# DS311
# Group Project Documentation
# FIFA World Cup Data Set Analysis Plan

The FIFA World Cup is the most prestigious football tournament in the world. The championship has been awarded every four years since the start of the tournament in 1930. The current format involves a qualification phase, which takes place over the preceding three years, to determine which teams' quality for the tournament. In the tournament, 32 teams, including the host nation, compete for the title at different stadiums in the host country.

The reigning champion is France, which beat Croatia in the 2018 tournament in Russia. Qatar will host the 2022 tournament, for which the first match will be played in November. This dataset provides a complete overview of all international soccer matches played since the 90s. On top of that, the strength of each team is provided by incorporating actual FIFA rankings as well as player strengths based on the EA Sport FIFA video game.

Here are some interesting questions you may want to answer with this dataset.

I. Can you predict what team is most likely to win the 2022 FIFA World Cup?
   - What are the key variables explained which team has higher chance winning?
   - Can we calculate the expected winning rate for each team based on the history?

II. What team has the strongest defense, midfield, and offense players?
   - Can we quantify those statistics?
   - How we define strongest?

III. Is there really such a thing as a home team advantage?
   - Do we find any evidence of home team advantage?

IV. Do teams with stronger offense players score more goals? And do teams with stronger goalkeepers receive fewer goals?
   - Use the data to support your hypothesis.

V. What team has the longest winning streak?
   - Visualize the result.

VI. Does the best team always win? Can you explain why a weaker team sometimes win?
   - What are the reasonings that could possibly explain your findings?

You can choose three of the above questions to answer in your project. The second part of the project is to come up with three other questions and answer them in the similar fashion.

For each question your team is answering in this project, the answers must be supported by the data, which can be tables or graphs. However, you should be aware that you never have a chance in the real-world presentation to flash out 15 plots and 20 tables because no stakeholders will wait for your explanations to all the plots and tables. Be wise on selecting the right presenting material and make sure telling a story from each of them.