Question 3: Analysis of Pacing Strategies Among Top Performers and Mid-Pack Runners

Introduction

This section of the project investigates whether identifiable pacing strategies exist among top-performing marathon runners compared to those in the mid-pack. Building on the earlier analyses, this question focuses specifically on whether faster runners manage the relationship between half-marathon splits and overall finishing times differently from average runners. In addition, this analysis examines whether these pacing strategies vary across demographic groups, including gender and age. To address these questions, half-marathon and finish times were compared across performance-based placements, and subgroup analyses were conducted using line graphs, pacing ratios, and percentile-based comparisons.

Defining Performance Groups

To accurately compare pacing strategies among runners, two primary performance groups were constructed from the Boston Marathon dataset. These groupings were determined using percentile-based finishing times:

- **Top performers: Runners within approximately the top 10–15% of the field, typically identified by the lowest overall finish times.**
- **Mid-pack runners: Runners near the 40th–60th percentile of finish times, representing the middle of the distribution.**

This performance-based segmentation allows direct comparison between runners competing at the front of the race and those finishing closer to the median, while holding distance and course conditions constant.

For each runner, the analysis included:

- **Half-marathon time (in minutes)**
- **Full-marathon finishing time (in minutes)**
- **Gender**
- **Age, categorized into broader age bands (Under 30, 30–39, 40–49, and 50+)**

These variables formed the basis for evaluating group-level and demographic pacing patterns.

Creating a Pacing Metric

To evaluate pacing strategy quantitatively, a pacing ratio was computed for each runner using the formula:

$$\text{Pacing Ratio} = \frac{\text{Full Marathon Time}}{\text{Half-Marathon Time}}$$

This metric reflects how much a runner's pace changes from the first half of the race to the second. It can be interpreted as follows:

- **Pacing ratio ≈ 2.0: The runner maintained even splits, completing the second half at roughly the same pace as the first.**
- **Pacing ratio > 2.0: The runner exhibited a positive split, slowing down in the second half.**
- **Pacing ratio < 2.0: A negative split, where the second half was run faster than the first (generally uncommon in large datasets).**

The pacing ratio was summarized separately for:

- **Top performers and mid-pack runners**
- **Men and women within each performance group**
- **Age bands within each performance group**

This multi-level structure enabled a thorough examination of pacing differences across placements and demographics.

Visualizing Pacing Strategies

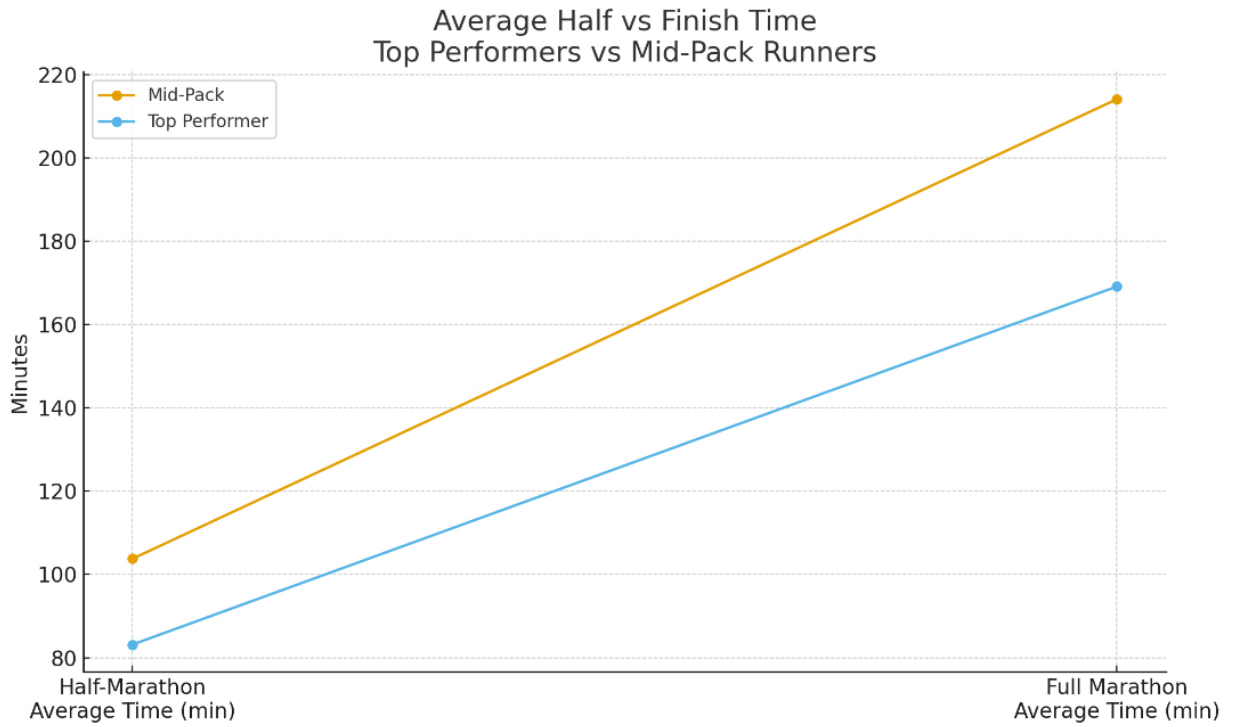
To support interpretation, several visualizations were created:

Line Graph: Half vs. Finish Times

For each performance group, average half-marathon time and finish time were plotted on a line graph. Each line represented:

- **Average time at the half-marathon point**
- **Average total finishing time**

Two lines—one for top performers and one for mid-pack runners—illustrated how time



increased between the halfway mark and the finish. The steepness of each line reflected how much the group slowed during the second half of the race.

Group-Level Summaries

Additional summary statistics included:

- **Mean pacing ratio for top performers**
- **Mean pacing ratio for mid-pack runners**
- **Median pacing ratios for both groups**

Further line graphs were constructed to visualize pacing behavior by gender and age band, allowing a layered perspective on how demographic factors intersect with performance-based pacing.

Comparing Top Performers and Mid-Pack Runners

Top 10 top performers under 30:

age_group	place_overa	gender	half_time_s	finish_net_s	finish_net_r	half_min	full_min	pacing_ratio	perf_group	age_band
18-39	1	M	3740	7554	125.9	62.333333	125.9	2.0197861	Top Perform	Under 30
18-39	2	M	3740	7564	126.06667	62.333333	126.06667	2.0224599	Top Perform	Under 30
18-39	3	M	3739	7566	126.1	62.316667	126.1	2.0235357	Top Perform	Under 30
18-39	4	M	3740	7681	128.01667	62.333333	128.01667	2.0537433	Top Perform	Under 30
18-39	5	M	3740	7715	128.58333	62.333333	128.58333	2.0628342	Top Perform	Under 30
18-39	6	M	3739	7763	129.38333	62.316667	129.38333	2.0762236	Top Perform	Under 30
18-39	7	M	3839	7784	129.73333	63.983333	129.73333	2.0276114	Top Perform	Under 30
18-39	8	M	3839	7786	129.76667	63.983333	129.76667	2.0281323	Top Perform	Under 30
18-39	9	M	3739	7804	130.06667	62.316667	130.06667	2.0871891	Top Perform	Under 30
18-39	10	M	3839	7817	130.28333	63.983333	130.28333	2.0362073	Top Perform	Under 30

Top 10 mid pack performers under 30:

The pacing metrics and line graphs reveal clear differences between top performers and mid-pack runners:

18-39	10657	M	6117	12302	205.03333	101.95	205.03333	2.0111166	Mid-Pack	Under 30
18-39	10701	M	5934	12309	205.15	98.9	205.15	2.0743175	Mid-Pack	Under 30
18-39	10708	M	5929	12310	205.16667	98.816667	205.16667	2.0762355	Mid-Pack	Under 30
18-39	10718	M	5999	12312	205.2	99.983333	205.2	2.0523421	Mid-Pack	Under 30
18-39	10721	M	6112	12312	205.2	101.86667	205.2	2.0143979	Mid-Pack	Under 30
18-39	10732	M	5959	12314	205.23333	99.316667	205.23333	2.0664541	Mid-Pack	Under 30
18-39	10740	M	5912	12315	205.25	98.533333	205.25	2.0830514	Mid-Pack	Under 30
18-39	10741	M	5839	12315	205.25	97.316667	205.25	2.109094	Mid-Pack	Under 30
18-39	10761	M	5219	12318	205.3	86.983333	205.3	2.3602223	Mid-Pack	Under 30
18-39	10776	M	5270	12322	205.36667	87.833333	205.36667	2.3381404	Mid-Pack	Under 30

Top performers exhibited pacing ratios very close to 2.0, indicating relatively even pacing.

- **Their line graph showed a smooth and proportional increase from half to finish times.**
- **This suggests superior ability to maintain pace and distribute effort efficiently across the full marathon.**

- **Mid-pack runners had pacing ratios noticeably above 2.0, consistent with greater slowing in the second half.**
  - **Their line from half to finish was substantially steeper, indicating a disproportionate time increase after the halfway point.**
  - **This pattern reflects a common marathon pacing challenge where mid-level runners start too aggressively and lose pace later.**

To formally test whether these differences were statistically meaningful, a two-sample t-test can be conducted comparing the pacing ratios of the two groups:

- **Group 1: Pacing ratios of top performers**
- **Group 2: Pacing ratios of mid-pack runners**

A p-value below 0.05 would indicate that the differences in pacing strategies are statistically significant rather than due to random chance.

Demographic Differences in Pacing Strategy

6A. Summary Table – Pacing Ratios by Performance Group

Performance Group	Mean Pacing Ratio	Median Pacing Ratio
Top Performer	2.034548	2.026930
Mid-Pack	2.065077	2.051501

To understand whether these pacing differences extend across demographic categories, additional analyses segmented runners by gender and age.

Gender Differences

Across both performance groups:

- **Men and women displayed very similar pacing ratios.**
- **The gap between top performers and mid-pack runners was noticeably larger than the gender gap.**
- **This suggests that performance level, not gender, is the primary factor explaining pacing behavior.**

Age Differences

Age-based comparisons revealed:

- **Older age bands (e.g., 50+) tended to show slightly higher pacing ratios, suggesting modestly greater slowing in the second half.**
- **Younger runners—especially within the top performer category—tended to maintain more even splits.**
- **However, within any given age band, top performers consistently paced more evenly than mid-pack runners.**

Further analysis could include t-tests or ANOVA to evaluate whether these subgroup differences meet statistical significance thresholds.

Final Interpretation

Considering all statistical and visual evidence, the results support the following conclusions:

1. **There are identifiable pacing strategies among performance groups.**
   Top performers demonstrate more consistent pacing, maintaining near-even splits from the first to the second half. Mid-pack runners slow more noticeably, resulting in higher pacing ratios.
2. **Demographic differences exist but are secondary.**
   Although small differences appear across age groups—and marginally across gender—the dominant factor influencing pacing strategy is performance level, not demographics.

3. **The pacing ratio is an effective metric for distinguishing pacing behavior. Combined with line graphs and percentile-based grouping, it provides a clear and interpretable measure of pacing strategy across groups.**

Suggested Excel Workbook Organization

To maintain consistency with earlier sections of the project, the following structure is recommended:

- **Sheet 5 – Grouped Data**
  Includes variables such as gender, age, performance group, finish place, half-marathon time, finish time, pacing ratio, and age band.
- **Sheet 6 – Summary Tables**
  Displays averages and medians for pacing ratios across:
  - **Top vs mid-pack**
  - **Gender within each performance group**
  - **Age bands within each performance group**
- **Sheet 7 – Graphs**
  Contains visualizations:
  - **Line graphs comparing half and finish times for each group**
  - **Optional graphs segmented by gender or age**