# Forced Displacement Prediction Somalia

This is a project for predicting Forced Displacement in Somalia region using Satellite imagery along with other data sources.

## Problem Statement

This project aims to quantify the effect of climate anomalies on forced displacement and/or violent conflict through satellite imagery analysis.

## Data Sources

- Satellite imagery : Landsat 8 level 1 imagery : https://earthexplorer.usgs.gov/

- UNHCR-NRC forced displacement data : https://unhcr.github.io/dataviz-somalia-prmn/index.html

- FSNAU Early Warning-Early Action Dashboard : http://dashboard.fsnau.org/

- Conflict data: only violent conflict and fatalities: https://www.acleddata.com/data/

## The Process:

Firstly we scrapped the data from fsnau website mentioned above in the jupyter file named "**scraping_fansu.ipynb**". An account on the same website is required to scrape the data. Used python code to login and scrape the data from desired page into a dictionary and then put into lists for easier operation and data saved into "**fsnau_data.csv**".

Then we used data from "**fsnau_data.csv**" and "**d1.csv**" (d1.csv is data downloaded directly from UNHCR_NRC website mentioned above in the "**full_data**" jupyter file. Wherein we performed some data cleaning and manipulations and used geolocator api to find lat/lon coordinates of regions and added that data to the dataframe, all the changes were then saved into "**all_region_data.csv**"

Then in "**hargeysa_data.ipynb**", EDA was performed and a particular region Hargeysa was chosen randomly to further analyse the data, so as to work on small data first before moving onto complete data. "**hargeysa.csv**" is the data created for same.

Then we experimented, visualised the data and built a baseline RNN model on the "hargeysa_data.csv" to see if can come up with some results but it was not successful.

We then worked on satellite images. Since Landsat Images were rotated, we wrote python code to transform the image and calcite the NDVI values from it. It is done in "***Image Transform & NDVI***" jupyter notebook. This was done on random sample Images.

A script was then written "***usgs.py***" which was uploaded on server, that connected to earthexplorer website to automate the downloading of Images of all regions one by one firstly into the server which was then uploaded to s3 storage.

All extracted and built data was then used in "***displacement_pred_usingLR***" jupyter notebook, EDA was performed and data was fed into Linear Regression Model which gave pretty good displacement predictions with score of `0.08823723862729238` (mean square error)