

Predicting UEFA Euro 2020 results

Kristo Markov, Pelle Holden Porila



BACKGROUND

Euro is the primary association football competition contested by the senior men's national teams of the members of Union of European Football Associations (UEFA), determining the continental champion of Europe.



AIM



Our aim is to try and predict so many games correctly as possible. For example we would be happy if we could predict the outcome of 15 games out of 51 games. If we predict the game winner and loser or if the game ends in a draw then we have been successful.

DATA



Data contains variables such as: date (when did the game take place), home team, away team, home score (how many goals did home team score), away score (how many goals did away team score), tournament, city, country, neutral.

We used Margus Jürisoo's dataset „International football results from 1872 to 2020“ from kaggle. From the dataset we used games which were played 2000-2020.

BUILDING THE MODEL



Firstly we gave every game a weight. We used a formula which took into account on what year the game took place.

Then we created the model data and implemented Poisson regression on it.

RESULTS



For first time we predicted the tournament outcome using all the games between the participating countries from 2000 until 2020. Out of 1000 tournaments most wins got Spain(656) most goals scored Spain(8981), most goals conceded North Macedonia (4905). Average goals per game was 1.38.

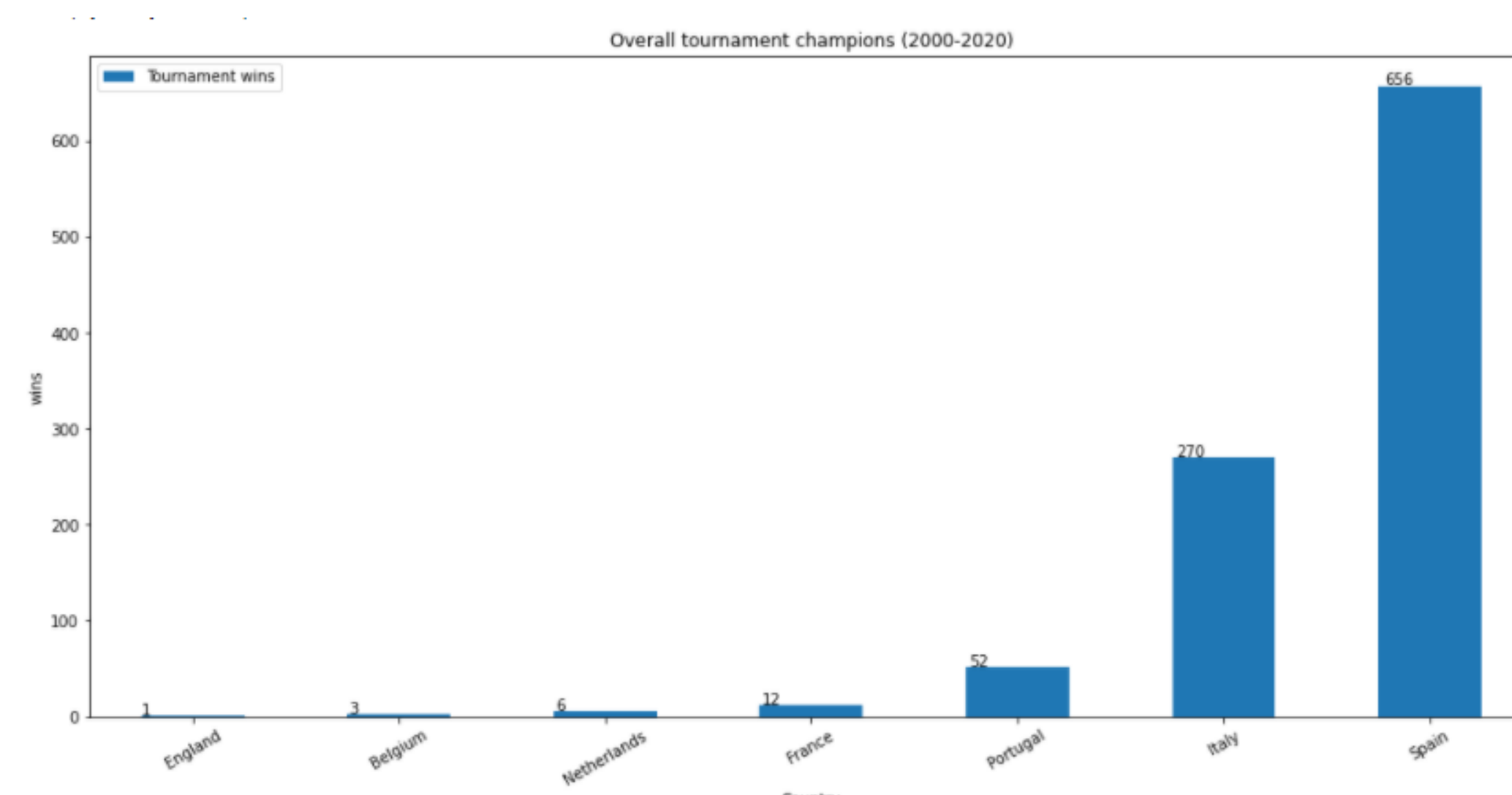


Figure 1. Overall tournament champions (2000-2020). Who won how many tournaments out of 1000

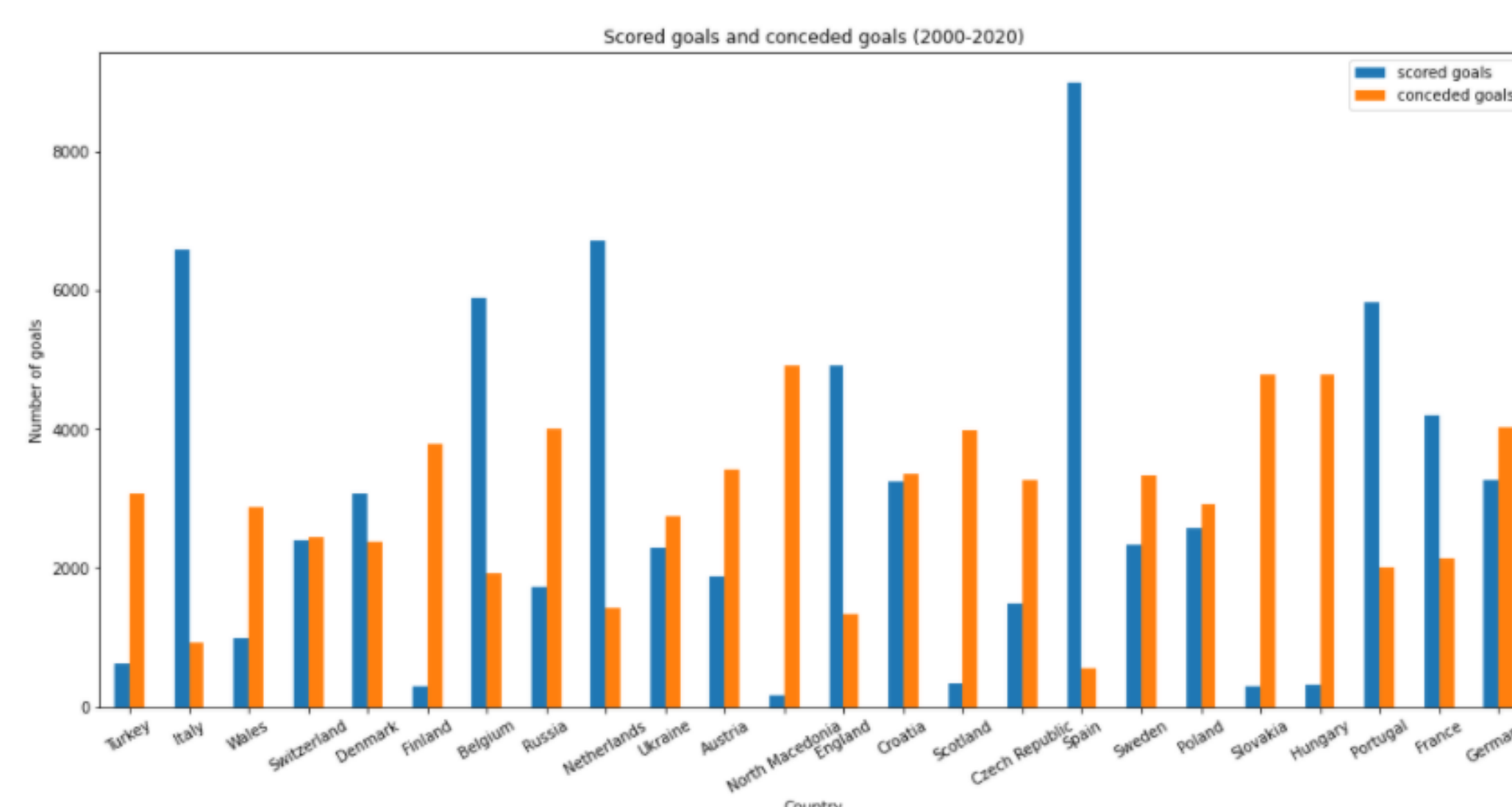


Figure 2. Scored and conceded goals (2000-2020). Who scored and conceded how many goals during 1000 tournaments playthrough.

The second time we used games from 2010-2020. Out of 1000 tournaments most wins got Spain(674), most goals scored Spain(9348), most goals conceded North Macedonia (5007). Average goals per game was 1.43.

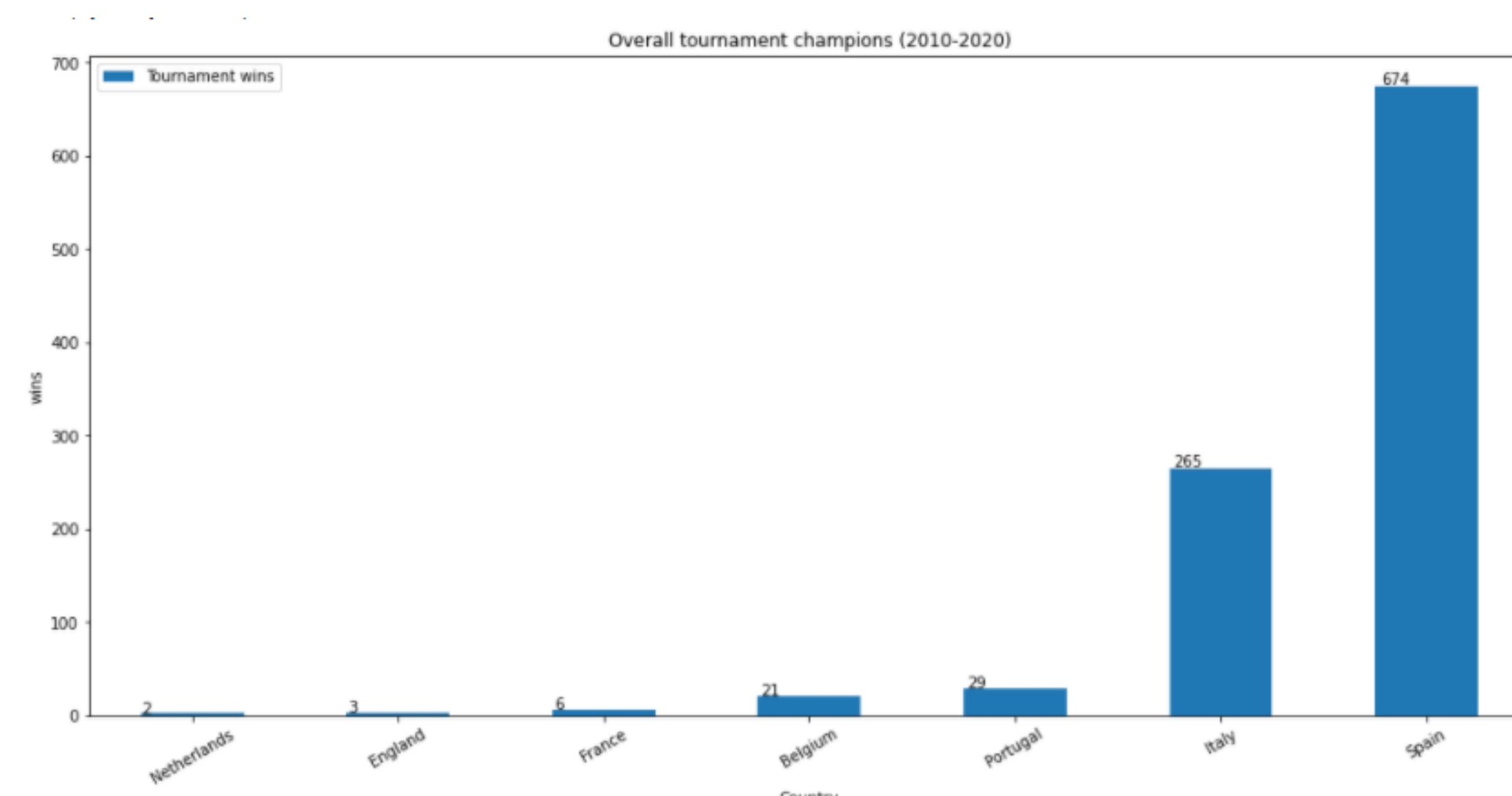


Figure 3. Overall tournament champions (2010-2020). Who won how many tournaments out of 1000

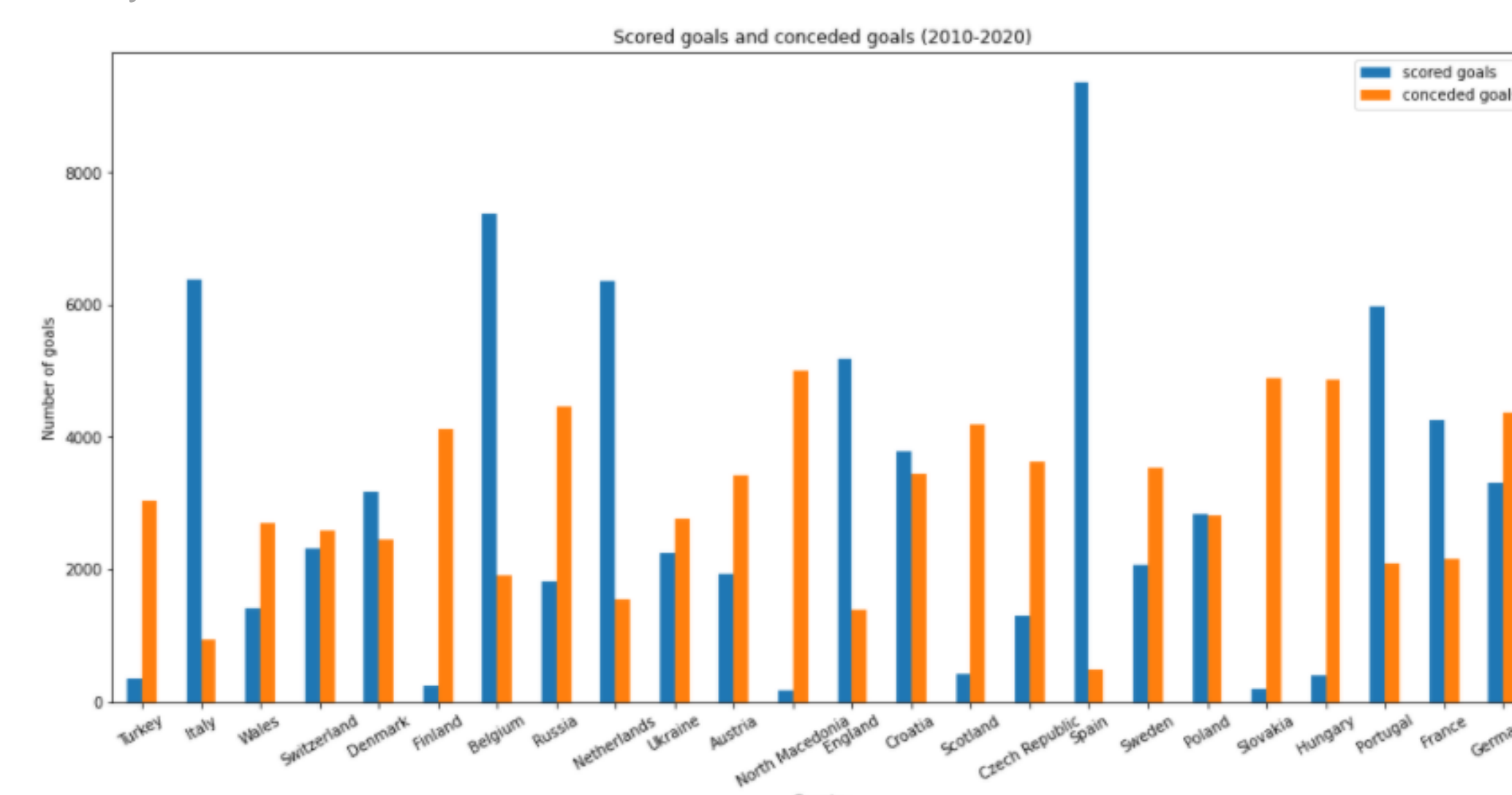


Figure 4. Scored and conceded goals (2010-2020). Who scored and conceded how many goals during 1000 tournaments playthrough.

CONCLUSION

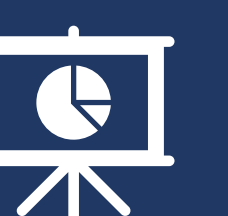


In conclusion our model predicted both times that Spain has the biggest chance to win the tournament, which was a little unexpected that they won with such a big lead because Spain hasn't been at the top of the FIFA leaderboard for a long time.

But more surprising was that Italy still won 265 tournaments out of 1000 when considering the games 2010-2020 because these years they haven't performed well according to FIFA rankings. France tournament victories between 2000-2020 were so much higher because France football so called „Golden generation“ played in the years 2000-2006.

The exact same story applies to Portugal. Belgium results have been a lot better between 2010-2020 than 2000-2020, it's because Belgium have been in the top 5 of FIFA leaderboard for the past 6 years. and before that they didn't have much success. Overall Spain isn't the favorite of this tournament but them winning this is very realistic.

EXAMPLE TOURNAMENT



	points	diff	goals
Italy	9	3	3
Turkey	6	1	2
Switzerland	3	-1	1
Wales	0	-3	0

	points	diff	goals
Belgium	9	5	6
Denmark	6	0	2
Russia	3	-1	2
Finland	0	-4	0

	points	diff	goals
Netherlands	9	6	7
Ukraine	6	0	2
Austria	3	-1	2
North Macedonia	0	-5	0

	points	diff	goals
England	9	4	4
Croatia	6	0	2
Czech Republic	3	-1	1
Scotland	0	-3	0

	points	diff	goals
Spain	9	4	4
Sweden	6	1	2
Poland	3	-1	1
Slovakia	0	-4	0

	points	diff	goals
Portugal	9	3	4
Germany	6	3	4
France	3	0	2
Hungary	0	-6	1

Figure 5. Possible outcome of tournament group stage.

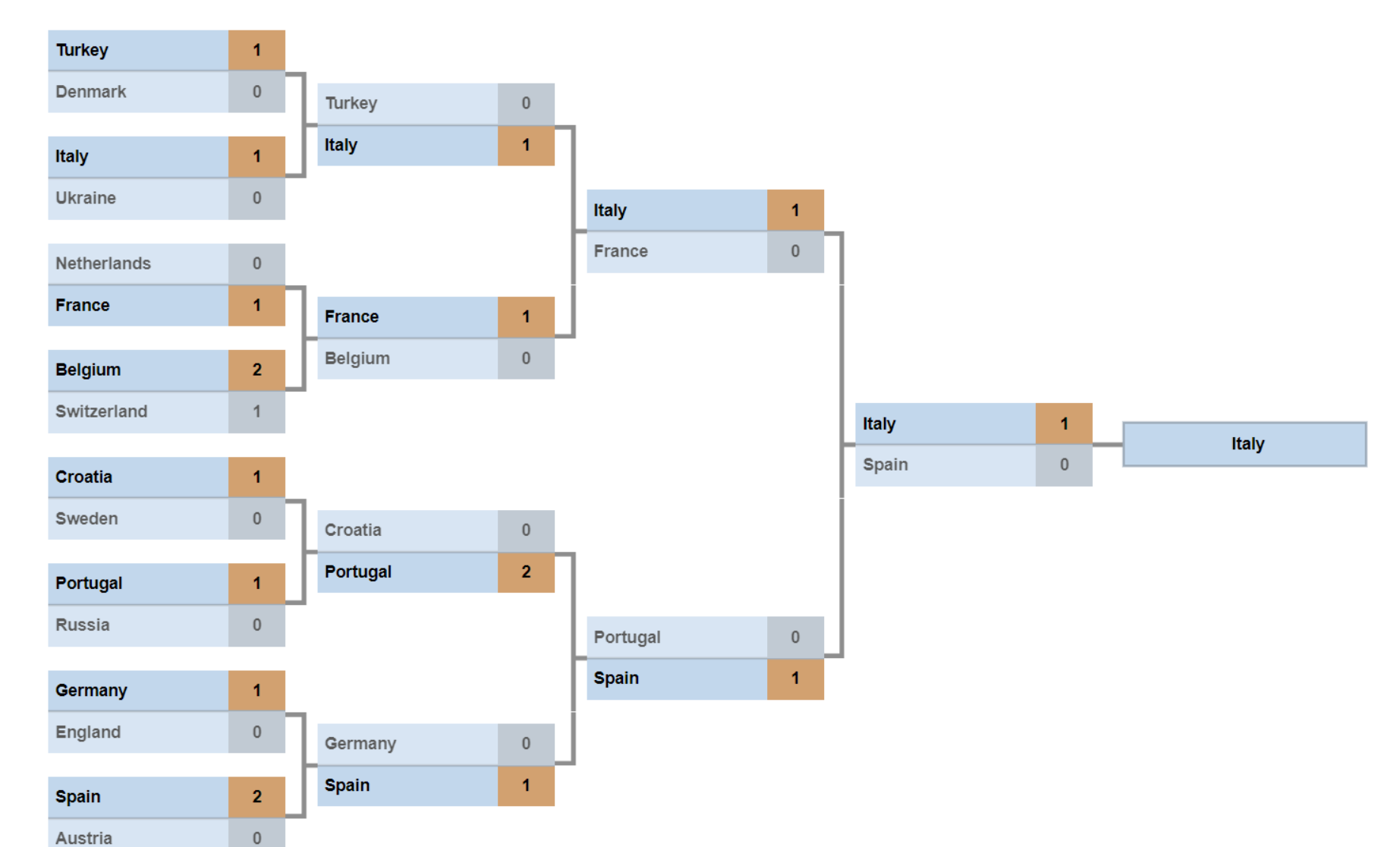


Figure 6. Possible outcome of knockout stage