Summary

- The goal is to expand the soccer literature to simulate the FIFA tournament and group phases. This entails a holistic approach that starts from the qualifiers and ends at the World Cup final.
- Individual matches are predicted by a Poisson model, which is a model that fits well for rare events.
- A concrete final result is the probability of each team moving through the rounds in the 2022 World Cup. Some favorite teams already show up in the ranking!
What would a holistic approach look like?

1. The qualifiers define the 32 teams that are going to enter the tournament
2. A couple of months before the tournament FIFA draws the 8 groups (A - H). For Qatar, that will be in April 2022
3. The tournament takes place in 6 rounds

The goal is then to select the 32 qualified teams and to simulate the World Cup 10,000 times starting from the group draw.
Qualifiers and Group Draw

**Qualifiers**

Goal: Get the 32 teams that are going to qualify for the World Cup

- AFC (4)
- CONMEBOL (6)
- CAF (5)
- CONCACAF (4)
- UEFA (13)

**Group Draw**

Create 4 Pots
(Not random: FIFA 2021 ranked)

Create 8 Groups
Random

- Pot 1
- Pot 2
- Pot 3
- Pot 4
- Group A
- Group B
- Group C
- Group D
- Group E
- Group F
- Group G
- Group H
The tournament is simulated for each randomized group draw. This allows us to assess what is the probability of each team reaching each round regardless of the group they fall into in April.

### Tournament Stages

<table>
<thead>
<tr>
<th>Stage</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group Stage</td>
<td>Teams play each other within the group</td>
</tr>
<tr>
<td>Round 16</td>
<td>Winner of a group plays runner up of following group</td>
</tr>
<tr>
<td>Quarter-finals</td>
<td>Winner plays winner of following group</td>
</tr>
<tr>
<td>Semi-finals</td>
<td>Winner plays winner of following group</td>
</tr>
<tr>
<td>Final</td>
<td>Final match</td>
</tr>
</tbody>
</table>

Each match:

- **Poisson Regression for individual matches**
  - For each match, predict the number of goals for each team and check who won.
Goals in individual matches are well modeled by Poisson distributions. Poisson distributions are frequently used to represent rare events, they are similar to binomial distributions for a large number of trials with a small probability of success.

Trained with FIFA match data from 1960 to 2018.

\[ y_{ijkt} \sim \text{Pois}(\lambda_{ijkt}) \]

\[ \ln(\lambda_{ijkt}) = \beta_0 + \beta_1 \text{CupYear} + \beta_2 (\text{Elo}_{i,t-1} - \text{Elo}_{j,t-1}) + \beta_3 \text{AvgGoalsScored}_{i,t-1} + \beta_4 \text{AvgGoalsTaken}_{j,t-1} + \beta_5 \text{StageNum}_{kt} + \beta_6 \text{HomeAdvant}_{ijkt} \]
The Poisson regression predicted the result of 70% of the matches correctly for the 2014 World Cup.
Preliminary Results for 2022

<table>
<thead>
<tr>
<th>Teams</th>
<th>Round of 16</th>
<th>Quarter finals</th>
<th>Semi finals</th>
<th>Final</th>
<th>Champion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belgium</td>
<td>100%</td>
<td>70%</td>
<td>60%</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>France</td>
<td>100%</td>
<td>50%</td>
<td>40%</td>
<td>30%</td>
<td>0%</td>
</tr>
<tr>
<td>Portugal</td>
<td>100%</td>
<td>40%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Spain</td>
<td>90%</td>
<td>40%</td>
<td>20%</td>
<td>10%</td>
<td>10%</td>
</tr>
<tr>
<td>Brazil</td>
<td>90%</td>
<td>90%</td>
<td>50%</td>
<td>10%</td>
<td>0%</td>
</tr>
<tr>
<td>Argentina</td>
<td>90%</td>
<td>20%</td>
<td>10%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Italy</td>
<td>80%</td>
<td>30%</td>
<td>20%</td>
<td>10%</td>
<td>0%</td>
</tr>
<tr>
<td>Denmark</td>
<td>80%</td>
<td>60%</td>
<td>40%</td>
<td>20%</td>
<td>20%</td>
</tr>
<tr>
<td>Netherlands</td>
<td>80%</td>
<td>40%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Uruguay</td>
<td>70%</td>
<td>50%</td>
<td>20%</td>
<td>10%</td>
<td>0%</td>
</tr>
<tr>
<td>Mexico</td>
<td>70%</td>
<td>30%</td>
<td>30%</td>
<td>30%</td>
<td>20%</td>
</tr>
<tr>
<td>England</td>
<td>60%</td>
<td>40%</td>
<td>30%</td>
<td>10%</td>
<td>0%</td>
</tr>
<tr>
<td>Germany</td>
<td>60%</td>
<td>20%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Colombia</td>
<td>60%</td>
<td>10%</td>
<td>10%</td>
<td>0%</td>
<td>0%</td>
</tr>
</tbody>
</table>

- 10 first simulations
- Each line represents the number of times the respective country played in the round
- **Belgium** seems to be one of the favorites for 2022! Indeed, as of 2021, it had the second highest Elo Score
Conclusion and Next Steps

● The project is broken down into two steps:
  ○ Monte Carlo that simulates the World Cup from qualifiers to tournament
  ○ Poisson model that predicts individual matches using historical FIFA data

Next Steps:
● Run the simulation for more rounds (computationally intensive)
● Check the relationship between Elo scores and being qualified for the World Cup