Final Exam Dataset

The data set I used for my project was the Olympics. I stared working on the data by cleaning it. Since the other table is just to identify the region codes, there was no need to have an id number. I removed both the ID column and the ID# column since they are both unnecessary to the way the data was going to be used. Also, for the medal column I removed all NA values since they were also not applicable to the data set. The age, height and weight column were all listed as text columns because of the N/A data information. So to converted them to whole numbers I performed a transform data that made all the N/A in to 0’s. I honestly wasn’t sure that was the best way to convert the data, but it was the only thing I could think of to solve the issue. For the noc\_regions table I delete the notes column since it also serves no purpose. I then set the table to use it’s first row as header since it did not correct that by default. At this point the data is cleaned and minimized so I load it to power bi and begin working on adding my column, measures and table.

Power Bi was able to automatically detect the related columns of NOC. It automatically created a one to many relationship from the tables athlete\_events to noc\_regions. This relationship will help the athelet\_events table be able to corelate the NOC column to the actual region it correlates to. The column I created was in the athlete\_events table I called it “over 30”. I used the IF DAX function so it would return a true or false statement based on the age criteria. I created this column because it can be used to get a quick result of the players age range when they won. The measure I chose to preform was to find out the total medals earned by the name of the person. I did this by doing a quick measure and specifying count of medal average by name. I then created a calculated table by using the distinct function for cities. I then moved onto creating my visualizations.

My technique of analyzing the data was to first look over the source material. If there was any term, I was unfamiliar with I looked it up. That way I was able to have a complete understanding of what the data meant. For example, NOC I have never heard of before, but it means National Olympics Committee. The NOC varies by area and can help determine the region the player is from. After familiarizing myself with the data sets, I then proceeded to clean up any missing or irregular information. I did not alter the original data though I made sure all my changes were only in PowerBI. I removed any unnecessary columns as well. Once I cleaned the data I started to think about what I wanted to learn from it. I quickly determined I was interested in the data behind gold medal winners. Going in I knew the data wouldn’t be 100% since there was missing information from the original data set. I had to make sure all my visualizations were not counting the 0’s I had to use to replace the N/A information.

The first visualization I worked on that helped me get the information I needed to filter all the others, was a line chart. I used a line chart to find out which countries had the most gold medals. I then filtered the whole sheet down to the top ten countries I found from this visualization. I then decided to do a card that showed the average age of the people who won gold medals in the top ten countries. This I found was interesting because it can be used to know when people are in theory, at their peak physical performance. I then created a stacked bar chart so I can see how which Olympians won the most gold medals in the top ten countries. The nice thing about the bar chart is that it automatically sorted them from most to least gold medal wins. It made it very easy to tell who the top Olympians were. My fourth and final visualization was a map chart. This map chart averaged the gold medal winners’ heights by country.

The information I got form the data is more than I was expecting. To start with the map chart based around heights I could conclude height isn’t necessary. In just the top 10 countries I was looking at the range in height varied greatly. Now this is just a general statement due to the many different Olympic events that medals can be awarded at it. Some sports probably require more height than others. With the card I was able to find out the average age of gold medal winners is 26 years old. This information I found very interesting because it was higher than I expected. I always assumed most Olympic gold medal winners would be in their early 20’s. The line chart visualization showed something sad but no unexpected. The countries with the most gold medals all are considered to have citizens that are well off and have a good GDP. The stacked bar chart shows all the top Olympians in the ten countries. Also, unsurprising If looked at closely, you can also tell that only three females have won over seven gold medals. So, a majority of the top gold medal winners are male.

The information from this data can be used by other countries and trainers. It can help them scout and train their Olympians, to match the performance in these top ten countries. By knowing who the top gold medal winners are, the trainers can research information on the players and their regimen. This information can be used to help them train their Olympians in the same way. By knowing the average age of gold medal winners, it can help determine the best age to cut their loss if their players have not won yet. The height information really varies by which sport the players are preparing for. Though It can be easily modified to show what the average height of the winning gold medalist are for each sport as well.