INTRODUCTION TO DATA MANAGEMENT

PROJECT REPORT

5TH Semester August-December 2021

DATA ANLYSIS ON OLYMPICS

Submitted by

Swaraj Senapati

Registration No.: 11902728

Programme: Bachelor's in Technology

Section: KM009

Course Code: INT 217

Under the Guidance of

Name of faculty: Sandeep Kaur

U Id: 23614

Discipline of CSE/IT

Lovely School of Computer Science & Engineering

Lovely Professional University, Phagwara

Transforming Education Transforming India

CERTIFICATE

This is to certify that I Swaraj Senapati bearing 11902728 has completed INT 217 project titled, **"Data Analysis on Olympics"** under my guidance and supervision. To the best of my knowledge, the present work is the result of his original development, effort and study.

Signature and Name of the Supervisor

Designation of the Supervisor

School of Computer Science and Engineering
Lovely Professional University
Phagwara, Punjab.

Date: …… December 2021

DECLARATION

I, Swaraj Senapati, student of bachelor's 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: 10 December 2021

Name of the student: Swaraj Senapati Signature

Registration No.: 11902728

ACKNOWLEDGEMENT

I would like to express my gratitude towards all the people who have contributed their precious time and effort to help me. Without whom it would not have been possible for me to understand and complete the project.

I would like to thank Professors of Department of Computer Science and Engineering, my Project Supervisor for his guidance, support, motivation and encouragement through out the period this work was carried out. His readiness for consultation at all times, his educative comments, his concern and assistance even with practical things have been invaluable.

We are grateful to Department of Computer Science and Engineering for providing necessary facilities in the department.

Name: Swaraj Senapati

Registration No.: 11902728

Signature:

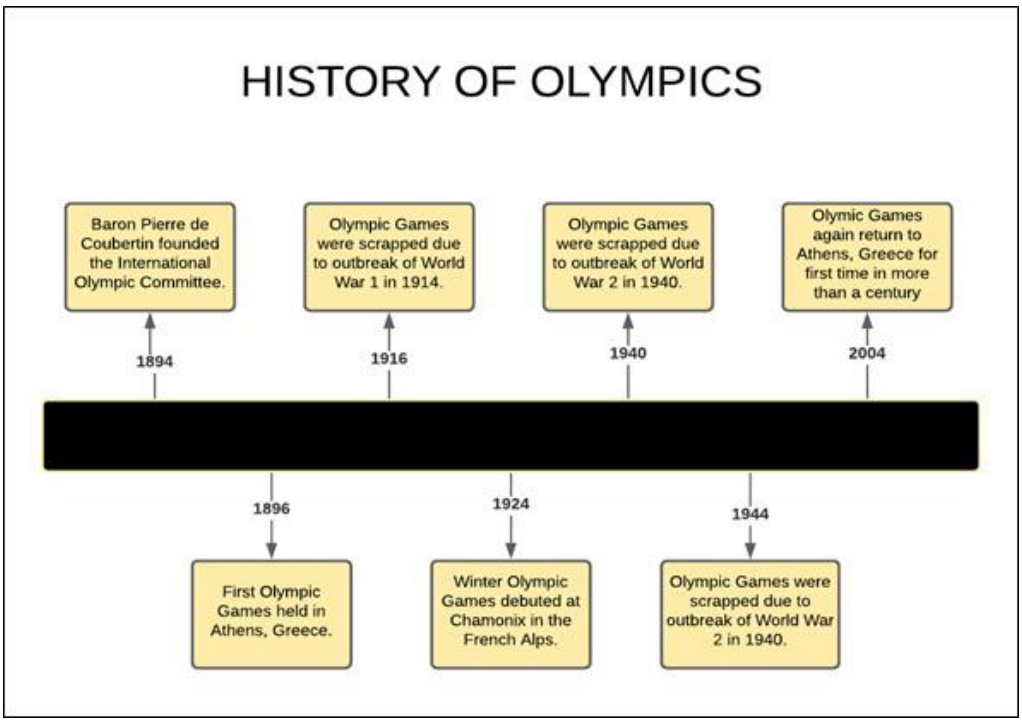
Table of Content

SL NO.	Content	Page No.
1	Cover Page	1
2	Certificate	2
3	Declaration	3
4	Acknowledgement	4
5	Index	5
6	Introduction	6
7	Objective/Scope of Analysis	7
8	Source of dataset	8
9	ETL Process	9 – 10
10	Analysis on dataset	11 – 25
11	Objective 1 (Olympic View)	11 – 13
12	Objective 2 (Country View)	14 – 16
13	Objective 3 (Sports View)	16 – 19
14	Objective 4 (Age View)	19 – 21
15	Objective 5 (Event View)	21 – 23
16	Objective 6 (India View)	23 – 25
17	List of analysis	26 – 27
18	References	28
19	Bibliography	29

INTRODUCTION

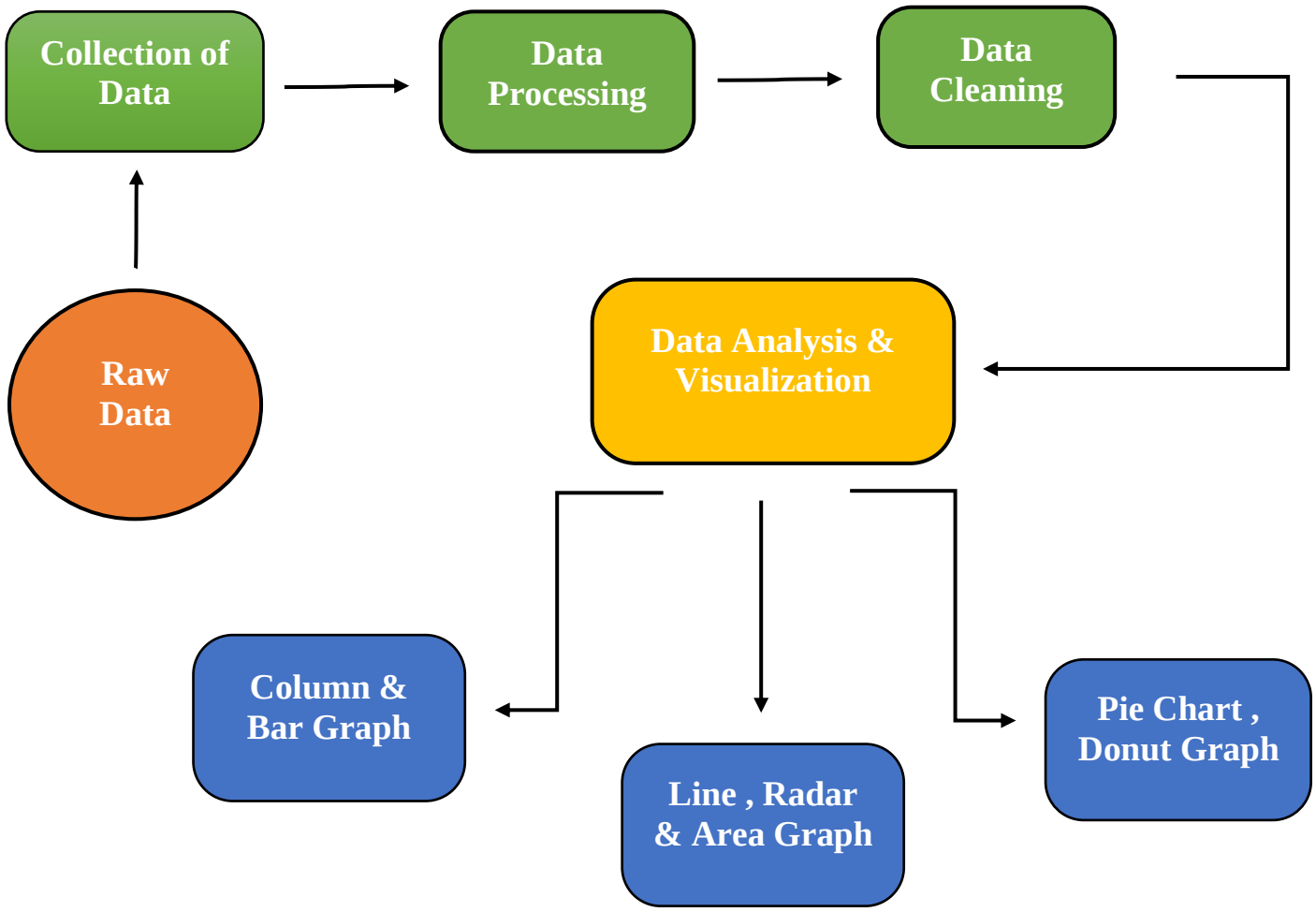
Olympics games are considered as one of the most prime events which provides a valid and common platform for players across different countries to show their talent and skills. Modern Olympic Games were originated by taking inspiration from Ancient Olympic Games held in Olympia, Greece from the 8th Century BC to 4th Century AD. The following timeline outlines the main events in the history of Modern Olympic Games. The Olympics consists of various games (Approximately 45) in which players from various countries (Approx. 205) participate to win a medal for their century. Olympics has a great history of evolution. From 13 participating nations in 1st Olympics (1896) to 207 participating nations in 2016 Rio Olympics, the Olympics have come across a long way.

There are various scenarios which comes in our mind when we look into Evolution of Olympic Games over the years. These scenarios are: Increase in number of participating nations, increase in number of participating Athletes, Increase/Decrease in number of events, increase in the expenditure cost of event, improvement in performance of particular country, improvement in performance of a particular player, Increase in women participation, Participation Ratio of Men to Women, improvement in medication facilities during competition, the effect of pandemic (if any) on the performance of the players. Analysis over these scenarios would depict the evolution of the Olympics over the years. This analysis would help in the future prediction of the number of participating countries, players; winners of various games; Women participation and many more. This type of Analysis can also serve as a performance indicator of a particular country or player.



OBJECTIVE / SCOPE OF ANALYSIS

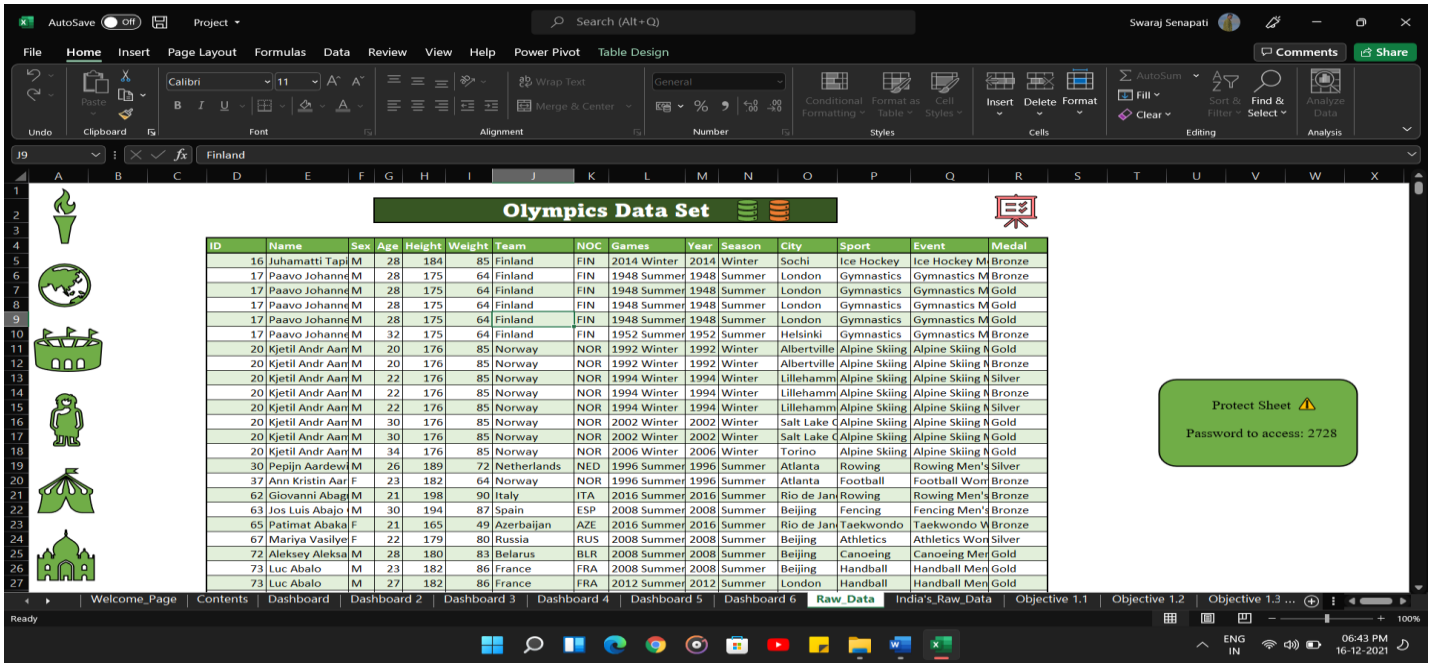
The main objective of this study is to analyze the various factors mentioned above which plays a vital role in the evolution of Olympic Games over the years. The Analysis will include the visualization and explanation of the change in the trends of various factors over the years which help to predict the information of future Olympic Games. As Olympics Games are one of the most important sporting events across the world, each country and each player tries to give their best performance in the event In order to improve their performance, every country should perform such Analysis which would help them in the improvement of their policies and strategies by providing current statistics to them.



SOURCE OF DATASET

Source of Dataset: https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results?select=athlete_events.csv

Olympic Dataset: It is the complete dataset of Olympics from 1896 to 2016.



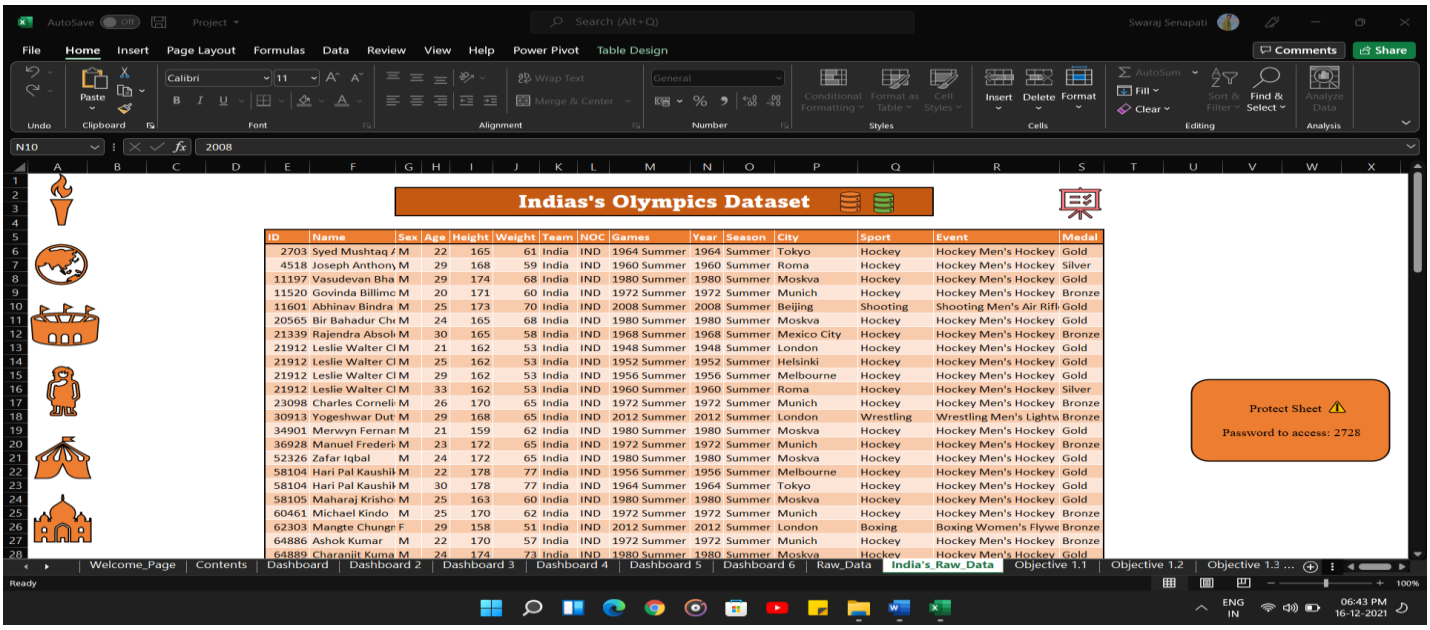
Olympics Data Set

ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal
16	Juhamatti Tapi	M	28	184	85	Finland	FIN	2014 Winter	2014	Winter	Sochi	Ice Hockey	Ice Hockey M	Bronze
17	Paavo Johanne	M	28	175	64	Finland	FIN	1948 Summer	1948	Summer	London	Gymnastics	Gymnastics M	Bronze
17	Paavo Johanne	M	28	175	64	Finland	FIN	1948 Summer	1948	Summer	London	Gymnastics	Gymnastics M	Gold
17	Paavo Johanne	M	28	175	64	Finland	FIN	1948 Summer	1948	Summer	London	Gymnastics	Gymnastics M	Gold
17	Paavo Johanne	M	28	175	64	Finland	FIN	1948 Summer	1948	Summer	London	Gymnastics	Gymnastics M	Gold
17	Paavo Johanne	M	32	175	64	Finland	FIN	1952 Summer	1952	Summer	Helsinki	Gymnastics	Gymnastics M	Bronze
20	Kjetil Andr Aan	M	20	176	85	Norway	NOR	1992 Winter	1992	Winter	Albertville	Alpine Skiing	Alpine Skiing N	Gold
20	Kjetil Andr Aan	M	20	176	85	Norway	NOR	1992 Winter	1992	Winter	Albertville	Alpine Skiing	Alpine Skiing N	Bronze
20	Kjetil Andr Aan	M	22	176	85	Norway	NOR	1994 Winter	1994	Winter	Lillehamm	Alpine Skiing	Alpine Skiing N	Silver
20	Kjetil Andr Aan	M	22	176	85	Norway	NOR	1994 Winter	1994	Winter	Lillehamm	Alpine Skiing	Alpine Skiing N	Bronze
20	Kjetil Andr Aan	M	22	176	85	Norway	NOR	1994 Winter	1994	Winter	Lillehamm	Alpine Skiing	Alpine Skiing N	Silver
20	Kjetil Andr Aan	M	30	176	85	Norway	NOR	2002 Winter	2002	Winter	Salt Lake C	Alpine Skiing	Alpine Skiing N	Gold
20	Kjetil Andr Aan	M	30	176	85	Norway	NOR	2002 Winter	2002	Winter	Salt Lake C	Alpine Skiing	Alpine Skiing N	Gold
20	Kjetil Andr Aan	M	34	176	85	Norway	NOR	2006 Winter	2006	Winter	Torino	Alpine Skiing	Alpine Skiing N	Gold
30	Pepijn Aardew	M	26	189	72	Netherlands	NED	1996 Summer	1996	Summer	Atlanta	Rowing	Rowing Men's	Silver
37	Ann Kristin Aar	F	23	182	64	Norway	NOR	1996 Summer	1996	Summer	Atlanta	Football	Football Wom	Bronze
62	Giovanni Abagn	M	21	198	90	Italy	ITA	2016 Summer	2016	Summer	Rio de Jan	Rowing	Rowing Men's	Bronze
63	Jos Luis Abajo	M	30	194	87	Spain	ESP	2008 Summer	2008	Summer	Beijing	Fencing	Fencing Men's	Bronze
65	Patimat Abaka	F	21	165	49	Azerbaijan	AZE	2016 Summer	2016	Summer	Rio de Jan	Taekwondo	Taekwondo W	Bronze
67	Mariya Vasilye	F	22	179	80	Russia	RUS	2008 Summer	2008	Summer	Beijing	Athletics	Athletics Wor	Silver
72	Aleksey Aleksa	M	28	180	83	Belarus	BLR	2008 Summer	2008	Summer	Beijing	Canoeing	Canoeing Mer	Gold
73	Luc Abalo	M	23	182	86	France	FRA	2008 Summer	2008	Summer	Beijing	Handball	Handball Men	Gold
73	Luc Abalo	M	27	182	86	France	FRA	2012 Summer	2012	Summer	London	Handball	Handball Men	Gold

Protect Sheet

Password to access: 2728

India Dataset: It is the filter out dataset from the above original dataset for analyzing India in Olympics.



Indias's Olympics Dataset

ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal
2703	Syed Mushtaq A	M	22	165	61	India	IND	1964 Summer	1964	Summer	Tokyo	Hockey	Hockey Men's Hockey	Gold
4518	Joseph Anthony	M	29	168	59	India	IND	1960 Summer	1960	Summer	Roma	Hockey	Hockey Men's Hockey	Silver
11197	Vasudevan Bha	M	29	174	68	India	IND	1980 Summer	1980	Summer	Moskva	Hockey	Hockey Men's Hockey	Gold
11520	Govinda Billimo	M	20	171	60	India	IND	1972 Summer	1972	Summer	Munich	Hockey	Hockey Men's Hockey	Bronze
11601	Abhinav Bindra	M	25	173	70	India	IND	2008 Summer	2008	Summer	Beijing	Shooting	Shooting Men's Air Rifl	Gold
20565	Bir Bahadur Ch	M	24	165	68	India	IND	1980 Summer	1980	Summer	Moskva	Hockey	Hockey Men's Hockey	Gold
21339	Rajendra Absol	M	30	165	58	India	IND	1968 Summer	1968	Summer	Mexico City	Hockey	Hockey Men's Hockey	Bronze
21912	Leslie Walter Cl	M	21	162	53	India	IND	1948 Summer	1948	Summer	London	Hockey	Hockey Men's Hockey	Gold
21912	Leslie Walter Cl	M	25	162	53	India	IND	1952 Summer	1952	Summer	Helsinki	Hockey	Hockey Men's Hockey	Gold
21912	Leslie Walter Cl	M	29	162	53	India	IND	1956 Summer	1956	Summer	Melbourne	Hockey	Hockey Men's Hockey	Gold
21912	Leslie Walter Cl	M	33	162	53	India	IND	1960 Summer	1960	Summer	Roma	Hockey	Hockey Men's Hockey	Silver
23098	Charles Cornell	M	26	170	65	India	IND	1972 Summer	1972	Summer	Munich	Hockey	Hockey Men's Hockey	Bronze
30913	Yogeshwar Dut	M	29	168	65	India	IND	2012 Summer	2012	Summer	London	Wrestling	Wrestling Men's Lightw	Bronze
34901	Merwyn Fernar	M	21	159	62	India	IND	1980 Summer	1980	Summer	Moskva	Hockey	Hockey Men's Hockey	Gold
36928	Manuel Frederi	M	23	172	65	India	IND	1972 Summer	1972	Summer	Munich	Hockey	Hockey Men's Hockey	Bronze
52326	Zafar Iqbal	M	24	172	65	India	IND	1980 Summer	1980	Summer	Moskva	Hockey	Hockey Men's Hockey	Gold
58104	Hari Pal Kaushi	M	22	178	77	India	IND	1956 Summer	1956	Summer	Melbourne	Hockey	Hockey Men's Hockey	Gold
58104	Hari Pal Kaushi	M	30	178	77	India	IND	1964 Summer	1964	Summer	Tokyo	Hockey	Hockey Men's Hockey	Gold
58105	Maharaj Krisho	M	25	163	60	India	IND	1980 Summer	1980	Summer	Moskva	Hockey	Hockey Men's Hockey	Gold
60461	Michael Kindo	M	25	170	62	India	IND	1972 Summer	1972	Summer	Munich	Hockey	Hockey Men's Hockey	Bronze
62303	Mangte Chungn	F	29	158	51	India	IND	2012 Summer	2012	Summer	London	Boxing	Boxing Women's Flywe	Bronze
64886	Ashok Kumar	M	22	170	57	India	IND	1972 Summer	1972	Summer	Munich	Hockey	Hockey Men's Hockey	Bronze
64889	Charanjit Kuma	M	24	174	73	India	IND	1980 Summer	1980	Summer	Moskva	Hockey	Hockey Men's Hockey	Gold

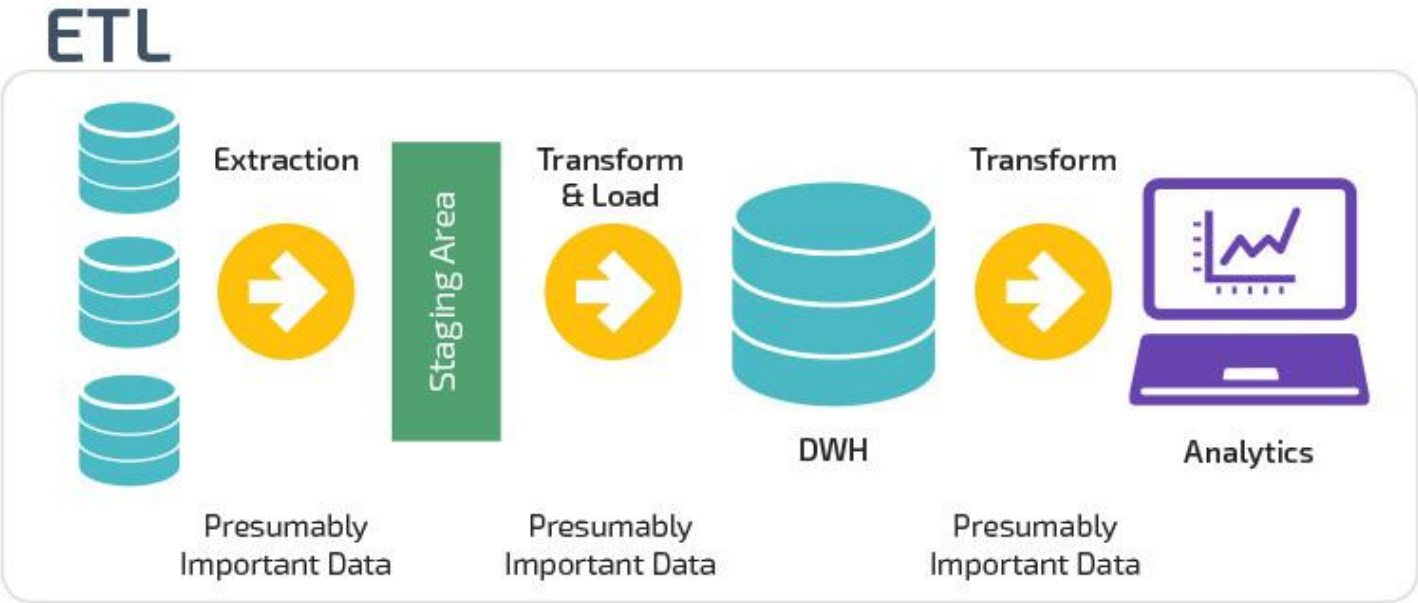
Protect Sheet

Password to access: 2728

ETL PROCESS

Data Wearhouse: Data Warehouse is a relational database management system (RDBMS) construct to meet the requirement of transaction processing systems. It can be loosely described as any centralized data repository which can be queried for business benefits. It is a database that stores information oriented to satisfy decision-making requests. It is a group of decision support technologies, targets to enabling the knowledge worker (executive, manager, and analyst) to make superior and higher decisions. So, Data Warehousing support architectures and tool for business executives to systematically organize, understand and use their information to make strategic decisions.

ETL is a process in Data Warehousing, and it stands for Extract, Transform and Load. It is a process in which an ETL tool extracts the data from various data source systems, transforms it in the staging area, and then finally, loads it into the Data Warehouse system.



Detailed concepts or analyze on the steps of ETL process are:

Extraction:

The first step of the ETL process is extraction. In this step, data from various source systems is extracted which can be in various formats like relational databases, No SQL, XML, and flat files into the staging area. It is important to extract the data from various source systems and store it into the staging area first and not directly into the data warehouse because the extracted data is in various formats and can be corrupted also. Hence loading it directly into the data warehouse may damage it and rollback will be much more difficult. Therefore, this is one of the most important steps of ETL process.

Transformation:

The second step of the ETL process is transformation. In this step, a set of rules or functions are applied on the extracted data to convert it into a single standard format. It may involve following processes/tasks:

Filtering – loading only certain attributes into the data warehouse.

Cleaning – filling up the NULL values with some default values, mapping U.S.A, United States, and America into USA, etc.

Joining – joining multiple attributes into one.

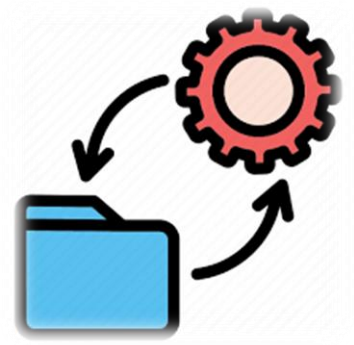
Splitting – splitting a single attribute into multiple attributes.

Sorting – sorting tuples based on some attribute (generally key-attribute).

Loading:

The third and final step of the ETL process is loading. In this step, the transformed data is finally loaded into the data warehouse. Sometimes the data is updated by loading into the data warehouse very frequently and sometimes it is done after longer but regular intervals. The rate and period of loading solely depends on the requirements and varies from system to system.

ETL Tools: Most commonly used ETL tools are Sybase, Oracle Warehouse builder, Clover ETL, and Mark Logic.

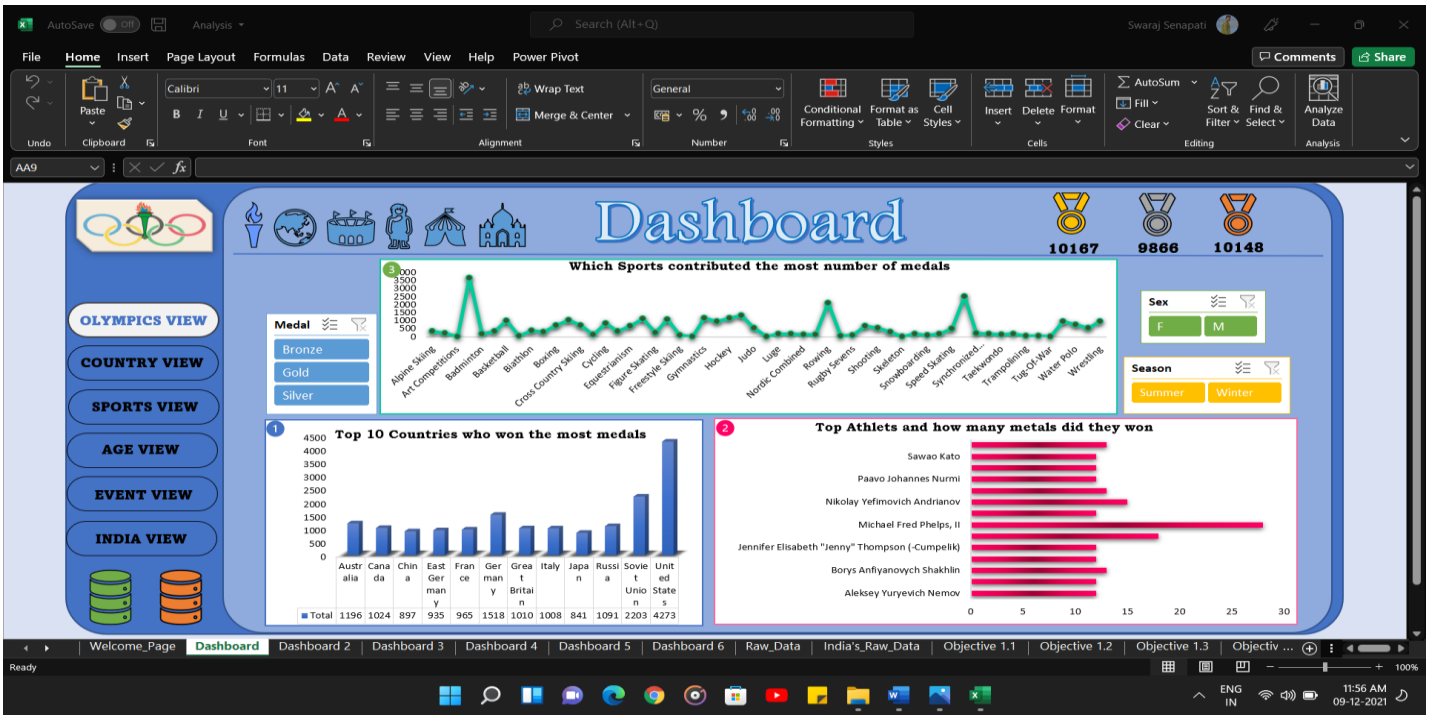


ANALYSIS ON DATASET

In this analysis of Olympic dataset, I have collected the dataset or can raw data from Kaggle. And then examine the data and process for cleaning. Cleaning like removing the unnecessary rows and column and also the NA values that are present in the dataset by the help of filter and sort functions based upon the requirement of the analysis. Then the analysis is divided into 6 parts or objective in the dashboards. The six main objective and their sub objectives are explained and visualized with their results below:

1. Olympic View:

In this part I have analysis the Olympic view of the dataset. I have made a dashboard containing three sub objectives in this and the count of medals i.e., Gold, Sliver & Bronze. Also creating some icon with proper hyperlinks to navigate the sheets. Also included different types of charts and slicers.

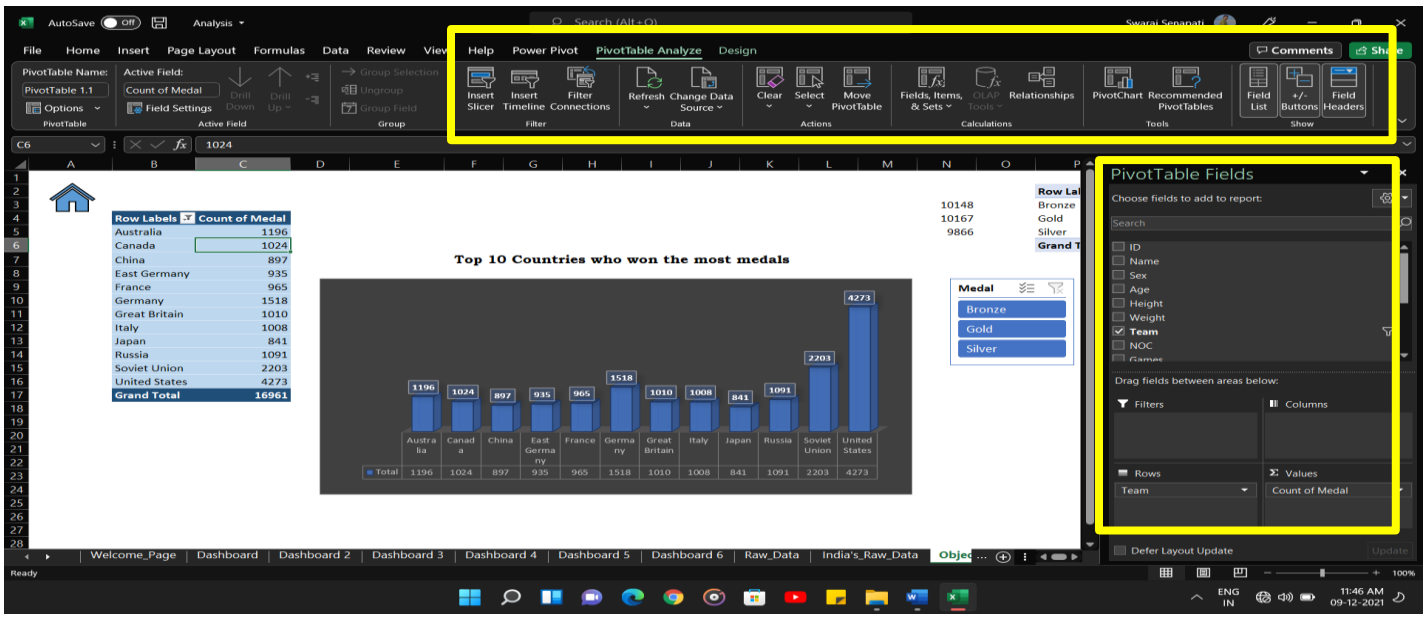


The sub objectives for the Olympics view are:

1.1. Top countries who have won most number of medals.

In this part I have analyze in Olympic who are the top countries medals wise and how many medals these countries have won. First, I have created the pivot table from insert ribbon and analysis in the table format and then by the help of slicer I have join the sex, medal & season box with the table so that we can find how many males or females athletes got gold, silver & bronze medals in a particular summer or winter season of Olympics. And then visualize the data with the column graph.

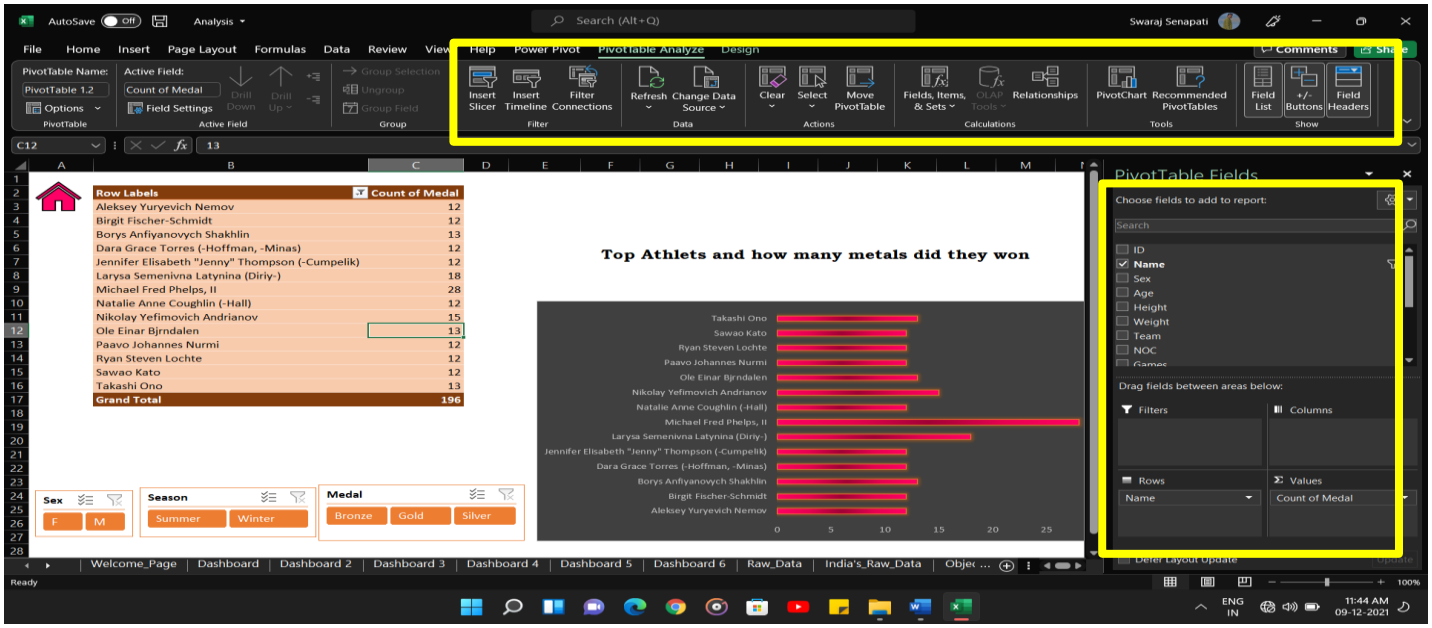
Result: As we can see from the analysis that, United States is leading in count od medals i.e., 4273 overall.



1.2. Top Athletes in Olympics and how many medals they have won.

In this part I have analyze who are the top athletes by names and how many medals they have won. First, I have created the pivot table from insert ribbon and analysis in the table format and then by the help of slicer I have join the sex, medal & season box with the table so that we can find who are the males or females' athletes that got gold, silver & bronze medals in a particular summer or winter season of Olympics. And then visualize the data with the bar graph.

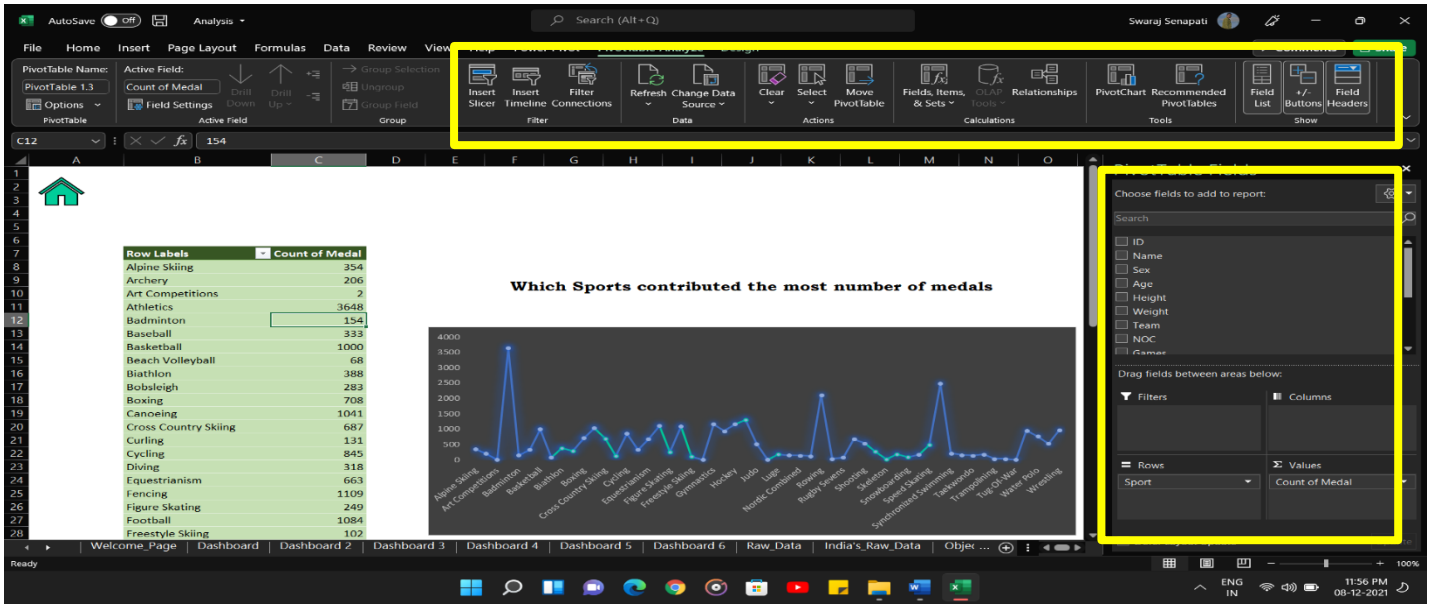
Result: As we can see from the analysis that, Michael Fred Phelps, II had won most medals i.e.,28 overall.



1.3. Which sports contributed most number of medals.

In this part I have analyze which sports is contributing a greater number of medals in the Olympic games. First, I have created the pivot table from insert ribbon and analysis in the table format and then by the help of slicer I have join the sex, medal & season box with the table so that we can find how many males or females' athletes that got gold, silver & bronze medals in a particular summer or winter season of Olympics in a particular sport. And then visualize the data with the line graph.

Result: As we can see from the analysis that, Athletics is leading the list of sports and contributed more number of medals i.e., 3648 overall.



2. Country View:

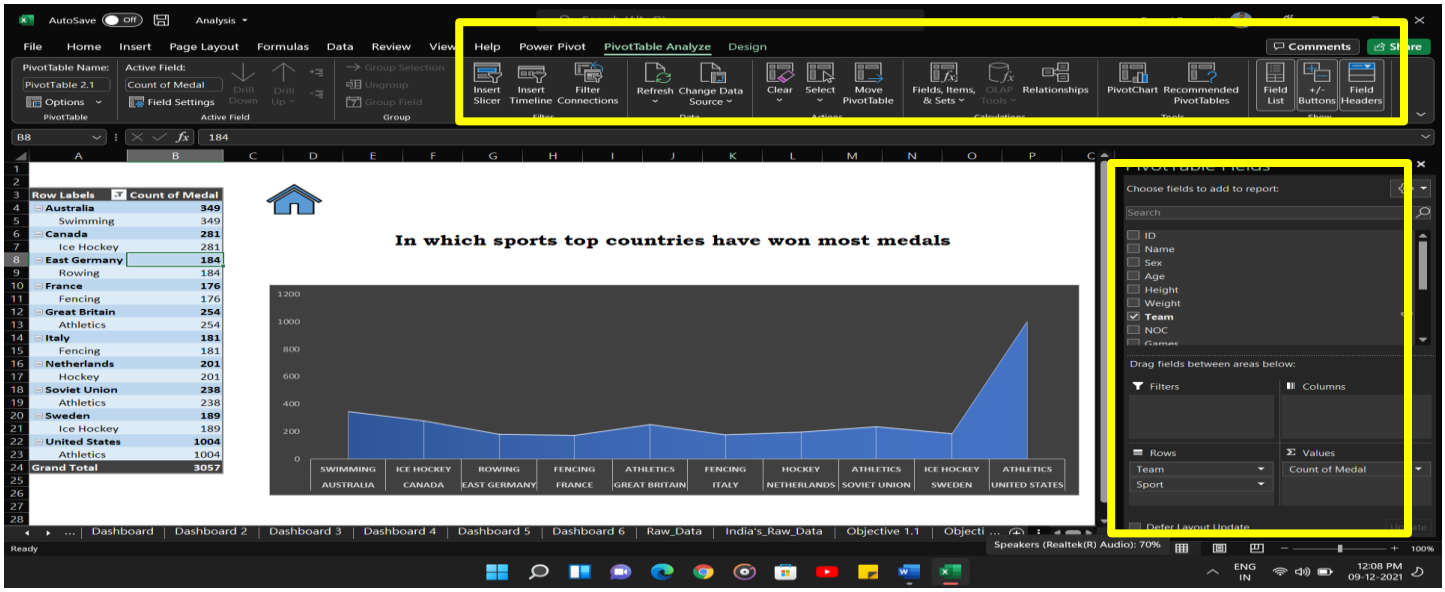
In this part I have analysis the country view of the dataset. I have made a dashboard containing three sub objectives in this and the count of medals i.e., Gold, Sliver & Bronze. Also creating some icon with proper hyperlinks to navigate the sheets. Also included different types of charts and slicers.

The sub objectives for the Olympics view are:

2.1. In which sports top countries have won most number of medals.

In this part I have analyze the top countries and the sports by the count of medals. First, I have created the pivot table from insert ribbon and analysis in the table format and then by the help of slicer I have join the sex & season box with the table so that we can find how many males or females' athletes that got medals in a particular summer or winter season of Olympics in a particular sport. And then visualize the data with the area graph.

Result: As we can see from the analysis that, United States is leading in Athletics by a margin of 1004 medals overall.



2.2. Who are the top athletes in the country in the count of medals.

In this part I analyze the list of top athletes in their countries and how many medals they have won. First, I have created the pivot table from insert ribbon and analysis in the table format and then by the help of slicer I have join the sex & season box with the table so that we can find how many males or females' athletes that got medals in a particular summer or winter season of Olympics for their country. And then visualize the data with the donut graph.

Result: As we can see from the analysis that, Michael Fred Phelps, II from United States toped in the count of medals for a country i.e., 28 overall.

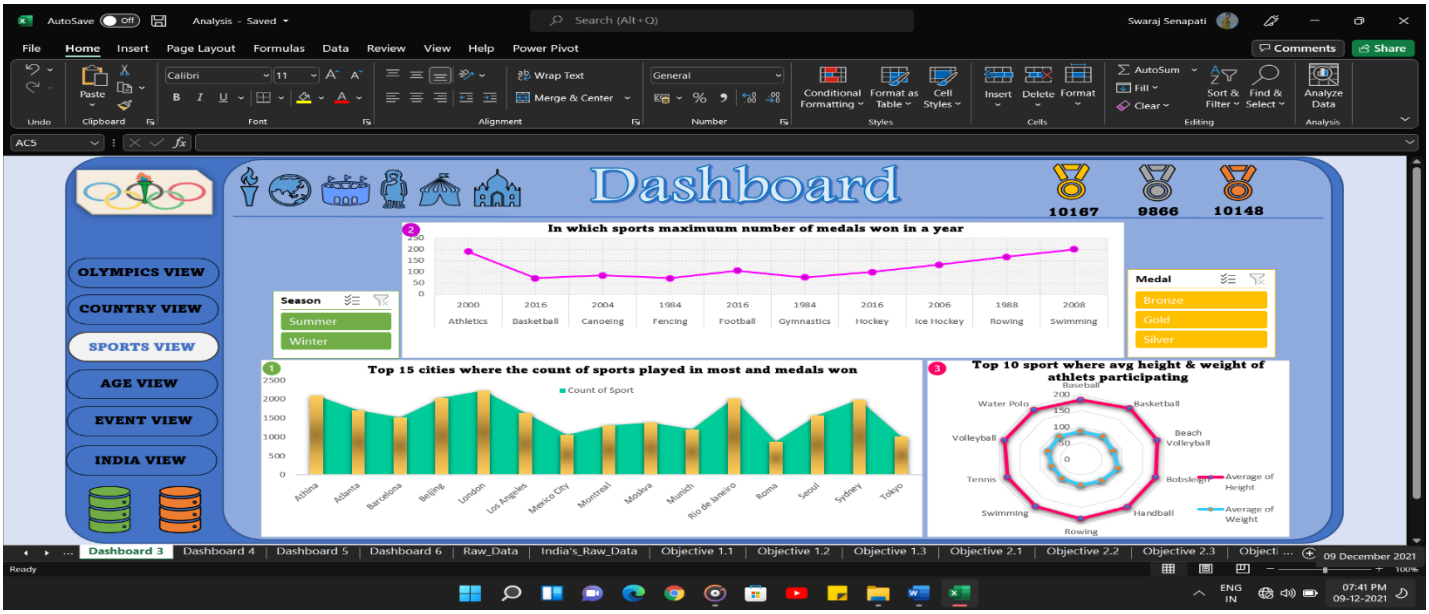
2.3. Top countries where maximum count of sports played & average number of players participated.

In this part I have analyze the top countries where the maximum count of sports played & average number of players/athletes participated. First, I have created the pivot table from insert ribbon and analysis in the table format and then by the help of slicer I have join the season box with the table so that we can find which country in a particular summer or winter season of Olympics played the sport and the average no. of athletes participates. And then visualize the data with the combo graph.

Result: As we can see from the analysis that the count of sports played is maximum in United States i.e., 4273 and the average number of athletes participated is from Soviet Union i.e., 74054 overall.

3. Sports View:

In this part I have analyze the sports view of the dataset. I have made a dashboard containing three sub objectives in this and the count of medals i.e., Gold, Sliver & Bronze. Also creating some icon with proper hyperlinks to navigate the sheets. Also included different types of charts and slicers.



The sub objectives for the Olympics view are:

3.1. Top 15 cities where the count of sports played is most and medals won.

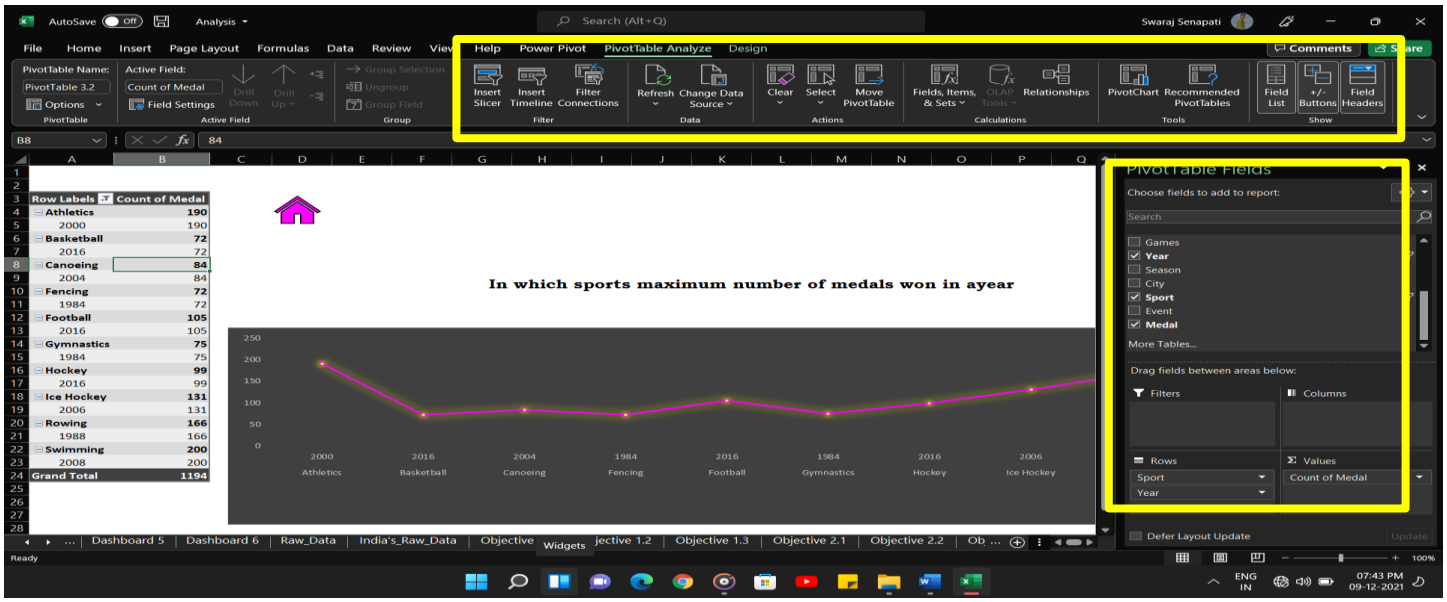
In this part I have analyze the top 15 cities in Olympics where the sports played is the most and medals won. First, I have created the pivot table from insert ribbon and analysis in the table format and then by the help of slicer I have join the season box with the table so that we can find which country in a particular summer or winter season of Olympics and who are top cities where most number of sports played and medals won. And then visualize the data with the combo graph.

Result: As we can see from the analysis that in London most number of sports played and medals won i.e., 2231 overall.

3.2. In which sports maximum number of medals won in a year.

In this part I have analyze which sports is contributing a greater number of medals in the Olympic games. First, I have created the pivot table from insert ribbon and analysis in the table format and then by the help of slicer I have join the medal & season box with the table so that we can find how many athletes that got gold, silver & bronze medals in a particular summer or winter season of Olympics in a particular sport in a year. And then visualize the data with the line graph.

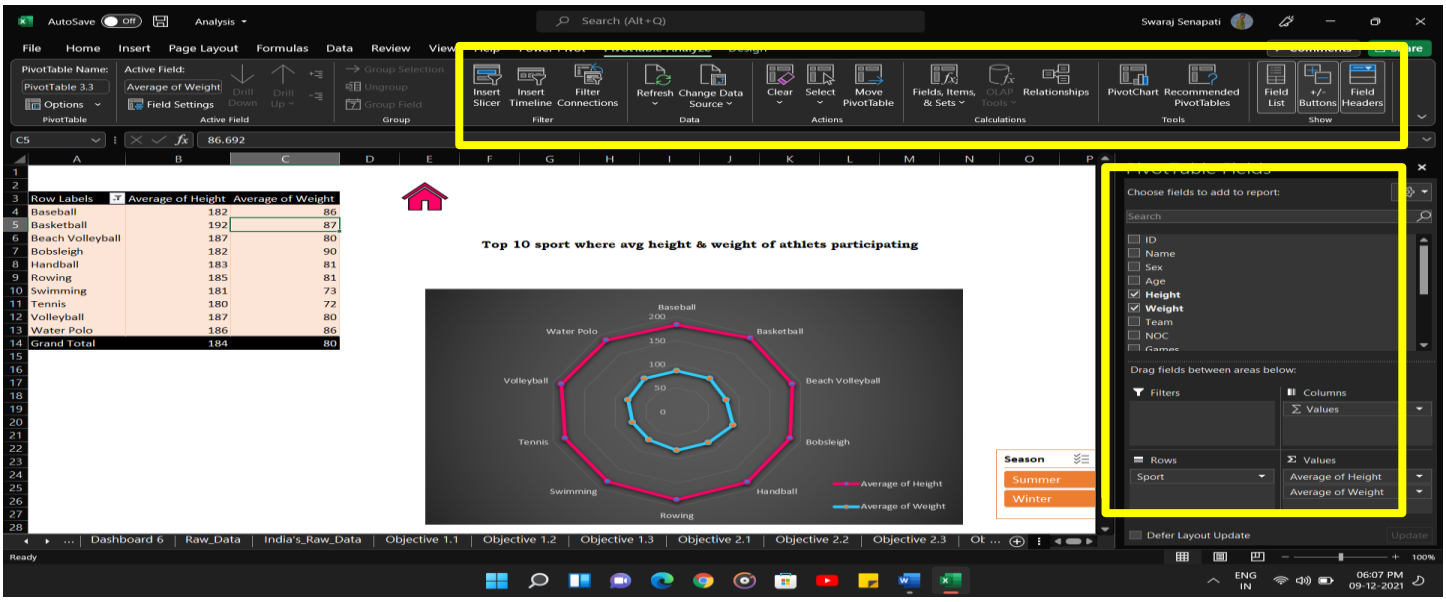
Result: As we can see from the analysis that in 2008 Swimming had acquire most number of medals i.e., 200 overall.



3.3. Top 10 sport where avg height & weight of athletes participating.

In this part I have analyze the top 10 sports where the average height and weight of athletes participating. First, I have created the pivot table from insert ribbon and analysis in the table format and then by the help of slicer I have join the medal & season box with the table so that we can find how many athletes that got gold, silver & bronze medals in a particular summer or winter season of Olympics the average height and weight of athlete played the sports. And then visualize the data with the radar graph.

Result: As we can see from the analysis that avg. height of athletes are from Basketball & avg. weight of athletes are from Bobsleigh i.e., 192 & 90 overall.



4. Age View:

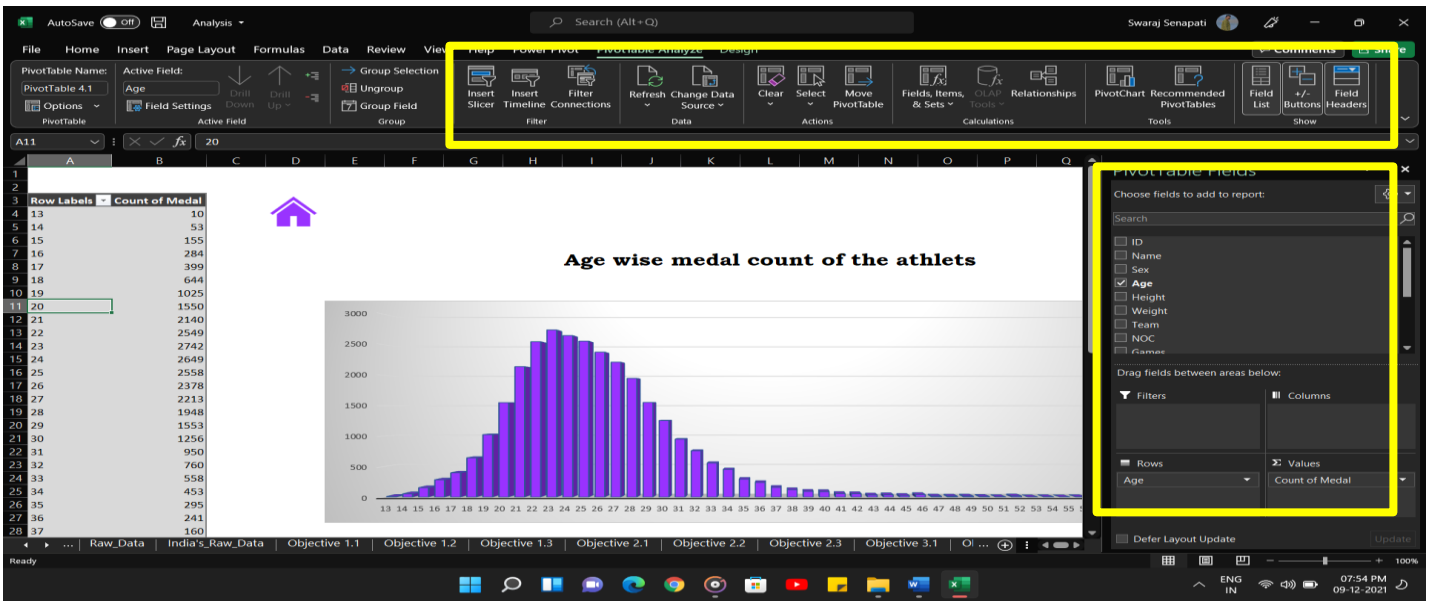
In this part I have analyze the age view of the dataset. I have made a dashboard containing three sub objectives in this and the count of medals i.e., Gold, Sliver & Bronze. Also creating some icon with proper hyperlinks to navigate the sheets. Also included different types of charts and slicers.

The sub objectives for the Olympics view are:

4.1. Age wise medal count of the athletes.

In this part I have analyze the age wise medals count of the athletes consisting of male & female. First, I have created the pivot table from insert ribbon and analysis in the table format and then by the help of slicer I have join the medal & sex box with the table so that we can find how many male and female athletes that got gold, silver & bronze medals in Olympics games. And then visualize the data with the column graph.

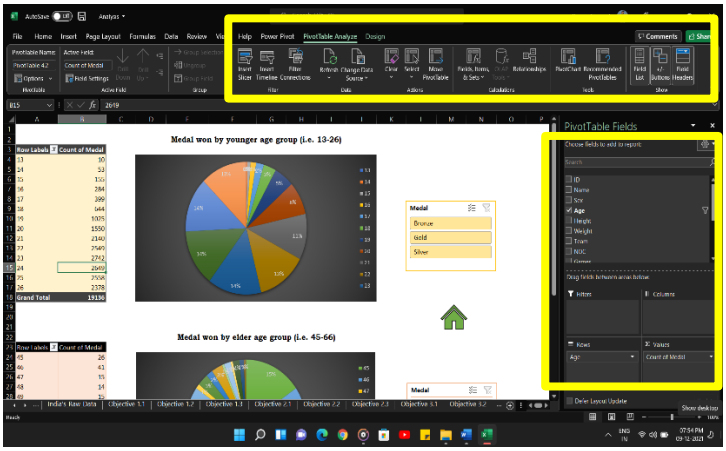
Result: As we can see from the analysis that, 23 years age group of athletes acquire most number of medals i.e., 2742 overall.

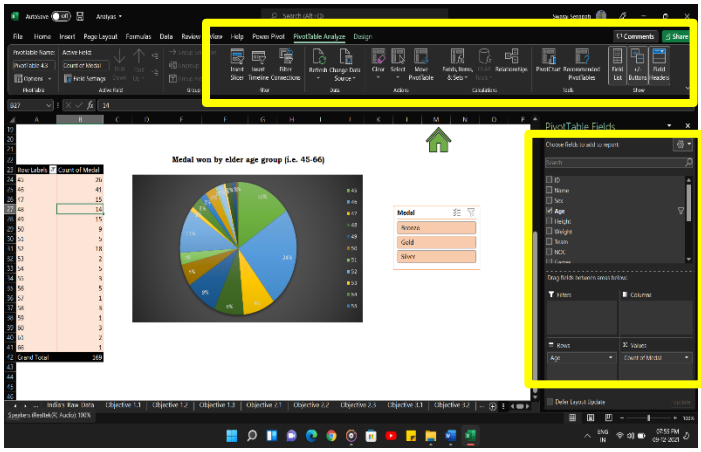


4.2. Medals won by young & elder age groups.

In this part I have analyze the different age group that getting most number of medals. It is divided into two parts young age group i.e., from 13-26 & elder age group i.e., from 45-66. First, I have created the pivot table from insert ribbon and analysis in the table format and then by the help of slicer I have join the medal & season box with the table so that we can find how many male and female athletes that got gold, silver & bronze medals in Olympics games. And then visualize the data with the pie chart.

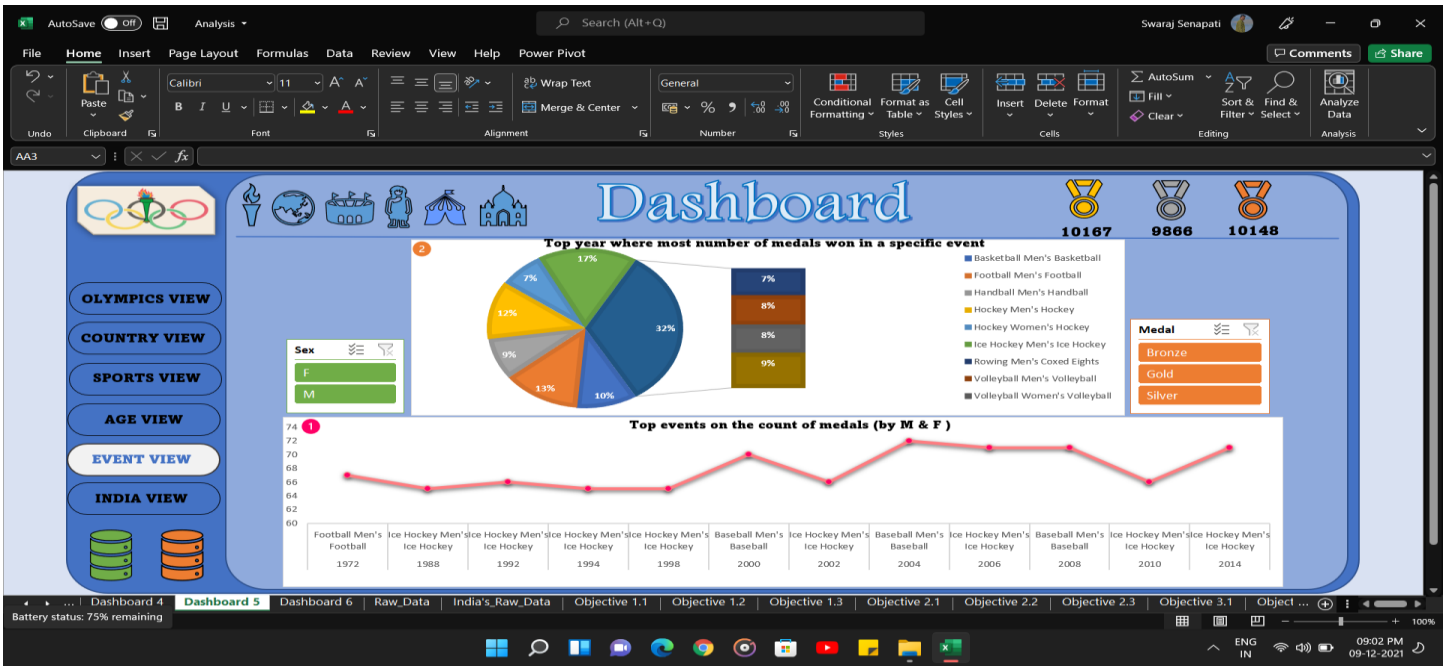
Result: As we can see from the analysis that medal won by younger age group are higher overall.





5. Event View:

In this part I have analyze the event view of the dataset. I have made a dashboard containing three sub objectives in this and the count of medals i.e., Gold, Sliver & Bronze. Also creating some icon with proper hyperlinks to navigate the sheets. Also included different types of charts and slicers.

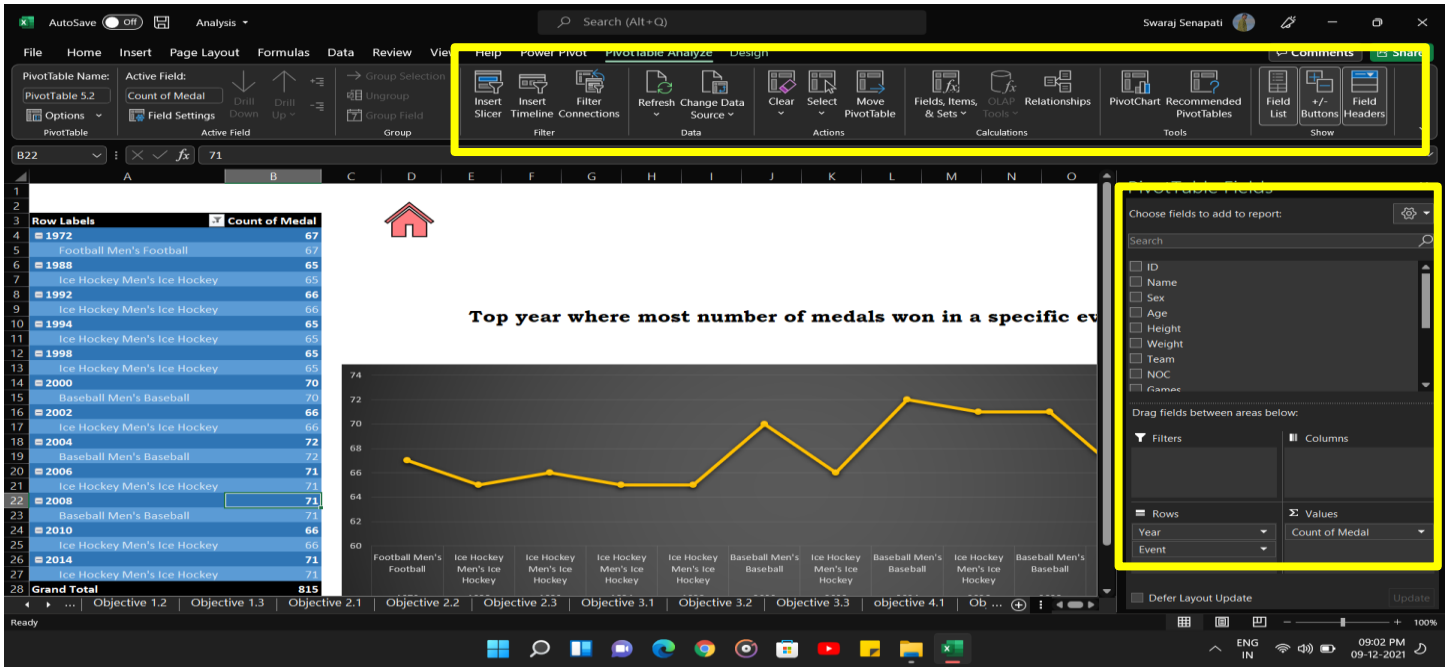


The sub objectives for the Olympics view are:

5.1. Top events on the count of medals (by Male & Female).

In this part I have analyze the top event on the count of male and female in a particular year of Olympics. First, I have created the pivot table from insert ribbon and analysis in the table format and then by the help of slicer I have join the sex & medals box with the table so that we can find how many male and female athletes that got gold, silver & bronze medals in Olympics games and which are the top events and in that particular year. And then visualize the data with the line graph.

Result: As we can see from the analysis that in Men's Hockey maximum number of medals won i.e.,1001 overall.



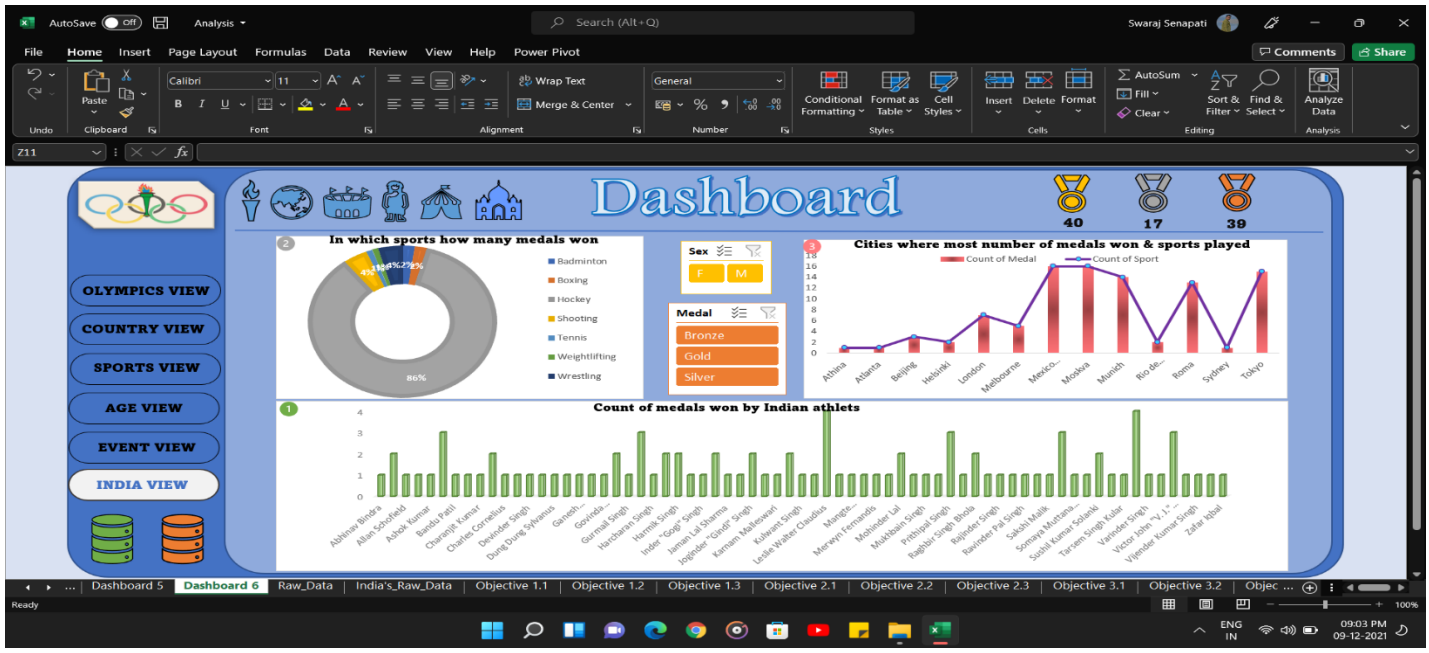
5.2. Top years where most number of medals won in a specific event.

In this part I have analyze the top years where most numbers of medal won in a specific event of Olympics. First, I have created the pivot table from insert ribbon and analysis in the table format and then by the help of slicer I have join the sex & medals box with the table so that we can find how many male and female athletes that got gold, silver & bronze medals in Olympics games and which are the top years where the specific event held. And then visualize the data with the 2D pie chart.

Result: As we can see from the analysis that in 2004 Men's Baseball had won maximum medals i.e., 72 overall.

6. India View:

In this part I have analyze the India view of the dataset. First by filter out the data of India and make it to be analyze and then I have made a dashboard containing three sub objectives in this and the count of medals i.e., Gold, Sliver & Bronze. Also creating some icon with proper hyperlinks to navigate the sheets. Also included different types of charts and slicers.

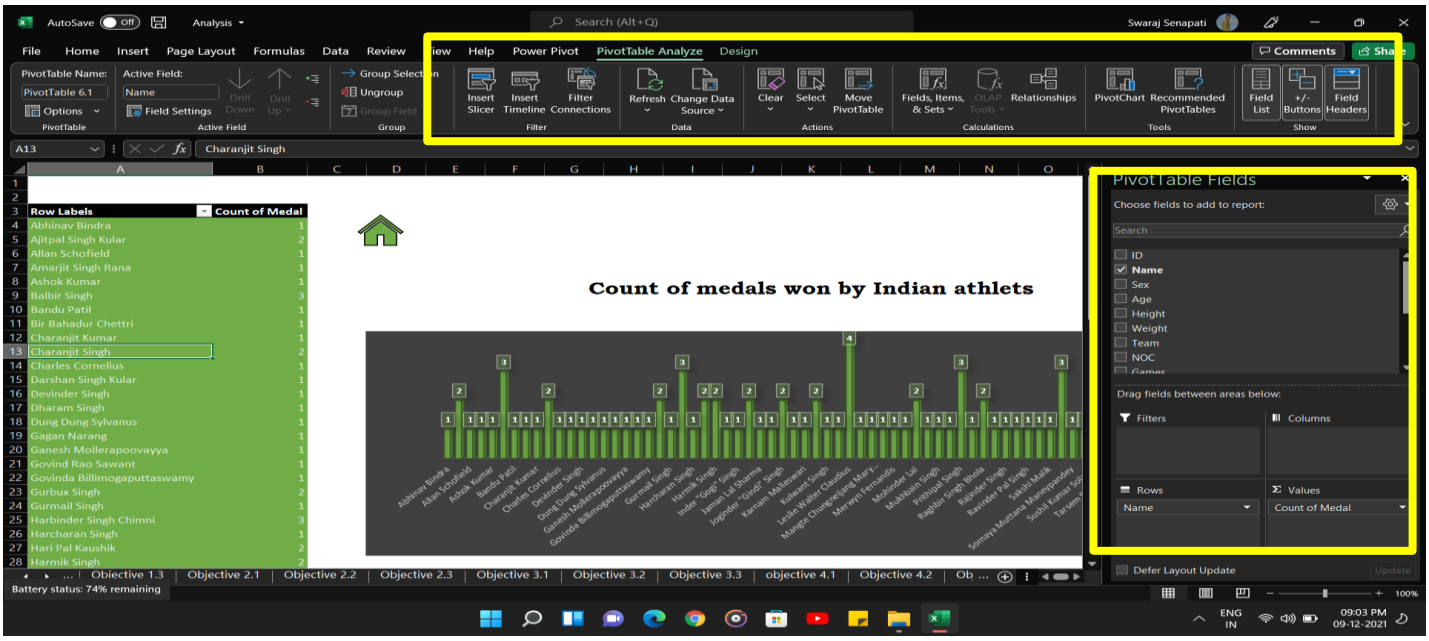


The sub objectives for the Olympics view are:

6.1. Count of medals won by Indian athletes.

In this part I have analyze the count of medals won by Indian athletes by their names. First, I have created the pivot table from insert ribbon and analysis in the table format and then by the help of slicer I have join the sex & medal box with the table so that we can find how many males or females athletes got gold, silver & bronze medals Olympics. And then visualize the data with the column graph.

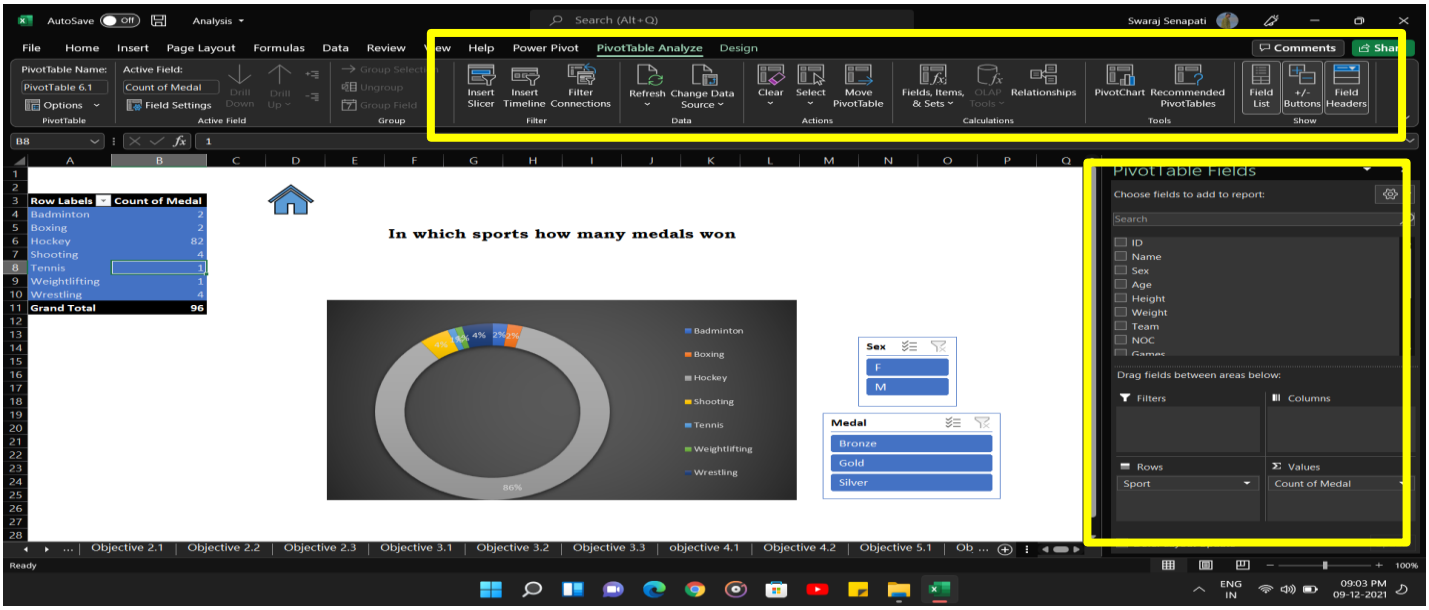
Result: As we can see from the analysis that Leslie Walter Claudius & Undam Singh Kolar had won most number of medals i.e., 4 overall.



6.2. In which sports how many medals won by Indian athletes.

In this part I have analyze which sports in contributing more number of medals to the Indian athletes. First, I have created the pivot table from insert ribbon and analysis in the table format and then by the help of slicer I have join the sex & medal box with the table so that we can find how many males or females athletes got gold, silver & bronze medals Olympics and which sport in contributing more number of medals. And then visualize the data with the donut graph.

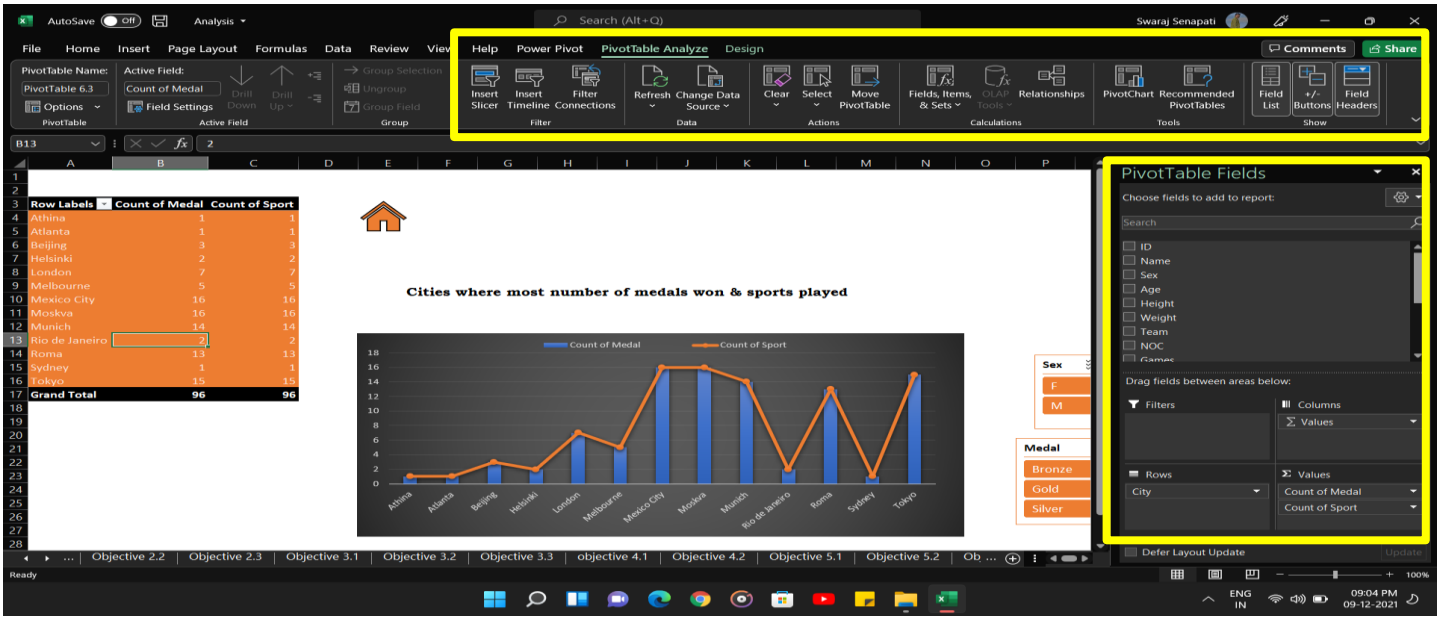
Result: As we can see from the analysis that, Hockey is contributing maximum medals i.e., 82 overall.



6.3. Cities where most number of medals won & match played by Indian athletes.

In this part I have analyze which are the cities where most number of medals won and match played by Indian athletes. First, I have created the pivot table from insert ribbon and analysis in the table format and then by the help of slicer I have join the sex & medal box with the table so that we can find how many males or females athletes got gold, silver & bronze medals Olympics and which city in contributing more number of medals. And then visualize the data with the combo graph.

Result: As we can see from the analysis that in Mexico City & Moskva most number of sports played and medals won i.e., 16 overall.



LIST OF ANALYSIS

In this project or can say analysis of Olympic data set I have used 6 analysis or objectives containing 6 different dashboards with proper formatting, charts, hyperlinks, icons etc. So, the list of analysis are:

1. Olympic View

1.1. Top countries who have won most number of medals.

1.2. Top Athletes in Olympics and how many medals they have won.

1.3. Which sports contributed most number of medals.

2. Country View

2.1. In which sports top countries have won most number of medals.

2.2. Who are the top athletes in the country in the count of medals.

2.3. Top countries where maximum count of sports played & average number of players participated.

3. Sports View

3.1. Top 15 cities where the count of sports played is most and medals won.

3.2. In which sports maximum number of medals won in a year.

3.3. Top 10 sport where avg height & weight of athletes participating.

4. Age View

4.1. Age wise medal count of the athletes.

4.2. Medals won by young (i.e., 15-26) & elder (i.e., 45-66) age groups.

5. Event View

5.1. Top events on the count of medals (by Male & Female).

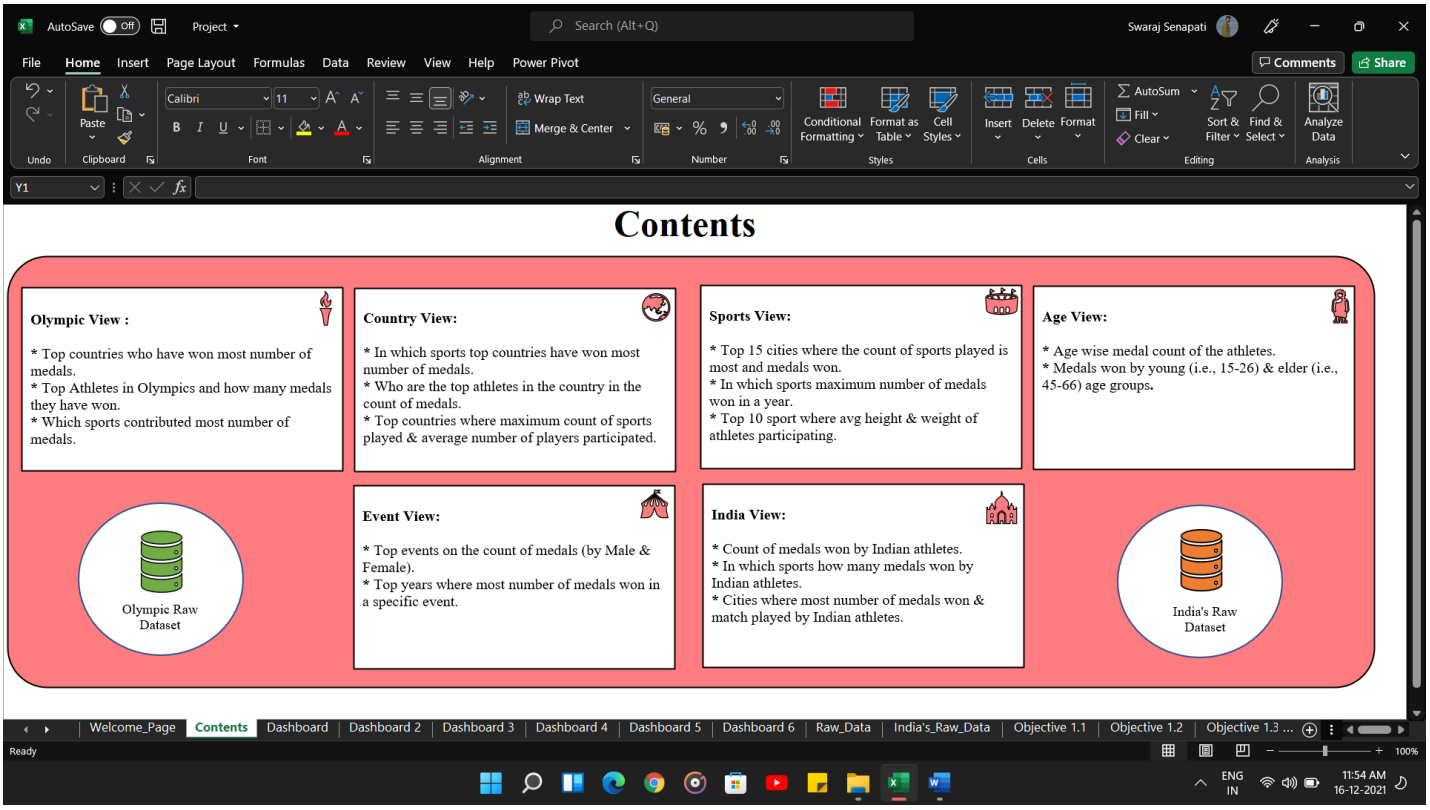
5.2. Top years where most number of medals won in a specific event.

6. India View

6.1. Count of medals won by Indian athletes.

6.2. In which sports how many medals won by Indian athletes.

6.3. Cities where most number of medals won & match played by Indian athletes.



REFERENCES

For completing this project or analysis I have refer from:

- The ppts that we have received by our section teachers that help in the building concepts for the analysis.
- I have also referred some video lecture from Tutorials Point

  The playlist from the lecture I have referred:

  https://youtube.com/playlist?list=PLWPirh4EWFpEpO6NjjWLbKSCb-wx3hMql
- I have also referred from the video recording of our class that was taught by our concerned faculty.
- For ETL process I have referred from Geeks for Geeks website for further concepts.

  Link: https://www.geeksforgeeks.org/etl-process-in-data-warehouse/

BIBLIOGRAPHY

- Source of Dataset

https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results?select=athlete_events.csv

- Links regarding analysis

https://www.kaggle.com/benhamner/starter-120-years-of-olympic-history-e5a32948-3

https://www.kaggle.com/dmauricio/starter-120-years-of-olympic-history-f04377fb-2

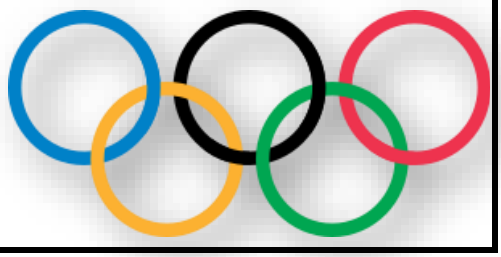
- Links regarding report creation

https://towardsdatascience.com/120-years-of-olympic-games-56411bc4bd53

https://www.geeksforgeeks.org/etl-process-in-data-warehouse/

- Links of images used in the report

https://olympics.com/ioc/olympic-rings

https://panoply.io/data-warehouse-guide/etl-tutorial/

https://svitla.com/blog/data-warehouse-vs-database