

CAPSTONE PROJECT – UNLEASHING THE FOOTBALL INSIGHTS

STEPS FOLLOWED – BRIEF SUMMARY

- **MS EXCEL** – Using the MS EXCEL analysed all the five excel files given to us.
- **PYHTON** – Using the jupyter notebook, performed the data pre-processing which includes the data understanding, data merging, data formatting, data cleansing and completed visualization for the given business questions framed.
- **MY SQL** – Using the SQL Workbench, created the football database, imported the csv files into the tables and checked the accuracy of data.
- **TABLEAU** – Performing visualizations and dashboard creation for the rest of the business questions framed.
- **POWER POINT PRESENTATION** – Summarized all the business questions framed along with their visualizations from both python as well tableau and drawn conclusions and key findings.
- **MS WORD** – consolidated results and interpretations obtained

Performance Analysis

Geoff Cameron tops the list with 2 red cards with him

Aron Johhansson tops the list with 128 goals

MARKET VALUE

1) How does key performance metrics (like goals, assists) influence the market value of player

Here goals 0, 1 and 2 have high market value which is evenly spread and comparatively 2, 3, 4 are not spread much and does not have, for the assists we have the goals 0, 1, 2 where equally it is spread over a range for the market value.

2) how does metrics like goals , assists correlate with highest market value fluctuations

Here the metrics goals and assists correlation with the highest market value shows that then goals 0 and 1 are highly correlated with highest market value where as 4 is least and coming to assists all 3 values 0, 1, 2 are highly correlated with highest market value.

Team Comparison

1) how does overall performance of both teams in terms of goals scored in past seasons

Here overall performance by home club team is represented using line chart in terms of goals scored in past seasons , so coming to the home club goals it is seen that in the year 2012 is the high goal hit and in 2015 there is a gradual dip in the hit of the goal.

2) Team comparison based on goals for both the teams

Above representation shows the count of goals of both the team goals namely home club goals and away club goals , in home club team it is seen that 1 and 2 are the goals hit more comparatively to other goals and in away club team 1 goal was hit prominently lot more times

Attendance and Stadium Analysis

1) attendance based on the positions of the home club team and away club team

Attendance by the home club based on their positions in the matches and the attendance followed by it is seen above so when their position was no.1 it is seen that they has a lot of people attending the match and for away club their attendance peaked up when their position was no.17

2) find out the highest attendance count of a football match in different seasons

It is seen that from 2012 - 2019 the attendance was constant throughout consecutively with almost 80,000 counts of people and there is a decline with 30,000 count only in the year 2020.

Referee Analysis

1) how the decisions made by referees like red cards yellow cards impact the overall outcome of matches for home club

Match outcome based on yellow card and red card is shown above for home club team , it is seen that home club position with the position number 8is the highest for the number of yellow cards 0, 1 and 2 and for red cards again it is the 8th position but only for the 0 number of red cards

2) how the decisions made by referees like red cards yellow cards impact the overall outcome of matches for home club

the yellow cards 0, 1 and 2 it is the 8th position highest one and the same for red card.

Substitution Patterns

1) Which position of players is substituted the most frequently?

The line chart provides a clear visualization of which player positions are substituted most frequently. This information can be crucial for understanding managerial tactics, player performance management, and how teams adapt during a game

2) How does the average number of substitutions per game vary across different competitions?

The circle chart reveals how the average number of substitutions per game varies across different competitions, with larger circles indicating competitions where teams make more frequent substitutions.

Event Analysis

What are the patterns and impacts of different types of game events on the outcome of matches?

The analysis of game events and their impact on match outcomes reveals several key insights. By examining the frequency and effects of different types of game events—such as substitutions, yellow cards, red cards, and assists—we can uncover significant patterns in how these events influence match results

What is the impact of game events on the final match result?

The chart will show the distribution of match results based on the types of game events, helping you analyze the impact of those events on the final outcome of the matches. The visualization provides actionable insights into how specific game events, such as goals, red cards, and substitutions, influence match results.

Competition Analysis

What are the impacts creating substitute players from different countries in these competitions?

players from a particular country consistently score more goals or provide more assists, it might indicate higher effectiveness or skill levels among those substitutes.

Which competition types see the most strategic substitutions in terms of goals scored afterward?

Average of Goals After Substitution for each Competition Type. Color shows details about Competition Type. The marks are labeled by average of Goals After Substitution. The data is filtered on Season, which ranges from 2013 to 2020.

Player Attributes and Demographics

How do a player's demographics influence their on-field performance during a season?

The in-depth analysis of player demographics reveals how various factors like nationality, age, position, height, and dominant foot intricately shape on-field performance. Players from certain countries consistently lead in key metrics such as goals, assists, and overall minutes, highlighting that certain regions may produce talent particularly suited to high-impact roles

Does a player's market value or demographics influence the manager's substitution decisions during matches with significant attendance?

The visualization reveals key insights into how player market value and demographics influence managerial substitution decisions during high-attendance matches. Players with higher market values appear to be substituted more frequently

Contract Management

Does the presence of high-value players, based on their contract terms, affect the attendance at games?

The analysis reveals a nuanced picture of how high-value players impact game attendance. If the results show that games featuring high-value players consistently attract larger crowds, it suggests that fans are more drawn to matches where star players are present, potentially enhancing the overall game experience and increasing ticket sales.

Do players from certain countries have higher market values?

The visualization of the average market value of players by country reveals notable disparities in player valuations across different nationalities. Players from countries with strong footballing traditions and high-performance leagues, such as Brazil or France, tend to have higher average market values, reflecting their established reputation and demand in the global football market.

SUMMARY

- The dynamics of football have changed in recent seasons due to notable performances and interesting trends. With 119 goals, Aron Johannsson is the most deadly scorer in the league, and Christian Pulsic has shown himself to be an important facilitator with 32 assists. The attendance figures at stadiums speak for themselves: Old Trafford had the highest average attendance of 75,112 people, while May 2019 had the highest attendance of 44,731. Referee Dr. Felix Brych has officiated 79 games on 1 field, making him the busiest, which shows how important he is to keeping the game moving.
- It's interesting to note that during the game, match events like goals and substitutions are dispersed more evenly, with goals

peaking at the 60th minute. Player demographics show that age ranges are consistent amongst contests, with most players peaking in their early 30s.

- Nevertheless, the pandemic's effects are evident because 2020 saw a dramatic drop in objectives and replacements, underscoring the disruption it brought about. Despite this, players' market values continue to be important, especially for younger players in their early 20s who fetch greater fees before starting to decline as they get older.
- Football clubs must strategically manage substitutions, player performance metrics, and market values to gain a competitive edge. Frequent substitutions, player versatility, and game events impact match outcomes, while market value tied to player origin and crowd dynamics informs recruitment and fan engagement.
- High-value players boost both performance and attendance, highlighting their dual importance. Additionally, consistent leadership and strategic management based on data-driven insights help clubs optimize tactics, enhance stadium experiences, and make informed decisions, leading to sustained success in a competitive environment.

Business conclusions :

Performance Analysis

Identify and develop top-performing players to improve our team's overall success

Action:

Christian Pulisic is the top scorer, highlighting a significant goal contribution disparity among players.

The analysis reveals potential areas for improvement by incorporating additional metrics like assists and minutes played .

Team Comparison

Improve our team's defensive performance and analyse the performance disparity between home and away games?

Action:

Teams with better defensive records conceded fewer goals, suggesting effective defensive strategies.

The home field advantage is significant, impacting overall team performance and strategic planning .

Referee Analysis

Referee decisions impact player behaviour and team performance

Action:

Referee bias and player behavior are influenced by inconsistent rule application, affecting game outcomes.

A correlation exists between key events, such as substitutions following goals, which can alter match dynamics .

Player Attributes and Demographics

Player attributes and positions correlate with performance metrics like goals and assists?

Action:

Position Correlation: A strong positive correlation exists between certain positions (e.g., attackers, midfielders) and key performance metrics such as goals and assists. This indicates that these positions are more involved in scoring opportunities.

Competition Analysis

football competitions be increased to match the popularity of domestic leagues

Action:

Dominance of Domestic Leagues: The analysis showed that domestic leagues have the highest participation rates. Efforts should be focused on boosting

engagement in other competitions, such as charity leagues, by tailoring marketing strategies to target specific audience segments.

Income Bracket Insights: Attendance data revealed that the 30k-50k income brackets are the most significant contributors to stadium audiences, suggesting that these groups should be a focus for future marketing campaigns.

Market Value Analysis

distribution of individual market values within our portfolio affect our overall market strategy

Action:

Market Value Distribution: The analysis highlighted the varying levels of market value contributions among players. Top performers like John Anthony Brooks and DeAndre Yedlin were identified as significant assets, suggesting that focusing on high-value individuals could optimize market strategies.

Strategic Insights: Understanding these market value distributions allows for better strategic planning, helping to focus resources on players who offer the greatest financial returns.

Contract management analysis

Trends can be observed in contract expiration values from 2019 to 2023, and how might these trends impact future contract renewal strategies?

The contract expiration values displayed significant fluctuations over the years, with peak periods observed in 2021 and 2022. These peaks indicate crucial times for contract renewals and strategic planning.

The analysis highlighted players with upcoming contract expirations, such as Alejandro Bedoya and DeAndre Yedlin, emphasizing the need for timely contract negotiations to maintain team stability.

Stadium and attendance analysis :

attendance data across different countries to identify key markets and optimize event planning and marketing strategies?

Action:

The analysis showed that countries like the United States, Germany, and the United Kingdom have the highest stadium attendance, making them key markets for event planning and marketing efforts.

Historical data from 2013 to 2020 revealed consistent trends in stadium attendance, which can be used to optimize future event scheduling and marketing strategies.

Substitution analysis

substitution patterns be analyzed to optimize team strategy and player performance?

Action

The analysis identified specific periods, such as during the 2023 season, where substitution counts spiked. Understanding these patterns can help in planning strategic substitutions to maintain player fitness and match momentum.

Players like Brad Guzan were identified as having the highest number of substitutions, indicating a potential focus area for optimizing player rotation and usage.

Event analysis :

strategies can be implemented to improve the club's win rate in both home and away games based on historical match outcome data

Actions :

The data revealed a stronger win rate for home games, indicating a pronounced home-field advantage. This suggests that strategies should focus on replicating this success in away games.

Certain seasons showed better performance than others, likely due to changes in team composition or management. These insights can inform future team-building and tactical decisions.

KEY FINDINGS :

- Aron Johhansson is the leading goal scorer with 119 goals

- Christian Pulisic assists charts with 32 assists
- Highest attendance was seen in 2012 – 2019 and declined in 2020
- Dr Felix Brych was the lead referee for almost 80 games and is among top referees from analysis
- Player age ranges are from 25-35 for the domestic league, domestic cup and international cup
- Attackers and midfielders have highest average market value and goalkeepers with low averages comparatively
- Home club is generally scored more than the away club from the analysis
- Substitution peaked in 2012 and declined at 2020 just like seasonality of the attendance of matches