Identification of Training Patterns in Race Preparation

By Mark Kennedy
Supervisor: Dr. Aonghus Lawlor

University College Dublin
An Coláiste Ollscoile, Baile Átha Cliath

The Growing Popularity Of Running

- Proven benefits for a healthy lifestyle

  (Physical & Mental).
- Suitable for people of all ages/genders.
- Caters for all levels of fitness.
- Low participation costs

  (No expensive equipment required).
- Highly satisfying (Individual goals).

Unprecedented demand as 2019 Dublin Marathon sells out **THE IRISH TIMES**

Ever Evolving Fitness Apps

- Recent growth in the popularity of Running.
- Advancements in technology

  (e.g iPhones, Wearable Technology).
- Such apps are heavily endorsed by major companies

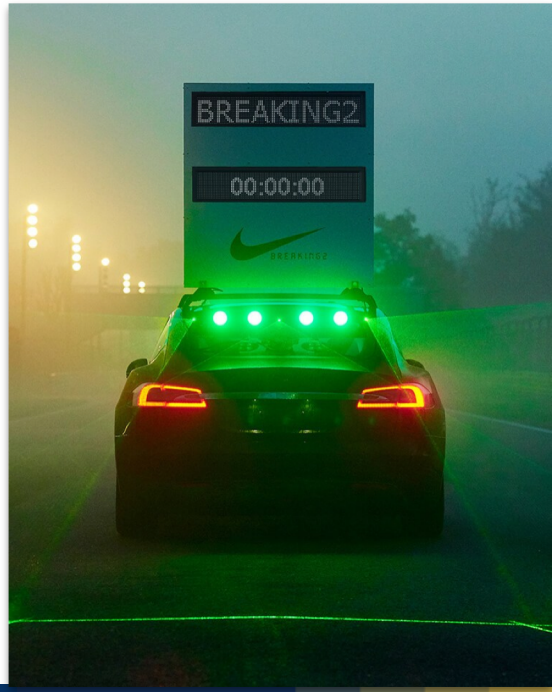
  (e.g. Nike, Under Armour).

myfitnesspal

Research Objectives

1. How do people really train for marathons?
2. What is the most effective way to prepare for a marathon?

   (i.e. Most important features/patterns to include in your training programme)
3. Can I use my findings to recommend more suitable training programmes for individual athletes, based on their desired finish time, and sessions completed to date?



The Data

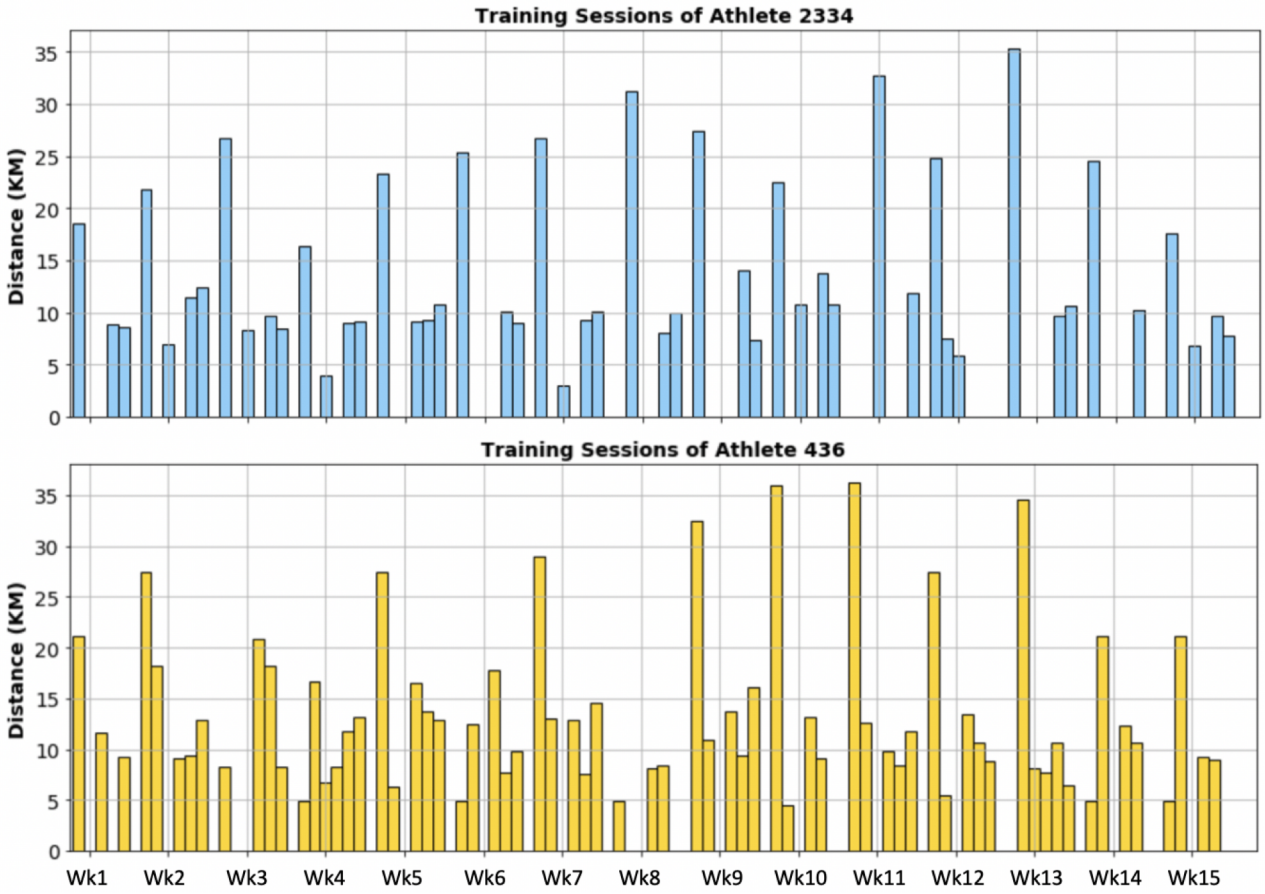
Source: STRAVA

Marathon: London 2015

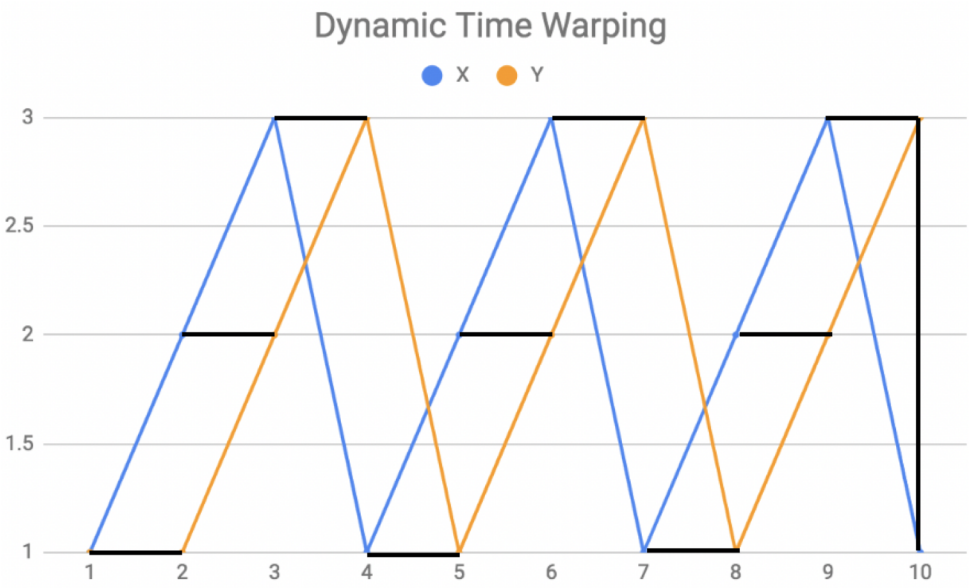
4487 Athletes

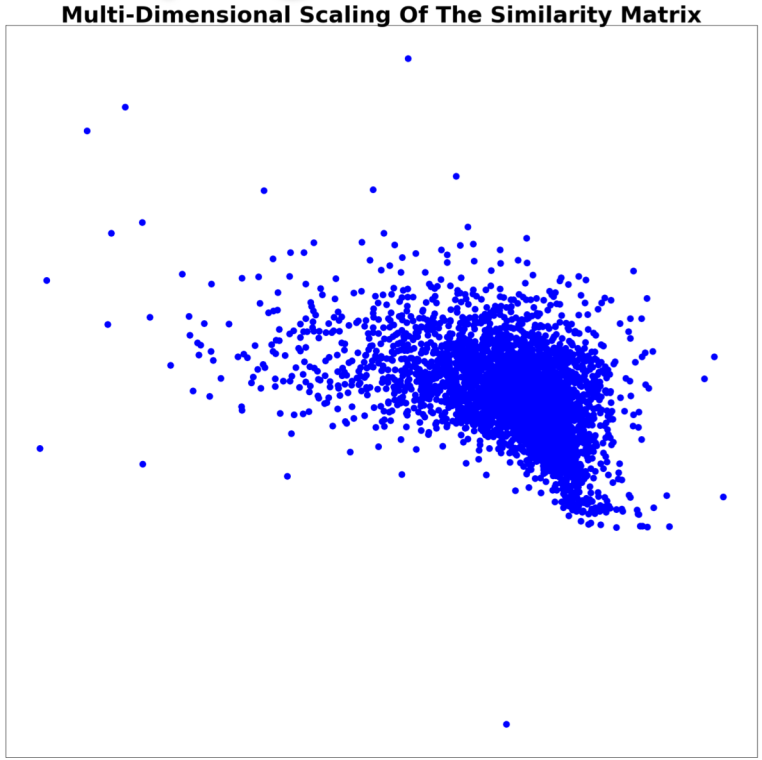
550,000+ Sessions

20 Features

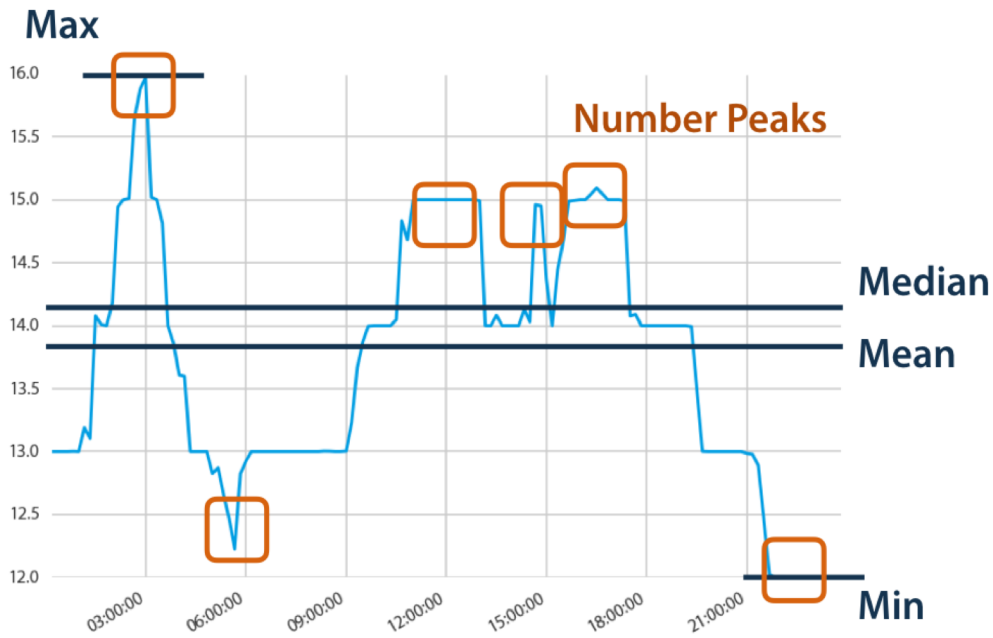


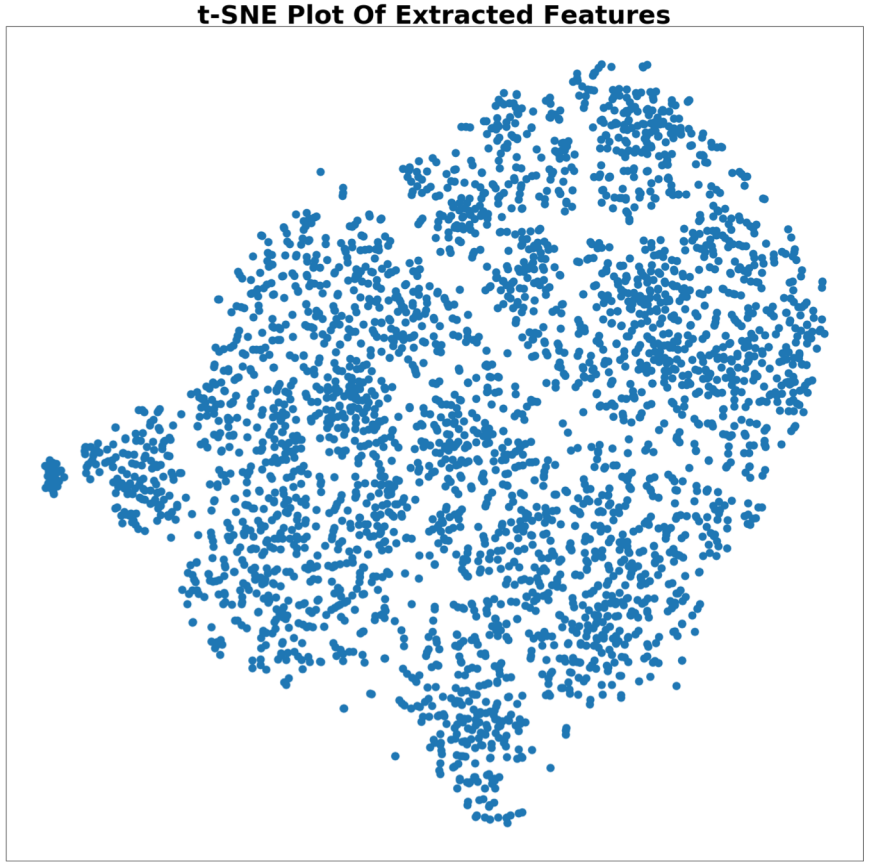
Approach 1: Dynamic Time Warping & MDS



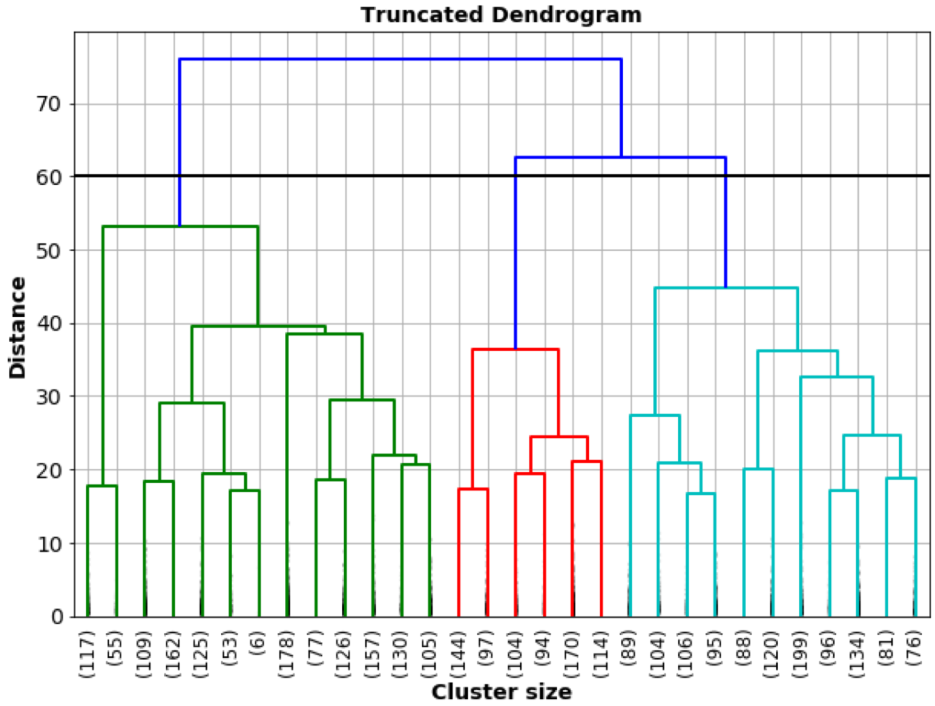


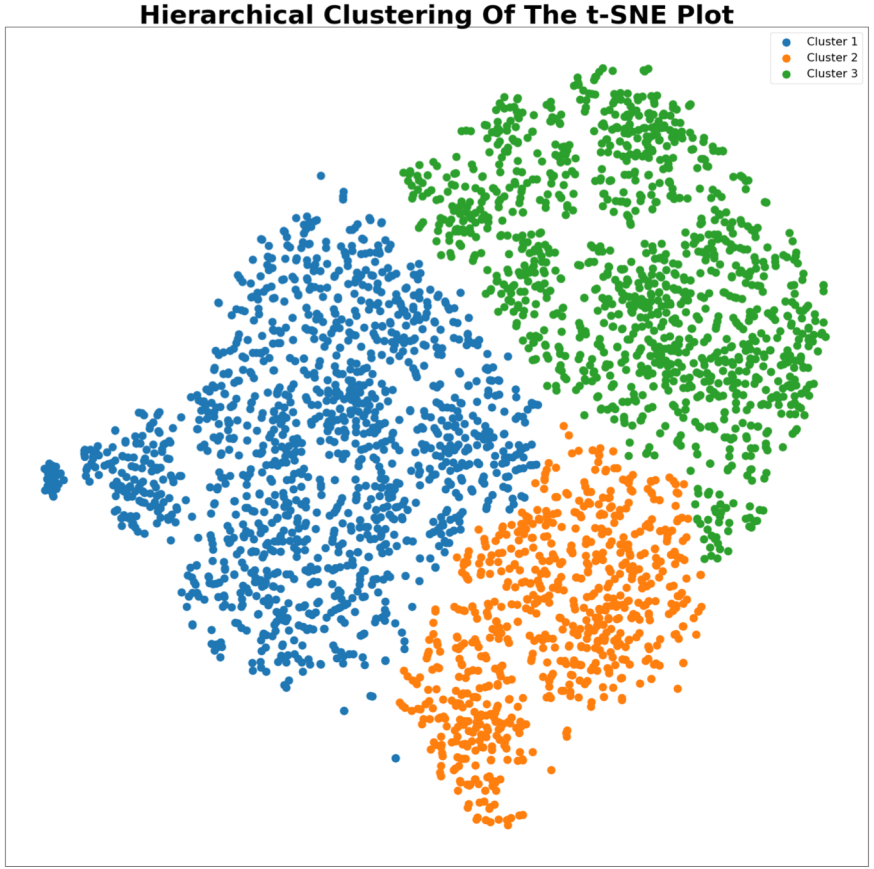
Approach 2:
Feature Extraction & TSNE



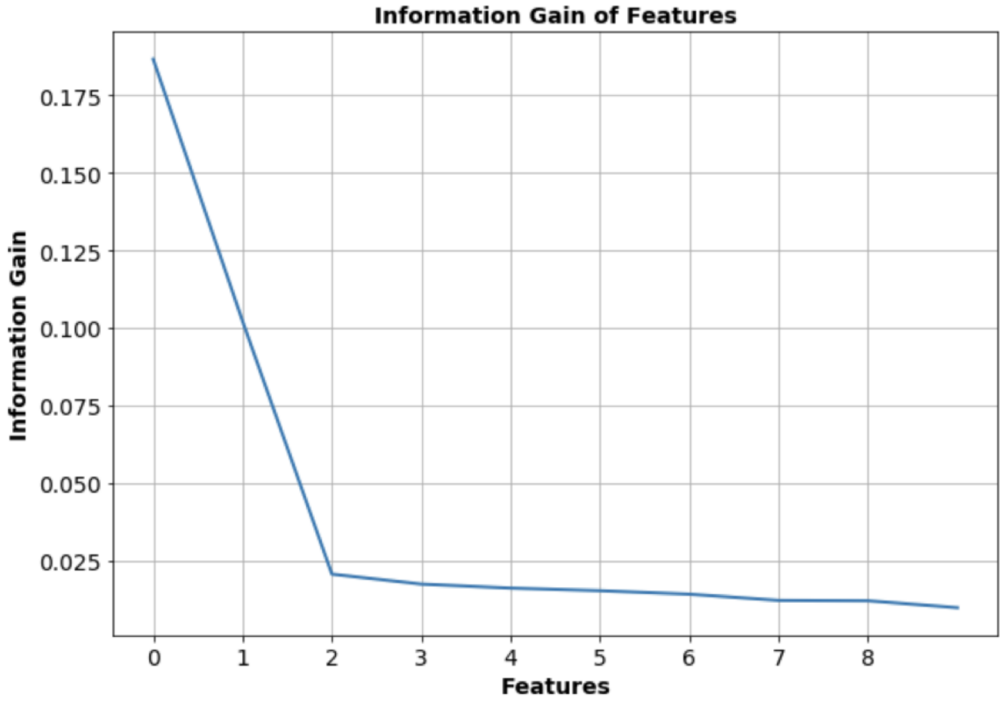


Hierarchical Clustering





Information Gain Of Extracted Features

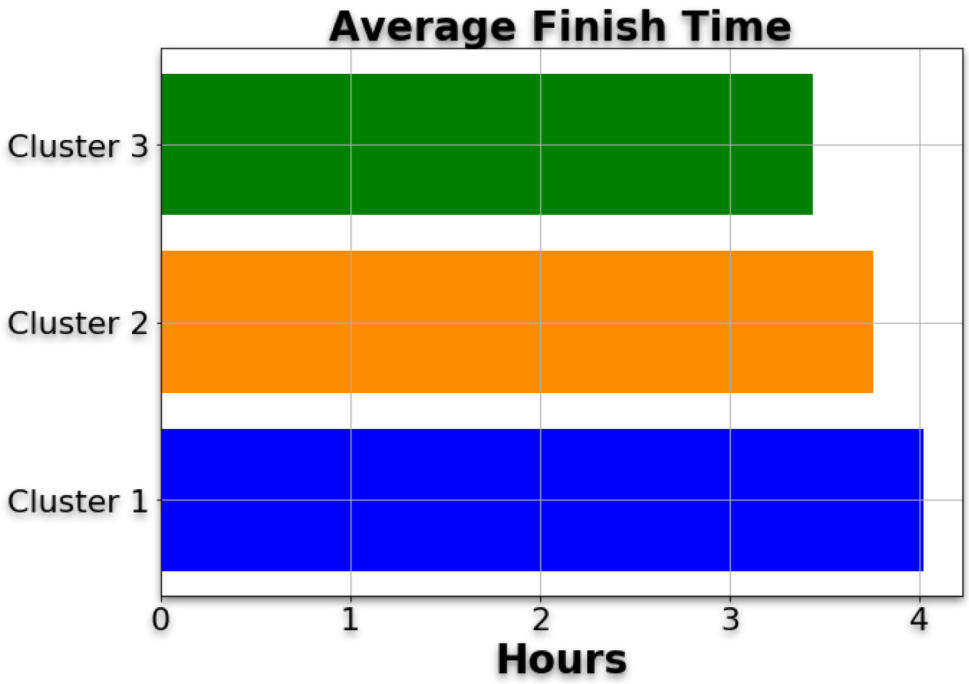


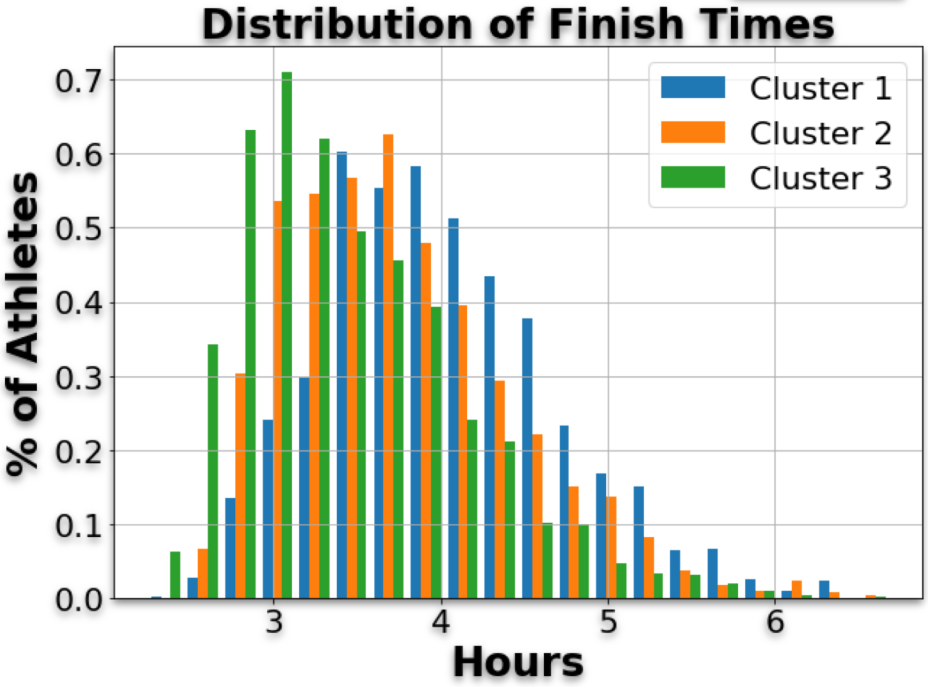
	Feature	Importance
0	session_pace__large_standard_deviation__r_0.2	0.186727
1	totaldistance__large_standard_deviation__r_0.25	0.102147
2	session_pace__ratio_beyond_r_sigma__r_3	0.020503
3	session_pace__large_standard_deviation__r_0.25	0.017303
4	session_pace__approximate_entropy__m_2__r_0.3	0.016004
5	totaldistance__ratio_beyond_r_sigma__r_3	0.015167
6	session_pace__binned_entropy__max_bins_10	0.014042
7	totaldistance__symmetry_looking__r_0.1	0.012038
8	session_pace__symmetry_looking__r_0.05	0.011920

Boolean variable denoting if the standard dev of x is higher than 'r' times the range = difference between max and min of x. Hence it checks if

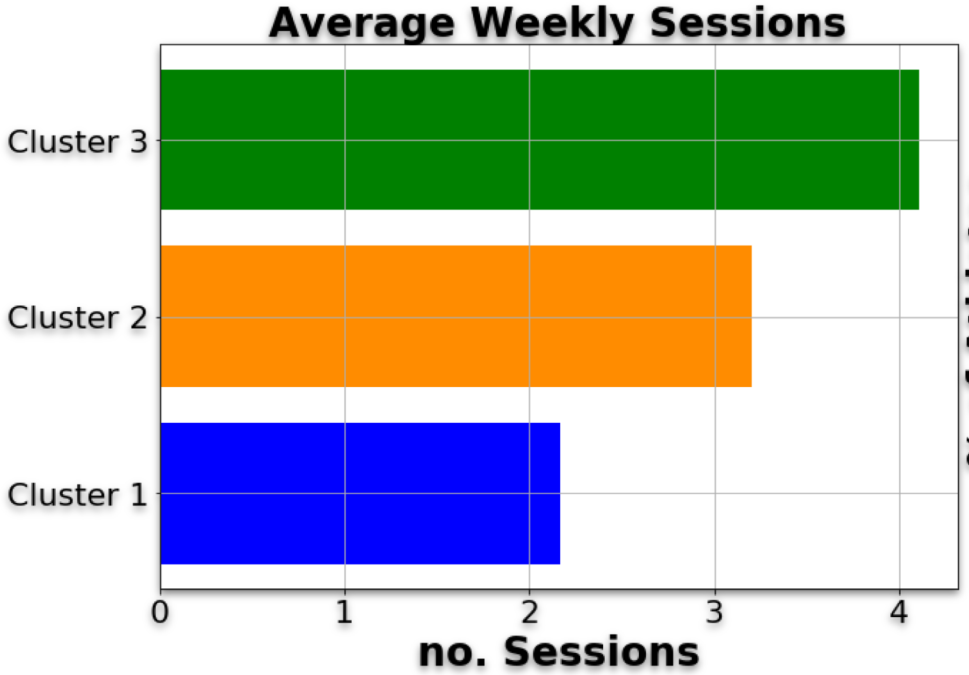
$$std(x) > r * (max(X) - min(X))$$

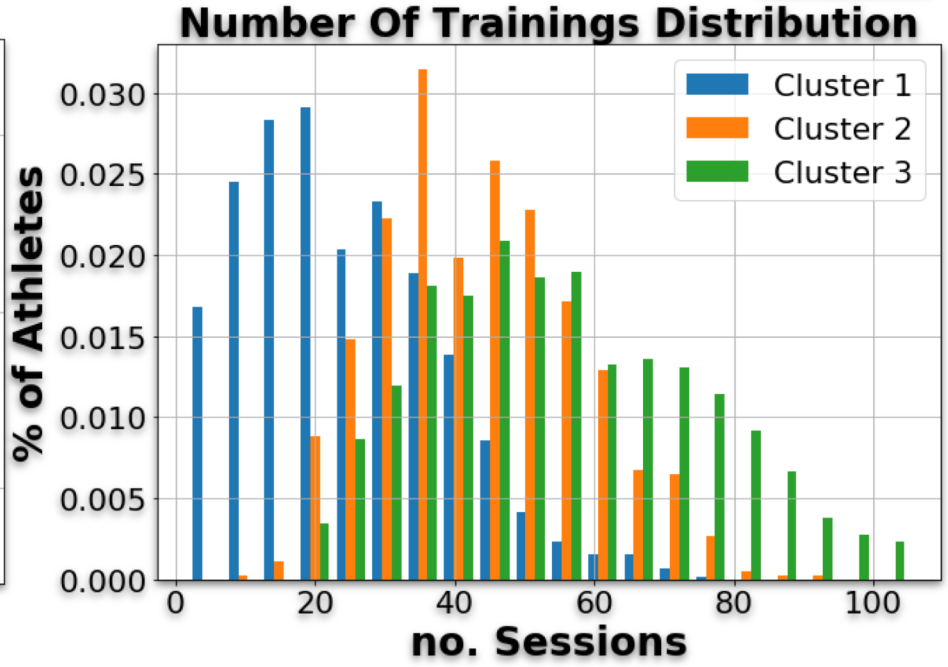
Cluster Analysis



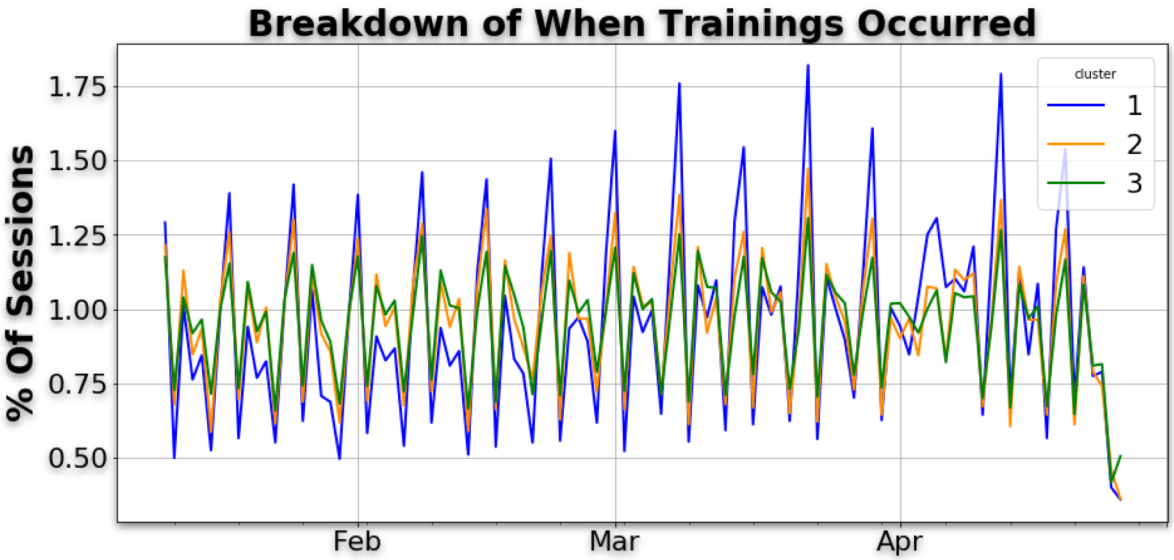


Cluster Analysis





Cluster Analysis



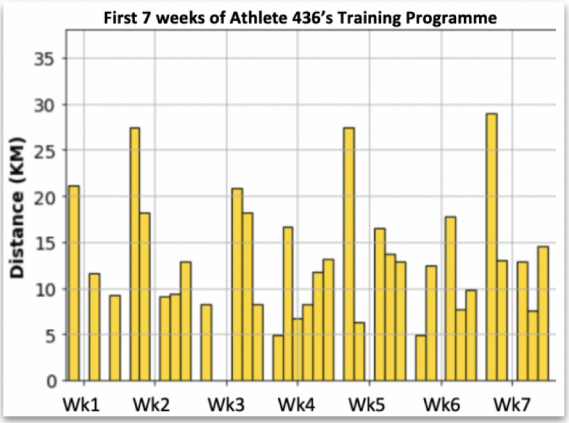
And There's Much More!

In the report I examine…

- Additional features (e.g. distance, pace)
- Correlations between certain features and marathon finish times.
- The extracted features.
- The credibility of some popular theories around training for a marathon.

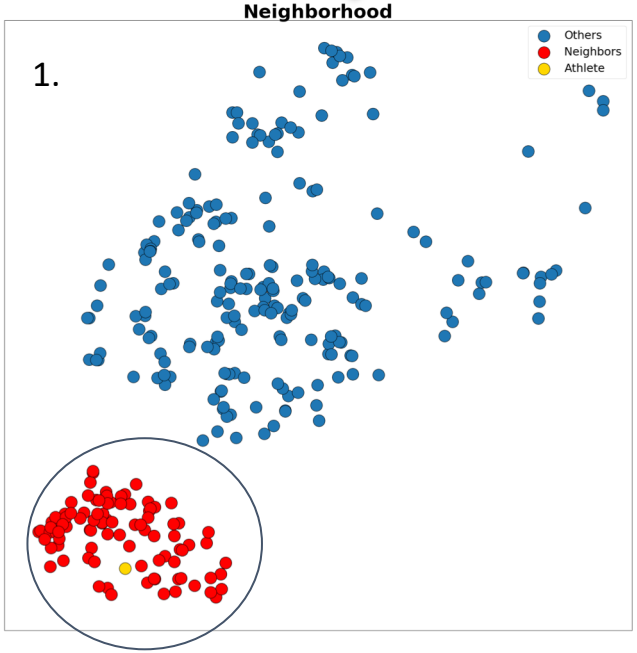
Training Plan Recommender

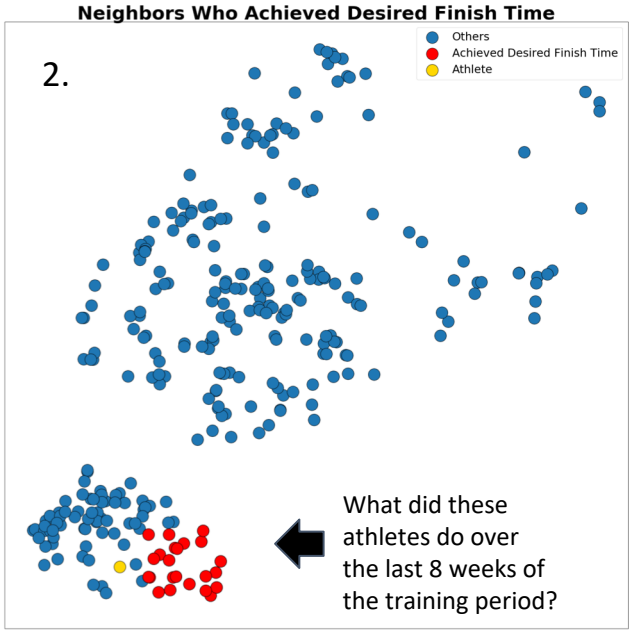
Sample Input:



Desired Finish Time: 2:50:00

Process Steps:

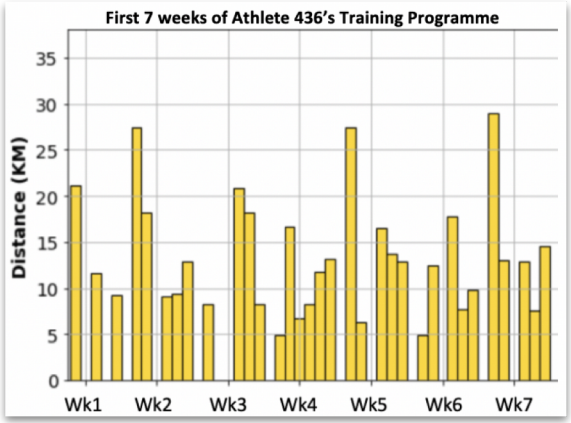




What did these athletes do over the last 8 weeks of the training period?

Training Plan Recommender

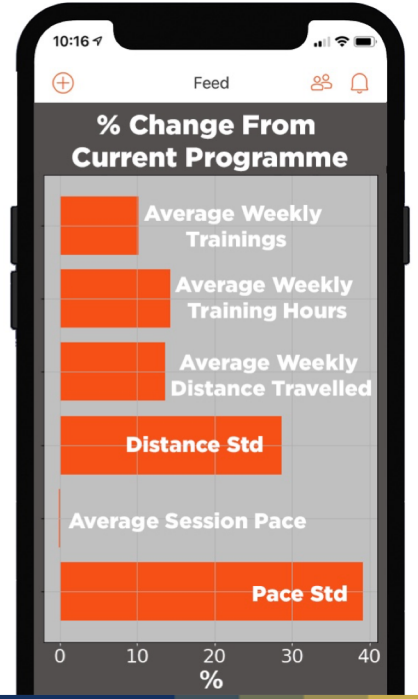
Sample Input:



Desired Finish Time: 2:50:00

Sample Output:

Targets For Last 8 Weeks



Q&A