

# Process Book

Boston Marathon 2015-2017



By Team Sacoche

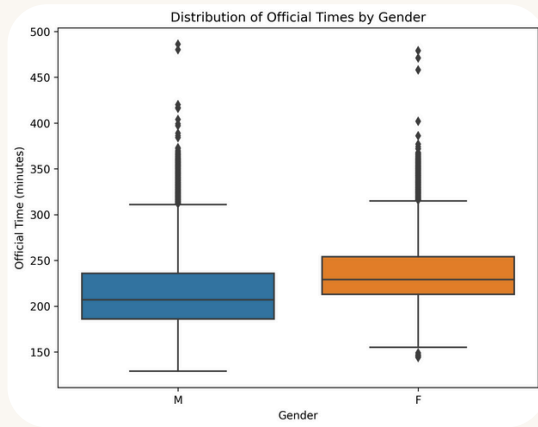
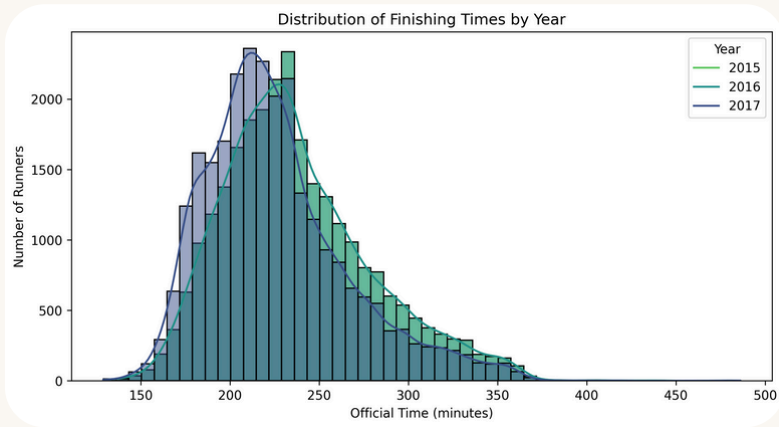
# Project Goal

The goal of this project is to analyze pacing patterns and the diversity among Boston Marathon runners over the 2015–2017 races. We focused on understanding how runners of different backgrounds perform and how their pace changes throughout the marathon. The Boston Marathon is an ideal case study for this, as it attracts thousands of runners each year with varied ages, genders, and skill levels

## Dataset

For our analysis, we used official Boston Marathon results for the years 2015, 2016, and 2017 (sourced from Kaggle). These data were provided in CSV files, each containing every finisher’s details: finish time, average pace, split times, age, gender, and more. In total, the dataset covers tens of thousands of runners, providing a rich ground for exploration. We augmented the raw data by writing Python scripts (Jupyter Notebooks) to compute additional metrics and generate JSON files (pace profiles for race segments and general statistics) for use in our visualizations. We also incorporated a .gpx route file of the Boston Marathon course. This file supplies the exact course path and elevation profile, enabling us to map runner progress along the route in our visuals.

## Exploratory Data Analysis



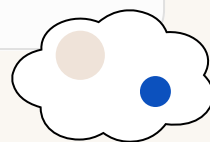
Preliminary analysis revealed clear differences in finishing times across years and between genders. These disparities led us to implement filters for year, gender, and runner level across the site. They also motivated the creation of the Pace Calculator page, where users can explore personalized pacing strategies depending on their gender and goal time, reflecting the variability we observed during our initial exploration.

## Website Colour Scheme

For our colour scheme, we first wanted to implement the one from the Boston Marathon logo which is dark blue and yellow. After few tests, we decided to replace the yellow that was too imposing with cream. In fact, with colorful graphs it was better to have neutral background. We chose blue for headers, and important information. It contrasts well with the body background, and draws attention on it.

Then, simple black is used for text to enhance readability.

For the graphs, and homepage we chose a colour code that represents well the runner level : red, orange, blue, green. From Elite to beginner.



## Website Layout

The site's layout and navigation were designed for clarity and ease of use. We organized the content into three main pages: Home, Stats, and Pace Strategy. A simple menu allows users to switch between these sections. The Home page provides an overview and an eye-catching introduction to the data. The Stats page is a deep-dive dashboard with multiple charts and filters for exploration. The Pace Strategy page offers an interactive tool for users to input different pace in order to refine their strategy. By separating the project into these sections, each page can focus on a specific aspect (general story, detailed stats, personal exploration) while maintaining a cohesive overall experience.

# Storytelling

We adopted a storytelling approach to guide users through the data, centered around three key questions:

1. Who are the runners? – We introduce the participants' demographics and backgrounds. Visualizations show the composition of runners by gender, age group, and other attributes, giving a sense of who takes part in the Boston Marathon.
2. What are their paces? – We present the distribution of finishing times and paces. Charts like histograms and density plots reveal how finish times spread out from the fastest elites to the back-of-the-pack runners. Users can see the typical marathon times and the range of performances.
3. How should pace evolve by type of runner? – We compare pacing patterns across different categories of runners. We provide indications on speed variation depending on elevation. By categorizing runners (by experience level, gender, age, etc.), we highlight variations in pacing strategies. For example : Beginners tend to accelerate at the end whereas Elite runners are decelerating.

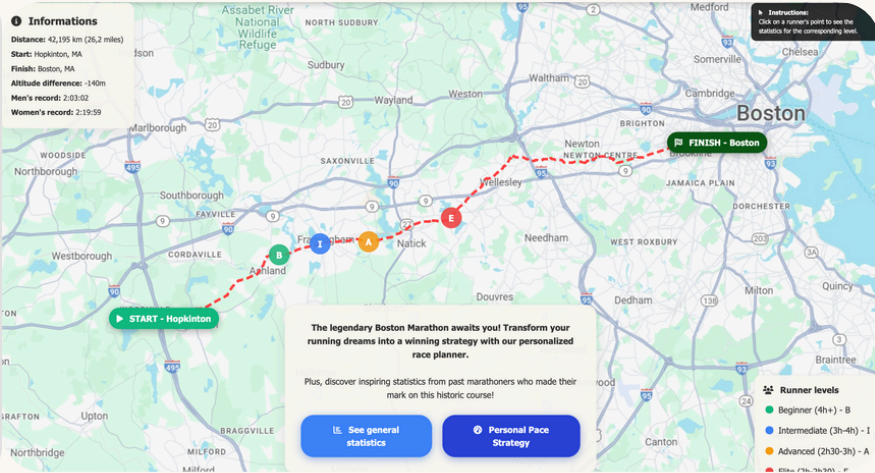
Each of these questions is answered through targeted visualizations and text on our site, creating a narrative flow. Crucially, interactivity is built into our visuals to engage the user in exploring these questions. All charts and graphics are dynamic: users can filter data by year (2015 - rainy/2016- sunny/2017- cloudy) or by subsets of runners (e.g., view only female runners, or a specific age range) to immediately see updated results. Many elements feature animation or incremental reveal of data; for instance, the Home page includes an animation of runners moving along the course (bringing the race to life), and the Stats page charts have hover tooltips and clickable legends to toggle data series. This interactive storytelling means that instead of passively reading facts, the user can actively investigate the marathon data, making discoveries at their own pace.

## Visualisations

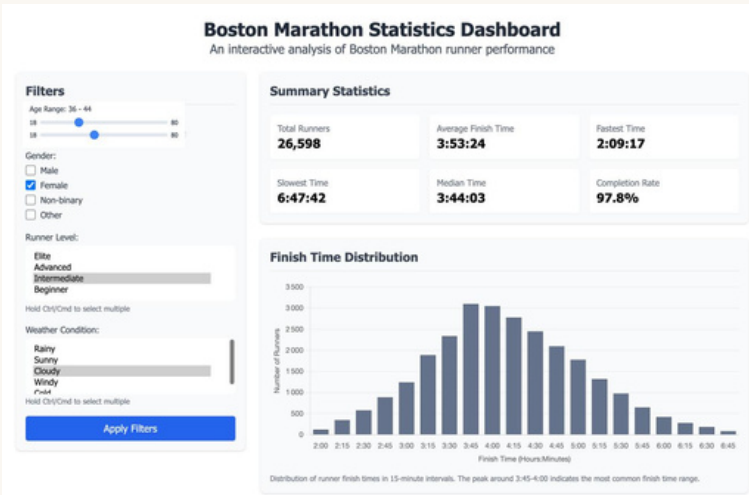
For our JavaScript implementation, we chose to use Chart.js instead of D3.js because our team had prior experience with Chart.js.

Home page

The landing page of our site features an animated visualization of the Boston Marathon route. Using official GPX data, we recreated the 42.2 km journey from Hopkinton to Boston and animated runner icons along the path. Each icon moves at a speed proportional to a typical finishing time (e.g., elite: ~2:10, intermediate: ~3:30, beginner: ~4:30), allowing users to visually compare pacing differences in real time.



An additional interactive feature allows users to click on a runner icon, which dynamically displays relevant statistics and visualizations associated with that performance tier (e.g., distribution of finishing times, pace profiles). At the bottom of the page, two buttons allow users to navigate directly to the Stats Dashboard and the Personal Pace Strategy tools, ensuring intuitive access to the rest of the site. This makes the homepage both an introduction and a point of entry for deeper exploration across the rest of the dashboard.



This dashboard was our prototype for Milestone 2 prototype. At the time, it served as the homepage, but we felt it lacked visual impact and engagement. For the final version, we created a new homepage focused on storytelling and animation to better draw users into the experience.



# Personal Pace Strategy

This page allows users to generate a personalized pacing plan by selecting their gender and a target finishing time. Upon clicking the submit button, the tool returns a suggested segment-by-segment strategy based on similar runners from previous Boston Marathons. Behind the scenes, a Jupyter notebook processes all results from the 2015–2017 editions. When a user, for instance, selects “female” and “3h25,” the tool filters all female finishers across the three years with finish times between 3h22min30 and 3h27min30.

Get personalized pacing strategy based on real Boston Marathon data

Race Configuration

Gender

♂ Male

♀ Female

Target Time (3h10 - 8h00)

3

h

15

min

Minutes in 5-minute increments (0, 5, 10, 15, etc.)

Calculate My Paces

Intermediary Times

Cumulative times at each milestone

5 km

22:35

4:31 min/km

10 km

0:45:10

4:31 min/km

15 km

1:07:55

4:33 min/km

20 km

1:31:00

4:37 min/km

25 km

1:54:10

4:38 min/km

30 km

2:18:00

4:46 min/km

35 km

2:42:25

4:53 min/km

40 km

3:06:35

4:50 min/km

42.2 km

3:17:09

4:49 min/km

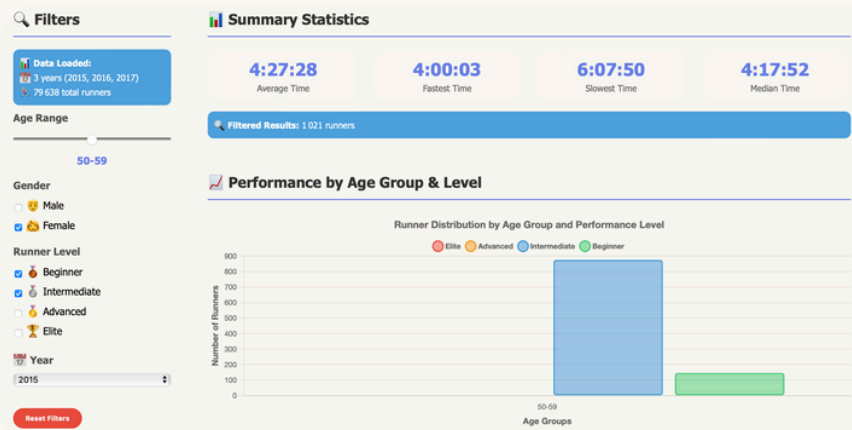
Pro Tip: These paces are based on successful Boston Marathon finishers. Start conservatively and adjust based on how you feel during the race!

From this subgroup, segment-by-segment pace profiles are extracted and averaged, after excluding the 5% of runners with the highest pace variability (to remove unstable or erratic pacing patterns). The result is a clean and realistic split suggestion that mirrors the behavior of stable runners with similar profiles. This data-driven personalization helps users understand what typical pacing might look like for their own goals, while grounding the tool in actual marathon performance patterns.

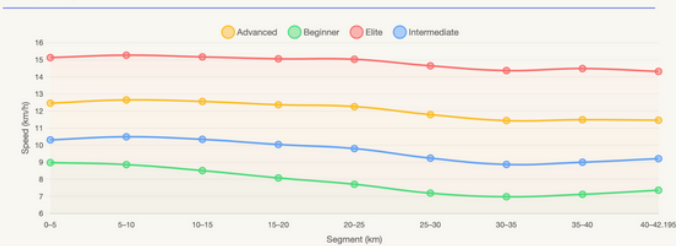
The second graph plots the calculated pace against the elevation profile of the Boston Marathon course. This comparison highlights how uphill sections (such as Heartbreak Hill) may correlate with pace drops. The elevation data was derived directly from the official GPX course file, allowing precise matching of terrain and pacing behavior. Together, these visualizations give users a more nuanced understanding of how real-world course conditions affect pacing across different runner profiles.



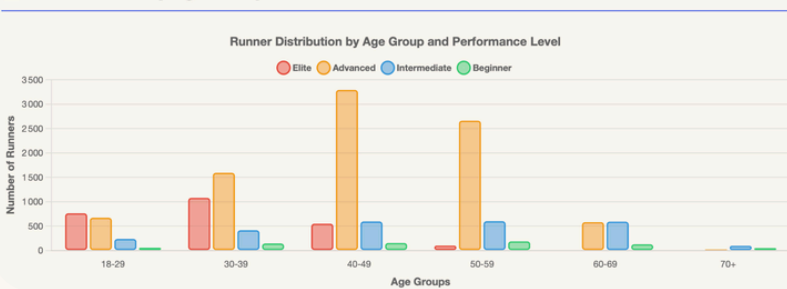
The Stats page is the most data-rich section of the site. It features several interactive charts that allow users to explore marathon performance across age, gender, year, and runner level.



Split Speed Evolution



Performance by Age Group & Level



The Stats page is the most data-rich section of the site. It features several interactive charts that allow users to explore marathon performance across age, gender, year, and runner level. One key visualization is the bar chart of runner distribution by age group and level, which reveals where most runners fall. Just below, the Split Speed Evolution graph shows how average speed varies across the course depending on runner level. While elevation is not explicitly shown, this chart highlights general fatigue patterns, for example, beginner runners tend to slow down much more dramatically after the 25 km mark, whereas elite runners maintain a more stable profile. These segment-by-segment speed comparisons are based on real race data and emphasize how pacing differs structurally between performance levels. Additionally, the ability to toggle between different years makes inter-annual comparisons possible, especially meaningful since the weather conditions varied greatly across editions, which likely influenced overall times and pacing behaviors. Together, these charts offer users a comprehensive view of how marathon pacing interacts with demographic and contextual variables.

# Challenges

## 1. Real Map Integration on the Homepage :

Our initial vision for the homepage was to incorporate a real map of the Boston Marathon route as the background. This would allow users to zoom, pan, and engage with the geographic context more dynamically. However, during implementation, we encountered significant constraints with map APIs. While Google Maps and Mapbox free usage limits were too restrictive, a clean and interactive solution would have required a paid plan.

To avoid compromising user experience with low-resolution alternatives, we opted for a custom SVG-based path over a stylized static background. Although it meant sacrificing some interactivity, this decision gave us full design control and better performance.

## 2. Filter Interface Complexity in the Dashboard

We wanted to design cleaner and more modern filters, like the ones on fashion websites. Instead of using regular checkboxes for gender and runner level, we tried using clickable text. When a filter was selected, the word would go bold to show it was active.

But while setting this up, we faced layout problems and bugs with how it connected to our dashboard and Chart.js. Because of time limits, we decided to stick with simple checkboxes to keep everything working reliably.

# Peer Assessment

### Clara

- Led the data analysis and preprocessing
- Wrote the Jupyter notebooks for JSON generation
- Developed the Pace Calculator page and implemented the logic to compute personalized pacing strategies based on gender and goal time

### Jules

- Designed and built the Stats Dashboard with interactive filters and dynamic charts
- Created the global CSS style guide to ensure design consistency
- Unified the visual layout across all pages

### Amaury

- Developed the interactive Home Page, including the animated runner simulation
- Integrated the official GPX route data and synced animation with pacing logic