Understanding the Olympics Through Data Analysis

VHARIHARA SESHU AND NS PRANAV

Introduction to the Olympics

01 What are the Olympics?

02 Brief history and significance

03 Types: Summer & Winter Olympics

The Olympics: A Global Celebration of Sport and Unit

Exploring the Origins, Significance, and Evolution of the Summer and Winter Games

Data Collection

- Sources of Olympic data (e.g., IOC databases, Kaggle, sports APIs)
- Types of data collected (medals, countries, athletes, sports, years, etc.)
- Challenges in collecting sports data

DATA COLLECTION

Click here to edit subtitle

```
# ----------------------
# Import libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from google.colab import files

# Upload CSV file
print("Please upload the 'summer.csv' file...")
uploaded = files.upload()

# Read the CSV file into a DataFrame
df = pd.read_csv("summer.csv")
print("Data Loaded Successfully!")

# Display basic info
print("\nDataFrame Shape:", df.shape)
print("\nFirst 5 Rows:")
print(df.head())
```

```
Please upload the 'summer.csv' file...
Choose Files summer.csv
• summer.csv(text/csv) - 2573921 bytes, last modified: 5/14/2025 - 100% done
Saving summer.csv to summer.csv
Data Loaded Successfully!

DataFrame Shape: (31165, 9)

First 5 Rows:
   Year    City     Sport Discipline             Athlete Country Gender  \
0  1896  Athens  Aquatics   Swimming       HAJOS, Alfred     HUN    Men
1  1896  Athens  Aquatics   Swimming    HERSCHMANN, Otto     AUT    Men
2  1896  Athens  Aquatics   Swimming   DRIVAS, Dimitrios     GRE    Men
3  1896  Athens  Aquatics   Swimming  MALOKINIS, Ioannis     GRE    Men
4  1896  Athens  Aquatics   Swimming  CHASAPIS, Spiridon     GRE    Men
```

Data Transformation

- [ ] Cleaning and preprocessing raw data
- [ ] Handling missing values and inconsistencies
- [ ] Structuring data for analysis (e.g., CSV, DataFrames)

Data TRANSFORMATION

This image teases our innovative product, redefining industry standards.

```
# ---------------------------

# Check for missing values
print("\nMissing Values:")
print(df.isnull().sum())

# Drop rows with missing values (optional)
df.dropna(inplace=True)

# Normalize column names (optional step)
df.columns = df.columns.str.strip().str.lower()

# Convert year to int just in case
df['year'] = df['year'].astype(int)

# Check unique cities
unique_cities = df['city'].unique()
print(f"\nTotal unique cities that hosted Summer Olympics: {len(unique_cities)}")
print("\nList of cities:")
print(unique_cities)
```

```
Missing Values:
Year          0
City          0
Sport         0
Discipline    0
Athlete       0
Country       4
Gender        0
Event         0
Medal         0
dtype: int64

Total unique cities that hosted Summer Olympics: 22

List of cities:
```

Data Analysis Techniques

Techniques used: descriptive stats, visualization, etc.

Key questions analyzed: which country wins the most? Trends over time?

Tools used: Excel, Python, Power BI, etc.

Insights and Findings

Summary of insights derived from data analysis

Importance of data in understanding and predicting Olympic trends

Call to Action