**Project Title**: FIFA World Cup 2018 Championship Predictor

**Team Members**: Long Ngo, Connor Ladd, Ramin Jabbari, Jason Douangsanith

**Project Description/Outline**

To research historical and current data and determine which country will most likely advance to the next rounds and who will most likely win the 2018 FIFA World Cup championship.

**Research Questions to Answer**

1. Who is likely to win the 2018 FIFA World Cup?
   1. How does social media sentiment predict the outcome/results of the tournament?
   2. How does historical data/statistics predict the outcome/results of the tournament?
   3. How does current team/player data predict the outcome/results of the tournament?
2. How does our data prediction closely match with the gambling spreads/odds?
3. Which teams will most likely advance from the group level and quarter/semi/finals?
4. Who is most likely to receive the Golden Boot?
5. Does a country’s economic indicators factor in the prediction of who will likely win the tournament?
6. Analyze 2014 World Cup Data

**Data Sets to be Used**

* Kaggle
  + <https://www.kaggle.com/agostontorok/soccer-world-cup-2018-winner/notebook>
* <https://www.jokecamp.com/blog/guide-to-football-and-soccer-data-and-apis/>
* FIFA website
* FiveThirtyEight
* Twitter/Instagram
* Soccerstats.com
* <https://www.fifa.com/worldcup/news/take-the-fifa-world-cup-bracket-challenge>
* Sklearn
* <https://github.com/openfootball/world-cup>
* <https://www.oddschecker.com/football/world-cup/winner>
* <https://en.wikipedia.org/wiki/List_of_FIFA_World_Cup_finals>

**Rough Breakdown of Tasks**

|  |  |  |
| --- | --- | --- |
| Task | Assignment | Notes |
| Research historical data of previous world cup winners and determine which stats correlate with who is more likely to win the championship. | Ramin | South American and European countries have only won the previous world cups |
| Research current player/team data to determine who is likely to advance to the next rounds | Jason |  |
| Observe social media trends/discussion to see which teams are discussed positively/negatively | Long | Disclosure: limited to English  Celebrities vs. General Public vs. Sports Analyst |
| Determine which statistical variables we will use for prediction  Simulate Monte Carlo |  | Shots on goal; star players; home field advantage; passes; penalties; offsides; red/yellow cards/ejections; club team records/championships; age; how many times a country has appeared in a world cup; goalie performance; top defenders; highest scoring teams  Correlation between winner of the group of death to advancing |
| Validate our data predictions against current analysis from experts |  |  |
| Research country’s economic indicators and correlation with winning |  | GDP; population; sports popularity; strongest league; which continents bring out the most winners |
| Data Visualizations |  |  |
| Summary/Predictions |  |  |