Segmenting and Grouping Neighborhood Vulnerable to Ebola Virus Diseases

1. Introduction

This Section present a description of the problem and a discussion of the background.

1.1 Background

According to World Health Organization (WHO, 10th February 2020) news on her website:

- Ebola virus disease (EVD), formerly known as Ebola haemorrhagic fever, is a rare but severe, often fatal illness in humans. The virus is transmitted to people from wild animals and spreads in the human population through human-to-human transmission. The average EVD case fatality rate is around 50%. Case fatality rates have varied from 25% to 90% in past outbreaks.
- Relapse-symptomatic illness in someone who has recovered from EVD due to increased replication of the virus in a specific site is a rare event, but has been documented. Reasons for this phenomenon are not yet fully understood. Studies of viral persistence indicate that in a small percentage of survivors, some body fluids may test positive on reverse transcriptase polymerase chain reaction (RT-PCR) testing for Ebola virus for longer than 9 months.

International tourists and other business people are unaware of the real locations of the six species of Ebola Viruses which has been identified as: Zaire, Bundibugyo, Sudan, Taï Forest, Reston and Bombali. Therefore it's of merits to segment and group the location of these Ebola Virus species.

1.2 Problem Statement

Data that can be used to identify, subdivide and group similar or dissimilar species of Ebola Viruses outbreak in countries are Years of outbreak, Country, Ebola Virus Disease Species (EVD), Cases, Deaths, Case fatality, Latitude and Longitude of location so that International tourists and business people are certain about real location of EVD in the world map. This project aimed at segmenting and clustering neighborhood vulnerable to Ebola viruses in the world map.

1.3 Significance of the Study

International main bodies will have to get interested in identifying the locations of each of these species of Ebola viruses.

Academia and other scientists may be motivated to do more research to identify why these Ebola species are common in such location and devise appropriate solutions to mitigate the Ebola outbreak in such neighborhood.

Helps the country to get organizations that are willing to support the victims of Ebola viruses and devise methods of controlling the outbreak of the disease through sensitization of the community.

1.4 Interest

The most target audience are the community, tourists, students, international health workers and other workers in the location. Governmental and non-governmental organization care about this problem of Ebola Virus Disease outbreak in such location.

2. Methodology

This section entails description of the data and how it will be used to solve the problem.

2.1. Data

Data will be scrapped from WHO website: https://www.who.int/news-room/fact-sheets/detail/ebola-virus-disease . The table contains records with the following columns: Years of outbreak, Country, Ebola Virus Disease Species (EVD), Cases, Deaths and Case fatality. However, since this project deals with location data, the researcher will use Foursquare data to locate the latitude and longitude of each country to be used in segmenting and clustering of neighborhood vulnerable to Ebola Virus diseases.

2.2. Approaches

After acquiring data, data cleaning and features selection are to be performed with python.

Thereafter, the researcher will import the required libraries for Geocode and append the Latitude and Longitude columns in the dataframe. Finally, explore the data, use Foursquare to get at least top five countries with serious Ebola virus diseases, cluster using K-means algorithm and visualize the resulting clusters.

3. References

- WHO (10th Feb. 2020), https://www.who.int/news-room/fact-sheets/detail/ebola-virus-disease