

# Data Project Proposal

Aadi Bhandary

2024-11-25

## 1. The names of the students who will be contributing to the group project.

Aadi Bhandary, Alvin Do, Milan Patel

## 2. A tentative title for the project.

"Modeling Probability of Getting a Medal at the Olympics Based On Athlete Characteristics"

## 3. Description of the data file (what they contain including number of variables and number of records). You do not necessarily have to list all the variables, but at least mention those of greatest importance.

-This dataset has 271,116 records and 15 variables.

-Some of the important variables are: Sex, Age, Height, Weight, Sport, Season

## 4. Background information on the data sets, including specific citation of their source (so that I can also access it).

-<https://www.kaggle.com/datasets/heesoo37/120-years-of-olympic-history-athletes-and-results>  
(<https://www.kaggle.com/datasets/heesoo37/120-years-of-olympic-history-athletes-and-results>)

-This dataset was sourced from Kaggle. It is a compilation of historical Olympic data.

-It can be used to analyze athletic performance and any relationships to athlete characteristics.

## 5. A brief statement of the business, science, research, or personal interest you have in the data set which you hope to explore.

- All of us have a general interest in sports (Baseball, Football, and Basketball). This data set is neutral among us three.
- We are going to focus on: Designing a binary classifier (Medal or No medal) as well as predicting multiple outcomes (Gold, Bronze, Silver).

-The data set spans 120 years, which allows us to see how the factors that play into earning a medal change over time.

## 6. Evidence that the data can be loaded into R. Load the data, and print the first few values of the response variable as evidence.

```
library(readr)

dat = read.csv("./athlete_events.csv")
head(dat)
```

##	ID	Name	Sex	Age	Height	Weight	Team	NOC
## 1	1	A Dijiang	M	24	180	80	China	CHN
## 2	2	A Lamusi	M	23	170	60	China	CHN
## 3	3	Gunnar Nielsen Aaby	M	24	NA	NA	Denmark	DEN
## 4	4	Edgar Lindenau Aabye	M	34	NA	NA	Denmark/Sweden	DEN
## 5	5	Christine Jacoba Aaftink	F	21	185	82	Netherlands	NED
## 6	5	Christine Jacoba Aaftink	F	21	185	82	Netherlands	NED

  

##	Games	Year	Season	City	Sport
## 1	1992 Summer	1992 Summer	Barcelona	Basketball	
## 2	2012 Summer	2012 Summer	London	Judo	
## 3	1920 Summer	1920 Summer	Antwerpen	Football	
## 4	1900 Summer	1900 Summer	Paris	Tug-Of-War	
## 5	1988 Winter	1988 Winter	Calgary	Speed Skating	
## 6	1988 Winter	1988 Winter	Calgary	Speed Skating	

  

##	Event	Medal
## 1	Basketball Men's Basketball	<NA>
## 2	Judo Men's Extra-Lightweight	<NA>
## 3	Football Men's Football	<NA>
## 4	Tug-Of-War Men's Tug-Of-War	Gold
## 5	Speed Skating Women's 500 metres	<NA>
## 6	Speed Skating Women's 1,000 metres	<NA>