

FIFA match results and predictions for the last three decades. Delivery

Date: February 10, 2023

Student: Sebastián Vega Uribe

Table of Contents

1. Data entry	3
2. Description of the data subject	3
3. Definition Relational model:	4
4. Table relationship diagram	5
5. List of tables	6
6. List of tables by columns	7
7. Transformations performed	8
8. Board analysis	11
9. Calculated measurements	13
10. Conclusions	14

1. Data entry

The FIFA World Cup is the most prestigious soccer tournament in the world. The championship has been held every four years since the tournament's inception in 1930.

The current game format consists of a qualifying phase, which takes place during the three previous years, to determine which teams can participate in the tournament where 32 teams compete for the title in different stadiums in the host country.

The past champion was France, who beat Croatia in the 2018 tournament in Russia and in the Qatar 2022 version, Argentina emerged champions in an agonizing final against Le blues former champion France.

This dataset provides a complete overview of all international soccer matches played since the 1990s, plus the strength of each team is provided by incorporating actual FIFA rankings.

2. Description of the data subject

The data used comes from an Excel database, which has almost 24,000 entries of all FIFA matches played since 1993, considering the matches of each World Cup to take into account the qualification for the next World Cup by score, by host country, direct qualification or playoffs. In Africa-CAF, Asia-AFC, Central America-CONCACAF, Europe-UEFA, Latin America-CONMEBOL and Oceania- OFC.

Having a database with such an extensive timeline and with such detailed information, it is possible to perform several types of analysis, below, we will see a brief analysis of how to apply each one to our topic:

- Descriptive: What happened, the database has more than twenty-three thousand matches played since 1993, it tells us in detail which nations played against each other, where and when.

- **Diagnosis:** Why did it happen? The database also includes the number of points to date, as well as the tournament in which they played each other, providing information to create a diagnosis by continent.
- **Predictive:** what will happen, thanks to the extensive database, an average of 113 games per selection can provide trends, averages, modes and medians.
- **Prescriptive:** what should I do, based on past results that resulted in a win, various scenarios that can help this outcome (in addition to winning all or most of the games) can be realized.

3. Definition Relational model:

Project objective:

Thousands of records of matches played during the last 30 years in various competitions, but all with the same goal, the FIFA World Cup, are counted with the date, place and result of each one.

With this information, you can analyze the trend in the results of each game, whether you win, lose or draw at home and how the number of goals scored or conceded influences the results.

Scope and user:

Additionally, the trends of the countries with the most games won and the difference in goals scored will be analyzed, thus getting an idea of how much scope a certain team can have according to the trends and projections of the last matches played. By having the information of all the participating teams to date, anyone interested in the sport can use the information, but its focus is for the bookmakers' administrators,

This is a predictive analytics application that analysts use at the operational level for

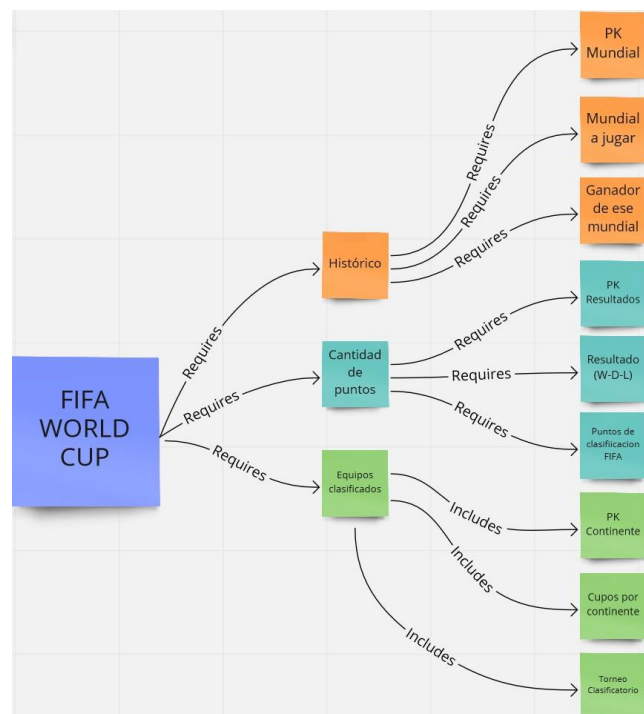
establish the winning and losing margins for each bet, in addition to providing observational information to coaches and athletes.

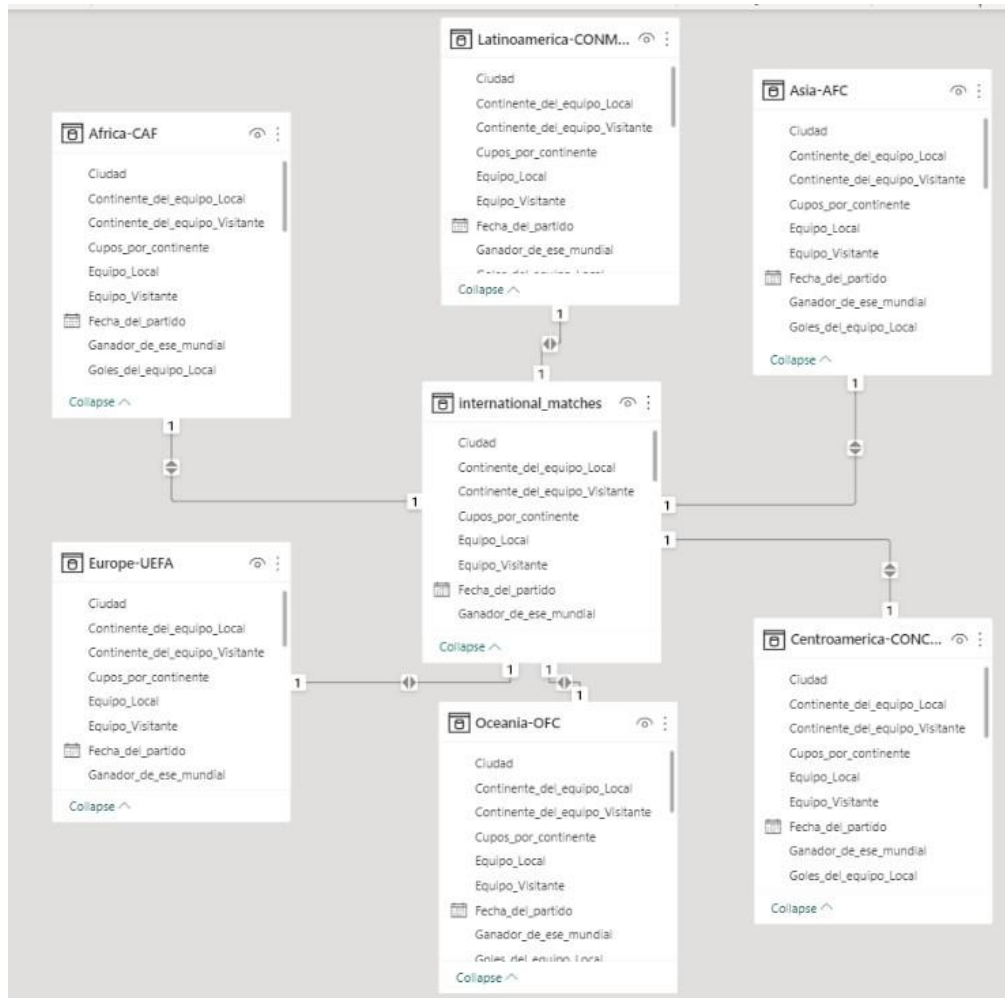
Among the data in the database, in addition to the tables (which will be analyzed one by one later), we have precise information on location, city, country and continent. The qualifying tournament, the result of the match, the number of goals conceded and scored, and the points obtained for each match.

4. Table relationship diagram

The table relationship diagram was created based on the qualification to the World Cup regardless of the number of slots or playoffs, starting from the historical basis to analyze trends, the number of points needed to advance and the teams qualified according to continent.

Each table has a primary key, which can define the probabilities of classification according to the results, within each table, the quotas and the qualifying tournament were taken into account as a factor to be considered, which is further evidenced in the modeling of the database tables.





As mentioned above, the database was divided today into tables according to the continent and the tournament to be played for the respective classification.

5. List of tables

Next, we are going to visualize the tables found in the History base

- PK_ID_MUNDIAL
- World_to_play
- Winner_of_that_world_winner

Number of points

- PK_ID_Result_Local_result
- Local_team_results
- FIFA_FIFA_Local_team_ranking_points

Qualified teams

- PK_ID_LOCAL_CONTINENT
- Quotas_per_continent
- Qualifying_Tournament

6. List of tables by columns

In each table we see the type of field, its key and if it is Null.

TABLE 1 Historical			
Field	Field type	Type of key	Null ?
PK_ID_MUNDIAL	int	PK	N
World_to_play	int	X	N
Winner_of_that_world_winner	int	X	N
PK_ID_Result_Local_result	int	FK	N
Local_team_results	int	X	N
FIFA_team_rankings_points_Local	int	X	N
PK_ID_LOCAL_CONTINENT	int	FK	N
Quotas_per_continent	int	X	N
Qualifying_Tournament	int	X	N

Table 1: in this first table we obtain a preview of the historical winners of the World Cups played during the last 30 years, understanding as determining factors the results of the matches played during the qualification to the tournament, as well as the FIFA ranking points. In this way, we will obtain the history for each country, match and results, according to the quotas per continent.

TABLE 2 Number of points			
Field	Field type	Type of key	Null?
PK_ID_Result_Local_result	int	PK	N
Local_team_results	int	X	N
FIFA_FIFA_Local_team_ranking_points	int	X	N

Table 2: The second table defines the total number of points regardless of the wins, losses or defeats that the teams experienced throughout the tournament and pre-qualifying rounds, the purpose of this table is to compare with the historical records and against the data of the winner of the respective world cup.

TABLE 3 Qualified teams			
Field	Field type	Type of key	Null?
PK_ID_LOCAL_CONTINENT	int	PK	N
Quotas_per_continent	int	X	N
Qualifying_Tournament	int	X	N

Table 3: the third table shows the teams ranked by each continent according to points, games played, games won, goals scored and conceded, as well as point difference. The purpose of this table is to show the trend for each continent and championship where Europe has more places compared to the rest of the world and the trends they show.

7. Transformations performed

The first part of the dashboard consists of a general summary of the information contained in the database. As only one table was available, it was divided by continent and its respective confederation, taking into account the number of quotas that each one has per world cup. This information is based on the current system of

32 quotas in total, CAF in Africa with 5 quotas, AFC in Asia with 4.5, CONCACAF in Central America with 3.5, CONMEBOL in South America with only 4.5 and UEFA in Europe with 13 quotas.

In addition to these changes, calculations were also made in both the base and the board, to deduct the points obtained with a win-loss and draw system, providing 3, 2 and 1 respectively, as well as totals, general totals categorized by country, continent and result.

And finally, there is a detailed analysis of the current champion Argentina, with a history of the last 30 years, segmentation by dates and others that will be analyzed later.

- **International Matches Table**

A consecutive number is generated for each World Cup since 1994, as well as a number is assigned for each result, be it win, lose or draw (these calculations were applied to each divisional sheet by continent). For this purpose, calculated columns were created to present the relevant information for each country.

- **Africa-CAF Table**

The African soccer confederation has 54 participants and only 5 quotas, the country with the most matches played is South Africa with 229 matches, the winningest team is Egypt with 163 matches, South Africa again appears as the losingest team with 48 matches and the one with the most draws with 54. This gives us a total of 5885 matches played by all the teams over the last 30 years.

- **Asia-AFC Table**

the Asian Football Confederation has 46 teams and only 4.5 slots for the tournament, with a total of 5302 matches, the most successful team is Arabia.

Saudi Arabia with 164 matches, with 64 matches lost Qatar and the Arab Emirates occupy the first place with the most matches lost and with the most draws is Japan with 63 matches, which coincidentally is the same number of matches lost.

- **Central America-CONCACAF Table**

The North Central America and Caribbean Association Football Confederation has 34 participants and 3.5 quotas for the World Cup, with a total of 2772 matches since 1994, the team with the most defeats is El Salvador with 65, with the most ties is Mexico with 66 and at the same time the most matches played with 316. The team with the most victories is the United States with 194.

- **Europa-UEFA table**

The Union of European Football Associations has 55 countries, but a total of 13 quotas for the tournament with a total of 7593 matches played during the last 5 tournaments where the team with the most appearances is France with 221, the winningest teams with 145 matches are France and Spain, the losers are Luxembourg and Malta with 95 matches and with 53 matches the team with the most draws is Italy.

- **Latin America-CONMEBOL Table**

The South American soccer confederation has 4.5 quotas to the World Cup, where 10 teams participate, with a total of 1839 matches, the team with the highest number of participations is Brazil with 233, being at the same time the most winning team with 181 matches, Bolivia has the highest number of draws with 39 matches and Venezuela with 46 lost matches occupies the last place.

- **Oceania-OFC Table**

Finally, with 0.5 quotas, the Oceania soccer confederation has 7 teams which have played 530 matches, with Australia being the biggest participant with 169 matches and the biggest winner with 112 matches, and the one that has lost the most.

with 31 and the most tied with 26.

8. Board analysis

Below is an analysis of the data presented in each tab of the Power Bi project, what can be found in each graph and how an analysis of which team is Argentina's champion was performed.

Tab 1: Cover

Resultados y predicciones partidos FIFA últimas 3 décadas, Argentina campeón.



A través de una base de datos que contiene partidos que datan desde 1994, de todos los encuentros FIFA oficiales realizados hasta la fecha (antes del mundial Catar 2022), se realizó un análisis predictivo que en su momento dio como resultado a un país latinoamericano campeón (Brasil).

En este documento se analizará las métricas de Argentina, durante los últimos 30 años.

Curso Data Analytics
Comisión 47645

Estudiante: Sebastian Vega Uribe



Tab 2: Global data entry

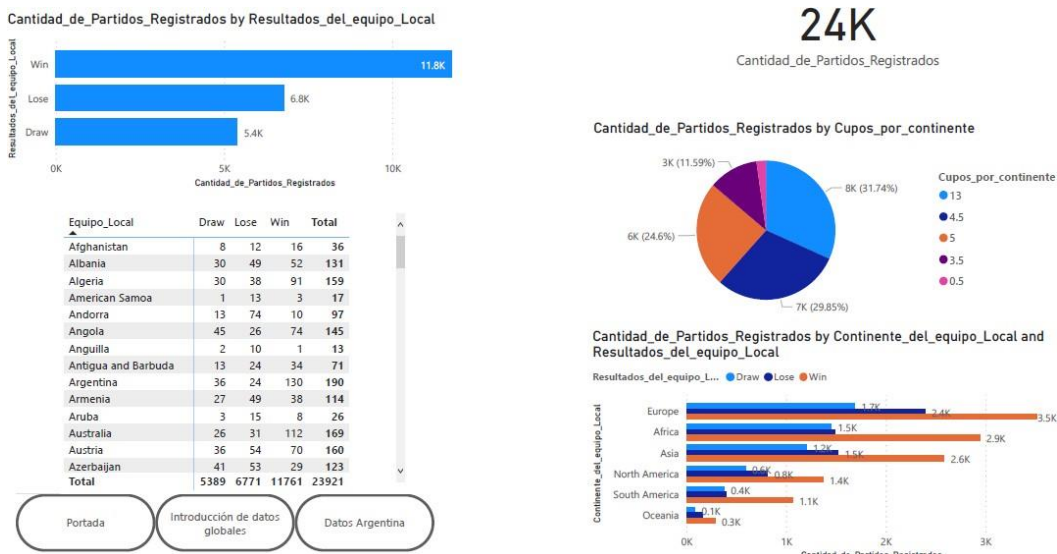
In the first tab we find a general summary where it mentions that during the last 3 decades around 24,000 matches have been played (23,922 to be exact), where 64,048 goals have been scored, of which 38,494 are home and away goals. 25,554 visitors.

Presentación de datos globales



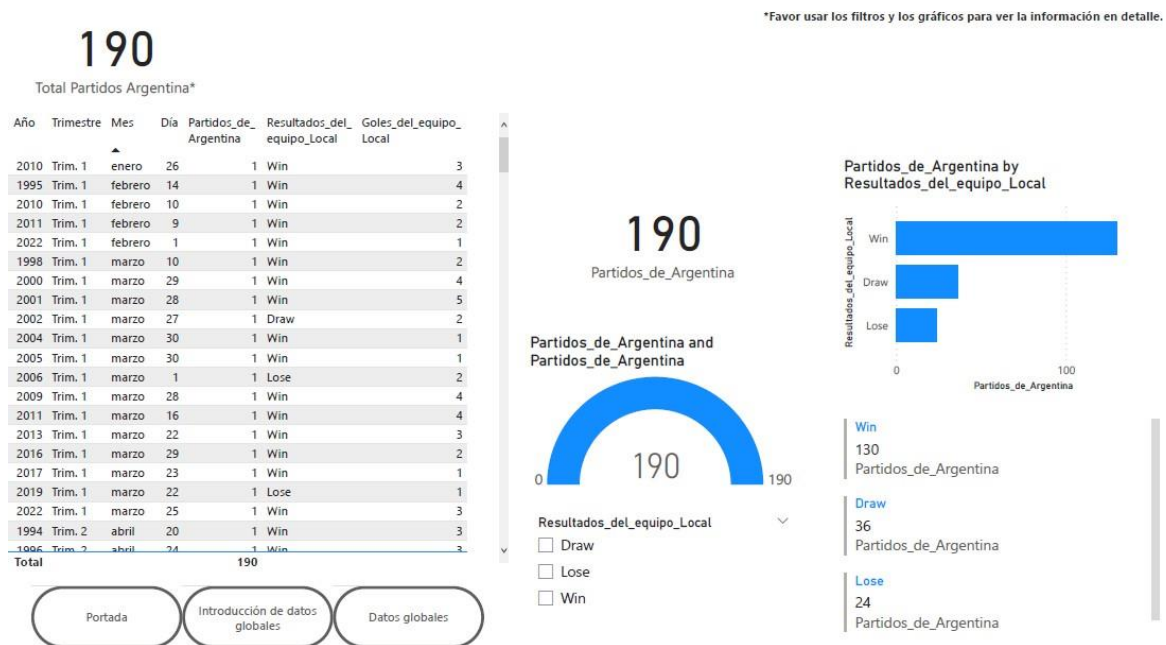
Tab 3: Global data

In the next tab we see in detail how the team is divided between home and away, where the information displayed is of the home team, and we see each category of winning, losing or drawing, at a glance you can see the advantage that the teams have playing at home. Likewise, we see how these categories are distributed by continent and how over the years the quotas are playing a percentage in each cup.



Tab 4: Argentina data

Finally, we see in detail the information of the champion team, since 1994 Argentina has played 190 matches, of which 130 resulted in victory. This tab has interactive information when you click on any of the graphics, today you can see the date, the number of goals, the number of matches and the result of that match.



9. Calculated measurements

As mentioned above, the largest amount of calculated data is found in the Argentina analysis tab, there is a calculated field that adds up all the matches played (190) since 1994, as a curious fact of these 190, they have won 130, drawn 36 and only lost 24.

Within the calculated fields there is a percentage measure of the matches won by the Albiceleste national team, as well as a specific field for this country only, below are the formulas:

Argentina_matches =

CALCULATE([Number_of_Registered_Matches],international_matches[Local_Team] = "Argentina")

Percentage_Measure =

CALCULATE([Number_of_Registered_Matches],international_matches[Local_Team_Results] = "Argentina")

And finally there is a table of dates where you can visualize year by year grouped by semester the day the result and the number of goals scored, which shows a trend of victory is quite high during the last 30 years.

10. Conclusions

Making any type of prediction in sports includes random factors impossible to forecast and considering that the most popular sport on the planet moves hundreds of masses and generates millions in bets, merchandise, travel and tickets, today it is difficult to be rational in front of this type of data, which generates a higher degree of difficulty to this type of projects. Initially the results showed Brazil as possible champion, following the trend that the last World Cup played in Asian territory was won by Brazil, the second point to consider was that it could be a Latin country. Of the two criteria, only one was achieved, despite having more than 24,000 match entries since 1994.

Although it is possible to perform trend analysis and analyze the information in detail, chance factors have a direct influence on all models. The clear example of this is Australia and Egypt. The former has the highest number of matches played and won in its tournament, but has only participated in 5 of 21 World Cups, where it has never disputed a title or reached a final, while Egypt is the team that has won the most matches, today it has only participated in 3 World Cups. While Morocco, with only 6 participations in the World Cup Y

124 matches won out of 193 and made history by becoming the first African and Arab country to reach even more World Cup semifinals with only one goal conceded and whose initial odds as a member of Group F were not the best as they were up against the likes of Croatia and Belgium.

On the other hand, the finalists of the last World Cup had high odds as France as champion of the Russia 2018 World Cup and Argentina eternal favorite played a match to the delight of the planet, whose odds of meeting were high, but favored the European team.