Fuad Hassan

CIS 3715

Final Project Progress Report I

April 9, 2024

Country Clustering Analysis Using Unsupervised ML

## TODO:

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

## Introduction:

This week I have worked on visualization the data that I have to better understand the relationships between features. I have created an interactive map for each feature, that helps you to visualization each of the feature’s impact each country.



## Transformation:

Next, Data seemed to be very positive skewed. My goal is to transform them to a normal destitution; however, a normal distribution is not requirement but can help achieve better results since outliers might skew the cluster centroids. For this I will perform BoxCox Transformation, which seems to work the best for our data.

A chart of blue dots

Description automatically generated with medium confidence

## Model Development

The first model I am planning on preformation is K mean. After that we will perform couple other models to evaluate and pick the best model.