Group: TKDM

Trinity, Keegan, Devin, Mads

02/11/25

Data Set: **Olympic Swimming 1924 to 2020**

Project Task Plan

The Olympic swimming dataset provides a comprehensive look at the performance of swimmers across multiple Summer Olympic Games. It contains 606 rows and 10 variables, with each row representing a swimmer who completed their race. Key variables include the year, location, stroke type, race distance, gender, country (team), and the swimmer's result time and rank. This dataset allows us to explore performance trends over time, compare gender differences, and analyze which countries have dominated Olympic swimming from 1924 to 2020.

1. How has the performance of Olympic swimmers evolved over across different strokes for 100 m? (time) trend chart, different variables, race graph, flourish

- This question allows you to analyze trends in swimmer performance by comparing the average race times across different time periods (early vs. recent), strokes, and distances. We plan to explore whether advancements in training, technology, and technique have led to faster completion times and whether certain strokes or distances show more improvement over time.

2. Is there a significant difference in performance between male and female swimmers across strokes and time periods? (regression)

time =
Beta 0 +

- This question investigates **g**ender-based performance differences by comparing the completion times of male and female swimmers. We will analyze how these differences vary across strokes and whether the performance gap has changed over time, providing insights into gender equity in sports and the evolution of athletic ability.

3. Which countries have dominated Olympic swimming over time, and how has their performance changed between the early and recent time periods? (race graph)

- This question focuses on country-level performance by evaluating which nations consistently achieve top ranks. We will examine whether the distribution of winning

countries has become more or less diverse over time and identify emerging powerhouses or fading giants in Olympic swimming.

Task Planning for Olympic Swimming Data Analysis

1. How has the performance of Olympic swimmers evolved over time across different strokes and distances?

Steps to Complete This Task:

1. **Data Preparation:**
    a. Load and explore the dataset to understand the structure and identify missing or inconsistent data.
    b. Filter the dataset by year, stroke, and distance for more focused analysis.
2. **Performance Trends Analysis:**
    a. Calculate the average race times for each stroke and distance across both early (1924-1972) and recent (1976-2020) time periods.
    b. Compare how performance has changed over the years by visualizing average results for each stroke and distance using line plots or box plots.
3. **Identify Key Milestones:**
    a. Highlight major performance shifts by identifying years where significant improvements occurred.
    b. Investigate whether these changes align with key events like technological advancements (e.g., swimsuit innovations) or training improvements.
4. **Sub-Questions:**
    a. Are swimmers getting faster over time, and if so, which strokes and distances show the greatest improvement?
    b. Do performance trends vary consistently between the early and recent time periods?

2. Is there a significant difference in performance between male and female swimmers across strokes and time periods?

Steps to Complete This Task:

1. **Data Segmentation:**
    a. Separate the data by gender and further segment by stroke and time period.
    b. Ensure consistent comparison by controlling for variables like distance and Olympic year.
2. **Statistical Comparison:**

a. Calculate and compare the mean race times for male and female swimmers across all strokes and time periods.
b. Conduct statistical tests (e.g., t-tests or ANOVA) to determine if gender-based performance differences are statistically significant.

3. **Visualization of Differences:**
    a. Create side-by-side box plots or histograms comparing male and female race times for each stroke and time period.
    b. Use scatter plots to show how performance for each gender has evolved across decades.
4. **Sub-Questions:**
    a. Is there a measurable performance gap between male and female swimmers across different strokes?
    b. Has the gender performance gap changed over time, and if so, in which strokes or time periods is the change most apparent?

3. Which countries have dominated Olympic swimming over time, and how has their performance changed between the early and recent time periods?

Steps to Complete This Task:

1. **Data Grouping by Country:**
    a. Group the dataset by Team (country code) and time period to analyze national performance.
    b. Identify which countries have the most top finishes (Rank = 1) in each period.
2. **Analyze Performance Consistency:**
    a. Calculate the average rank and number of podium finishes for each country across all years.
    b. Identify countries with consistent dominance and those with emerging or declining performance.
3. **Visualize Country Trends:**
    a. Create a stacked bar chart or heatmap showing the distribution of top finishes by country and time period.
    b. Use a line graph to show how the performance of major swimming nations has evolved over time.
4. **Sub-Questions:**
    a. Which countries have dominated Olympic swimming historically, and have these dominant nations changed over time?
    b. Are there any emerging nations gaining prominence in recent Olympic Games?

This structured approach will allow a thorough analysis of performance trends, gender differences, and national dominance in Olympic swimming.