Olympics Data Analysis & Insights (1896 - 2016)

Exploring Demographics, Performance & Trends

Rimsha Iram| Python | JupyterLab | 2025

Introduction & Objective

- Dataset: IBM-recommended Olympics dataset
- Objective: Analyze athlete demographics, participation patterns, and medal trends
- Goal: Support athlete selection, training, and strategic planning
- Tools: Python, Pandas, Matplotlib/Seaborn

ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal
1	A Dijiang	M	24.0	180.0	80.0	China	CHN	1992 Summer	1992	Summer	Barcelona	Basketball	Basketball Men's Basketball	NaN
2	A Lamusi	M	23.0	170.0	60.0	China	CHN	2012 Summer	2012	Summer	London	Judo	Judo Men's Extra-Lightweight	NaN
3	Gunnar Nielsen Aaby	M	24.0	NaN	NaN	Denmark	DEN	1920 Summer	1920	Summer	Antwerpen	Football	Football Men's Football	NaN
4	Edgar Lindenau Aabye	M	34.0	NaN	NaN	Denmark/Sweden	DEN	1900 Summer	1900	Summer	Paris	Tug-Of-War	Tug-Of-War Men's Tug-Of-War	Gold
5	Christine Jacoba Aaftink	F	21.0	185.0	82.0	Netherlands	NED	1988 Winter	1988	Winter	Calgary	Speed Skating	Speed Skating Women's 500 metres	NaN

Hypotheses

- Taller/heavier athletes succeed in strength-based sports
- Female participation has increased over time
- Developed countries dominate medal counts
- Older athletes may hold an experience advantage

Data Preparation

- Cleaned dataset: removed duplicates, handled missing values
- Final dataset ready with: Athlete demographics + Sports + Medals
- Exploratory visualizations to uncover trends

Before:

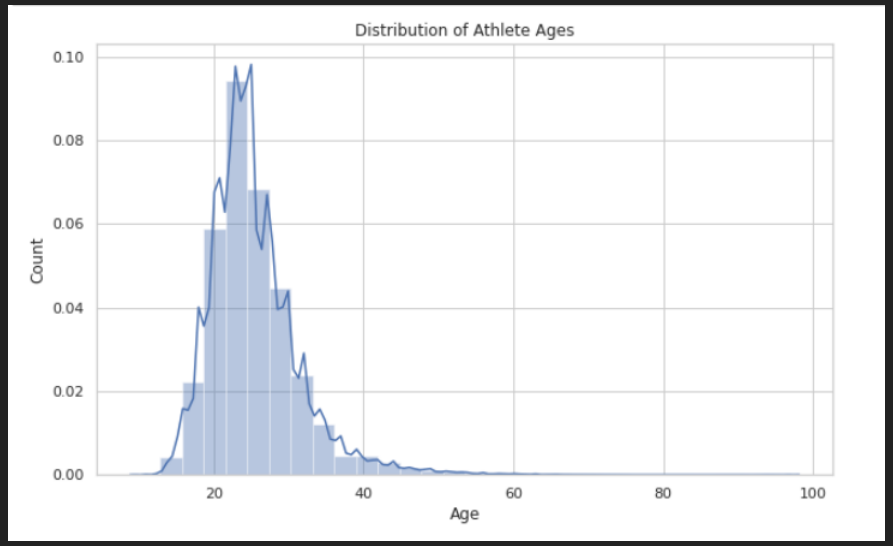
	Missing_Age	Missing_Height	Missing_Weight	Missing_Medal
0	9474	60171	62875	231333

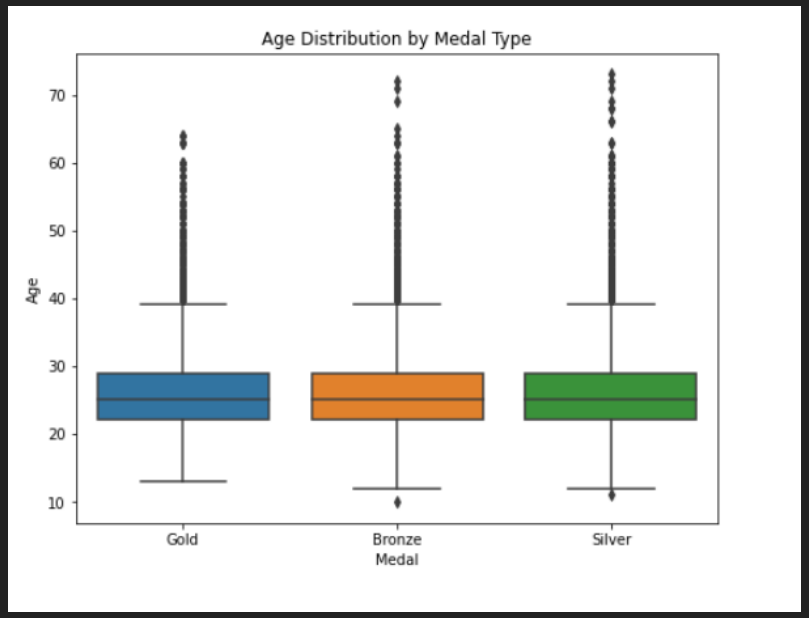
Afetr:

	Missing_Age	Missing_Height	Missing_Weight	Missing_Medal
0	0	0	0	231333

Descriptive Statistics – Age

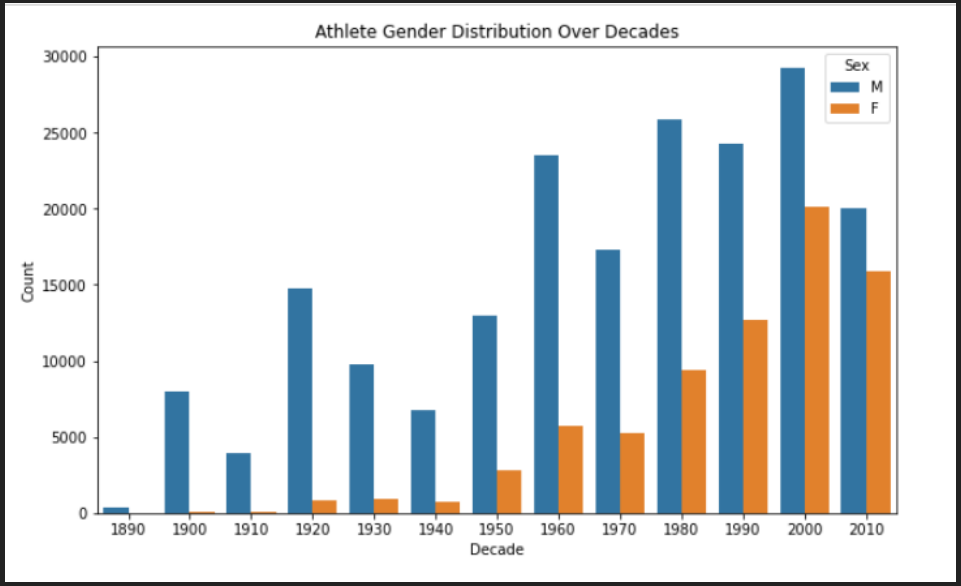
- Analysis of **age distribution among medalists** shows a **mean age of 25.9 years**
- Younger athletes dominate agility-based sports
- Older athletes succeed in experience-heavy sports (e.g., Judo, Basketball)





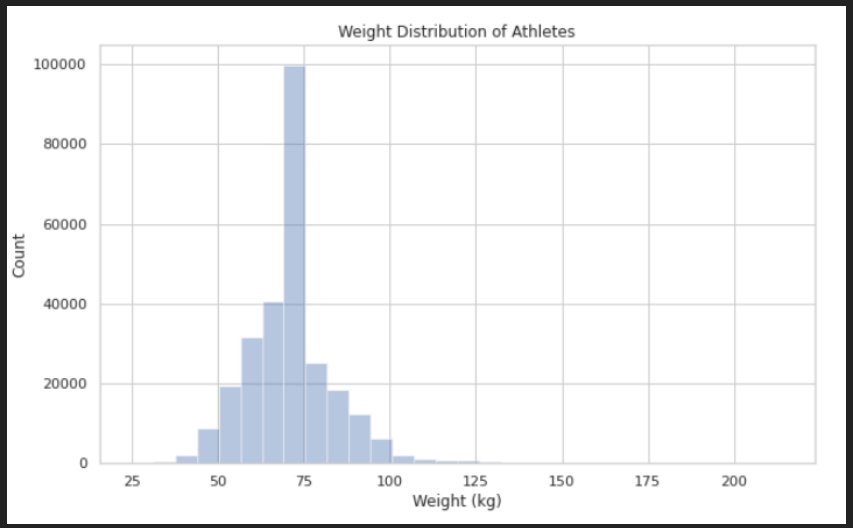
Descriptive Statistics – Gender

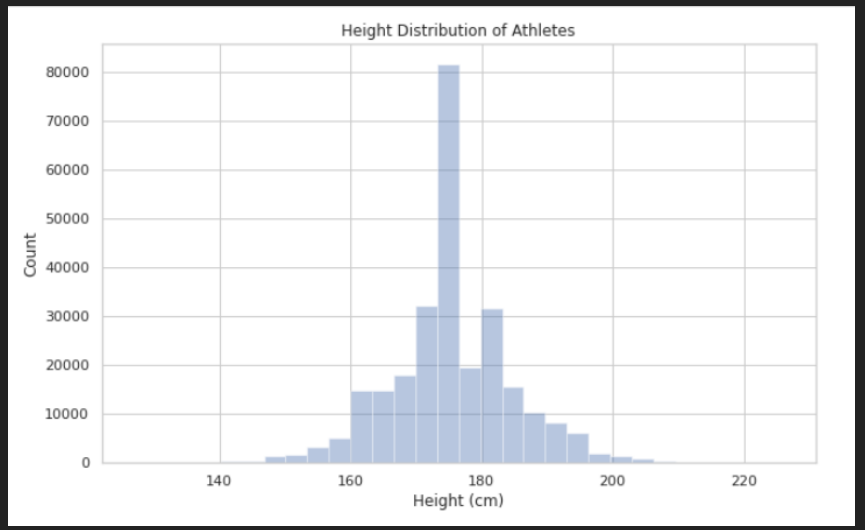
- Male vs Female participation: steady rise in female athletes
- Key trend: Inclusivity improving over decades
- Still male-dominated in some sports



Descriptive Statistics – Height & Weight

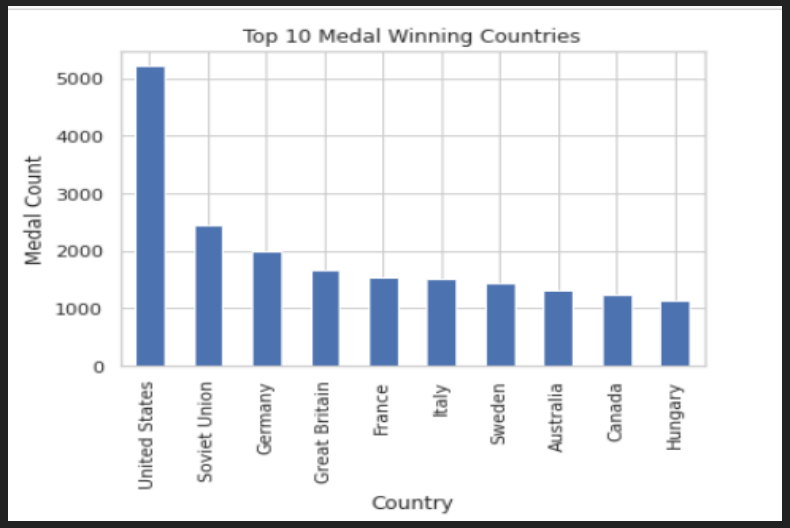
- Height/Weight distributions show natural variation
- Extreme values mostly from sport-specific requirements
- Example: Basketball players are taller & heavier; Gymnasts are shorter & lighter





Medal Trends

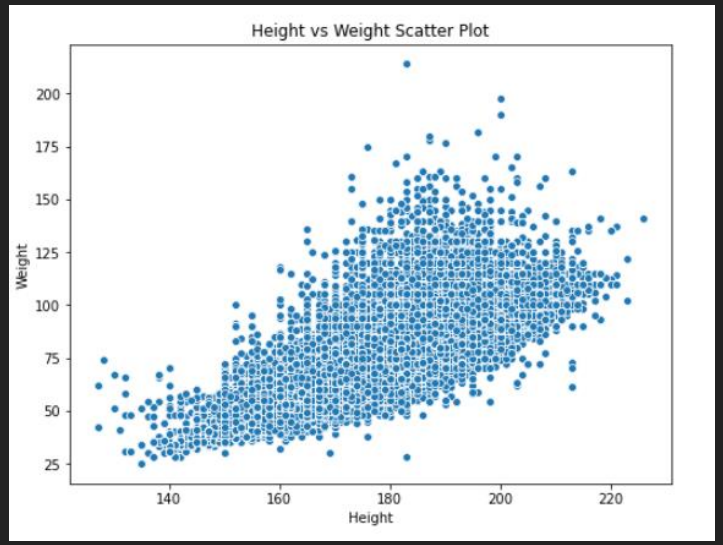
- USA and URS dominate overall medal counts
- Smaller countries show high medal efficiency
- Summer Games dominate over Winter in athlete & medal counts



	NOC	Total_Athletes	Total_Medals	Medal_Efficiency
139	URS	5685	2503	0.440281
47	GDR	2645	1005	0.379962
3	ANZ	86	29	0.337209
40	EUN	864	279	0.322917
141	USA	18853	5637	0.298998
145	WIF	20	5	0.250000
112	RUS	5143	1165	0.226521
49	GER	9830	2165	0.220244
117	SRB	392	85	0.216837
100	PAK	562	121	0.215302

Correlations & Deeper Insights

- Height & Weight correlation = 0.796 overall
- Stronger correlation in Basketball (0.87) & Speed Skating (0.88)
- Confirms physique-sport alignment



Gender & Country Patterns

- Female participation up → more medals for women
- Medal efficiency highlights: URS, GDR, ANZ
- Insights: Some countries focus on quality over quantity

	NOC	Total_Athletes	Total_Medals	Medal_Efficiency
139	URS	5685	2503	0.440281
47	GDR	2645	1005	0.379962
3	ANZ	86	29	0.337209
40	EUN	864	279	0.322917
141	USA	18853	5637	0.298998
145	WIF	20	5	0.250000
112	RUS	5143	1165	0.226521
49	GER	9830	2165	0.220244
117	SRB	392	85	0.216837
100	PAK	562	121	0.215302

New Metrics Developed

- Medal Efficiency = Total Medals ÷ Total Athletes
- Age Advantage = Avg Medalist Age − Avg Athlete Age (≈ +0.37 years)
- Strategic insights beyond raw medal counts

	Year	NOC	Athlete_Count
0	1896	AUS	5
1	1896	AUT	8
2	1896	DEN	15
3	1896	FRA	26
4	1896	GBR	25
5	1896	GER	94
6	1896	GRE	148
7	1896	HUN	18
8	1896	ITA	1
9	1896	SUI	8

Outliers & Discoveries

- Extreme medal efficiency → small delegations outperforming
- Height/Age outliers → natural variation, not significant
- Key finding: Medal efficiency is the most powerful differentiator

	NOC	Total_Athletes	Total_Medals	Medal_Efficiency
3	ANZ	86	29	0.337209
47	GDR	2645	1005	0.379962
139	URS	5685	2503	0.440281

Key Insights & Recommendations

- Physique matters: Taller/heavier athletes succeed in strength sports
- Experience counts: Slight age advantage for medalists
- Inclusivity rising: Female participation steadily growing
- Efficiency matters: Some countries outperform with fewer athletes
- Recommendation: Use metrics like Medal Efficiency in scouting & training

Future Directions

- Apply regression models to predict medal outcomes
- Build interactive dashboards (Power BI/Tableau) for storytelling
- Extend analysis to textual/narrative Olympic data

Thank You!

Turning Data Into Insights for Smarter Sports Strategies

GitHub link: Rimsha-Iram (Rimsha Iram)

LinkedIn profile: www.linkedin.com/in/rimsha-iram-analytics