Clustering African Capitals

Mohamed Abu Bakr

July-2019

1. Introduction :

1.1 Background:

Africa is the world's second largest and second most-populous continent, being behind Asia in both categories. At about 30.3 million km2 (11.7 million square miles) including adjacent islands, it covers 6% of Earth's total surface area and 20% of its land area. With 1.2 billion people as of 2016, it accounts for about 16% of the world's human population.

The continent is surrounded by the Mediterranean Sea to the north, the Isthmus of Suez and the Red Sea to the northeast, the Indian Ocean to the southeast and the Atlantic Ocean to the west. The continent includes Madagascar and various archipelagos. It contains 54 fully recognized sovereign states (countries), nine territories and two de facto independent states with limited or no recognition. The majority of the continent and its countries are in the Northern Hemisphere, with a substantial portion and number of countries in the Southern Hemisphere.

The African Union (AU) is a 55-member federation consisting of all of Africa's states. The union was formed, with Addis Ababa, Ethiopia, as its headquarters, on 26 June 2001. The union was officially established on 9 July 2002 as a successor to the Organization of African Unity (OAU).

1.2 Problem:

Considering diversity of African countries, this study aims to **cluster African capitals into three groups**. Each group will combine capitals with similar economic, demographical, and venues features.

1.3 <u>Interest:</u>

The African Union as a continental foundation with a lot of funding and construction plans on hand , would be highly interested in such study which may help greatly in resources allocations and fair funds distribution between countries based on current condition for each group of countries.

Also any investors aiming to start business in Africa, would make use of domestic, and demographic measures provided by study.

2. Data Acquisition and cleaning:

2.1 <u>Data Requirements:</u>

A lot of features could be used to cluster African capitals. This would include social, economic, geographic, and demographical features. Unfortunately not all of these data available online. Therefore, our clustering factors will include GDP, Population density, and average life in years of individuals. In addition to, different venues (hotels, stations, stores, etc.) located in the capital area.

Each cluster would combine African capitals with obvious similarity of the above mentioned features.

2.2 <u>Data Sources:</u>

Demographic measures including population density, average birth rate …etc., were scrapped from following webpage (http://doheth.co.uk/info/countries-of-the-world.php).

Location data (Latitude, and longitude) was scrapped from following web-page (http://techslides.com/list-of-countries-and-capitals).

Then different venues (including hotels, stations, shops, etc.) in each capital as a measure of how facilities exist and distributed will be obtained from foursquare platform.

Clustering will be limited to capitals where all above data is available (some capitals not covered by above online sources).