INT 353 CA-1

NAME – R SAHIL SHARMA

REG. NO. – 12015815

SECTION – K20CH

ROLL NO. – RK20CHA01

ABOUT THE DATASET

This dataset contains data about the participants and the participating regions in the Olympics from the year 1896 to 2016. It contains over 200 thousand rows and 14 columns. The columns contain attributes such as height, weight, age, gender, Team, Year, Season, NOC etc. There are a lot of null values in the dataset, and I have separated it into two parts medalists and non-medalists then I filled the null values using median/mode values of the players by grouping them based on gender and sporting event.

DATASET CONTENT

The file athlete_events.csv contains 271116 rows and 15 columns; Each row corresponds to an individual athlete competing in an individual Olympic event (athlete-events). The columns are as follows:

#	Column	Non-Null Count	Dtype
0	Name	271116 non-null	object
1	Sex	271116 non-null	object
2	Age	261642 non-null	float64
3	Height	210945 non-null	float64
4	Weight	208241 non-null	float64
5	Team	271116 non-null	object
6	NOC	271116 non-null	object
7	Games	271116 non-null	object
8	Year	271116 non-null	int64
9	Season	271116 non-null	object
10	City	271116 non-null	object
11	Sport	271116 non-null	object
12	Event	271116 non-null	object
13	Medal	39783 non-null	object

WHY DID I CHOSE THIS DATASET

In the past decade there has been an exponential growth in the number of people who watch Olympic Games. Olympics can be considered as one of the most watched sporting events around the world. As it is so popular around the world it also generates a huge amount of revenue and broadcasters along the with the IOC use certain tactics for the timings of the conduct of various sporting events so that they can get the maximum amount of viewers and maximise their profit. Player sponsors also use data of the players from the Olympics to figure out which players can be a future prospect for them.

OBJECTIVES

- Examine/clean the dataset
- Explore distributions of single numerical and categorical features via statistics and plots
- Explore relationships of multiple features via statistics and plots
- Figure out the patterns in the physical attributes of players who have excelled in the different sporting events.
- Which sporting events are being dominated by players of certain regions.
- Which are the countries which have been dominating specific sporting events.