

Cluster Analysis on Algerian economic data

Description:

The project aims to perform a cluster analysis on Algerian economic data based on a set of economic indicators. The study should mainly focus on two aspects:

- data visualization
- and prediction/classification by clustering approach.

Students are asked to realize the following parts.

1- Data acquisition

The source is the World Bank website. Use the link: <http://wdi.worldbank.org/tables>

Students have to download files and construct two datasets:

- The first dataset contains the following economic indicators
 - From the **Data World tab**, download the **size of the economy** file. The 3 first columns (**population, surface area, population density**) are not used as economic indicators.
 - From the **economy tab**, add attributes that are included in: **growth of output, structure of output, structure of manufacturing and structure of merchandise exports** files.
 - From **people tab**, add **unemployment** data.
 - From **poverty and shared prosperity**, include **poverty data at international line** (2 files).
- The second dataset concerns the different resources (human, natural) and potentials. The indicators used in this dataset are:
 - **Population, surface area and population density** from **size of the economy** file.
 - From **environment tab**, add data from: **rural environment and land use, agricultural inputs, fresh water, energy production and use, contribution of natural resources to gross domestic product** files.
 - From **people tab**, add **population dynamics, labor force structure, participation in education and education completion and outcomes** data.

2- Data visualization

Students have to use the **ggplot2 package** to propose and produce data graphics of the two datasets.

3- Clustering

Students have to apply the **k-means** algorithm that has been implemented previously with the **Euclidian distance**. Produce two clustering results using the datasets.

Clustering results should also be presented with graphics using ggplot2 package.

4- Reporting

Based on the previous results, students should prepare an analysis report illustrating a:

- Comparison between results of clustering.
- Prediction of economic classification based on resources and potentials clustering.

Students are divided into teams:

	Team 1	Team 2	Team 3
Leader	Bouhadda	Labiod	BOUHOUITA GUERMECH
Members	- AZOUZ - REFFAI - HAMMOUCHE - BENDJEDDOU	- LIHOUM - OUARTI - RAHIM	- SISSAOUI - M'HAMDI - BOUGUERRA - BELLAGHMES

The rest of students are asked to form teams of 3 to 5 members by Thursday, March 4th at 23:59.

Deadlines:

- The first part (1st clustering) should be delivered by Sunday, March 7th at 23:59.
- The project completion deadline is Tuesday, March 9th at 23:59.