DATA ANALYSIS AND VISUALIZATION ASSIGNMENT -160123737146

Olympic Medals Dataset - Data Analysis Report

Dataset Overview

The dataset contains information on Olympic medals awarded across different countries, disciplines, and dates. Key columns include:

- country: Country name
- medal_type: Type of medal (Gold, Silver, Bronze)
- medal_code: Encoded value for medal types
- medal_date: Date the medal was awarded
- gender: Gender of the participant
- discipline: Sport or event
- code: Unique athlete or country code

Data Cleaning Steps

- Handled missing values using forward fill and median replacement.
- Converted the medal_date column to datetime format.
- Removed irrelevant columns like gender and discipline after analysis.
- Encoded the gender column using 0 (Male) and 1 (Female) where necessary.
- Added extra columns to support grouped analysis.

Key Analysis & Insights

1. Top Countries by Medal Count

- The top 10 countries were identified using a bar plot.
- Countries like the **USA, Russia, and China** dominated the medal tally.

2. Medal Type Distribution

- Gold medals appeared to be slightly more frequent than Bronze or Silver.

- Countplot visualized the frequency of each medal type.

3. Medals by Country and Type

- A pivot table and heatmap showed which countries excelled in which medal category.
- For example, **Germany and Japan** had high gold medal counts in specific disciplines.

4. Statistical Summary

- Mean, Median, and Standard Deviation of medal codes were calculated.
- Descriptive statistics provided a general overview of the numerical data.

Visualization Highlights

- **Bar Plot**: Showed top 10 countries with highest medal counts.
- **Countplot**: Showed frequency distribution of medal types.
- **Heatmap** *(Optional addition)*: Helped visually compare countries and medal categories.

Final Output

- A cleaned dataset was exported as: cleaned_medal_data.csv
- This file is ready for further analysis or visualization.

✅ **Conclusion**

This project provided a comprehensive look into Olympic medal trends using Python and data visualization tools. With proper cleaning, encoding, and grouping, we extracted meaningful insights about top-performing countries and medal distribution.