

CONCACAF World Cup Qualifying Match Predictions

Project by Aaron O'Neal

A dark blue diagonal gradient bar that starts from the bottom left and extends towards the top right, covering the lower half of the slide.

The FIFA World Cup



Every 4 years, the world comes together to watch the most prestigious international competitions of the world's most popular sport, the World Cup. The competition is hosted by FIFA (Fédération Internationale de Football Association) and is comprised of countries from all 6 qualifying regions (AFC, CAF, CONCACAF, CONMEBOL, OFC, UEFA). The 211 member countries of FIFA participate in a multi-round round-robin tournament in their specified region in hopes of advancing and qualifying as one of the 32 teams that participate in the FIFA World Cup.

Some background on soccer/football and round-robin tournaments



In soccer/football, there are 3 possible outcomes of a match (win, loss, draw). In round-robin tournaments, teams play a specified number of matches against the other teams in their group. For each match, a team either earns 3 points for a win, 1 point for a draw, or 0 points for a loss. After all matches are played, the team with the most points is declared the winner of the group and advances in the tournament. However, if there is a tie in points after all of the matches are played, the winner between those teams is determined by which team has the highest goal differential then by which team has scored the most goals in the tournament.

All of the points and goal stats are kept on a table to track which team is leading.

Problem Statement/Goal of this Project



Predicting events and outcomes in soccer/football has proven that the sport is one of the most difficult sports to predict. If you follow the sport, Leicester City (an English Premier League team) were given 5,000:1 odds to win the Premier League in the 2015-2016 season and they did.

With low-scoring games and so many different factors that determine the outcome, it is very difficult to predict a win/loss/draw outcome let alone the final score of a match.

Although difficult, the goal of this project is to predict the scores in the CONCACAF World Cup Qualifying matches and help gambler's gain slight edge when betting on soccer matches.

The Data

- Results of International Soccer Matches played from 11/30/1872 to present day.
(<https://www.kaggle.com/martj42/international-football-results-from-1872-to-2017?select=results.csv>)
- FIFA Rankings since 12/31/1991
(<https://www.kaggle.com/cashncarry/fifaworldranking> and <https://www.fifa.com/fifa-world-ranking/>)
- Upcoming/completed CONCACAF World Cup Qualifying Matches
(<https://www.fifa.com/tournaments/mens/worldcup/qatar2022/qualifiers/concacaf>)
- The current CONCACAF World Cup Qualification table
(<https://www.fifa.com/tournaments/mens/worldcup/qatar2022/qualifiers/concacaf>)

Cleaning and Preprocessing the data

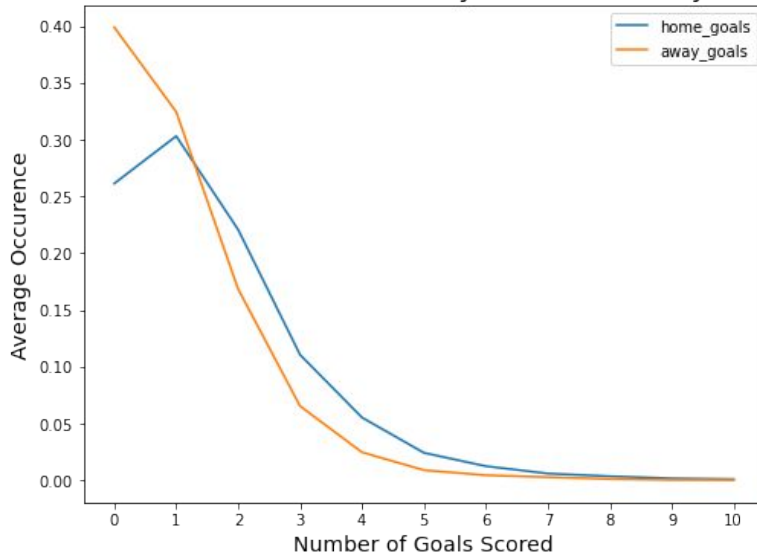
- Bringing in and creating features to add to the dataframe of results (rankings, goals for/against avgs, whether the match was played at a neutral site, and what type of match)
- Removal of matches prior to 12/31/1992 (date of first FIFA rankings) as well as matches of teams prior to their first FIFA ranking
- Removal of non-FIFA associated teams
- Writing functions to bring a teams most recent rank on or prior to that matchday as well as functions to aggregate a teams previous 10 game averages of goals scored/conceded when playing at home/away

Exploratory Data Analysis – Goals Scored



- Home teams score 1 goal most often on average
- Away teams score 0 goals most often on average
- Both home and away teams are less likely to score higher amounts of goals, however the home team has a greater chance to score multiple goals than the away team does

Distribution of Goals Scored by Home and Away teams



Exploratory Data Analysis – Wins and Losses

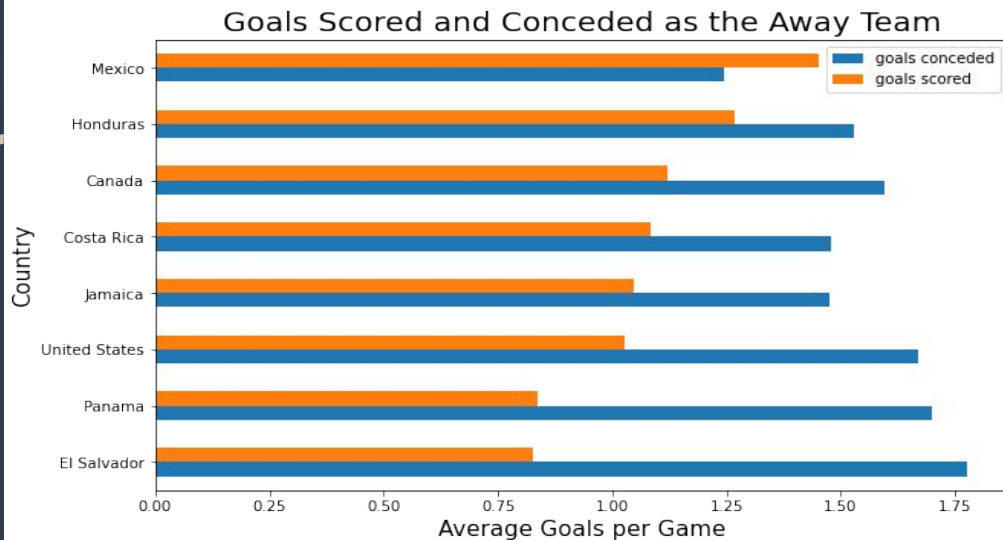
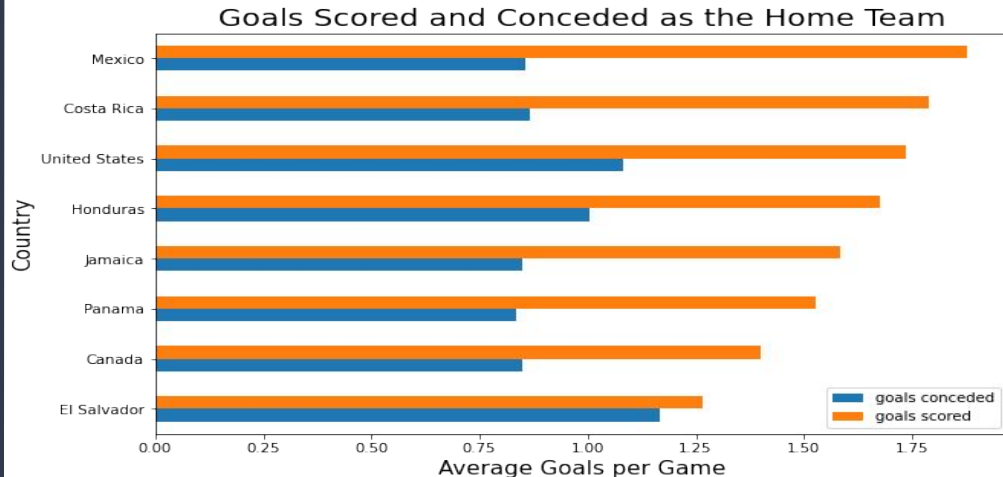
outcome	average_occurrence
home_win	48%
away_win	28%
draw	24%

outcome	average_occurrence
higher_rank_wins	55%
lower_rank_wins	21%
draw	24%

Exploratory Data Analysis - CONCACAF teams

World Ranking going into WCO:

Mexico - 9th
United States - 10th
Costa Rica - 44th
Jamaica - 50th
Canada - 59th
Honduras - 63rd
El Salvador - 64th
Panama - 74th



Modeling to predict home and away scores



Two separate models created (one to predict home_score and the other to predict away_score):

- Both models used 4 layer neural networks with a linear regression output that was then rounded to an integer for predictions
- Features used in both models: neutral location, home and away ranks, 10 game averages of home/away goals for/against, and the type of match being played
- Home_score model finished with r2 score of 0.18 and MAE of 1.0
- Away_score model finished with r2 score of 0.11 and MAE of 0.81

Predicted Scores of CONCACAF World Cup Qualifiers

date	home_rank	home_team	home_score	away_score	away_team	away_rank
2021-09-02	59	Canada	2	0	Honduras	63
2021-09-02	74	Panama	1	1	Costa Rica	44
2021-09-02	9	Mexico	3	0	Jamaica	50
2021-09-02	64	El Salvador	0	2	United States	10
2021-09-05	50	Jamaica	2	0	Panama	74
2021-09-05	64	El Salvador	2	0	Honduras	63
2021-09-05	44	Costa Rica	1	2	Mexico	9
2021-09-05	10	United States	2	0	Canada	59
2021-09-08	59	Canada	2	0	El Salvador	64
2021-09-08	74	Panama	1	3	Mexico	9
2021-09-08	44	Costa Rica	1	1	Jamaica	50
2021-09-08	63	Honduras	0	2	United States	10
2021-10-07	63	Honduras	1	1	Costa Rica	44
2021-10-07	10	United States	3	1	Jamaica	50
2021-10-07	9	Mexico	2	0	Canada	59
2021-10-07	64	El Salvador	1	0	Panama	74
2021-10-10	50	Jamaica	1	1	Canada	59
2021-10-10	74	Panama	1	2	United States	10
2021-10-10	44	Costa Rica	1	1	El Salvador	64
2021-10-10	9	Mexico	4	0	Honduras	63
2021-10-12	10	United States	2	0	Costa Rica	44
2021-10-13	59	Canada	2	0	Panama	74
2021-10-13	63	Honduras	1	1	Jamaica	50
2021-10-13	64	El Salvador	0	2	Mexico	9
2021-11-11	10	United States	1	1	Mexico	9
2021-11-12	63	Honduras	2	0	Panama	74
2021-11-12	64	El Salvador	1	1	Jamaica	50
2021-11-12	59	Canada	1	1	Costa Rica	44

Predicted Scores of CONCACAF World Cup Qualifiers

date	home_rank	home_team	home_score	away_score	away_team	away_rank
2021-11-16	50	Jamaica	1	1	United States	10
2021-11-16	74	Panama	1	1	El Salvador	64
2021-11-16	44	Costa Rica	2	1	Honduras	63
2021-11-16	59	Canada	0	2	Mexico	9
2022-01-26	50	Jamaica	0	2	Mexico	9
2022-01-26	63	Honduras	1	1	Canada	59
2022-01-26	10	United States	3	0	El Salvador	64
2022-01-26	44	Costa Rica	2	1	Panama	74
2022-01-29	63	Honduras	1	0	El Salvador	64
2022-01-29	74	Panama	1	2	Jamaica	50
2022-01-29	59	Canada	0	2	United States	10
2022-01-29	9	Mexico	2	0	Costa Rica	44
2022-02-01	50	Jamaica	1	1	Costa Rica	44
2022-02-01	10	United States	3	0	Honduras	63
2022-02-01	9	Mexico	4	0	Panama	74
2022-02-01	64	El Salvador	1	1	Canada	59
2022-03-23	50	Jamaica	1	0	El Salvador	64
2022-03-23	74	Panama	1	1	Honduras	63
2022-03-23	9	Mexico	1	1	United States	10
2022-03-23	44	Costa Rica	1	1	Canada	59
2022-03-26	10	United States	3	0	Panama	74
2022-03-26	63	Honduras	0	2	Mexico	9
2022-03-26	64	El Salvador	1	1	Costa Rica	44
2022-03-26	59	Canada	1	1	Jamaica	50
2022-03-29	9	Mexico	3	0	El Salvador	64
2022-03-29	50	Jamaica	2	0	Honduras	63
2022-03-29	74	Panama	1	1	Canada	59
2022-03-29	44	Costa Rica	1	1	United States	10

Comparing predictions vs. actual outcomes

- 6 correct home and away score predictions in 32 matches (18.75%)
- 17 correct predictions of outcome in 32 matches (53.13%)
- 12 correct predictions of home team score (37.50%)
- 11 correct predictions of away team score (34.38%)
- 12 correct predictions of total goals scored (37.50%)

Actual Outcome

Predicted Outcome

date	home_team	home_score	away_score	away_team	preds_home_team	preds_home_score	preds_away_score	preds_away_team	preds_date
2021-09-02	Canada	1.0	1.0	Honduras	Canada	2	0	Honduras	2021-09-02
2021-09-02	Panama	0.0	0.0	Costa Rica	Panama	1	1	Costa Rica	2021-09-02
2021-09-02	Mexico	2.0	1.0	Jamaica	Mexico	3	0	Jamaica	2021-09-02
2021-09-02	El Salvador	0.0	0.0	United States	El Salvador	0	2	United States	2021-09-02
2021-09-05	Jamaica	0.0	3.0	Panama	Jamaica	2	0	Panama	2021-09-05
2021-09-05	El Salvador	0.0	0.0	Honduras	El Salvador	2	0	Honduras	2021-09-05
2021-09-05	Costa Rica	0.0	1.0	Mexico	Costa Rica	1	2	Mexico	2021-09-05
2021-09-05	United States	1.0	1.0	Canada	United States	2	0	Canada	2021-09-05
2021-09-08	Canada	3.0	0.0	El Salvador	Canada	2	0	El Salvador	2021-09-08
2021-09-08	Panama	1.0	1.0	Mexico	Panama	1	3	Mexico	2021-09-08
2021-09-08	Costa Rica	1.0	1.0	Jamaica	Costa Rica	1	1	Jamaica	2021-09-08
2021-09-08	Honduras	1.0	4.0	United States	Honduras	0	2	United States	2021-09-08
2021-10-07	Honduras	0.0	0.0	Costa Rica	Honduras	1	1	Costa Rica	2021-10-07
2021-10-07	United States	2.0	0.0	Jamaica	United States	3	1	Jamaica	2021-10-07
2021-10-07	Mexico	1.0	1.0	Canada	Mexico	2	0	Canada	2021-10-07
2021-10-07	El Salvador	1.0	0.0	Panama	El Salvador	1	0	Panama	2021-10-07
2021-10-10	Jamaica	0.0	0.0	Canada	Jamaica	1	1	Canada	2021-10-10
2021-10-10	Panama	1.0	0.0	United States	Panama	1	2	United States	2021-10-10
2021-10-10	Costa Rica	2.0	1.0	El Salvador	Costa Rica	1	1	El Salvador	2021-10-10
2021-10-10	Mexico	3.0	0.0	Honduras	Mexico	4	0	Honduras	2021-10-10
2021-10-12	United States	2.0	1.0	Costa Rica	United States	2	0	Costa Rica	2021-10-12
2021-10-13	Canada	4.0	1.0	Panama	Canada	2	0	Panama	2021-10-13
2021-10-13	Honduras	0.0	2.0	Jamaica	Honduras	1	1	Jamaica	2021-10-13
2021-10-13	El Salvador	0.0	2.0	Mexico	El Salvador	0	2	Mexico	2021-10-13
2021-11-11	United States	2.0	0.0	Mexico	United States	1	1	Mexico	2021-11-11
2021-11-12	Honduras	2.0	3.0	Panama	Honduras	2	0	Panama	2021-11-12
2021-11-12	El Salvador	1.0	1.0	Jamaica	El Salvador	1	1	Jamaica	2021-11-12
2021-11-12	Canada	1.0	0.0	Costa Rica	Canada	1	1	Costa Rica	2021-11-12
2021-11-16	Jamaica	1.0	1.0	United States	Jamaica	1	1	United States	2021-11-16
2021-11-16	Panama	2.0	1.0	El Salvador	Panama	1	1	El Salvador	2021-11-16
2021-11-16	Costa Rica	2.0	1.0	Honduras	Costa Rica	2	1	Honduras	2021-11-16
2021-11-16	Canada	2.0	1.0	Mexico	Canada	0	2	Mexico	2021-11-16

Who will qualify for
the World Cup from
the CONCACAF
region?

Current Table

rank	team_name	matches_played	wins	draws	losses	goals_for	goals_against	goal_differential	points
1	Canada	8	4	4	0	13	5	8	16
2	USA	8	4	3	1	12	5	7	15
3	Mexico	8	4	2	2	11	7	4	14
4	Panama	8	4	2	2	11	9	2	14
5	Costa Rica	8	2	3	3	6	7	-1	9
6	Jamaica	8	1	4	3	6	10	-4	7
7	El Salvador	8	1	3	4	4	10	-6	6
8	Honduras	8	0	3	5	5	15	-10	3

Predicted Table

	team_name	matches_played	wins	draws	losses	goals_for	goals_against	goal_differential	points
rank									
1	Mexico	14	12	2	0	33	4	29	38
2	United States	14	10	4	0	28	6	22	34
3	Jamaica	14	4	7	3	15	16	-1	19
4	Canada	14	3	7	4	13	15	-2	16
5	Costa Rica	14	2	9	3	14	17	-3	15
6	El Salvador	14	2	5	7	8	19	-11	11
7	Honduras	14	2	4	8	8	23	-15	10
8	Panama	14	0	4	10	8	27	-19	4

Conclusions, Recommendations, and Next Steps

- With the limited amount of games being predicted on, low scores of the models, and limited amount of data for these international matches, obtaining more information on match statistics and a different approach at predicting soccer events may help predict scores better.
- Given more time, I would also like to test using a Random Forest Regressor for predictions to see if that would improve predictions.
- Finally, I may look into predicting the features of home/away goals for/against averages for future games to then predict the scores of those games instead of using the average prior to those predictions.

Thank you!!!

And GO USA!!

