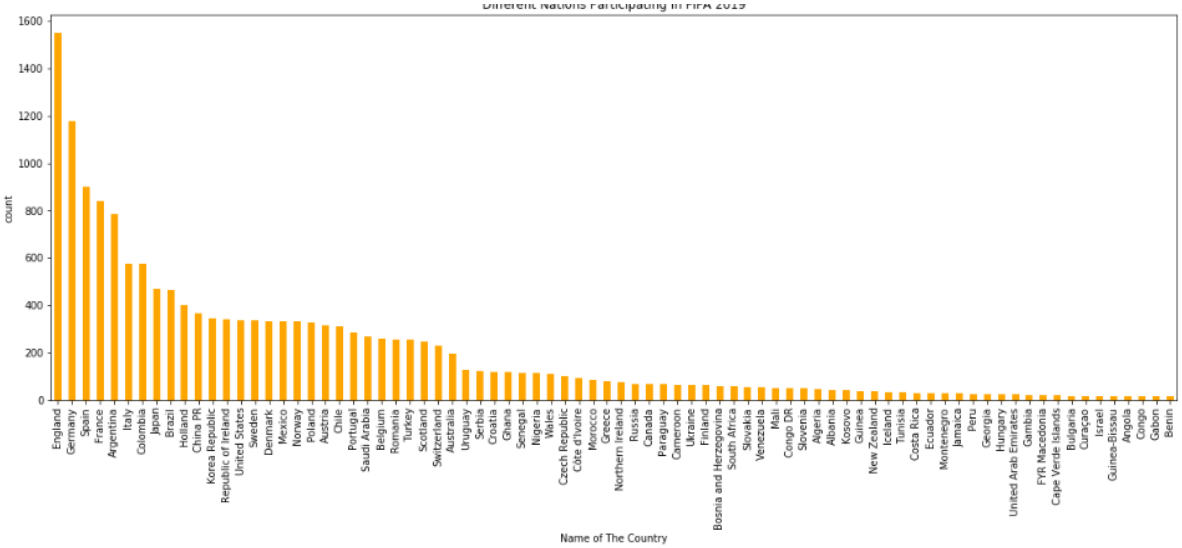
UVA_2019_SemesterProject_Learning FIFA players' traits - What makes an elite soccer player?

In 2019, the FIFA roster includes a great number of players from a total of 158 countries. The top three participating countries based on the numbers of players are England, Germany and Spain. Using an interactive world map, it is possible to hover the cursor over any participating country of interest to ascertain its average "Overall Rating".

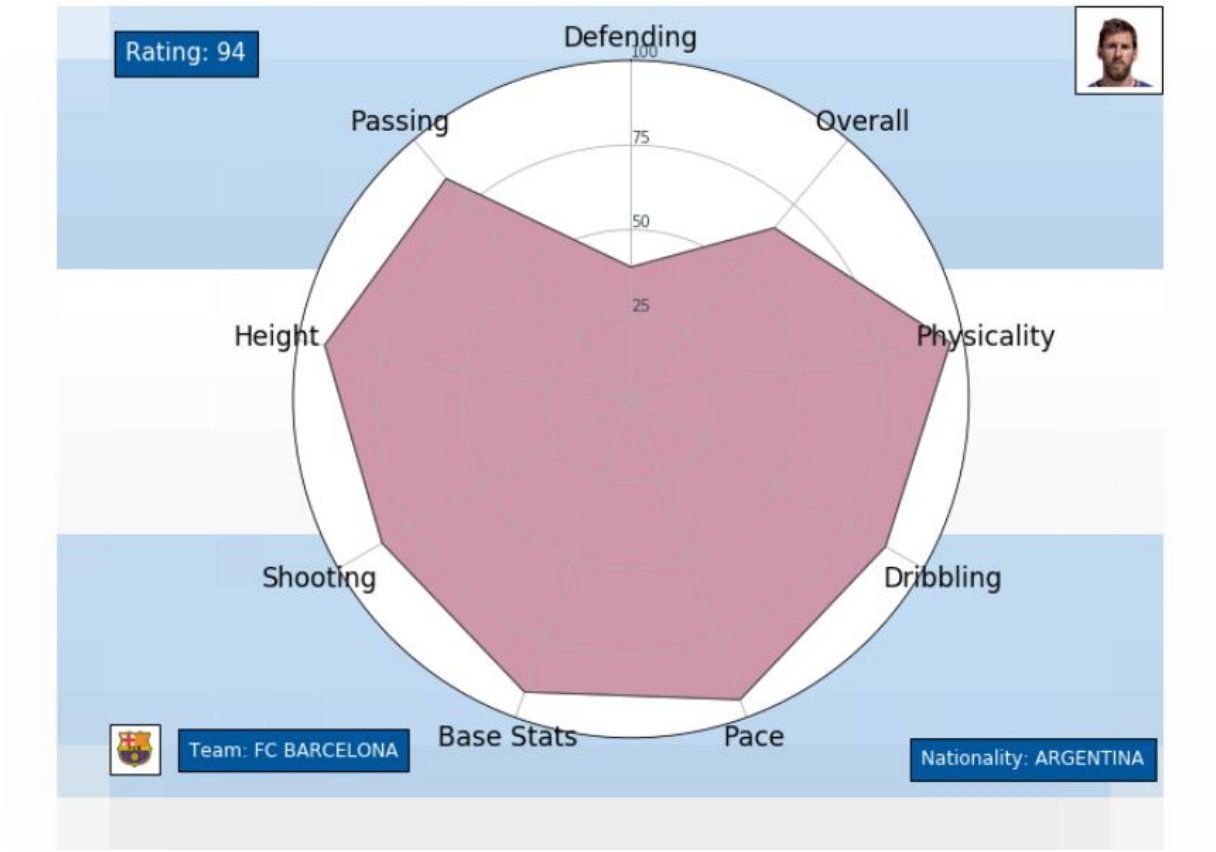
A variety of player attributes are used to understand the individual skills needed to be an elite level player. Since the game of soccer involves many different field positions, each of which favor players with certain innate attributes/features as well as mastery of certain skills, the combination of position, features and skills were identified and analyzed. **First** is the correlations between players attributes and skills, **second** is the features needed to be an elite player in each position group or position. **Third** is each country's strategic field formation used by the national team.

This study focuses on FIFA players, their nationalities and the traits they possess. The data was statistically analyzed by making correlations between players performance and their physical features and skills. The data was presented in both tabular and graphical formats. The summary of the results of players features is offered.

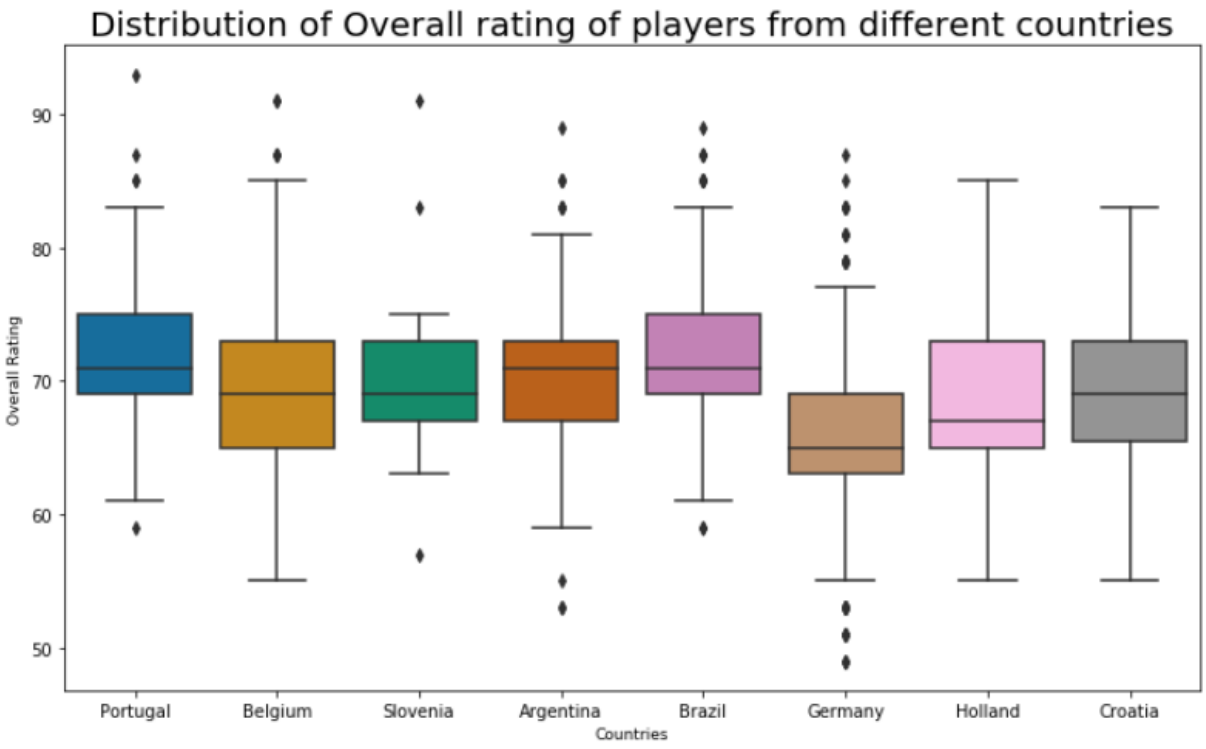


Exploratory Data Analysis:

The top FIFA soccer player based on overall performance is the Argentinian Lionel Messi. According to the data provided for the analysis, Messi's most prominent qualities are: physicality, pace, passing and shooting ability.



The overall top country ranking based on the analysis were shown to be Brazil.



What is required for a country to achieve a high status or success in international soccer? Upon analyzing the correlation between overall player ratings and attributes, the strongest three are passing, dribbling and base stats. Because the purpose of this study is to try and understand overall player ratings based on nationality, the attributes of actual field position and a broader position groups must be included in order to establish what makes the best players based on their field position.

Using tuple to get the top index

```
# defining the features of players
player_features = ('Weak Foot', 'Pace', 'Shooting',
                   'Passing', 'Dribbling', 'Defending',
                   'Physicality', 'Base Stats',
                   'In Game Stats')

# Top four features for every position in football

for i, val in data.groupby(data['Position'])[player_features].mean().iterrows():
    print('Position {}: {}, {}, {}'.format(i, *tuple(val.nlargest(4).index)))
```

Postion

* Position CAM: Base Stats, Pace, Dribbling

* Position CB: Physicality, Defending, Base Stats

* Position CDM: Base Stats, Physicality, Defending

* Position CF: Base Stats, Pace, Dribbling

* Position CM: Base Stats, Dribbling, Pace

* Position GK: Base Stats, Dribbling, Pace

* Position LB: Base Stats, Pace, Physicality

* Position LF: Base Stats, Dribbling, Pace

* Position LM: Pace, Base Stats, Dribbling

* Position LW: Pace, Base Stats, Dribbling

* Position LWB: Pace, Base Stats, Physicality

* Position RB: Base Stats, Pace, Physicality

* Position RF: Pace, Dribbling, Base Stats

* Position RM: Pace, Base Stats, Dribbling

* Position RW: Pace, Base Stats, Dribbling

* Position RWB: Pace, Base Stats, Physicality

* Position ST: Base Stats, Pace, Physicality

```
# defining the features of players
player_features = ('Weak Foot', 'Pace', 'Shooting',
                   'Passing', 'Dribbling', 'Defending',
                   'Physicality', 'Base Stats',
                   'In Game Stats')

# Top four features for every position in football

for i, val in data.groupby(data['Position'])[player_features].mean().iterrows():
    print('Position {}: {}, {}, {}'.format(i, *tuple(val.nlargest(4).index)))
```

Position Group: Attacker, Defender, Goal Keeper and Midfielder

* Position Attacker: Pace, Dribbling, Shooting, Physicality

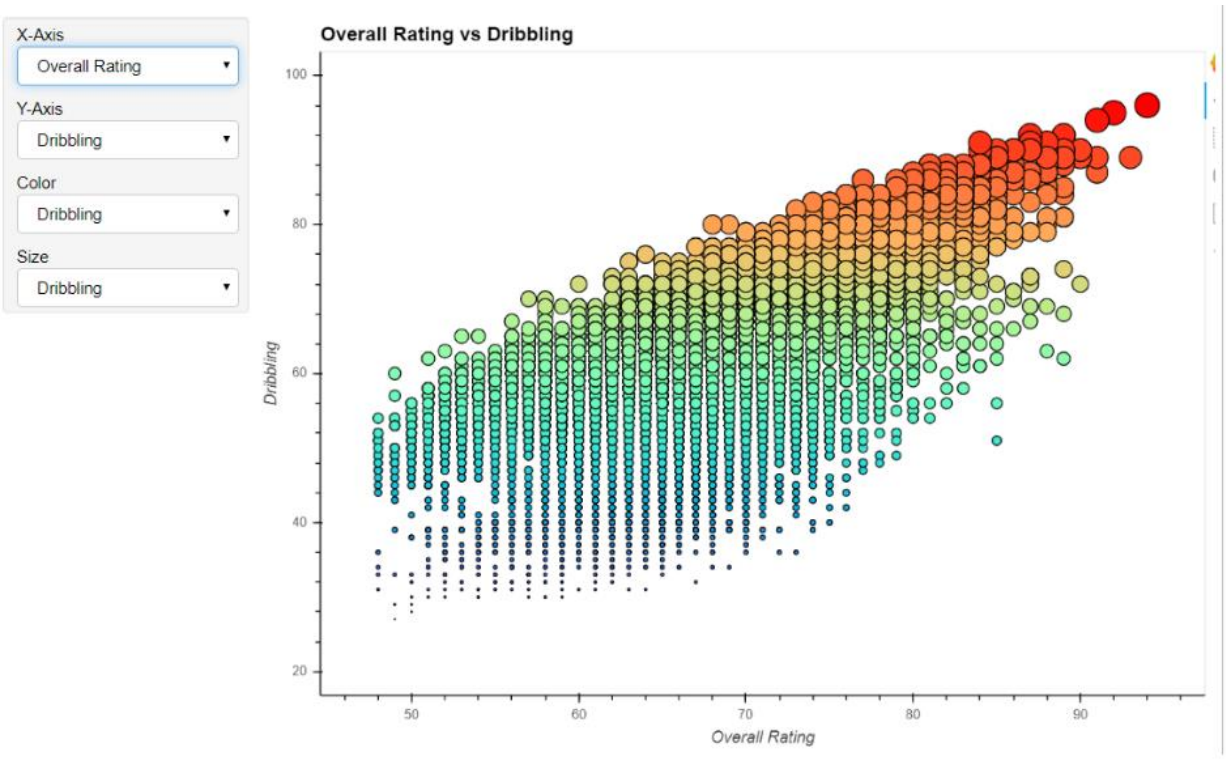
* Position Defender: Physicality, Defending, Pace, Dribbling

* Position Goal Keeper: Dribbling, Pace, Physicality, Shooting

* Position Midfielder: Pace, Dribbling, Physicality, Passing

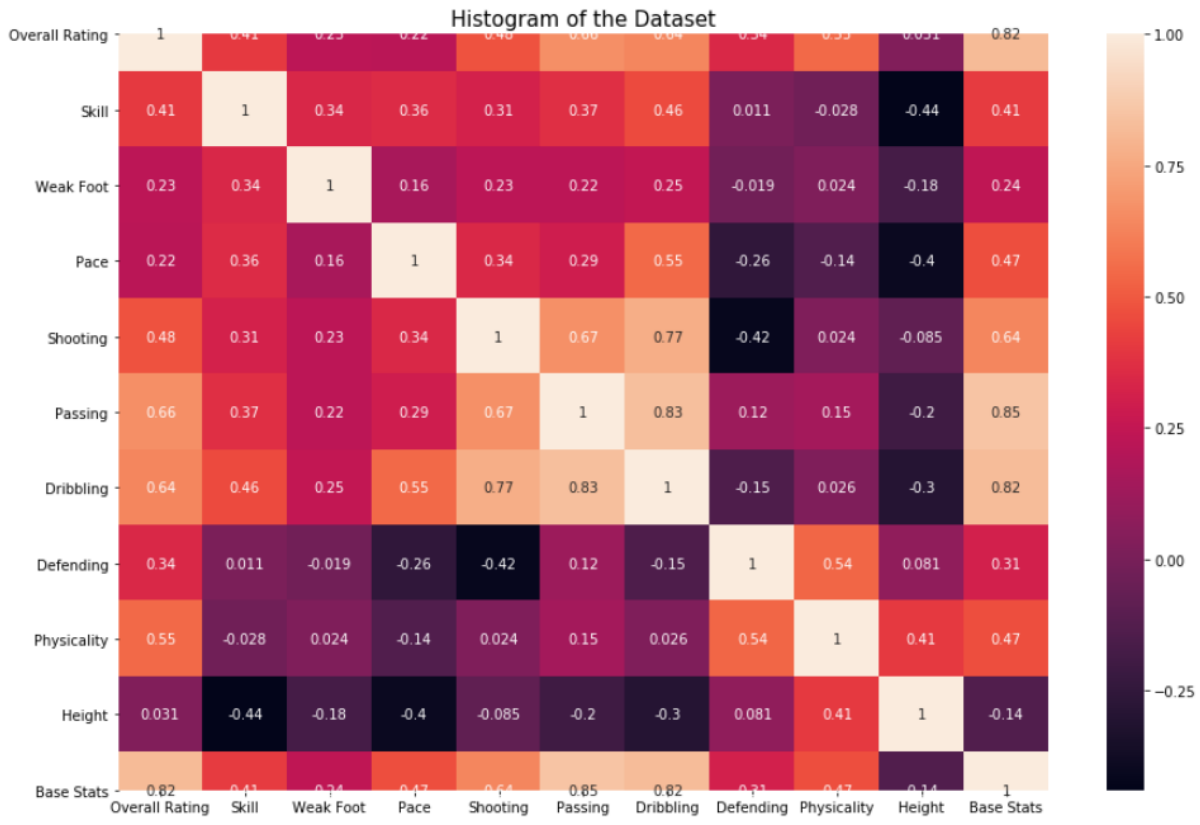
```
# defining the features of players

player_features = ('Pace', 'Shooting', 'Passing',
                   'Dribbling', 'Defending', 'Physicality',
)

# Top five features for every position in football

for i, val in df.groupby(df['Position Group'])[player_features].mean().iterrows():
    print('Position {}: {}, {}, {}, {}'.format(i, *tuple(val.nlargest(4).index)))
```

Interactive scatter plots are made to allow the selection of attributes of interest to check the correlation, the color and size are also the variables which allow users to select.

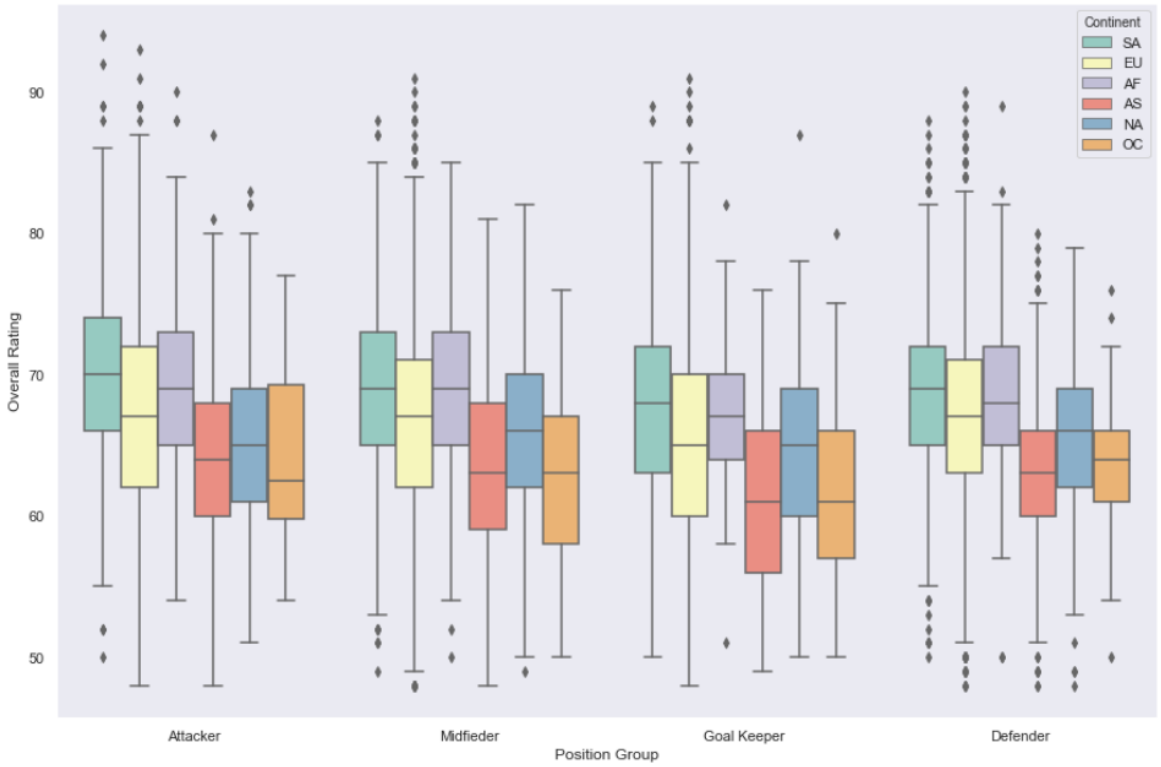


Correlation between player attributes

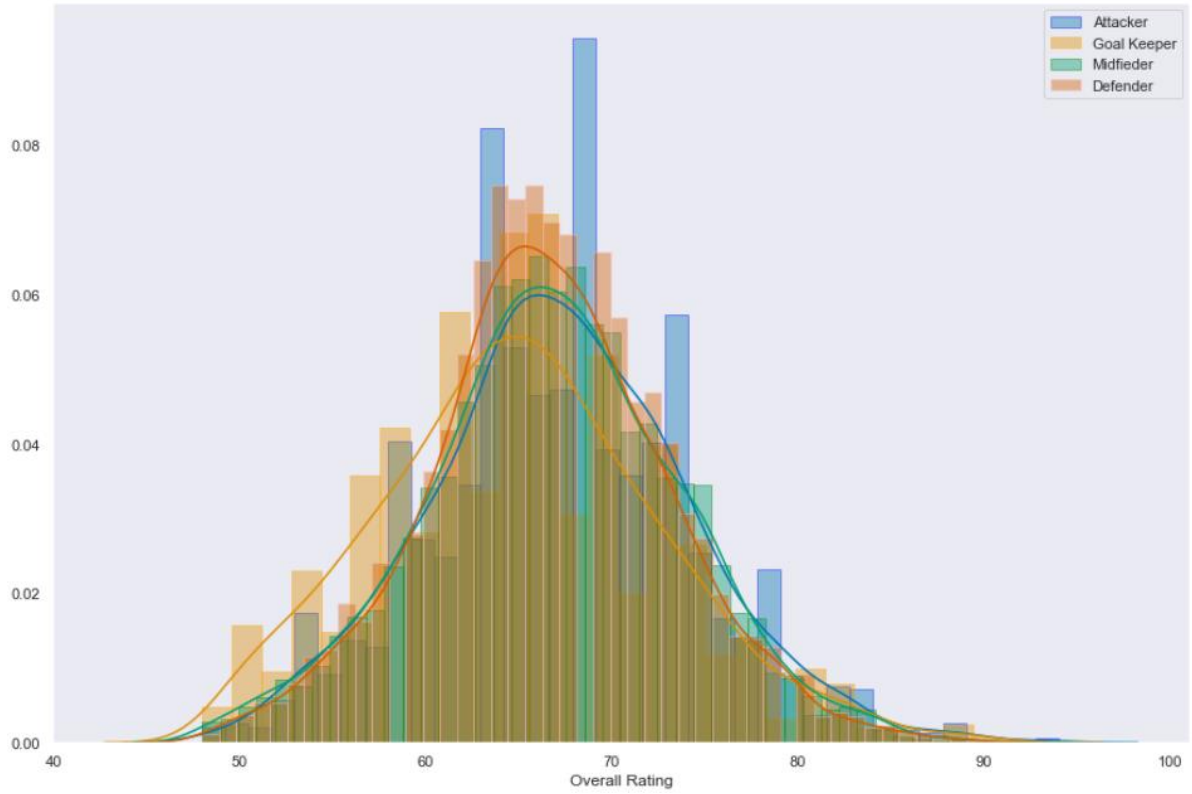
A correlation matrix was plotted and displayed as a heatmap to understand correlation between each attribute. The heatmap is color-coded scale of correlation coefficient gives a quick overview of which skills are highly correlated and which are negatively correlated.



Since there are a total of 136 countries in FIFA in 2019, a further understanding of each field position group; player strength based on the continent, the top three performing continents in the positions of Attacker. Midfielder, Goal Keeper and Defenders are South America, Europe and Africa



The histogram distribution shows that the overall ratings for the position groups. It can be seen that the Attackers and Midfielders score higher than that Defenders and Goalkeepers. The highest count is shown by the Defenders followed by the Midfielder and the Attackers. The lowest count was displayed by the Goalkeepers.



The strategies and formations used by the respective top national teams are analyzed. In addition, pertinent team formations were analyzed in order to understand the correlation between formation and success and to score these formations on a national basis.

It seems the national teams favoring more traditional formations, like 3-4-3 or 4-4-2, achieve success because of the balance between the safer defensive and the riskier offensive play-styles. An ideal defense will never lose a match. However, to win competitions teams need to win matches and that involves taking risks and scoring goals.

Squad = ['3-4-3', '4-4-2', '4-3-1-2', '4-3-3', '4-2-3-1']

```
squad_343_strict = ['GK', 'CB', 'CB', 'CB', 'RB|RWB', 'CM|CDM', 'CM|CDM', 'LB|LWB', 'RM|RW', 'ST|CF', 'LM|LW']
squad_442_strict = ['GK', 'RB|RWB', 'CB', 'CB', 'LB|LWB', 'RM', 'CM|CDM', 'CM|CAM', 'LM', 'ST|CF', 'ST|CF']
squad_4312_strict = ['GK', 'RB|RWB', 'CB', 'CB', 'LB|LWB', 'CM|CDM', 'CM|CAM|CDM', 'CM|CAM|CDM', 'CAM|CF', 'ST|CF', 'ST|CF']
squad_433_strict = ['GK', 'RB|RWB', 'CB', 'CB', 'LB|LWB', 'CM|CDM', 'CM|CAM|CDM', 'CM|CAM|CDM', 'RM|RW', 'ST|CF', 'LM|LW']
squad_4231_strict = ['GK', 'RB|RWB', 'CB', 'CB', 'LB|LWB', 'CM|CDM', 'CM|CDM', 'RM|RW', 'CAM', 'LM|LW', 'ST|CF']
squad_list = [squad_343_strict, squad_442_strict, squad_4312_strict, squad_433_strict, squad_4231_strict]
squad_name = ['3-4-3', '4-4-2', '4-3-1-2', '4-3-3', '4-2-3-1']
```

Squad Overall

Nationality (Country, Squad, Rating)

* Germany 3-4-3 85.36

* Germany 4-4-2 84.82

* Germany 4-3-1-2 85.64

* Germany 4-3-3 85.91

* Germany 4-2-3-1 85.73

```
Germany = pd.DataFrame(np.array(get_summary_n(squad_list, squad_name, ['Germany'])).reshape(-1,3), columns = ['Nationality',
Germany.set_index('Nationality', inplace = True)
Germany['Overall'] = Germany['Overall'].astype(float)
print (Germany)
```

```
             Squad  Overall
Nationality
Germany        3-4-3    85.36
Germany        4-4-2    84.82
Germany      4-3-1-2    85.64
Germany        4-3-3    85.91
Germany      4-2-3-1    85.73
```

Squad Overall

Nationality (Country, Squad, Rating)

* France 3-4-3 85.64

* France 4-4-2 85.36

* France 4-3-1-2 86.36

* France 4-3-3 85.64

* France 4-2-3-1 85.55

```
France = pd.DataFrame(np.array(get_summary_n(squad_list, squad_name, ['France'])).reshape(-1,3), columns = ['Nationality', '
France.set_index('Nationality', inplace = True)
France['Overall'] = France['Overall'].astype(float)
print (France)
```

```
             Squad  Overall
Nationality
France         3-4-3    85.64
France         4-4-2    85.36
France       4-3-1-2    86.36
France         4-3-3    85.64
France       4-2-3-1    85.55
```

Squad Overall

Nationality (Country, Squad, Rating)

* Spain 3-4-3 86.36

* Spain 4-4-2 86.18

* Spain 4-3-1-2 86.91

* Spain 4-3-3 86.82

* Spain 4-2-3-1 86.82

```
Spain = pd.DataFrame(np.array(get_summary_n(squad_list, squad_name, ['Spain'])).reshape(-1,3), columns = ['Nationality', 'Sq
Spain.set_index('Nationality', inplace = True)
Spain['Overall'] = Spain['Overall'].astype(float)
print (Spain)
```

```
              Squad  Overall
Nationality
Spain           3-4-3    86.36
Spain           4-4-2    86.18
Spain         4-3-1-2    86.91
Spain           4-3-3    86.82
Spain         4-2-3-1    86.82
```

Squad Overall

Nationality (Country, Squad, Rating)

* Brazil 3-4-3 85.82

* Brazil 4-4-2 84.64

* Brazil 4-3-1-2 85.09

* Brazil 4-3-3 85.91

* Brazil 4-2-3-1 85.73

```
Brazil = pd.DataFrame(np.array(get_summary_n(squad_list, squad_name, ['Brazil'])).reshape(-1,3), columns = ['Nationality', '
Brazil.set_index('Nationality', inplace = True)
Brazil['Overall'] = Brazil['Overall'].astype(float)
print (Brazil)
```

```
              Squad  Overall
Nationality
Brazil          3-4-3    85.82
Brazil          4-4-2    84.64
Brazil        4-3-1-2    85.09
Brazil          4-3-3    85.91
Brazil        4-2-3-1    85.73
```

Squad Overall

Nationality (Country, Squad, Rating)

* England 3-4-3 82.82

* England 4-4-2 83.00

* England 4-3-1-2 83.00

* England 4-3-3 83.27

* England 4-2-3-1 83.18

```
England = pd.DataFrame(np.array(get_summary_n(squad_list, squad_name, ['England'])).reshape(-1,3), columns = ['Nationality',
England.set_index('Nationality', inplace = True)
England['Overall'] = England['Overall'].astype(float)
print (England)
```

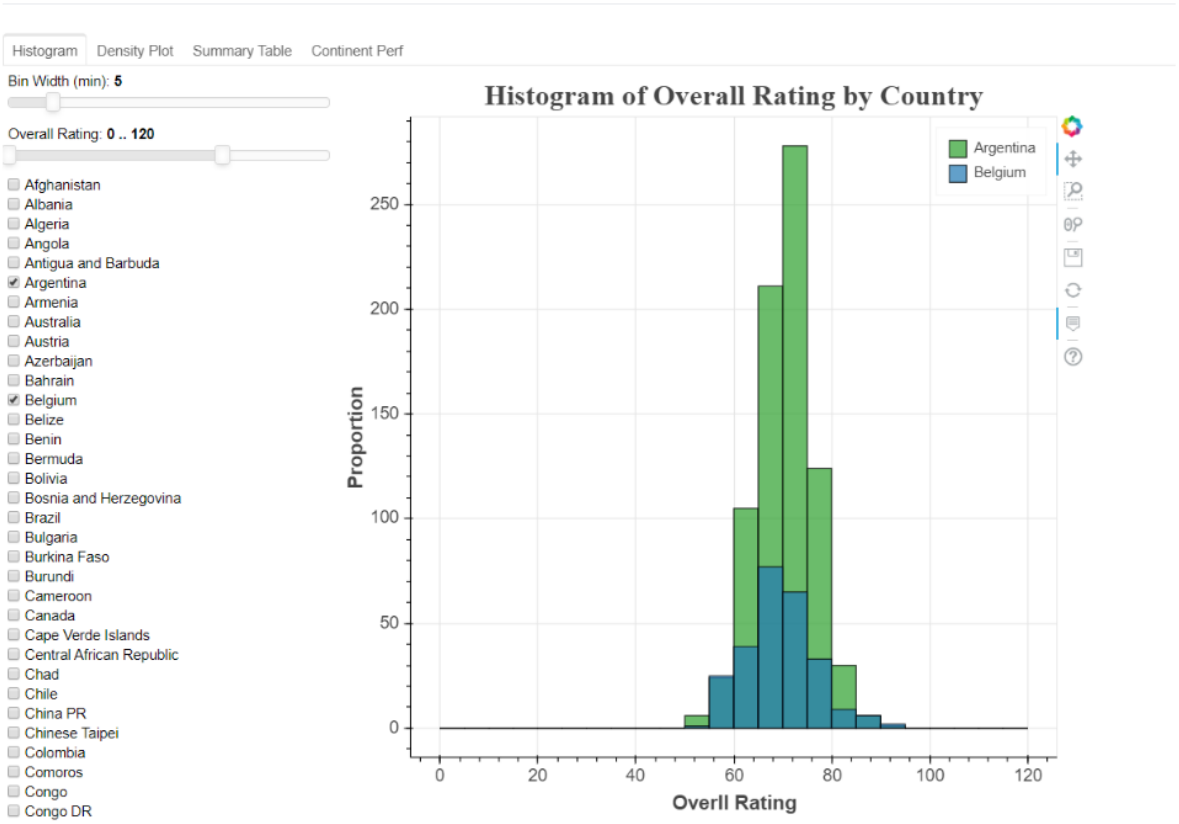
```
              Squad  Overall
Nationality
England         3-4-3    82.82
England         4-4-2    83.00
England       4-3-1-2    83.00
England         4-3-3    83.27
England       4-2-3-1    83.18
```

After obtaining an understanding of each country's formation strategy, it is worth comparing each country's overall rating. Therefore, a Bokeh is used to make interactive plots, which allow the user to select countries of interest and compare the overall ratings between these different countries (Tab1). The density plot allows the user to visualize the statistical distribution between each country (Tab2). The summary table provides the user with the counts associated with each country, namely: Min overall Rating, Mean Overall Rating, Medium Overall Rating, and Max Overall Rating.

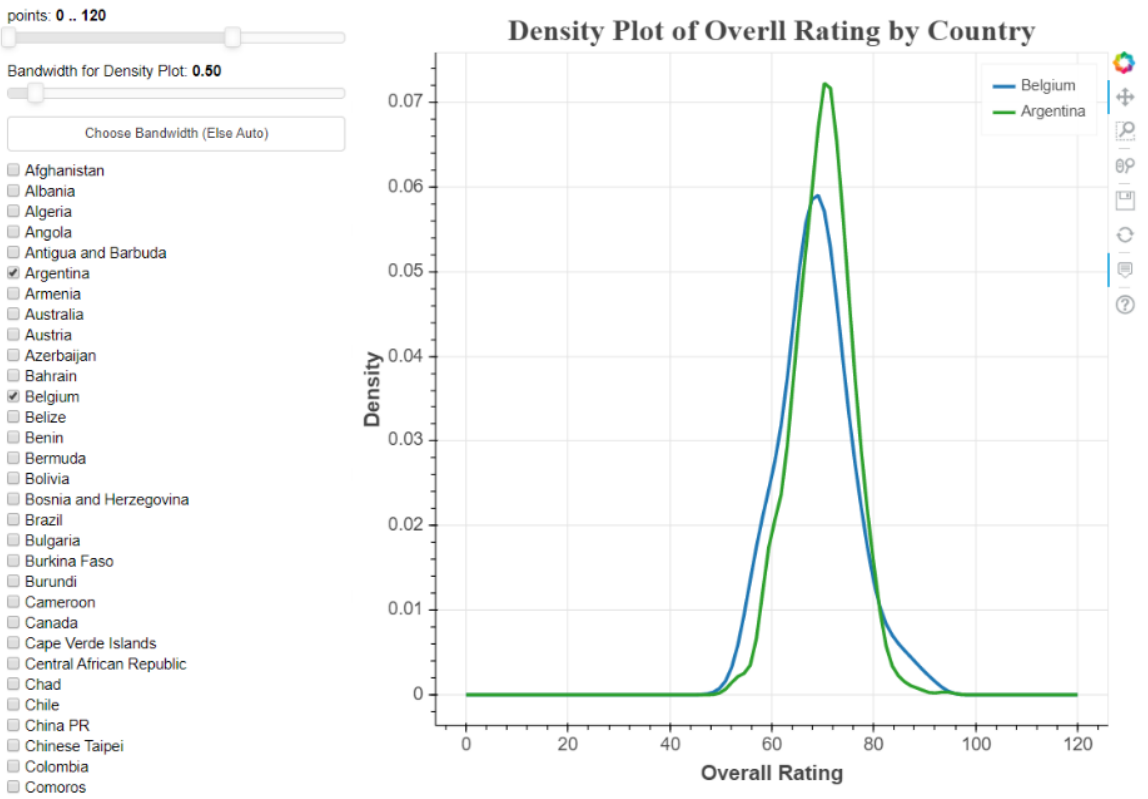
From the previous study, it was shown that the counts of Attackers, Defenders, Midfielders and Goalkeepers are different, for example, some countries might not have any Goalkeepers. Therefore, if a particular country does not have any Goalkeepers, they might want to obtain the player from another country in the same continent.

Therefore, Tab 4 is used to group countries into continents to see the overall rating of players vs. continent. The distribution of each country in the same continent is grouped together to see the overall ratings. For example, Bolivia can obtain a Midfielder from Argentina or Brazil to form their team.

Tab1



Tab2

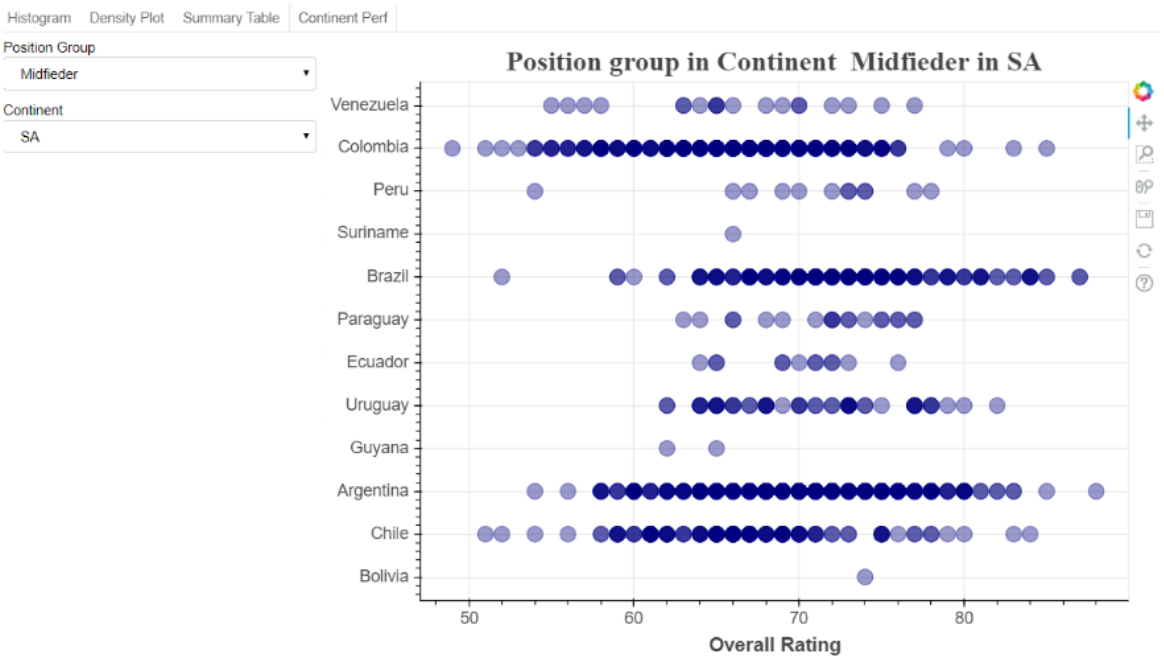


Tab3

Histogram | Density Plot | Summary Table | Continent Perf

#	Country	Counts	Min Overall Rating	Mean Overall Rating	Median Overall Rating	Max Overall Rating
0	Afghanistan	2	60	62	62	64
1	Albania	39	49	66.56	68	82
2	Algeria	46	59	71.76	73	84
3	Angola	15	62	69.6	71	78
4	Antigua and Barbuda	6	49	60.67	62	69
5	Argentina	785	52	69.97	70	94
6	Armenia	7	62	68.86	67	81
7	Australia	196	50	62.85	63.5	80
8	Austria	313	52	65.7	65	85
9	Azerbaijan	6	50	62.83	64	70
10	Bahrain	1	72	72	72	72
11	Belgium	257	52	68.88	69	91
12	Belize	1	66	66	66	66
13	Benin	14	63	67.93	68	76
14	Bermuda	3	59	63.67	62	70

Tab4


Histogram
Density Plot
Summary Table
Continent Perf
Position Group
Midfieder
Continent
SA
Position group in Continent Midfieder in SA
Venezuela
Colombia
Peru
Suriname
Brazil
Paraguay
Ecuador
Uruguay
Guyana
Argentina
Chile
Bolivia
50
60
70
80
Overall Rating