IST 718 Project Proposal

**Research Questions**

Our proposed project takes a deep dive into historical Olympic data to answer the following questions. The answers to these questions will help countries and individual teams with recruitment and potential ways to achieve competitive advantage in future Olympic games.

* Q1: What factors are related to a country’s number of Olympic medals? Can we predict how many Olympic medals a country may win based on demographic, economic, and development measures?
* Q2: Based on past trends, what might the height, weight, and age of Olympic medal-winning athletes in certain events be in the future? This will help teams with advance planning and recruitment.

**Data Sources**

The key dataset for these analyses is a historical dataset on 120 years of Olympic Games ([Kaggle](https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results?select=athlete_events.csv)). It includes both athlete and event results data. There are 271,116 rows and 15 columns. The dataset contains the following attributes:

* 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

To help us answer the first set of research questions, we will link certain country-level demographic and economic data and other measures we may determine meaningful (Source TBD) to the Olympic data.

**Team members/tentative roles**

* Andrea Leitereg (tentative Q1)
* Tamilselvan Tamilmani (tentative Q2)