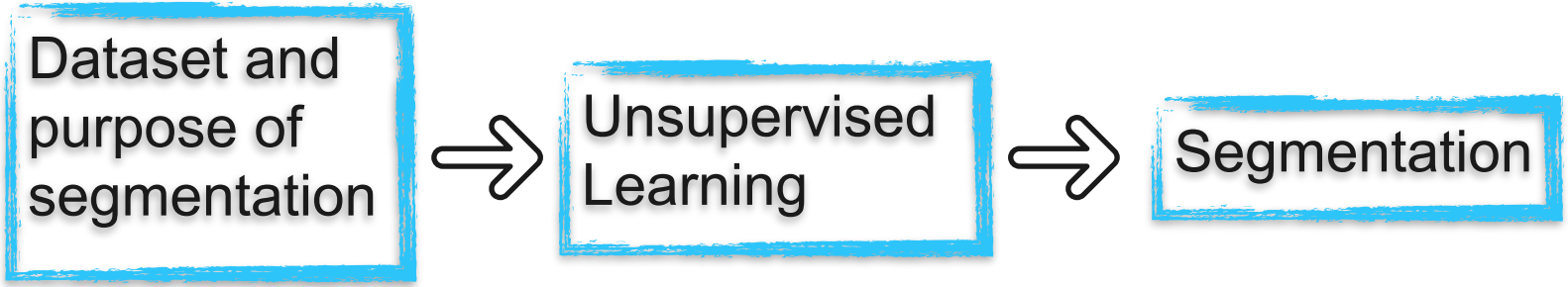
Project 6 - Clustering

Finding patterns in your data

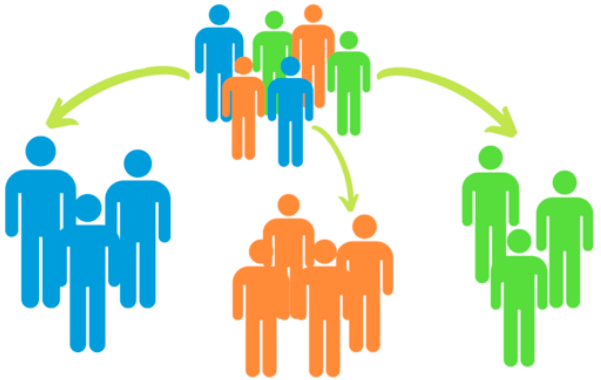
Main Objectives

Use cluster analysis to identify the groups of similar characteristics in your dataset.



Specific Objectives - Technical

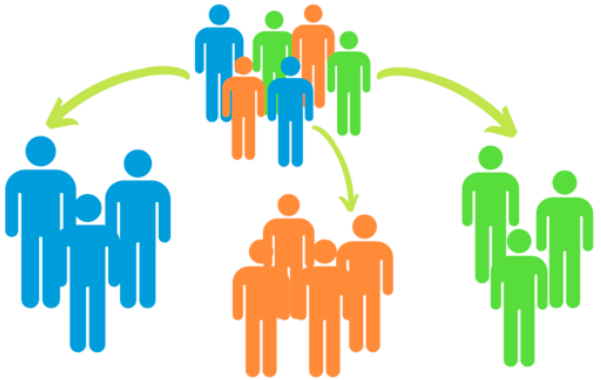
1. You should use both a **partitioning technique** (K-Means, DBScan, etc) and a hierarchical technique to obtain your segmentation.
2. Explain the reason you are using the **distance** you've chosen for the hierarchical algorithm.



Specific Objectives - Technical

3. You should use **PCA** and **interpret** its results.

4. **Visualize** the results of your **PCA and** the **clusters**.



Questions

- **Interpretation:** Is it possible to explain what each cluster represents? Do your features enable a meaningful interpretation of the clusters? Do the compositions of the clusters seem to make sense?
- **Tuning**: How did you choose the parameters of your model? Don’t think solely on the machine learning metrics, think about what you are trying to solve.
- **Evaluation**: How did you check whether your choice was good enough?

Hints

- Use the unsupervised learning techniques not only to achieve the final result (the segmentation itself) but also to understand your data.
- Don’t forget what your observations are and what is the purpose of your segmentation.
- This is an open-ended assignment. Your only constraints are: execute high-quality and justifiable clustering technique, provide your rationale for the decisions you made, and produce meaningful clusters.

Deliverables

- Jupyter Notebook - Documented with rationale of decisions

Datasets

You can choose from one the following options:

1. The dataset you've obtained in the Data Gathering project.
2. The data you think of using for your final project.
3. YOUNGS survey dataset
4. FIFA Dataset

When

- Turn in until 9AM


IRON
HACK