**Project Title:**

120 Years of Olympic History Data analysis

**Team Members:**

Mitchell Hatchett

Andrew Giannettino

Tetyana Spielmann

Jaylen Whittaker

David Scott

**Project Description/Outline:**

This dataset contains historical information on the modern Olympic Games, covering all the Games from Athens 1896 to Rio 2016. The dataset consists of 271,000 records, with each row representing an individual athlete participating in a specific Olympic event. The data was sourced from the [www.sports-reference.com](http://www.sports-reference.com/) website. The significance of this dataset is heightened as the USA is gearing up to host the next Olympic Games in 2028.

The dataset contains 15 key features:

* ID - Unique number for each athlete
* Name - Athlete's name
* Sex - M or F
* Age - Integer
* Height - In centimeters
* Weight - In kilograms
* Team - Team name
* NOC - National Olympic Committee 3-letter code
* Games - Year and season
* Year - Integer
* Season - Summer or Winter
* City - Host city
* Sport - Sport
* Event - Event
* Medal - Gold, Silver, Bronze, or NA

Visualizations:

1. Heat map to represent the countries that have won the most medals
2. Chart: Top 10 countries with gold medals
3. Filters: countries, sex, season
4. Graph: Athletes that won most medals
5. Chart: most medals sorted by age

**Datasets to be Used:** [**https://www.kaggle.com/code/suhass19/120-years-of-olympic-history-data-analysis**](https://www.kaggle.com/code/suhass19/120-years-of-olympic-history-data-analysis)

**Rough Breakdown of Tasks:**

Each group member with collaborate on tasks such as cleaning, processing, and analyzing the data; storing the data into adatabase; creating an ETL workflow; generating visualizations; working on documentation; and presenting a summary.