EXPERIMENT – 8

PART - A

AIM :

Create Dashboard using Power BI for the Spots Data Analysis with
the following information:

- Which club(s) have the maximum share of players from England?
- Which club(s) have the maximum share of players from Spain?
- Which club(s) have the maximum share of players from Germany?
- Check it out for players with age Between 16 to 28?
- Create visuals and draw meaning full insights from the analyzed data

Python Scripts for Data Preprocessing

1. **Loading the Dataset**:

```
import pandas as pd

# Load the dataset from the specified path
df = pd.read_csv(r"D:/Power BI /Week 7/EXP-7 FILES/Olympics Dataset.csv")
```

2. **Handling Missing Values in the 'Age' Column**:

```
import pandas as pd
import numpy as np

# Assuming 'dataset' is the table provided by Power BI or loaded in Python
# Replace error values in the 'Age' column (e.g., NaN, -1) with the median
median_age = dataset[dataset['Age'] > 0]['Age'].median()

# Replace NaN and other error values in the 'Age' column with the median
dataset['Age'].replace([np.nan, -1], median_age, inplace=True)

# Return the cleaned dataset
dataset
```

3. **Handling Missing Values in 'Height' and 'Weight' Columns**:

```
import pandas as pd
import numpy as np

# Convert 'Height' and 'Weight' columns to numeric, forcing errors to NaN if any
```

```
dataset['Height'] = pd.to_numeric(dataset['Height'], errors='coerce')
dataset['Weight'] = pd.to_numeric(dataset['Weight'], errors='coerce')

# Calculate the mean of the 'Height' and 'Weight' columns, ignoring NaN values
mean_height = dataset['Height'].mean()
mean_weight = dataset['Weight'].mean()

# Replace NaN values in 'Height' and 'Weight' columns with their respective means
dataset['Height'].fillna(mean_height, inplace=True)
dataset['Weight'].fillna(mean_weight, inplace=True)

# Return the cleaned dataset
Dataset
```

DAX Code for Power BI Analysis

1. **Calculating the Share of English Athletes**:

```
MaxEnglandAthletes =
VAR TotalEnglandAthletes =
   CALCULATE(
      COUNTROWS(df),
      FILTER(df, df[Team] = "England")
   )
VAR TotalAthletes =
   CALCULATE(
      COUNTROWS(df)
   )
RETURN
   DIVIDE(
      TotalEnglandAthletes,
      TotalAthletes
   )
```

2. **Creating a Sports Table with Hardcoded English Athletes**:

```
SportsEnglandTable =
VAR SportsList =
   VALUES(df[Sport])
RETURN
   ADDCOLUMNS(
```

```
    SportsList,
    "MaxEnglandAthletes", 12
  )
```

3. **Calculating the Share of Spanish Athletes**:

```
MaxSpainAthletes =
VAR TotalSpainAthletes =
  CALCULATE(
    COUNTROWS(df),
    FILTER(df, df[Team] = "Spain")
  )
VAR TotalAthletes =
  CALCULATE(
    COUNTROWS(df)
  )
RETURN
  DIVIDE(
    TotalSpainAthletes,
    TotalAthletes
  )
```

4. **Calculating the Share of German Athletes**:

```
MaxGermanyAthletes =
VAR TotalGermanyAthletes =
  CALCULATE(
    COUNTROWS(df),
    df[Team] = "Germany"
  )
VAR TotalAthletes =
  COUNTROWS(df)
RETURN
  DIVIDE(
    TotalGermanyAthletes,
    TotalAthletes,
    0    // Handle division by zero
  )
```

1. Which club(s) have the maximum share of players from England?

1. In Power BI, import the merged data from the full outer join.
2. Create a measure to calculate the **share of players from England** for each club. Use the MaxEnglandAthletes column to identify the English athletes.
3. Create a **bar chart** or **pie chart** to visualize the clubs and their respective shares of English players.

- **Visual**: A **bar chart** showing clubs on the x-axis and the share of English players on the y-axis, sorted in descending order.

2. Which club(s) have the maximum share of players from Spain?

1. Create a measure for the **share of players from Spain** using the MaxSpainAthletes column.
2. Create a similar **bar chart** or **pie chart** for Spanish players' share by club.

- **Visual**: A **pie chart** to show the proportion of Spanish athletes across different clubs.

3. Which club(s) have the maximum share of players from Germany?

1. Create a measure for the **share of players from Germany** using the MaxGermanyAthletes column.
2. Use a **bar chart** or **donut chart** for better comparison.

- **Visual**: A **donut chart** showing the clubs and their share of German players.

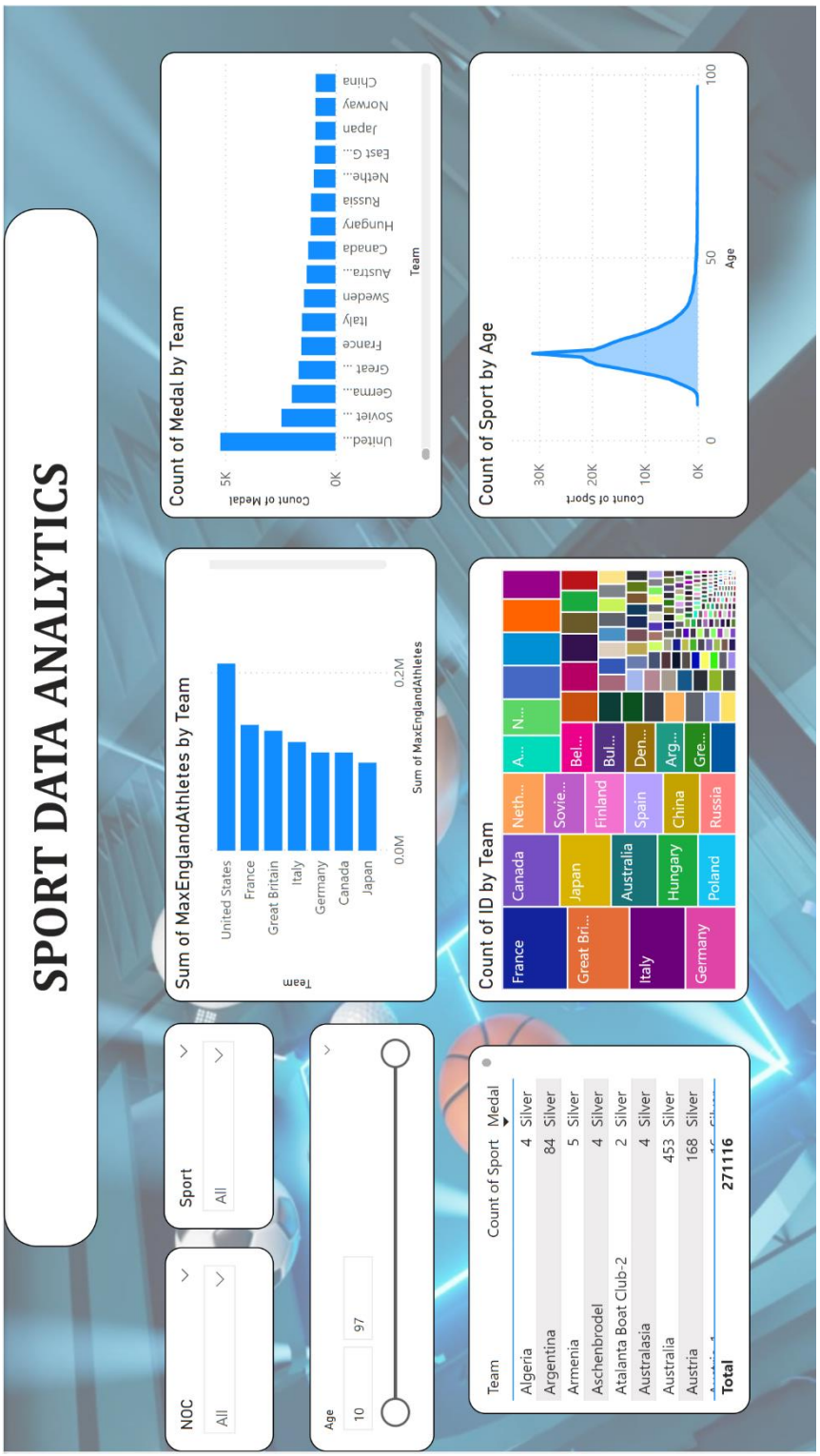
4. Check it out for players with age Between 16 to 28?

1. Add a **filter** in Power BI to limit the data to athletes aged between 16 and 28 using the age column.
2. Apply this filter to the previous visuals to see how the shares change when only considering this age group.

- **Visuals**: Apply the **slicer/filter** to all charts (for England, Spain, Germany) to focus only on players aged 16 to 28.

5. Create visuals and draw meaningful insights:

1. Add a **scatter plot** to analyze the relationship between age and sum of weight or height.
2. Create a **stacked column chart** for medal distribution by country (use the Medal and NOC columns).
3. Use a **map visual** to plot the locations of events based on the city and team columns to explore geographical patterns.
4. Create a **line chart** for year-wise analysis of the number of athletes using the sum of year column.
5. Analyze **performance by sport** using the Sport column to visualize the athlete count per sport.



Result :

Sports Data Analytics Dashboard is constructed based on the given requirements and additional insights are generated.