Fuad Hassan

CIS 3715

Final Project Progress Report I

April 7, 2024

Country Clustering Analysis Using Unsupervised ML

## Introduction:

The past we I have worked on exploring the data and understanding the data as much as possible. My data has total of 10 features about each country. They include:

* **country:** Name of the country
* **child\_mort:** Child mortality rate (per 1000 live births)
* **exports:** Exports of goods and services (% of GDP)
* **health:** Total health spending (% of GDP)
* **imports:** Imports of goods and services (% of GDP)
* **income:** Net income per person
* **inflation:** The measurement of the annual growth rate of the Total GDP
* **life\_expec:** Life expectancy at birth (in years)
* **total\_fer:** The number of children that would be born to each woman if the current age-fertility rates remain the same
* **gdpp:** Gross Domestic Product per capita

My goal is to use these features to categorize all the counties into 3 different sectors. They are

* **Highly Developed**
* **Least Developed**
* **Lower-Middle Developed**
* **Upper-middle developed**

## ERD:

### Data sample:

**Data table:**

A table with numbers and text

Description automatically generated

|  |  |
| --- | --- |
| **Individual country:** | **Data Type:** |

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

**Data Correlation:**

A table of numbers and symbols

Description automatically generated with medium confidence

List of things that already has been done:

1. Exploration of data
2. Handling missing value
3. Handling outliers
4. Exploration statistic of data
5. Finding correlations

List of things that needs to be done:

1. Data visualization
2. Data Transformation
3. Feature extraction
4. PCA
5. Model Development

List of things that will be done this week:

1. Data visualization
2. Data Transformation
3. Feature extraction