Storytelling Project - The Next World Cup Winner

Cyro Estevão Freire de Lima

Introduction:

The analyses of The Next Word Cup Winner helps to understand complex data and what are the factors most important in predicting a World Cup Winner. Soccer is the most popular and highest paid sport in the world, and many people bet money on sports or are fans enthusiastic about the World Cup. These analyses are not only storytelling that predict the winner, they can also be used for teams to understand who will be the difficult opponents and what are their strengths and weaknesses. Soccer data apps and websites have been very popular for a while, where professionals and amateurs try to understand the data to analyze competitions, and the World Cup is the biggest competition of all. The information visualization is designed to lead the spectator through the best teams in World Cup history, using treemap to show hierarchy in the overall best teams. On each slide the data zooms in and specifies recent periods of competition for deep analyses, using a Gantt chart with a timeline of all World Cups that uses a filter to highlight the strongest winner teams in the past ten Word Cups. The storytelling starts with the overall data from all of the cups, going through the last ten cups. The story ends by zooming in on the data in a dashboard for this current year of 2021. The dashboard uses bar graphs, tables, and filters for better comparison of the teams (stats, rank, and performance), using filters and images to highlight the scenarios of the teams in the competitions for better explanation of the data through storytelling. These visuals make clear the possibilities of winning for several different teams, showing the team with the highest probability of winning the next World Cup.

Dataset:

For this project two different data sets were used. The first data set has data for the soccer teams in all of the World Cups, exploring overall data from the competitions. This data set provided the first place winners, runners-up, and third place for each year of the competition as well as an overall score (number of matches won) for every team. It was used in the first three slides.

The second data set used was gathered from the FIFA official website and fotmob app, which is a professional app that contains soccer data from across the globe. The FIFA data was used for the Rank and the Points of each team as well as for the images of the World Cup brackets in the text filter, where you can click on the text filter and the images of the brackets will appear on the screen along with the table chart for that year. The images are for every year of the last seven competitions.

The fotmob app has the data for the Euro Cup, America's Cup, and the Qualifiers all from 2020 and 2021. It contains the stats number for each of the best teams. Color was used to highlight the winner of these two cups gathered from the data set, as well the team (Italy) that has not qualified yet for the next World Cup. Other values gathered are the numbers for goals, goals conceded, including overall

and per match. The last key from the data is the overall rating for the best teams from these competitions.

The data abstraction used for the first slide is the number of overall number of matches won from all the World Cups by country/team, filtering and highlighting the best ones.

The second and third slides are a timeline where the y axis is a table in alphabetic order with the winner of a world cup in a column, followed by the runners up (second place), and third place in the same row. The x axis is the year, both x and y form a Gantt map with a timeline of each group of teams that won for each year of the competition. The data also show the count for how many times a country won or was in second or third place.

The fourth slide is the table chart with the Rank Points for each one of the best countries and the most important countries in a World Cup, using traffic colors to highlight the winner, second, and third place, of the World Cup in that year.

The last slide has a table chart with the FIFA Rank in 2021 and also the Rating of the best teams in the Euro Cup and America's cup. The data also show in a bar chart the number of "win", "draw", and "loss" for each one of the most important teams in the Qualifiers as well as a second bar chart with the number of goals, goals conceded, overall and per match for each one of the best teams.

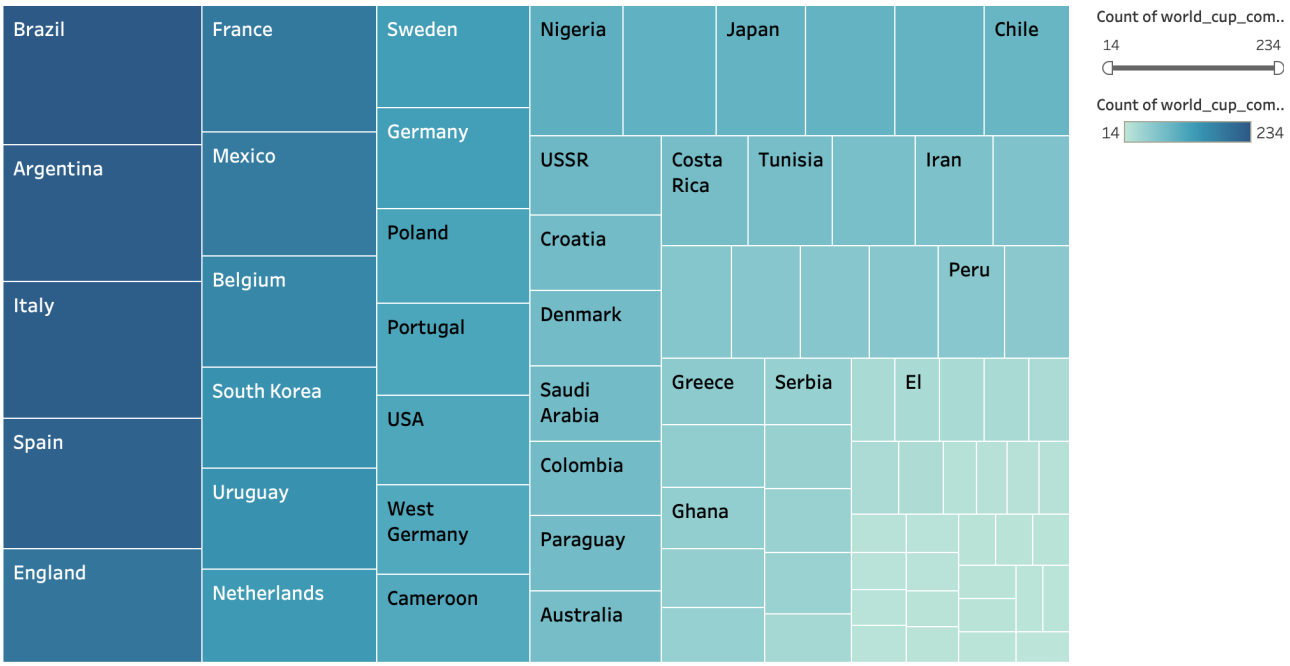
<u>Domains Task and Tasks Abstraction:</u>

- What are the best teams overall
    - Iterate through the Overall Performance statistics in the World Cup history
    - Compare Winners from all World Cups
        - Identify the best places and analyze what they have in common
        -
- Best current teams
    - Current most relevant winners
        - Filter the last ten World Cup winners and runners up
    - Rank odds for the teams and the luck factor
        - Attributes comparison for the ten last Wold Cup
        - Discover and identify the patterns using infovis and filters
        - Use Brackets image and text to explain the luck factor
    - The best teams (Rank/ Stats) in 2021
        - Filter the main attributes for team comparison
        - Target and highlight the favorites
- Conclusion
    - Summarize the previous domain and tasks abstraction
    - Provide The Next World Cup Winner

Visual Designs (Treemap, Gantt Chart, H. Bar Charts, Table Charts, V. Bar Charts):

The first idiom is a Treemap where white text, blue color, hue, and size show the best teams overall in World Cup history. The tooltip interactivity shows the score value of the teams (number of winning games overall). The reason a Treemap was chosen is because it is a good idiom to show hierarchical data, therefore highlighting the best teams.

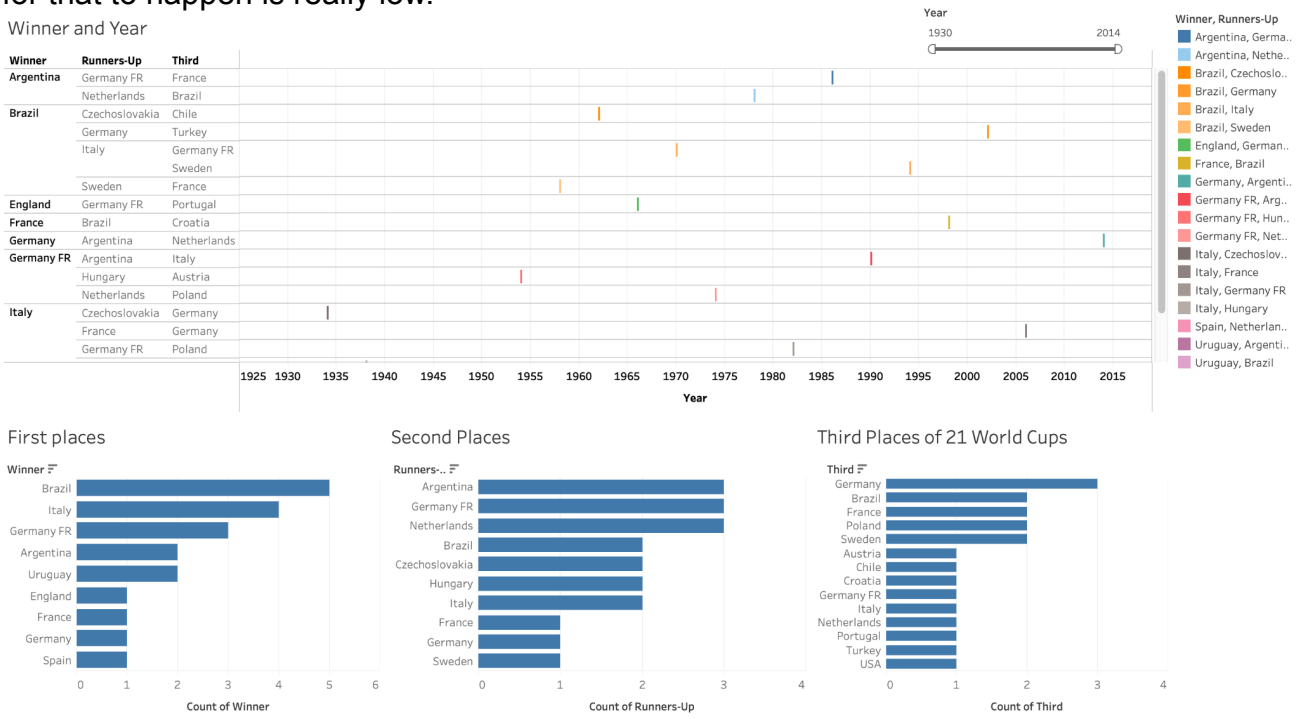
The Treemap explains about the teams that have a winning tradition in soccer and what that means. This is a fact that will weigh for the decision of the next World Cup winner. A filter on the right side can be used to redraw the Treemap with only the best teams. The tooltip shows the count of wins for every team in all the World Cups.



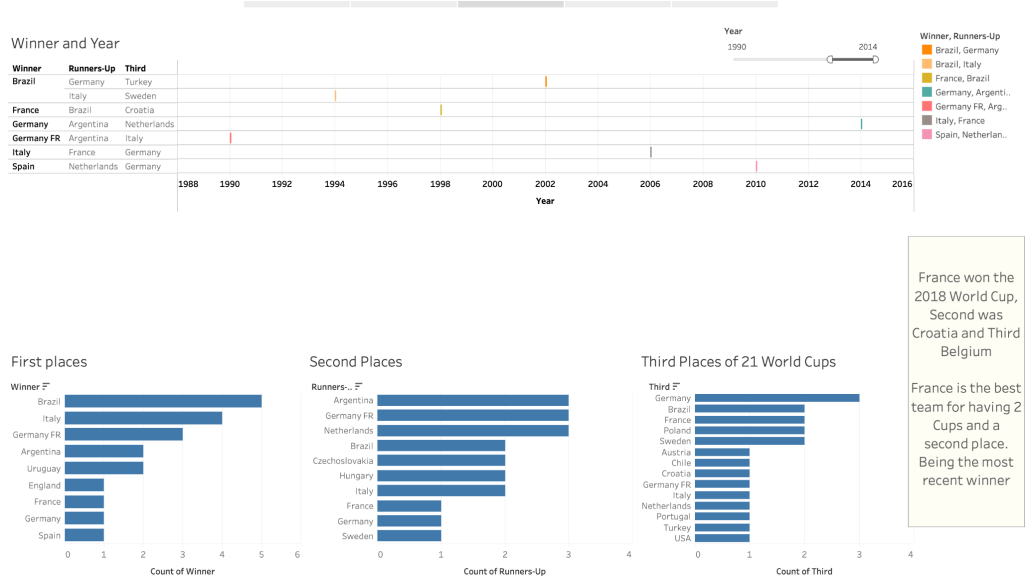
The second slide has values for the teams from 1930 - 2014 World Cups. In this dashboard I used a Gantt chart as the main idiom, using Winner and Year as keys. Usually a Gantt chart is used for project schedule, but in this case it has an easy way to see the colored bars in the timeline and the keys (Winner/Year). The bars and the colors shows a timeline of each year and the team who won that World Cup. The table on the left contains the Winner values followed by the Runners-up (second), and the Third place in alphabetic order.

Underneath the dashboard are three horizontal bar charts showing the count of the attribute winner, second, and third place for each value, sorted in descending order of attributes. The reason the horizontal bar chart was used is because it saves space on the dashboard at the same time as organizing the attributes in descending order of value, highlighting the ones with the biggest values. On the right side is a filter for the Gantt map where you can select the teams who won the World Cup

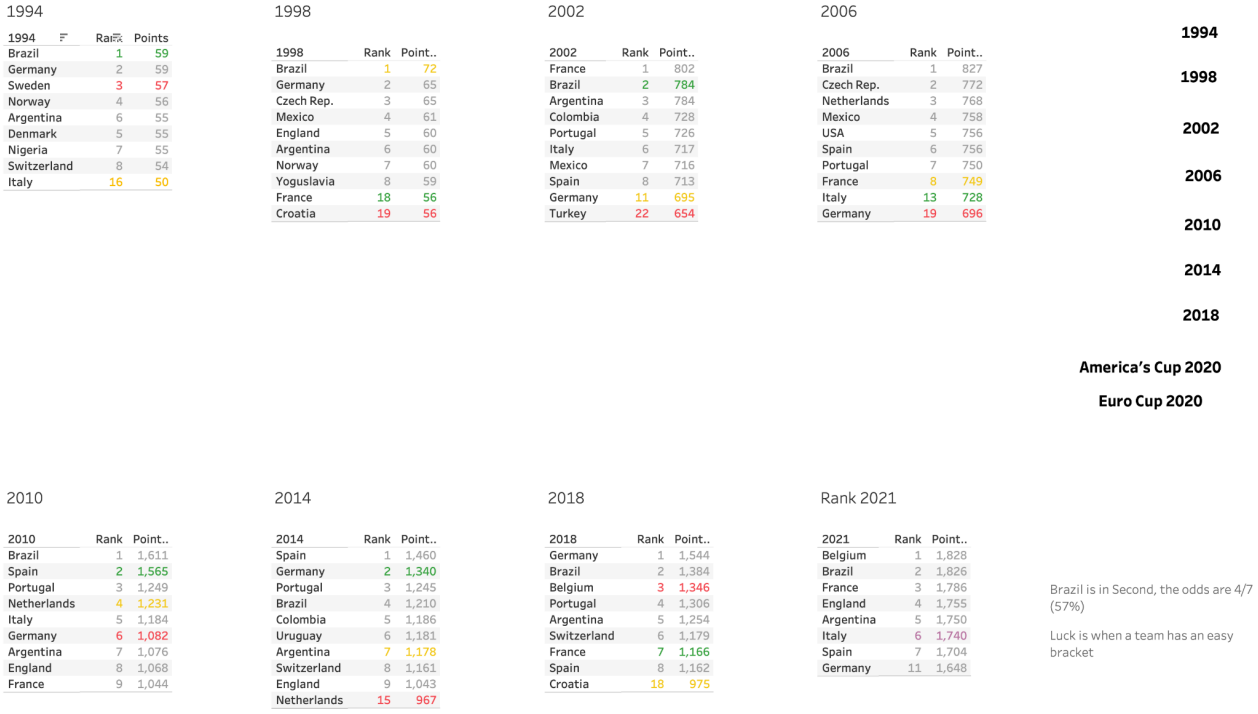
sorted by name, second, and third place. The values in the filter are organized by alphabet, color and hue for each one of the key Winners. On top of the dashboard is a filter named "Year" where you can filter the years that you want to check and consequently changing all the other idioms and filters, zooming in on the data. The hue and the color followed by the key makes the Gantt chart very well organized, and it is easy to filter by the Winners. This is a detail that makes a big difference in the visual organization of this amount of data. Seeing this visual idiom it is possible to notice that recently no team has won two World Cups in a roll, therefore the odds for that to happen is really low.



The third slide is the same as the second but with the filter activated, highlighting the best teams in the last ten World Cups. It is easy to see who are the three best teams (Brazil, France, Germany). In this data set the 2018 World Cup is missing, therefore it is explained in the storytelling and in the slides with a text.



In the fourth slide, the storytelling filters the last ten World Cups and point out the best team for these years (France, Germany, Brazil), presented in a dashboard. This slide is part of a second data set used in the project, it is based on the FIFA Rank for the last seven World Cups, including the last Euro Cup, America's Cup and the first part of the 2020/2021 Qualifiers for South America and Europe.



1994

1994	Rank	Points
Brazil	1	59
Germany	2	59
Sweden	3	57
Norway	4	56
Argentina	6	55
Denmark	5	55
Nigeria	7	55
Switzerland	8	54
Italy	16	50

1998

1998	Rank	Point..
Brazil	1	72
Germany	2	65
Czech Rep.	3	65
Mexico	4	61
England	5	60
Argentina	6	60
Norway	7	60
Yoguslavia	8	59
France	18	56
Croatia	19	56

2002

2002	Rank	Point..
France	1	802
Brazil	2	784
Argentina	3	784
Colombia	4	728
Portugal	5	726
Italy	6	717
Mexico	7	716
Spain	8	713
Germany	11	695
Turkey	22	654

2006

2006	Rank	Point..
Brazil	1	827
Czech Rep.	2	772
Netherlands	3	768
Mexico	4	758
USA	5	756
Spain	6	756
Portugal	7	750
France	8	749
Italy	13	728
Germany	19	696

1994
1998
2002
2006
2010
2014
2018
America's Cup 2020
Euro Cup 2020

2010

2010	Rank	Point..
Brazil	1	1,611
Spain	2	1,565
Portugal	3	1,249
Netherlands	4	1,231
Italy	5	1,184
Germany	6	1,082
Argentina	7	1,076
England	8	1,068
France	9	1,044

2014

2014	Rank	Point..
Spain	1	1,460
Germany	2	1,340
Portugal	3	1,245
Brazil	4	1,210
Colombia	5	1,186
Uruguay	6	1,181
Argentina	7	1,178
Switzerland	8	1,161
England	9	1,043
Netherlands	15	967

2018

2018	Rank	Point..
Germany	1	1,544
Brazil	2	1,384
Belgium	3	1,346
Portugal	4	1,306
Argentina	5	1,254
Switzerland	6	1,179
France	7	1,166
Spain	8	1,162
Croatia	18	975

Rank 2021

2021	Rank	Point..
Belgium	1	1,828
Brazil	2	1,826
France	3	1,786
England	4	1,755
Argentina	5	1,750
Italy	6	1,740
Spain	7	1,704
Germany	11	1,648

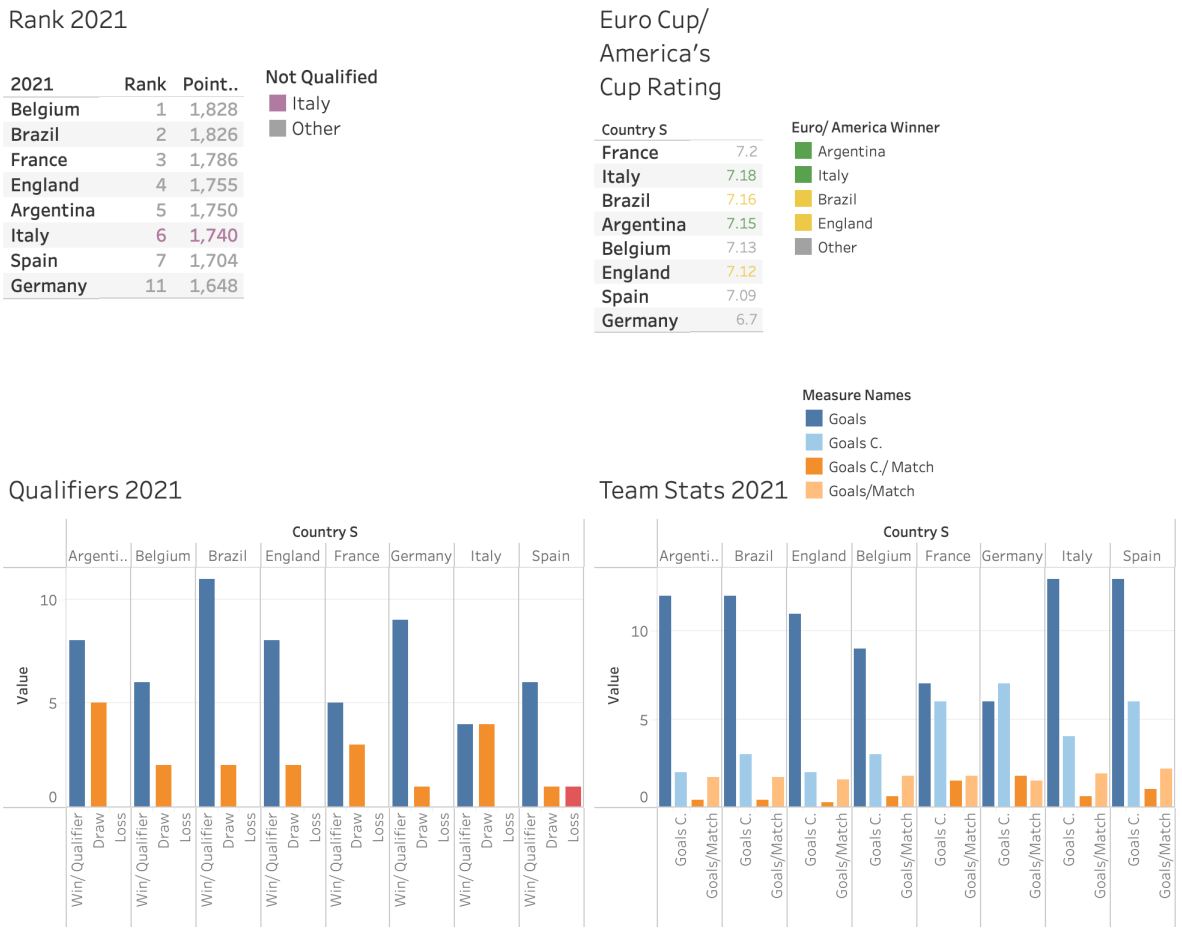
Brazil is in Second, the odds are 4/7 (57%)

Luck is when a team has an easy bracket

The Table charts are used for faster visualization and comparison of the rank and point numbers, with the attribute "countries" organized by the year of the competition and ordered by Rank position. On the right side it is possible to see the text with the years of the World Cups and the 2020 local Cups. The text is a filter you can click on and it shows the knockout round brackets for every item available. The data on the tables are filtered, so only the teams that are important for the data and storytelling are showed. For that reason, the key Rank does not necessarily have the same quantity of values, it usually shows the best eight teams, after that most of the teams do not have importance. The traffic light color is used to show the first, second and third place positions for that competition.

The storytelling goes through the odds of the second place. The second position team won four times in seven World Cups, having a percentage of 57% rate of winning. The story describes what is the luck factor that some of the teams needs to go far in the knockout round, and how that can affect a team for good or bad. The example given is Croatia who finished in second place in 2018 despite having a bad team statistically going into the Cup, and France who lost the final to Italy who was lucky in the brackets sort, while France had bad luck facing four difficult teams in 2006 and losing the final match for a lower ranked team. The luck factor can help a team with a tradition in winning the World Cup, but not those who do not have such a tradition.

The fifth slide is a dashboard with two tables and two bar graphs. The storytelling zooms in to the year of 2021, showing the Rank of the teams for this year and the Rating for the last Euro Cup and America's Cup, also pointing out the stats for the first part of the Qualifiers for the 2022 World Cup. The data is focused on the current year and on the main attributes of the team: number of goals, goals conceded, performance (win, draw, loss), overall rating, and the FIFA Rank Points.



The Bar Graphs are used with color and hue to highlight the main values and its keys (Country/ teams) for easier comparison of the two different color bars. The keys are in descending order (from left to right), organized by the performance in the America's/ Euro Cup, and in order to fulfill the storytelling.

Example: Argentina is in first because they won the America's Cup and they had a good performance in the Qualifiers. Italy is almost the last because they won the Euro Cup, but they are the only team in the graph that did not qualify for the World Cup yet, having a good stats in the Euro Cup but a poor performance in the Qualifiers (most recent competition), potentially not to qualify for the next World Cup.

Possible Improvements:

The improvement that could be done to the project is to add more data specific to the individual teams performance in 2021 such as attack, defense,

goalkeeper, team strategy (defensive/ offensive formation), as well as best players and goals scored by them. This would show the stats and numbers for easier comparison among the teams with the intent to further explain the quality of each team. That data would help to reinforce the quality of the teams showing, and further explaining the best teams, adding more context to the storytelling for the spectator.

The design for improvement can be a dashboard with the main teams using line up bar to show all the stats and attributes about the teams. The dashboard would contain the best players for each team and the user could also click on a button that is a filter, that would show an image with all the players for that team, organized in a field with the tactical formation.

One of the graphs in the dashboard can be a Rank with the best Strikers/ Players from the competitions (Euro/ America's Cup, Qualifiers). The interactivity would be the line up chart, that you can use the a bar to scroll through the data, and also the filter that would show the data about the team, as previously described.

Examples for improvement using a dashboard:

When you clicked on a team, an image would show the players and the stats (number) for that player (image on the left without stats number). In the middle is the rank for the best players of the competition and their stats number. On the right is found the line up chart with the teams attributes. Clicking on the team name would show the image on the left. The last image is a Team Comparison filter, to compare the attributes of two or more teams.

Results (Insights, Storytelling):

To create the storytelling I had a main idea to use the data for the prediction of the next World Cup winner. However, while I was working on the project I did many different research queries to check and answer questions like "Why one team is the

best and not the other?" The first insight appeared when I checked the FIFA Rank and it was clear that there is a pattern between the position in the Rank and the Winner of the World Cup.

The second insight was when I was trying to answer the question "Why some teams that were not good in the FIFA Rank won the World Cup?" That is when the tradition and the overall quality of the teams in all the World Cups appeared from my data set, and I realized that only teams with tradition ever had a comeback from a bad position and won the World Cup.

The third and last insight was when I analyzed the brackets and saw that some teams had luck in the sorting of the knockout round brackets, and in some of the Cups that affected who was the winner, second and third place. However, never a team that was among the ones with no tradition in soccer ever won the Cup.

The storytelling is linear, focusing on the main keys and attributes to highlight the best teams, giving multiple views of the many World Cups with the intent to fill any gaps in the story to provide the best result. After all the insights provided by the data I filtered the attributes and teams, to focus only in the data that would provide the possible winner of the 2022 World Cup.

The first slide uses a Treemap to highlight the best teams overall the World Cups, explaining the these teams have tradition in soccer and that is a factor that weighs for winning a World Cup.

The second slide is a dashboard with a Gantt chart as main idiom that explains that only twice has a team won two World Cups in a roll in total of 21 World Cups, and that was when the soccer competition was different. That never happened in modern times.

The third slide is a filter of the second where the data is zoomed in on the last 10 World Cups, showing that France is the best team along with Germany and Brazil due to the number of first place and second place over these years. The first and second places are shown in a horizontal bar chart in descending order in the bottom of the dashboard, highlighting the teams with the biggest number of victories for that key (first, runners up, third places).

The fourth slide has many table charts with the FIFA Rank and points for the team of that year of the World Cup. The data used is the last one available right before the World Cup. In the tables it is possible to identify that the second position in the Rank is the one with the highest probability of winning a Cup (57%), and for 2021 Rank this team is Brazil. The text filter (Years/ Competitions) in this dashboard you can click on and shows an image of the brackets for the specific year or competition. With this image you can analyze the matches for the knockout round and it is explained how France had a bad luck and lost the 2006 World Cup, and Croatia had a good luck and finished in second place in the 2018 World Cup. Also that a non traditional team with a bad ranking can have luck in a World Cup, but only a team with tradition can win a World Cup having luck and a bad ranking for that year.

The fifth slide is a dashboard focusing only in the years of 2020/2021, where happened the Euro/ America's Cup and the Qualifiers. It shows the FIFA Rank for

the end of the year and the best attribute (rating) for the best teams. Then, the data goes deeper and shows the attributes goals, goals conceded, overall and per match in the Euro/ America's Cup, showing which country has the best attack/defense, comparing this data with the Qualifiers attributes (win, draw, loss). Using this data it is possible to filter the best teams from the worst teams and one more time Brazil shows in one of the best positions. After filtering the teams that do not have much chance due to their attributes, only three teams are left: Brazil, Argentina, and England.

Now considering the tradition, all three of these teams have it. Considering best team in the last ten World Cups, only Brazil appears. Considering Rank, Brazil is in the best position followed by Argentina together with England. Analyzing the year of 2021 the three teams are the Best with Argentina being better than the other two.

As Brazil appears in the best situation for many of the criterias, it is the favorite team followed by Argentina, who won the America's Cup, and England who was second place in the Euro Cup.

After my project and storytelling were finished, many articles about the next winner appeared in newspapers like The Guardian, and always being the favorite team to win the next World Cup, is Brazil. Many of these articles point to France as one of the favorites, but the data show that France does not match many of the criterias to win, and that is another reason why data projects are important.

That is the link for the video presentation:

https://youtu.be/UvhddOXr75Q