# Berlin Marathon Project

# Research Questions –

## 1. Evolution of performance

• How have winning times evolved in the Berlin Marathon from 1974 to 2024?

• Are world records being broken more frequently in Berlin than in other marathons?

## 2. Gender gap

• What is the time difference between male and female winners?

• Has the gender gap decreased over the years?

## 3. Country dominance

• Which countries have produced the most winners in the Berlin Marathon?

• Has the dominance of Kenya and Ethiopia increased in recent decades?

## 4. Female participation

• How has the percentage of female runners changed since 1974?

• Is the finishing time distribution of female runners approaching that of male runners?

## Hypotheses

1. Winning times in the Berlin Marathon have significantly improved over the years.

2. The performance gap between men and women has decreased over time.

3. A few countries (mainly Kenya and Ethiopia) dominate the winners.

4. Female participation has increased since 1974.

## Methodology

1. Data Collection – gather marathon results from 1974–2019 (Kaggle), extend with winners 2020–2024, and include country data from 2019 onwards.

2. Data Cleaning – standardize column names, convert times to timedelta, handle missing values (e.g., 2020 race cancelled due to COVID).

3. Data Merging – combine datasets to create a continuous timeline from 1974 to 2024.

4. Exploratory Data Analysis – compute descriptive statistics, analyze by gender, country, and year.

5. Visualization – use line charts for performance trends, bar charts for country dominance, histograms for age and participation.

6. Interpretation – test the hypotheses and summarize findings.

## Project Structure

1. Introduction – motivation and objectives of the project.

2. Dataset Description – sources, size, key variables.

3. Data Cleaning and Preparation – steps taken to format and merge datasets.

4. Analysis – answering research questions with statistics and visualizations.

5. Conclusions – main findings, limitations, and final remarks.

6. Future Work – possible extensions (e.g., compare Berlin with Boston or NYC).