**Name: Saurabh Shrestha**

**Dataset and Project Name: Summer Olympic (1976-2008)**

**Date: June 5, 2021**

**Assignment: Summary of Statistical/Hypothetical Question**

For my exploratory data analysis project, I collected the data from Kaggle. My dataset consisted of Summer Olympic data from 1976 till 2008. The columns were city, year, athletes who won, gender of the athlete, city where the Olympic event was held and few more. One example of my histogram variable was on gender, especially women participant from united states. I wanted to see how many women participated in the summer Olympic held from 1976 till 2008. As it turns out there were few women participants during 1976 Olympics - around 50 women. The number started to increase from 1992 event going from 100 women participant to 160 u.sa women participant in 2008 summer Olympic.

While working on the project, I felt I could have worked on other types of analysis as well. Since I had to focus on only the assigned task for the final project, I couldn’t work on other dissection or analysis such as quantity of medal and overall athlete’s contribution through all the Olympic event.

Yes, the challenges I faced was there were not many numeric columns in my dataset. At the beginning I thought I could work on it. But my dataset consisted of City, Discipline, Country and so on. The only numeric column was Year which was the time when the summer Olympic was held. Later when working on the dataset and reading and researching I found it difficult and struggled at the to work on the project. Reason being I my project had exercise which were to work on variables of numeric in nature. For which I filtered medals won. In my original dataset there were medal won by countries such as gold, silver, bronze. I create dictionary medals such as for gold medal and others. This was categorized based on the countries. So, each country had columns of total medal won (gold, silver, bronze and total). I wish my original dataset had columns with total number of men and women in each Olympic.

There were many challenges I faced during the course of completing this assignment.

One of it was working on the graph for PMF. While relearning, I found there were many different ways to implement PMF. Book mentioned one way while others on the net taught other way. I still do not understandable the part is in PMF section where from the book first.MakeFrames(). I tried different ways to work with this approach but failed.