

### Introduction to Data Management PROJECT REPORT

(Project Semester August-December 2021)

**PROJECT REPORT ON**

## 120 Years of OLYMPICS History (1896-2016)

Submitted by

### Banagani Prashanth

### Reg.No : 11906733

Programme : **Bachelor of Technology**

Section : KM002

Course Code: INT217

Under the Guidance of

**Maneet Kaur**

**Discipline of CSE/IT**

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

# DECLARATION

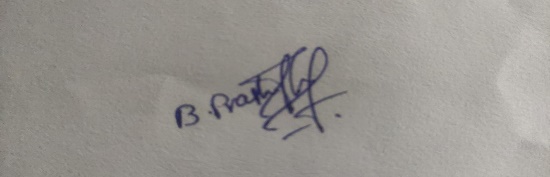
I, Banagani Prashanth, student of Computer Science & Engineering under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

**Date: 15/12/2021**

**Banagani Prashanth**

**Registration No: 11906733**

**Signature:**



# ACKNOWLEDGEMENT

The success and final outcome of this project required a lot of guidance and assistance from many people. All that I have done is only due to such supervision and assistance and I would not forget to thank them.

I respect and thank **Mr. Ashok Mittal** for providing me an opportunity to do the **project in Data Management** and giving me all the support and guidance, which made me complete the project successfully. I am extremely thankful to him for providing such a nice support and guidance.

I owe my deep gratitude to my project trainee **Mrs. MANEET KAUR** who took keen interest on my project and guided me all along, till the completion of my project work by providing all the necessary information.

I am thankful and fortunate enough to get constant encouragement, support and guidance from the teaching staff of **“School of Computer Science and Engineering”** which helped me successfully completing my project.

Banagani Prashanth

**CONTENTS:**

|  |  |  |
| --- | --- | --- |
| **Sr No.** | **Title** | **Page No.** |
| 1 | Introduction | 5 |
| 2 | Objectives/Scope of the Analysis | 6 |
| 3 | Source of dataset | 7 |
| 4 | ETL Process | 9 |
| 5 | Analysis of dataset | 12 |
| 7 | Final Dashboard | 27 |
| 8 | References | 28 |

**INTRODUCTION**

**Data management** is the practice of collecting, keeping, and using data securely, efficiently, and cost-effectively. The goal of data management is to help people, organizations, and connected things optimize the use of data within the bounds of policy and regulation so that they can make decisions and take actions that maximize the benefit to the organization. A robust data management strategy is becoming more important than ever as organizations increasingly rely on intangible assets to create value.

The use of Excel is widespread in the industry. It is a very powerful data analysis tool and almost all big and small businesses use Excel in their day-to-day functioning. Microsoft Excel is one of the top tools for data analysis and the built-in pivot tables are arguably the most popular analytic tool.

Using excel we can perform many operations on a given dataset. Few of the important operation are:

* Basic functions like arithmetic as well as various logical functions
* Formatting rows and columns
* IF and the nested IF functions
* Data filtering
* Use of Pivot tables with categorical as well as numerical data
* Line, Bar and Pie charts

Pivot charts

# OBJECTIVES/SCOPE OF ANALYSIS

After analysis of the dataset, the aim of this project is to give answer of given objectives in easy way:

* Distribution of Medals with in sports
* Top 10 Countries with most Medals
* Top 10 players with most Medals
* Top 10 sports with most Medals
* Analysis of Medals according to gender
* Female vs Male
* Age vs Medals

# SOURCE OF DATASET:

### Source of dataset:

https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results

The dataset is based on OLYMPICS from 1896 to 2016. This dataset contains more than 27,000 rows and 14 columns.

The columns included in the dataset are given below:

* ID

Id of the player

* NAME

Name of the player

* SEX

Gender female of male

* AGE
* HEIGHT,
* WEIGHT
* MEDAL

Gold , Silver or Bronze

* WEIGHT
* TEAM

Country Name

* NOC

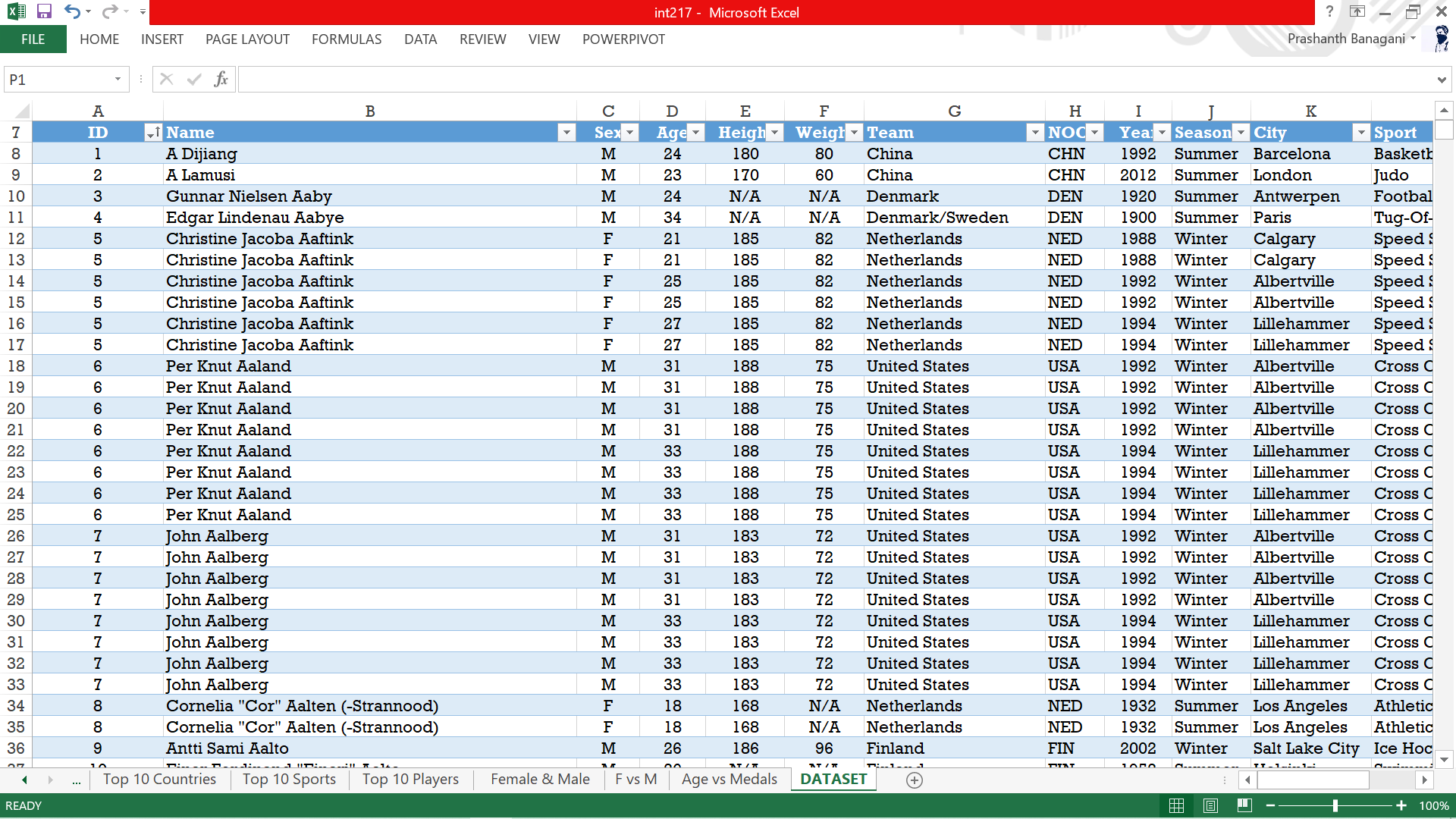
Country code

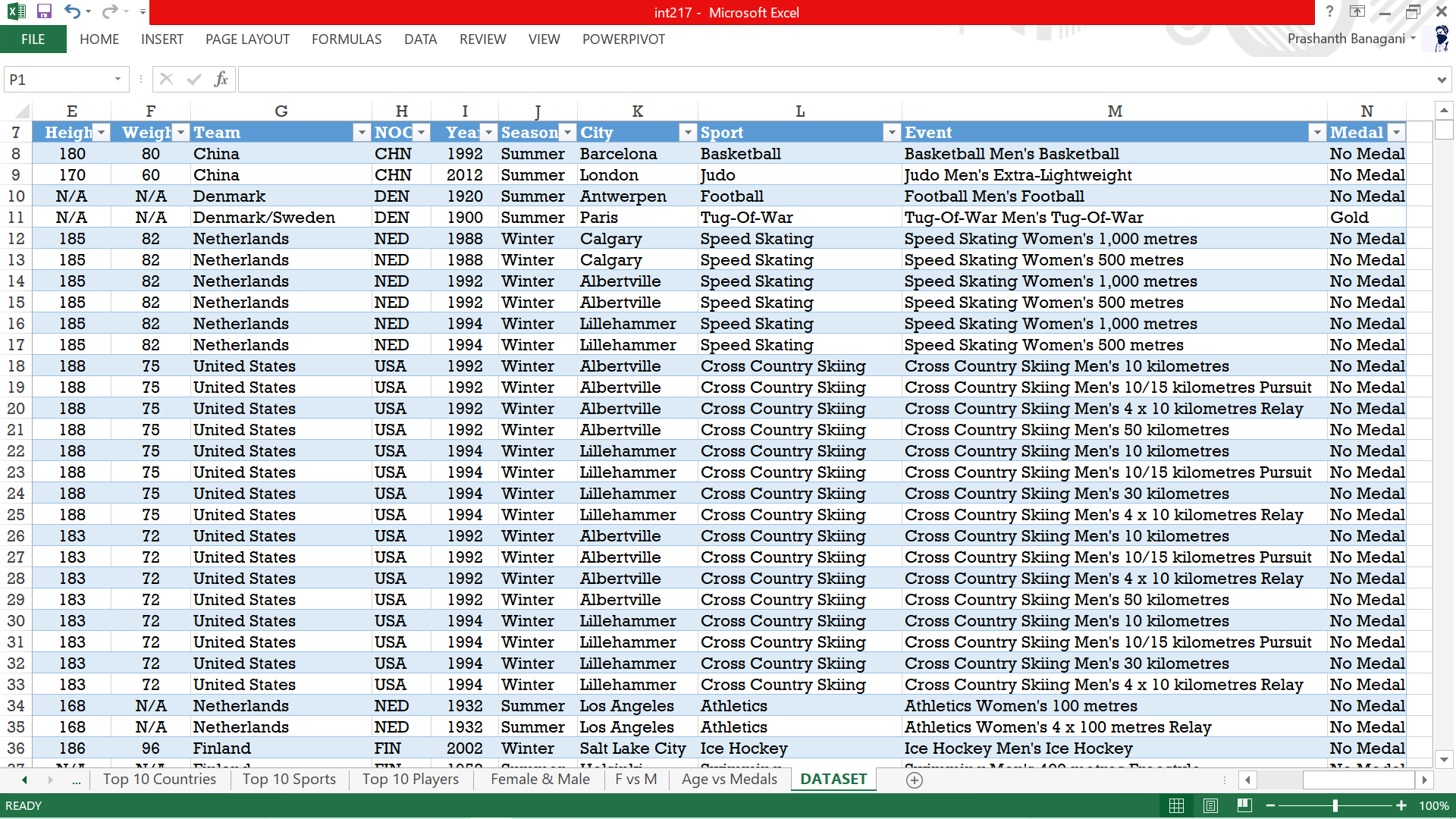
* YEAR
* Season

Winter or Summer

* City
* SPORT
* Event Name

**Sample of dataset with data fields is given below:**



****

# ETL PROCESS:

* **ETL** is a process that extracts the data from different source systems, then transforms the data (like applying calculations, concatenations, etc.) and finally loads the data into the Data Warehouse system.
* Full form of ETL is Extract, Transform and Load.
* The triple combination of ETL provides crucial functions that are many times combined into a single application or suite of tools that help in the following areas:
  + Enhances Business Intelligence solutions for decision making.
  + Allows verification of data transformation, aggregation and calculations rules.
  + Allows sample data comparison between source and target system.
  + Helps to improve productivity as it codifies and reuses without additional technical skills.

## Steps taken to clean dataset thorough ETL process

### Step 1:

* Firstly, open blank excel file and go to Data from Tab and select get data, then we choose from file and then use work book to get data from different excel book.

### Step 2:

### Many columns had null values so I tried fixing those by replacing the nulls with the

**value of N/A**

**Step 3:**

### Then loaded the file back after cleaning

### Step 4:

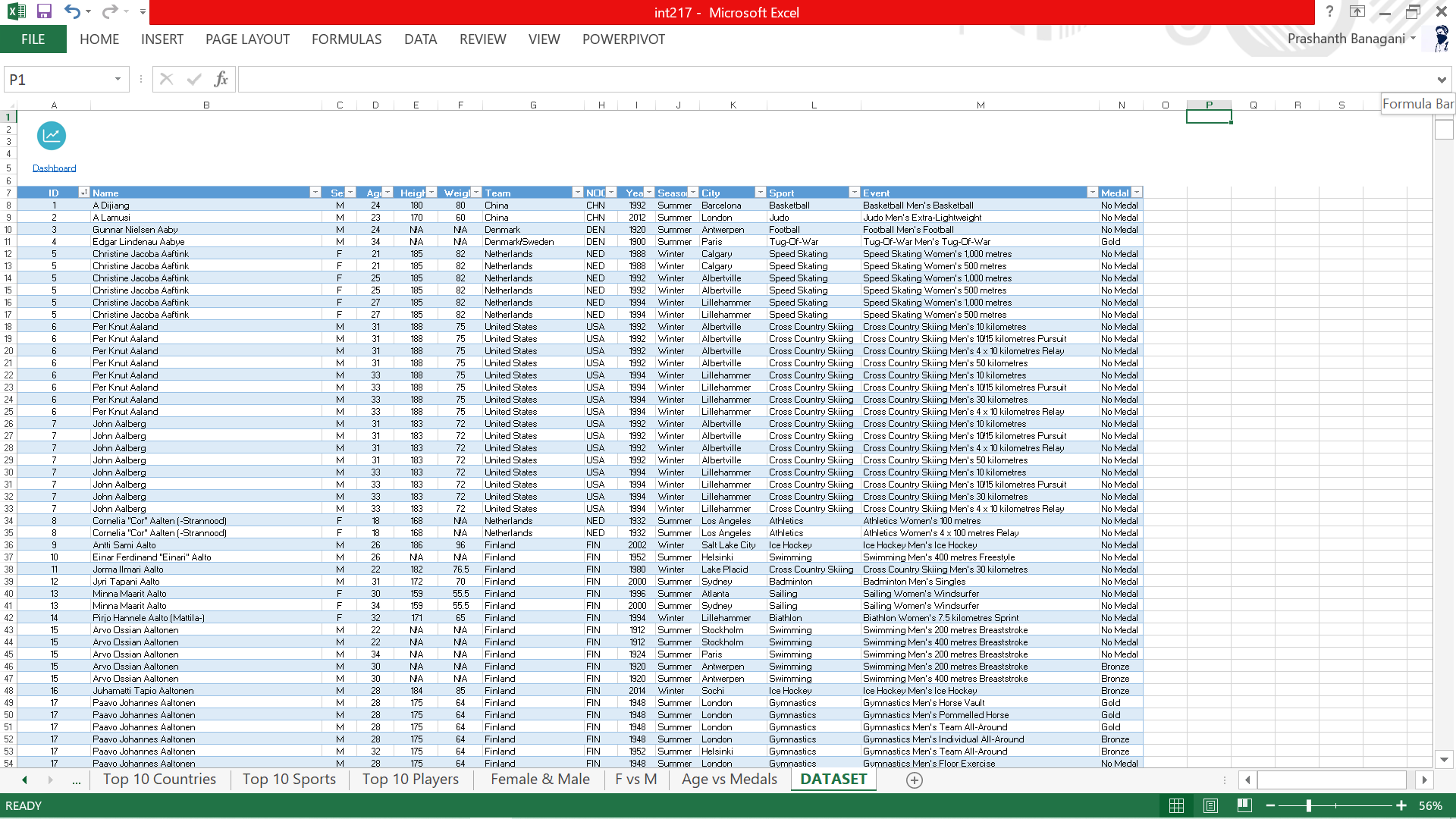
* Now opening data in excel I cleaned the Olympics dataset where I had to change the data and removed all the unwanted columns .

### Step 5:

### To filter the null values we can either remove the rows or we can replace those nulls.

### Step 6:

* Finally, the modified and loaded data is ready to perform our analysis.



**Analysis on the dataset**

## Distribution of Medals with in sports

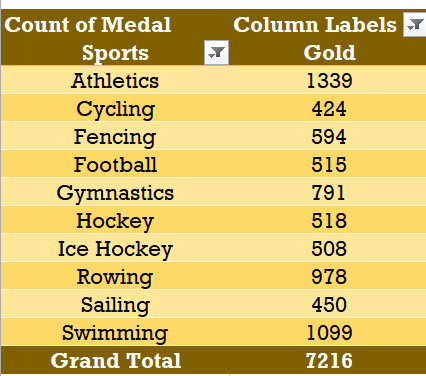
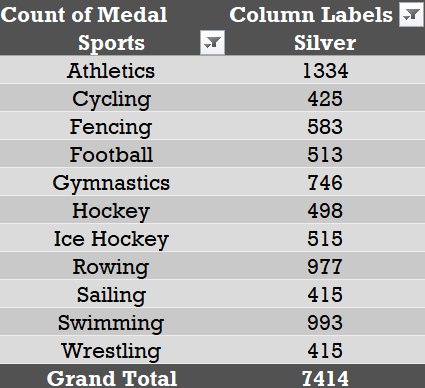
### Introduction

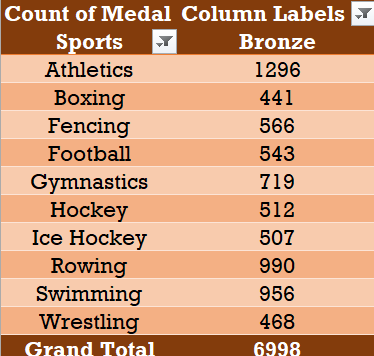
* + By performing this analysis, we get the names of top 10 sports that have won the most medals in GOLD, SILVER, BRONZE.

### Description

* + The analysis is based on top SPORTS and how many medals they have.

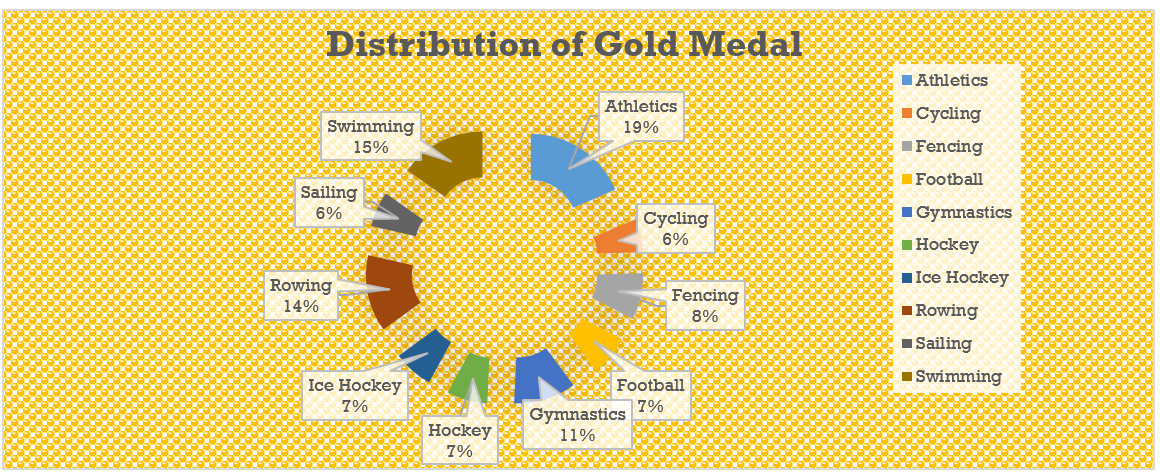
### Specific requirements, functions and formulas

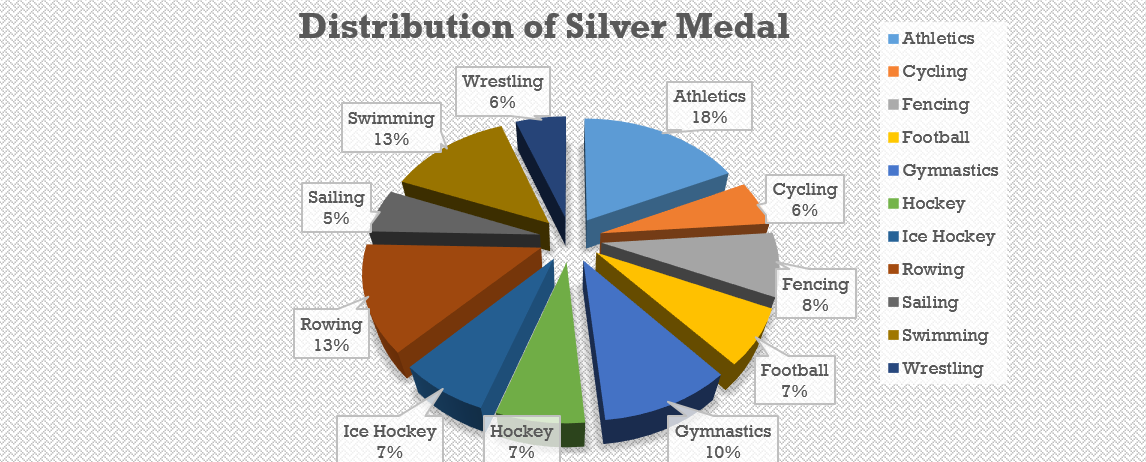
* + Pivot table is used for the analysis.
  + Count function is used in pivot table for the Count of the medals in the pivot table.  



### Visualization

* I have used Doughnut chart for Gold medals ,3D pie chart for Silver medals and Bar of Pie chart for Bronze medals





### C:\Users\Asus\Pictures\Screenshots\Screenshot (147).png

## Top 10 Countries with most medals.

### Introduction

* + By performing this analysis, we will get countries names that have most medals

### Description

* + The analysis based on the Team names and the count of medals.

### Specific requirements, functions and formulas

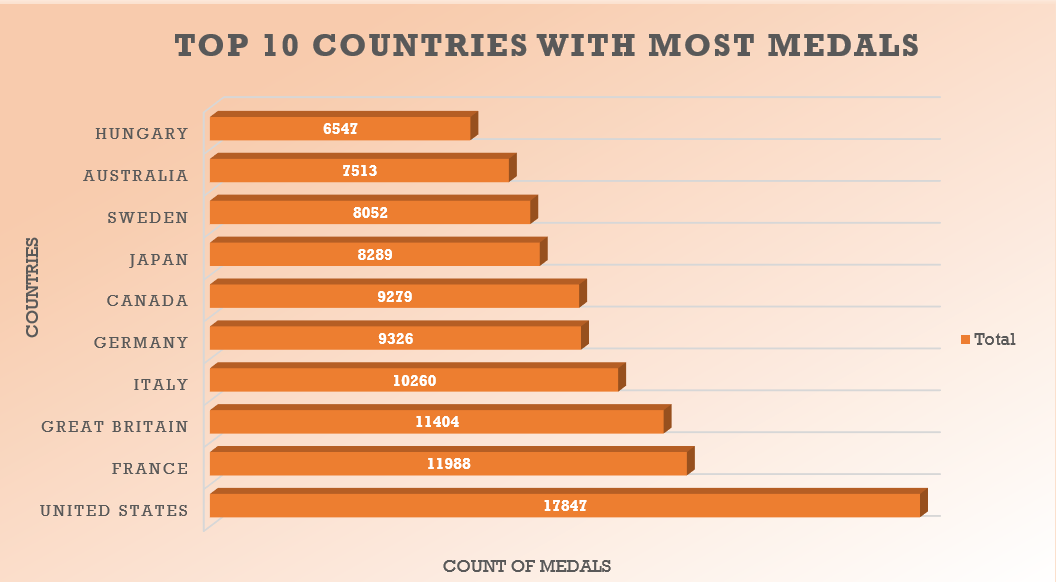
* + Pivot table is used for the analysis.
  + Count function is used in pivot table for the Count of the medals in the pivot table

### Analysis results:

### C:\Users\Asus\Pictures\Screenshots\Screenshot (149).png

### Visualization:

* I have used Clustered Bar chart to represent this analysis as it is the best way to compare values of different Constituency.



## Top 10 Players with most Medals.

### Introduction

* + By performing this analysis, we will get Top 10 Player names with most medals

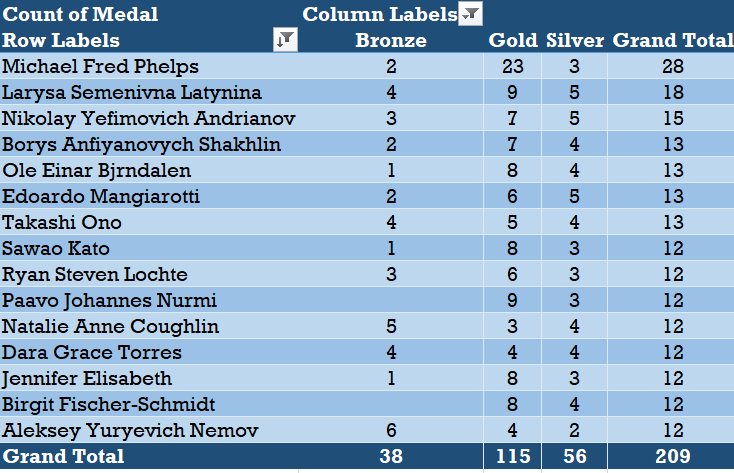
### Description

* + The analysis based on Player names, Medals.

### Specific requirements, functions and formulas

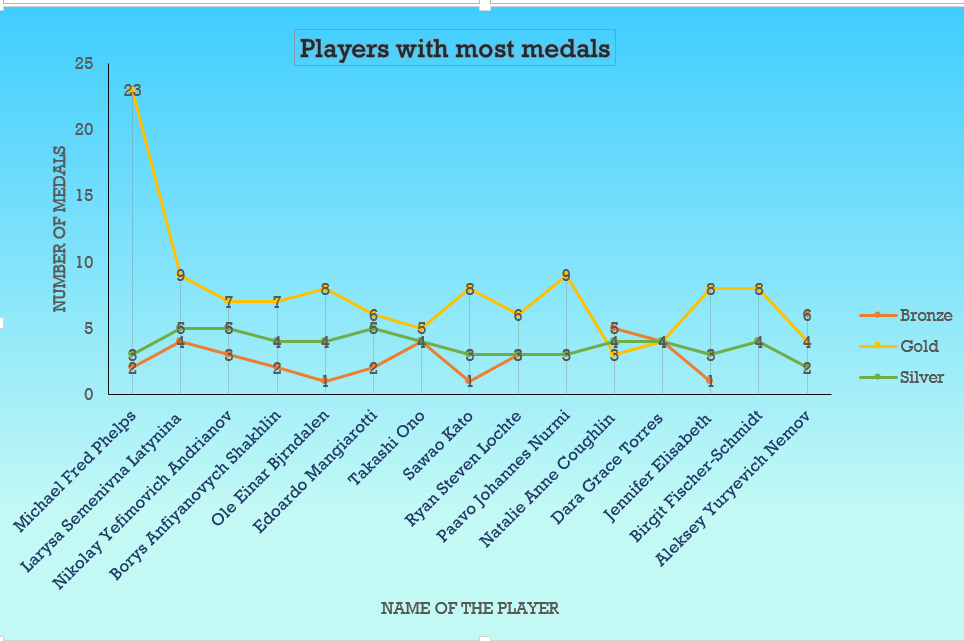
* + Pivot table is used for the analysis.
  + Count function is used to count the total number of Medals

### Analysis results:



**Visualization:**

* I have used Line Chart to represent this analysis as the data-point are huge stacked line graphs suits the best.



## Analysis to find the Top 10 Sports with most medals

### Introduction

* + By performing this analysis, we will get top 10 Sports with most number of medals.

### Description

* + The analysis based on Sports names Medals.

### Specific requirements, functions and formulas

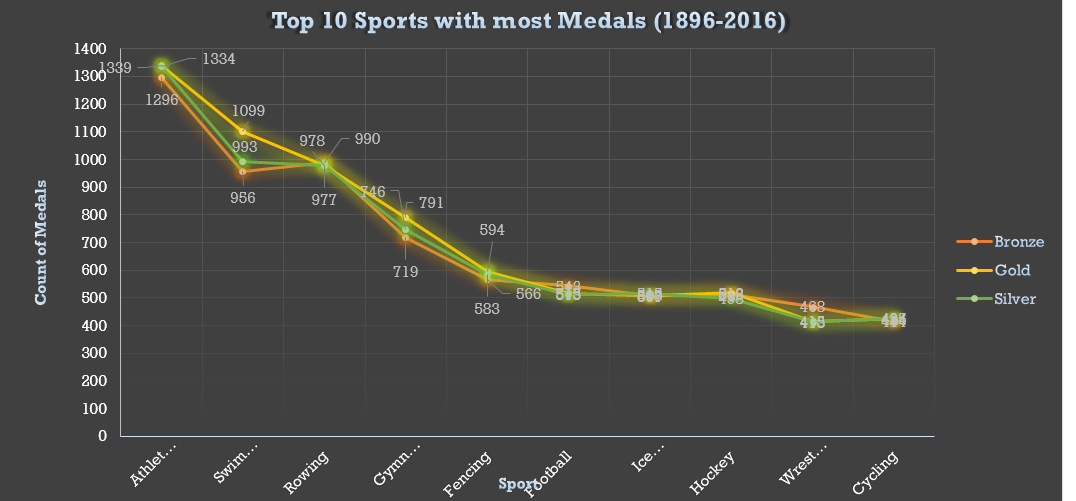
* + Pivot table is used for the analysis.
  + Sum function is used to calculate the total number of Medals.

### Analysis results:

### C:\Users\Asus\Pictures\Screenshots\Screenshot (151).png

**Visualization:**

* I have used Line chart to represent this analysis as the data-point are huge stacked line graphs suits the best.



## Analysis of medals according to Gender.

### Introduction

* + By performing this analysis, we will get the top 10 sports with most medals with respect to gender

### Description

* + The analysis is based SEX , SPORTS , MEDALS..

### Specific requirements, functions and formulas

* + Pivot table is used for the analysis.
  + Count function is used in pivot table for the COUNT of the medals.

### Analysis results:

### FOR FEMALE:

### C:\Users\Asus\Pictures\Screenshots\Screenshot (155).png

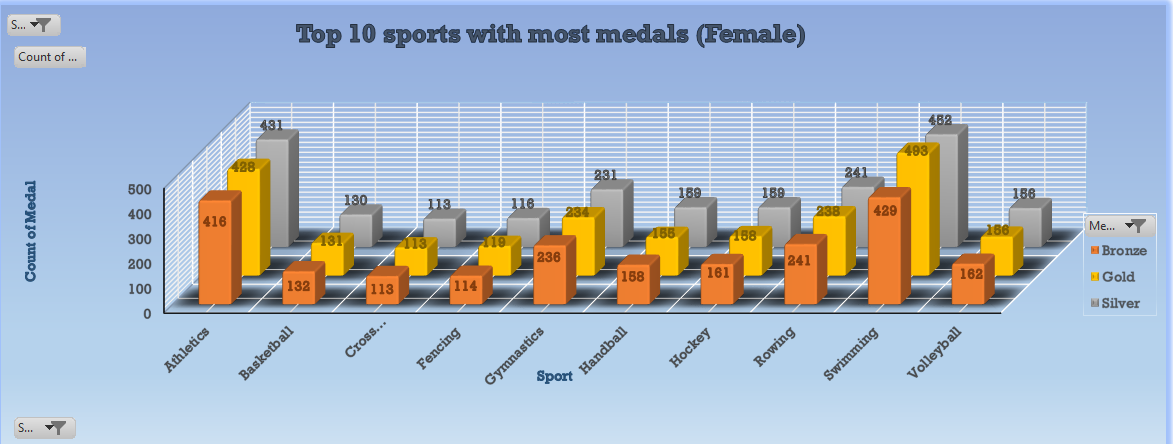
### FOR MALE

### C:\Users\Asus\Pictures\Screenshots\Screenshot (157).png

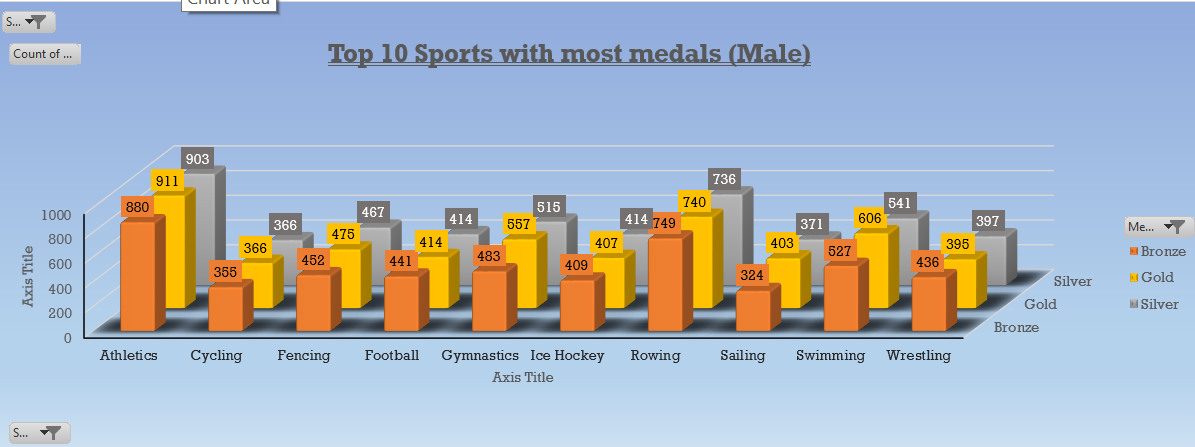
**Visualization:**

* I have used 3D Column chart to represent this analysis .

FOR FEMALE:



FOR MALE:



## FEMALE VS MALE (MEDALS)

### Introduction

* + By performing this analysis, we will get the comparison between male and female

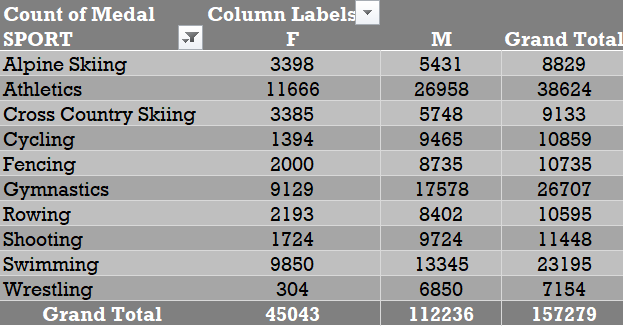
### Description

* + The analysis based on the SEX, SPORTS, MEDALS.

### Specific requirements, functions and formulas

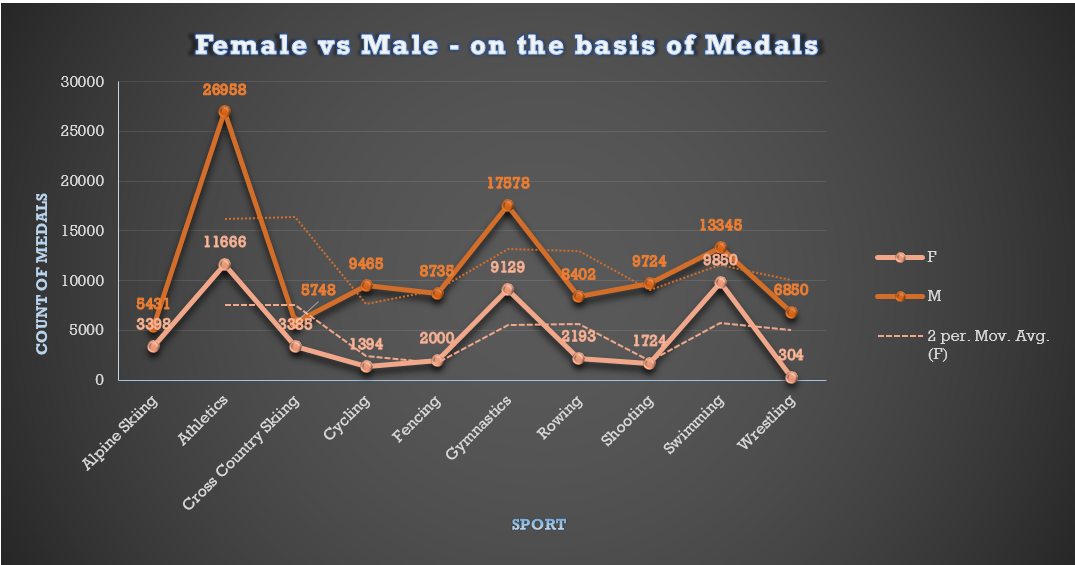
* + Pivot table is used for the analysis.
  + Count function is used in pivot table for the Total number of medals.

### Analysis results:



**Visualization:**

* I have used Line chart (line with markers) because this chart is very useful during comparison



## AGE VS MEDALS (with respect to medals).

### Introduction

* + By performing this analysis, we will get that at what age players getting most medals.

### Description

* + The analysis based on the YEAR, SPORT, AGE , MEDALS.

### Specific requirements, functions and formulas

* + Pivot table is used for the analysis.
  + Count function is used in pivot table for the Count of the MEDALS in the pivot table

### Analysis results:

### C:\Users\Asus\Pictures\Screenshots\Screenshot (161).png

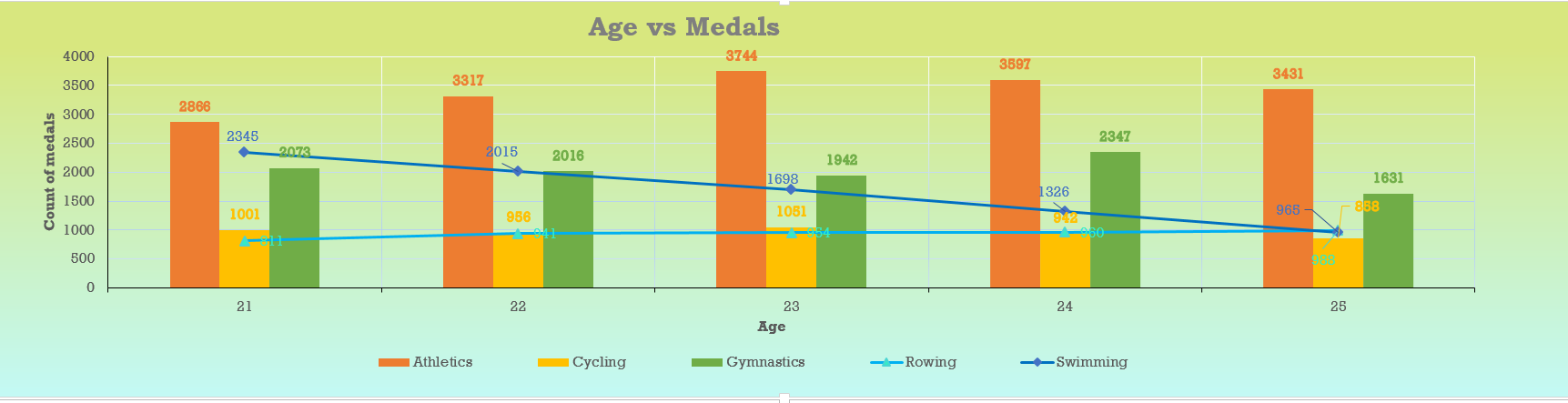
### 

### Visualization:

* I have used custom combo chart to represent this analysis as it is the best way to compare values of age and the number of medals .

### Slicers :

* I have Inserted slicer for state so that we can get the data of any particular state year wise.



### FINAL DASHBOARD:



### REFERENCES:

* https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results
* YOUTUBE
* GOOGLE

