

Draft Session (50m)

```
/kaggle/input/2021-olympics-in-tokyo/EntriesGender.xlsx
/kaggle/input/2021-olympics-in-tokyo/Teams.xlsx
/kaggle/input/2021-olympics-in-tokyo/Athletes.xlsx
/kaggle/input/2021-olympics-in-tokyo/Coaches.xlsx
/kaggle/input/2021-olympics-in-tokyo/Medals.xlsx
```


Input

[+ Add Input](#) [⬆ Upload](#)

- ▶ 2021-olympics-in-tokyo

Output (72KB / 19.5GB)

- ▶  /kaggle/working

Session options

Schedule a notebook to run

Code Help

 Find code help

Notebook

Input

 Upload

DATASETS

- 2021-olympics-in-tokyo

Output (72KB / 19.5GB)

- /kaggle/working

Session options

Schedule a notebook to run

Code Help

 Find code help

```
# Load the datasets
```

```
athletes = pd.read_excel('/kaggle/input/2021-olympics-in-tokyo/Athletes.xlsx')
coaches = pd.read_excel('/kaggle/input/2021-olympics-in-tokyo/Coaches.xlsx')
entries_gender = pd.read_excel('/kaggle/input/2021-olympics-in-tokyo/EntriesGender.xlsx')
medals = pd.read_excel('/kaggle/input/2021-olympics-in-tokyo/Medals.xlsx')
teams = pd.read_excel('/kaggle/input/2021-olympics-in-tokyo/Teams.xlsx')
```

```
datasets = {
  "Athletes": athletes.head(),
  "Coaches": coaches.head(),
  "EntriesGender": entries_gender.head(),
  "Medals": medals.head(),
  "Teams": teams.head()
}
```

datasets

```
/opt/conda/lib/python3.10/site-packages/openpyxl/styles/stylesheet.py:237: UserWarning: Workbook contains no default style, apply openpyxl's default
warn("Workbook contains no default style, apply openpyxl's default")
/opt/conda/lib/python3.10/site-packages/openpyxl/styles/stylesheet.py:237: UserWarning: Workbook contains no default style, apply openpyxl's default
warn("Workbook contains no default style, apply openpyxl's default")
```

	Name		NOC	Discipline		
0	AALERUD Katrine	Norway	Cycling Road			
1	ABAD Nestor	Spain	Artistic Gymnastics			
2	ABAGNALE Giovanni	Italy	Rowing			
3	ABALDE Alberto	Spain	Basketball			
4	ABALDE Tamara	Spain	Basketball,			
'Coaches':						
	Name		NOC	Discipline Event		
0	ABDELMAGID Wael	Egypt	Football	NaN		
1	ABE Junya	Japan	Volleyball	NaN		
2	ABE Katsuhiko	Japan	Basketball	NaN		
3	ADAMA Cherif	Côte d'Ivoire	Football	NaN		
4	AGEBA Yuya	Japan	Volleyball	NaN,		
'EntriesGender':						
	Discipline		Female	Male	Total	
0	3x3 Basketball	32	32	64		
1	Archery	64	64	128		

notebook6a6086f4dc Draft saved

File Edit View Run Add-ons Help

+ ✂ 📄 📄 ▶ ▶▶ Run All Code

Draft Session (1h:1m)

```
# Inspect data types and missing values for each dataset

# Athletes dataset
athletes_info = {
    "Data Types": athletes.dtypes,
    "Missing Values": athletes.isnull().sum()
}

# Coaches dataset
coaches_info = {
    "Data Types": coaches.dtypes,
    "Missing Values": coaches.isnull().sum()
}

# EntriesGender dataset
entries_gender_info = {
    "Data Types": entries_gender.dtypes,
    "Missing Values": entries_gender.isnull().sum()
}

# Medals dataset
medals_info = {
    "Data Types": medals.dtypes,
    "Missing Values": medals.isnull().sum()
}

# Teams dataset
teams_info = {
    "Data Types": teams.dtypes,
    "Missing Values": teams.isnull().sum()
}

# Combine all information into a single dictionary
data_info = {
    "Athletes": athletes_info,
    "Coaches": coaches_info,
    "EntriesGender": entries_gender_info,
    "Medals": medals_info,
    "Teams": teams_info
}
```

Share Save Version 1

Notebook

Input

+ Add Input Upload

DATASETS

2021-olympics-in-tokyo

Output (72KB / 19.5GB)

/kaggle/working

Session options

Schedule a notebook to run

Code Help

Find code help

8 updates available

kaggle

Create

Home

Competitions

Datasets

Models

Code

Discussions

Learn

More

Your Work

VIEWED

2021 Olympics in Tokyo

Olympic Games 2021 ...

Tokyo 2020 Olympics ...

Tokyo 2021 Olympics...

Paris 2024 Olympic Su...

notebook6a6086f4dc Draft saved

File Edit View Run Add-ons Help

+ Code + Markdown Run All

```
[6]: # Step 1: Handle missing values in the 'Event' column of the Coaches dataset
# For simplicity, we'll fill missing values with 'Unknown'
coaches['Event'].fillna('Unknown', inplace=True)

# Step 2: Check for duplicates and remove them
athletes_clean = athletes.drop_duplicates()
coaches_clean = coaches.drop_duplicates()
entries_gender_clean = entries_gender.drop_duplicates()
medals_clean = medals.drop_duplicates()
teams_clean = teams.drop_duplicates()

# Step 3: Generate summary statistics for numerical columns
athletes_stats = athletes_clean.describe(include='all')
coaches_stats = coaches_clean.describe(include='all')
entries_gender_stats = entries_gender_clean.describe()
medals_stats = medals_clean.describe()
teams_stats = teams_clean.describe(include='all')

# Compile statistics into a dictionary
statistics = {
    'Athletes': athletes_stats,
    'Coaches': coaches_stats,
    'EntriesGender': entries_gender_stats,
    'Medals': medals_stats,
    'Teams': teams_stats
}

statistics
```

```
[6]: {'Athletes':
count      11084      Name      11084      NOC Discipline
unique      11062      206      46
top      GANNA Filippo United States of America Athletics
freq          2          615      2068,
'Coaches':
count      393      Name      NOC Discipline      Event
unique      381      61      9      7
```

Share Save Version 1

Notebook

Input

+ Add Input Upload

DATASETS

2021-olympics-in-tokyo

Output (72KB / 19.5GB)

/kaggle/working

Session options

Schedule a notebook to run

Code Help

Find code help

8 updates available

notebook6a6086f4dc Draft saved

File Edit View Run Add-ons Help

+ ✂ 📄 📁 ▶ ▶▶ Run All Code

Draft Session (1h:1m)

min	1.000000	0.000000	0.000000	0.000000	1.000000	1.000000
25%	24.000000	0.000000	0.000000	1.000000	2.000000	23.000000
50%	46.000000	1.000000	1.000000	2.000000	4.000000	47.000000
75%	70.000000	3.000000	4.000000	5.000000	11.000000	66.000000
max	86.000000	39.000000	41.000000	33.000000	113.000000	77.000000
'Teams':						
	Name	Discipline	NOC Event			
count	743	743	743			
unique	146	20	84	36		
top	Japan	Swimming	Japan	Men		
freq	45	113	48	120		

```
[11]: # Load Medals dataset
medals_df = pd.read_excel('/kaggle/input/2021-olympics-in-tokyo/Medals.xlsx')

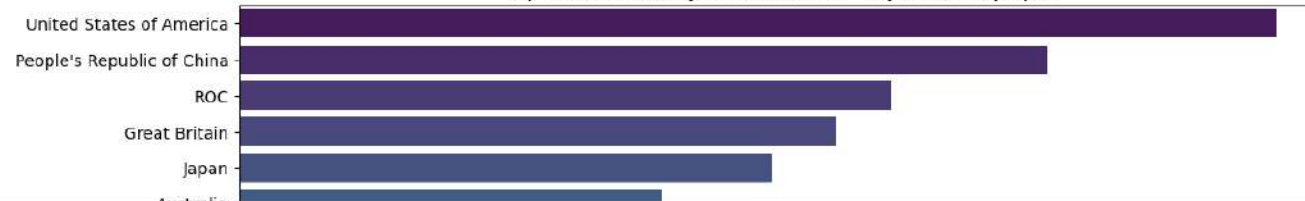
# Sort by total medals
medals_df = medals_df.sort_values(by='Total', ascending=False)

# Plotting
plt.figure(figsize=(12, 8))
sns.barplot(x='Total', y='Team/NOC', data=medals_df.head(20), palette='viridis')
plt.title('Top 20 Countries by Total Medals in Tokyo 2021 Olympics')
plt.xlabel('Total Medals')
plt.ylabel('Country')
plt.show()
```

/opt/conda/lib/python3.10/site-packages/openpyxl/styles/stylesheet.py:237: UserWarning: Workbook contains no default style, apply openpyxl's default

warn("Workbook contains no default style, apply openpyxl's default")

Top 20 Countries by Total Medals in Tokyo 2021 Olympics



Share Save Version 1

Notebook

Input

+ Add Input Upload

DATASETS

2021-olympics-in-tokyo

Output (72KB / 19.5GB)

/kaggle/working

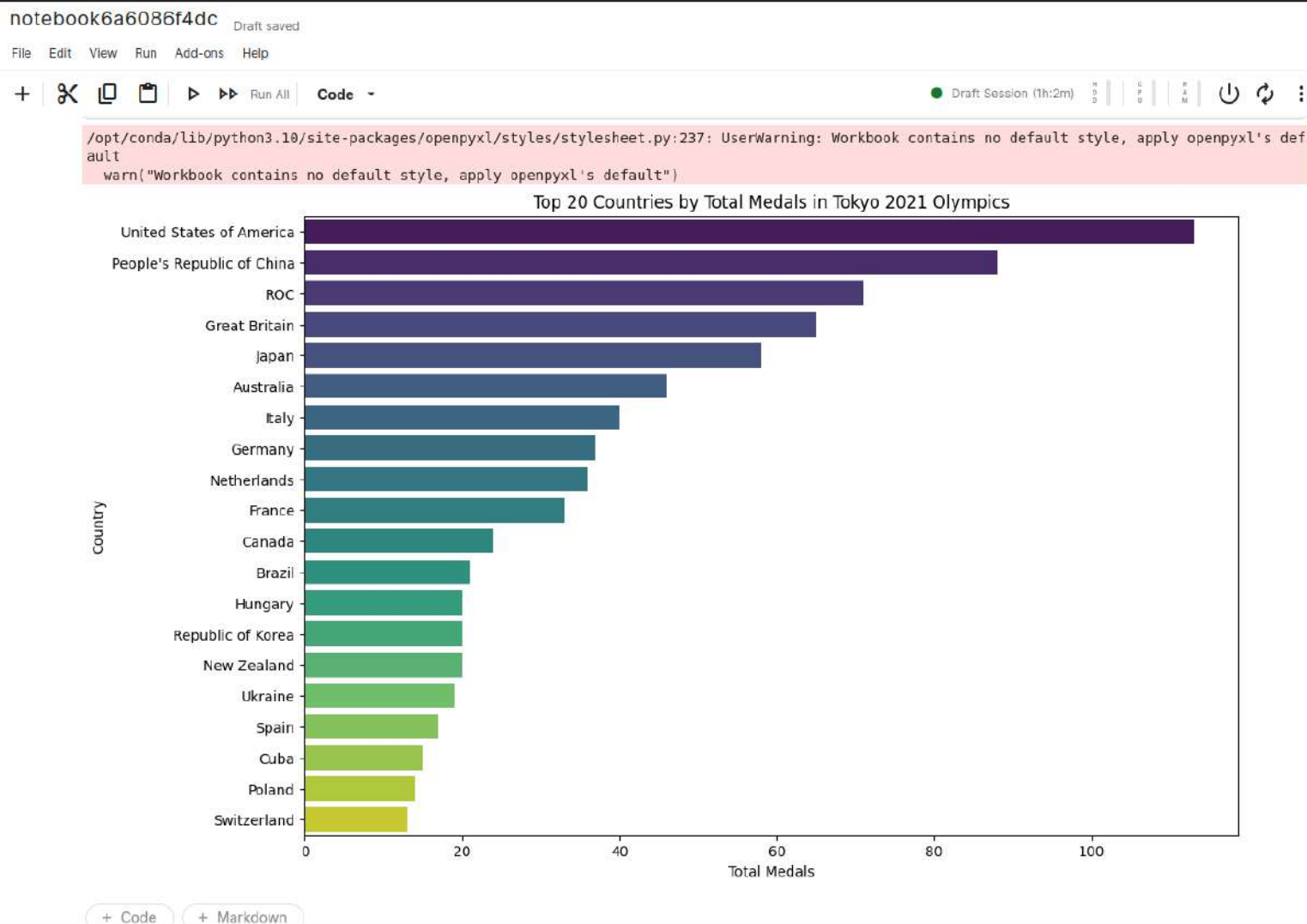
Session options

Schedule a notebook to run

Code Help

Find code help

8 updates available



```
[26]: # Load datasets
athletes_df = pd.read_excel('/kaggle/input/2021-olympics-in-tokyo/Athletes.xlsx')
entries_gender_df = pd.read_excel('/kaggle/input/2021-olympics-in-tokyo/EntriesGender.xlsx')

# Merge athletes and entries_gender dataframes on Discipline
merged_df = pd.merge(athletes_df, entries_gender_df, on='Discipline', how='left')

# Calculate percentage of male and female athletes per country
merged_df['Female Percentage'] = (merged_df['Female'] / merged_df['Total']) * 100
merged_df['Male Percentage'] = (merged_df['Male'] / merged_df['Total']) * 100

# Group by country (NOC) and calculate average percentages
country_gender_percentages = merged_df.groupby('NOC')[['Female Percentage', 'Male Percentage']].mean().reset_index()

# Plotting
plt.figure(figsize=(12, 8))
sns.scatterplot(x='Male Percentage', y='Female Percentage', data=country_gender_percentages, hue='NOC', palette='viridis', s=100)
plt.title('Gender Representation in Tokyo 2021 Olympics')
plt.xlabel('Male Percentage')
plt.ylabel('Female Percentage')
plt.show()
```



Notebook

Input

+ Add Input Upload

DATASETS

2021-olympics-in-tokyo

Output (72KB / 19.5GB)

/kaggle/working

Session options

Schedule a notebook to run

Code Help

Find code help

8 updates available

kaggle

Create

Home

Competitions

Datasets

Models

Code

Discussions

Learn

More

Your Work

VIEWED

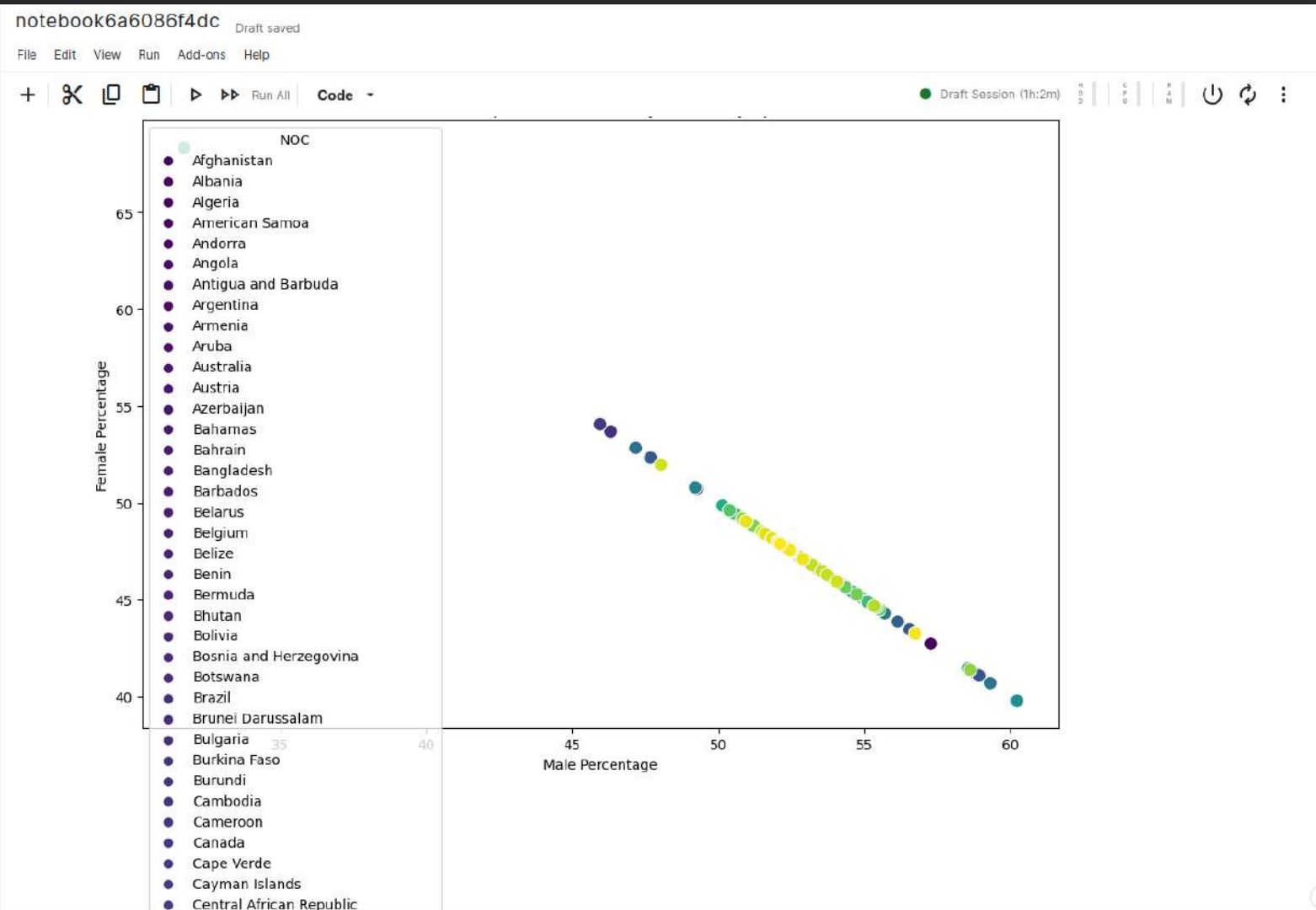
2021 Olympics in Tokyo

Olympic Games 2021 ...

Tokyo 2020 Olympics ...

Tokoyo 2021 Olympics...

Paris 2024 Olympic Su...



Share Save Version 1

Notebook

Input

+ Add Input Upload

DATASETS

2021-olympics-in-tokyo

Output (72KB / 19.5GB)

/kaggle/working

Session options

Schedule a notebook to run

Code Help

Find code help

8 updates available

Draft Session (1h:2m)

[14]:

```
/opt/conda/lib/python3.10/site-packages/openpyxl/styles/stylesheet.py:237: UserWarning: Workbook contains no default style, apply openpyxl's default
warn("Workbook contains no default style, apply openpyxl's default")
```

Country	Number of People (in millions)
Japan	127.1
United States of America	126.8
Spain	126.8
Australia	126.8
Canada	126.8
Italy	126.8
ROC	126.8
Egypt	126.8
South Africa	126.8
People's Republic of China	126.8

Input

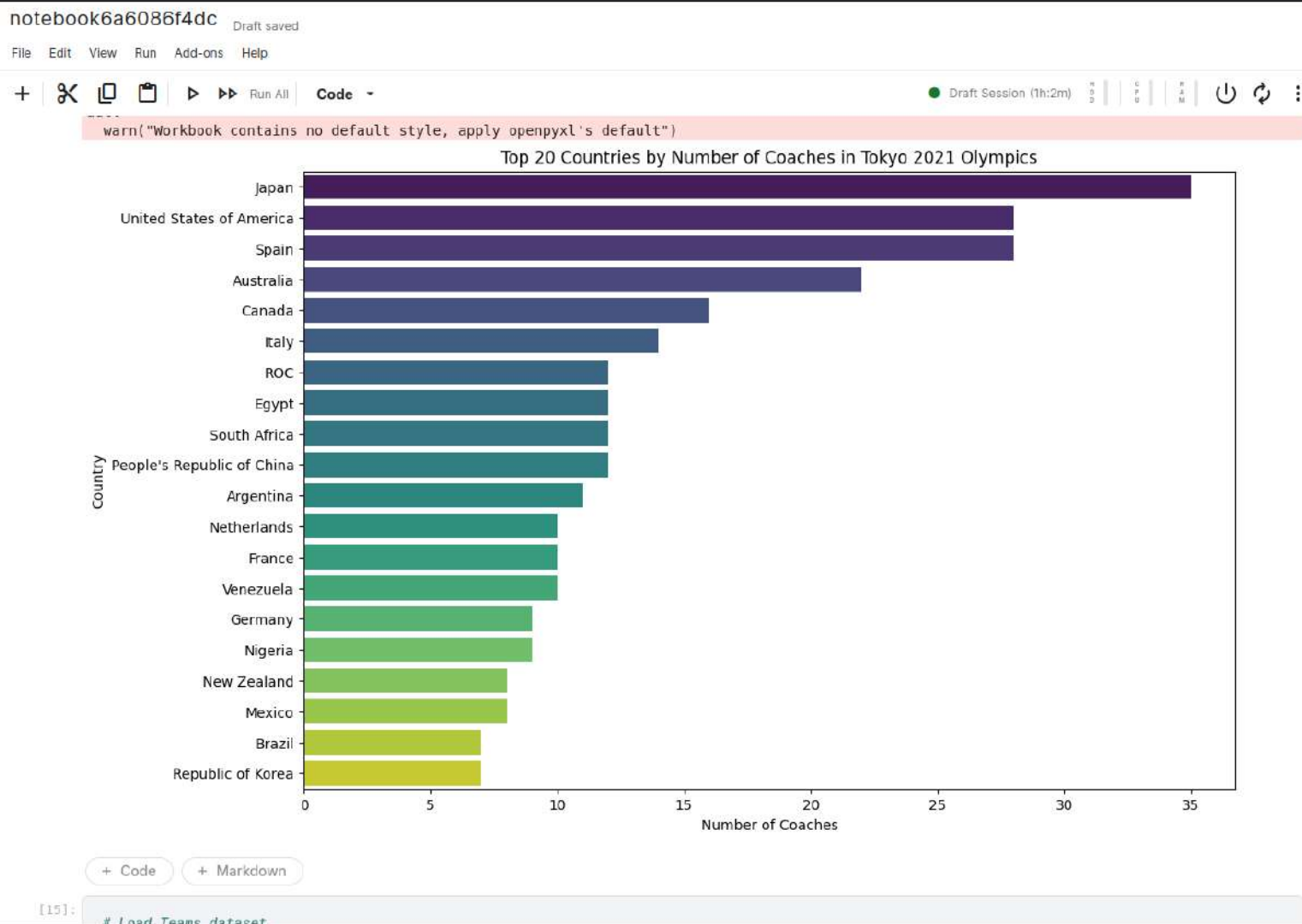
DATASETS

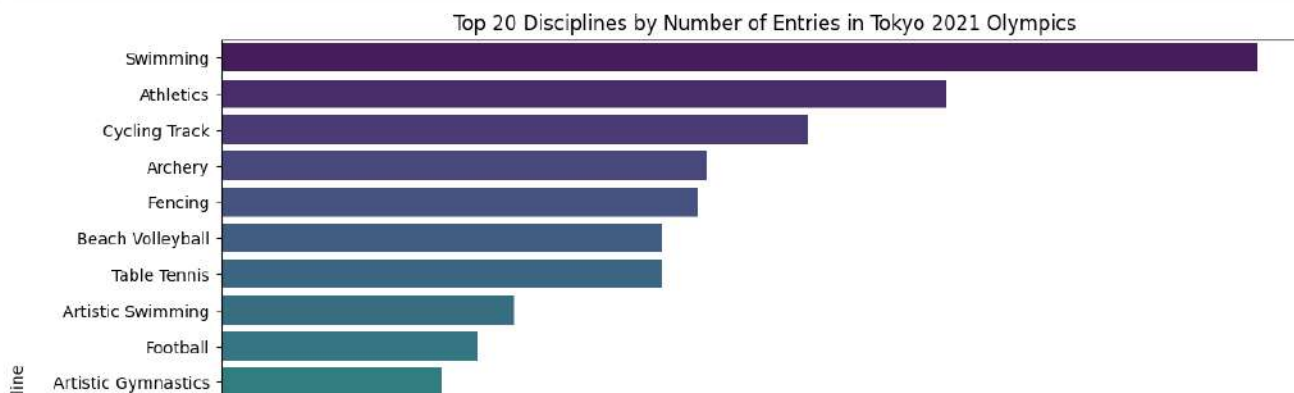
Output (72KB / 19.5GB)

Session options

Schedule a notebook to run

 Find code help





Input

+ Add Input ⬆️ Upload

▶ 2021-olympics-in-tokyo

Output (72KB / 19.5GB)

-  /kaggle/working

Session options

Schedule a notebook to run

Code Help

 Find code help

8 updates available

notebook6a6086f4dc

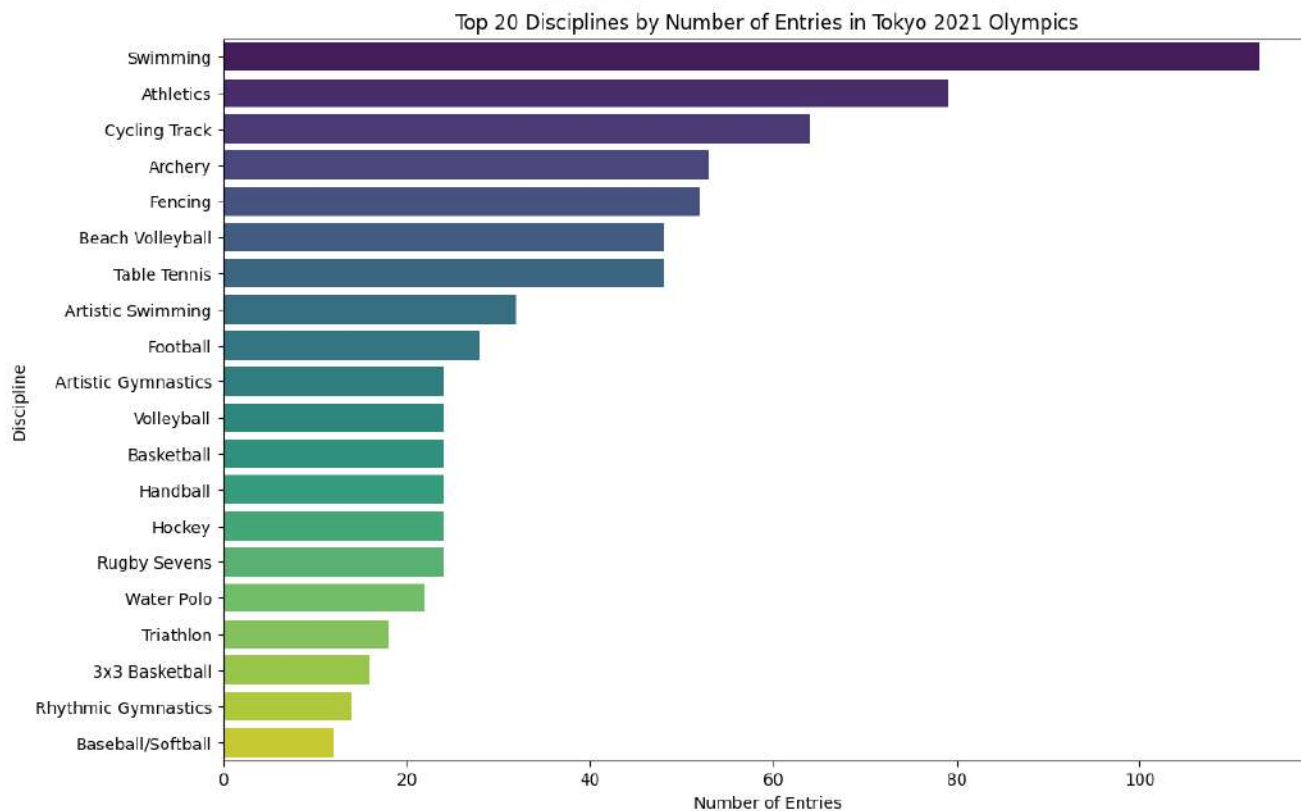
Draft saved

File Edit View Run Add-ons Help

+ ✂ 📄 📄 ▶ ▶▶ Run All Code

Draft Session (1h:3m)

```
plt.xlabel('Number of Entries')
plt.ylabel('Discipline')
plt.show()
```



Notebook

Input

+ Add Input Upload

DATASETS

2021-olympics-in-tokyo

Output (72KB / 19.5GB)

/kaggle/working

Session options

Schedule a notebook to run

Code Help

Find code help

8 updates available

Draft saved

Share Save Version 1



[16]:

Correlation Between Gender Representation and Medal Counts



Input

[+ Add Input](#) [⬆ Upload](#)

DATASETS

► 2021-olympics-in-tokyo

Output (72KB / 19.5GB)

-  /kaggle/working

Session options

Schedule a notebook to run

Code Help

Find code help

notebook6a6086f4dc

Draft saved

File Edit View Run Add-ons Help

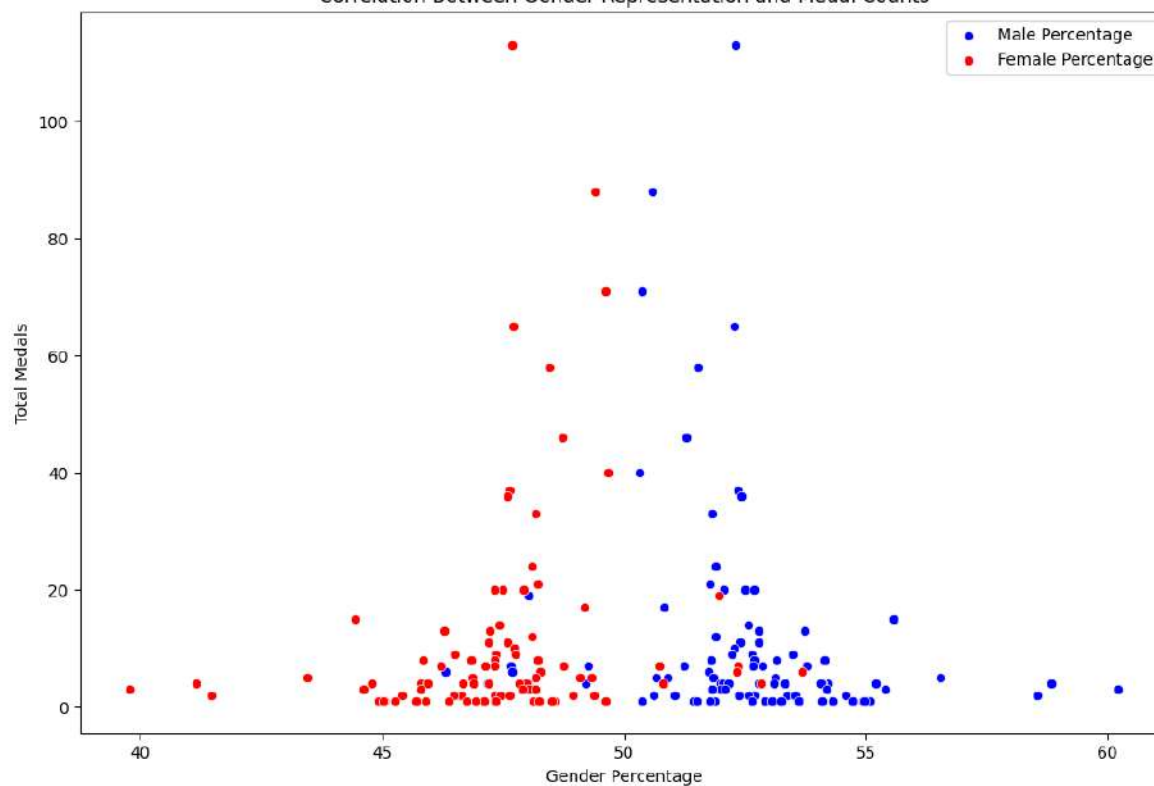
 Share

Save Version 1

 Run All
 Code

● Draft Session (1h:3m)

Correlation Between Gender Representation and Medal Counts



+ Code

+ Markdown

[17]:

```
# Merge medal counts with coaches data
medals_coaches_df = pd.merge(medals_df, coaches_count, left_on='Team/NOC', right_on='NOC', how='left')
```

Notebook

Input

+ Add Input

 Upload

DATASETS

- ▶ 2021-olympics-in-tokyo

Output (72KB / 19.5GB)

- ▶ /kaggle/working

Session options

Schedule a notebook to run

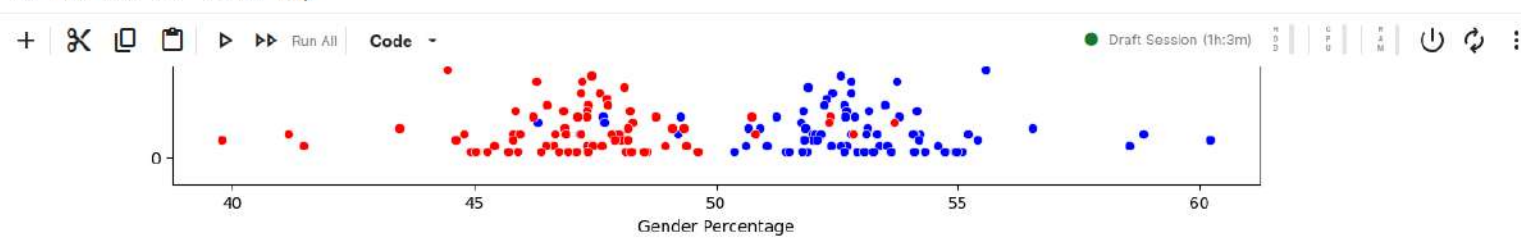
Code Help

 Find code help

8 updates available

notebook6a6086f4dc Draft saved

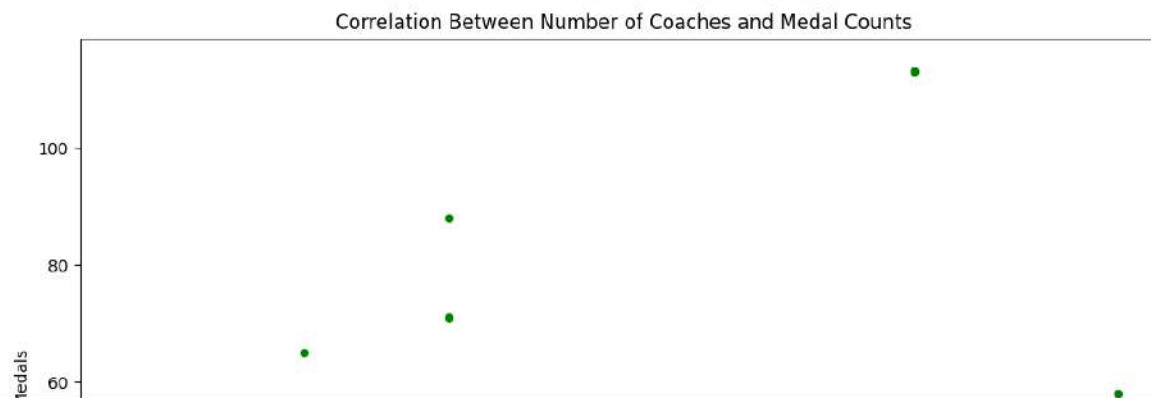
File Edit View Run Add-ons Help



+ Code

+ Markdown

```
[17]:  
# Merge medal counts with coaches data  
medals_coaches_df = pd.merge(medals_df, coaches_count, left_on='Team/NOC', right_on='NOC', how='left')  
  
# Plotting  
plt.figure(figsize=(12, 8))  
sns.scatterplot(x='Name', y='Total', data=medals_coaches_df, color='green')  
plt.title('Correlation Between Number of Coaches and Medal Counts')  
plt.xlabel('Number of Coaches')  
plt.ylabel('Total Medals')  
plt.show()
```



Share Save Version 1

Notebook

Input

+ Add Input

Upload

DATASETS

2021-olympics-in-tokyo

Output (72KB / 19.5GB)

/kaggle/working

Session options

Schedule a notebook to run

Code Help

Find code help

8 updates available

kaggle

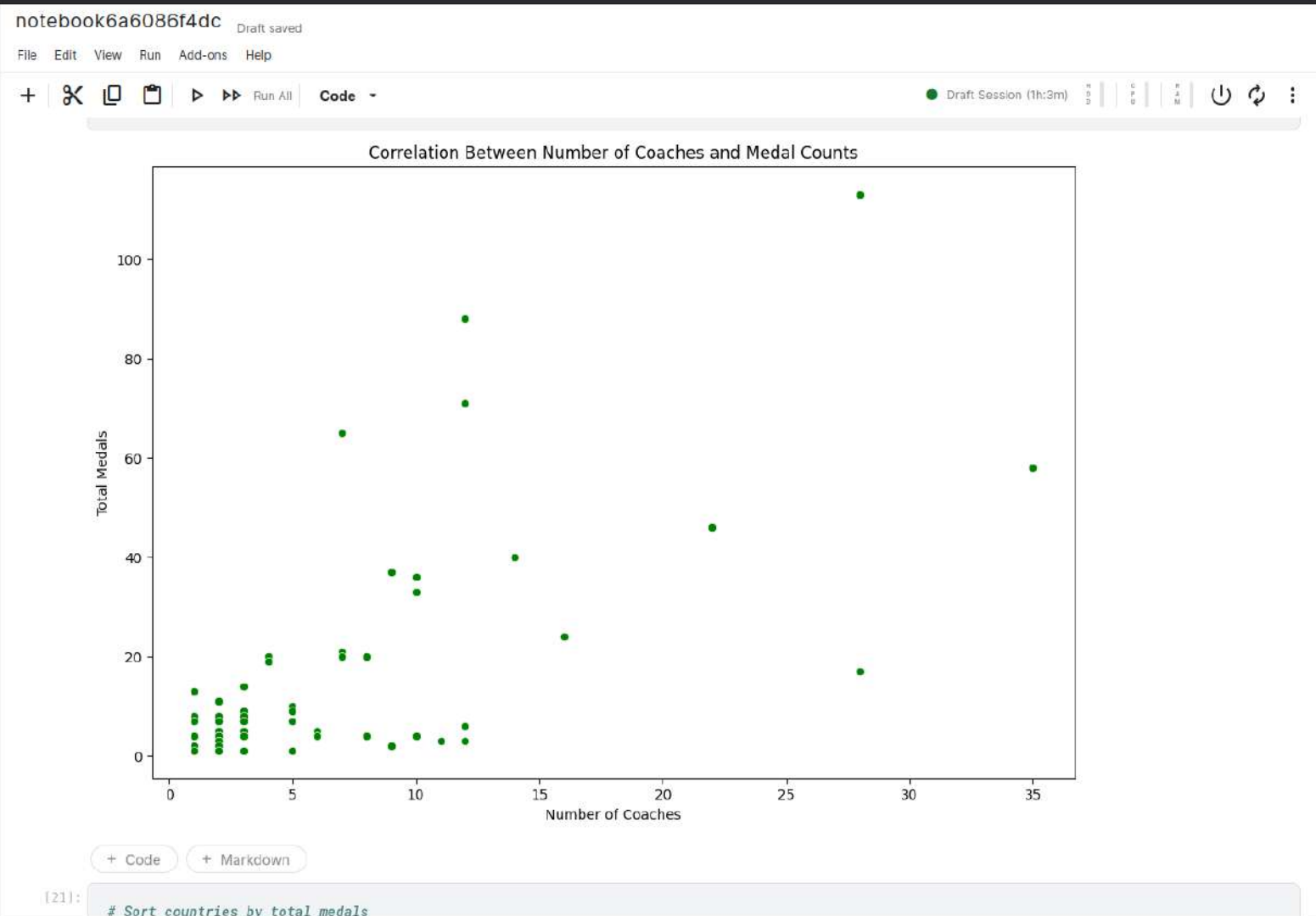
Create

- Home
- Competitions
- Datasets
- Models
- Code
- Discussions
- Learn
- More

Your Work

VIEWED

- 2021 Olympics in Tokyo
- Olympic Games 2021 ...
- Tokyo 2020 Olympics ...
- Tokoyo 2021 Olympics...
- Paris 2024 Olympic Su...



Share Save Version 1

Notebook

Input

+ Add Input Upload

DATASETS

- 2021-olympics-in-tokyo

Output (72KB / 19.5GB)

- /kaggle/working

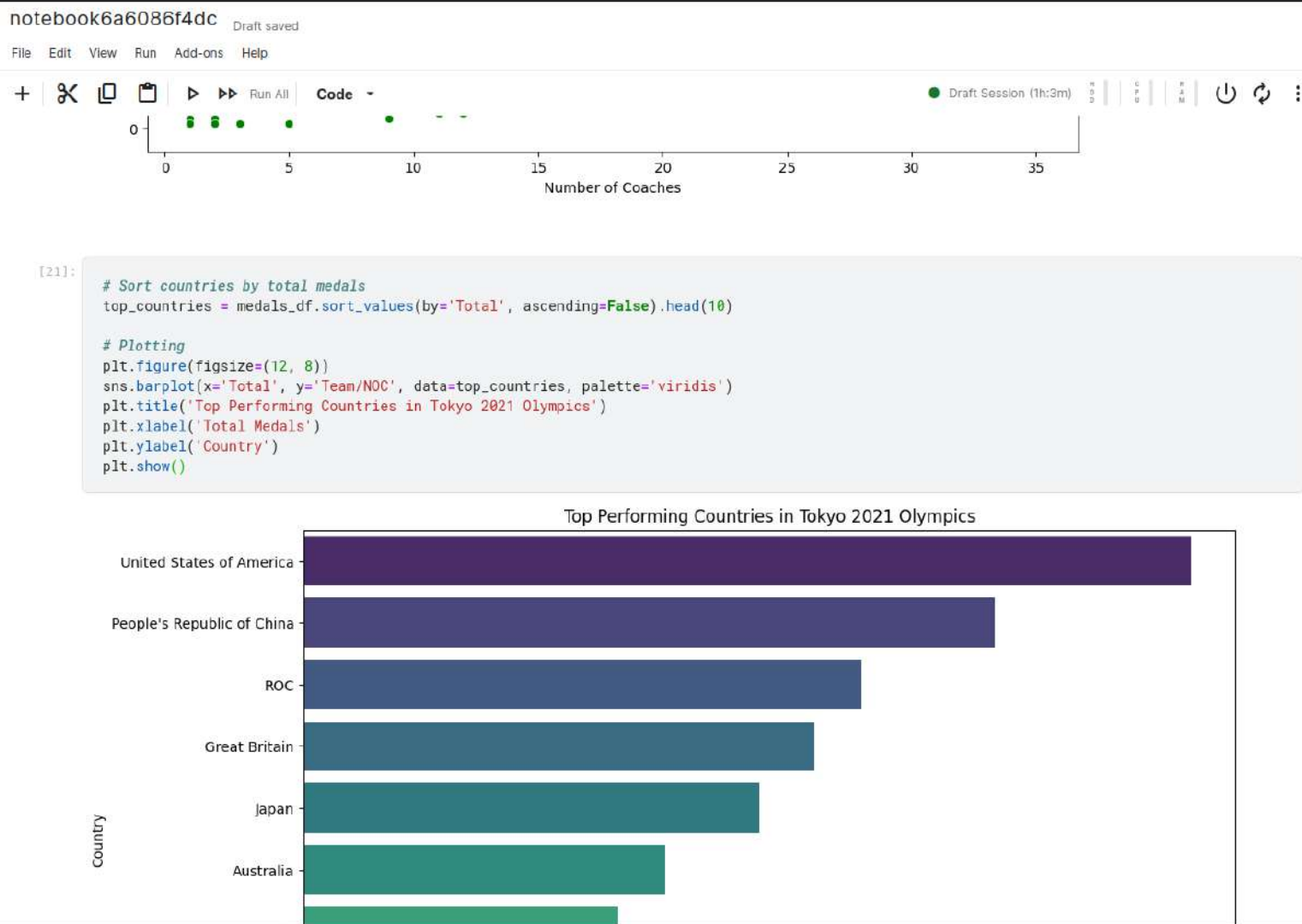
Session options

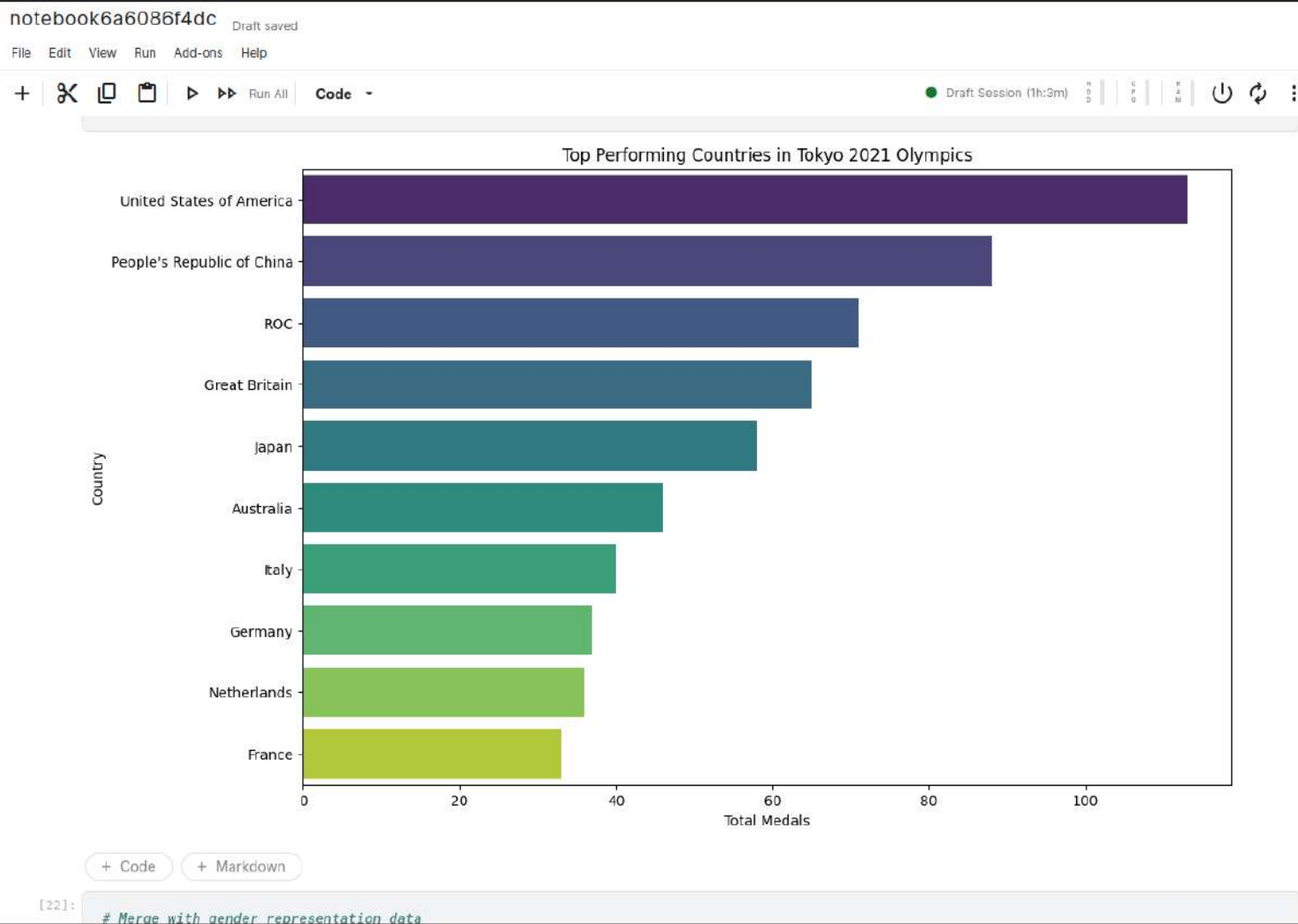
Schedule a notebook to run

Code Help

Find code help

8 updates available





kaggle

+ Create

- Home
- Competitions
- Datasets
- Models
- Code
- Discussions
- Learn
- More

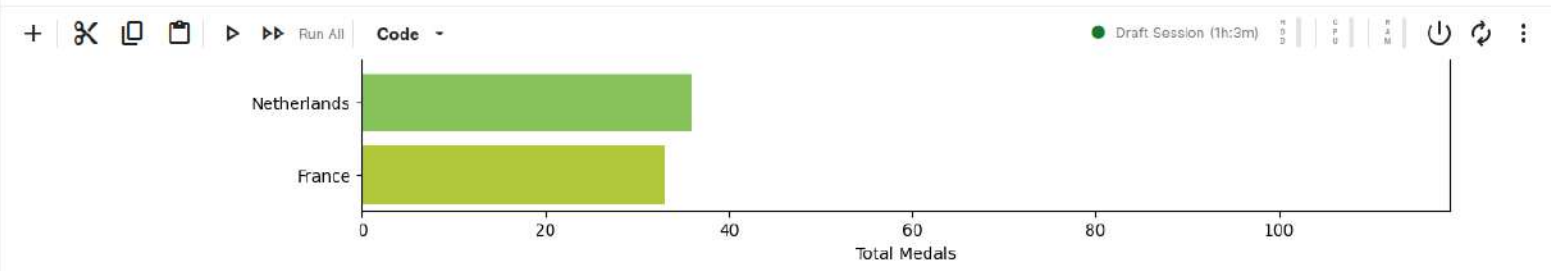
Your Work

VIEWED

- 2021 Olympics in Tokyo
- Olympic Games 2021 ...
- Tokyo 2020 Olympics ...
- Tokoyo 2021 Olympics...
- Paris 2024 Olympic Su...

notebook6a6086f4dc Draft saved

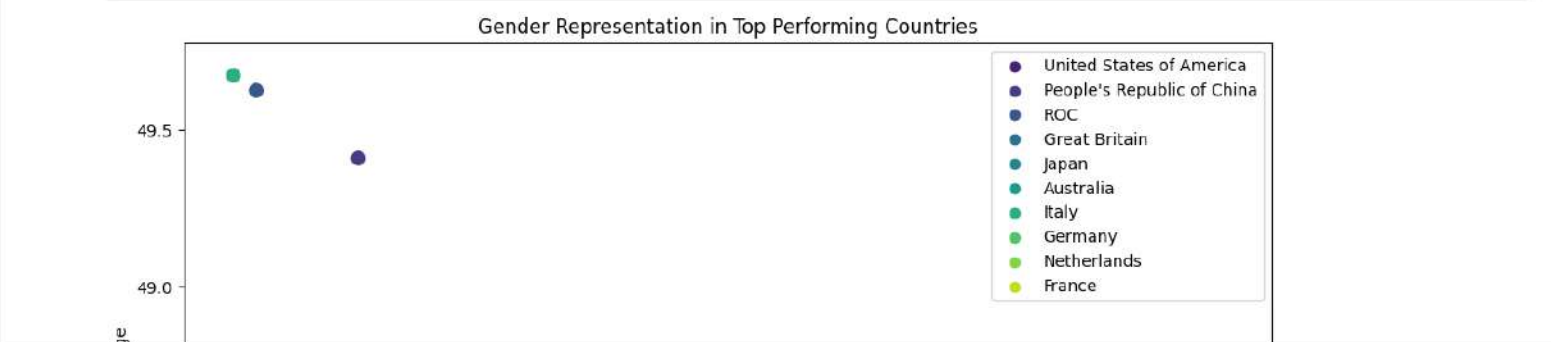
File Edit View Run Add-ons Help



+ Code + Markdown

```
[122]: # Merge with gender representation data
top_countries_gender = pd.merge(top_countries, country_gender_percentages, left_on='Team/NOC', right_on='NOC', how='left')

# Plotting
plt.figure(figsize=(12, 8))
sns.scatterplot(x='Male Percentage', y='Female Percentage', data=top_countries_gender, hue='Team/NOC', palette='viridis', s=100)
plt.title('Gender Representation in Top Performing Countries')
plt.xlabel('Male Percentage')
plt.ylabel('Female Percentage')
plt.legend()
plt.show()
```



Share Save Version 1

Notebook

Input

+ Add Input Upload

DATASETS

- 2021-olympics-in-tokyo

Output (72KB / 19.5GB)

- /kaggle/working

Session options

Schedule a notebook to run

Code Help

Find code help

8 updates available

