

PROJECT 2

Data Mining in Engineering

Objective

The primary objective of this project is to practice implementing unsupervised learning techniques. You have been tasked to consult a running footwear & apparel company on strategic decisions regarding their running product & marketing. You will be using clustering to identify if there exist key running consumers using data from a recent survey to base your recommendations on. Deliverables include end-to-end clustering analysis with in-depth annotations to back your recommendations.

Project Details

- You may work solo or collaborate with up to 3 other people in the class. Regardless of whether you choose to work solo or collaborate, you must submit your work individually. Duplicate files in the class will not be accepted and original work is expected. You must list all names of individuals who you collaborated with.
- You are free to use any online resources and/or resources from class as reference.
- **Dataset will be released at 6PM on Wednesday 12/06. You will have the duration of the class to work on the project. Project submissions are due in 48 hours by 6PM Friday 12/08.**

Project Details

The dataset will contain survey data coming from a pool of participants that answered various questions regarding their running habits. You will receive an excel file containing this data along with a datamap that describes each field.

Deliverables

Grading will be on completeness, thoroughness, and your justifications for steps taken and recommendations made to the company. By 6PM on Friday 12/08, you must submit a Python Google Colab Notebook that contains (at minimum) the following sections:

1. **Objective:** paragraph explaining what this notebook covers.
2. **Team Members:** list of people you collaborated with (or indicate if you worked solo).
3. **Data Preparation:** format the data so that it's best suited to your analysis.
4. **Similarity Metric & Matrix:** justify an appropriate similarity metric and calculate.
5. **Clustering Algorithm:** justify methods, iterate as needed.
6. **Interpret Results:** what (if any) clusters of consumers were found? What recommendations do you have for this company regarding its product portfolio and marketing strategies? Could this company personalize its approach to key consumers in any way?