# DATA MANAGEMENT PROJECT REPORT

(Project Semester: August-December 2020)

***SUMMER OLYMPICS STATISTICS 1976-2008***

Submitted by P Sravanthi

Registration No. 11805578

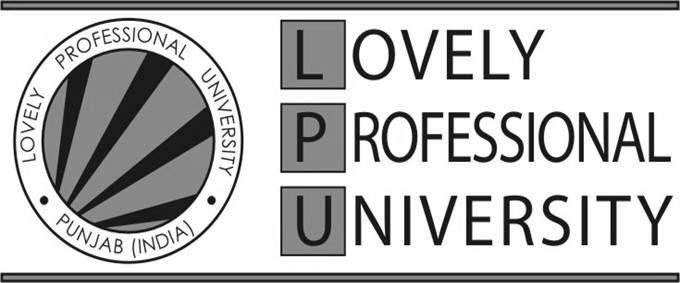
Programme and Section: B.Tech (Computer Science), Course Code: INT 217

Under the Guidance of

### Savleen Kaur, 18306

**Discipline of CSE/IT**

**Lovely School of Computer Science and Engineering Lovely Professional University, Phagwara**



**CERTIFICATE**

This is to certify that P SRAVANTHI bearing Registration no. 11803529 has completed INT 217 project titled, **“SUMMER OLYMPICS STATISTICS (1976-2008)”** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

### Savleen Kaur, 18306

School of Computer Science and Engineering Lovely Professional University

Phagwara, Punjab.

Date: 19 December, 2020

**DECLARATION**

I, P Sravanthi student of Bachelor in Technology under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 19 December 2020

Registration No. 11805578 PSravanthi

**ACKNOWLEDGEMENT**

I would like to express my special thanks of gratitude to my teacher Ms. Savleen Kaur who gave me the golden opportunity to do this wonderful project of analysis of the data of a superstore namely “ SUMMER OLYMPICS STATISTICS: 1976-2008” which also helped me in doing a lot of research and I came to know about so many new things. I am thankful to them. Secondly, I would also like to thank my friends and faculties who helped me a lot in finalizing this project objectives within the limited time frame.

## Table of Content

1. Introduction
2. Scope of the Analysis
3. Source of dataset
4. ETL process
5. Analysis on dataset (for each analysis)
   1. Introduction
   2. General Description
   3. Specific Requirements, functions and formulas
   4. Analysis results
   5. Visualization
6. References

## INTRODUCTION

**Data Analysis** is a process of inspecting, cleansing, transforming, and modelling data with the goal of discovering useful information, informing conclusions, and supporting decision- making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, while being used in different business, science, and social science domains.

In this project WORLD SUMMER OLYMPICS SATISTICS (1976-2008),

This dataset is a list of all the medal winners in the Summer Olympics from 1976 Montreal to 2008 Beijing. It includes each and every medal awarded within the period. This dataset is intended for beginners so that they can get a taste of advanced Excel functions which is perhaps one of the key skills required to be a great data scientist. I too got my hands dirty with the dataset and played with some advanced Excel functions. Further, this dataset can also be used for a predictive model as to which country is likely to fetch the highest number of gold in a particular sports category (just an example), etc

## SCOPE OF ANALYSIS

I wanted to analyze the SUMMER OLYMPICS data collected from all over the world (1976-2008 year). The main objective of this data entry is to contribute to an informed, open debate about the ways to increase participation in sports by understanding it in a deeper level .So,that not only men’s women’s also need to participate more in sports. The project aims to find and explain some as like gender, cities and medals across countries. I cleaned, analyzed and presented it in the form of a dashboard for it to be more understandable for users and to get more meaningful insights.

## SOURCE OF DATASET

The data is being taken from the Kaggle.

<https://www.kaggle.com/divyansh22/summer-olympics-medals>

**About Kaggle**: Kaggle is an Airbnb for Data Scientists – this is where they spend their nights and weekends. It’s a crowd-sourced platform to attract, nurture, train and challenge data scientists from all around the world to solve data science, machine learning and predictive analytics problems. It has over 536,000 active members from 194 countries and it receives close to 150,000 submissions per month. Started from Melbourne, Australia Kaggle moved to Silicon Valley in 2011, raised some 11 million dollars from the likes of Hal Varian (Chief Economist at Google), Max Levchin (PayPal), Index and Khosla Ventures and then ultimately been acquired by the Google in March of 2017. Kaggle is the number one stop for data science enthusiasts all around the world who compete for prizes and boost their Kaggle rankings. There are only 94 Kaggle Grandmasters in the world to this date.

## ETL PROCESS

In computing, extract, transform, load (ETL) is a process in database usage to prepare data for analysis, especially in data warehousing. Data extraction involves extracting data from homogeneous or heterogeneous sources, while data transformation processes data by transforming them into a proper storage format/structure for the purposes of querying and analysis; finally, data loading describes the insertion of data into the final target database such as an operational data store, a data mart, or a data warehouse. A properly designed ETL system extracts data from the source systems, enforces data quality and consistency standards, conforms data so that separate sources can be used together, and finally delivers data in a presentation-ready format so that application developers can build applications and end users can make decisions.

Precisely, ETL is defined as a process that extracts the data from different RDBMS source systems, then transforms the data (like applying calculations, concatenations, etc.) and finally loads the data into the Data Warehouse system. ETL stands for Extract, Transform and Load.

Through the process of ETL, we are going to clean the dataset and bring all the entities to their proper data format.

### Step 1: Removing the blank cells from the dataset.

For this, select the whole dataset. Go to Find and Select in the Home tab of excel. Select Go to Special from the drop-down menu and then tick the blank option. All the blank cells will be selected. Then go to Delete option in the home tab again and select Delete Rows from the drop- down menu. This will remove any rows with blank cells.

### Step 2: Removing columns which are not properly defined or not crucial to our analysis.

For this we will columns which are redundant like the column with just the index numbers. For this we will select that particular column and then go to delete option in the home tag and then select Delete Columns from the drop-down menu.

### Step 3: Giving proper and appropriate column names.

The dataset does not have proper columns so our next step would be to giver proper column names to the columns wherever required.

### Step 4: Excluding the NULL values from the data.

We’ll be using Tableau prep for this work as it’ll make the work simple and faster because we might not know how many null values could be there in this huge data set. Tableau helps us doing one step cleaning with ease.

### 

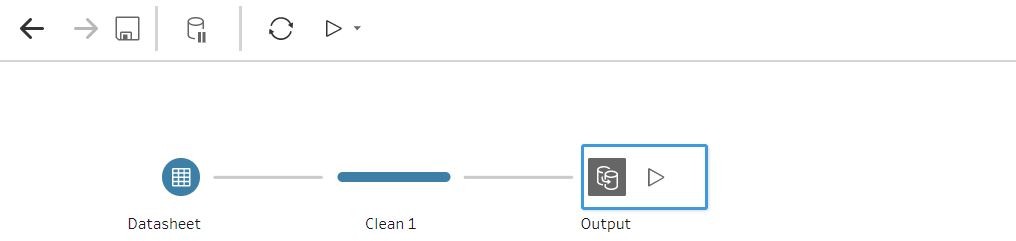
### Step 5: Improvising Proper Data Formatting

Without proper Data Formatting, proper analysis will not take place. So, we will bring down certain columns to their proper format. For example, the dates should be in the date format and price and sales should be in currency format for better results.

### Step 6: Removing Duplicate Values if Any

It might be possible that our data may be containing duplicate values which may hinder in precise analysis. So, our last task in ETL will be removing duplicate values and making our data perfect for analysis.

**Cleaned Data after ETL process in Tableau prep:**



**ANALYSIS OF DATASET:**

1. Year Wise Medals:

**This chart shows Year wise no of medals like bronze, gold, silver.**

**I using slicer of years from 1976-2008 and countries slicer we can see the chart varies.**

**It will also be helpful if we want to check in which year which particular country got how many medals like that we can examine through the help of this line chart.**

1. **SPORTS WISE TOTAL MEDALS:**

1)This chart shows the analysis of sport wise totals no .of medals in that particular sport

2)This is also linked with country slicer so that we can examine country wise different sports and that sport consists of no of medals.

### 3)COUNTRY WISE MEDALS:

1. The above chart describes country wise total no of medals **(**Bronze,gold,silver)
2. It is helpful in analyzing data if we want from any of the particular country by using country code.

4.Athlete From Each City:

The above Bar chart shows no of athlete from each city of country .

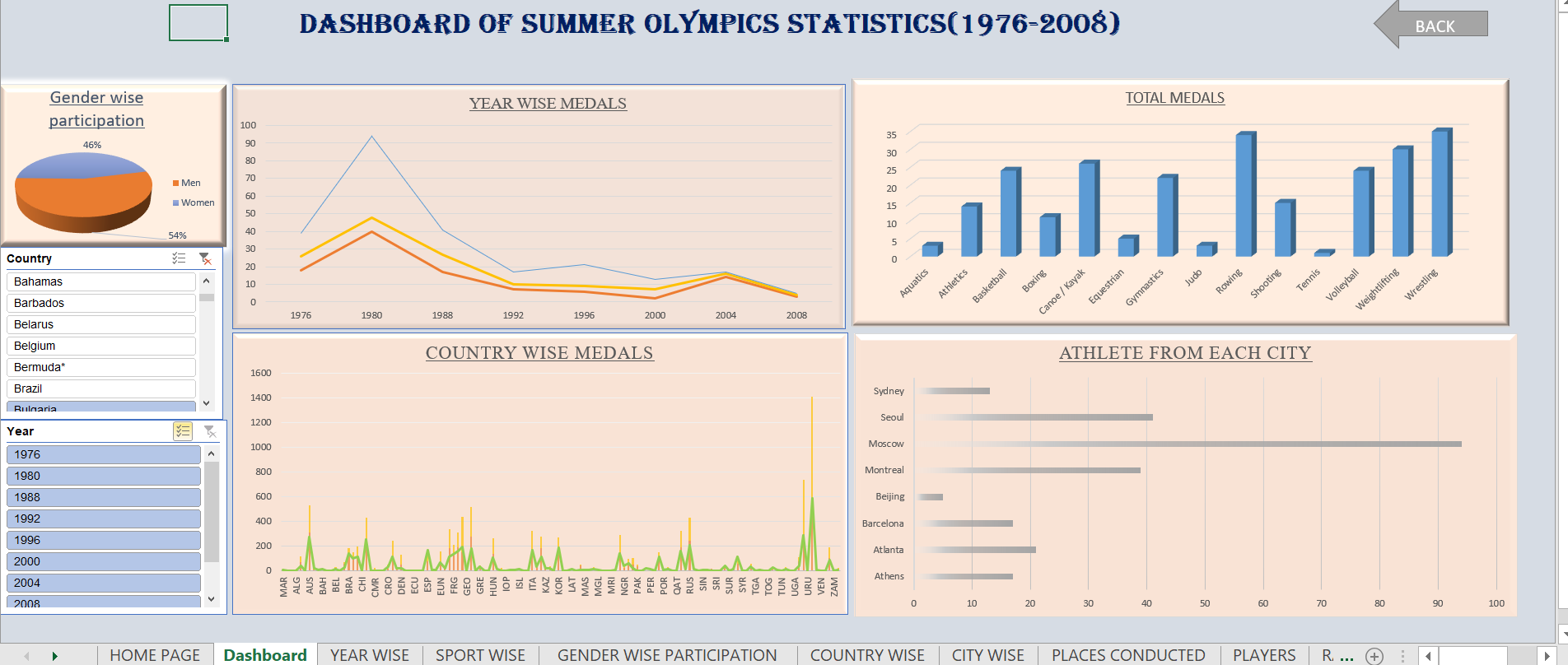
It is also linked with country slicer so that we can see particular cities total no of athlete.

### 5.Average number of suicides of each generation in different countries

This above chart shows the analysis on gender wise people participated from different countries.

So,it is also connected to country slicer through that we can come to know who is participating more and by see seeing this all women’s and men’s need to do participate in sports.

**DASHBOARD:**

****

**References**

* + YouTube
  + Ppt’s and notes
  + Easy Excel



[This Photo](https://picnicwithants.com/2017/10/06/thank-you-ever-so-much/) by Unknown Author is licensed under [CC BY-NC](https://creativecommons.org/licenses/by-nc/3.0/)