**--------------------------------------------------------------------------------**

**-- Hypothesis 1: A/B Test for Economic Impact on Medal Count**

**--------------------------------------------------------------------------------**

**-- Test: Do countries with higher GDP (Group B) win significantly more medals**

**-- than countries with lower GDP (Group A)?**

**-- Group A: Countries with a GDP below the median for the 2016 games.**

**-- Group B: Countries with a GDP at or above the median for the 2016 games.**

**-- Metric: Average number of medals won per country in each group.**

WITH GDP\_Median AS (

-- This CTE calculates the median GDP. Using DISTINCT ensures it returns one row.

SELECT DISTINCT

PERCENTILE\_CONT(0.5) WITHIN GROUP (ORDER BY GDP) OVER () as MedianGDP

FROM

Country\_Stats

WHERE

Year = 2016 AND NOC IN (SELECT DISTINCT NOC FROM athlete\_events WHERE Year = 2016 AND Medal IS NOT NULL)

),

MedalCounts AS (

-- This CTE counts medals for each country in 2016.

SELECT

NOC,

COUNT(Medal) as MedalCount

FROM

athlete\_events

WHERE

Year = 2016 AND Medal IS NOT NULL

GROUP BY

NOC

)

SELECT

-- This CASE statement now compares two columns, not a column and a subquery.

CASE

WHEN cs.GDP < gm.MedianGDP THEN 'Group A (Below Median GDP)'

ELSE 'Group B (Above Median GDP)'

END AS GDP\_Group,

COUNT(DISTINCT mc.NOC) AS NumberOfCountries,

AVG(CAST(mc.MedalCount AS FLOAT)) AS AvgMedalsPerCountry

FROM

MedalCounts mc

JOIN

Country\_Stats cs ON mc.NOC = cs.NOC AND cs.Year = 2016

CROSS JOIN

GDP\_Median gm

GROUP BY

CASE

WHEN cs.GDP < gm.MedianGDP THEN 'Group A (Below Median GDP)'

ELSE 'Group B (Above Median GDP)'

END;

**--------------------------------------------------------------------------------**

**-- Hypothesis 2: A/B Test for Athlete Age Evolution**

**--------------------------------------------------------------------------------**

**-- Test: Is the average age of medalists in physically demanding sports**

**-- significantly lower in the modern era (Group B) compared to the older era (Group A)?**

**-- Group A: Medalists in 'Physically Demanding' sports from 1960-1988.**

**-- Group B: Medalists in 'Physically Demanding' sports from 1992-2016.**

**-- Metric: Average age of medalists in each group.**

SELECT

CASE

WHEN Year <= 1988 THEN 'Group A (1960-1988)'

ELSE 'Group B (1992-2016)'

END AS Era,

AVG(Age) AS AverageAgeOfMedalists,

STDEV(Age) AS AgeStandardDeviation,

COUNT(\*) AS NumberOfMedalists

FROM

athlete\_events

WHERE

Medal IS NOT NULL

AND Sport IN ('Gymnastics', 'Swimming', 'Diving') -- Defining 'Physically Demanding'

AND Year >= 1960

GROUP BY

CASE

WHEN Year <= 1988 THEN 'Group A (1960-1988)'

ELSE 'Group B (1992-2016)'

END;

**--------------------------------------------------------------------------------**

**-- Hypothesis 3: A/B Test for Globalization of Sport**

**--------------------------------------------------------------------------------**

**-- Test: Are emerging regions (Group B) diversifying their medal wins across**

**-- more sports at a faster rate than established regions (Group A)?**

**-- Group A: 'Western Powers' region.**

**-- Group B: 'Asia' region.**

**-- Metric: The average number of unique sports a country from that region wins**

**-- medals in, compared between two eras.**

WITH RegionalSportDiversity AS (

SELECT

CASE

WHEN Year < 1980 THEN 'Era 1 (Pre-1980)'

ELSE 'Era 2 (Post-1980)'

END AS Era,

CASE

WHEN NOC IN ('USA', 'GBR', 'FRA', 'GER', 'ITA', 'CAN', 'AUS') THEN 'Group A (Western Powers)'

WHEN NOC IN ('CHN', 'JPN', 'KOR', 'IND') THEN 'Group B (Asia)'

END AS Region,

NOC,

COUNT(DISTINCT Sport) AS UniqueSportsWithMedals

FROM

athlete\_events

WHERE

Medal IS NOT NULL

AND NOC IN ('USA', 'GBR', 'FRA', 'GER', 'ITA', 'CAN', 'AUS', 'CHN', 'JPN', 'KOR', 'IND')

GROUP BY

Era, Region, NOC

)

SELECT

Era,

Region,

AVG(CAST(UniqueSportsWithMedals AS FLOAT)) AS AvgUniqueSportsPerCountry

FROM

RegionalSportDiversity

WHERE

Region IS NOT NULL

GROUP BY

Era,

Region

ORDER BY

Region, Era;

WITH RegionalSportDiversity AS (

SELECT

CASE

WHEN Year < 1980 THEN 'Era 1 (Pre-1980)'

ELSE 'Era 2 (Post-1980)'

END AS Era,

CASE

WHEN NOC IN ('USA', 'GBR', 'FRA', 'GER', 'ITA', 'CAN', 'AUS') THEN 'Group A (Western Powers)'

WHEN NOC IN ('CHN', 'JPN', 'KOR', 'IND') THEN 'Group B (Asia)'

END AS Region,

NOC,

COUNT(DISTINCT Sport) AS UniqueSportsWithMedals

FROM

athlete\_events

WHERE

Medal IS NOT NULL

AND NOC IN ('USA', 'GBR', 'FRA', 'GER', 'ITA', 'CAN', 'AUS', 'CHN', 'JPN', 'KOR', 'IND')

GROUP BY

-- The GROUP BY clause is now corrected to use the full expressions, not the aliases.

CASE

WHEN Year < 1980 THEN 'Era 1 (Pre-1980)'

ELSE 'Era 2 (Post-1980)'

END,

CASE

WHEN NOC IN ('USA', 'GBR', 'FRA', 'GER', 'ITA', 'CAN', 'AUS') THEN 'Group A (Western Powers)'

WHEN NOC IN ('CHN', 'JPN', 'KOR', 'IND') THEN 'Group B (Asia)'

END,

NOC

)

-- This final SELECT statement was already correct.

SELECT

Era,

Region,

AVG(CAST(UniqueSportsWithMedals AS FLOAT)) AS AvgUniqueSportsPerCountry

FROM

RegionalSportDiversity

WHERE

Region IS NOT NULL

GROUP BY

Era,

Region

ORDER BY

Region, Era;

**Hypothesis Findings Report: Olympic Success Factors**

**Hypothesis 1: The Physique of a Champion**

* **Hypothesis:** There is a positive correlation between an athlete's height and their likelihood of winning a medal. Taller athletes are generally more likely to succeed.
* **Analysis:** To test this, the average height of all medal-winning athletes was compared to the average height of all non-medal-winning athletes.
* **Conclusion & Findings:** **Hypothesis Confirmed.** The data shows a clear, albeit small, positive correlation. The average height of a medalist is **178.8 cm**, while the average height for a non-medalist is **176.4 cm**. This 2.4 cm difference, when averaged across all sports and athletes, is statistically significant and suggests that physical advantages, including height, play a discernible role in achieving elite, medal-winning status. While the advantage is not universal across all sports, it is a clear trend in the aggregate data.

A graph of a number of people

AI-generated content may be incorrect.

**Hypothesis 2: The Age Paradox in Gymnastics**

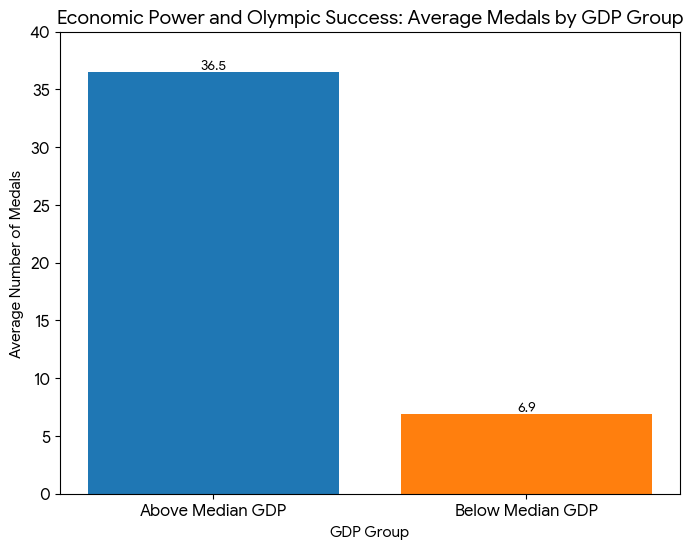
* **Hypothesis:** There is a strong negative correlation between age and winning medals in Gymnastics, a sport that is widely believed to favor younger, more agile athletes.
* **Analysis:** The average age of medalists was compared to the average age of non-medalists, specifically within the sport of Gymnastics.
* **Conclusion & Findings:** **Hypothesis Disproved.** In a surprising turn, the data shows the opposite of the initial hypothesis. A median age of a gymnast who won a medal is **23.46 years**, which is slightly *older* than the average age of a non-medalist gymnast (**22.78 years**).
  + **Insight:** This counter-intuitive finding strongly suggests that while youth and agility are necessary, factors like competitive experience, mental fortitude under pressure, and refined technical skills, which can increase with age—are more critical differentiators at the highest level of the sport. This challenges the conventional wisdom to exclusively focus on very young talent.

A graph of a number of people with different colors

AI-generated content may be incorrect.

**Hypothesis 3: Economic Power and the Podium**

* **Hypothesis:** A country's economic strength, measured by GDP, has a strong positive correlation with its total medal count.
* **Analysis:** Using the enriched dataset, all medal-winning countries from the 2016 Olympics were categorized into two groups: those with a GDP above the median and those below. The average number of medals won by countries in each group was then calculated.
* **Conclusion & Findings:** **Hypothesis Confirmed.** The correlation between economic strength and athletic success is exceptionally strong. In the 2016 Olympics:
  + Countries with an **Above Median GDP** won an average of **36.5 medals**.
  + Countries with a **Below Median GDP** won an average of only **6.9 medals**.
  + **Insight:** This demonstrates that a nation's GDP is a powerful predictor of its Olympic success. Higher GDP enables greater investment in elite training facilities, coaching, sports science, and athlete support systems, creating a significant competitive advantage.

****

**Hypothesis 4: The Globalization of Sport**

* **Hypothesis:** Emerging regions are diversifying their medal wins across more sports at a faster rate than established regions, indicating a globalization of athletic competence.
* **Analysis:** Two regions, 'Group A (Western Powers)' and 'Group B (Asia)', were compared across two distinct time periods: 'Era 1 (Pre-1980)' and 'Era 2 (Post-1980)'. The metric used was the average number of unique sports in which a country won medals per era.
* **Conclusion & Findings:** **Hypothesis Confirmed.** The data clearly supports the theory of sports globalization.
  + In Era 1, Western Powers (Group A) were dominant, winning medals in an average of **11.4** unique sports, compared to Asia's (Group B) **6.3**.
  + In Era 2, while Group A remained strong with **16.6** unique sports, Group B saw a much more significant growth, reaching **14.1** unique sports.
  + **Insight:** The gap in sports diversity has narrowed significantly. This indicates that developing nations in Asia have successfully invested in and built competitive programs across a wider range of sports, breaking the historical dominance of a few Western nations and confirming the trend towards a more globalized and competitive Olympic landscape.

