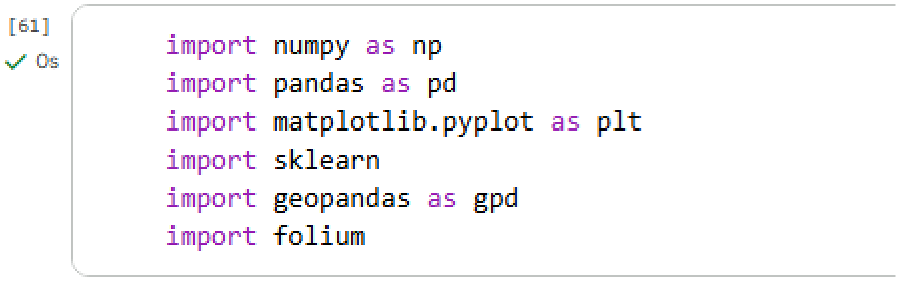
**Althea Maxene B. Viar**
**BSCS 3B**

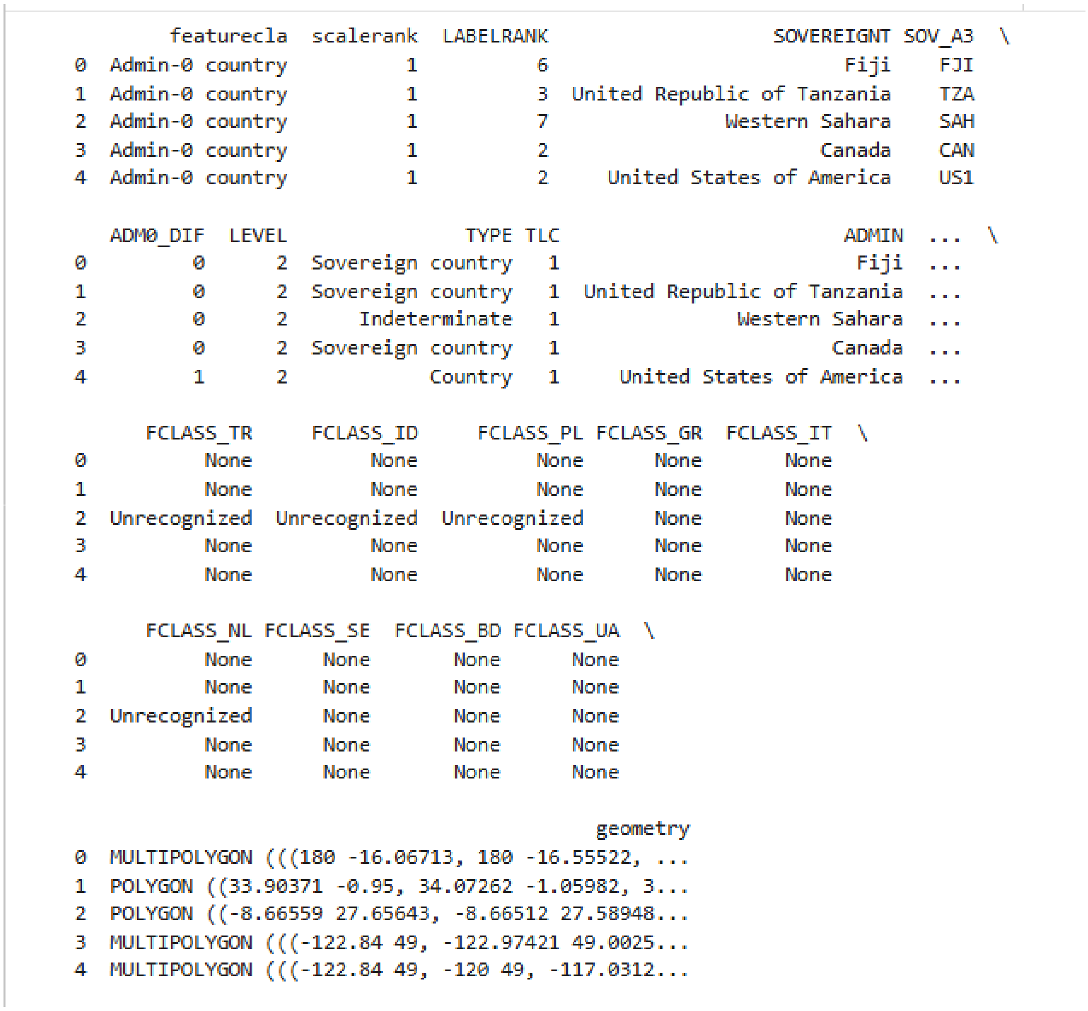
### Task 1: Import Required Libraries



```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import sklearn
import geopandas as gpd
import folium
```

### Task 2: Load NAtural Earth Dataset

1. How many columns does the dataset contain?
   - The dataset contains 169 columns. This information is directly visible in the output of world.head(), which displays [5 rows x 169 columns]. In a GeoDataFrame, columns typically represent attributes or properties of the geographical features, such as names, population, or other relevant data points for each country.
2. What type of geometries and included?
   - The geometry column primarily includes multipolygon and polygon types. A polygon represents a single, enclosed geographical area, like a country without any disconnected parts. A multipolygon is a collection of one or more polygon objects, used to represent countries that consist of multiple separate landmasses or islands.
3. What does the geometry column represent?
   - The geometry column is the core spatial component of a GeoDataFrame. It represents the geographical shapes and boundaries of each country. This column is crucial because it contains the actual spatial data (points, lines, or polygons) that allows for mapping, spatial analysis, and calculations such as area, perimeter, or relationships between different geographical features.



```
      featurecla  scalerank  LABELRANK                    SOVEREIGNT SOV_A3  \
0  Admin-0 country          1          6                          Fiji    FJI
1  Admin-0 country          1          3   United Republic of Tanzania    TZA
2  Admin-0 country          1          7                Western Sahara    SAH
3  Admin-0 country          1          2                        Canada    CAN
4  Admin-0 country          1          2      United States of America    US1

   ADM0_DIF  LEVEL               TYPE TLC                        ADMIN  ...  \
0         0      2  Sovereign country   1                         Fiji  ...
1         0      2  Sovereign country   1  United Republic of Tanzania  ...
2         0      2      Indeterminate   1               Western Sahara  ...
3         0      2  Sovereign country   1                       Canada  ...
4         1      2            Country   1     United States of America  ...

      FCLASS_TR     FCLASS_ID     FCLASS_PL FCLASS_GR FCLASS_IT  \
0          None          None          None      None      None
1          None          None          None      None      None
2  Unrecognized  Unrecognized  Unrecognized      None      None
3          None          None          None      None      None
4          None          None          None      None      None

      FCLASS_NL FCLASS_SE  FCLASS_BD FCLASS_UA  \
0          None      None       None      None
1          None      None       None      None
2  Unrecognized      None       None      None
3          None      None       None      None
4          None      None       None      None

                                            geometry
0  MULTIPOLYGON (((180 -16.06713, 180 -16.55522, ...
1  POLYGON ((33.90371 -0.95, 34.07262 -1.05982, 3...
2  POLYGON ((-8.66559 27.65643, -8.66512 27.58948...
3  MULTIPOLYGON (((-122.84 49, -122.97421 49.0025...
4  MULTIPOLYGON (((-122.84 49, -120 49, -117.0312...
```

-

**Task 3: Check the Coordinate Reference System**

1. What does EPSG:4326 represent?
    - EPSG:4326 represents the World Geodetic System 1984 (WGS84), which is a geographic coordinate system. It uses latitude and longitude coordinates to define locations on a 3D spherical or ellipsoidal model of the Earth. It's the most common coordinate system used globally, especially for GPS and web mapping applications.
2. Why is CRS important in spatial analysis?
    - A Coordinate Reference System (CRS) is crucial in spatial analysis because it defines how geographic coordinates (like latitude and longitude, or X and Y) relate to real-world locations. Without a defined CRS, spatial data would just be a set of meaningless numbers. CRS ensures:

**Task 4: Convert to Metric CRS for Area Calculation**

1. Why can't we compute area accurately using EPSG:4326?
    - You cannot compute area accurately using EPSG:4326 because it is a geographic (unprojected) coordinate system. Its units are degrees (latitude and longitude), not linear units like meters or kilometers. On a curved surface like the Earth, the length of a degree of longitude varies with latitude, and similarly, the area represented by one square

degree changes significantly across the globe. Therefore, calculations like area and distance performed directly on unprojected coordinates will be highly inaccurate, especially over large regions.

2. What unit is EPSG:3857 based on?

- EPSG:3857, also known as WGS 84 / Pseudo-Mercator or Web Mercator, is a projected coordinate system. It is based on a meter unit. This projection is widely used in web mapping services (like Google Maps, OpenStreetMap) because it preserves angles and shapes locally, making it suitable for navigation and display, although it distorts areas significantly as you move away from the equator.

**Task 5: Extract Centroid Coordinates**

1. What is a centroid?

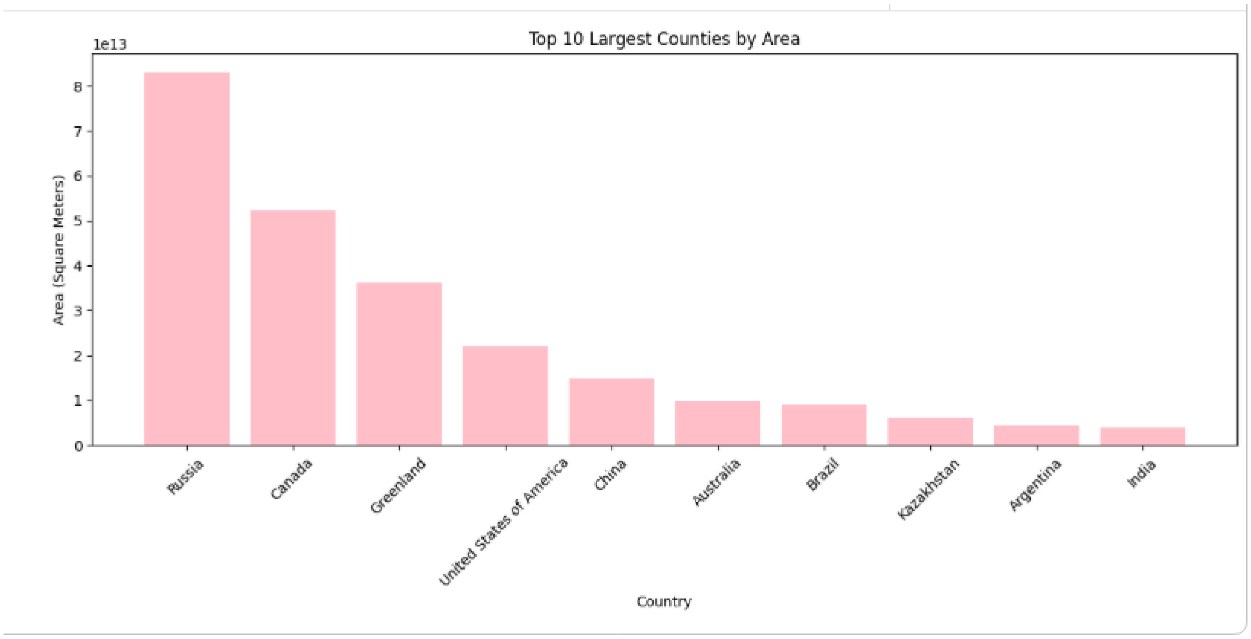
- In the context of a geographic feature (like a country polygon), a centroid is the geometric center of that feature. It's essentially the average of all the coordinates that define the shape. For a simple shape, it might be intuitively in the middle; for complex shapes or multipolygons, it represents a single, representative point location for the entire feature.

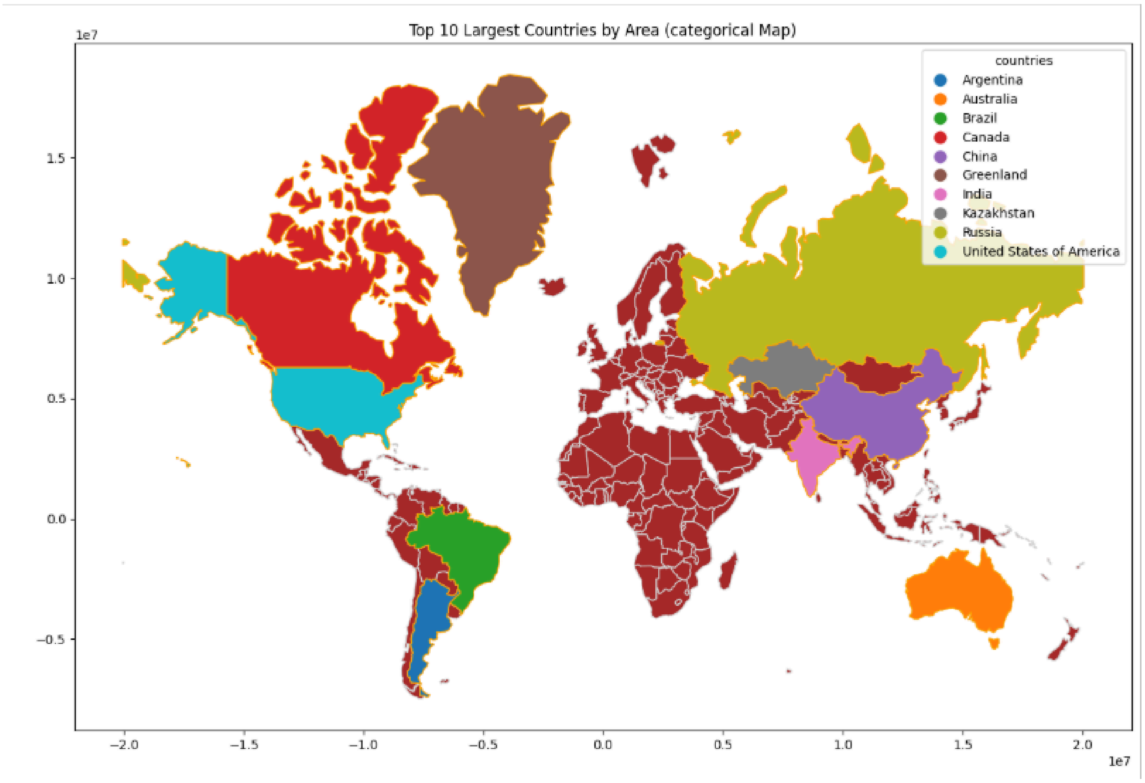
2. What can centroid coordinates be used in clustering?

- Centroid coordinates are highly valuable in clustering for several reasons:

- They provide a single, representative point for each geographic feature, simplifying the data for clustering algorithms.
- Clustering algorithms (like K-means) often rely on distance calculations between points. Centroids allow you to group geographic entities (e.g., countries, cities) based on their spatial proximity.

### Task 5: Plot the top 10 largest countries by area





## Reflection

What difficulties did you encounter?

- One significant difficulty encountered was understanding and correctly handling Coordinate Reference Systems (CRS). Initially, calculating the area of countries directly from the 'world' GeoDataFrame (which used EPSG:4326) would have resulted in inaccurate measurements because EPSG:4326 is a geographic CRS based on degrees, not a projected CRS suitable for area calculations. Forgetting to reproject to a metric CRS like EPSG:3857 would lead to errors in spatial analysis where units are critical, such as area calculation.

What did you learn about spatial data?

- I learned the critical importance of Coordinate Reference Systems (CRS) in spatial data analysis. Different CRS are suitable for different purposes; geographic CRS (like EPSG:4326) are best for representing locations on the Earth's surface using latitude and longitude, while projected CRS (like EPSG:3857) are essential for accurate measurements like area, distance, or buffering, as they maintain constant unit lengths across the map. Reprojection is a fundamental step to ensure the validity of geometric operations and quantitative analyses on spatial data. I also learned about working with GeoDataFrames, handling geometry columns, and extracting properties like centroids.