FIFA WORLD CUP-ANALYSIS

DETAILED PROJECT REPORT

MOHAMMAD ZAHID GUROO

UMIP15259

PROJECT DETAIL

PROJECT TITLE	FIFA WORLD CUP ANALYSIS
TECHNOLOGIES	DATA ANALYSIS
DOMAIN	SPORTS
PROJECT DIFFICULTIES LEVEL	ADVANCED
TOOLS USED	VS CODE(WITH JUPYTER INTEGRATION)

OBJECTIVE

- The Goal of this project is to analyse the FIFA World Cup competition ,based on a combination of features/stats that have taken place in the FIFA World Cup.

INTRODUCTION

- Full form of FIFA is Fédération Internationale de Football Association. The championship has been awarded every four years.

- There are held within the six FIFA continental zones (Africa, Asia, North and Central America and Caribbean, South America, Oceania, and Europe).

- The qualification process can start almost three years before the final tournament and last over a two-year period.

- The final tournament has been used since 1998 & features 32 national teams competing over the course of a month in the host nation(s). There are two stages: the group stage followed by the knockout stage

VARIABLES UNDERSTANDING FROM DATA SET

▸ Year - The year in which the match was played.

▸ Matches Played - Total matches played in the Tournament.

▸ Date - The Date on which the match was played.

▸ Attendance - Total attendance in the Tournament.

▸ Stage - The stage at which the match was played.

▸ Qualified Teams - Total participating teams.

▸ Stadium - Stadium name where the match was held.

▸ Goals Scored - Total goals scored in the world cup tournament.

▸ City - The city name, where the match was played.

▸ Fourth - Team who won the fourth place.

▸ Home team - Home team's country name.

▸ Third - Team who won the third place.

▸ Home Goals - Total goals scored by the home team.

▸ Runner-Up - Team who won the second place.

▸ Away Goals - Total goals scored by the away team.

▸ Winner - Team who won the FIFA world cup.

▸ Away Team - Away team's country name.

DATA VISUALIZATION (STEP-1)

- **Importing & Collecting The Data:**

- The first and most important step of data visualization is gathering data in large amounts. Only after we have substantial data, we can apply data visualization techniques on the collected data and get some helpful insights from it.

```python
#importing the necessary libraries
import numpy as np
import pandas as pd
import plotly.express as px



#importing the data
df1 = pd.read_csv("E:\Projects\FIFA WC Analysis\WorldCupMatches.csv")
df2 = pd.read_csv("E:\Projects\FIFA WC Analysis\WorldCupPlayers.csv")
df3 = pd.read_csv("E:\Projects\FIFA WC Analysis\WorldCups.csv")
```

DATA VISUALIZATION (STEP-2)

- **Cleaning Data**

- Data cleaning is an essential step to perform before creating a visualization. A bunch of data out of a large dataset that has inappropriate, empty, or false values may lead to adding erroneous visuals with anomalies in it.

- The output received from a data cleaning process is usually a dataset that is free of errors and anomalies etc. which gives much more accuracy when data is processed. Data cleaning is pretty much dependent on the dataset domain that you're working with.

```
#Checking for null values
df1.isnull().sum()

Year                     3720
Datetime                 3720
Stage                    3720
Stadium                  3720
City                     3720
Home Team Name           3720
Home Team Goals          3720
Away Team Goals          3720
Away Team Name           3720
Win conditions           3720
Attendance               3722
Half-time Home Goals     3720
Half-time Away Goals     3720
Referee                  3720
Assistant 1              3720
Assistant 2              3720
RoundID                  3720
MatchID                  3720
Home Team Initials       3720
Away Team Initials       3720
dtype: int64

#Dropping the null values
df1_cleaned = df1.dropna()
```
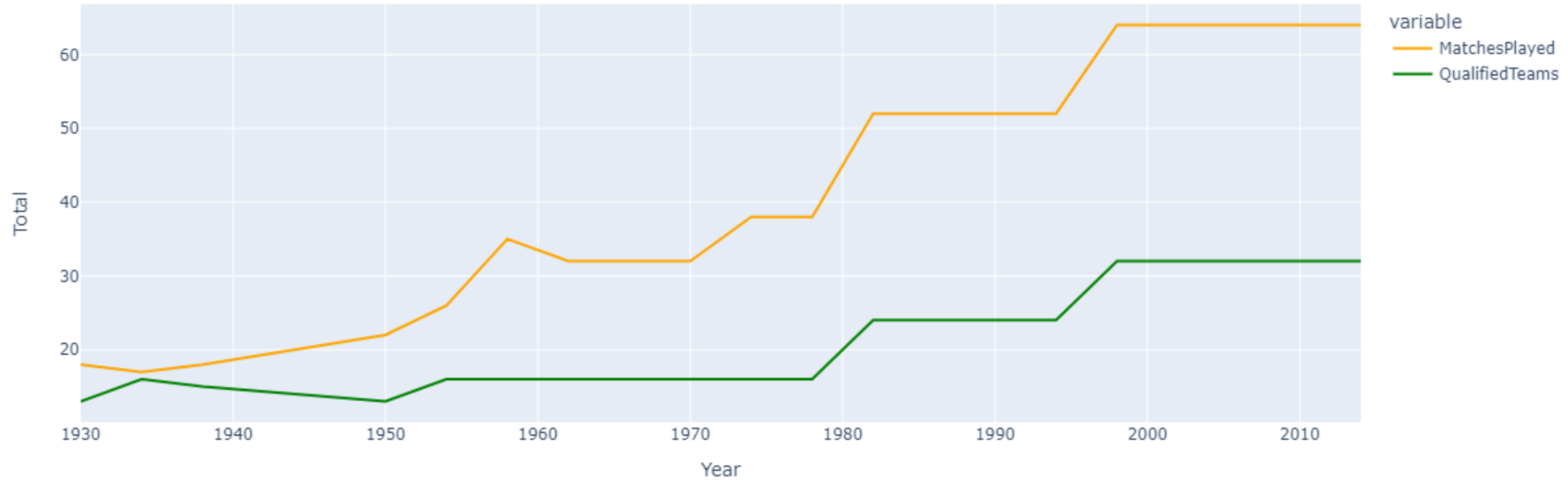
DATA VISUALIZATION (STEP-3)

- **Choose A Chart Type**

- Before choosing a visual chart or graph, it is important to understand your audience and then choose a chart or graph accordingly which will best communicate the message.

- Choosing a chart totally depends on what findings you need to convey to your audience.

DATA VISUALIZATION (STEP-3)

▶ **Choose A Chart Type**

▶ Before choosing a visual chart or graph, it is important to understand your audience and then choose a chart or graph accordingly which will best communicate the message.

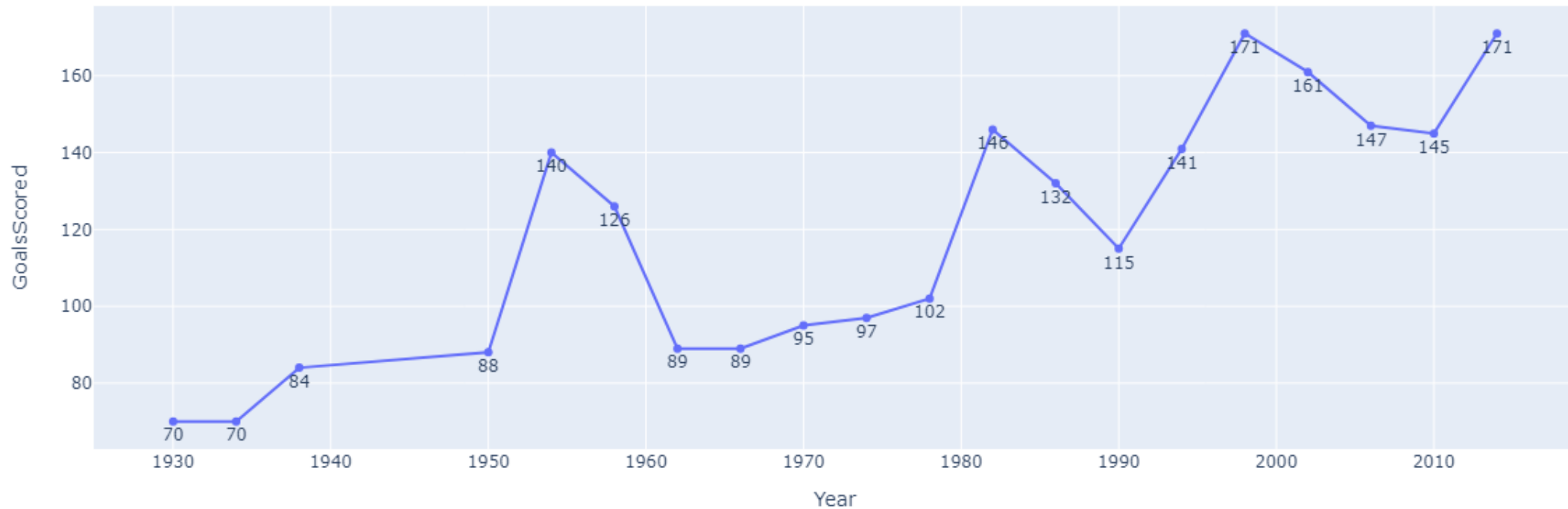▶ Choosing a chart totally depends on what findings you need to convey to your audience.

Total Qualified Teams and Matches Played Over The Years

Analysis

❖ As evident from the line plot the FIFA World Cup has had more and more teams qualifying over the years as a result showing upward trend & due to increase in the no. of teams the total matches have also increased . So we can say that the World Cup has become much more competitive over the years.
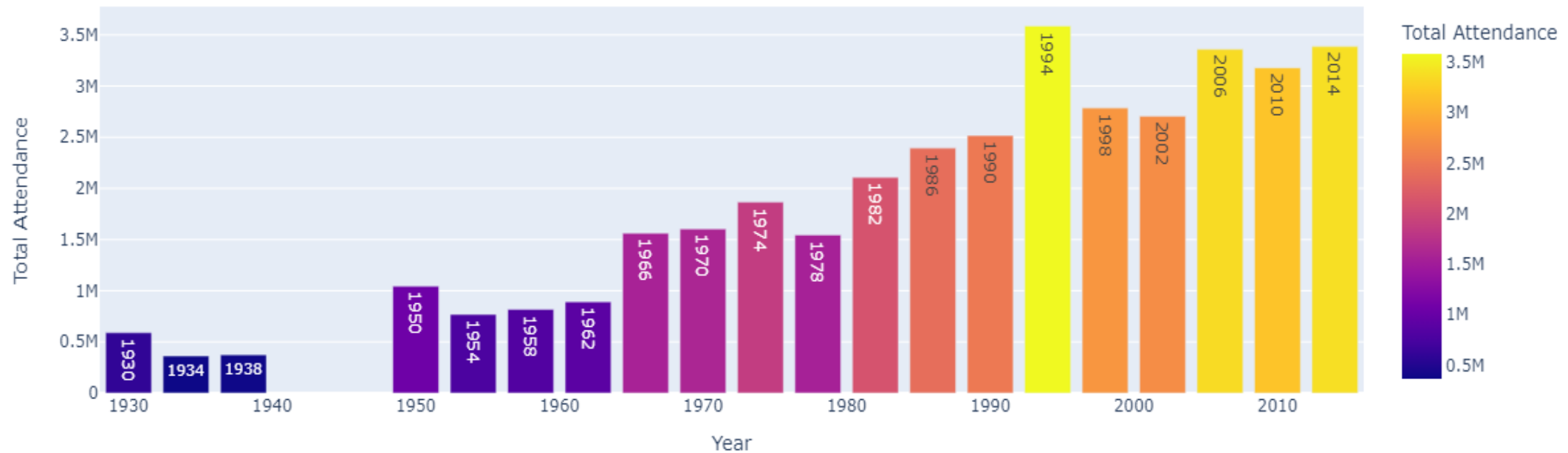
Goals Scored Over The Years

❖ **Analysis:**
From the following plot we can observe that the amount of goals scored in the FIFA World Cup over the years has been increasing . This is mainly due to increase in the number of games played . So we can say more the teams qualified , more will be the games played and more will be the goals scored.
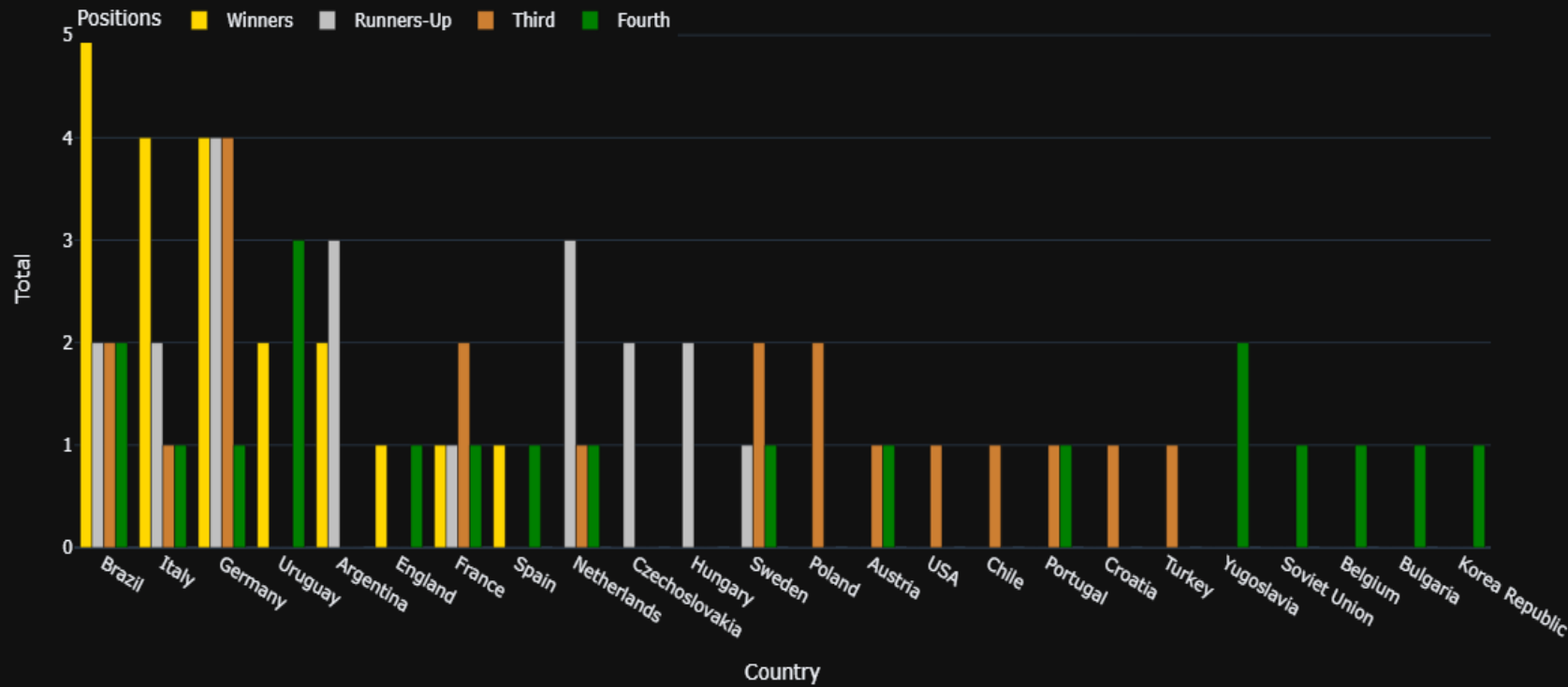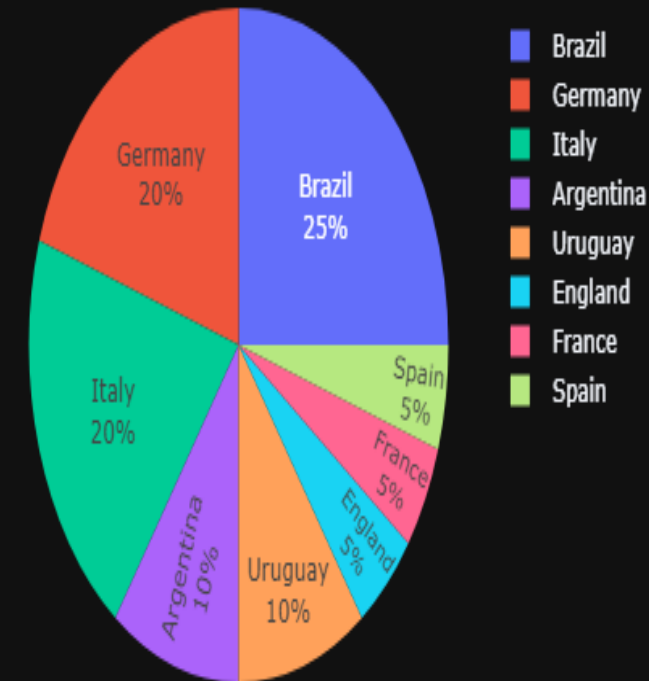
Total Attandance Over The World Cups

Analysis

❖

From the above graph its quite obvious that 1994 edition of the FIFA World Cup was the most popular and the most attended World Cup followed by the 2014 and 2006 respectively .It can be also seen from the graph that there has been a tremendous rise in the popularity of football(mainly the FIFA World Cup) over the years as number of attendees has massively increased.

Countries With Most Podium Finishes

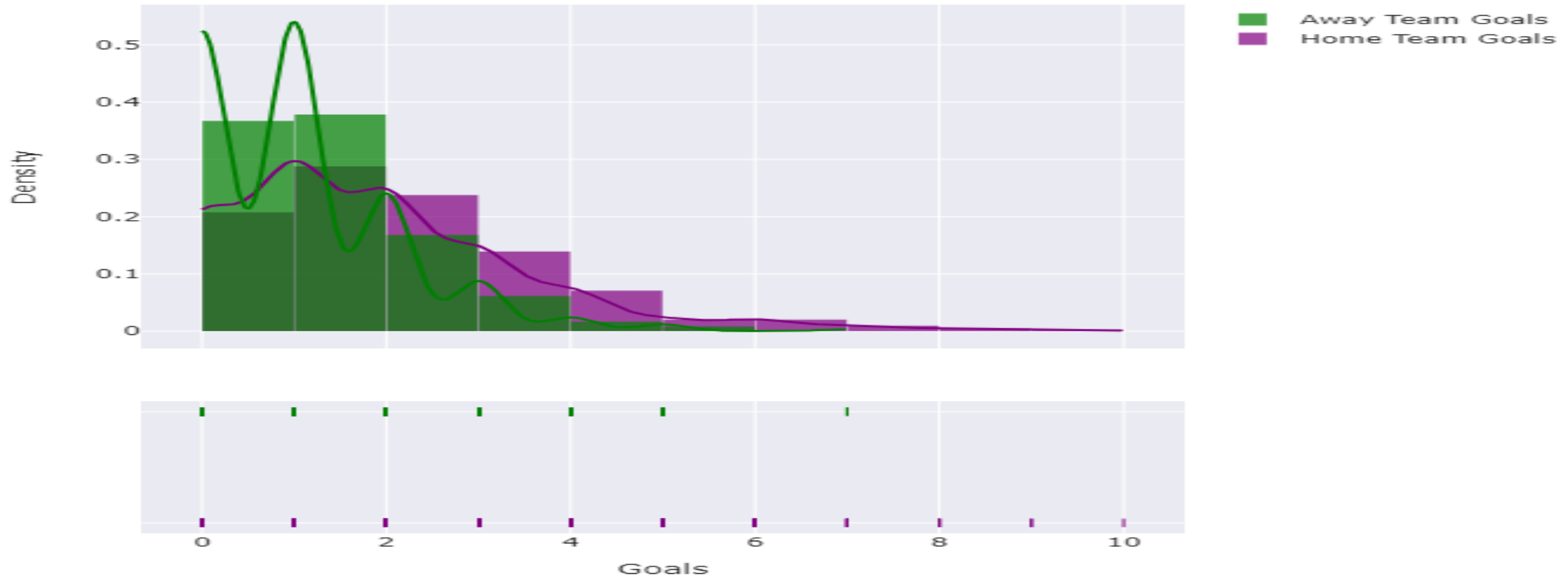Countries with Most World Cups

Analysis

From the above graphs we observe that Brazil is the most coveted nation as they have won the World Cup 5 times , however in terms of the most podium finishes Germany stands above the rest.
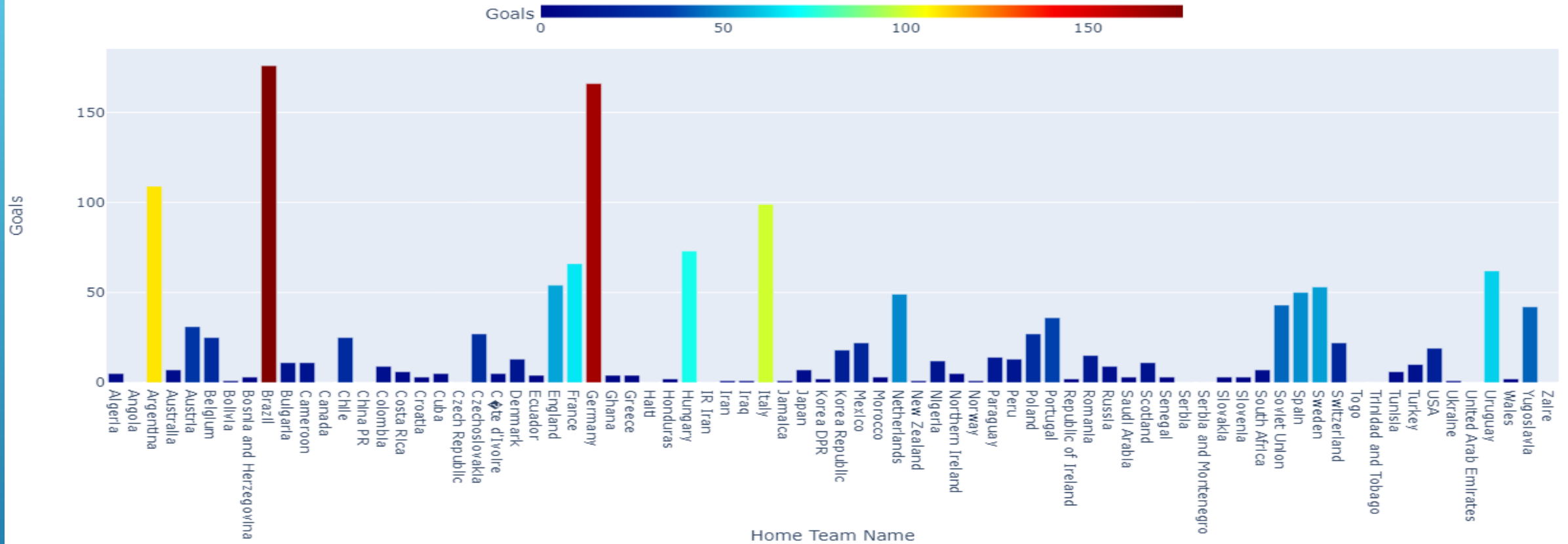
Home Team Goals Vs Away Team Goals

Analysis

The following distplot shows the distributions of home goals versus away goals where we can see that the both the graphs or skewed to the right meaning they have been less matches with high number of goals and more matches with less number of goals. The peak of the distribution of home team goals is around 2 suggesting this is the common number of goal scored by the home team, where is the peak of the distribution for the away team is around 1 suggesting this is the common number of goal scored by the team.
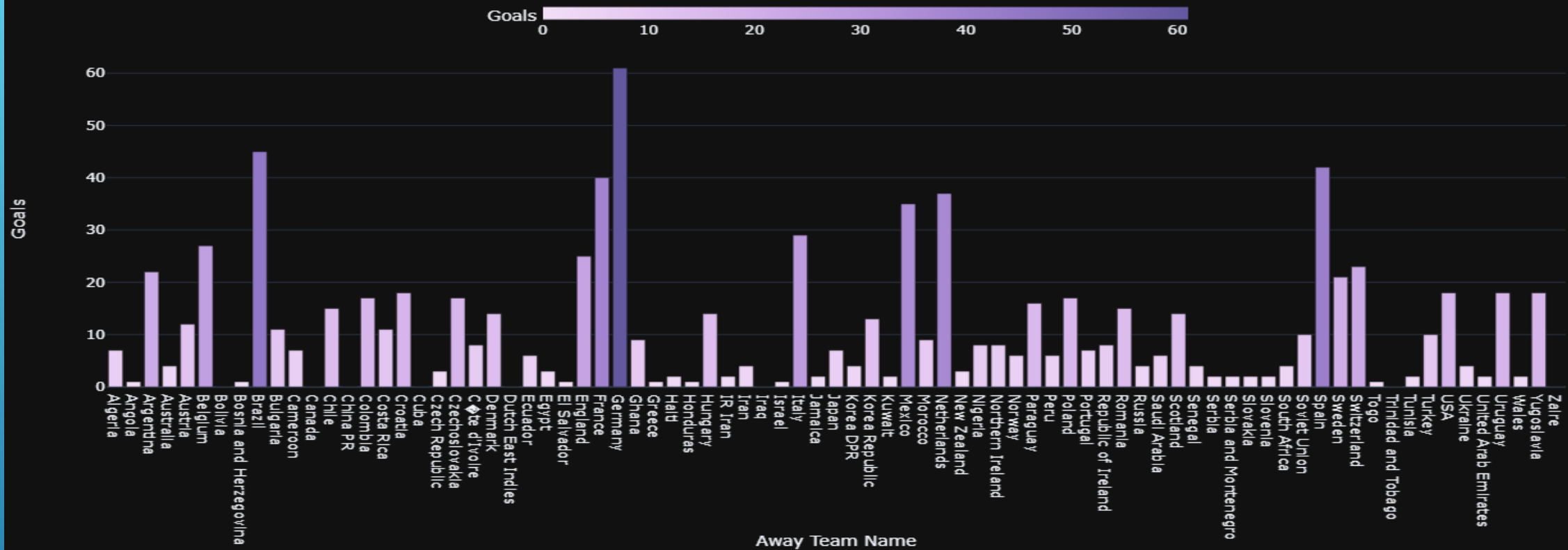
Total Goals By Home Teams

Analysis

❖ Above bar graph shows the total goals scored by teams while being the "Home Team" where it's quite evident that the teams Brazil & Germany have been much better than the rest as they are only teams which have scored more than 150 Goals making them an instant favourite to win while playing as "Home Team".
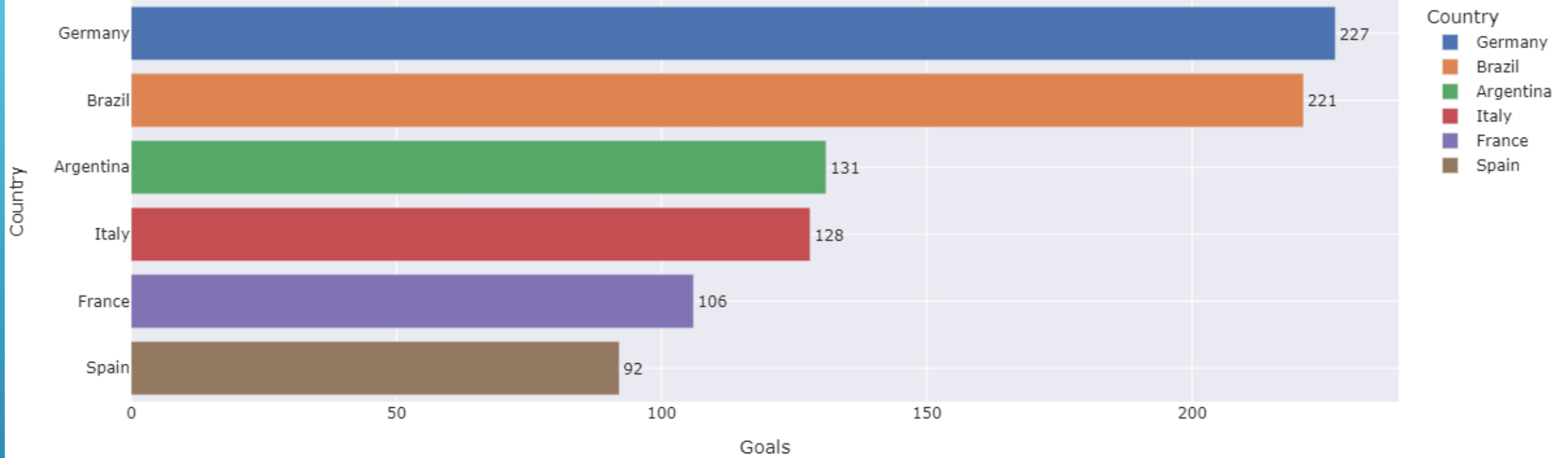
Analysis

From the graph we can see Brazil & Germany have scored most goals as an "Away Team" thus making them an automatic favourite to win the FIFA World Cup . Surprisingly we can also see that teams like Spain , Mexico , France & Netherlands score more goals being an "Away Team" whereas teams like Argentina have been less impressive as an "Away Team".
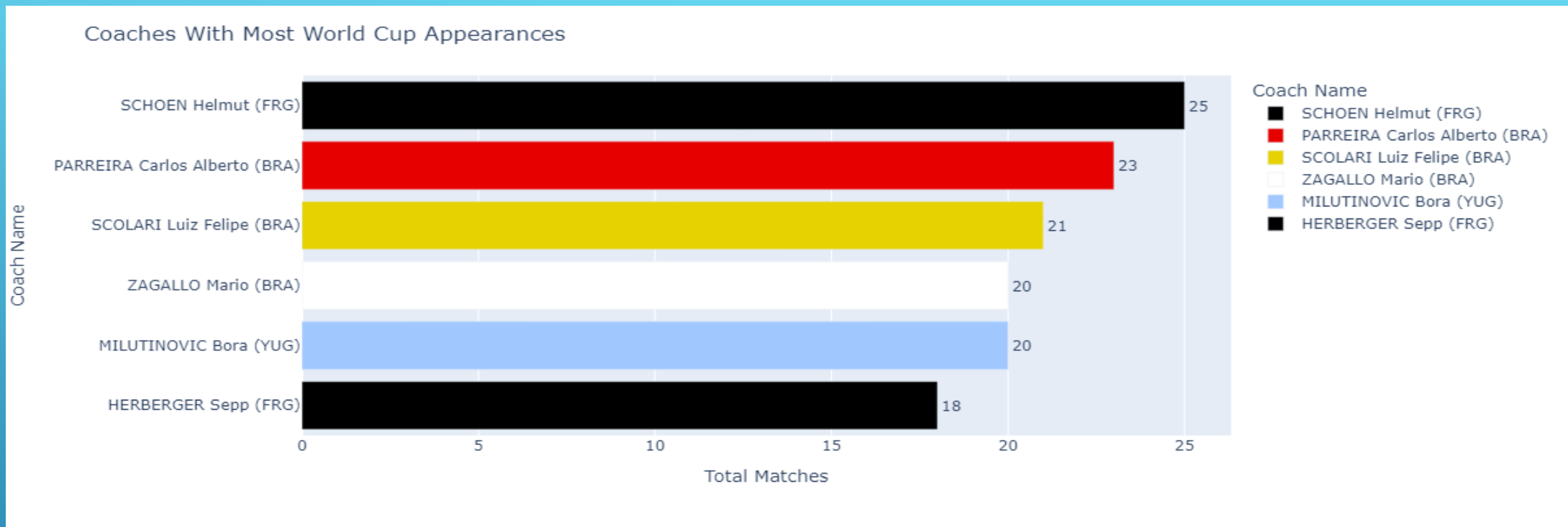
Top 6 Teams with Most Goals Scored
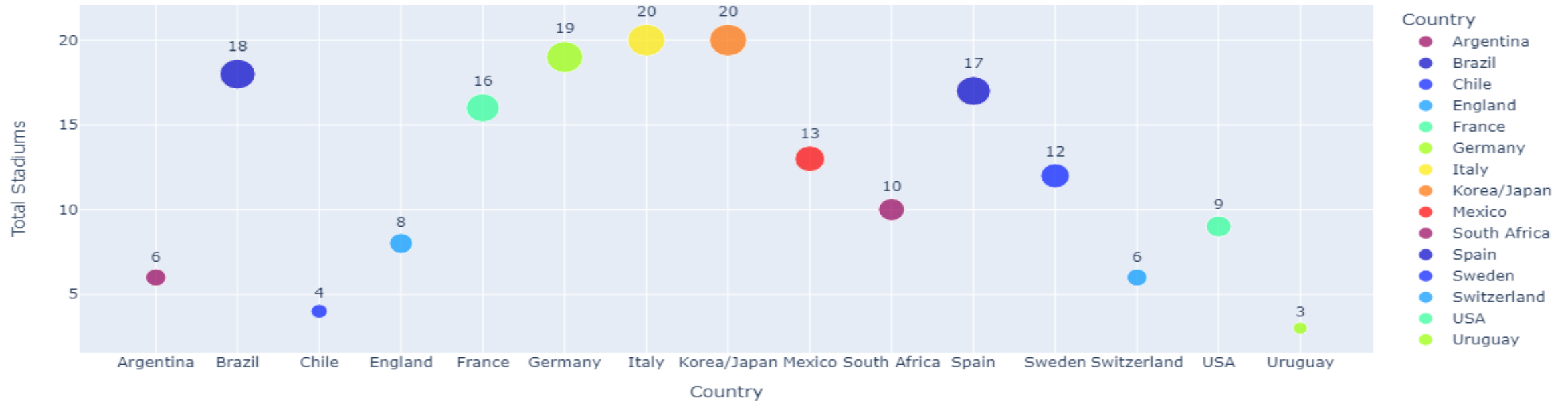
Analysis

❖

From the above plot we observe the Top 6 nations which have scored the most goals at the world cup . According to the plot Germany and Brazil have scored more goals than rest of teams indicating why they won the FIFA world cup the most amount of times and it also signifies the quality of players both the teams have had over the years for example Pele , Ronaldo , Muller & Klose.

Coaches With Most World Cup Appearances

Analysis

❖

The above plot shows the coaches who have coached the most games at the FIFA World Cup where Helmut & Carlos Alberto are leading the list indicating they have been the most successful and result bearing coaches at the World Cup.
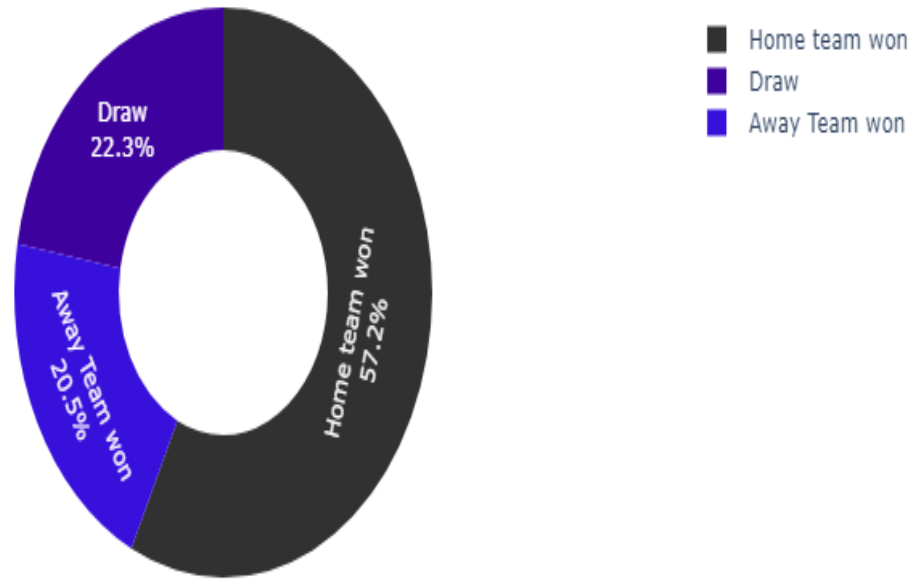
Total Stadiums Used per Country in the World Cup

Analysis

❖

The above scatter plot shows the number of different stadiums used per Country at the FIFA World Cup , where we can see that Italy and Korea/Japan world cup has the most stadiums used followed by Germany & Brazil thus giving the fans a varied and vibrant atmosphere.

Results over the World Cup

Home team won
Draw
Away Team won

Draw
22.3%

Away Team won
20.5%

Home team won
57.2%

```
#indicating the matches that have been settled at the extra time or penalties
results_after_draw=df1_without_duplicates["Win conditions"].nunique()
results_after_draw
```
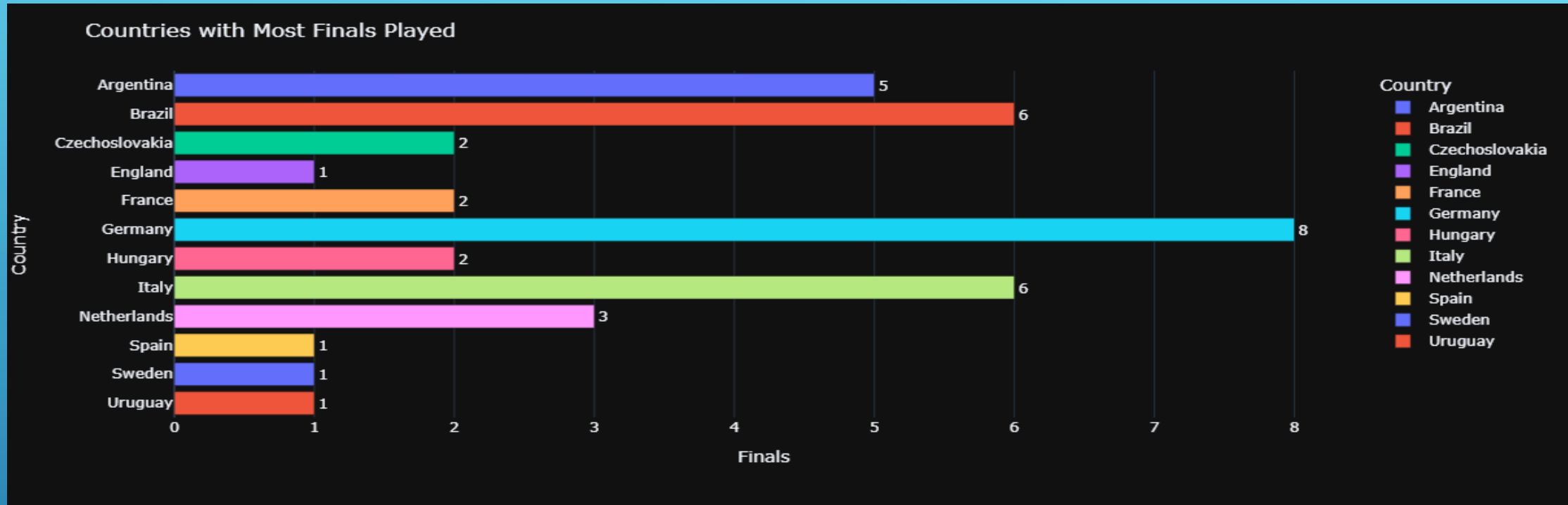
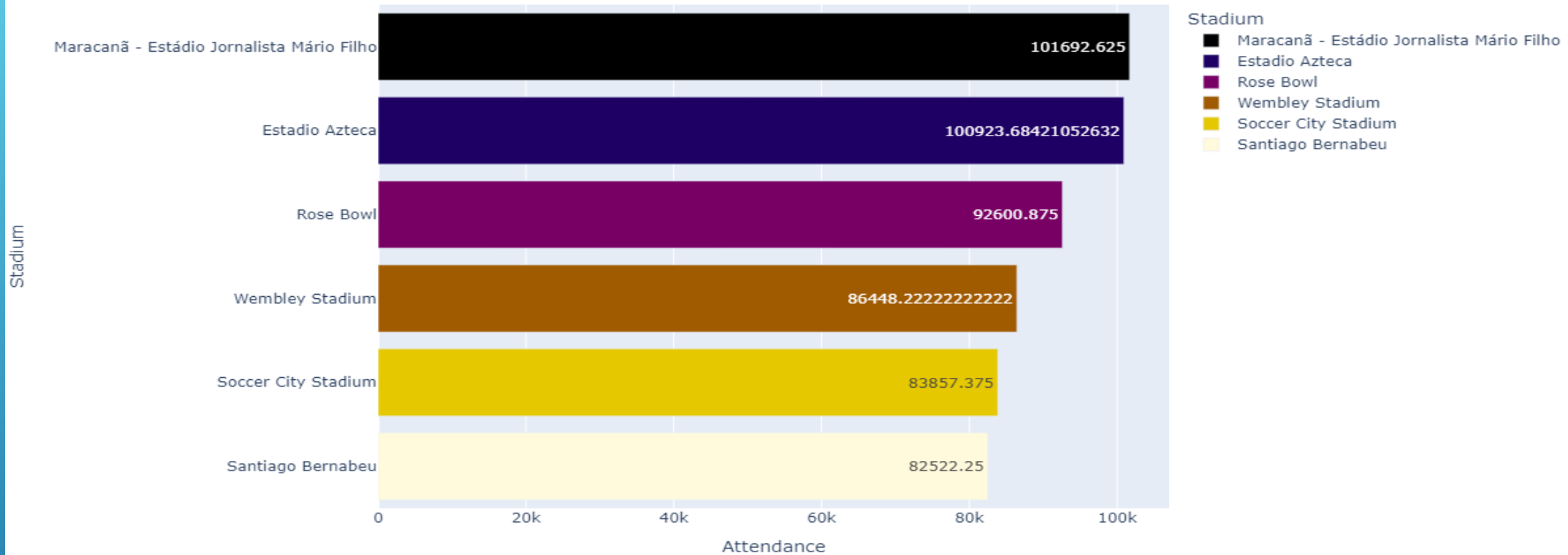✓ 0.0s

43

Analysis

The above Donut chart shows the percentage of all the results at the FIFA World Cup indicating teams representing the "Home Team" have won the most matches that being around 57% whereas the teams representing "Away teams" have only around 20% of the matches . Also from the 20% of draws(186 matches) 43 matches have had a result either at the Extra Time or at the Penalties.

Countries with Most Finals Played

Analysis

❖

The graph shows as that the countries who have played the most finals at the FIFA World Cup, from the above graph it's quite clear that Germany Brazil Italy and Argentina have played the most finals out of which obviously Brazil has won the most finals followed by Germany and Italy making these four teams a clear favourites during the start of a new World Cup Campaign

Stadiums with Highest Average Attendance

Analysis

The above graph shows the stadiums with the highest average attendance at the FIFA World Cup it is clear from the graph that the stadium Maracanā in Brazil has had the highest average attendance followed by Estadio Azteca and the Rose Bowl . The Maracanā is infamous for hosting the 2014 FIFA World Cup Final in which Germany were crowned as the Champions.

KEY FINDINGS FROM THE DATA SET

- ✓ The world cup has been played every 4 years except in the year 1942 and 1946; the period of the World War 2.

- ✓ The number of matches played has increased with each passing year as new countries participated in the World cup.

- ✓ 1994 World cup witnessed the highest number of attendees i.e. 35,87,538 while 1934 witnesses 3,63,000.

- ✓ Brazil holds the record of winning the maximum number of times, Germany is close behind. However, Germany has played most of the quarter finals, semi finals and finals. Furthermore, Germany has scored the highest number of goals in the championship.

- ✓ A major insight from the data was, on an average the home team scored more goals than the away goals, this behaviour has been shown in different rounds played in different years,

- ✓ The basic assumption of people that if matches played is more, the goals scored would also be high was changed in the year 1954. This year 140 goals were scored in just 26 matches.

- ✓ Looking at the various stadiums across the country, Italy has the maximum stadiums. Also the European countries have more stadiums that have hosted the World Cup compared to their Latin counterpart.