DSCI 235 - Project Proposal

By: Blake Davis and Elisabeth Ball

Topic Description and Interests:

For this project, we are analyzing data from the Tokyo 2020 Olympics and the Rio 2016 Olympics. The Olympic Games are a fascinating event, bringing the world together to appreciate excellence in athletics. The numerous different events, countries, and people competing in the Olympic Games give ample material to analyze. Finding patterns and interesting information in this Olympic data will likely change the way we see the games in the future and make us think about the nature of sports in a new way. This project will also be challenging and help us develop our teamwork, communication, and data wrangling skills.

Describing the Data set:

We are using 2 .csv files for our project. We have a main data set that is for the 2020 Summer Olympics in Tokyo. It includes data sorted by athlete name, with information on each athlete including age, country, sport, event, rank, and medal earned. The data set has multiple entries for each athlete as many athletes compete in multiple events. We have not decided how we are going to sort/filter the data for each individual athlete as they are each in the data set multiple times. We will use a 2016 Olympics in Rio De Janeiro data set to be able to determine if there were any repeat medalists and if there were more female or male medalists in 2016 vs. 2020.

Questions we want to answer:

1. What is the average age of an Olympic athlete in 2016? 2020? Compare.
2. What are the top medaling countries in 2016? 2020? Compare.
3. Does the United States have more female or male medalists in 2020?
4. Which Olympic athlete has the most medals in 2016? 2020?
    - Is it the same person each Olympics? Same event?
5. Who are the 5 oldest athletes and who are the 5 youngest athletes?
    - What country are they from?

- What sport do they compete in? What is the age distribution of their sport in 2020?

Clear Steps for Project:

The first step for the project will be to build a space for it to live on Github. Next, we will load the data and format it nicely for our purposes. Then we will work on each question one at a time, working to find meaningful answers and documenting as we go. Finally, we will write up a summary to include in the report. We will also create a video to show off our hard work.

Timeline:

We will work on this together at a planned time each week and finish the bulk of the project one week before it is due to ensure appropriate editing and clarifying adjustments can be made. There are 5 major questions/sections above that we want to answer, we want to finish at least 1 question/section per week, giving us plenty of time to write the summary and create the report.

Team Responsibilities:

We will meet once a week to work on the project so we make consistent progress each week.

- Consistent and Open Communication between Teammates