Project 3 - London Marathon Predictions

William Sorg.74

2025-04-27

Question Formulation

The London Marathon is one the most prestigious races in the world. As one of the Abbott World Marathon Majors, the London Marathon attracts over 50,000 runners to compete each year, along with approximately 750,000 spectators. Among the runners are some of the greatest elite distance runners in the world. Eliud Kipchoge, Mo Farah, Sifan Hassan, and the late Kelvin Kiptum have all raced the streets of London.

Despite the marathons's significance, it's difficult to find statistically driven predictions for the race. Nearly every other sport has some sort of win probability model, but there is a void for runnning. In a sport that can witness extreme unpredictability, I believe it would be interesting and informative to build a model that predicts the winner and top finishers of the London Marathon. Given that the 2025 London Marathon will begin at 4:00 a.m. EST on April 27th, this project will focus on predicting the results of the 2025 race. I will then briefly evaluate the accuracy of the predictions after the race concludes.

Data Selection

This project will use the 2018-2024 London Marathons to train and test different models with the goal of developing predictions for the 2025 London Marathon on April 27th. The London Marathon Elite runner results were scraped from each year's London Marathon results [website] (https://results.tcslondonmarathon.com/2024). Then, for each athlete, every past race result was scraped off of the World Athletics [website] (https://worldathletics.org/). From this data, statistics like PR and number of marathons run were calculated. All code for scraping can be found on my [GitHub] (https://github.com/williamsorg4/London-Marathon-Model). (Note, scraping the data was the biggest time requirement of this project.)

Exploratory Analysis

Analysis

Recommendations