Group Project Report

Chosen Dataset: Option 3 FIFA World Cup

-We chose this set as our group agreed to work on the sports data mainly because of our general interest in FIFA.

-Due to the next World Cup for the first time being held in 3 different countries it will be interesting to obtain insights

-The program we used to do a preliminary search of the dataset after obtaining it from the repository was R. R works as a great program to process data and stage preliminary checks to get a proper feel of the dataset.

-Here is some basic information about the FIFA dataset

- 23921 observations
- 25 Variables
- The data starts from the year 1993 and goes up until 2022 giving us the most recent data as the World Cup is held every four years
- The data set columns are very in-depth taking in scores of home versus away teams in terms of defense offense midfield and goalkeeper.

Some interesting things were observed about this dataset.

- An overall trend graph for goals was created showing the total goals over the years showing a steady increase of goals being scored until the 2000s hit where we can observe a steady fluctuation in goals
- The peaks in the generated chart are the years 2002 2006 and 2018
- The lowest drop include 1993 and 2020 oddly enough
- We noticed that 2020 was included or at least found from the dataset which leads us to believe qualifier matches are also included in the dataset.
- Brazil ends up being the team with the highest average FIFA rank we could explore that as a potential correlation

-Here is the generated graph for reference

Total Goals Over the Years