

Вам предоставлены данные с описанием ситуации в странах, данные содержат следующую информацию:

name - название страны

child_mort - смертность детей в возрасте до 5 лет на 1000 живорождений

exports - экспорт товаров и услуг на душу населения. Приведено в % от ВВП на душу населения

health - общие расходы на здравоохранение на душу населения. Указаны как % от ВВП на душу населения

Imports - импорт товаров и услуг на душу населения. Указано в % от ВВП на душу населения.

Income - чистый доход на человека

Inflation - измерение годового темпа роста общего ВВП

life_expec - ожидаемая продолжительность жизни

total_fer - ожидаемая рождаемость

gdpp - ВВП на душу населения

Представьте, что вы анализируете данные для международной благотворительной организации, которая оказывает поддержку наиболее нуждающимся странам, на основе этих данных выберите страны, которым помощь нужна в первую очередь.

1. Анализ и предобработка.
 1. Проанализировать данные (EDA).
 2. Предобработать данные.
 3. Скалировать/нормализовать данные.
2. Решите задачу с помощью следующих методов:
 1. K-means/mini batch k means
 2. Иерархическая кластеризация
 3. DBSCAN
3. Для каждого метода определить оптимальное количество кластеров (построить график)
4. Сделайте выводы, определите, каким странам нужно помогать в первую очередь.
Опишите выделите кластеры. Опишите эталонную страну в каждом кластере.