The codes submitted :

1. Preprocess.R

2. Building\_model.R

3. Visualizations.R

Datasets:

summer.csv

IOC\_Code.xlsx

indicator pwt7.1.xlsx

Medals\_country.xlsx

MenWomenComp.xlsx

(tot per year)Breakdown.xlsx

med per country.xlsx

ALL\_MEDALISTS.csv

Note 1. The path to include the datasets will have to be changed to the required directory in which the datasets exist in the

systems being used.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Using Preprocess.R

The dataset summer.csv was the dataset used for the initial preprocessing and binding the previous year performance with the dataset created i.e. mytable.

The IOC\_Codes and country names were also mapped and binded with mytable. Later mytable was exported to an excel file called Medal\_Country.

Note 2. The pre-processing of binding the GDP Data was done manually as it was not possible to write an R code for the values.

The data set "indicator pwt7.1.xlsx" was downloaded from gapminder and the missing data i.e. from 1896 to 1949 was added using the simple first repetition method of signal extension.

Then the columns in which the olympic games weren't held were removed. The resulting data set was "indicator pwt7.1.xlsx".

Later these columns were manually binded again to the Medal\_Country dataset and the resulting dataset was obtained.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Using Building\_model.R

The inbuilt libraries are included and the SVM model is built using all the attributes of the medal\_country dataset.

The predicted values are stored in pred\_svm and binded with the table. The squared errors in predicted and observed values are calculated, summed

and rooted to find the exact error.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Visualizations.R

The libraries plotly and ggplot were used to generate the visualizations.VISUALIZATION 1 and 4 are bar plots , VISUALIZATION 2 and 3 are line plots.

VISUALIZATION 1 uses MenWomenComp data set , 2 uses (tot per year)Breakdown , 3 uses med\_per\_country data , 4 uses ALL\_MEDALISTS data.