FIFA World Cup Analysis

Aarya Yogesh Pakhale
Data Analyst Intern
IIT KHARAGPUR

TABLE OF CONTENTS

INTRODUCTION

The FIFA World Cup, often simply called the World Cup, is an international association football competition contested by the senior men's national teams of the members of the Fédération Internationale de Football Association (FIFA), the sport's global governing body. The championship has been awarded every four years since the inaugural tournament in 1930, except in 1942 and 1946 when it was not held because of the Second World War. The current champion is Germany, which won its fourth title at the 2014 tournament in Brazil.

This report serves as a comprehensive exploration of the statistical landscape of the FIFA World Cup, spanning decades of matches, teams, and players. Through careful analysis and interpretation of data, we seek to uncover patterns, trends, and insights that illuminate the dynamics of this unparalleled sporting spectacle. This exploratory data analysis is carried out using the tool Tableau Public.

OBJECTIVE

The primary objective of this FIFA World Cup data analysis project is to examine and interpret the vast array of statistical data associated with the tournament. Through rigorous analysis, our aim is to:

1. Identify patterns and trends: We seek to uncover recurring patterns in team performance, player statistics, and match outcomes across various editions of the FIFA World Cup.
2. Gain insights into team strategies: By analyzing data on tactics, formations, and playing styles, we aim to understand the strategic approaches adopted by successful teams throughout World Cup history.
3. Examine historical context: We aim to contextualize our findings within the broader historical and socio-cultural context of the FIFA World Cup, considering factors such as host nations, tournament formats, and geopolitical influences.

4 Inform future decision-making: By drawing conclusions from our analysis, we aspire to provide valuable insights that may assist football stakeholders, including coaches, players, analysts, and administrators, in making informed decisions related to team strategies, player selection, and tournament preparation.

Overall, our objective is to present a comprehensive and insightful analysis of FIFA World Cup data that contributes to a deeper understanding of the tournament's dynamics and enriches discussions surrounding the world's most prestigious football event.
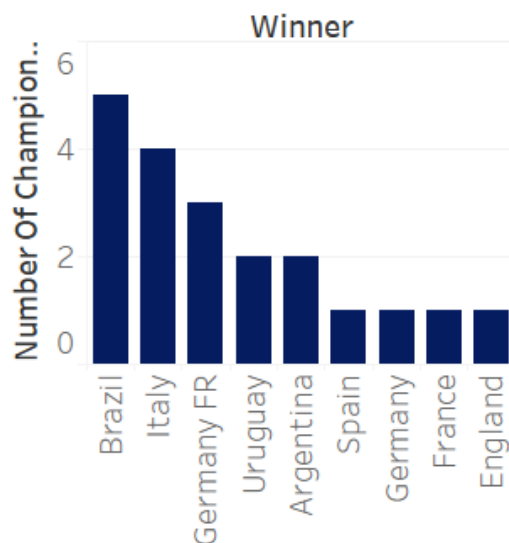
DATASET

The World Cups dataset show all information about all the World Cups in the history, while the World Cup Matches dataset shows all the results from the matches contested as part of the cups. The Folder consists of 3 files:

- WorldCupMatches.csv- This dataset contains details regarding the Year, date-time, stadium, city, home team and away team performance , the referees and their assistants, the win conditions and the match details.

- WorldCupPlayers.csv- This dataset provides information regarding, the player names, jersey number, position, event and lineup

- WorldCups.csv- This dataset gives insights regarding the host countries, attendance, the winners, the runner ups, number of qualified teams, goals scored in the different editions
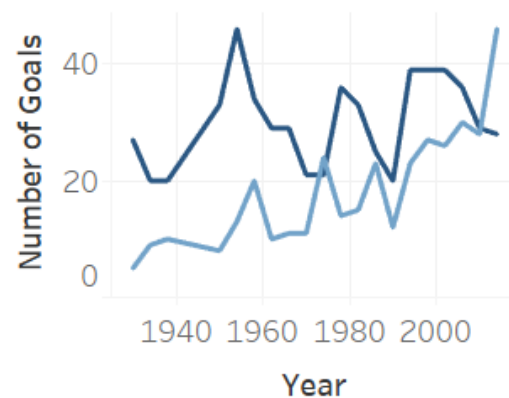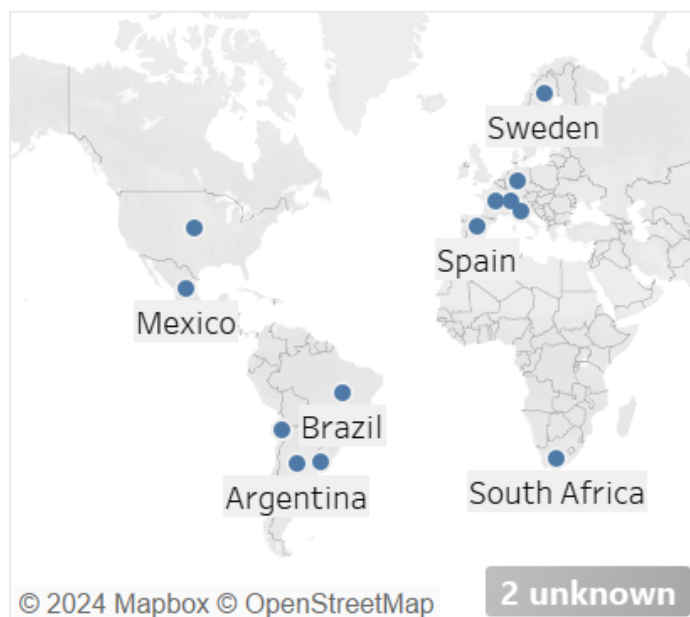
TABLEAU DASHBOARDS

Goals Scored by Winning Teams

Winner

Goals Scored: 600, 400, 200, 0

Brazil, Italy, Germany FR, Argenti..., Germany, France, Uruguay, Spain, England

Distribution of Shirt Number

0	6	11	12	14	15
	7				
1	8	16	5	17	20
2		18			
3	9		13		22
4	10	19	21		

Effect of Win Conditions

Win Condi..

win after ..

Value: 60, 40, 20, 0

Away Team Goals, Home Team Goals
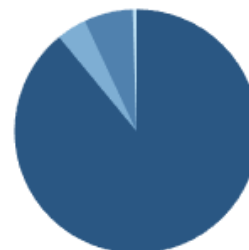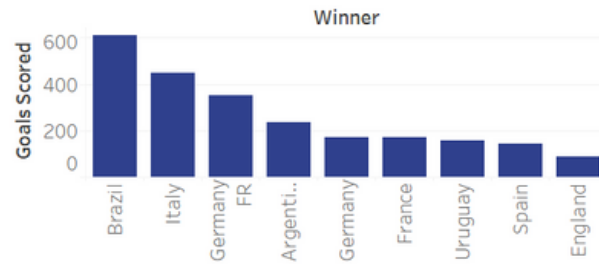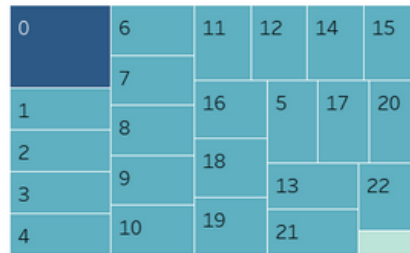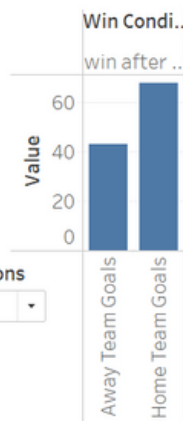
Line Up Distribution

Line-up

| N | 19,037 |
| S | 18,744 |

Choose Win Conditions

win after extra time

Goals Scored by Winning Teams
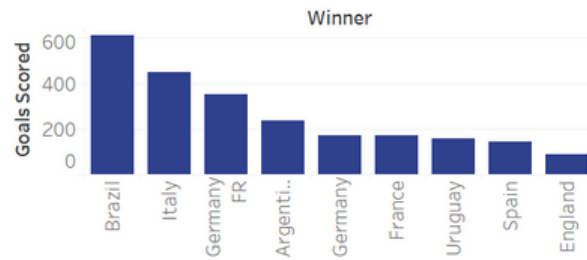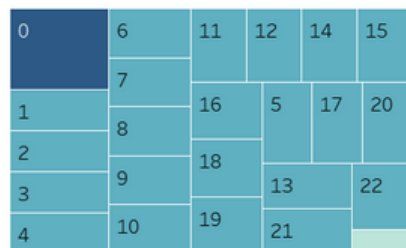
Winner

Goals Scored: 600, 400, 200, 0

Brazil, Italy, Germany FR, Argenti..., Germany, France, Uruguay, Spain, England

Distribution of Shirt Number

0	6	11	12	14	15
	7				
1	8	16	5	17	20
2		18			
3	9		13		22
4	10	19	21		

Effect of Win Conditions

Win Condi..

win on pe..

Value: 20, 10, 0

Away Team Goals, Home Team Goals
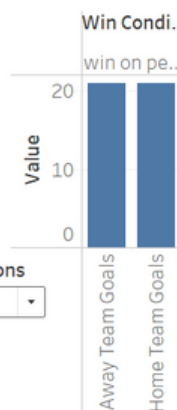
Line Up Distribution

Line-up

| N | 19,037 |
| S | 18,744 |

Choose Win Conditions

win on penalties

Goals Scored per Country
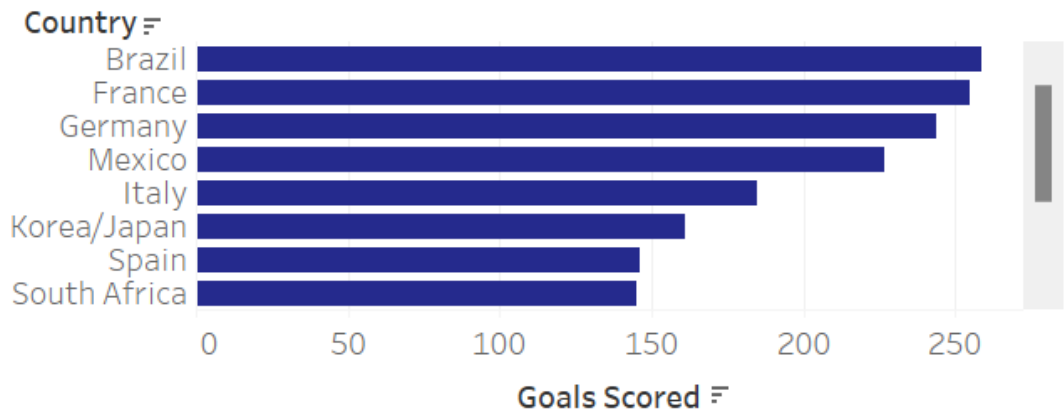
Country

Country	
Brazil	
France	
Germany	
Mexico	
Italy	
Korea/Japan	
Spain	
South Africa	

0   50   100   150   200   250

Goals Scored

Attendence over the Years

Attendance

4M

2M

0M

1940 1960 1980 2000

Year

Qualified Teams

Qualified Teams

30

20

10

0

1940 1960 1980 2000

Year

Stadiums With Highest Maximum Attendence

Stadium

Stadium	
Maracanã · Est..	
Estadio Azteca	
Wembley Stadium	
Camp Nou	
Rose Bowl	
Santiago Bernab..	
Soccer City Stadi..	
Stanford Stadium	
Olympiastadion	
Stade de France	

0K   40K   80K   120K   160K
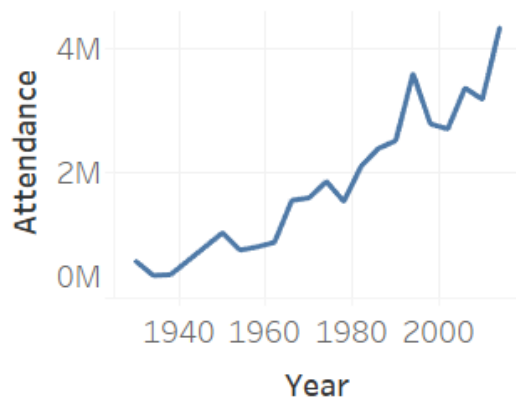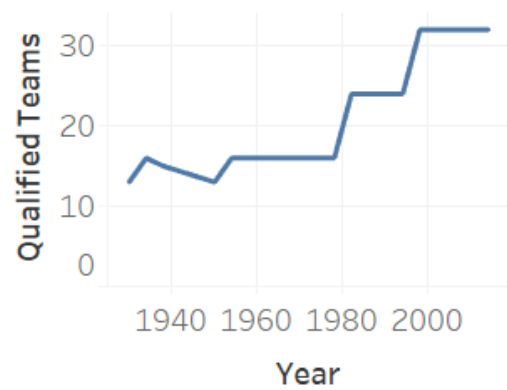
Max. Attendance
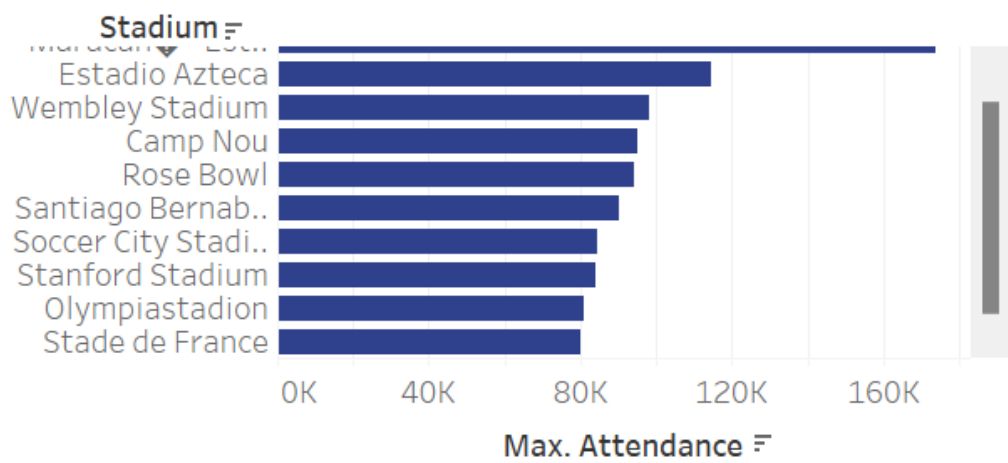
EXPLORATORY DATA ANALYSIS

Top Winners
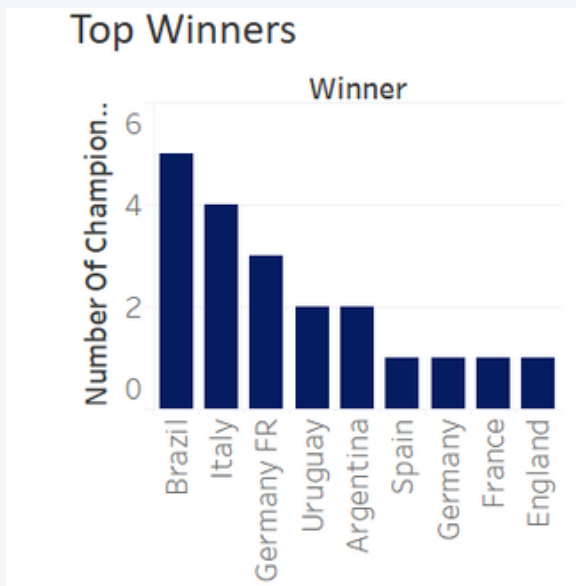
Brazil is the country which has won the title most number of times (5 times) followed by Italy (4 times), Germany FR (3 times), Uruguay (2 times), Argentina (2 times),Spain, Germany, France and England.

This graph shows the analysis of the number of goals scored till the half time by the Home team and the away team with each passing edition of the tournament. It indicates that the advantage of being the home team team has aggregately decreasing with time. Earlier the goals scored by the home team were usually higher than the goals by the away team due to the effect of the home support. But, today the trend shows that the away teams are performing much better than the home teams. One reason can be the increased attention to the mental and psychological strength of the players, so they are now less affected by the home crowd. One more possible reason can be the increase in the international support for the players and better connectivity for transport. Better exposure to better training methods to get accustomed to different playing fields.

Home Team Vs Away Team

Position Of Players
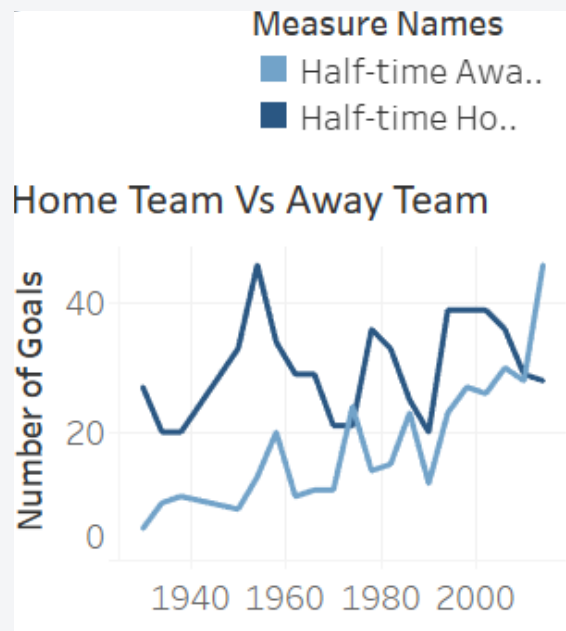
Position
- ■ Null
- ■ C
- ■ GK
- ■ GKC
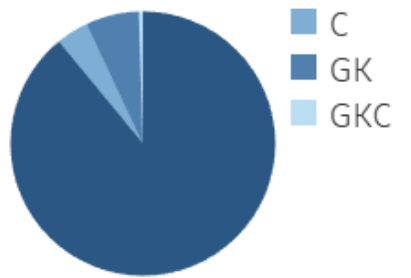
This graph shows the distribution of the position of players. Maximum players have no specific position. Goalkeepers and Captains are approximately equal and are the next big population among the players. Goalkeeper-cum- Captains are the least in population.

The N line is more common among the players as compared to the S lineup.

Line-up	
N	19,037
S	18,744

The preference for the "N" line over the "S" line in FIFA can be attributed to several factors:

- Tactical Advantage: The "N" line, typically referring to the central or middle area of the pitch, is strategically crucial in football. This area allows for better control of the game, as it provides access to both offensive and defensive plays. Teams often prioritize control of the midfield to dictate the pace of the game and create scoring opportunities.
- Key Player Positions: Many of the most influential players on the field operate within the "N" line, including central midfielders, playmakers, and defensive midfielders. These players play pivotal roles in orchestrating attacks, breaking up opposition plays, and maintaining possession. As a result, teams often focus on strengthening their presence in the midfield to gain a competitive edge.
- Transition Play: The "N" line serves as a crucial zone for transitional play, where teams transition from defense to offense and vice versa. By dominating the midfield, teams can disrupt their opponents' rhythm, regain possession, and launch quick counter-attacks. This ability to control transitions can significantly impact the outcome of matches.
- Defensive Stability: Strong defensive presence in the midfield can provide additional support to the defensive line, making it more difficult for opponents to penetrate and create scoring opportunities. By congesting the "N" line, teams can effectively neutralize opposition attacks and limit their ability to build sustained pressure.

Distribution of Shirt Number
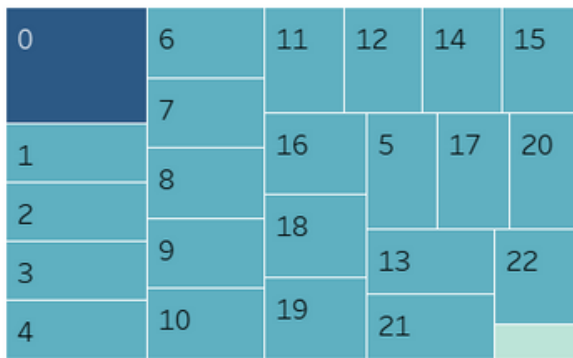
This tree plot shows the distribution of the Shirt Numbers of the players. The number '0' is the majority whereas the number '23' has the minority. Rest numbers are fairly evenly distributed.

The adjacent graph shows that the home teams have a higher win rate when it comes to win after extra time due to home crowd support and their preferred playing conditions. Away teams and Home teams perform equally good when it comes to win after penalties, i.e, the home teams don't have a better winning rate in win on penalties condition.
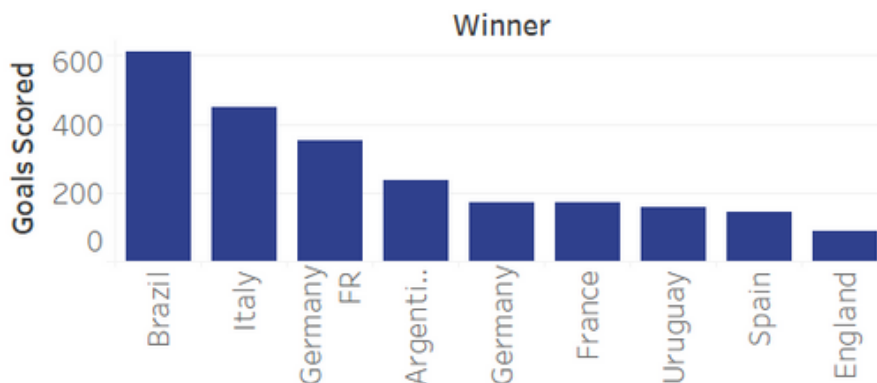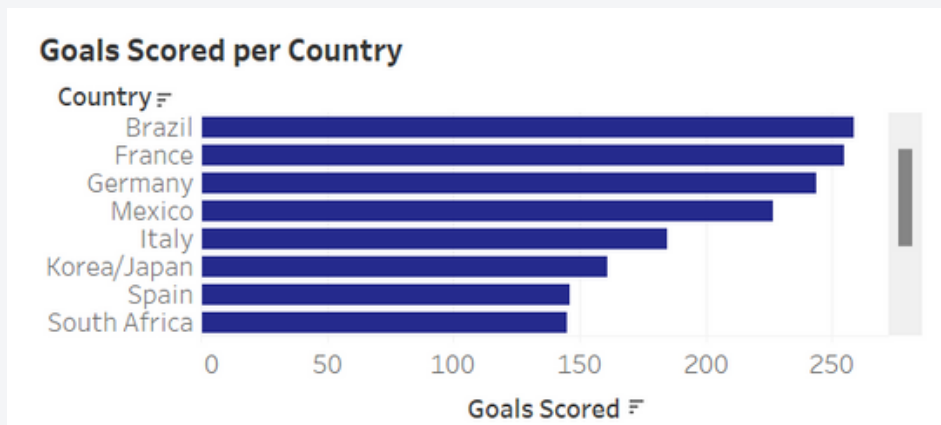
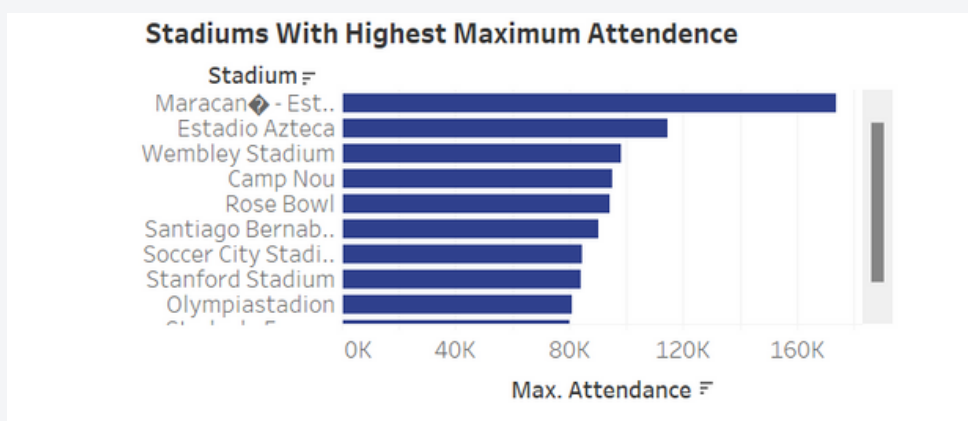### Effect of Win Conditions
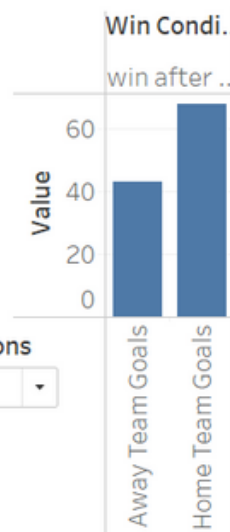
### Effect of Win Conditions
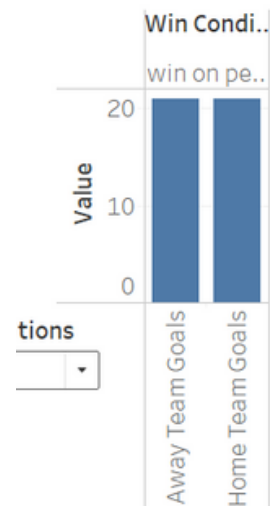
Goals Scored by Winning Teams

The above graph give the order of the goals scored by the winning teams over the years. It is found that this is directly proportional to the order of the top winners in the tournament. So, the idea remains simple the more you score, the more you win.

Goals Scored per Country

Country ⊏

Brazil	
France	
Germany	
Mexico	
Italy	
Korea/Japan	
Spain	
South Africa	

0   50   100   150   200   250

Goals Scored ⊏

In the total amount of goals scored by the countries in the tournament over the years, Brazil leads with over 250 goals, followed by France, Germany, Mexico, Italy, Korea/Japan, Spain and South Africa. One conclusion that can be drawn out of these is the teams scoring goals in their matches consistently over the years of the World cup have won more number of titles than the other teams. So, a key takeaway will be increasing the goals scored by the team per match. This will indirectly lead to winning of more titles.

Stadiums With Highest Maximum Attendence

Stadium ⊏

Maracan� - Est..	
Estadio Azteca	
Wembley Stadium	
Camp Nou	
Rose Bowl	
Santiago Bernab..	
Soccer City Stadi..	
Stanford Stadium	
Olympiastadion	

0K   40K   80K   120K   160K

Max. Attendance ⊏

The above figure shows the stadiums with the highest attendance. The highest being Maracan Stadium of Rio De Janeiro for the Match Brazil v/s Sweden in the year 1950. The probable reason can be the great sporting culture and Sport support for the home team Brazil which is highest winning team in the history of FIFA World Cup. It was Group 6 match which ended with a one-sided score of 7-1 in the favor of Brazil. Matches involving the teams who have won the highest cups often show up with a greater support and attendance. Countries which have a greater footballing culture will have greater crowd attendance than at the countries having a lesser sport culture. Stadium size, crowd accommodation capacity and accessibility and management also welcome a larger audience.

Attendence over the Years
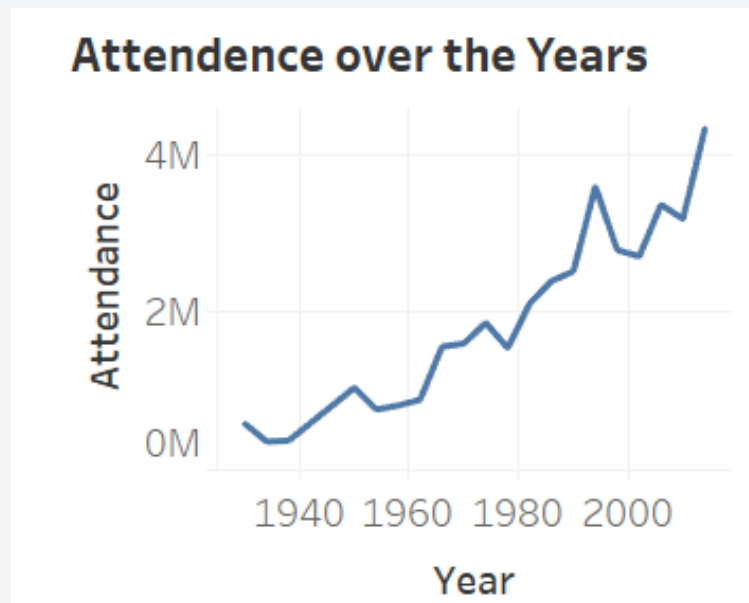
The attendance in the Stadiums has significantly. Overall, the rise in attendance at FIFA World Cup stadiums can be attributed to a combination of factors, including increased globalization, the growing popularity of football, stadium expansions, effective marketing, cultural significance, and fan engagement initiatives. As the tournament continues to captivate audiences worldwide, attendance figures are likely to remain high in the years to come.

Host Country
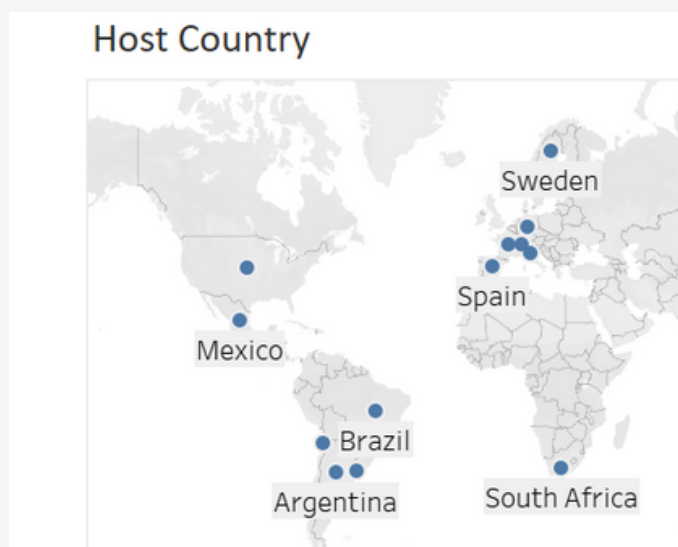
The adjacent figure shows the location of the host cities of the FIFA World Cup over the years on the World Map. It shows that the sport is more common in the eastern parts of Asia. and Southern parts of America.

RESULTS AND INSIGHTS

The following are the results and insights from our data analysis on the FIFA World Cup Data Set:

- Teams having won more number of cups have a higher crowd support.
- Earlier Home-team goals before the half time were greater than the half time goals of the away team but today, statistics that due to increased metal stability of the players and better training methods, away teams perform better.
- Home teams have a greater advantage when the win conditions are win after extra time but not in the case of win on penalties.
- Higher the number of goals scored by the country in the tournament, higher is the probability of the team to win the title.
- The attendance in stadiums for FIFA world cup matches has increased over the years due to increased popularity of the sport and enhanced connectivity.
- Teams prefer N line up over S line up in the world cup due to the above mentioned reasons.
- Popularity of football is more in the Eastern Asia and in the Southern America (judged by the hosting countries).

CONCLUSIONS

In conclusion, our data analysis project has offered a comprehensive examination of the FIFA World Cup, revealing a treasure trove of insights into the tournament's dynamics, trends, and historical context. Through meticulous scrutiny of statistical data spanning multiple editions, we've uncovered patterns, trends, and correlations that provide a deeper understanding of the beautiful game's most prestigious event.

Our findings have showcased the evolution of football tactics, from formations and playing styles to the strategic decisions that underpin success on the global stage. We've identified key factors influencing match outcomes, including the impact of individual player performances, team cohesion, and tactical adaptability.

Moreover, our analysis has highlighted the significance of historical context in shaping the narrative of the World Cup. From iconic moments to unforeseen upsets, each tournament has been shaped by a unique blend of socio-cultural, political, and sporting influences that have left an indelible mark on football history.

As we conclude this report, it's clear that the FIFA World Cup transcends mere statistics, embodying the passion, drama, and spirit of competition that captivates audiences worldwide. It serves as a unifying force, bringing together nations and cultures in celebration of the universal language of football.

We trust that our analysis has provided valuable insights for football enthusiasts, scholars, and stakeholders alike, offering a deeper appreciation of the World Cup's enduring legacy and inspiring further exploration into the intricacies of the world's most beloved sport.

VOCABULARY

S-LINE UP

In football, the term "S line up" isn't a common phrase used to describe a specific formation or lineup. It's possible that you might be referring to the defensive line, which is typically composed of defenders positioned in a line across the field to protect the goal.

A standard defensive lineup might include players such as:

1. Center-backs (CB): Positioned centrally in defense, responsible for marking opposition attackers and intercepting passes.
2. Full-backs (FB): Positioned on the left and right flanks of the defense, responsible for marking wingers and providing defensive cover.
3. Sweeper (SW): A less common position, positioned behind the center-backs to provide additional defensive support and cover.

N-LINE UP

In football, the "N lineup" isn't a standard term used to describe a specific formation or lineup. It's possible that you might be referring to the midfield lineup or the central area of the pitch where midfielders typically operate.

In a traditional football formation, the midfield consists of players positioned centrally on the field, between the defensive line and the forward line. Common midfield positions include:

1. Defensive Midfielder (DM): Positioned centrally in front of the defensive line, responsible for breaking up opposition attacks, intercepting passes, and providing defensive cover for the backline.
2. Central Midfielder (CM): Positioned in the center of the midfield, responsible for controlling the tempo of the game, distributing passes, and linking up play between defense and attack.
3. Attacking Midfielder (AM): Positioned in an advanced central role, responsible for creating scoring opportunities, providing assists, and supporting the forwards in attack.

The "N lineup" may also refer to a specific formation that emphasizes numerical superiority or control in the midfield area, such as a 4-3-3 or a 4-2-3-1 formation, where the midfield comprises three or four players respectively.