What Influences Marathon Performance? A Statistical Analysis of Boston Marathon Data

Jada Norris, Olivia Overholt, Carson Cochrane, AnnCathrine Beissel

Predicting Marathon Finish Times Using Regression

Using a sample of 200 runners, what is the estimated finish time for runners whose half marathon split in 62 minutes?

Why 62 minutes?

- The fastest half splits in the first 200 runners are around **1 hour 2 minutes (≈62 minutes)**.
- The slowest half splits in that group are closer to **1 hour 16 minutes (≈76 minutes)**.
- Most elite runners in the first 200 finishers hit the halfway mark between **1:02 and 1:14**.

Method and Calculation

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.937616546
R Square	0.879124787
Adjusted R Square	0.878514306
Standard Error	2.390396454
Observations	200

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	8228.455689	8228.455689	1440.052956	8.5531E-93
Residual	198	1131.371051	5.713995208		
Total	199	9359.82674			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	27.54373011	3.143090702	8.76326289	8.56016E-16	21.34550031	33.74195991	21.34550031	33.74195991
half_time_minutes	1.659753158	0.043737532	37.94802967	8.5531E-93	1.573501981	1.746004336	1.573501981	1.746004336

Intercept	27
Slope	2
Minutes	62
	151

Why this matters?

REAL TIME PACING DECISIONS.

GOAL SETTING STRATEGY.

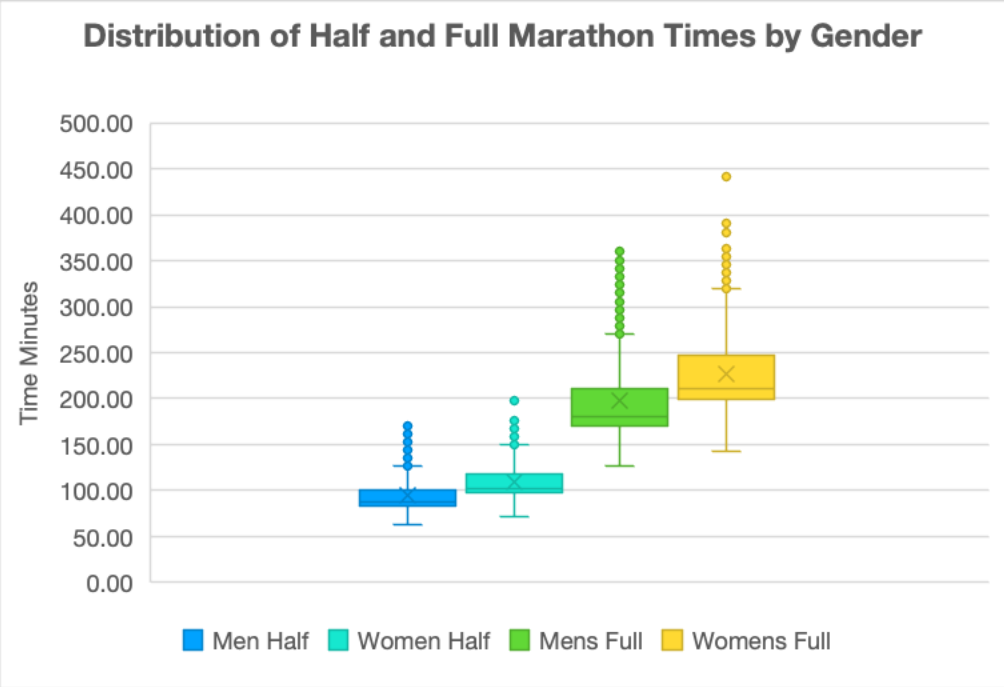
PERFORMANCE ANALYSIS.

How Age and Gender Influence Marathon Pacing Efficiency

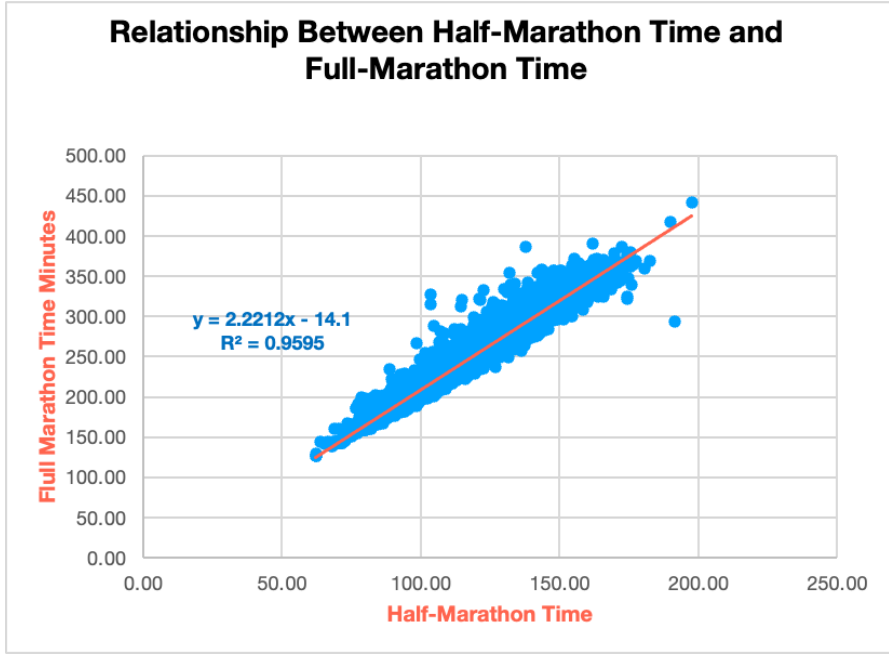
- How does age modify the relationship between half-marathon time and full-marathon finish time, and does this effect differ between men and women?

- Data: 13,152 Boston Marathon runners
- Variables analyzed:
  - Half time (min)
  - Full time (min)
  - Age group (18-34, 45-49, 50-54)
  - Gender
- Method: Boxplots + Scatterplot + Multiple Regression

Visualizing Half & Full Marathon Times



- Men and women pacing distributions overlap, meaning me and women pace themselves in a very similar way.
- Women are slightly higher on average for both – due to difference in starting pace, not gender itself
- No major gender differences in pacing



- Tight linear trend between half and full marathon times
- Runners who run the first half faster tend to finish the entire marathon faster as well.
- Orange line shows Best-fit Linear relationship, equation shows us every 1 minute added to the half marathon time, full marathon time increases by about 2.22 minutes

Regression Findings & What They Tell Us

Regression Statistics	
Multiple R	0.97995419
R Square	**0.96031021**
Adjusted R Square	0.96029511
Standard Error	8.91963891
Observations	13152

- **Half-marathon time: p < .001 (strongest predictor)**
- **Gender: NOT significant (p = .13)**
- **Age: small but significant effect (p = .03)**
- **Gender × Half Time: not significant**
- **Age × Half Time: not significant**

Key Take Away:

Pacing strategy — not age or gender — drives marathon performance.

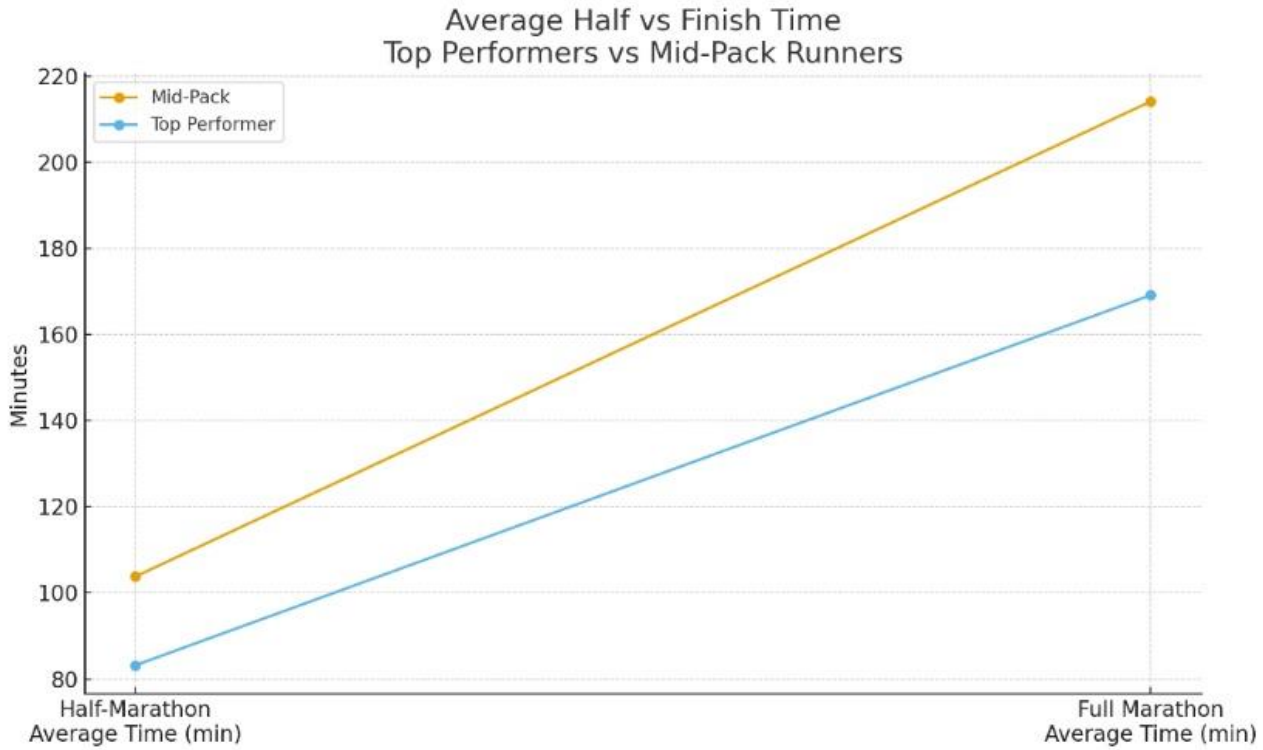
ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	5	25305846.5	5061169.29	63614.5292	0
Residual	13146	1045895.21	79.5599583		
Total	13151	26351741.7			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-15.325006	1.78549322	-8.5830658	1.0267E-17	-18.82483	-11.825181	-18.82483	-11.825181
Gender Dummy	1.48051121	0.9895091	1.49620778	**0.13462352**	-0.4590696	3.42009198	-0.4590696	3.42009198
age_mid	-0.1018859	0.0469689	-2.1692207	0.03008379	-0.1939517	-0.0098201	-0.1939517	-0.0098201
half_age	0.00047799	0.00042395	1.12746876	**0.25956492**	-0.000353	0.001309	-0.000353	0.001309
half_gender	0.01456961	0.00931125	1.56473207	**0.11766982**	-0.0036818	0.03282101	-0.0036818	0.03282101
Half Time Min	**2.2383109**	0.01642657	**136.261617**	**0.000000000**	2.20611245	2.27050935	2.20611245	2.27050935

Performance-Based Pacing Strategy

- Pacing ratio = Full Marathon Time ÷ Half-Marathon Time
- Top performers slow less → more efficient pacing
- Mid-pack pace drop shown by steeper line



Example Pacing Differences: Under-30 Runners

- Both are young, but performance differences still show stronger pacing in top group
- Reinforces pacing efficiency isn't just age-related

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

Summary of Pacing Ratio Analysis

- Top performers: pacing ratio ≈ 2.0 → consistent pacing
- Mid-pack: pacing ratio > 2.0 → more slowdown
- Gender differences minimal
- Slight fatigue increase in older runners
- **Performance level is the strongest pacing predictor**

Conclusion

- Pacing strategy is the strongest predictor of marathon performance, with a tight linear relationship showing that faster half-marathon splits consistently lead to faster overall finish times.
- Age and gender have minimal impact on performance once pacing is accounted for, reinforcing that strategic pacing—not demographics—is what truly drives results.
- Top performers maintain even pacing with a ratio near 2.0, while mid-pack runners slow down more in the second half, highlighting pacing efficiency as the key difference between performance levels.