

## Tech conference attendee segmentation using data-driven methods: Identifying clusters based on patterns of engagement

Georgios Lamprou\* February 2022

## Abstract

It is becoming increasingly common for companies to try to understand their customers and take various actions using data-driven methods. The data availability, the increased storage capacities and the improvements in computing have all contributed towards this direction. Web Summit, an Irish technology conference company, is no exception and has taken advantage of these developments.

In this thesis, an optimal division of attendees from three conferences into segments was the desired outcome and cluster analysis was the chosen approach. The segments would consist of homogeneous sub-populations which share common behavioural aspects regarding patterns of engagement per conference. The company's mobile app was the source of the various engagement metrics, which were subsequently aggregated.

To achieve the above, three clustering methods were chosen to be fitted, validated and compared. Those methods were K-Medoids, Agglomerative clustering with average-linkage and ("fuzzy") HDBSCAN. The data were transformed before clustering with UMAP, a fairly new dimensionality reduction method, after a simulation study which showed that UMAP-assisted clustering provides a performance advantage.

It was concluded that overall Agglomerative clustering with averagelinkage had a small advantage over the other methods. The cluster centroids revealed a shared pattern that underlines all clusters, namely the increase in engagement as the event approaches and the abrupt decrease of it when it ends. Most importantly it revealed spikes in engagement on different days, which is what makes the clusters distinct.

<sup>\*</sup>Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: education@glamprou.eu. Supervisor: Taras Bodnar.