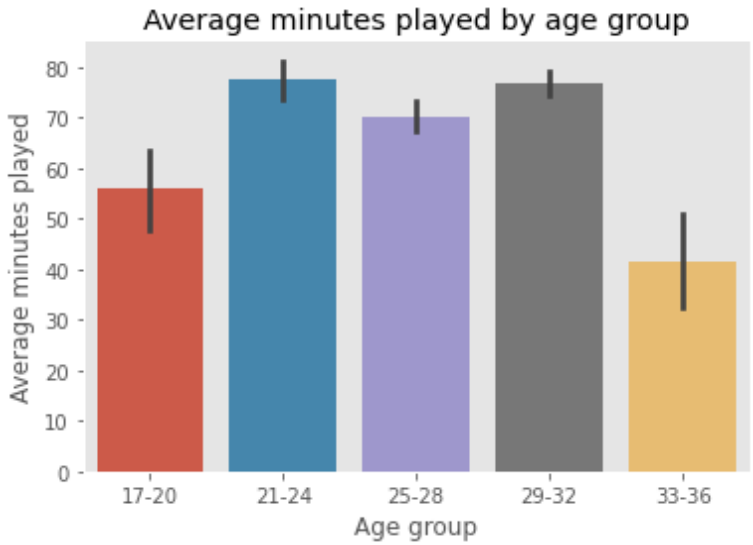
second group: 21-24

third group: 25-28

fourth group: 29-32

fifth group: 33-36

4- Exploratory data analysis (EDA):



5- Data modeling and analysis:

In this step, we performed an analysis of variance (ANOVA) to assess the impact of age on the number of minutes played. ANOVA is suitable when you have multiple groups (in this case, different age groups) and want to determine if there are significant differences between them.

6- Result Interpretation:

First we defined The null hypothesis states that there are no significant differences between the means of the groups being compared, and we got this results

```
F-statistic: 25.700058035329
P-value: 2.4356957809945952e-20
Lower critical F-value:  0.121017357498983
Upper critical F-value:  2.7979813415061896
```

As p-value is less than 0.05, we reject the null hypothesis and conclude that there is a significant difference between age and number of minutes played.

7- Communicating results:

Player Development and Retention:

- The analysis can help identify trends in player performance and playing time as they age. The club can assess whether players tend to play fewer minutes as they get older, indicating a decline in performance or physical abilities. This information can guide decisions on player contracts, such as negotiating suitable terms for older players or planning for potential replacements.

Youth Academy and Recruitment:

- If the analysis reveals that younger players tend to receive more playing time, it can emphasize the importance of investing in the club's youth academy and recruiting young talents. This knowledge can guide the scouting and recruitment strategy, focusing on promising young players who have the potential to contribute significantly and accumulate playing time over the years.

Squad Planning and Rotation:

- Understanding the relationship between age and playing time can help in squad planning and rotation strategies. The club can manage the workload of older players by strategically resting them during certain matches or substituting them with younger players. This approach can help maintain the performance and longevity of senior players while providing opportunities for younger talents to gain valuable experience and develop.

Ahmed Hafez Questions:

Question 1 (Exploratory):

How do Liverpool's players compare to other top players in the league in terms of key metrics such as goals, assists, and shots on target?

1- Define the problem:

The problem at hand is to compare Liverpool's players with other top players in the league based on key metrics such as goals, assists, and shots on target. This analysis aims to provide insights into the performance of Liverpool's players in relation to their peers in terms of offensive contributions. By examining these key metrics, we can evaluate the effectiveness and productivity of Liverpool's players in comparison to other top performers in the league.

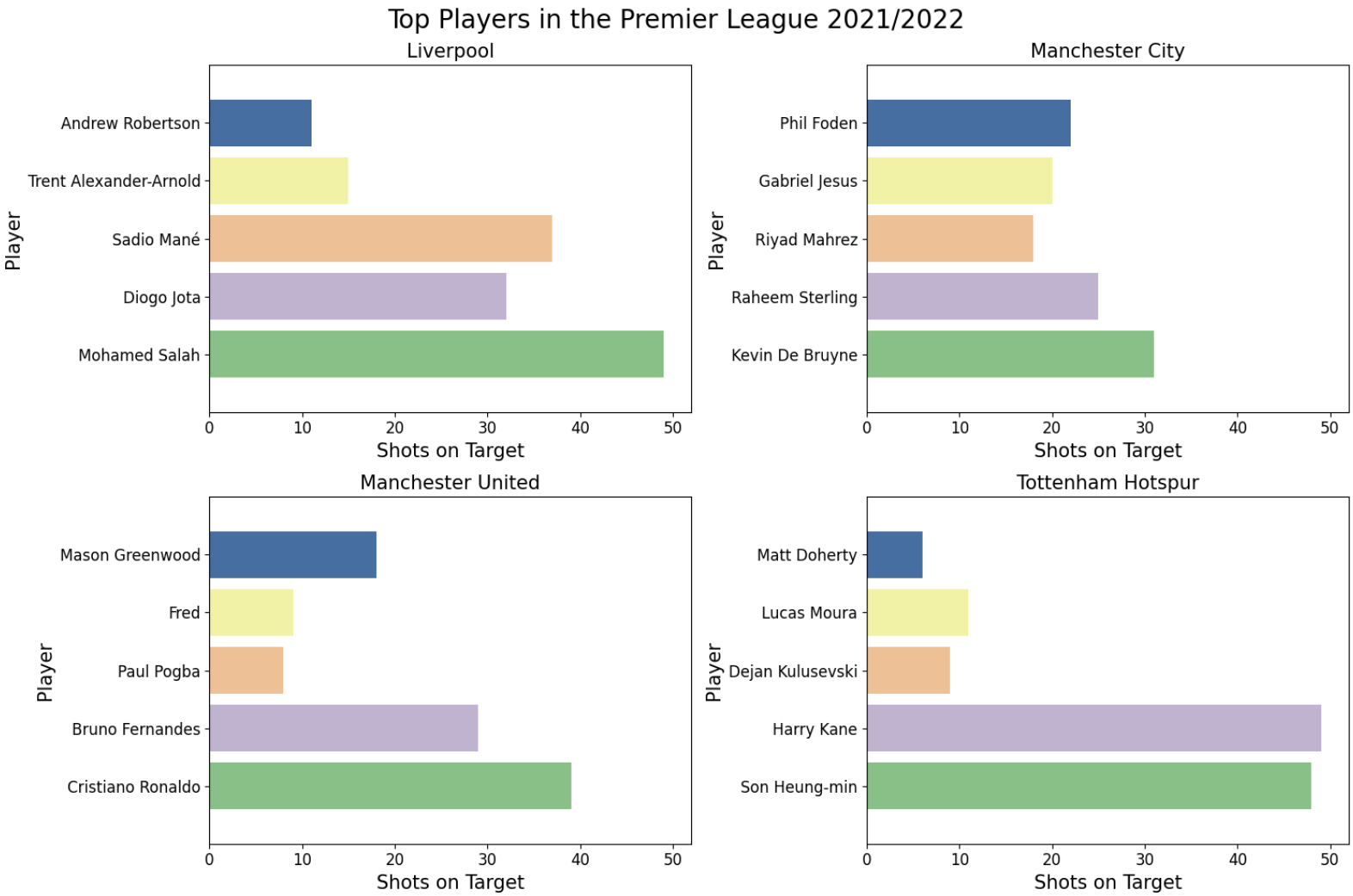
2- Data collection:

download data for top 4 teams in premier league from https://shorturl.at/dyJU9

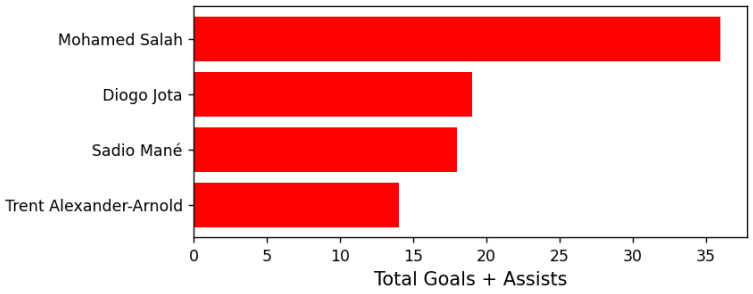
3- Data cleaning and preprocessing:

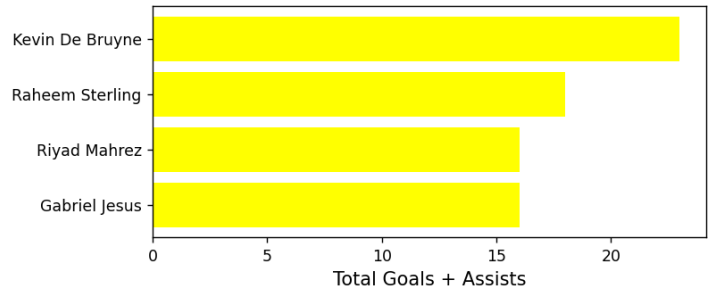
- dealt with non values rows
- select ['Player', 'Gls', 'Ast', 'SoT'] columns from the file
- merge data of the four teams
- form a column for total goals and assist for attackers
- sorting data & grouping every player to their team

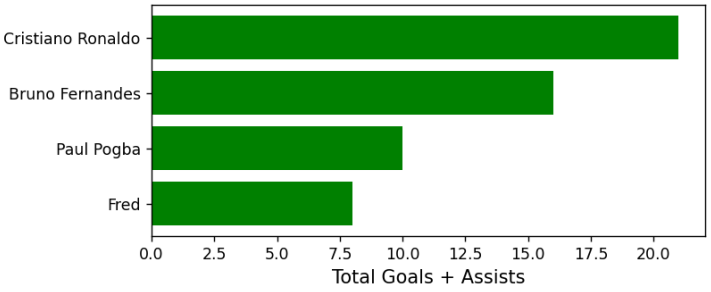
4- Exploratory data analysis (EDA):

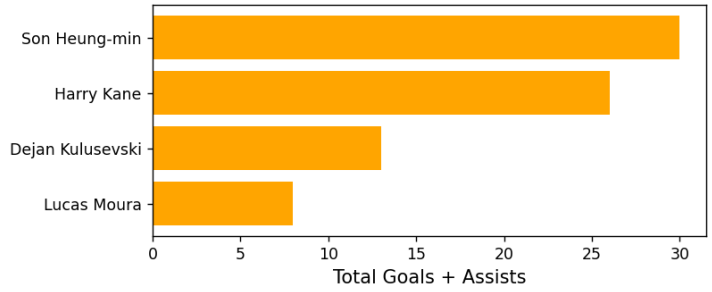


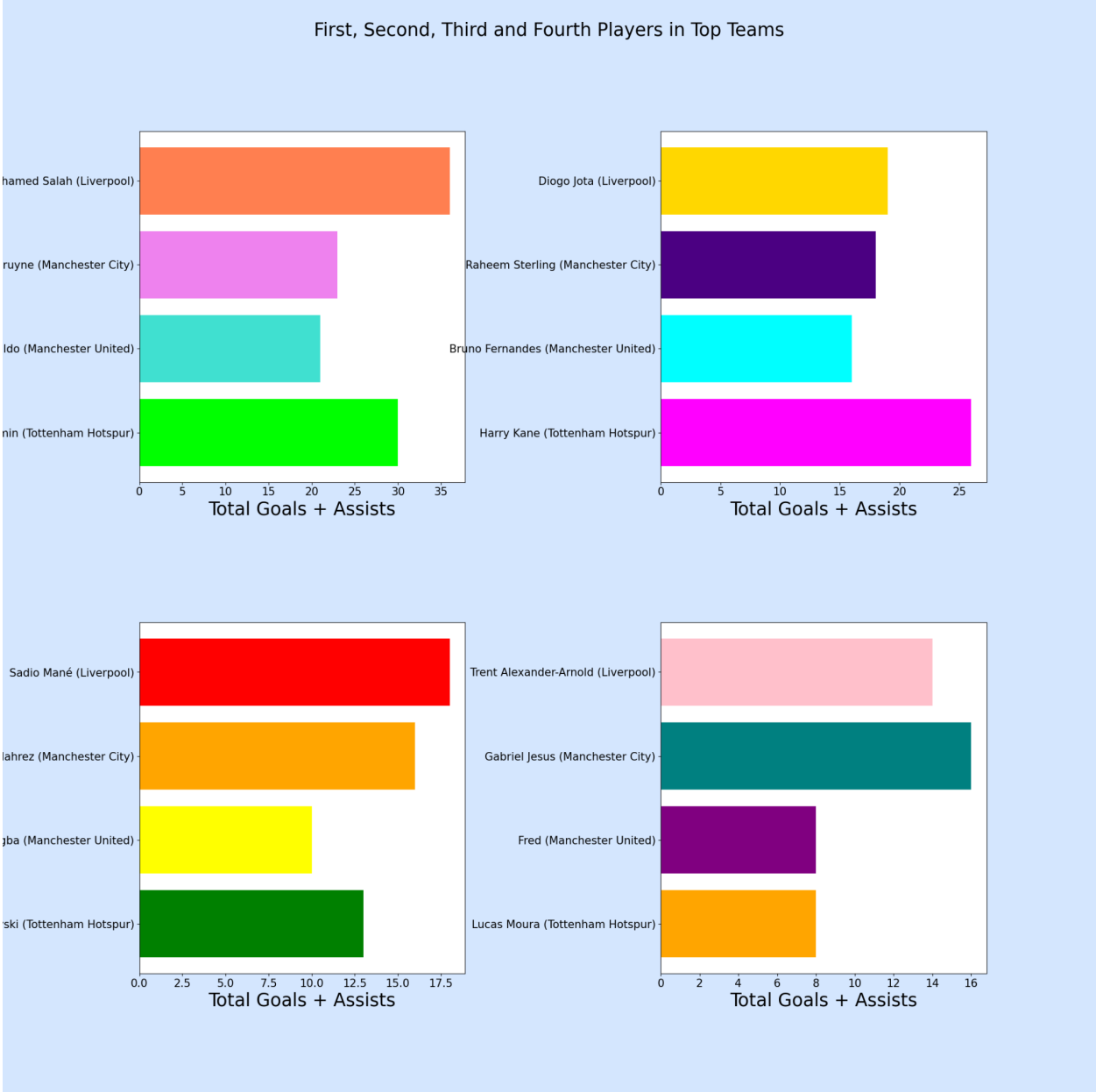
First, Second, Third and Fourth Players in Top Teams











5- Data analysis:

When examining the metric of **shots on target** (SOT), the analysis revealed that Liverpool's players performed exceptionally well. The plots clearly indicate that the majority of Liverpool's players consistently displayed a high number of shots on target throughout the season. **This suggests a strong offensive presence and accuracy in their shooting abilities**. However, it is important to note that there was one player

from another team who outperformed Liverpool's best player in terms of shots on target.

In the analysis of **total goals and assists**, it was found that Liverpool's players performed exceptionally well, with the team having the **highest** number of goals and assists compared to other players in the league. This indicates a **strong offensive prowess and a high level of productivity in terms of goal scoring and providing assists**. Liverpool's players have consistently demonstrated their ability to contribute significantly to the team's overall performance through their goal-scoring and playmaking abilities.

The resulting plots showcasing Liverpool's players' performance in shots on target provide valuable insights for both the business and the coach of the team.

Here's how they can make use of these plots:

- ☐ **Business Perspective**:

  Marketing and Merchandising: The plots can be utilized for marketing purposes, showcasing the exceptional shooting abilities of Liverpool's players. This can be leveraged to promote player merchandise, ticket sales, and attract sponsorships.

  Fan Engagement: The plots can be shared on social media and other platforms to engage with fans. Highlighting the impressive shots on target statistics can generate excitement and enhance fan loyalty.

  Negotiations and Transfers: The plots can serve as evidence of the players' performance when negotiating contracts or potential transfers. It can be used to showcase the value and

impact of Liverpool's players, potentially influencing contract negotiations and transfer decisions.

☐ Coaching Perspective:

Tactical Analysis: The plots can aid the coach in tactical analysis by identifying players who excel in shots on target. This information can be used to design strategies and game plans that maximize the team's shooting efficiency and capitalize on the strengths of individual players.

Player Development: The plots can help identify players who consistently perform well in shots on target. The coach can focus on developing these players' skills even further, working on their shooting techniques, positioning, and decision-making in order to optimize their goal-scoring capabilities.

Opponent Analysis: The plots can also be used to analyze opponents' weaknesses in terms of shots on target. By identifying teams or players who struggle to defend against shots on target, the coach can devise strategies to exploit these vulnerabilities during matches.

Overall, the resulting plots provide valuable insights for the business side of the team in terms of marketing and negotiations, while for the coach, they serve as tools for tactical analysis, player development, and opponent analysis. Leveraging this information can contribute to the team's success both on and off the field.

Ahmed Hafez Questions:

Question 2 (Exploratory):

What key performance indicators are most strongly correlated with team success (with respect to possession, shots on target).

1- Define the problem:

The problem at hand is to identify the key performance indicators that are most strongly correlated with team success, specifically in relation to possession and shots on target.

This will allow us to understand which aspects of possession and shots on target have the most influence on a team's success.

The findings can be used by team management, coaching staff, and analysts to develop strategies, make informed decisions, and optimize team performance in possession a*nd ge*nerating shots on target.

2- Data collection:

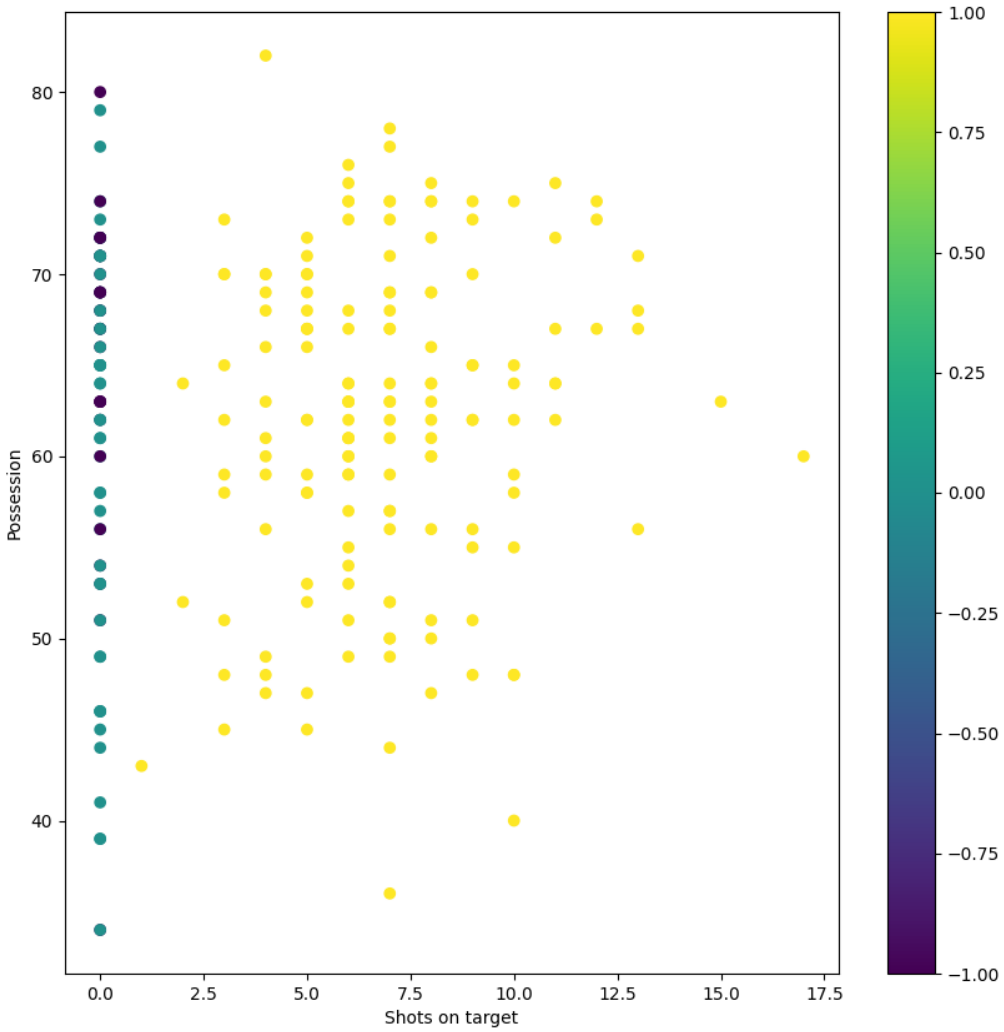
for possession: download the last 6 seasons from https://fbref.com/en/

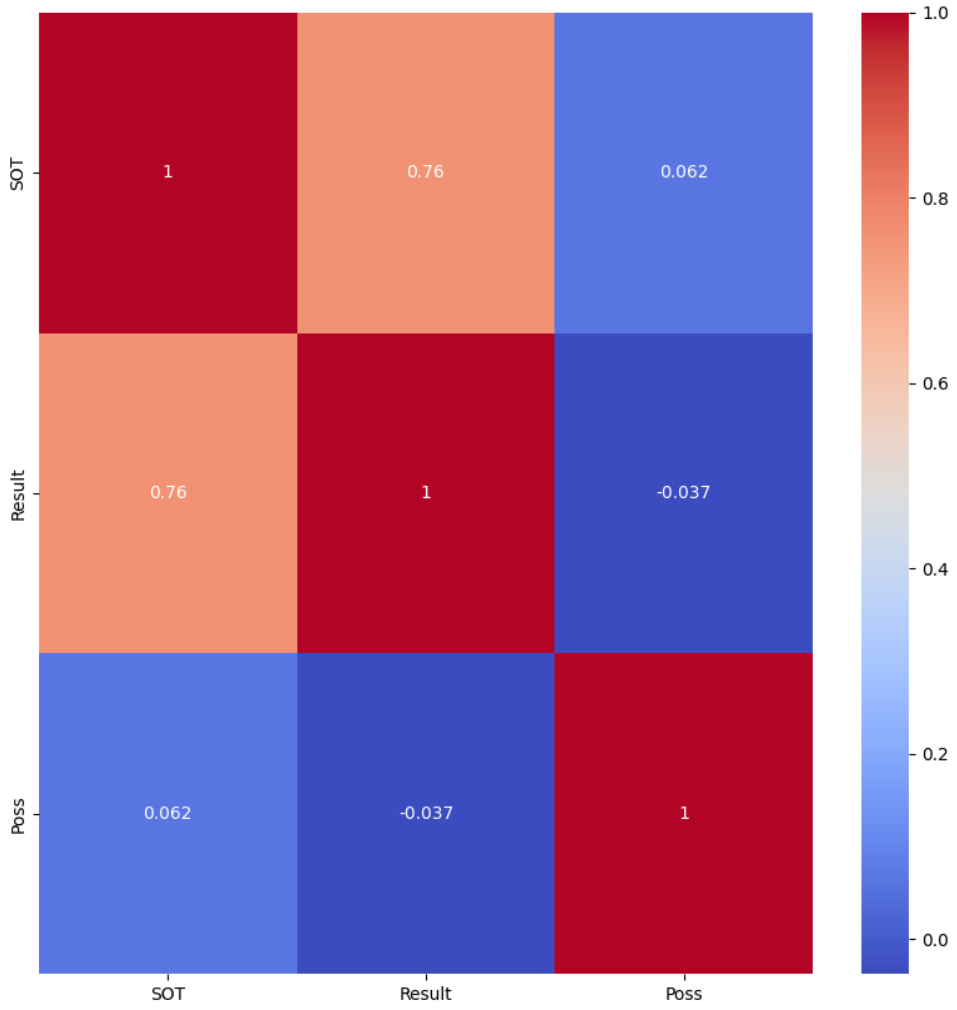
for shots (OT): download the last 6 seasons from https://fbref.com/en/

3- Data cleaning and preprocessing:

- for possession: concatenate all seasons and get poss and result and the match.
- for (SOT): concatenate all seasons and get (SOT)and result and the match and that after determining the season and either away/home.
- then merge the same matches together

4- Exploratory data analysis (EDA):




SOT
Result
Poss
1
0.76
0.062
0.76
1
-0.037
0.062
-0.037
1
SOT
Result
Poss
1.0
0.8
0.6
0.4
0.2
0.0

5- Data analysis:

Possession:

The analysis revealed a lack of significant correlation between possession and team success. **The correlation coefficient was found to be negative, suggesting that higher possession does not necessarily lead to greater team success**. The correlation coefficient value was **[-0.037].** This finding implies that possession alone may not be a reliable indicator of team success.

Shots on Target (SOT):

The analysis demonstrated a ***strong positive correlation between SOT and team success***. The correlation coefficient indicated a significant relationship between the number of shots on target and the likelihood of team success. A higher number of shots on target was associated with a greater probability of scoring goals and winning matches. The correlation coefficient value was **[.76].**

The strong positive correlation between SOT and team success **highlights the significance of focusing on the quality and accuracy of shots**.

6- Communicating results:

☐ For Coach:

Rethinking Possession Strategy: The lack of correlation between possession and team success challenges the notion that possession-based strategies alone guarantee favorable outcomes. Teams should reassess their approach to possession and focus on making the most of their possession time by creating meaningful scoring chances while maintaining a balance between offensive and defensive tactics.

Shot Accuracy: Coaches can emphasize the importance of shot accuracy and train their players to improve their ability to hit the target. By focusing on quality shots on target, teams can increase their chances of scoring goals and winning matches. Tactical ***Adjustments:*** Coaches can use the correlation between shots on target and team success to inform their tactical decisions. They can devise strategies that maximize their team's ability to create scoring opportunities and encourage their players to take more shots on target during matches.

☐ For Business:

Player Recruitment: Businesses involved in player scouting and recruitment can consider the correlation between shots on target and team success when identifying potential players. *Players with a track record of high shot accuracy and the ability to generate shots on target can be seen as valuable assets to improve team performance*.

Sponsorship and Advertising: Businesses that sponsor or advertise with teams can leverage the correlation between shots on target and team success to enhance their marketing strategies..

Ahmed Hafez Questions:

Question 3 (Inferential):

Is there a statistically significant relationship between Liverpool's performance(with respect to Goals Scored) and the number of fouls committed by their opponents, and if so, what is the nature of this relationship?

1- Define the problem:

The problem at hand is to investigate the existence of a statistically significant relationship between Liverpool's performance and the number of fouls committed by their opponents. Additionally, if a relationship is found, the nature of this relationship needs to be determined.

By conducting a statistical analysis, we can ascertain whether this relationship is statistically significant and evaluate the strength and direction of the associationUltimately, understanding the nature of this relationship can assist the team's management, coaching staff, and fans in optimizing their approach to matches and developing effective strategies.

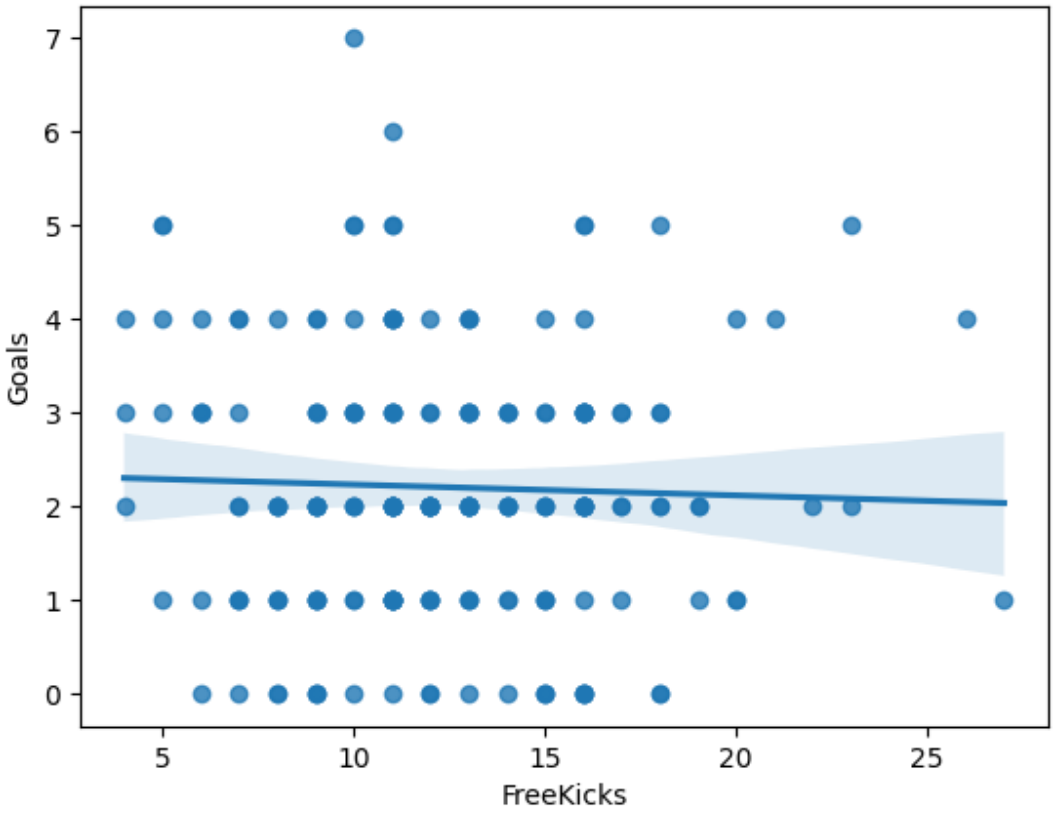
2- Data collection:

download data for the last 6 seasons for the premier league from https://shorturl.at/dyJU9

3- Data cleaning and preprocessing:

- extract the needed features from the dataset
- find all matches of liverpool either away/home
- get the goals scored and fouls committed by the opponents
- concatenate the 6 season

4- Exploratory data analysis (EDA):



5- Data analysis:

To investigate the relationship between Liverpool's performance (goals scored) and the number of fouls committed by their opponents, a **Pearson correlation coefficient** was calculated. The correlation coefficient measures the strength and direction of the linear relationship between two variables. Upon conducting the analysis, we found that there is **no statistically significant relationship between Liverpool's performance (goals scored) and the number of fouls committed by their opponents.** The Pearson correlation coefficient obtained was close to zero **(r = -0.03, p = 0.63)**, indicating a weak and non-significant correlation.

Pearson Correlation Coefficient: -0.034201995324845 78

p-value: 0.6394499141102308

6- Communicating results:

For the Coach:

- ☐ **Performance Analysis:** The analysis can provide valuable insights to the coach regarding the performance of their team in relation to the number of fouls committed by opponents. The coach can assess whether there is any correlation between fouls committed by opponents and the team's goal-scoring ability. This information can help the coach understand the impact of opponents' playing style and aggressiveness on their team's performance.
- ☐ **Tactical Adjustments:** The findings can guide the coach in making tactical adjustments during matches. If there is no significant relationship between fouls committed by opponents and goal-scoring, the coach may focus on other aspects such as team tactics, player positioning, or set-piece strategies to improve goal-scoring opportunities. It highlights that the team's success in scoring goals is not solely dependent on opponents' fouls.
- ☐ **player development:** coach will focus to train the players to get useful of the fouls committed by the opponents.

For the Business:

- ☐ **Market Advantage**: By targeting players who excel at drawing fouls, the business can gain a competitive advantage in their market. If opponents commit more fouls against the team, it may lead to more free-kick opportunities or penalties, which can significantly impact the team's ability to score goals. This strategy can create a distinct playing style for the team, attract fans, and generate excitement, ultimately benefiting the business through increased ticket sales, merchandise revenue, and brand visibility.

Abdullah Questions:

Question 1 (Descriptive):

1- Define the problem:

to analyze the impact of the formations used by opposing teams on the performance of Liverpool FC. By studying the outcomes of matches against different formations, the aim is to identify patterns, strengths, weaknesses, or tactical aspects that may influence Liverpool's ability to secure victories against specific formations.

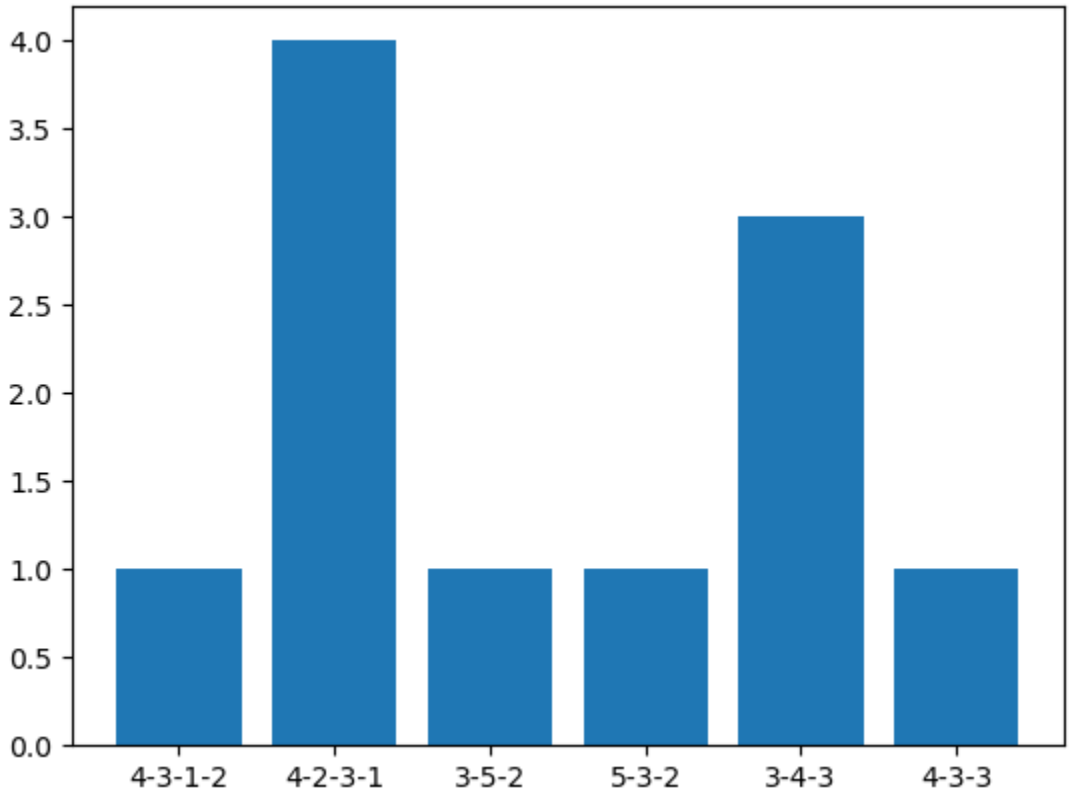
2- Data collection:

for formation: https://fbref.com/en/

3- Data cleaning and preprocessing:

- extract the needed features from the dataset
- Find the matches that liverpool has achieved a negative results
- aggregate on them

4- Result Interpretation:



5- Communicating results:

- **Tactical Insights: Analyzing the impact of opposing formations provides valuable insights into how Liverpool FC's tactical approach aligns or clashes with different setups. It helps identify patterns in terms of strengths and weaknesses in their playing style against specific formations.**

- **Performance Improvement: By understanding the challenges posed by certain formations, Liverpool FC can make strategic adjustments to exploit opponent weaknesses and mitigate their own vulnerabilities. This analysis can guide the team in developing effective game plans and making tactical changes during matches.**

Abdullah Questions:

Question 2 (Predictive):

1- Define the problem:

- **build a predictive model that forecasts the results of matches involving Liverpool FC. By utilizing historical match data, team performance indicators, and other relevant features, the aim is to develop a model that can accurately predict the outcomes of future matches involving Liverpool FC.**

2- Data collection:

- **for matches : the last six seasons from :** https://fbref.com/en/

3- Data cleaning and preprocessing:

- extract the needed features from the dataset
- handle the from different sources
- rename the fields to be convenient and readable
- standardization on the data to make the training faster

6- Data modeling and analysis:

- **Logistic Regression model is a popular machine learning algorithm used for multi classification tasks. In the context of predicting match results for Liverpool FC,we have built a logistic regression model to forecast whether the team will win, lose, or draw a given match based on various features and historical match data**

6- Result Interpretation:

- **The model has effectively captured the essence of home advantage by assigning a higher weight to the home team when predicting match results. This weight reflects the advantage and confidence that comes from playing in front of a home crowd, which can translate into increased motivation, familiarity with the pitch, reduced travel fatigue, and a more supportive atmosphere**

- **Number of shots on target: The number of shots on target by Liverpool FC during a match has proven to be a valuable predictor. This feature captures the offensive capabilities of the team and reflects their ability to create scoring opportunities, thus influencing the likelihood of a favorable outcome.**
- **Possession: The possession percentage of Liverpool FC in a match has demonstrated a strong impact on match results. This feature indicates the team's ability to control the game, maintain ball possession, and potentially dominate the play, which can lead to positive outcomes in terms of goals scored and overall match performance.**
- **The number of intelligent passes made by Liverpool FC players is an additional influential feature. Intelligent passes refer to well-timed and strategically placed passes that exploit the opponent's defensive weaknesses or create advantageous situations for the team. This feature captures the team's tactical awareness and ability to make decisive passes that lead to positive outcomes.**
- **Number of correct passes: The number of accurate passes made by Liverpool FC players during a match has shown a notable influence on match results. This feature reflects the team's ability to maintain fluidity, build effective attacking plays, and control the tempo of the game. Accurate passing contributes to maintaining possession and creating goal-scoring opportunities**

7- Communicating results:

- **for business**
    - **Betting and Fantasy Football: The model's predictions can be utilized by betting companies and fantasy football platforms to offer betting odds, player selection advice, and enhance user experience. This collaboration can**

generate revenue streams through partnerships or licensing agreements with such platforms

- **Accurate predictions of match results can attract sponsors and support marketing efforts. Sponsors often seek partnerships with successful teams, and the ability to predict positive outcomes can enhance the team's marketability and attract potential sponsors**

- **for coach**
  - **Tactical Decision-Making: The model provides insights to the coach regarding the potential outcomes of upcoming matches. This information can guide tactical decision-making, such as team selection, formation, and strategic adjustments, based on the predicted probabilities of winning, losing, or drawing.**
  - **Pre-Match Preparation: The coach can use the model's predictions to prepare the team for specific scenarios and opponents. By understanding the likelihood of different match outcomes, the coach can tailor training sessions, focus on key areas of improvement, and devise strategies that align with the predicted match results**

Abdullah Questions:

Question 3 (Predictive):

1- Define the problem:

- **build a predictive model that forecasts the number of support needed in the attacking position for Liverpool FC. The model aims to provide insights into the optimal number of players required to effectively support the attacking phase of the game, allowing the team to make informed decisions regarding player positioning, substitutions, and tactical adjustments.**

2- Data collection:

- **for statistic about the teams :** https://fbref.com/en/
- **for statistics about the transfers : https://transfermarkt.com/**

3- Data cleaning and preprocessing:

- extract the needed features from the dataset
- handle the from different sources
- joining the data from different sources
- standardization on the data to make the training faster

5- Data modeling and analysis:

- **A linear regression model was employed to predict the required number of support in the attacking position for Liverpool FC. Linear regression is a widely used statistical technique that establishes a linear relationship between the dependent variable (required number of support) and independent variables**

(attacking performance indicators, player positioning data, and tactical information).

6- Result Interpretation:

- **Number of minutes each player has played: The individual playing time of each player during a season has a strong influence on the required number of support in the attacking position. Players who have accumulated more minutes are likely to experience fatigue or require substitution, leading to the need for additional support to maintain the team's attacking effectiveness.**
- **Rank in the last season: The rank achieved by Liverpool FC in the previous season is an important factor that affects the required number of support players in the attacking position. Higher-ranked teams may face stronger opponents and require additional support to overcome defensive strategies, whereas lower-ranked teams may have more opportunities to dominate matches and require fewer support players.**
- **Rank in the last season: The rank achieved by Liverpool FC in the previous season is an important factor that affects the required number of support players in the attacking position. Higher-ranked teams may face stronger opponents and require additional support to overcome defensive strategies, whereas lower-ranked teams may have more opportunities to dominate matches and require fewer support players.**

7- Communicating results:

- **Resource Allocation: Accurate predictions from the model enable the team management to allocate resources effectively. By knowing the optimal number of support players needed in the attacking position, the club can make informed decisions**

regarding player recruitment, squad composition, and budget allocation to strengthen the team's attacking capabilities.