Data Source - External

Data collected through registration for and placement in competitions based on a format of how many tops and zones are reached by the competitor.

Summary
This data set is rock climbing results from the IFSC competitions over the two years of 2018 and 2019.

Why I chose this dataset
As a climber, data about climbing is interesting. I have been watching these competitions since 2018 and while there are patterns, there is a lot of chaos in who makes final rounds and which countries are most represented in the top climbers for various reasons including that each competition has different climbs. There is often speculation of rising stars in the climbing world but many never come to fruition or are often short lived and there are always newcomers on the scene. Getting a glimpse into trends of countries, climbers, and competitions will be a great exercise.

Something to note is that the year 2019 was the year of olympic qualifications. I think this changed the competition as well as the type of climbing athletes trained and participated in.

Limitations or Ethical Considerations
The main limitation is that the data is not the most recent data and that it only spans over 2 years. It also lacks some statistical data that would be interesting to explore such as climbers age, height, and or ape index. The names are public knowledge, but we are dealing with real athletes and humans.

Data cleaning Via Excel

- For the purpose of this project only data from the 13 IFSC bouldering world cups. (i.e. Youth Competitions and location specific competitions were removed)
- Column M (Category) was removed.
- Title columns were made consistent and named more clearly.
- Qualification columns were combined

- Last name format altered so only the first letter is capital
- First and last were combined
- Columns added where rounds are in the format of #.# showing tops.zones for comparison purposes (=LEFT(Q2,1)&"."&MID(Q2,3,1)
- Added a gender column
- Change Date Format

Inconsistencies	Missing Values
There are name misspellings for some foreign names replacing the correct keys with symbols. Changed names for any climbers that were in the finals round for any competition plus many more - several were left as is.	Null values are present when competitors that do not compete in the competition for that row. The null is not a missing data. All pertinent data is present. No rows deleted

Data Cleaning in Jupyter
General check of shape and description
Check for duplicates - No duplicates
Check values of results (float64) and see if date format can be read

Data Profile
2328 rows
10 columns

Competition Title 13 competitions str	Competition Date Ex. 7-8 June 2019	Full Name str	Nation str 58 Nations ex. FRA	Start Number int64
Rank From that comp int64	Gender M/F str	Qualification int64 ex. 3.4	Semifinal int64	Final int64

Questions to explore
Which country has the most climbers participating in the finals? Is this proportional to how many climbers they have participated in the competition compared to other countries? Is this consistent or vary from competition to competition? Year to year?

How much does a competitors starting number align with competition rank? How often do climbers outside of the top 25 make finals? How often do climbers within the top 25 never make a final?

If the start number is not aligned with rank, how often do those that are in the finals consistently make the finals? Does this vary by year?

How many tops do the top ranked athletes top? Is there consistency with the number of tops? Or the difference in the tops from the highest ranked athletes?

Which competitions are the most popular? Does this impact which climbers or countries make the finals?

How might having 2019 being the first Olympic qualifying year impact bouldering outcomes? Did mostly the same climbers participate in the bouldering world cups as similarly as the previous year, aka was there a big shift in who competed from 2018 to 2019? Did more participate? Did certain climbers participate more? Less?