ETL Project write-up
Team 13

Our project was on the history of the Olympic Games (Summer and Winter). We extracted data from 2 different sources on Kaggle. Both sources of data came as CSVs. One dataset we used was called “120 years of Olympic history: athletes and results”. The other dataset we used was called “World Population and Consumer Price Index”. First we imported the datasets into jupyter notebook. We had to transform the data to get the relevant columns we wanted and to have a common column to merge on. Some columns that were not of interest to us we would delete them from the tables. We also renamed some of the columns so someone could easily identify what the column was showing. After that we connected to our postgres database from jupyter notebook. Before checking the tables in jupyter notebook, we created the tables in postgres. Then after checking the tables in jupyter notebook, we loaded our tables into our postgres database from jupyter notebook. Now we could work in the postgres database and run some sql queries to join tables on relevant columns.

From the sql queries we did we found some interesting findings about the Olympic Games. We found the number of medals given out during the most recent 120 years of the olympic games was evenly distributed. More males appear in our dataset than females. This is likely because from our analysis we found that women did not participate in the Olympics before the year 1900. We looked up some well known Olympic athletes in the dataset to see how many medals they had won. For example, Michael Phelps known for swimming and his enormous amount of medals, had won 28 medals in total. We also looked up Roger Federer (known as one of the best tennis players in the world) and found out that he won 2 medals. Then we found how many gold medals the United States had won in total (including Summer and Winter). We were also able to find out how many medals the United States won in total throughout the 120 years. The United States is considered the leader out of all countries for the amount of medals we won. Finally, we did a query on age to find that most Olympic athletes are in their 20’s. This makes sense because most athletes are at top performance levels when they’re young.