

PHASE 2

Problem statement:

Conducting clustering analysis to identify patterns among different industrial categories and age groups is a valuable approach in data analysis. You can use techniques like k-means clustering or hierarchical clustering to group similar categories and age ranges based on dataset. This can help reveal insights and relationships within data that may not be apparent initially. If you have specific questions or need guidance on how to perform this analysis, feel free to ask for more detailed advice.

1. Data Collection: Begin by gathering data. You should have a dataset that includes information on industrial categories (e.g., types of industries) and age groups. Each data point should be associated with both a category and an age group.

Collected data from the given project to analysis the categories of the people distributed in various districts, ages ,working hours ,industry categories

2. Data Preprocessing:

- Ensure that data is clean and free from errors or missing values
- Standardize or normalize data if necessary to bring all variables to a common scale.
- Null value, age not stated these values are optimised

3. Choosing a Clustering Algorithm:

- Selecting an appropriate clustering algorithm. Two common choices are:
- K-Means Clustering: Divides data into 'k' clusters based on similarity.
- Hierarchical Clustering: Forms a tree-like structure of clusters, showing relationships at different levels of granularity.

4. Feature Selection and Engineering:

- Decide which features (variables) you want to use for clustering. In case, it's industrial categories and age groups.
- You might need to transform or combine features to make them suitable for clustering.

5. Clustering Analysis:

- Apply the chosen clustering algorithm to data.
- For K-Means, you'll need to specify the number of clusters (k) in advance.
- For Hierarchical Clustering, you can choose to cut the tree at a certain level to define clusters.

6. Interpretation and Visualization:

- Visualize the clusters using plots or graphs to understand how data points are grouped.
- Explore the characteristics of each cluster to interpret the patterns that emerge.

7. Validation and Refinement:

- Evaluate the quality of clusters using appropriate metrics (e.g., silhouette score, Davies-Bouldin index).
- If the results are not satisfactory, consider adjusting parameters or trying different algorithms.

8. Insights and Recommendations:

- Based on clustering results, draw insights and make recommendations. For example, you might discover that certain age groups are more prevalent in specific industries, which can inform marketing or policy decisions.

9. Documentation and Reporting:

- Documenting entire process, including data preprocessing, clustering parameters, and results.
- Presenting findings in a clear and understandable manner, possibly through reports or presentations.

10. Iterate if Needed:

- Depending on goals, you may need to refine analysis or perform additional iterations to gain deeper