Directory Structure

- `DATASET` : Contains the dataset files for the assignment.
- `SOURCE_CODE` : Contains the source code files for the assignment.

DATASET

- `country.csv` : The dataset for the assignment.

SOURCE_CODE

- `main.py` : The main script for the assignment.
- `requirements.txt` : Contains the required packages for the assignment.
- `makefile` : Contains the commands for running the project.

Usage

To run the project, execute the following command from terminal from SOURCE_CODE directory:

```
make model
```

To clean the project, execute the following command from terminal from SOURCE_CODE directory:

```
make clean
```

Setup Instructions

1. First Download the dataset and place it in the DATASET directory. Name the dataset as 'country.csv'.
2. Install the required packages using the following command:

```
pip install -r requirements.txt
```

3. Run the project using the following command:

```
make model
```

Outputs

A new directory named `OUTPUT` will be created in the main directory. The directory will contain the output files generated by the project.

These files containt the following information:

1. Total number of clusters.
2. For each of the clusters, the following information is provided:
    - Cluster number.
    - Number of countries in the cluster.
    - Attribute values for the centroid of the cluster.
    - List of countries in the cluster along with their attribute values.

- `cluster_output_3.txt` : Contains the output of the clustering model with 3 clusters.
- `cluster_output_4.txt` : Contains the output of the clustering model with 4 clusters.
- `cluster_output_5.txt` : Contains the output of the clustering model with 5 clusters.
- `cluster_output_6.txt` : Contains the output of the clustering model with 6 clusters.
- `cluster_output_hierarchical.txt` : Contains the output of the hierarchical clustering model.

The following files contain the cluster information for each of the clustering models as specified in the assignment and is stored in `SUBMISSION` directory:

- `divisive.txt` : Cluster information for divisive clustering.
- `kmeans.txt` : Cluster information for kmeans clustering.