

Project: World Cup 2022

Ken Dizon
MSDS670 - Data Visualization
October 15th, 2022



[GITHUB](#) | [YouTube](#)



Final Project



CONTENT

- Research Question
 - Data
 - Methodology
 - Results
 - Conclusion
-

Research Question

Sports and Data

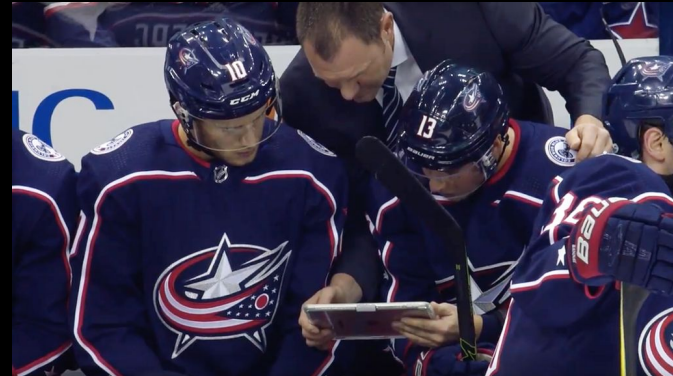
Most effective method of winning a soccer game?

- Goals win games
- Defensive is the best form of attack

Why the FIFA World Cup 2022?

Which teams have the highest chance to win?

- Popularity
- Four year time intervals



DATA

International football results from 1872 to 2022

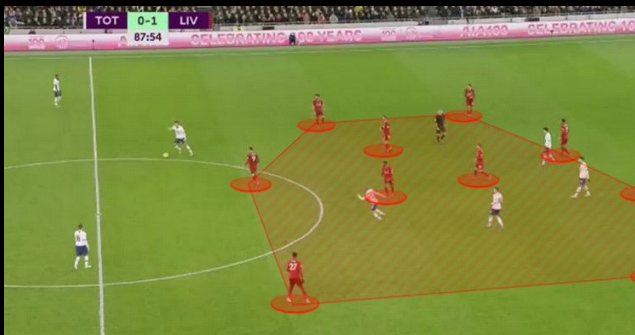
An up-to-date dataset of over 40,000 international football results



- Columns: 9 | Rows: 43,000
 - Terminology
 - Date, teams, score, venue, etc.
 - Sports specific
 - results , metrics per game, percentages, etc.
- Processing
 - Python: Data Preparation
 - Tableau: Data Visualization GUI

Methodology

- Time interval: > 2018 World Cup (July 2018)
- Categorical: group stage, region
- Numeric: results, scoreline, other
- Shape: 32 x 18



	date	home_team	away_team	home_score	away_score	tournament	city	country	neutral
0	1872-11-30	Scotland	England	0	0	Friendly	Glasgow	Scotland	False
1	1873-03-08	England	Scotland	4	2	Friendly	London	England	False
2	1874-03-07	Scotland	England	2	1	Friendly	Glasgow	Scotland	False
3	1875-03-06	England	Scotland	2	2	Friendly	London	England	False
4	1876-03-04	Scotland	England	3	0	Friendly	Glasgow	Scotland	False

Date Ranges

	date	home_team	away_team	home_score	away_score	tournament	city	country	neutral
40138	2018-07-22	Argentina	Croatia	2	1	World Cup	Madrid	Spain	True
40139	2018-08-04	France	Uruguay	2	0	World Cup	Madrid	Spain	True
40140	2018-08-04	France	Peru	1	0	World Cup	Madrid	Spain	True
40141	2018-08-15	France	Senegal	2	1	World Cup	Madrid	Spain	True
40142	2018-08-18	France	Poland	1	0	World Cup	Madrid	Spain	True
...
43729	2022-06-14	Argentina	Croatia	2	1	World Cup	Madrid	Spain	True
43730	2022-06-14	France	Uruguay	2	0	World Cup	Madrid	Spain	True
43731	2022-06-14	France	Peru	1	0	World Cup	Madrid	Spain	True
43725	2022-06-14	France	Senegal	2	1	World Cup	Madrid	Spain	True
43751	2022-06-14	France	Poland	1	0	World Cup	Madrid	Spain	True

3614 rows x 9 columns

- Split data
- Selected teams
- Dictionary/Mapping
- Define results
- Aggregation

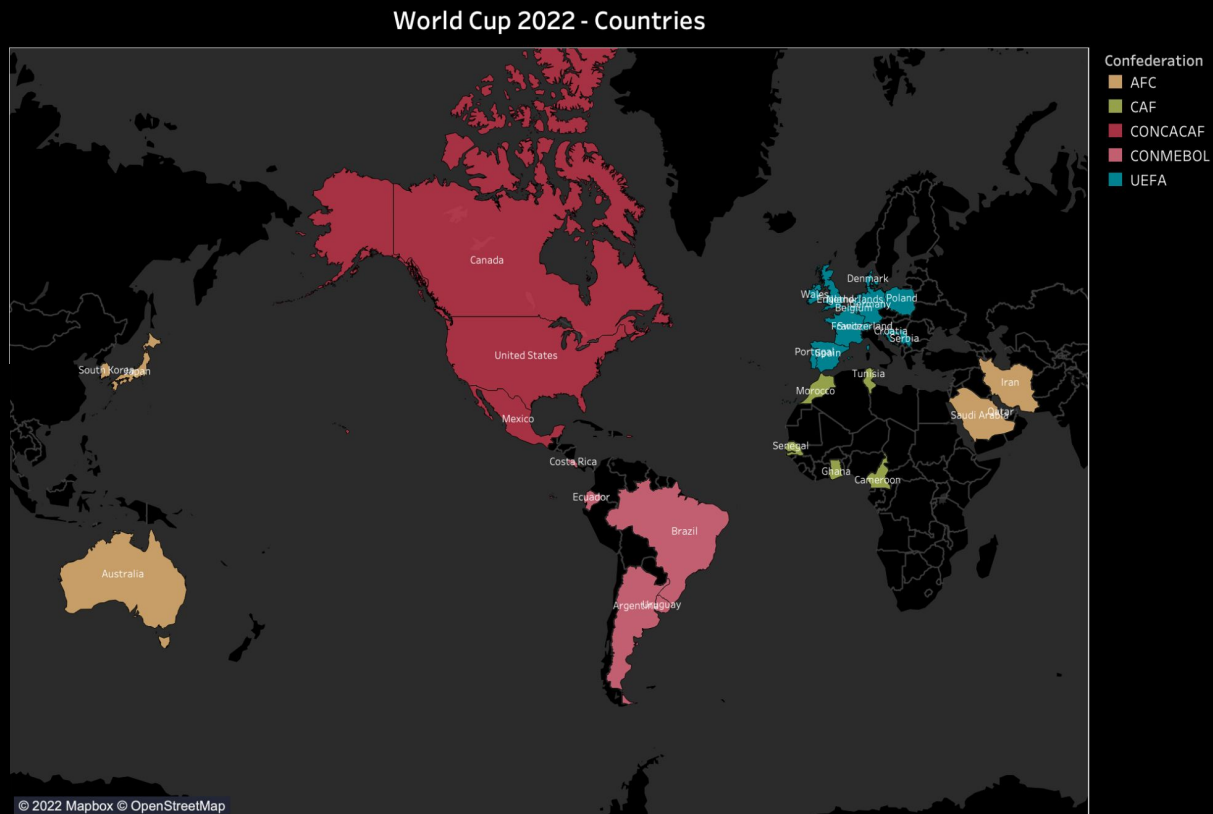
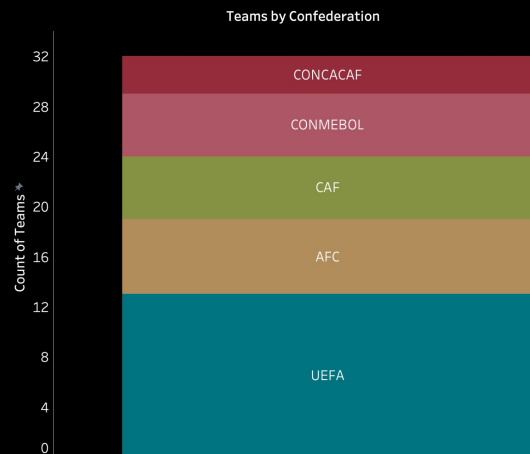
'home df'							
	team	games_played	W	D	L	goals_for	goals_against
0	Argentina	29	18	8	3	56	16
1	Australia	16	10	4	2	35	9
2	Belgium	25	19	3	3	76	22

'away df'							
	team	games_played	W	D	L	goals_for	goals_against
0	Argentina	18	12	5	1	31	11
1	Australia	14	8	2	4	29	9
2	Belgium	22	15	4	3	50	19

group	team	confederation	games_played	W	D	L	goals_for	goals_against	goals_diff	goals_scoring_ratio	goals_conceded_ratio	goals_per_game	goals_per_game_against	goals_per_game_diff	goals_per_game_ratio	goals_per_game_ratio_diff	goals_per_game_ratio_diff_ratio	
5	F	Canada	CONCACAF	42	29	4	9	120	29	91	4.14	2.86	0.69	2.17	0.85	0.69	-0.16	0.90
14	B	Iran	AFC	36	26	5	5	83	19	64	4.37	2.31	0.53	1.78	0.81	0.72	-0.09	0.86
2	F	Belgium	UEFA	47	34	7	6	126	41	85	3.07	2.68	0.87	1.81	0.76	0.72	-0.04	0.85
3	G	Brazil	CONMEBOL	48	35	10	3	103	16	85	5.72	3.16	0.36	1.77	0.76	0.73	-0.03	0.79

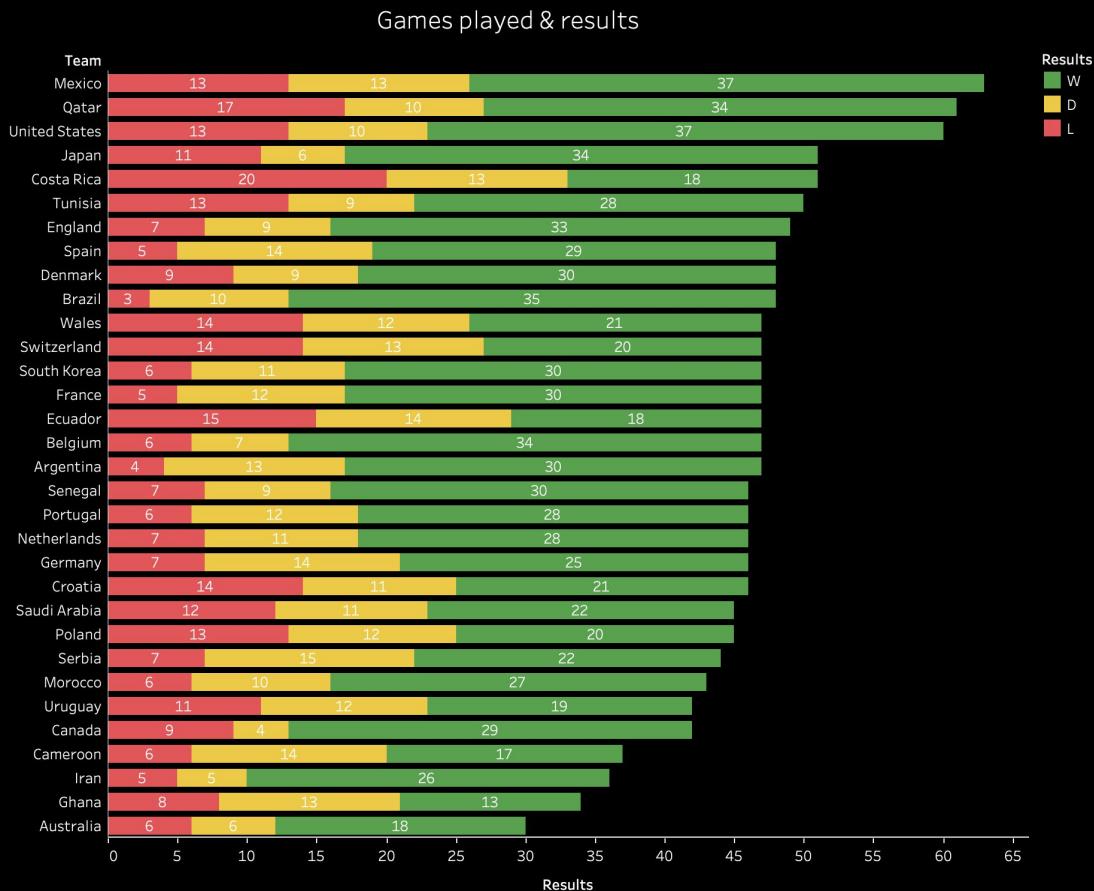
RESULTS: Teams

- Categorical
- Geospatial
- Qualification
- Scale



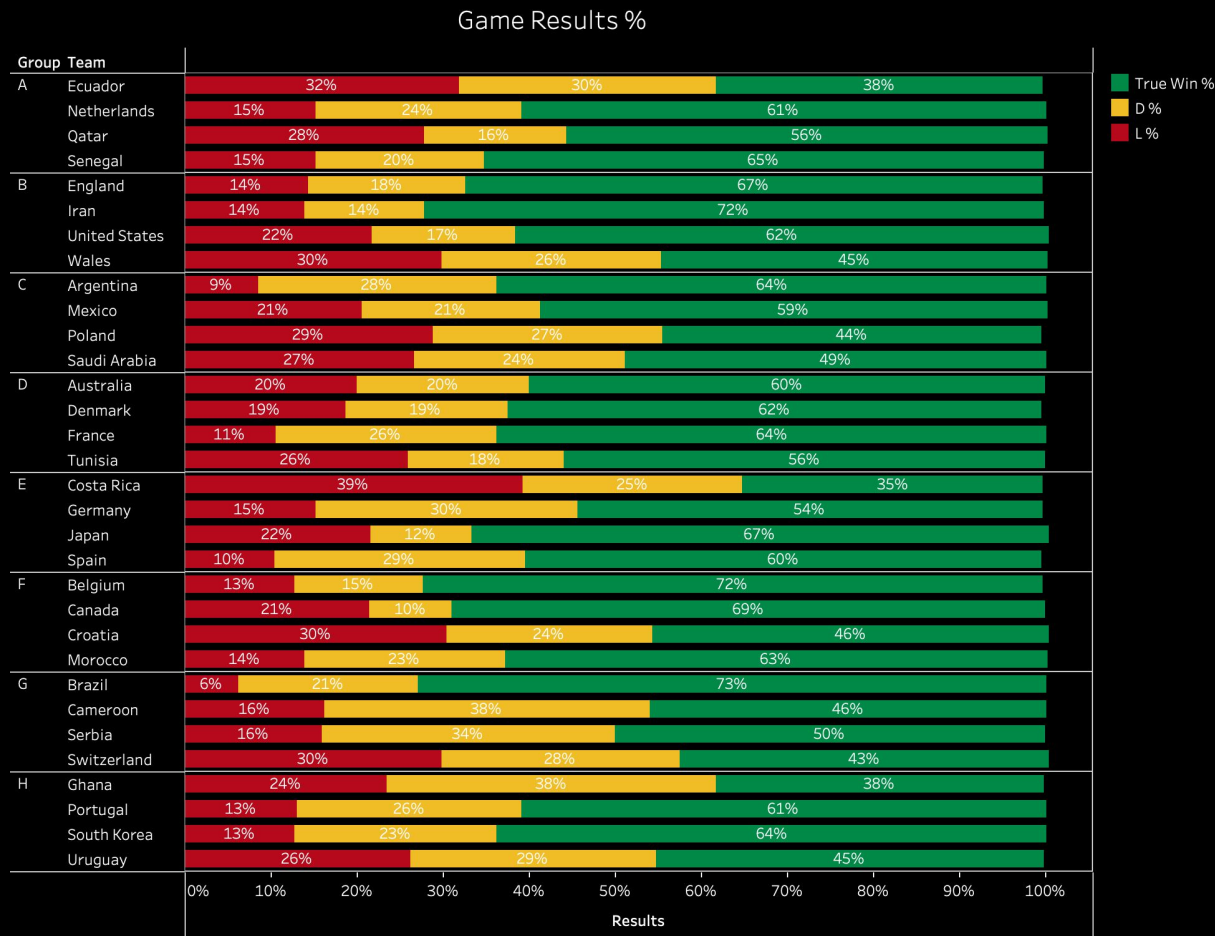
RESULTS: Fixtures

- Stacked bar chart
 - GP
 - W/D/L
- Game counts
 - Lowest < 40
 - Highest > 50
- Quantity of Games
- Quality of Results



RESULTS: Fixtures

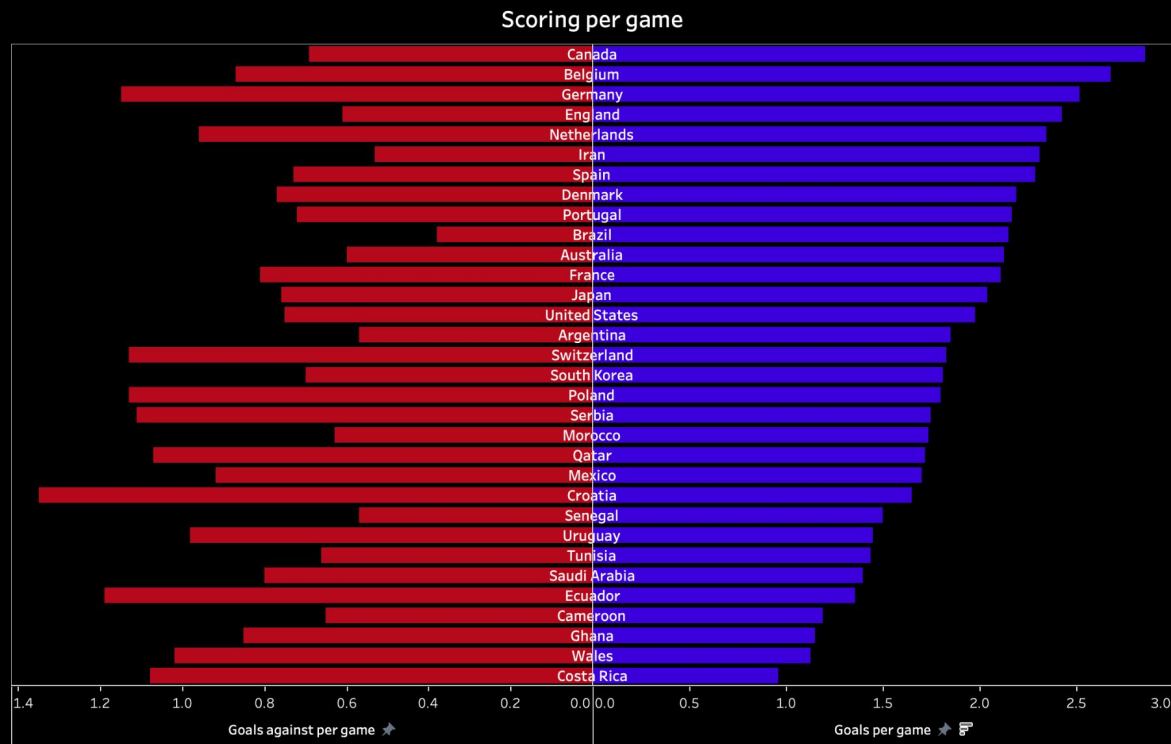
- 100 % Stacked bar chart
 - GP
 - W/D/L
- Proportions
- Three year outputs



RESULTS: Difference

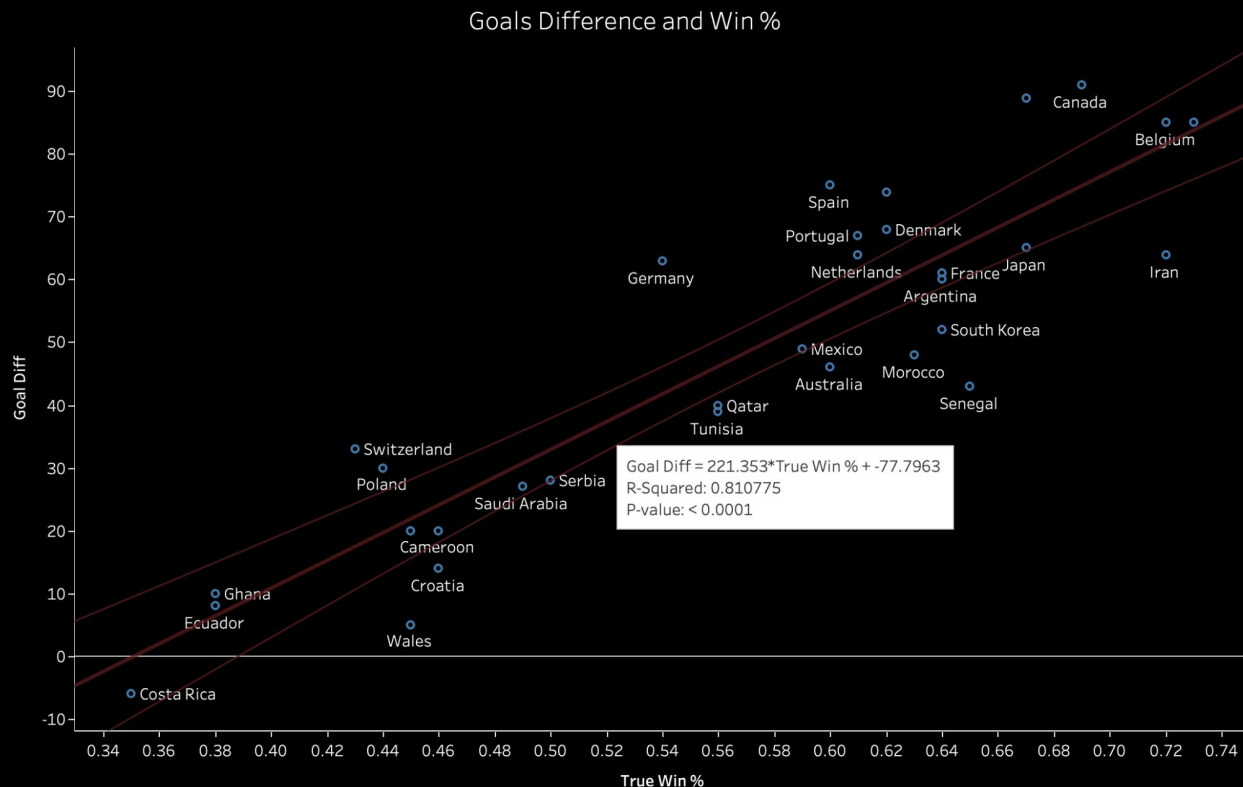
- Butterfly Bar Chart
 - Goals for
 - Goals against
 - Per game

- Attacking
- Defensive



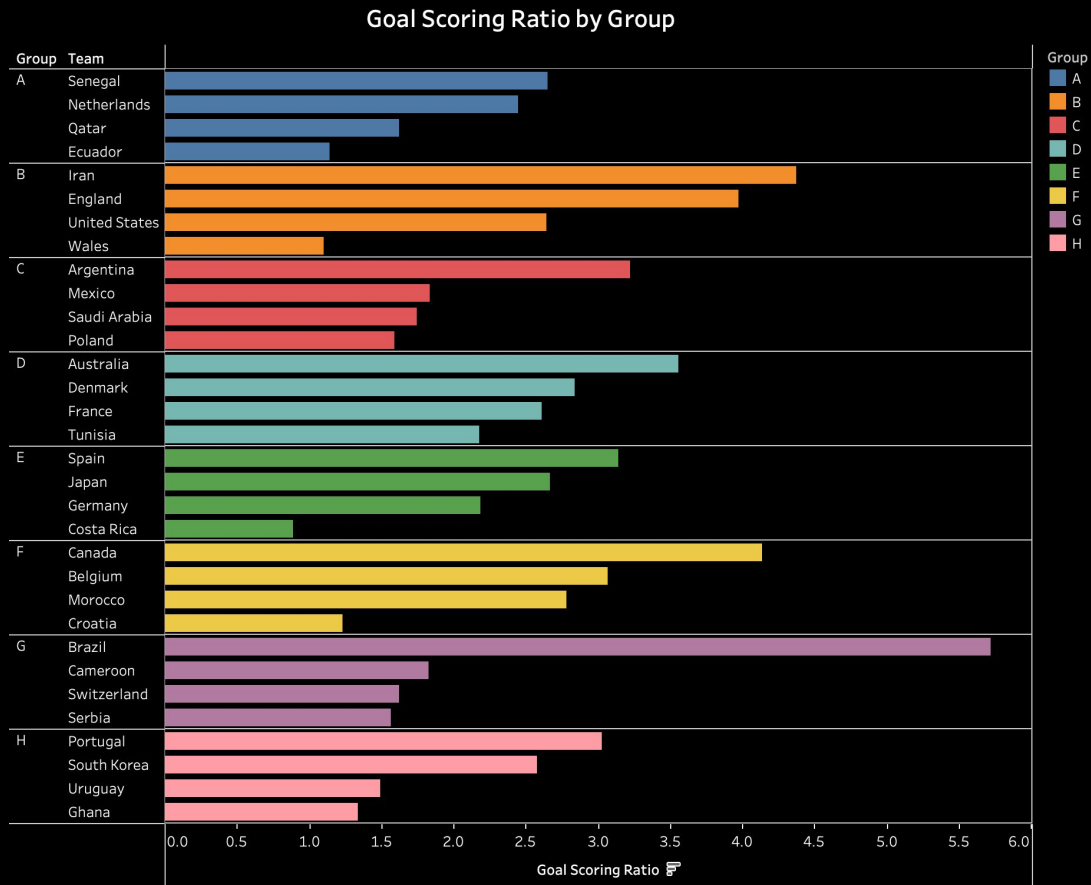
RESULTS: Difference

- Scatterplot
 - Relationship
 - Goal difference
 - Actual win %
- Linear regression
 - Strong correlation
 - Confidence intervals



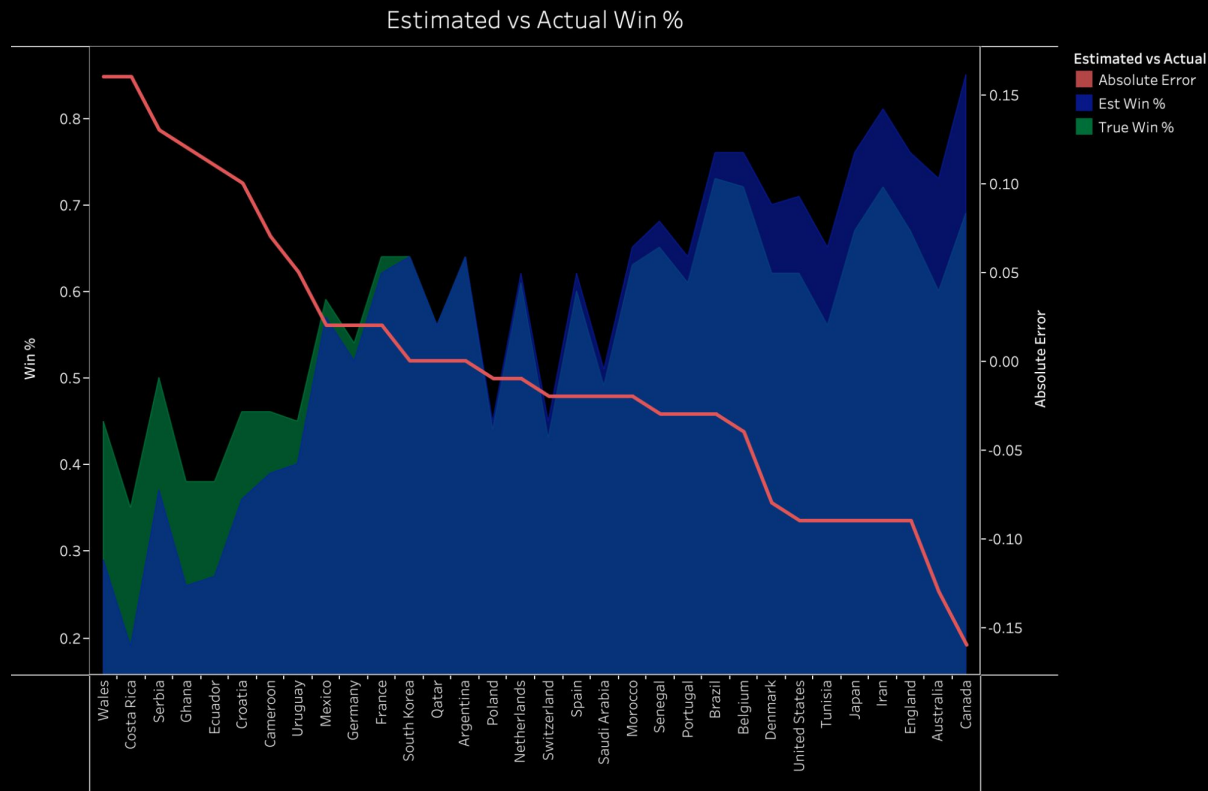
RESULTS: Goal Scoring

- Unstacked bar chart by group
 - Goal scoring ratio
 - Top 2 teams



RESULTS

- Dual axis
 - Area Graph
 - Estimated vs True
 - Line Graph
 - Absolute error
- Performance gauge
 - Under vs Over



CONCLUSION

- Low: Ghana Costa Rica, Ecuador
- Surprise: Iran
- Favorites: Brazil, Belgium

