

Question 2:

Effect of Age and Gender on Marathon Pacing Efficiency

1. Introduction

This analysis examines how age and gender influence the relationship between half-marathon performance and full-marathon finish time. Using a dataset of **13,152 Boston Marathon runners**, we sought to answer the following research question:

“How does age modify the relationship between half-marathon time and full-marathon finish time, and does this effect differ between men and women?”

To address this question, we combined multiple analytical approaches, including descriptive visualization, distributional comparisons by gender, and a multiple regression model with interaction terms. The goal was to determine whether age or gender significantly alters pacing behavior when accounting for how fast runners complete the first half of the marathon.

2. Data Preparation

Each runner’s record included:

- Half-marathon split time (minutes)
- Full-marathon finish time (minutes)
- Age group
- Gender

To evaluate age and gender interaction effects, we created additional variables:

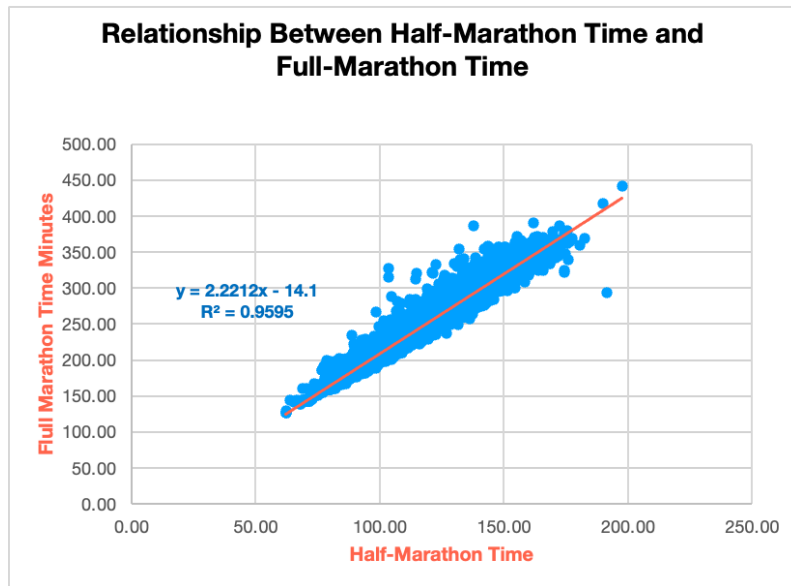
- **gender_dummy** (1 = male, 0 = female)
- **age_mid** (midpoint of the runner’s age group category)
- **half_age** interaction term ($\text{half_time_min} \times \text{age_mid}$)
- **half_gender** interaction term ($\text{half_time_min} \times \text{gender_dummy}$)

The final dataset included one row per runner with all variables needed for regression modeling.

3. Descriptive Visualizations

3.1 Scatterplot: Relationship Between Half and Full Marathon Times

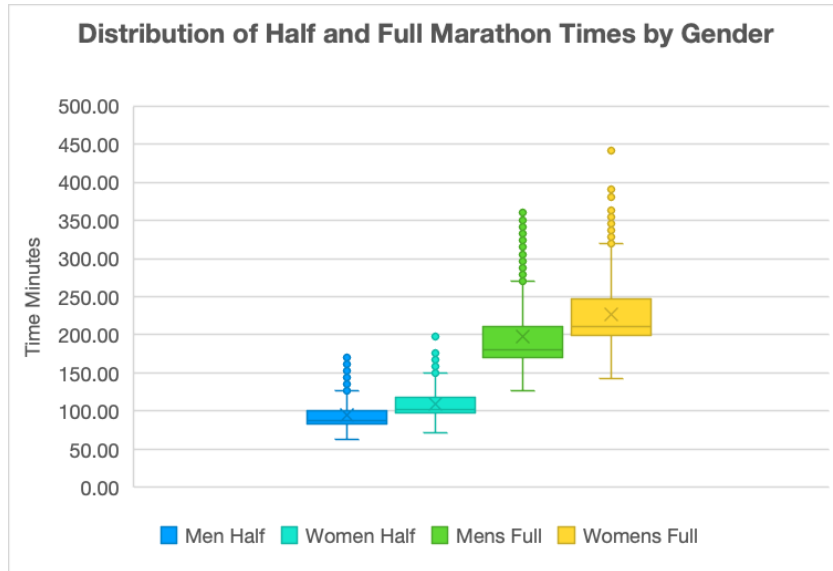
A scatterplot of half-marathon time vs. full-marathon time revealed a **strong linear pattern**, indicating that runners who run a slower half almost always run a slower full marathon. The points formed a tight upward-trending cloud with no major outliers. This suggests that half-marathon time is an excellent predictor of overall marathon performance a conclusion supported statistically in the regression model



This visualization helps establish the central relationship being analyzed: **pacing in the first half is highly connected to performance in the second half.**

3.2 Boxplot: Distribution of Half and Full Times by Gender

A box-and-whisker plot was used to compare men's and women's distributions for half-marathon and full-marathon times. The results showed:



- Women's times were slightly higher on average for both half and full marathon distances.
- However, the distributions for men and women overlapped extensively.
- The spread (variability) of finishing times was similar for both genders.

This supports the later regression finding that **gender does not meaningfully influence marathon performance once half-time and age are taken into account.**

4. Regression Analysis

To formally test whether age and gender modify pacing efficiency, we ran the following multiple regression model:

$$\text{Full Time} = \beta_0 + \beta_1(\text{Half Time}) + \beta_2(\text{Gender}) + \beta_3(\text{Age}) + \beta_4(\text{Half} \times \text{Gender}) + \beta_5(\text{Half} \times \text{Age}) + \epsilon$$
$$\text{Full Time} = \beta_0 + \beta_1(\text{Half Time}) + \beta_2(\text{Gender}) + \beta_3(\text{Age}) + \beta_4(\text{Half} \times \text{Gender}) + \beta_5(\text{Half} \times \text{Age}) + \epsilon$$

4.1 Model Fit

- $R^2 = 0.9603$
- Adjusted $R^2 = 0.9603$

The model explains **96% of the variation** in full-marathon finish times — indicating an exceptionally strong predictive model.

4.2 Key Coefficient Results

Variable	p-value	Interpretation
Half Time	< 0.001	Extremely strong predictor — every extra half-marathon minute adds ~2.24 minutes to full time
Gender Dummy	0.135	Not significant — men and women finish similarly after adjusting for half time and age
Age Mid	0.030	Small negative effect — older runners finish slightly faster when controlling for half time
Half × Age	0.260	Not significant — age does <i>not</i> change pacing relationship
Half × Gender	0.118	Not significant — gender does <i>not</i> change pacing relationship

5. Interpretation of Findings

5.1 The Half-Marathon Split Dominates Performance

The strongest finding is that half-marathon time overwhelmingly predicts full-marathon finish time. Runners who perform better in the first half nearly always maintain that advantage.

5.2 Age Has a Small, Surprising Effect

Even though older runners typically have slower absolute finishing times, the regression shows:

- When controlling for *how fast they ran the first half*,
- **Older runners finish slightly faster than younger runners.**

This suggests older runners pace themselves more evenly and avoid large positive splits.

5.3 Gender Does Not Significantly Impact Pacing

Despite small observed differences in raw times, once half-time and age are controlled for:

- Men and women display *nearly identical* pacing behavior
- Neither gender runs disproportionately slower in the second half
- Gender \times Half Time was not significant

This means **gender does not modify the relationship between half and full marathon times.**

5.4 No Interaction Effects

Both interaction terms (Half \times Age and Half \times Gender) were not significant, meaning:

- Age does not change how strongly half time predicts full time
- Gender does not change how strongly half time predicts full time
- Pacing efficiency is consistent across demographic groups

6. Conclusion

This study analyzed over 13,000 marathon runners to understand how age and gender influence pacing efficiency from the half-marathon split to the full-marathon finish.

Key conclusions:

1. **Half-marathon time is the single strongest predictor of full-marathon finish time.**
2. **Age has a small independent effect**, with older runners pacing slightly more efficiently.
3. **Gender does not significantly influence pacing behavior**, aligning with the overlapping boxplot distributions.

4. **No meaningful interaction effects** were found — neither age nor gender changes the fundamental pacing relationship between half and full marathon times.

Overall, this analysis shows that **pacing strategy, not demographic characteristics, is the primary driver of marathon performance once initial half-marathon speed is considered.**