OLYMPICS DATA ANALYSIS 2024

DEPI | 2024

ABSRTACT

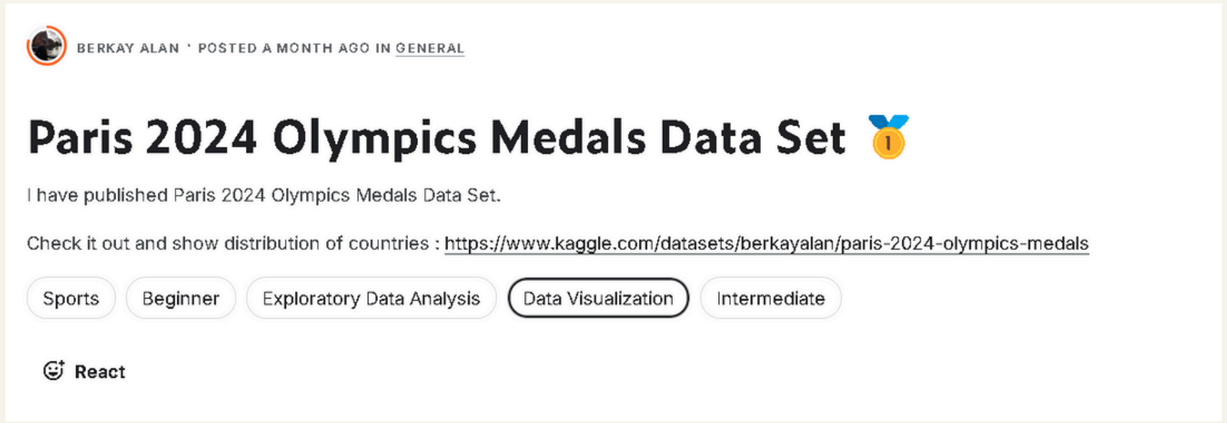
This project presents a simple analysis of (Olympic 2024) data to uncover trends and insights related to athlete country participation, and medal distribution. Using publicly available datasets, the data is cleaned and analyzed to explore key aspects such as medals, country rankings, gender participation,. Visualizations such as bar charts, dounat chart , and maps help illustrate findings,The project provides a clear, data-driven view of the Olympic Games' progression and its global impact.

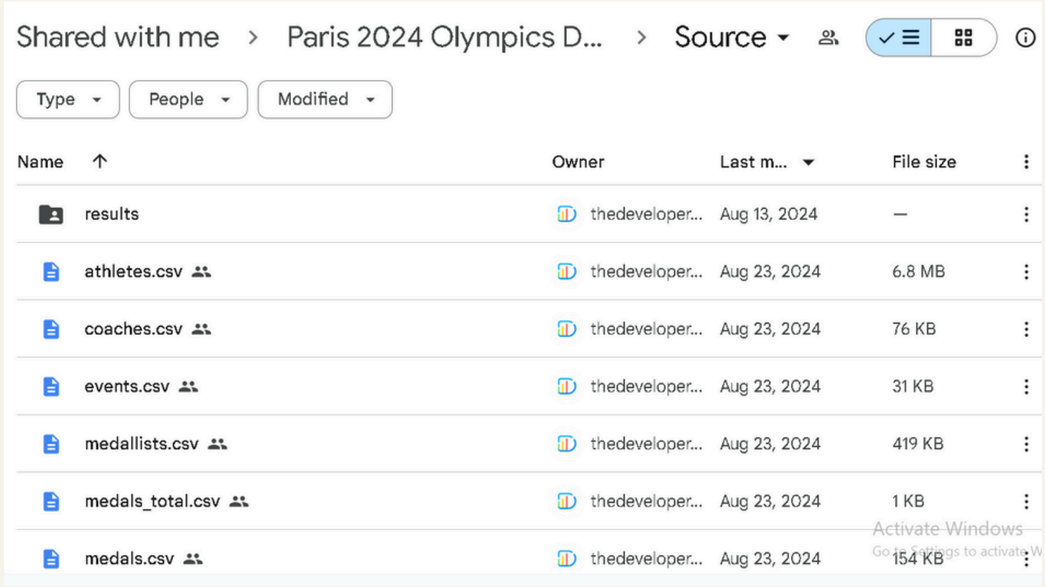
OVERVIEW

- Data Collection
- Questions
- tools
- Data Cleaning
- Visualization
- Thank you
- Transformaing
- Conclusion & Insights

DATA COLLECTION

Used publicly available datasets such as those International Olympic from kaggle , google drive





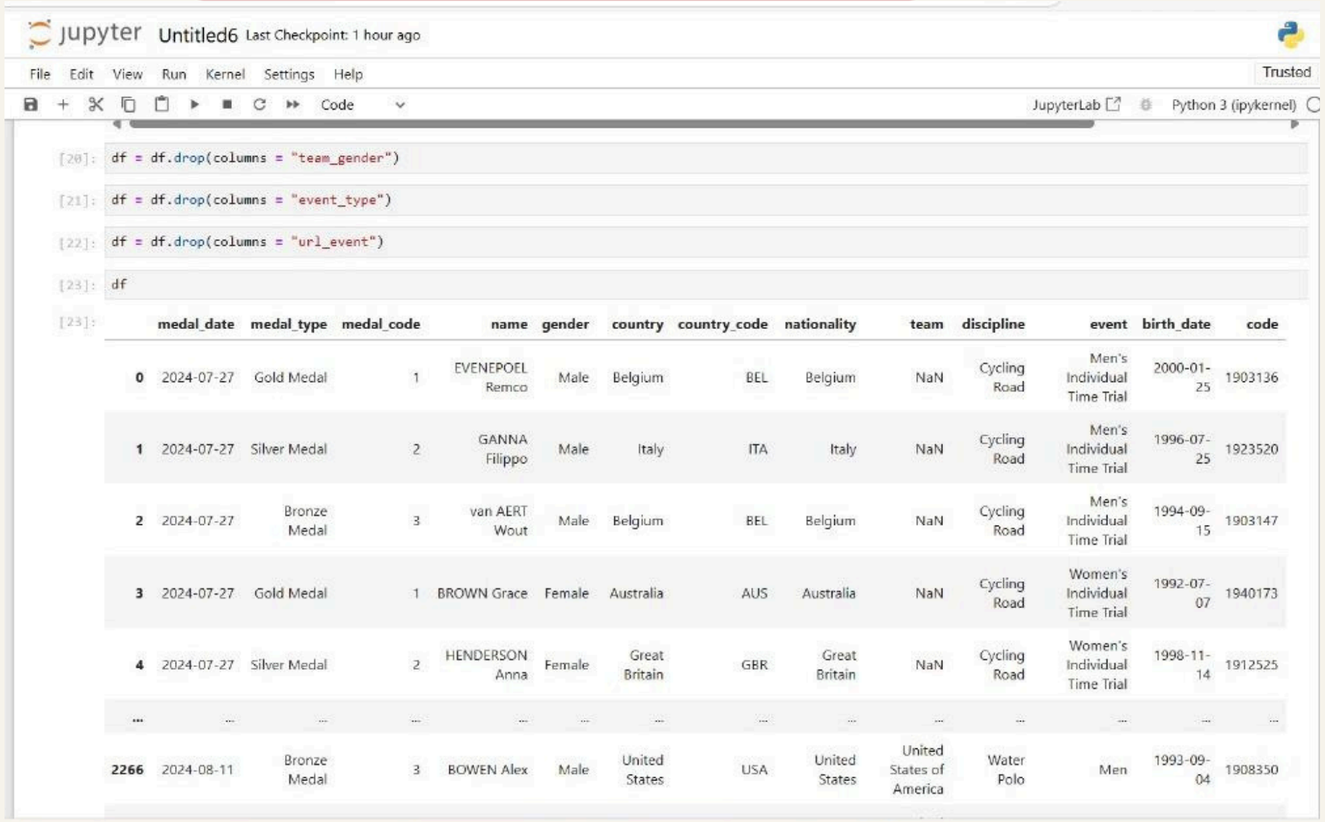
finally we used this dataset

Data cleaning

We faced several problems with the dataset,errors ,duplication, missing columns and Unimportant columns we dicide to transform this dataset using python

First Problem

Unimportant columns



jupyter Untitled6 Last Checkpoint: 1 hour ago

File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab

Python 3 (ipykernel)

```
[20]: df = df.drop(columns = "team_gender")
[21]: df = df.drop(columns = "event_type")
[22]: df = df.drop(columns = "url_event")
[23]: df
```

[23]:

	medal_date	medal_type	medal_code	name	gender	country	country_code	nationality	team	discipline	event	birth_date	code
0	2024-07-27	Gold Medal	1	EVENEPOEL Remco	Male	Belgium	BEL	Belgium	NaN	Cycling Road	Men's Individual Time Trial	2000-01-25	1903136
1	2024-07-27	Silver Medal	2	GANNA Filippo	Male	Italy	ITA	Italy	NaN	Cycling Road	Men's Individual Time Trial	1996-07-25	1923520
2	2024-07-27	Bronze Medal	3	van AERT Wout	Male	Belgium	BEL	Belgium	NaN	Cycling Road	Men's Individual Time Trial	1994-09-15	1903147
3	2024-07-27	Gold Medal	1	BROWN Grace	Female	Australia	AUS	Australia	NaN	Cycling Road	Women's Individual Time Trial	1992-07-07	1940173
4	2024-07-27	Silver Medal	2	HENDERSON Anna	Female	Great Britain	GBR	Great Britain	NaN	Cycling Road	Women's Individual Time Trial	1998-11-14	1912525
...	...	...	...	...	...	...	...	...	...	...	...	...	...
2266	2024-08-11	Bronze Medal	3	BOWEN Alex	Male	United States	USA	United States	United States of America	Water Polo	Men	1993-09-04	1908350

Data cleaning

second Problem

calculate Age

```
[58]: def calculate_age(birth_date):
          if pd.isnull(birth_date):
              return None
          today = date.today()
          age = today.year - birth_date.year - ((today.month, today.day) < (birth_date.month, birth_date.day))
          return int(age)

      df['age'] = df['birth_date'].apply(calculate_age)

      print(df[['birth_date', 'age']])
```

```
     birth_date   age
0    2000-01-25  24.0
1    1996-07-25  28.0
2    1994-09-15  30.0
3    1992-07-07  32.0
4    1998-11-14  25.0
...         ...   ...
2266 1993-09-04  31.0
2267 2003-04-05  21.0
```

Jupyter Untitled6 Last Checkpoint: 1 hour ago

Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

2271 rows × 14 columns

```
[4]: df['age'] = df['age'].fillna(0)
     df['age'] = df['age'].astype(int)
     print(df[['birth_date', 'age']])
```

```
     birth_date  age
0    2000-01-25   24
1    1996-07-25   28
2    1994-09-15   30
3    1992-07-07   32
4    1998-11-14   25
...         ...  ...
2266 1993-09-04   31
2267 2003-04-05   21
2268 2006-01-19   18
2269 1995-05-21   29
2270 1995-04-11   29

[2271 rows x 2 columns]
```

```
[6]: df
```

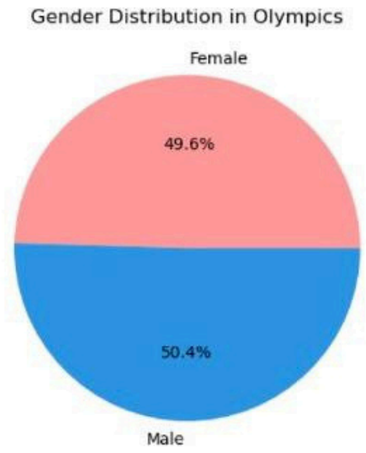
[6]:

	medal_date	medal_type	medal_code	name	gender	country	country_code	nationality	team	discipline	event	birth_date	code	age
0	2024-07-27	Gold Medal	1	EVENEPOEL Remco	Male	Belgium	BEL	Belgium	NaN	Cycling Road	Men's Individual Time Trial	2000-01-25	1903136	24
1	2024-07-27	Silver Medal	2	GANNA Filippo	Male	Italy	ITA	Italy	NaN	Cycling Road	Men's Individual Time Trial	1996-07-25	1923520	28
2	2024-07-27	Bronze Medal	3	van AERT Wout	Male	Belgium	BEL	Belgium	NaN	Cycling Road	Men's Individual Time Trial	1994-09-15	1903147	30

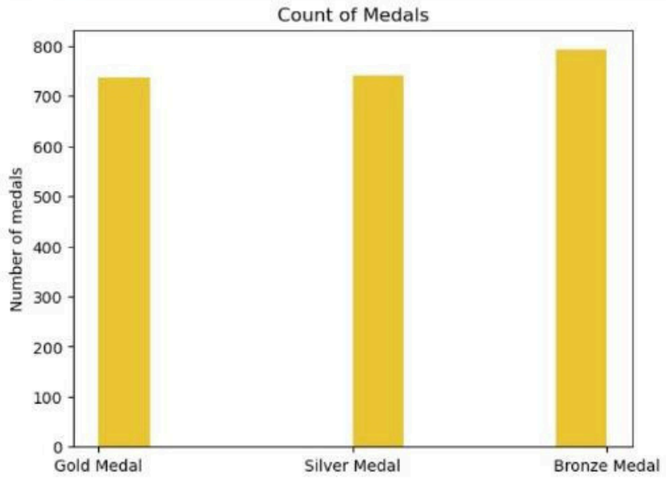
Transforming

transform data into visual with matplotlib

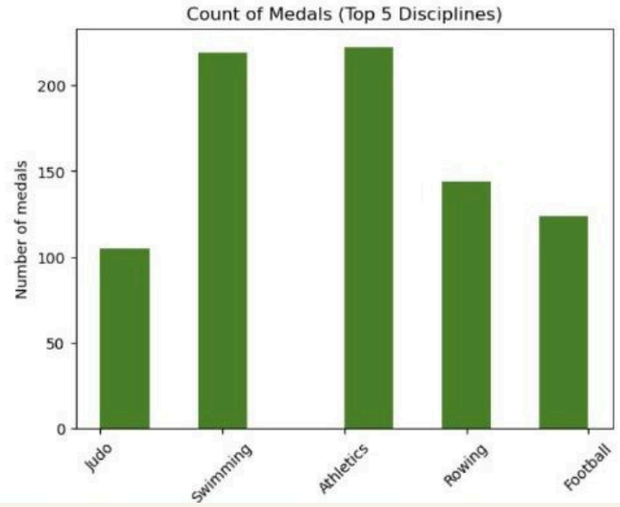

```
[150]: female_count = df.loc[df['gender'] == 'Female'].shape[0]
male_count = df.loc[df['gender'] == 'Male'].shape[0]
sizes = [female_count, male_count]
labels = ['Female', 'Male']
colors = ['#ff9999', '#2995e3']
plt.pie(sizes, labels=labels, colors=colors, autopct='%1.1f%%')
plt.title('Gender Distribution in Olympics')
plt.show()
```




```
[94]: plt.hist(df.medal_type, color='#ebc634')
plt.ylabel('Number of medals')
plt.title('Count of Medals')
plt.show()
```



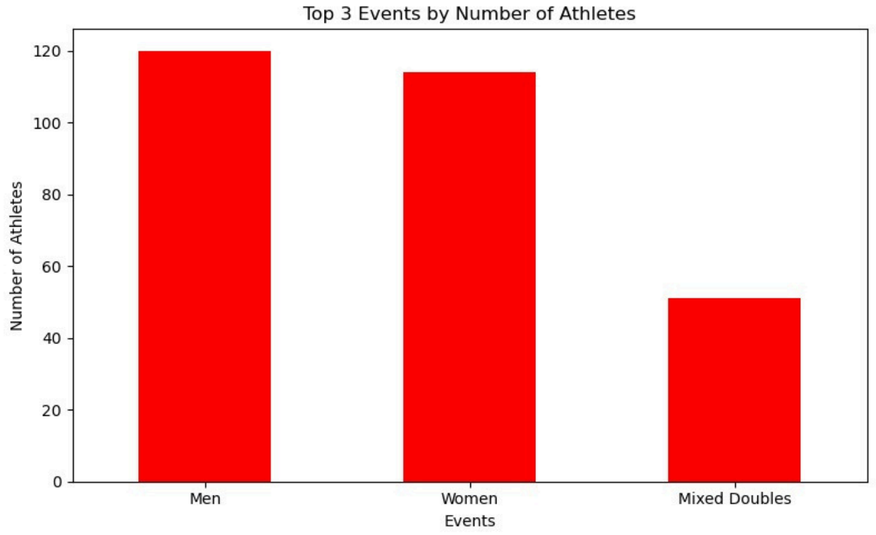

```
[102]: top_5_disciplines = df['discipline'].value_counts().nlargest(5).index
df_top_5 = df[df['discipline'].isin(top_5_disciplines)]
plt.hist(df_top_5['discipline'], color='#498028')
plt.ylabel('Number of medals')
plt.title('Count of Medals (Top 5 Disciplines)')
plt.xticks(rotation=45)  # Optional: rotate labels for better readability
plt.show()
```



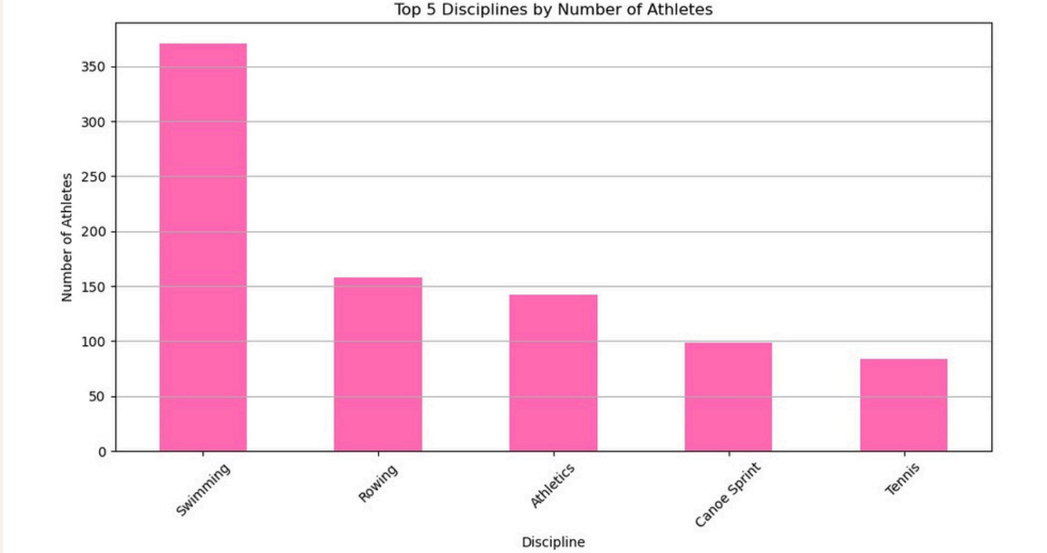
Transforming

transform data into visual with matplotlib

```
[36]: top_3_events = df['events'].value_counts().head(3)
plt.figure(figsize=(8, 5))
top_3_events.plot(kind='bar', color='red')
plt.title('Top 3 Events by Number of Athletes')
plt.xlabel('Events')
plt.ylabel('Number of Athletes')
plt.xticks(rotation=0)
plt.tight_layout()
plt.show()
```




```
[46]: top_5_disciplines = df['discipline'].value_counts().head(5)
plt.figure(figsize=(10, 6))
top_5_disciplines.plot(kind='bar', color='hotpink')
plt.title('Top 5 Disciplines by Number of Athletes')
plt.xlabel('Discipline')
plt.ylabel('Number of Athletes')
plt.xticks(rotation=45)
plt.grid(axis='y')
plt.tight_layout()
plt.show()
```



QUSETIONS

In this analysis, we focused on several key questions aimed at uncovering trends and insights. These questions helped guide the direction of the investigation, ensuring a deeper understanding of the data. Let's take a look at how these shaped the analysis."

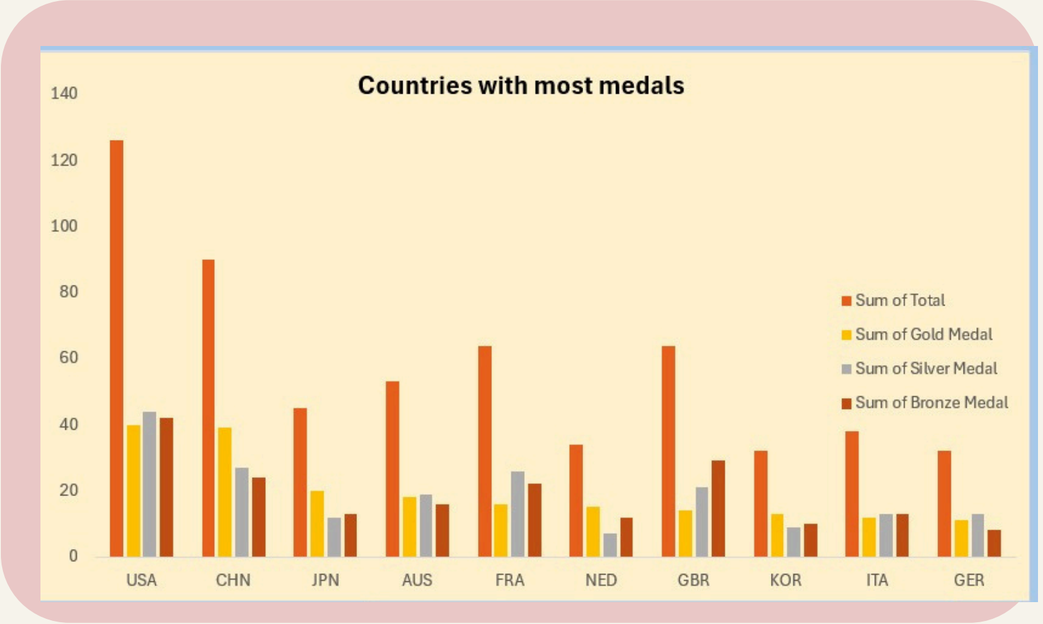
QUSETIONS

1. How many medals did each country and discpline win ?
2. How many medals did each country win in terms of medal type ?
3. what is the most game played by gender and medals type ?

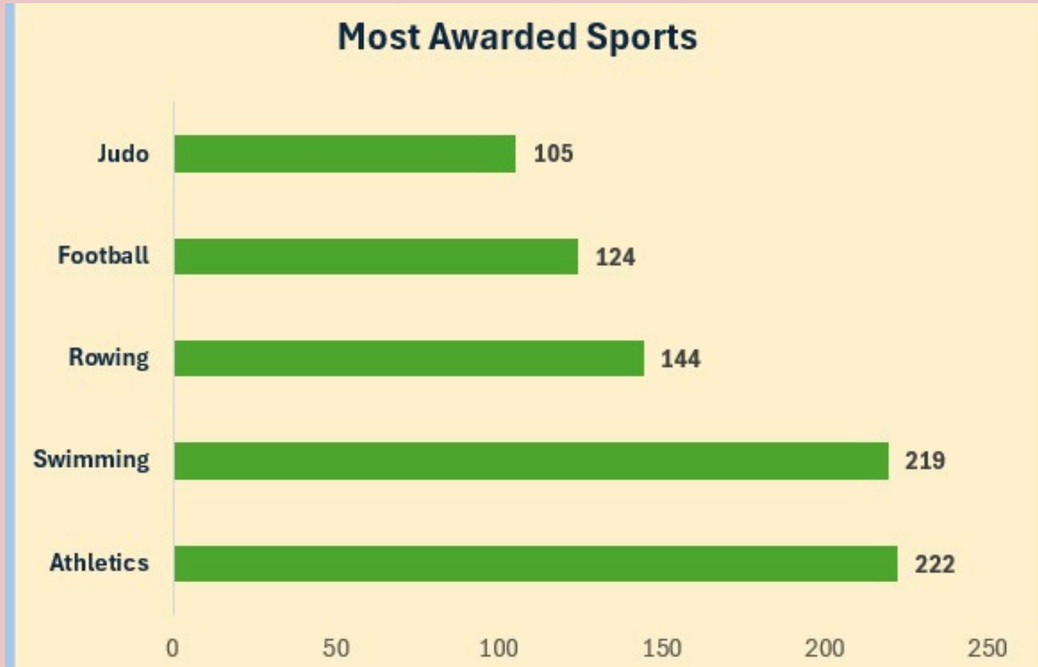
visulalization

- excel
- power bi

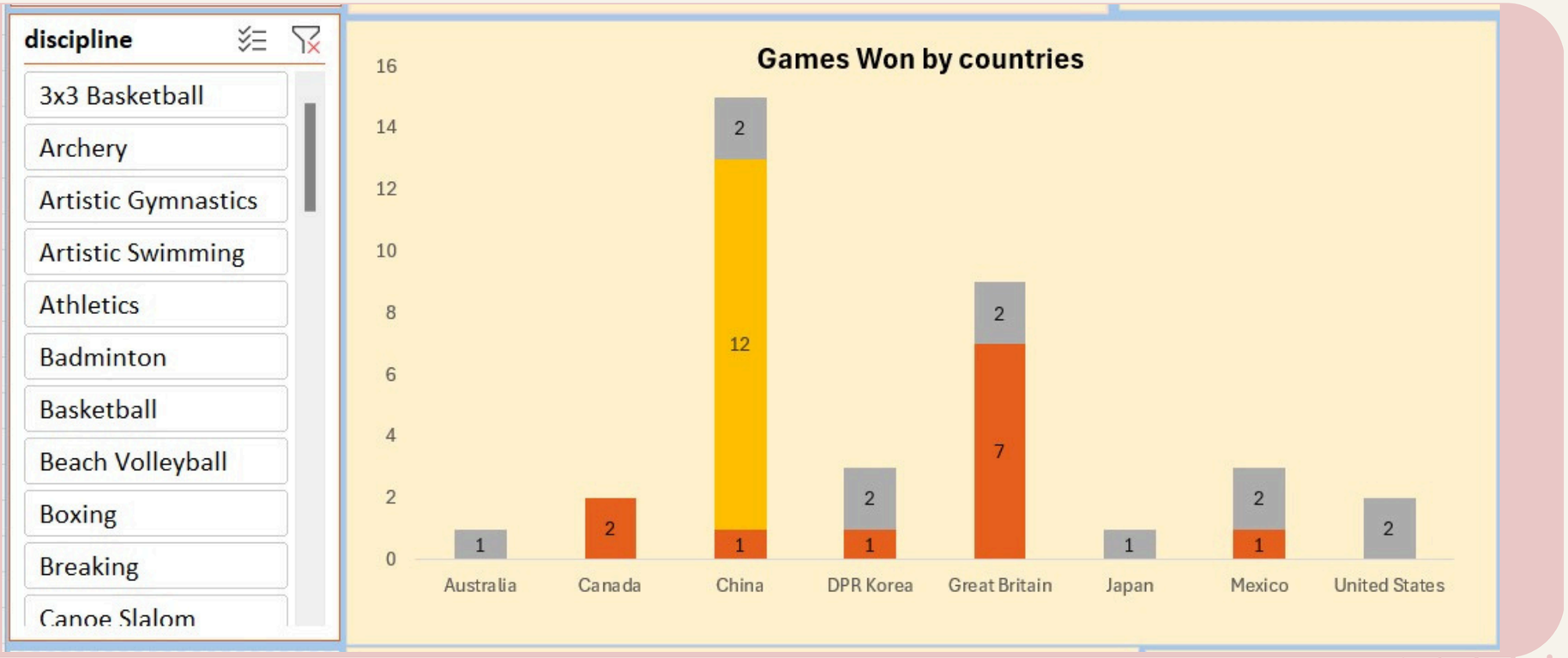
VISULAIZATION USEING EXCEL



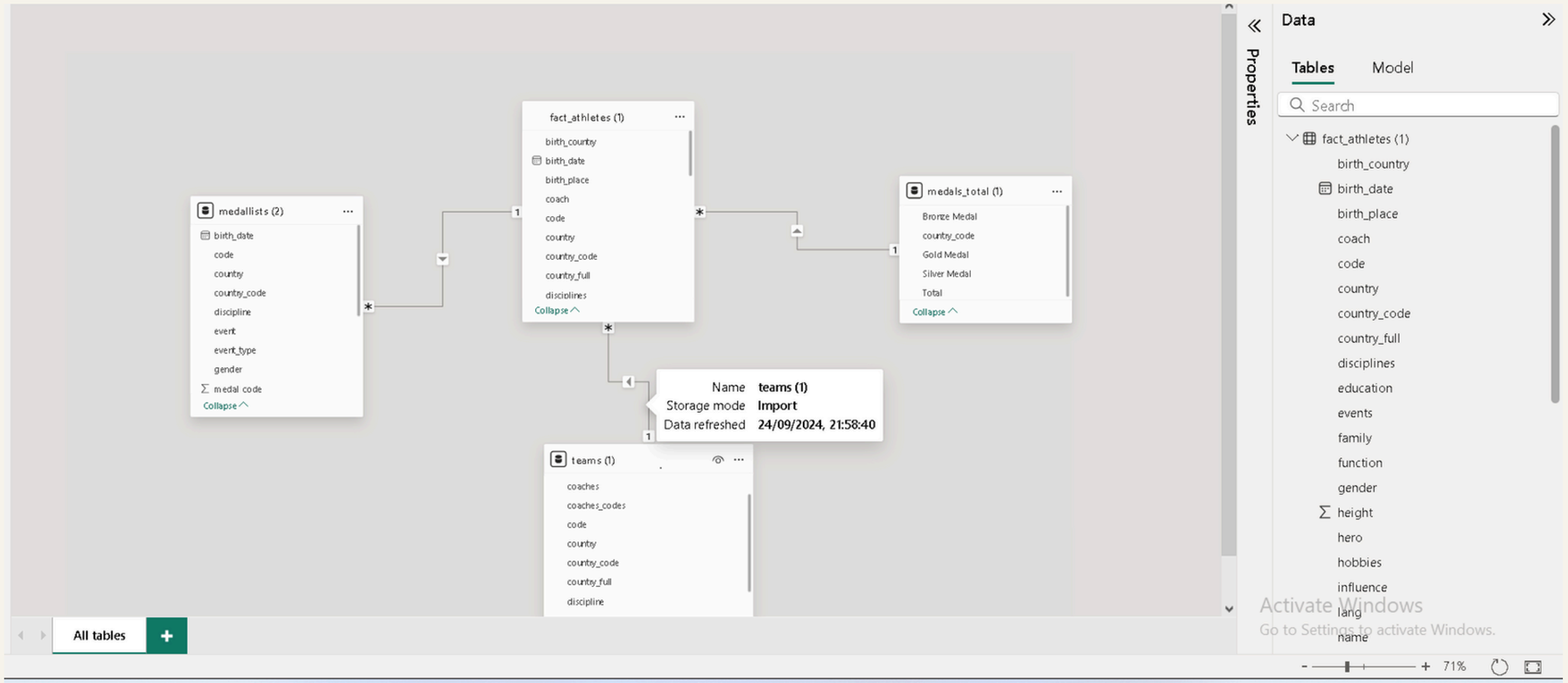
VISULAIZATION USEING EXCEL



VISULAIZATION USEING EXCEL



RELATIONSHIP



DAX &POWER QUERY

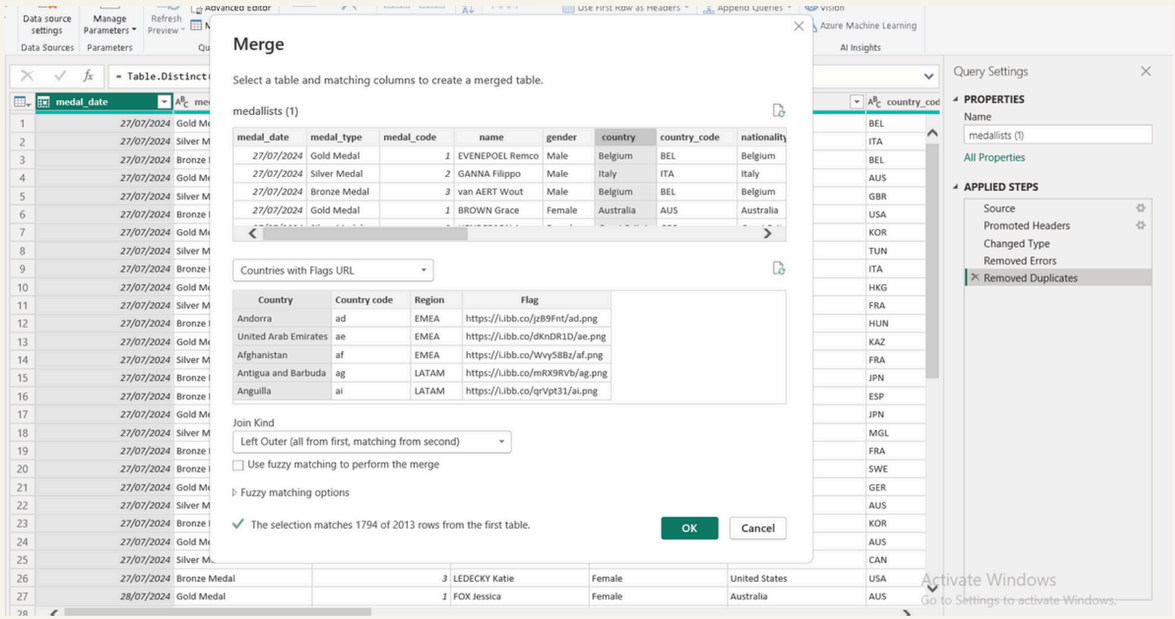

```
Age catogray =
SWITCH(
    TRUE(),
    'fact_athletes (1)'[Age]>=15 &&'fact_athletes (1)'[Age]<=20, "15-20",
    'fact_athletes (1)'[Age]>20 &&'fact_athletes (1)'[Age]<=25, "21-25",
    'fact_athletes (1)'[Age]>25 &&'fact_athletes (1)'[Age]<=30, "26-30",
    'fact_athletes (1)'[Age]>30 &&'fact_athletes (1)'[Age]<35, "31-35",
    'fact_athletes (1)'[Age]>35 &&'fact_athletes (1)'[Age]<=40, "36-40",
    'fact_athletes (1)'[Age]>40 &&'fact_athletes (1)'[Age]<=45, "41-45",
    'fact_athletes (1)'[Age]>45 &&'fact_athletes (1)'[Age]<=50, "46-50",
     'fact_athletes (1)'[Age]>50,"51+",
     "under 15"


)
```

Data

Search

- Countries with Flags URL
- fact_athletes (1)
  - Age
  - Age catogray
  - birth_country
  - birth_date
  - birth_place
  - coach
  - code
  - country

...ies	occupation	education	family	lang	coach	reason	hero	influence	philosophy	sporting_relatives	ritual	other_sports	Age	Age catogray
													24	21-25



INSIGHTS

Key Highlights

Total Athletes:11118, with 5659 male and 5459 female participants ,The male won 226 gold medals , and the female won 226 gold medals in the tournament , country USA is the highest-performing country in the tournament.

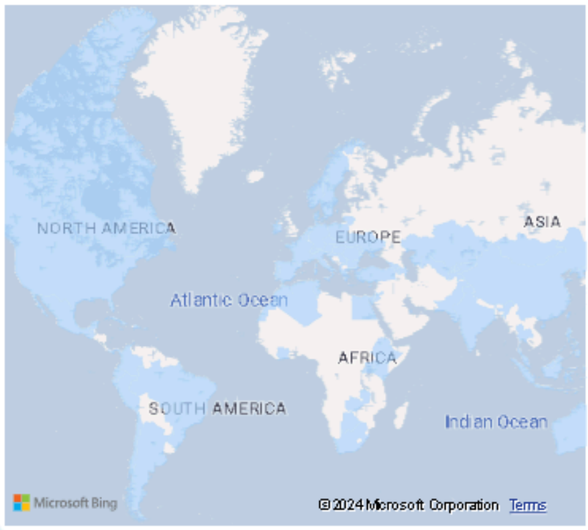
Gold 624

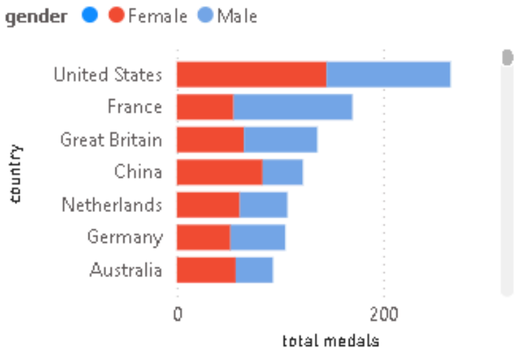
silver 667

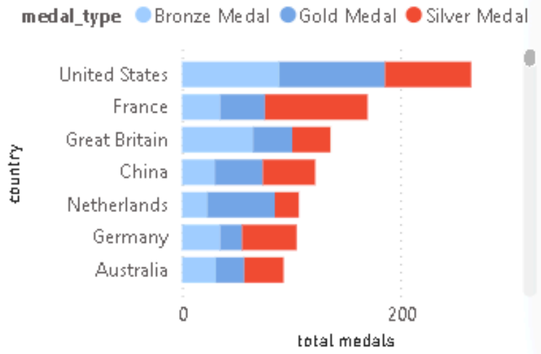
Bronze 722

AIN | Alban... | Algeria

11K	1698	206
Athletes	Teams	Country



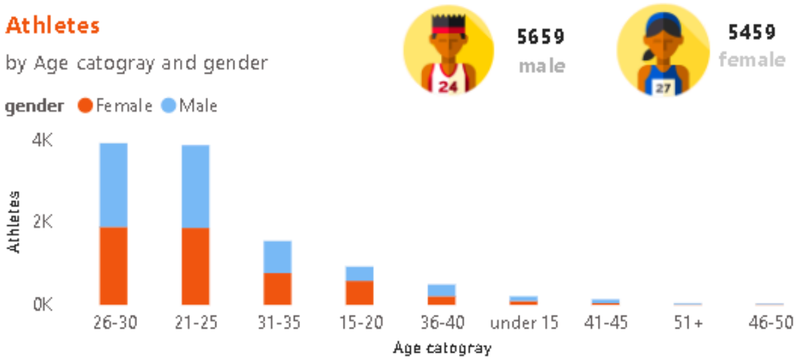




INSIGHTS

Athletes

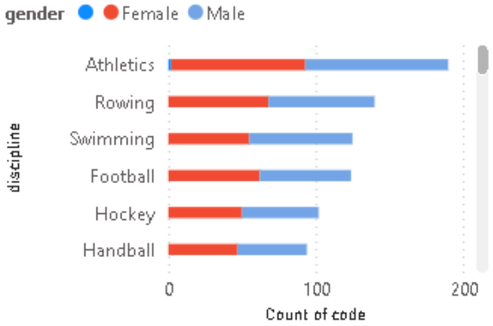
by Age catogray and gender



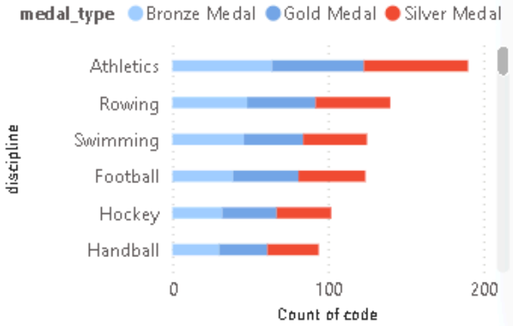
5659 male

5459 female

games by Gender



games by medals



Gold

Total 624 medals where male got 330 gold medals and female 293 medals

silver

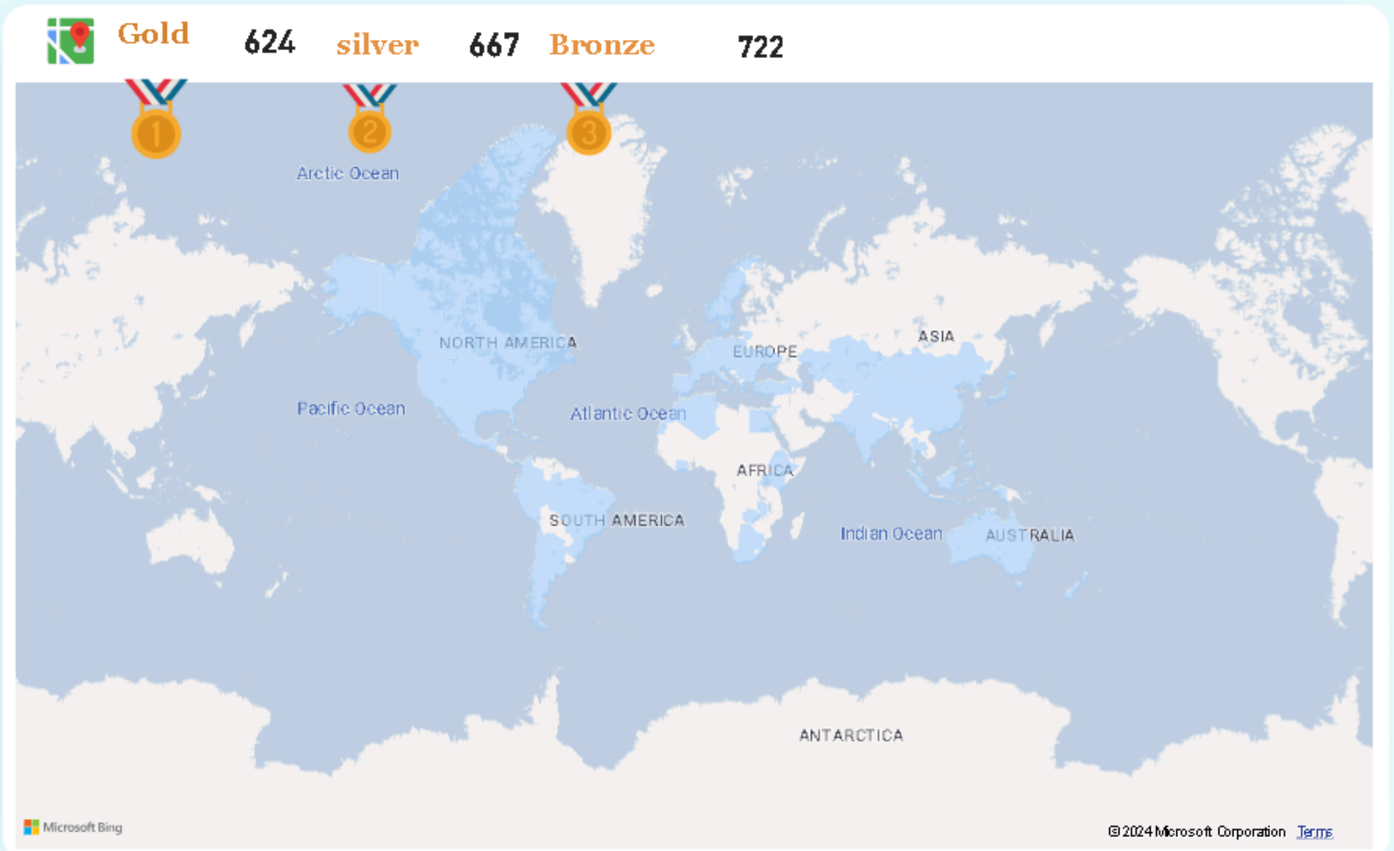
Total 667 medals where male got 337 gold medals and female 330 medals

Bronze

Total 772 medals where male got 370 gold medals and female 350 medals

INSIGHTS


AIN
Albania
Algeria
Argentina
Armenia
Australia
Austria
Azerbaijan


Gold 624
silver 667
Bronze 722
Arctic Ocean
NORTH AMERICA
EUROPE
ASIA
Pacific Ocean
Atlantic Ocean
AFRICA
SOUTH AMERICA
Indian Ocean
AUSTRALIA
ANTARCTICA
Microsoft Bing
© 2024 Microsoft Corporation Terms

TOOLS

- EXCEL
- jupyter notebook(matplotlib-pandas)
- power bi (power query)
- freepik
- YOUTUBE(DATASET)
- AI tools

DEPI | 2024

THANK YOU

Presented By :

Bassmla obaied

Mariam Mohamed Abd Elghany

Reem osama gamal

Salma Alaa Eldin Mohammed

Nataly Mohamed Khodairy