10Alytics

Specialization: *Unsupervised Learning*

Business Focus: *NGO*

Tool: *Pandas, Matplotlib, Seaborn, Sklearn, etc.*

Addressing social issues by Clustering the Countries for HELP International

Tools and Technology to be Used

10Alytics

Background

- HELP International is an international humanitarian NGO that is committed to fighting poverty and providing the people of developing countries with basic amenities and relief during the time of disasters and natural calamities.
- Also The NGO Aims to Improve The Health and Wealth Of Marginalized Nations all around the Globe As communicated in their Mission statement

10Alytics

Mission Statement

- The Organizations Problem Statements are outlined as follows:
  a. Fast Tracking Access to Health care For Vulnerable Persons in Developing Countries
  b. Addressing Prejudice Against Women And Children Welfare Especially in Developing Nations
  c. Improving The Income Capacity for Lower Income Individuals
  d. Addressing And Empowering Individuals to be goods and Service Providers As Opposed to being only Consumers in a bid to improve National Output

10Alytics

Problem Statement

- Help Has Been Collecting Data relevant to its problem statement in recent times.
- It Has been Decided by the CEO to Leverage Insights from data to identify nations that Embody the Problems they Are Trying to Address
- For this cycle, Donations Amounted up to $30M, The CEO wants you, the Partner Data scientist to look at the data and provide him with a group of countries that are at the epicenter of the problems they are trying to Address.
- Also the Benefactors, this time, want Countries with low output(gdpp) and lower life expectancy to be prioritized when Distributing AID to these countries.

10Alytics

Data Description

Help International collected data contains columns relevant to the problems it is trying to address:

General:

country: Name of Country

Socio-Economic Metrics:

exports: Exports of goods and services per capita. Given as percentage of the GDP per capita

Imports: Imports of goods and services per capita. Given as percentage of the GDP per capita

Income: Net income per person

Inflation: The measurement of the annual growth rate of the Total GDP

gdpp: The GDP per capita. Calculated as the Total GDP divided by the total population.

Health Metrics

life_expec: The average number of years a newborn child would live if the current mortality patterns are to remain the same

child_mort: Death of children under 5 years of age per 1000 live births

health: Total health spending per capita. Given as percentage of GDP per capita

total_fer: The number of children that would be born to each woman if the current age-fertility rates remain the same.

10Alytics

Task

- **The Case Study is divided in 2 folds:**
  a. **First group the country dataset optimally so we can successfully group and identify countries in critical conditions**
  b. **Then explore countries in the groups to identify the countries to prioritize when allocating resources.**

10Alytics

Project Workflow

STEP 1

Data Cleaning:

- *Handle any missing values.*
- *Remove duplicate records or irrelevant columns.*
- *Check for and correct anomalies.*

STEP 2

Exploratory Data Analysis (EDA):

- *Visualize distributions and relationships between features.*
- *Identify patterns, trends, and potential anomalies.*
- *Form hypotheses based on data insights.*

STEP 3

Data Preprocessing:

- *Scale/normalize numerical features and encode categorical data.*
- ***Feature Engineering***

STEP 4

Model Training:

- *Select and train machine learning models on the data.*
- *Elbow Method to find the optimal number of clusters (k).*
- *Experiment with different algorithms and assess performance.*

STEP 5

Model Evaluation (Interpreting Results):

- *Visual Inspection: visualize the clustering results or reduced features to evaluate if meaningful patterns emerge.*

10Alytics

READY TO DELVE IN?