FOOTBALL ANALYSIS

KAARJALA LAKSHMIKANTH

PROBLEM STATEMENT

- Analyze the distribution of yellow and red cards among players.
- Identify any patterns in team strategies based on the competition type.
- Determine if players who play more minutes tend to have higher or lower goal and assist numbers.
- Identify key players who have a significant impact on their team's performance.
- Investigate if certain referees are associated with more yellow or red cards.
- Explore whether managerial changes during a season have a significant impact on team performance.

OBJECTIVE

- Data have 26 columns (variables).
- There are 3568 rows (observations).
- Data have some missing information’s.
- Build models to predict future outcomes, such as match results or player performance, based on historical data.
- Identify key players based on their contributions to the team.
- Evaluate referee performance by analyzing their involvement in various games.

Data Cleaning & Pre processing

❖ Checking for Null values

```
In [18]: # checking the number of missing values in each column of the dataframe
         df.isnull().sum().sort_values(ascending=True)
```

❖ Displaying the dataset information

```
In [106]: print(df.info())    # data information
```

❖ Replace missing values

```
In [89]: x = df["away_club_position"].mean()
         df["away_club_position"].fillna(x, inplace = True)

In [93]: df["home_club_manager_name"].fillna(0, inplace = True)
```

❖ Merge all files

```
In [68]: merged_df = pd.merge(df1, df2, on='game_id', suffixes=('_df1', '_df2'))
         merged_df1 = pd.merge(merged_df, df3, on='game_id', suffixes=('_merged', '_df3'))
         merged_df = pd.merge(merged_df, df4, on='game_id', suffixes=('_merged', '_df4'))
         merged_df = pd.merge(merged_df, df5, on='player_id', suffixes=('_merged', '_df5'))
```

❖ Dropping unwanted columns

```
In [84]: df=df.drop(['home_club_formation','away_club_formation'],axis=1) # this two column removed
```

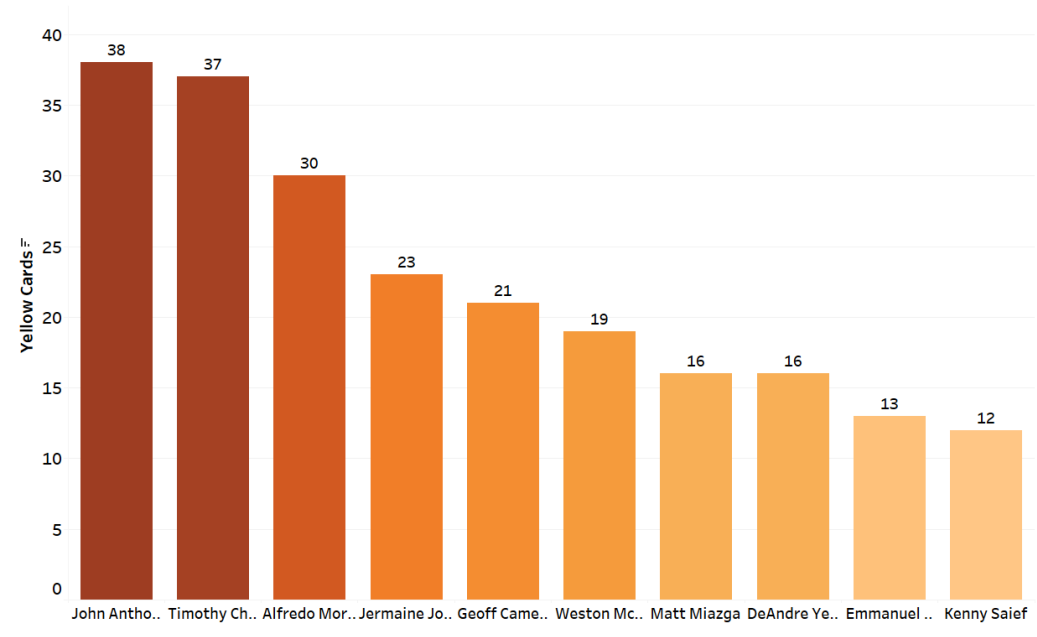
Observations

- home_club_formation column is more then 90% missing values there.That reason drop the two columns.

❖ Display the names of all columns

```
In [105]: print(df.columns)    # all columns name
```

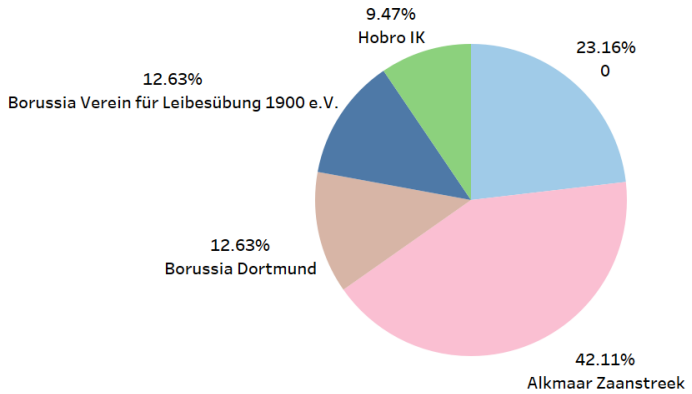
THE TOP 10 FOOTBALL PLAYERS WITH THE MOST YELLOW CARDS



OBSERVATIONS:-

John Anthony Brooks is the player with the most yellow cards, while Kenny Saief is also in the top 10 for receiving yellow cards.

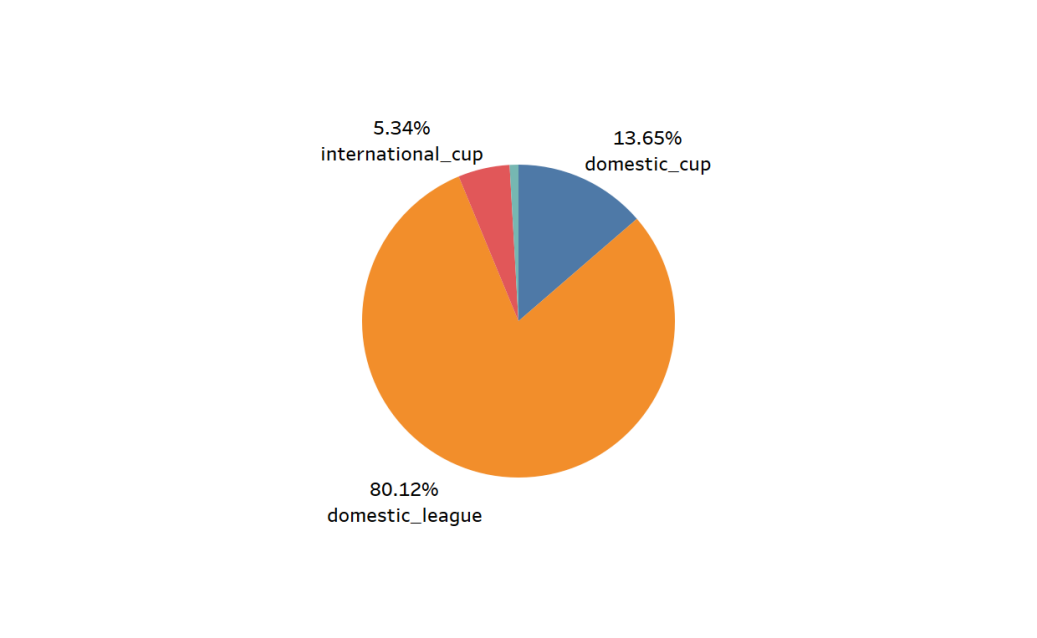
THE LEADING GOAL-SCORING CLUBS REVEALED



OBSERVATIONS:-

42.11% of the goals scored are received by the Alkmaar Zaanstreek club.

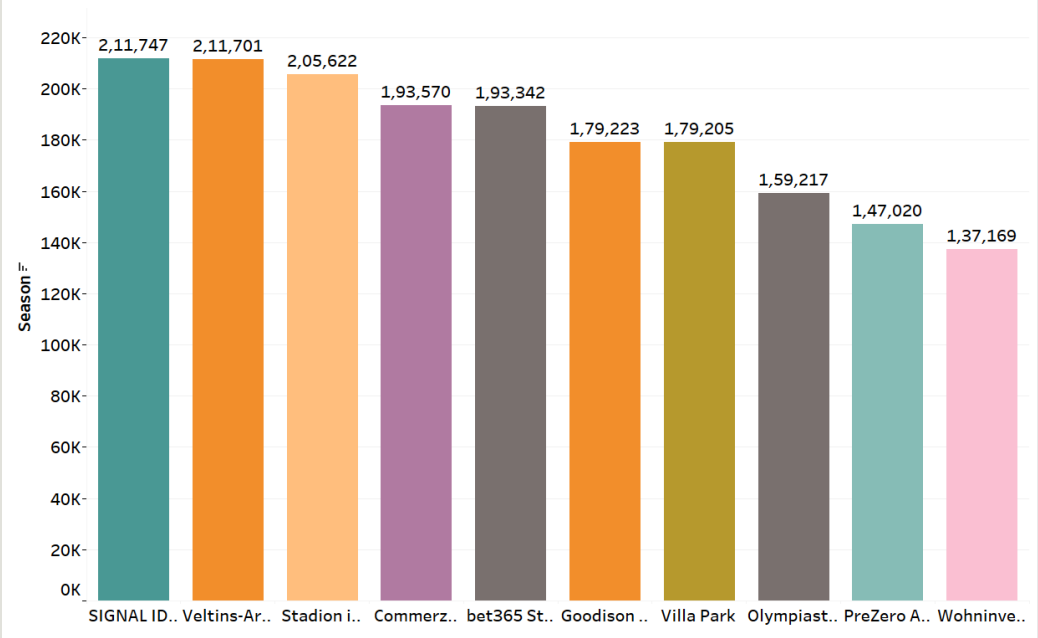
THE LEADING GOAL SCORERS BY COMPETITION



OBSERVATIONS:-

Majority of 80.12% goal for the domestic league.

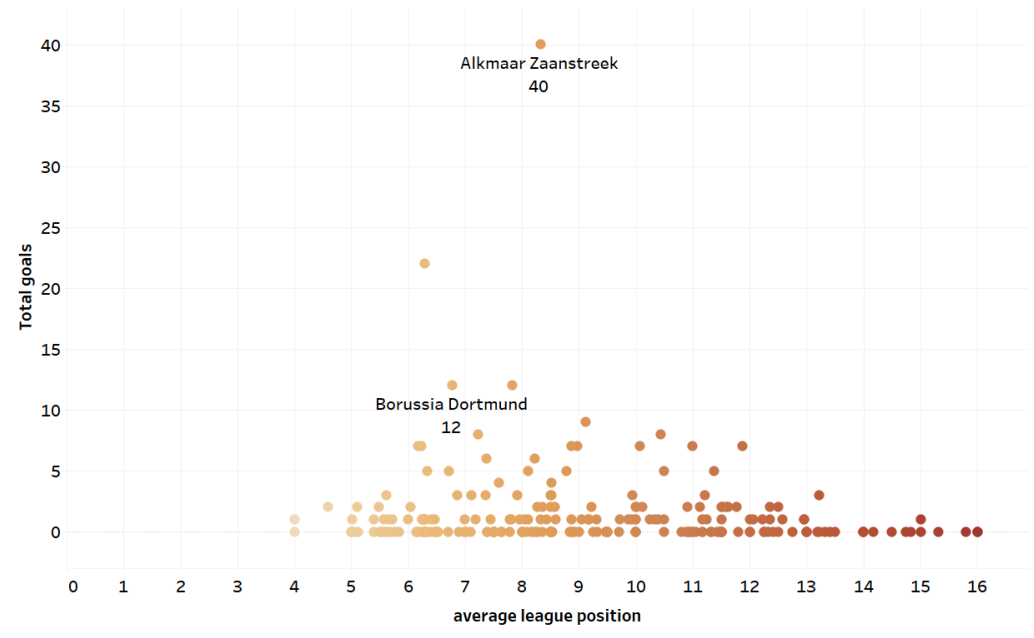
Seasonal Peaks in Football Stadium Performance



OBSERVATIONS:-

That Signal Iduna Park consistently performs well throughout the season.

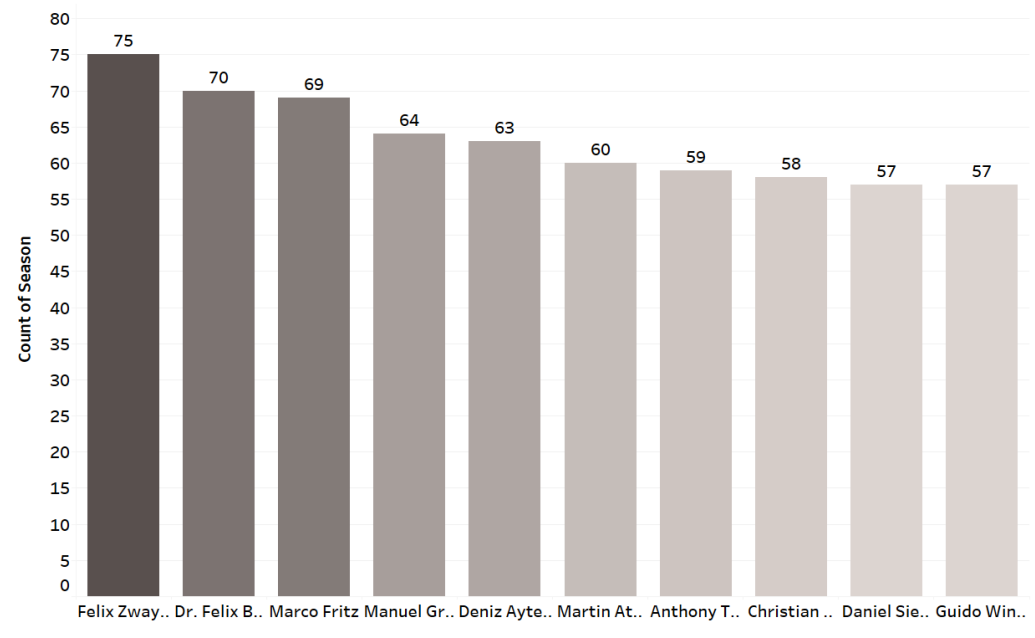
Which clubs have consistently performed well in terms position in the league



OBSERVATIONS:-

The Alkmaar Zaanstreek club consistently performs well.

Seasonal Referee Performance Breakdown



OBSERVATIONS:-

75 matches of refereeing, Felix Zwayer has accumulated a significant amount of experience in officiating matches.

THE MOST YELLOW CARDS IN A 2013 SEASON

home_club_name	total_yellow_cards
besiktas jimnastik kulubu	4
Stoke City	4
1.FC Nuremberg	4
Sunderland AFC	3
Panthessalonikios Athlitikos Omilos Konstantinou...	3
0	3
FC Schalke 04	3
Chelsea Football Club	3
Football Club de Nantes	2
Paris Saint-Germain Football Club	2

OBSERVATIONS:-

Besiktas jimnastik kuludu , Stoke city these two home club get the highest.

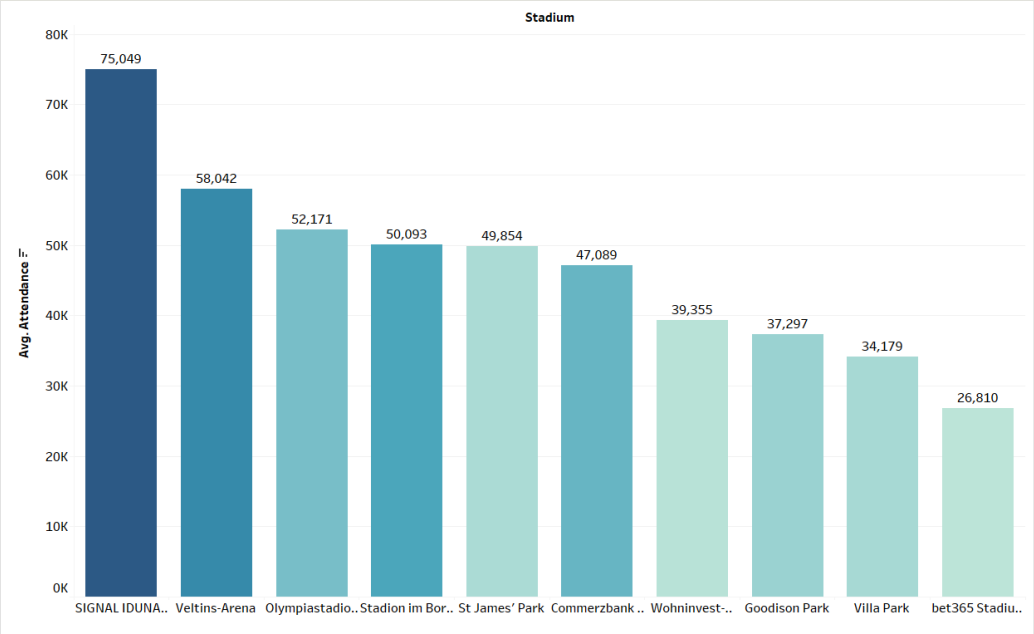
THE HIGHEST AVERAGE HOME ATTENDANCE

home_club_name	avg_home_attendance
Manchester United Football Club	75088.21052631579
Borussia Dortmund	75048.50635891038
Real Madrid Club de Football	67739
FC Bayern Munich	66763.61966474152
Futbol Club Barcelona	61768.5
club atletico de madrid sad	61023
Arsenal Football Club	58701.50161236144
FC Schalke 04	58041.72660428364
The Celtic Football Club	52546.4
Hamburger SV	51176.489795918365

OBSERVATIONS:-

Manchester united Football club is highest average home attendance.

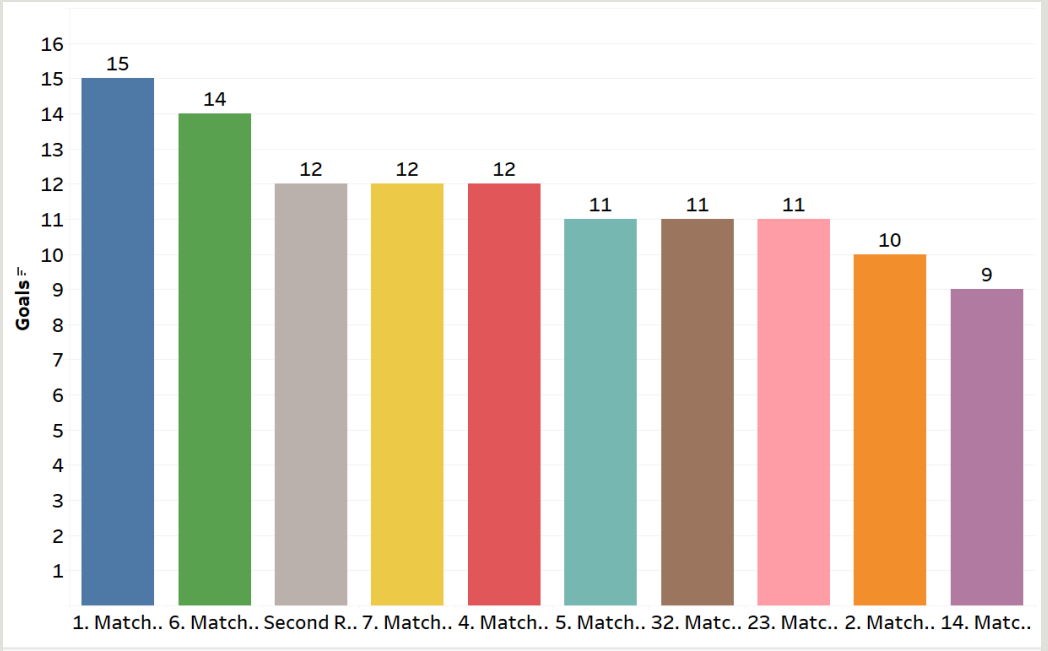
AVERAGE ATTENDANCE AT LEADING FOOTBALL VENUES



OBSERVATIONS:-

The average attendance at Signal Iduna Park is 75,049.

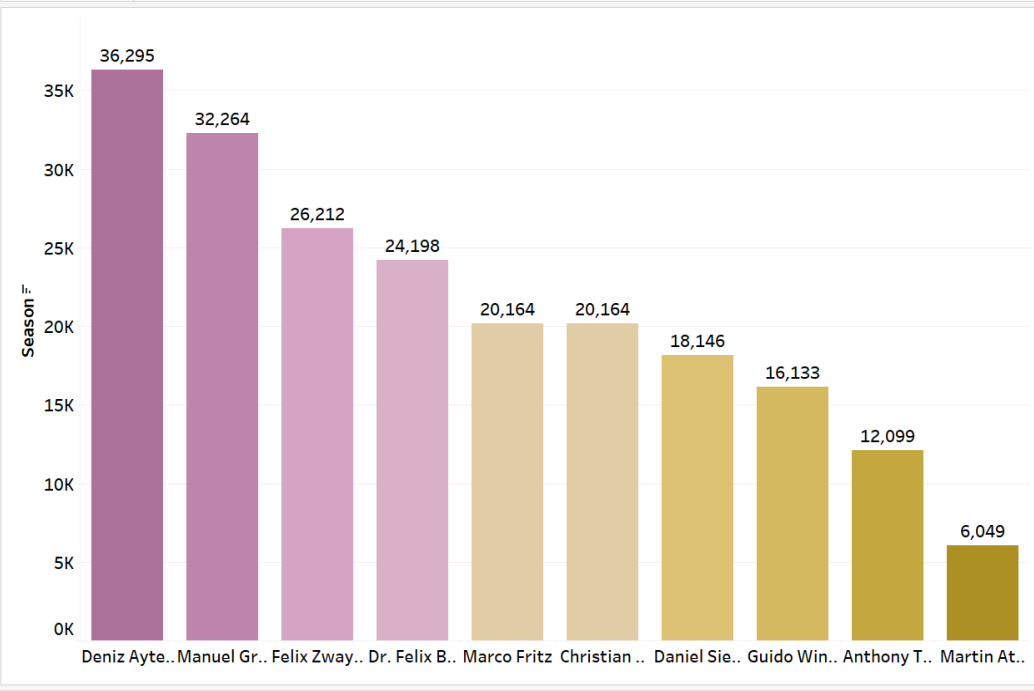
THE HIGHEST GOAL DISTRIBUTION WISE DIFFERENT ROUNDS



OBSERVATIONS:-

The 1st matchday round highest goals.

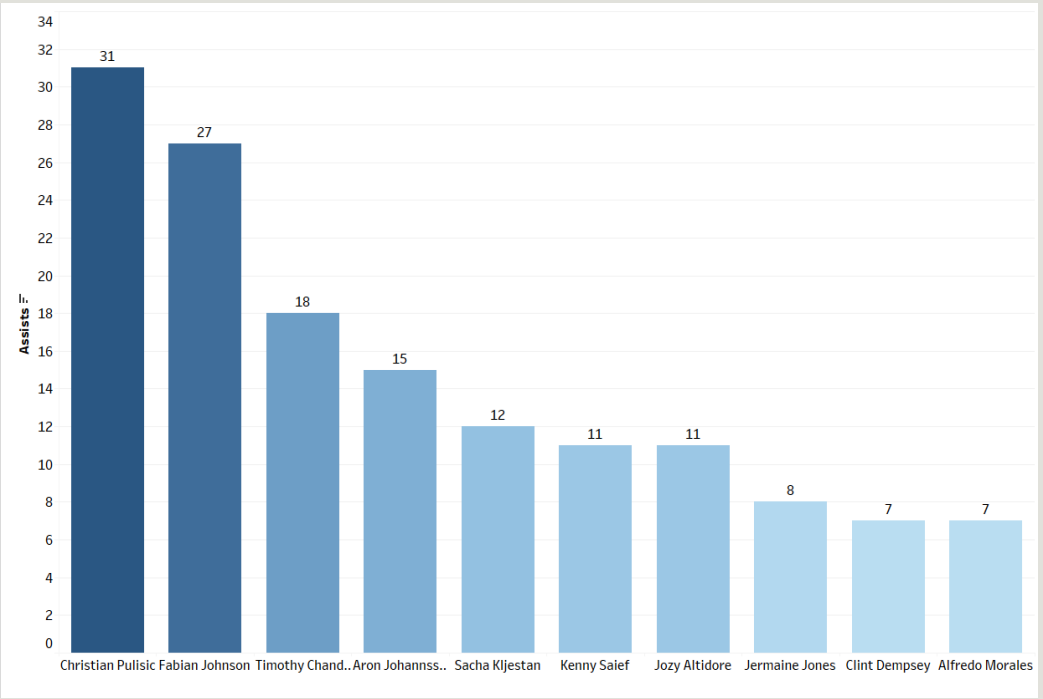
TRACKING THE TOP REFEREES 2017 YEAR



OBSERVATIONS:-

Deniz Aytekin 36,295 matches in 2017 year.

ASSISTS WISE TOP 10 PLAYES



OBSERVATIONS:-

Christian Pulisic 31 assists.

THE INFORMATION ABOUT MATCHES REFEREED BY A LEE MASON

game_id	home_club_name	away_club_name	referee
2225471	Aston Villa Football Club	Swansea City	Lee Mason
2225471	Aston Villa Football Club	Swansea City	Lee Mason
2332184	Newcastle United Football Club	Everton Football Club	Lee Mason
2332288	Aston Villa Football Club	Southampton FC	Lee Mason
2578002	Arsenal Football Club	Everton Football Club	Lee Mason
2578267	Everton Football Club	Norwich City	Lee Mason
3050436	West Ham United Football Club	Fulham Football Club	Lee Mason
2225511	Southampton FC	Tottenham Hotspur Football Club	Lee Mason
2225511	Southampton FC	Tottenham Hotspur Football Club	Lee Mason
2332038	Swansea City	Everton Football Club	Lee Mason

OBSERVATIONS:-

lee mason top 10 refereed.

THE MATCHES WHERE BOTH TEAMS SCORED THE MOST GOALS

game_id	home_club_name	away_club_name	home_club_goals	away_club_goals
2378598	Alkmaar Zaanstreek	Sportclub Heerenveen	9	8
2835837	Borussia Verein für Leibesübung 1900 e.V.	eintracht frankfurt Football ag	7	8
3319622	0	Fortuna Dusseldorf	8	7
2699117	0	FC Ingolstadt 04	7	8
3096455	1. FC Koln	FC Schalke 04	6	7
3192360	Liverpool Football Club	Chelsea Football Club	7	6
2862155	Borussia Dortmund	FC Bayern Munich	6	7
2755715	Borussia Dortmund	0	8	4
3122484	Borussia Dortmund	Sportverein Werder Bremen von 1899	5	7
3195519	0	SC Paderborn 07	5	7

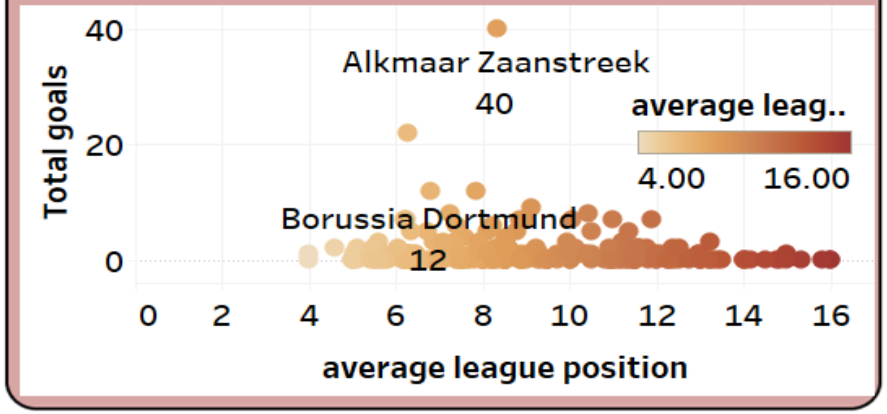
OBSERVATIONS:-

The Alkmaar Zaanstreek club consistently performs well.

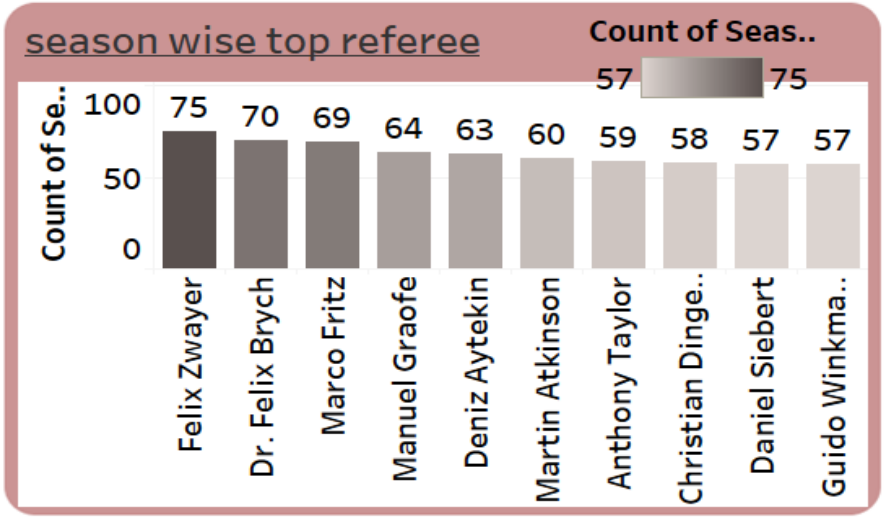
DASH BOARD

FOOT BALL ANALYSIS

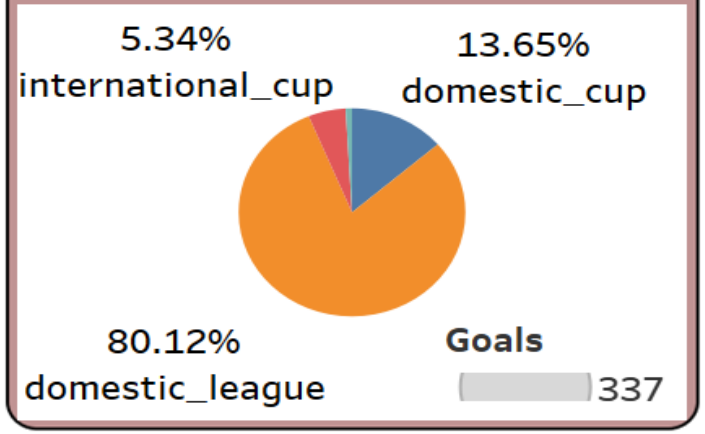
Which clubs have consistently performed well in terms position in the league



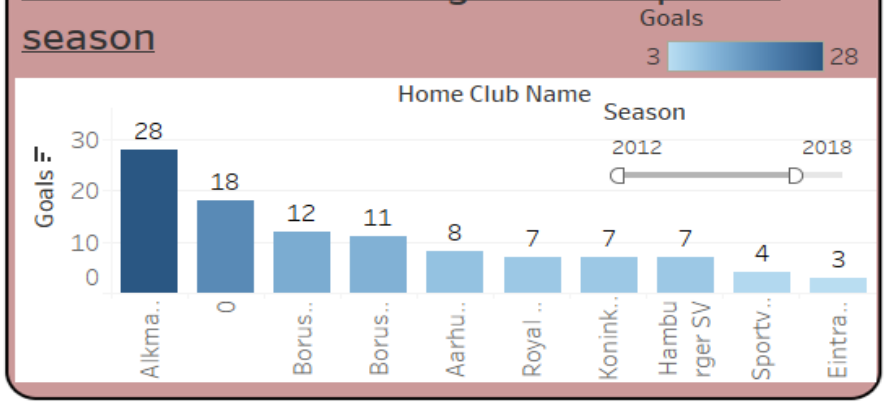
season wise top referee



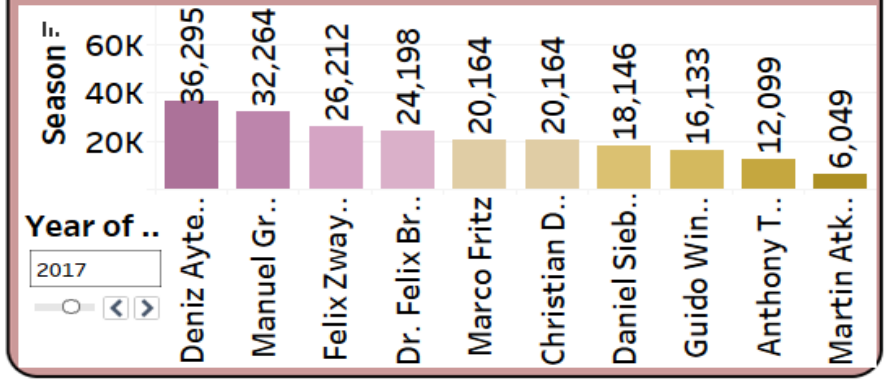
Competition type wise goals



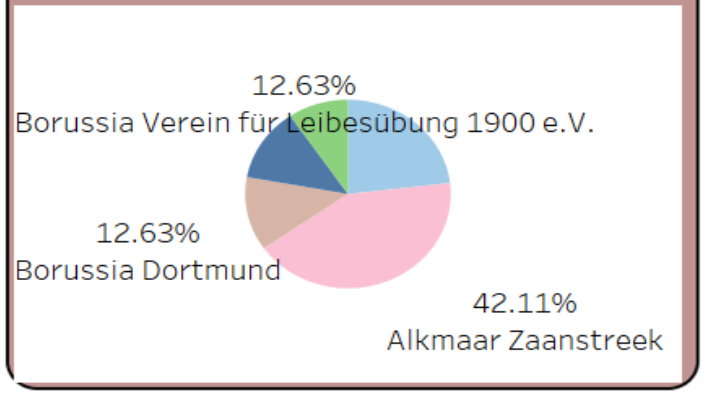
clubs scored the most goals in a specific season



Tracking the Top Referees Year by Year



Distribution of Goals by Club



SUMMARY

- The football season during which the game took place.
- Number of goals scored by the home team.
- Position of the home team in the competition.
- Number of spectators or audience present at the game.
- Aggregate score in case of a two-legged match.

CONCLUSION

- Evaluate the outcome of each game through home club goals and away club goals.
- Identify players and their appearance details in a match, essential for player-focused analysis.
- Identify the managers of the home and away teams, which could influence team strategy and performance.
- Represents the date of the match, facilitating time-based insights.
- Provide the league position of the home and away teams, aiding in understanding team standings.

Thank you