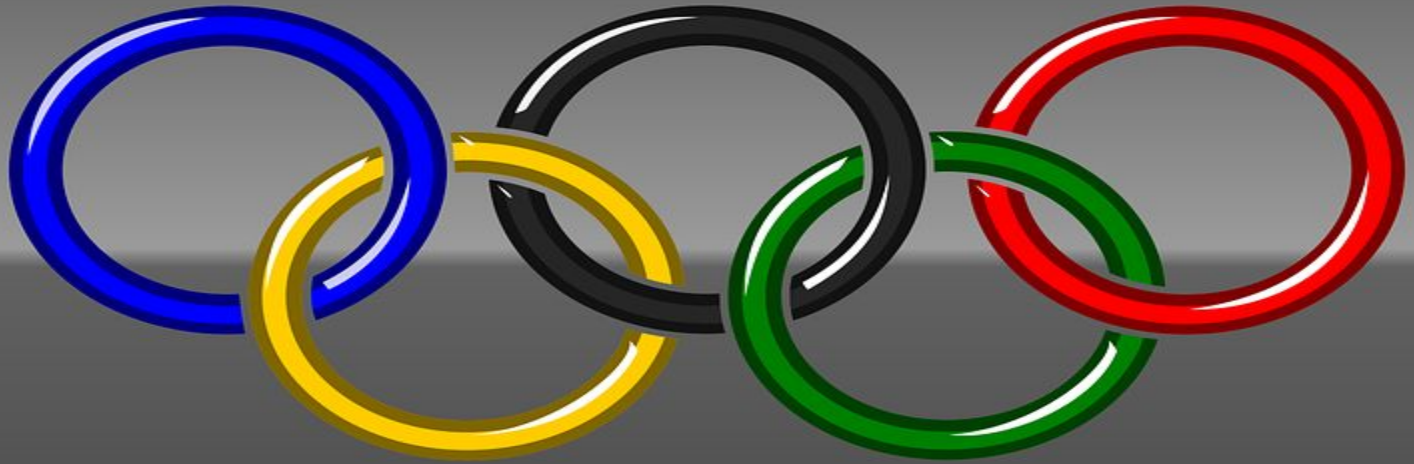
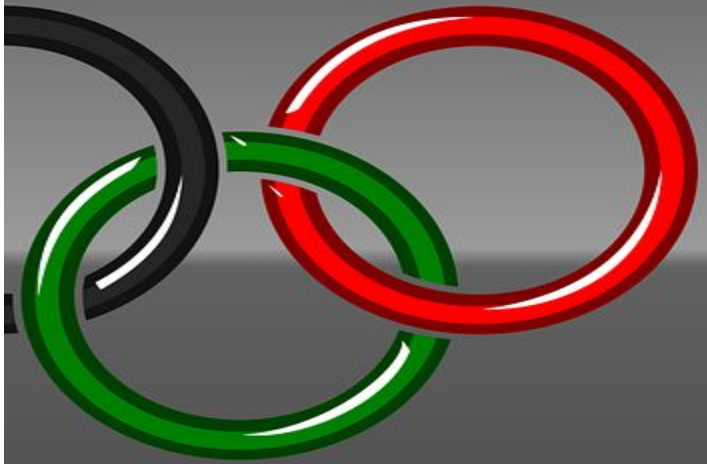
Predicting Medal Placement in the Olympics Based on BMI



Analysis By April Bostic and Margaret Lang

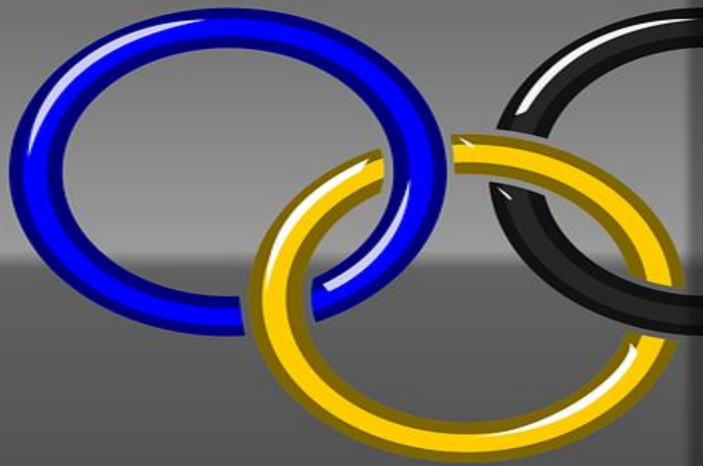
Topic Choice

The occurrence of the 2020 Summer Olympics in Tokyo was the reason behind choosing this topic.



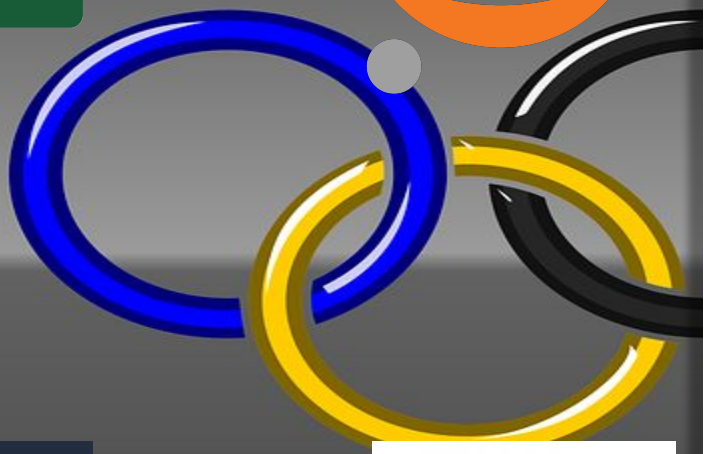
Sports no longer in the Olympics

Data



Title of DataSet: 120 years of Olympic History

DataSet Link: Data on athletes and medal results from Athens 1896 to Rio 2016 from Kaggle

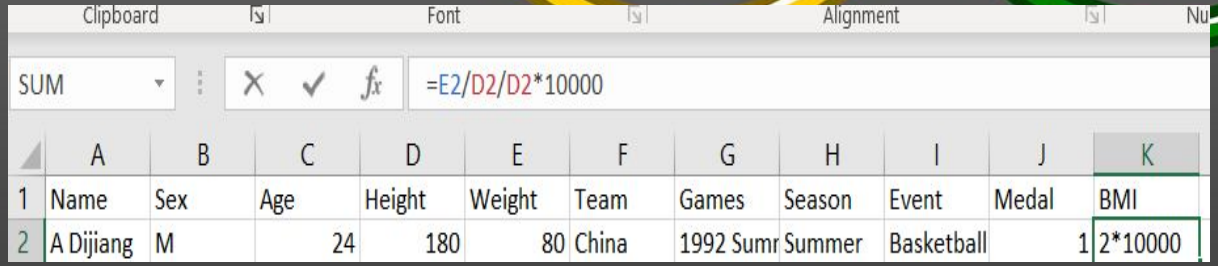


Data Exploration

- Cleaned Data with Pandas in Jupyter Notebook and Excel
- Calculated BMI in Excel
- Created Database in PGAdmin
- Database hosted on AWS

Data Exploration

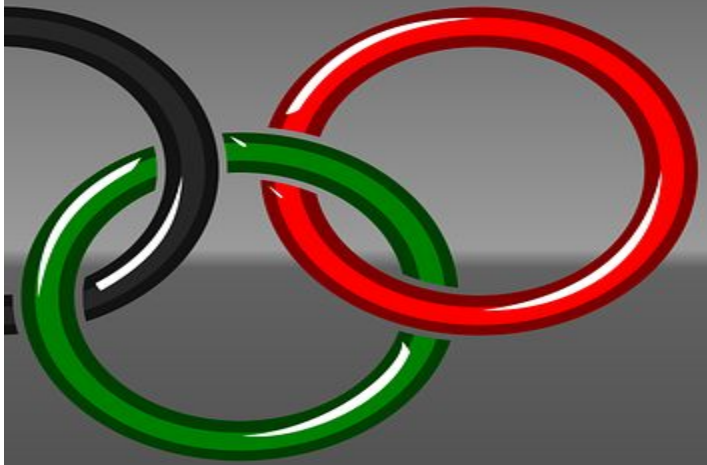
```
In [15]: # Determine if there are any missing values in the data.
         olympic_data_df.count()

Out[15]: Name      271116
         Sex       271116
         Age       261642
         Height    210945
         Weight    208241
         Team      271116
         Games     271116
         Season    271116
         Event     271116
         Medal      39783
         dtype: int64

In [20]: # Drop the NaNs.
         clean_olympic_data = olympic_data_df.dropna()

In [21]: clean_olympic_data.to_csv(r'.\cleandata.csv')
```



	A	B	C	D	E	F	G	H	I	J	K
1	Name	Sex	Age	Height	Weight	Team	Games	Season	Event	Medal	BMI
2	A Dijiang	M	24	180	80	China	1992 Sumr	Summer	Basketball	1	2*10000

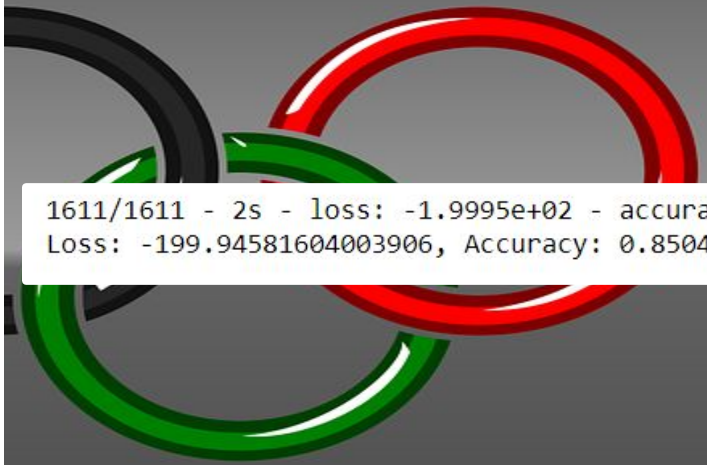
Questions for the Data

- Can we predict Olympic winners based on BMI?
- What is the variance by season or sex?



Data Analysis

- Encoded Data in Medal column
- Dropped useless columns
- Neural Network
  - Initial NN model tried to predict medal type won with low accuracy (0.35) and high loss
  - Switched NN to predicting if athlete will win any medal or no medal with high accuracy and high loss
- Features include Age and BMI
- Target is the Medal column



```
1611/1611 - 2s - loss: -1.9995e+02 - accuracy: 0.8505
Loss: -199.94581604003906, Accuracy: 0.8504714369773865
```

Further Analysis

With more time, we could have analyzed performance by BMI within individual sports. Premature analysis showed some sports (Gymnastics, swimming, etc) had lower average BMI than others (weightlifting, judo, etc).