Technical report  
Data Analysis on Olympics 2024

-Olympics data analysis Using excel, python, power bi

- project by: Basmalla Mahmoud – Reem Osama – Salma Alaa – Mariem Mohamed – Natalie Mohamed

- Date: 9/29/2024

2. Table of Contents

Example structure:

1. Executive Summary

2. Introduction

3. Data Description

4. Methodology

5. Analysis and insights

6. Conclusion

---

**1. Executive Summary (Abstract)**

This project involved the analysis of data related to the 2024 Olympics, obtained from Google, with a focus on cleaning, transforming, and visualizing the data to uncover valuable insights. Python was used extensively for data cleaning, which included removing unnecessary columns, eliminating duplicates, and calculating athletes' ages from their birthdates. The cleaned data was imported into Power BI and Excel for comprehensive visualization, utilizing advanced features like slicers and interactive charts. Key insights were derived from both Python and visualization tools, providing a deeper understanding of athlete demographics and event participation

**2. Introduction**

- Background: With the upcoming 2024 Olympics, analyzing athlete and event data offers an opportunity to understand patterns in participation, performance, and demographics. Such insights can provide stakeholders with valuable information about trends in international sports.

- Objective: The goal of this project was to process and analyze data related to the 2024 Olympics, with a focus on cleaning the dataset, preparing it for analysis, and generating insights. The data cleaning process in Python was crucial to ensure the accuracy of further analysis. Once the data was prepared, visualizations in Power BI and Excel were created to uncover trends in athlete ages, event participation, and country-wise representation  
  
  
**3. Data Description**

- Data Source: The dataset used is about the 2024 Olympics and was obtained from Google we used four datasets ( athletes – medallists – total\_medals – teams).

- Data cleaning: we removed duplicates from data and dropped unnecessary columns such as ( team\_gender, event\_type, url\_event, name\_tv , short\_name, height, weight, residence place,) to focus on important data to answer our questions to get best insights

- Data Cleaning: we also calculated ages of athletes by using birth\_date column

**4. Methodology**

**- Data Preprocessing in Python**:

- **Unneeded Columns**: we used library pandas by using code: import pandas as pd

- **Duplicates**: we removed duplicates ensure data quality and avoid skewing the analysis,

by using codes such as: ( df\_cleaned = df.drop\_duplicates(subset=['Country Code']) )

- **Age Calculation**: we calculated age from birth\_date column to analyze age distribution and identify the oldest / youngest participants.  
 by using this code:

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']])

We also used this code to remove decimals from age

df['age'] = df['age'].fillna(0)

df['age'] = df['age'].astype(int)

print(df[['birth\_date', 'age']])

-**Tools Used for Visualization:**

- **Python (Matplotlib, Pandas, NumPy):   
 -**Bar Chart (Top 5 Disciplines by Medal Count): Displayed the count of medals won in the top 5 sports disciplines, providing insight into which sports had the highest performance.

- Bar Chart (Oldest Participant): Visualized the age of the oldest participant, helping to highlight age diversity in the event.

- Pie Chart (Gender Distribution): Showed the distribution of male and female participants, giving a clear view of gender representation in the Olympics.

-Bar chart (Top three events by number of athletes): illustrates the **top 3 events** in the dataset based on the **number of athletes** participating in them. Specifically:

1. **Men's Events**: These have the largest number of athletes, with a count exceeding 120.
2. **Women's Events**: Close behind men's events, women's events also have a high number of athletes, just under 120.
3. **Mixed Doubles Events**: These have significantly fewer athletes, with around 60 participants.

-Bar chart(Top 5 disciplines by number of athletes): displays the "Top 5 Disciplines by Number of Athletes." Here's what it shows:

1. **Swimming** has the highest number of athletes, significantly more than the other disciplines, with over 350 participants.
2. **Rowing** comes second, with around 150 athletes.
3. **Athletics** closely follows rowing, also with around 150 athletes.
4. **Canoe Sprint** has fewer participants, with roughly 100 athletes.
5. **Tennis** is the fifth most popular discipline in terms of athlete participation, with around 100 athletes.

- **Power BI:** we used power query in power bi ,Dax measures for multiple calculations in our dashboard to show cards , also on power query we merged between two tables with primary key (country flags with medalist (left join)) to show the flags on slicer , our model is star model with one fact table and others is dimension .

- **Excel**: we used different slicers to show age of athletes who won medals and what type of medals did they win and second slicer to show games won by countries and kind of discipline

We also did multiple chart to analyze different insights as most/ least awarded sports and countries of most medals   
  
  
**5. Analysis and Insights**

Key Findings: our analysis shows most played/ awardes sports are athletics and swimming with 222 medal for athletics and 219 medal for swimming leading the Olympics.

-We have ages for oldest winning Olympics at age of 58 with silver medal also youngest Olympic at age of 14 with gold medal

-We also have countries with most participants with **USA** at lead with 619 olympic

- countries leading count medals were USA at lead with126 medal and following China with 90 and Japan with 45 medals

-

**6. Conclusion**

In conclusion, this project successfully leveraged data analysis techniques to provide meaningful insights into the 2024 Olympics. Through data cleaning in Python and the use of Power BI and Excel for advanced visualization, the project highlighted key patterns in athlete demographics and event participation. The combination of these tools enabled a deeper, more interactive understanding of the data, offering valuable perspectives for future analyses and decision-making processes related to sporting events. This comprehensive approach demonstrated the power of data-driven insights in enhancing our understanding of complex datasets