Billy Fryer

USOPC Data Analytics Fellow Project

When I got this assignment, my first thought was to compare different NGBs to see which ones were similar, particularly across Summer and Winter sports. One benefit from this is that we could budget our money in similar manners between these National Governing Bodies. For example, we wouldn't want to appropriate money the same way for USA Basketball, a team that wins gold quite often and has multiple levels of player development to build future generations of basketball players, as we would for the US Biathlon Association who is much smaller and doesn't have the same development system. I was surprised about some of the connections that were made by defining the "Similar NGBs" by the bucket variables in the "NGBHealthDataOutput Extract" data set. Ideally, I would have liked to use K-Means Clustering to get better results, but under the time constraints, I am very happy with what I was able to get done.

One of the challenges I had in this project was getting the data in a usable format. The NGB Health data in particular was really messy with lots of missing information. I ended up programming in R to alleviate the majority of these issues. I was able to group and summarize to aggregate this data to get it into a tidy csv file. This was fine until I nailed down how I wanted to create my dashboard and the data needed to be manipulated even further. This required me to break apart the data that I had just finished putting together into smaller data sets. This led to the biggest challenge I had which was relating all of these smaller data sets to one another in Tableau. Once I figured that out, from there it was relatively smooth sailing with the exception of a few times I got stuck on formatting the different tables I created.

The most interesting thing that I learned during this project was how to find data from so many different sources. During my work with NC State Athletics, most of the women's basketball or baseball information was easily findable with just one or two websites. With this data, I had to find information from all over the place, compiling records from Lima 2019, Tokyo 2020 and Beijing 2022 just to get medal counts. Then I had to figure out how to associate all of these different data sets together. That's why I decided to do my presentation in Part 2 about Structuring Data.

I think one of the cool things I did with this project was (hopefully) make it very engaging and somewhat interactive for the end user. The project required a Data Story, but I wanted the user to be able to be able to draw their own conclusions from the data, the same way I was able to about the US Ski and Snowboarding Federation in my data story. By creating an interactive dashboard as part of this Data Story, the end user can change the dashboard to show what NGB they are most interested in learning about, and hypothetically could make their own data story without having to do any of the tedious data manipulation I did. I tried to continue this theme of interactivity in Part 2 of the project. Obviously, making a YouTube video interactive is difficult, but I tried my best to relate the concepts I was teaching back to funny stories (such as Austria vs Australia) as well as have other repeated phrases so that the concepts could be reinforced throughout the presentation. Thank you so much for this opportunity to play with this data and learn some new skills as well as practice some old ones. I had a lot of fun doing this and I hope to hear from you soon!