

Owen George

PARKRUN PERFORMANCE PREDICTOR





INTRODUCTION

Parkrun offers a free, weekly, 5km event for people of all abilities in parks worldwide.

Many runners want to improve their performance, but it can be hard to set realistic targets.

This project uses historic data and weather conditions to generate data-driven, personalised target parkrun times.

23

Countries

839

UK locations

3.5M

UK Participants

60M

UK Finishes

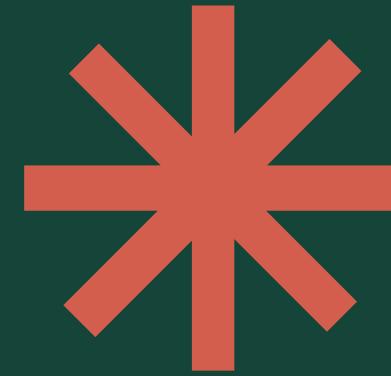
A photograph showing a man and a woman jogging on a bridge. The man is in the foreground, wearing a dark jacket and shorts, captured mid-stride. Behind him, a woman in a dark coat and skirt walks at a slower pace. The bridge has metal railings and is set against a backdrop of modern city buildings, including a prominent glass skyscraper. The lighting suggests it's either early morning or late afternoon.

BUSINESS CASE

This analysis could be used by running clubs, training programmes, and fitness apps, to drive engagement through personalized running goals and performance insights.

While built on parkrun data, the analysis can scale to other running events and fitness programmes.

DATA SOURCING AND CLEANING



Scraped parkrun results

- Scraped the parkrun results web pages for a specific location and saves the results as a DataFrame.

Added weather data

- Weather data was added for each parkrun using the OpenMeteo API.

Processing / Cleaning

- Columns were converted to appropriate data types.
- Calculated additional fields from previous times and dates.
- Filtered the data by appearance count, age, etc.

DATA SUMMARY

After cleaning

The popularity of the Brighton parkrun increased dramatically since it began in 2007. Although participation dropped after a 16 months pause over the pandemic.

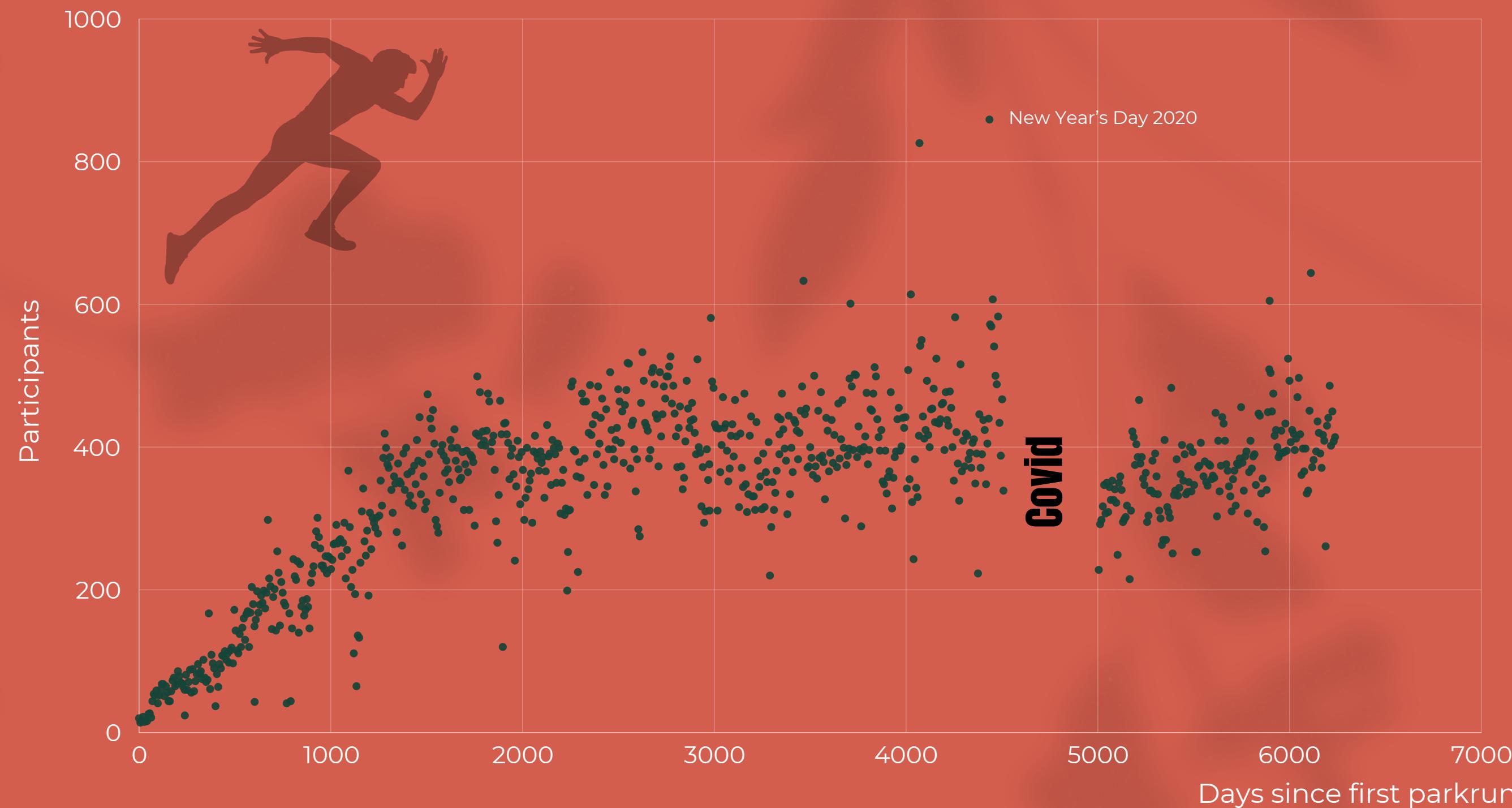
Since it returned it has been increasing again, and is now largely in line with pre-pandemic levels.

826
Events

**12 year
9 month**
Period

158K
Results

7,244
Unique runners



RUNNER DATA



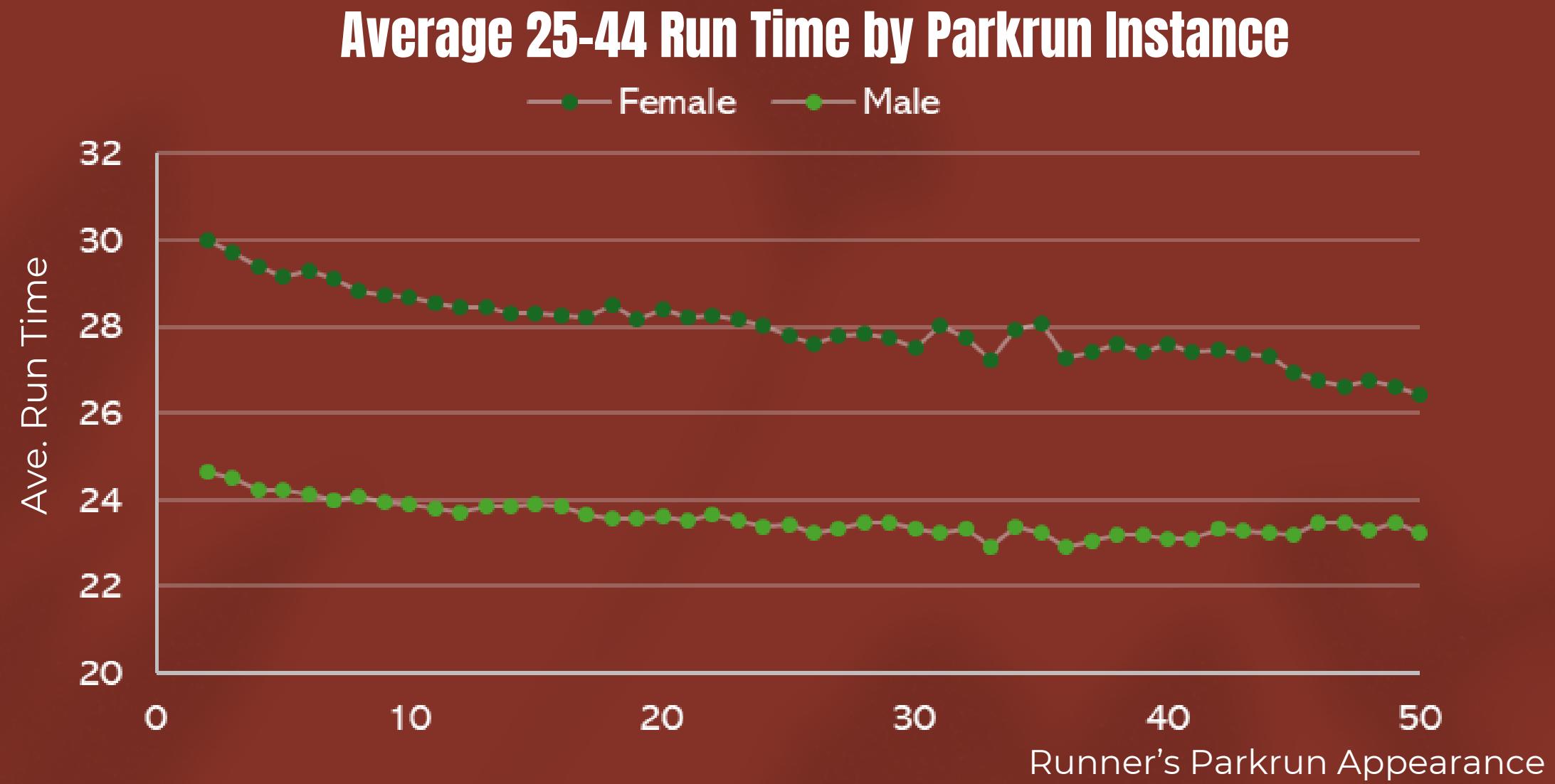
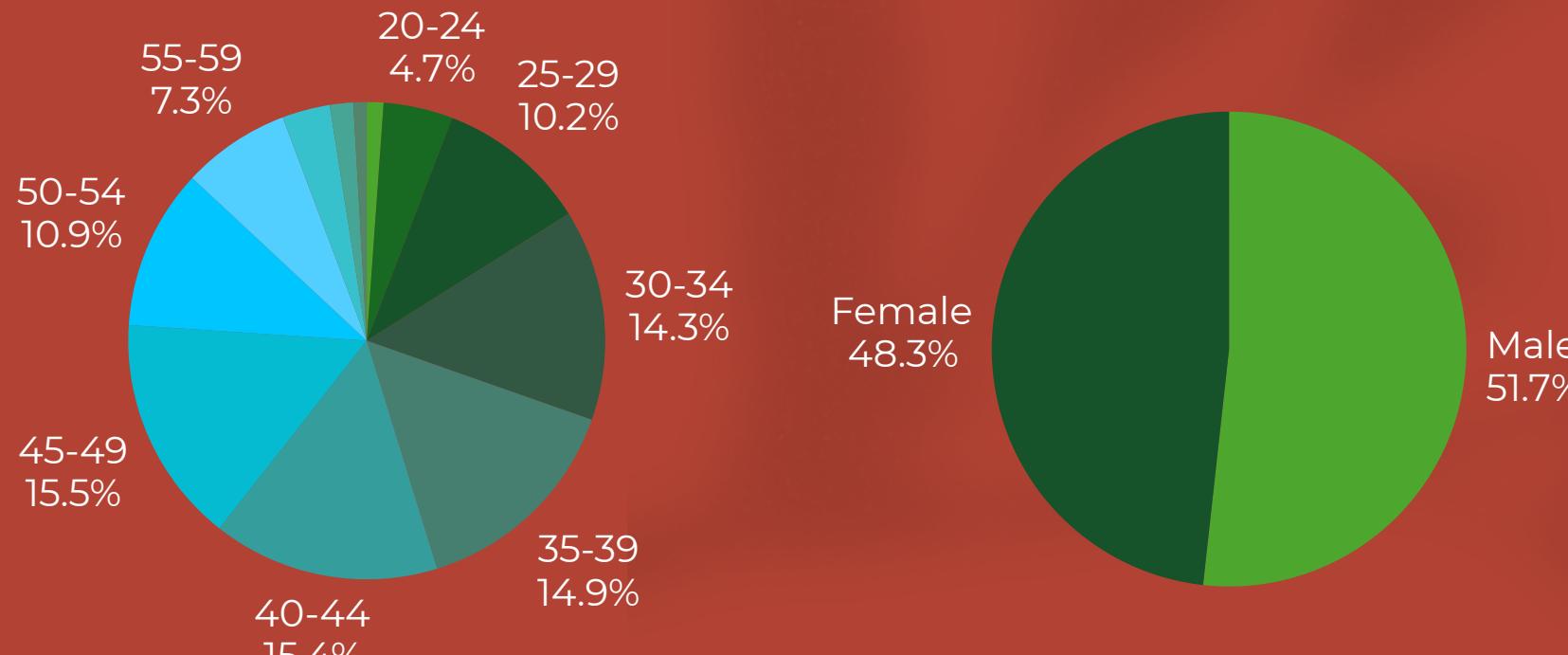
Improvement

We can see that, especially for the first few parkruns, there is a clear improvement in run times

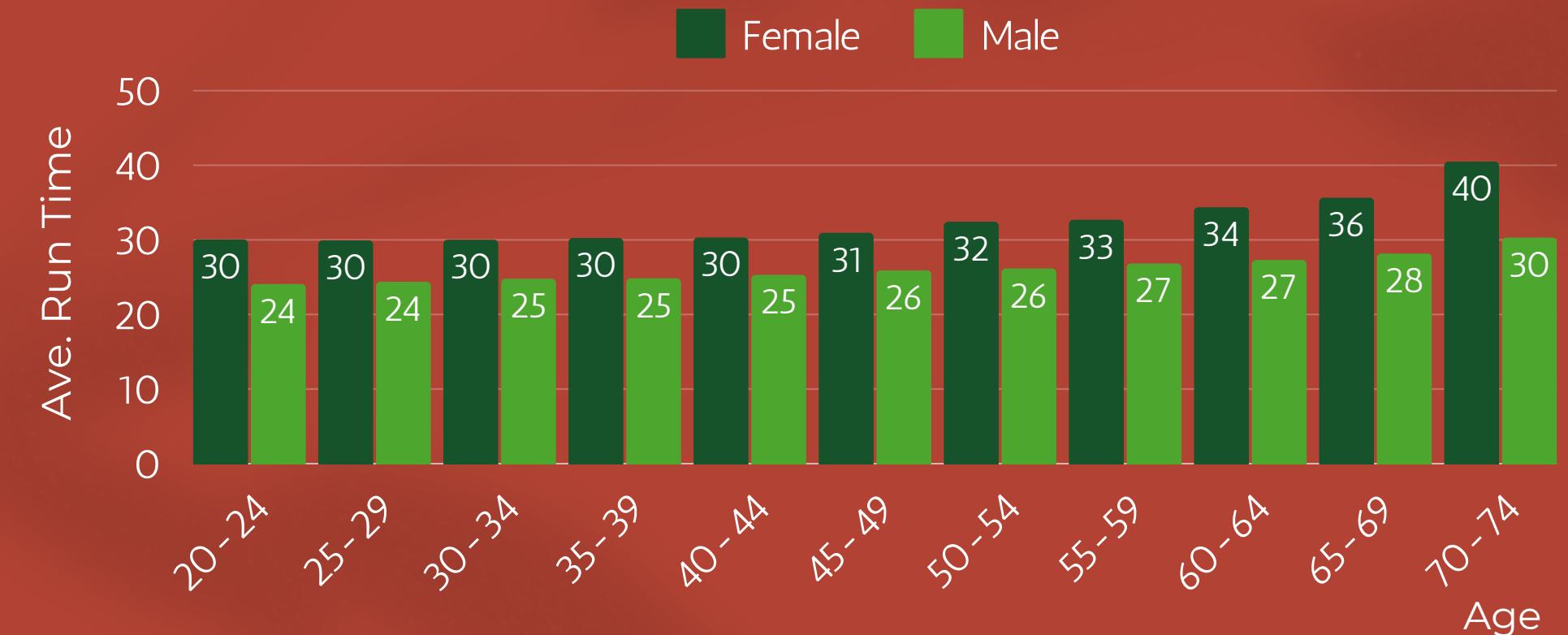
Age / Gender

Male run times tended to be quicker than female

Older runners get slower average times

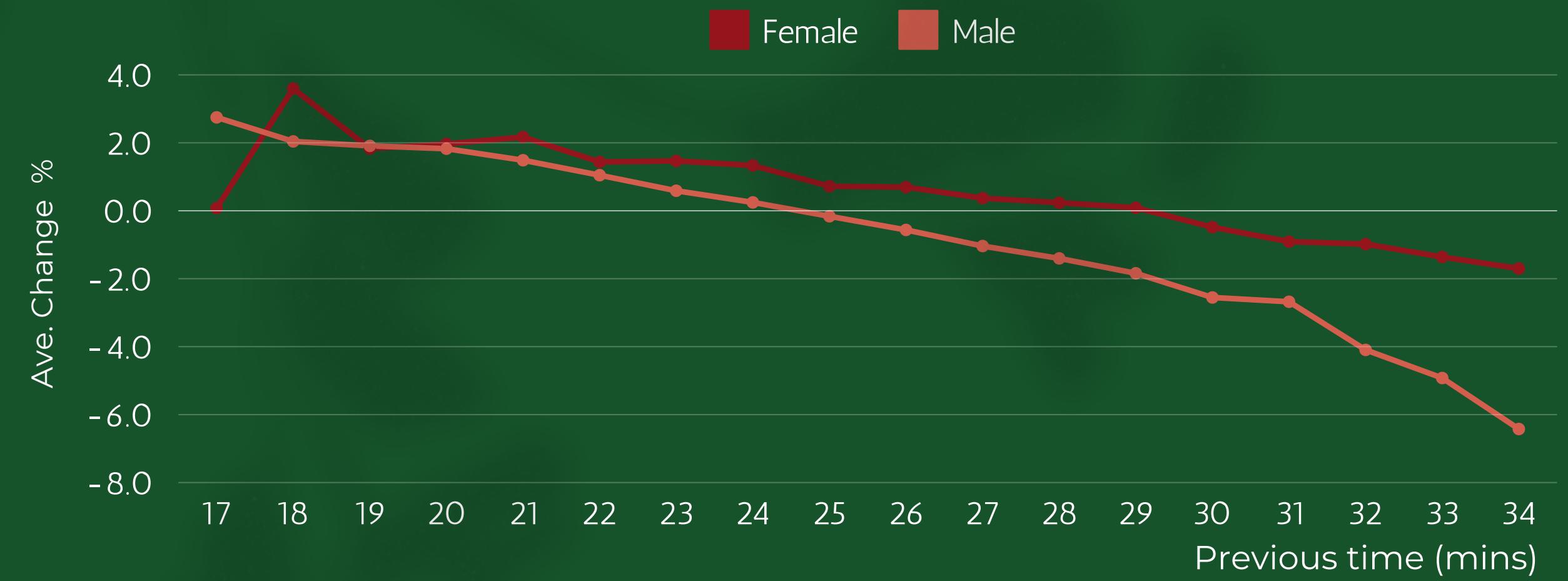
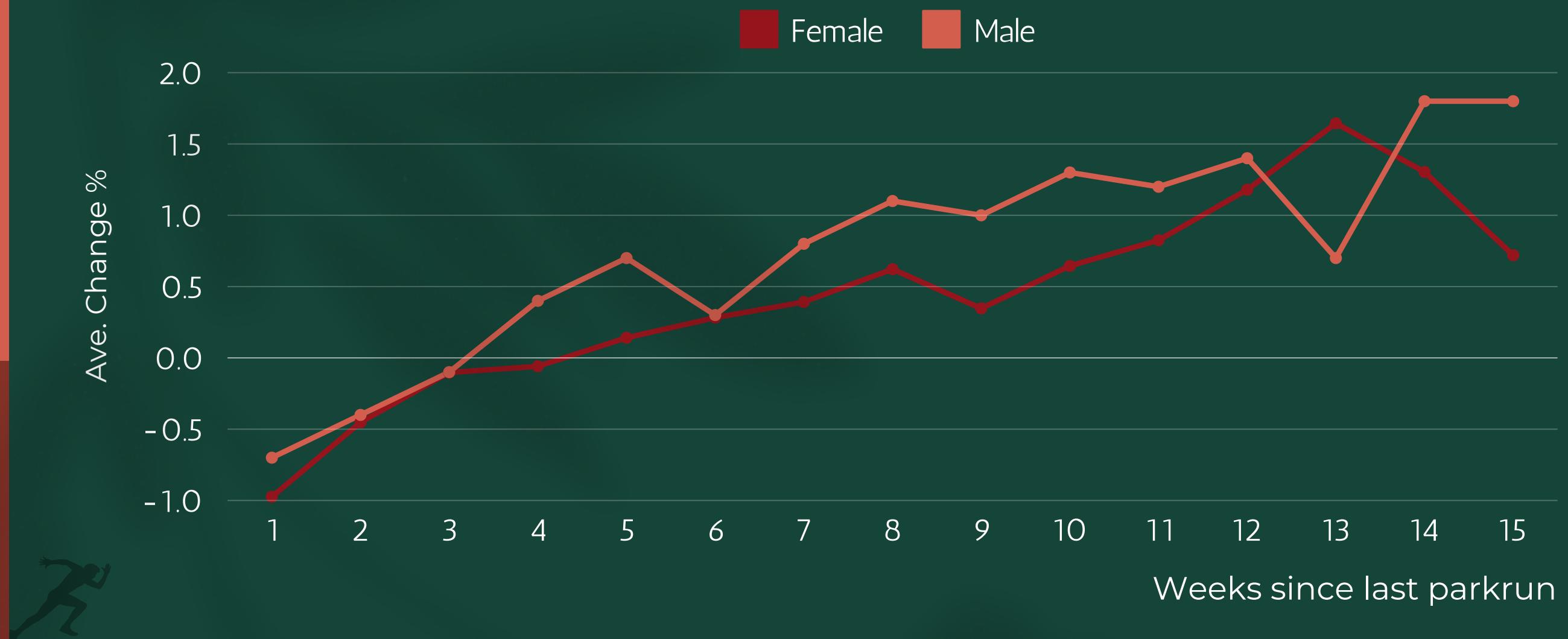


Ave. Run Time by Age/Gender



CHANGE IN RUN TIME

- For change in performance, the most important factors were time since last run and previous run times.
- Shorter gaps between parkruns resulted in greater improvements, and gaps over a month tended to result in slower next run time.
- Slower run times were more likely to be improved upon, and by a larger margin, than faster run times.

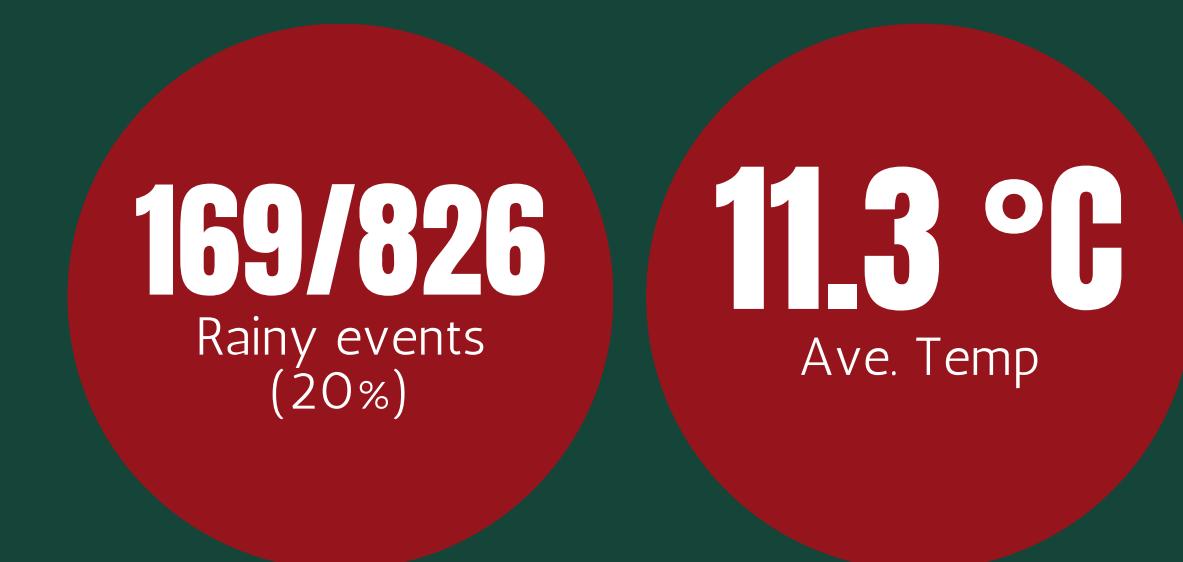
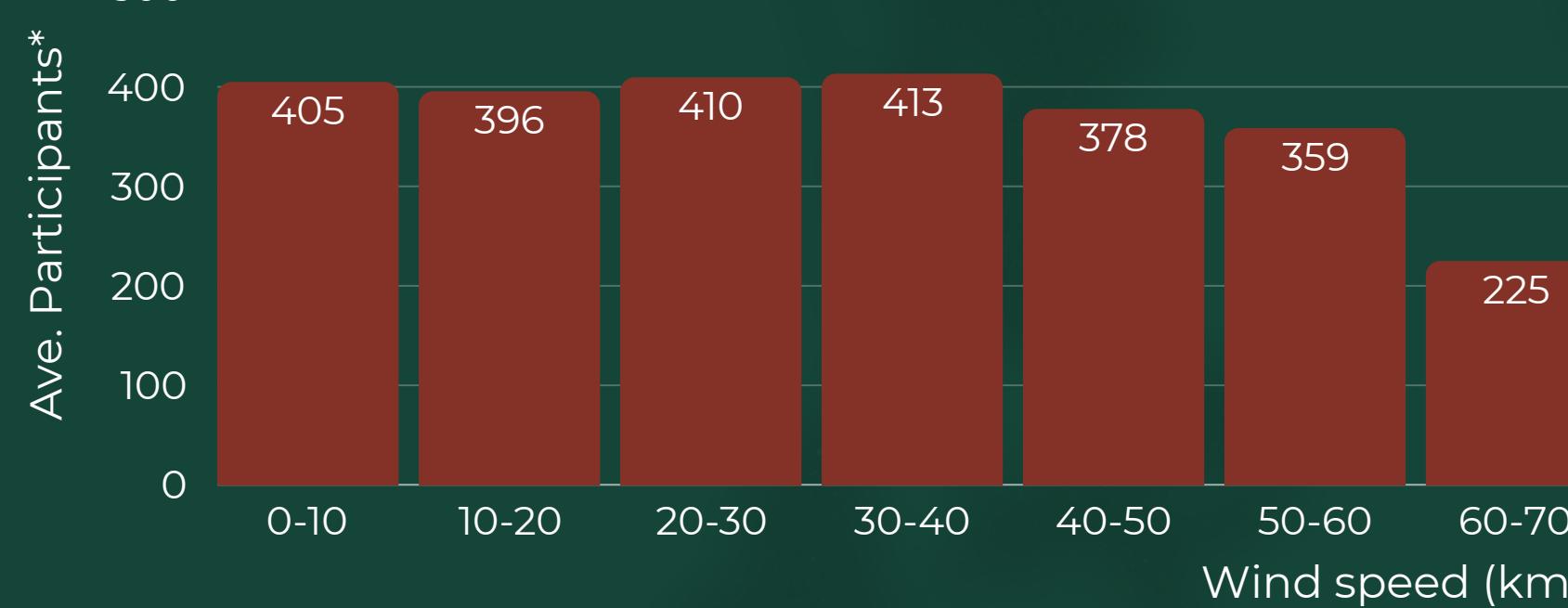
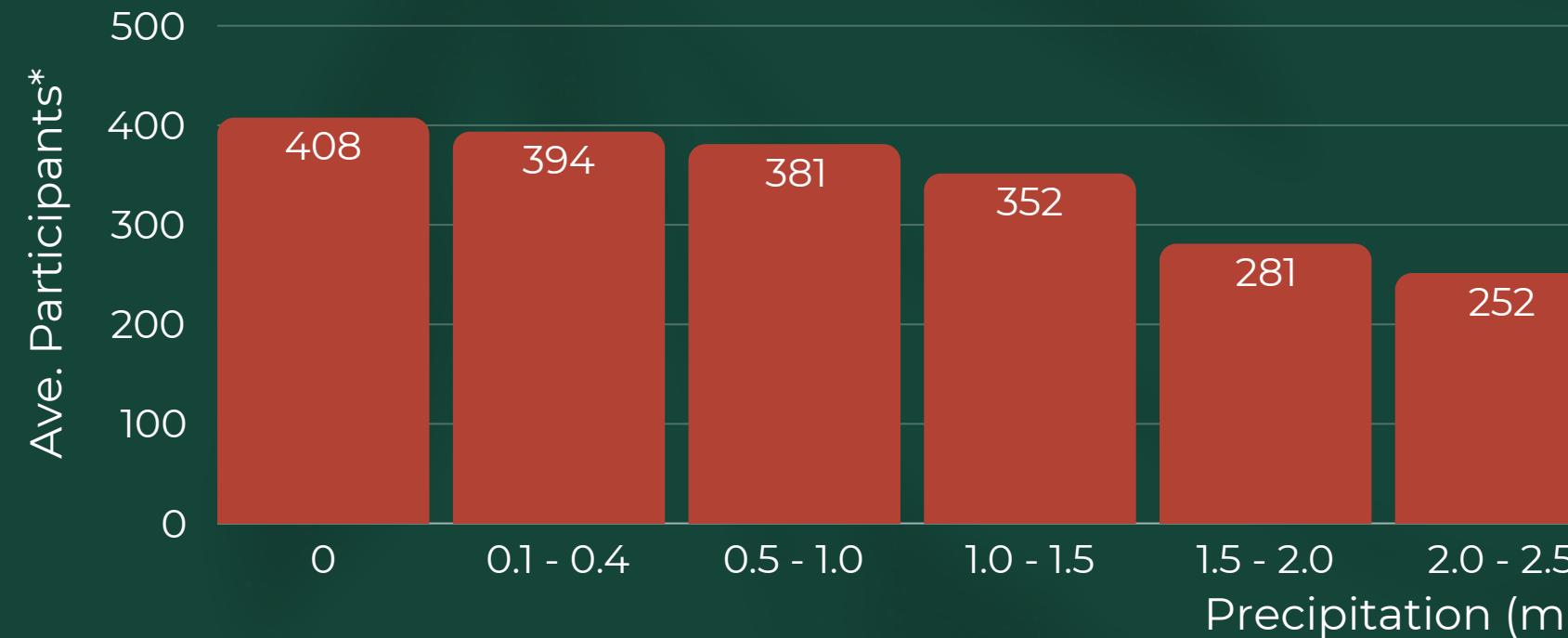
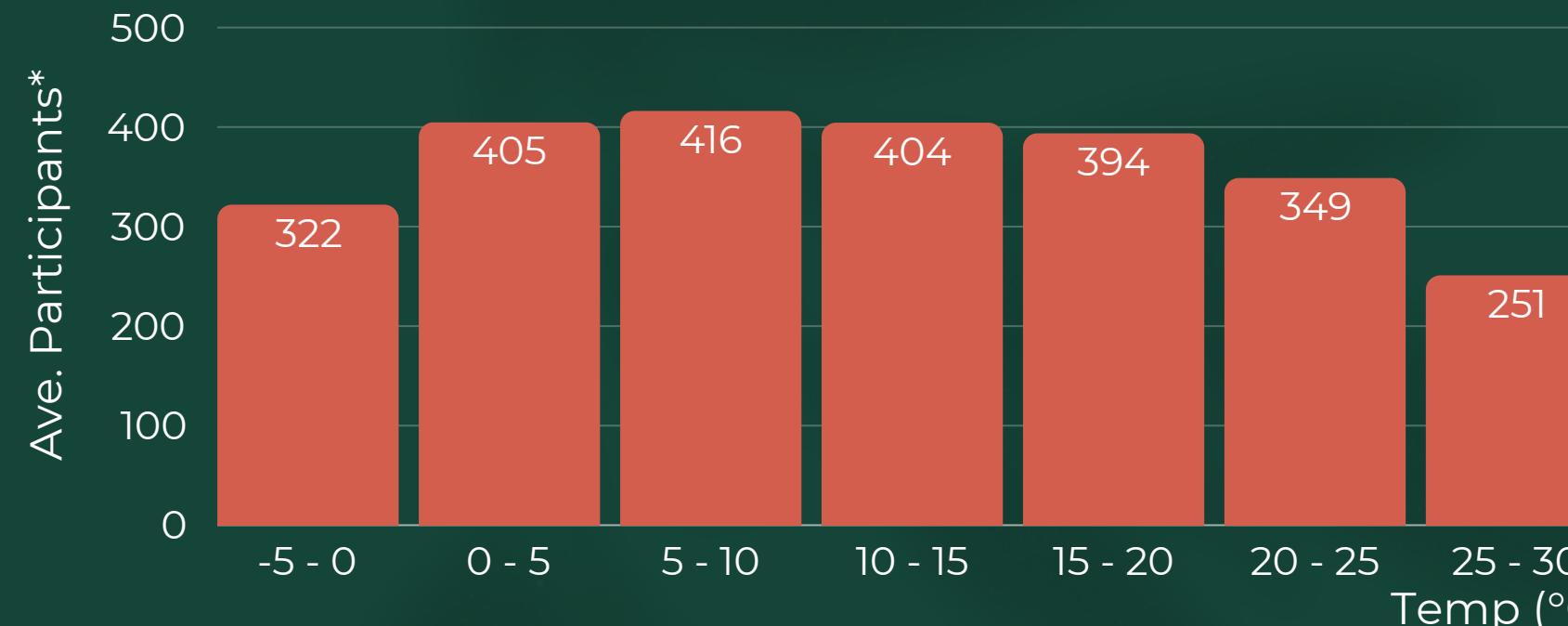




WEATHER CONDITIONS



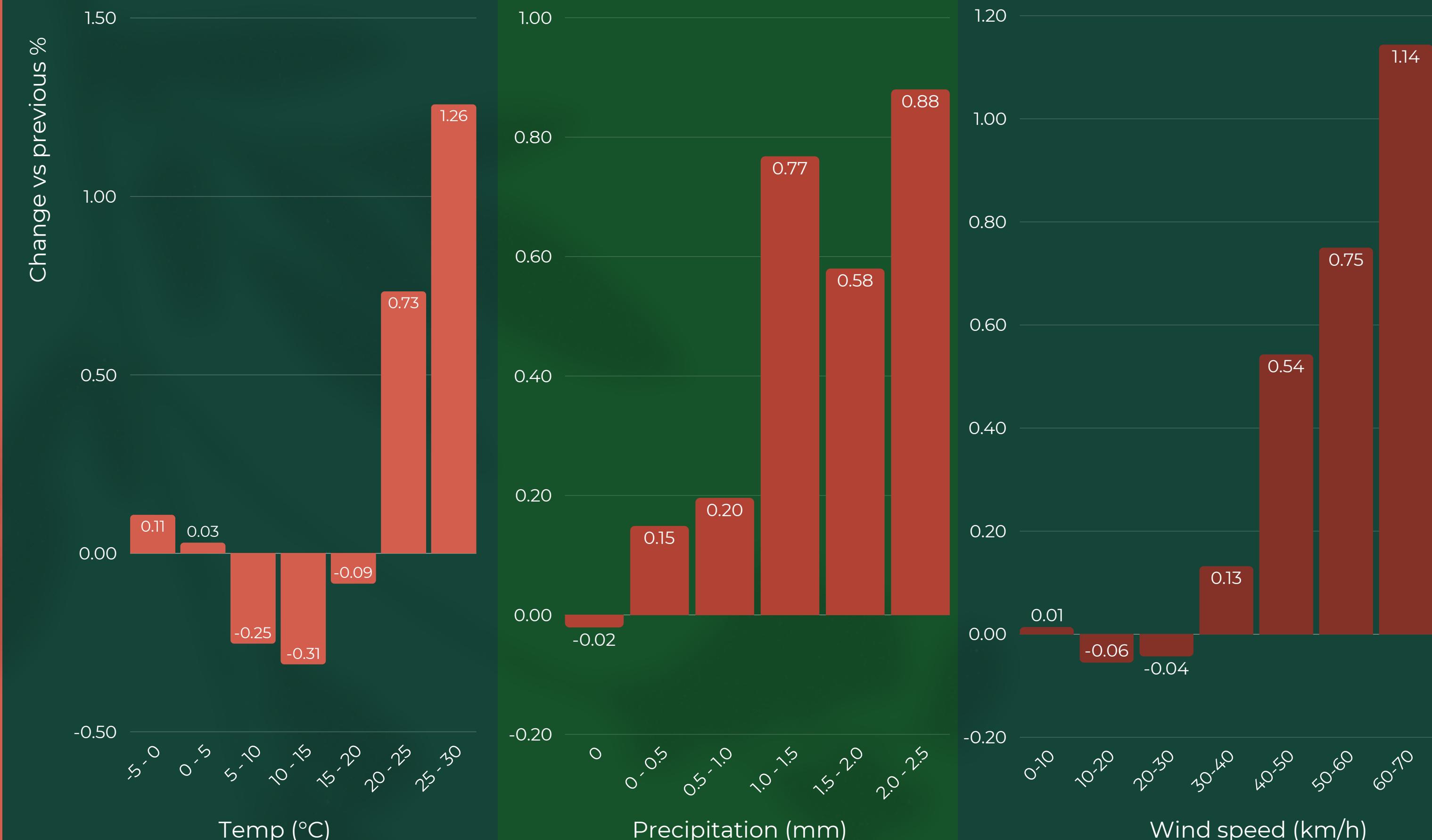
*2014 onwards



Weather conditions strongly impact participation, with lower attendance at more extreme conditions



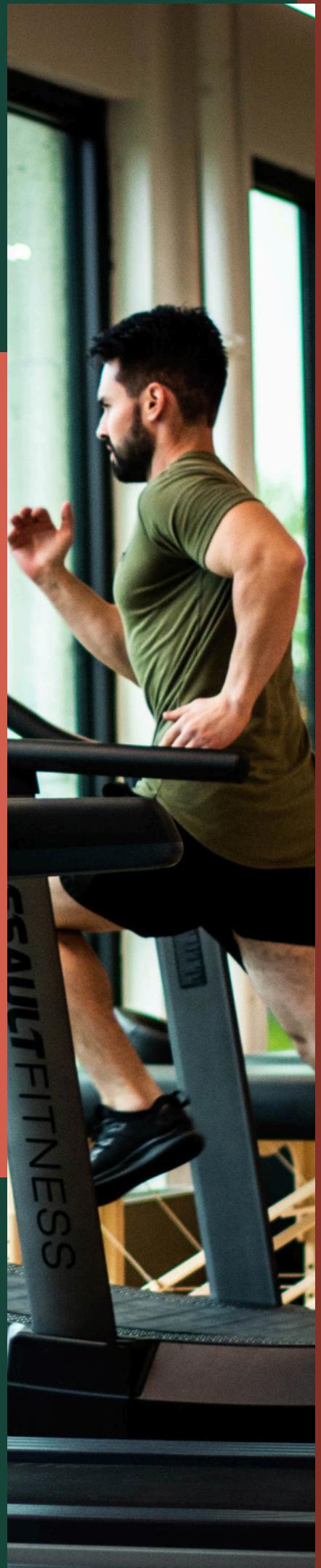
WEATHER CONDITIONS



Whilst weather conditions strongly impacted participation level, the effect on run time was more subtle, usually below 1% change, with more extreme conditions likely to result in worse times.



MODEL BUILDING



Data processing

- Outliers removed from “run time” and “previous run time”
- Age-group and gender converted to numeric values
- Dropped unnecessary columns:
 - Date, Position, Name, ID, etc.

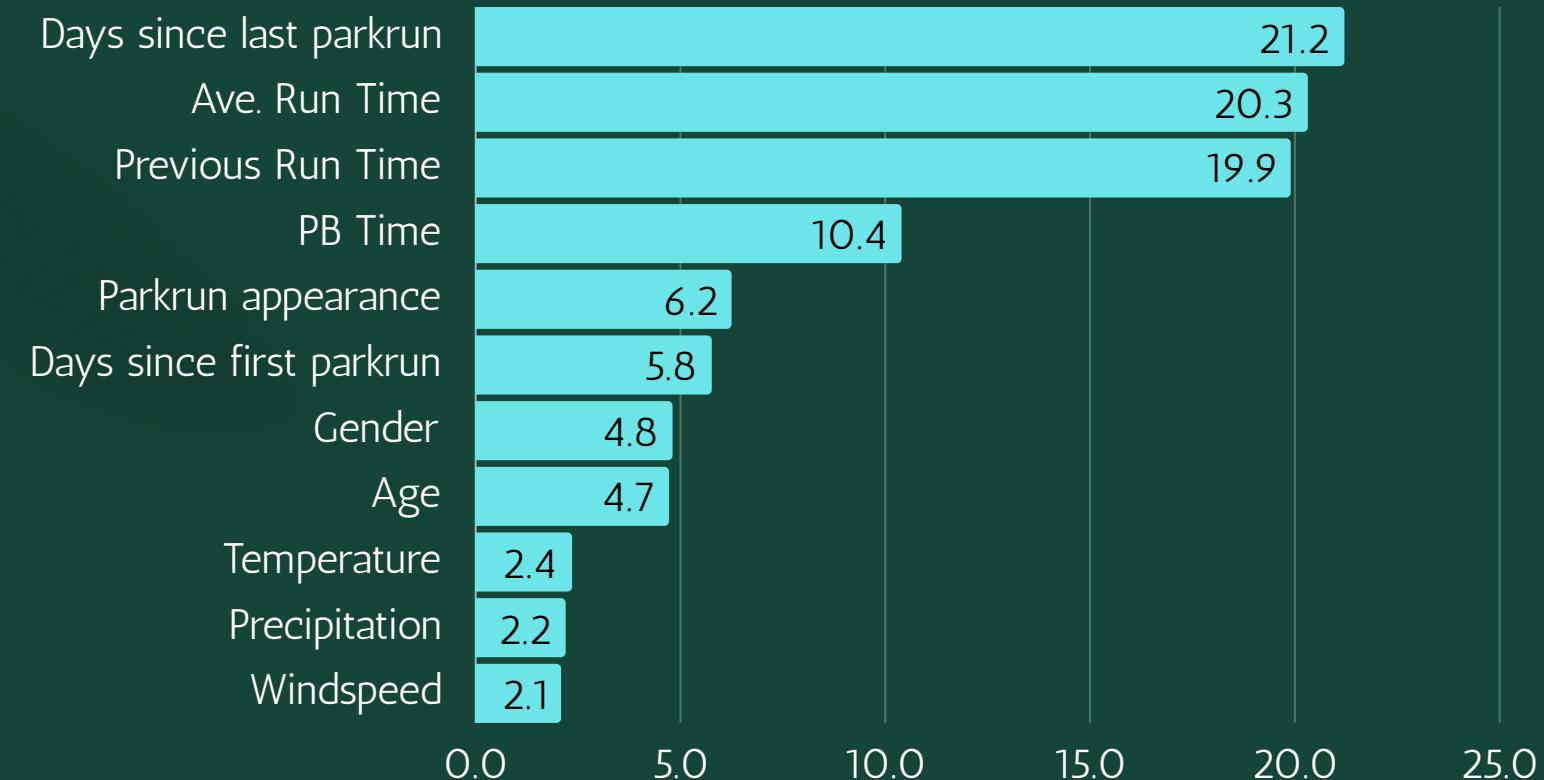
Training models

Various models were built, using a range of methods. Initial models overestimated times for faster runners and underestimated for slower runners. Therefore these were adjusted to predict the **time-change-index** (new time / previous time), and then calculate the time.

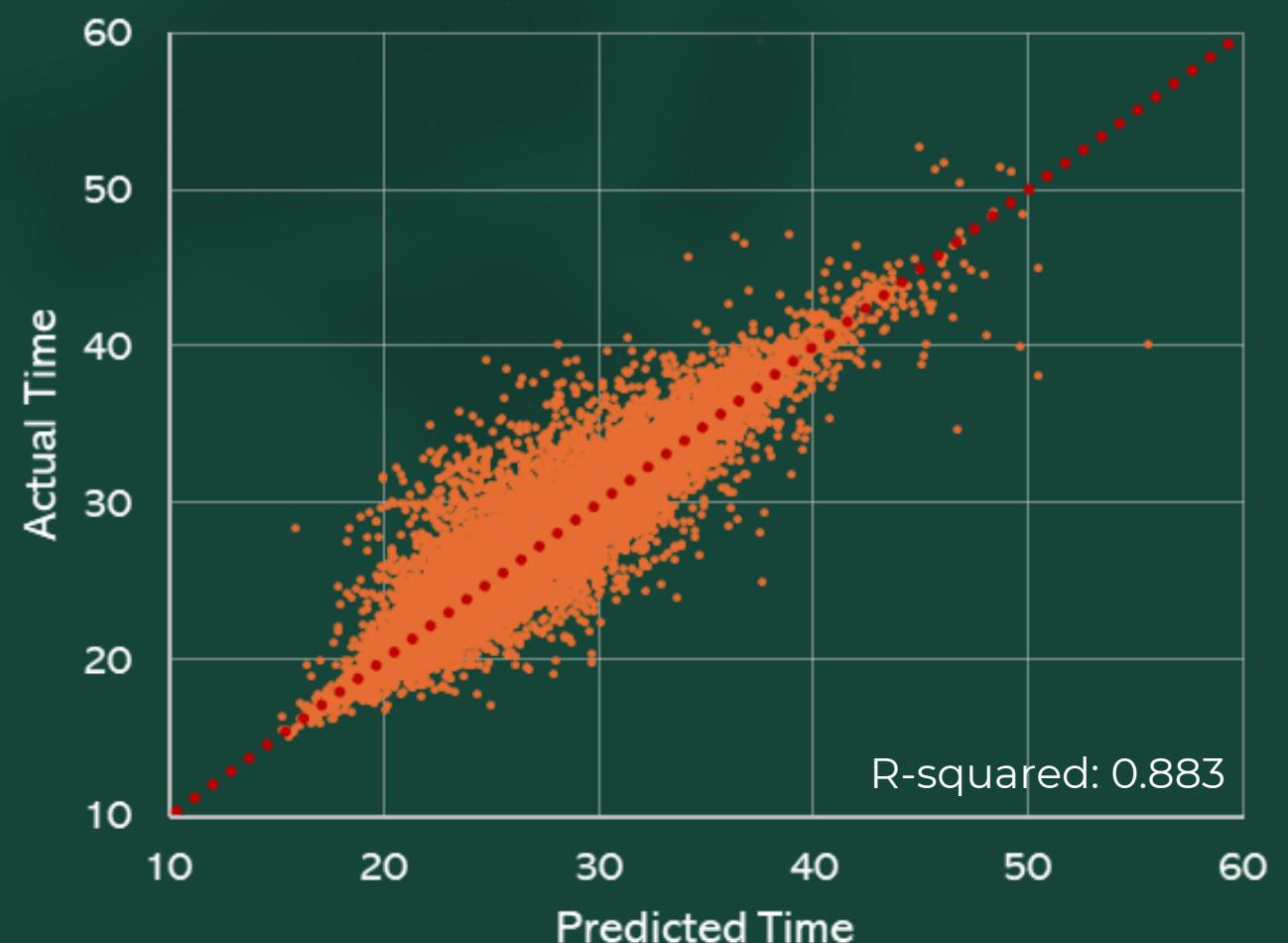
Optimisation

XGBoost was the method of the initial models that produced the best predictions, so this was further optimised with optuna hyperparameter tuning

Optimised Model - Feature Importance



Optimised Model Test - Prediction vs Actual



USING THE PREDICTOR

There is a notebook predictor designed so that you can manually input your stats and it will produce a target time

OR

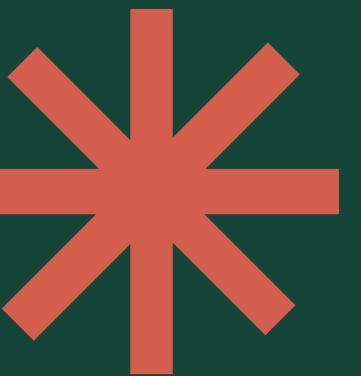
If you have your parkrun ID, you can use a Streamlit app to input weather conditions.

The app will scrape your stats from the parkrun website and use them to generate a prediction.



Demo

POTENTIAL IMPROVEMENTS



- Current reliance on scraping from the parkrun website makes the functionality susceptible to format changes. Supplementary data sources or full integration with park run data could future-proof the model.
- Further tuning and optimisation of the models could be completed, or incorporation of additional features such as terrain type, gradient, etc. could enhance predictions.
- Extend the model to support other running distances beyond 5km park runs.
- Incorporate weather forecasting API to automatically factor into targets.
- Create a more user-friendly web-interface or app with instant predictions.





THANK YOU

[GitHub](#)