I downloaded data from Kaggle for Olympics. Of the two CSV files, ‘athlete events’, contained multiple data columns and attributes of the athlete’s events, date, sex, region, etc. The other CSV file, ‘noc\_regions’ contained data of the NOC (National Olympic Committee) mapping with the region. In the first data set I removed a column that was a duplicate of the season and year of the games which is already depicted in unpivoted columns year and season. The second data set didn’t have headers; hence I promoted the first row as column headers to ‘noc\_regions’.

There was an opportunity to create a relationship between the two data sets by connecting the ‘NOC’ column from each table. I created a few measures based on sex and NOC which is also the region and linked that information with the medals to further see what sex and region was producing the most medals. I also created few relationships based on the region and teams which links to NOC. Based on that information I grouped gold medal winners, teams, and NOC to understand the locations which have the highest rate of Gold metals. I created a table called teams based on distinct teams. From the new table I grouped by ID of the person, location, region, NOC, and added a column to count the number of metals by group.

I tried to change the data type of the Year to ‘Date’ however, it changed it to mm/dd/yy format. Even though I found a selection to change the date to ‘yyyy’ once I did that the data in the table changed to a single year which was incorrect, hence, I changed it back to data type of ‘whole number.

For my analysis I was looking to find the regions that had the most metals. I also wanted to see what season the region was playing or participating the most. Finding the results based on sex was also interesting and shows that men get more metals than women. I found that the US followed by few European counties have received the most metals throughout all the years with Canada not far behind. The US, Bulgaria, and Sweden have had 20 or more gold metals as the top ranked countries since 1999.

My technique was to show the countries with the most medals over all the years of data. The highest regions are US and Europe. I created groups by metals, gold, silver, bronze, and no value or no metal. Then I created a distinct count-based number of metals by region and by participant ID. I also created a calculated table of teams that showed columns grouped by NOC region, ID, metal group, and distinct metal count.

I would recommend really drilling down the data to look for trends such as what regions participate in summer vs winter sports and finding out which events or sports have the most medals. It is interesting to see how much information you can gain. It shows both men and women earned more Silver medals, followed by Gold, then Bronze. Winter sports are ¼ of the number of regions that participate in summer season event sports. Using filters and slicers is helpful to drill down and see differences in the visuals. The data set was so large however, that a few times my filtering was taking quite some time! Overall, it was interesting viewing the data! The US showed to have the most metals by region not grouped by metal type.

Though the analysis of all the years, Finland, Hungry, and Japan have the most gold medals over time from the year 1900. See more visuals on page 2, ‘Gold Metals’. I was also interested to see what was the highest age of a person that received a gold medal. Age 64 was the highest for rouge and shooting sports. Age 13 was the youngest with 3 Gold medals in diving, 2 in rowing, and 1 in short track speed skating and swimming.

I created a fourth table to group the participants with the sports and metals along with the year. Then I created a quick measure based on correlation of ID, sport, medal, and year. I created another page called participants to look at how many attempts there were by all the participants and the number of metals and or ‘NA’ attempts. It is clear to see many with attempts but no metal, while a small population has metals in each category, and another population with metals in one category.