# Group Members

Lewis Gitonga Muriungi SCT222-0123/2020

Jemimah Asiko SCT-0204/2020

Ron Mbatia SCT222-0205/2020

Ted Wainoga SCT222-0247/2020

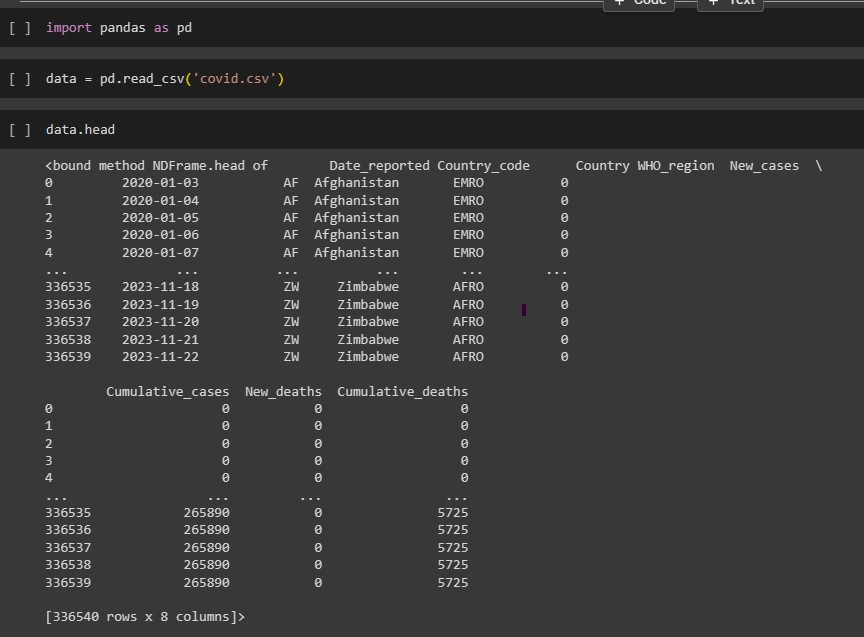
# Flow

## Getting Data

We scoured the internet to find a good dataset that fit the following criteria:

1. Had Kenyan data
2. Had death cases
3. Had confirmed cases
4. Had recovery cases – ideally

The data we found was <https://covid19.who.int/WHO-COVID-19-global-data.csv> from World Health Organization



## Pre-Processing the Data

The library used here was Pandas

The first step was to read the data

We explored the dataset

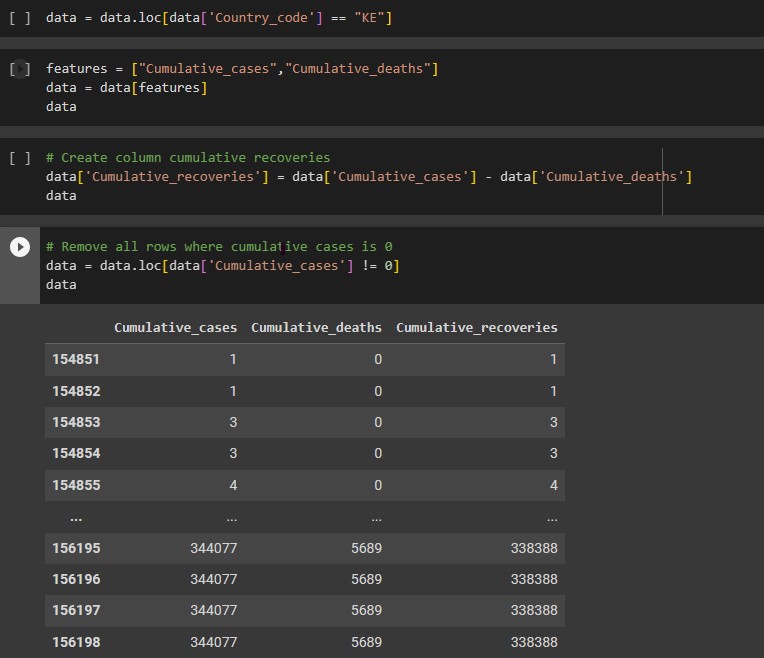
We filtered the data to include only Kenyan data

We chose the columns we would use

We added the recoveries column by subtracting deaths from cases

We filtered to remove 0s from the cases column

Lastly, we renamed the columns for easier use



## Model training and visualization

The library involved was sckit-learn

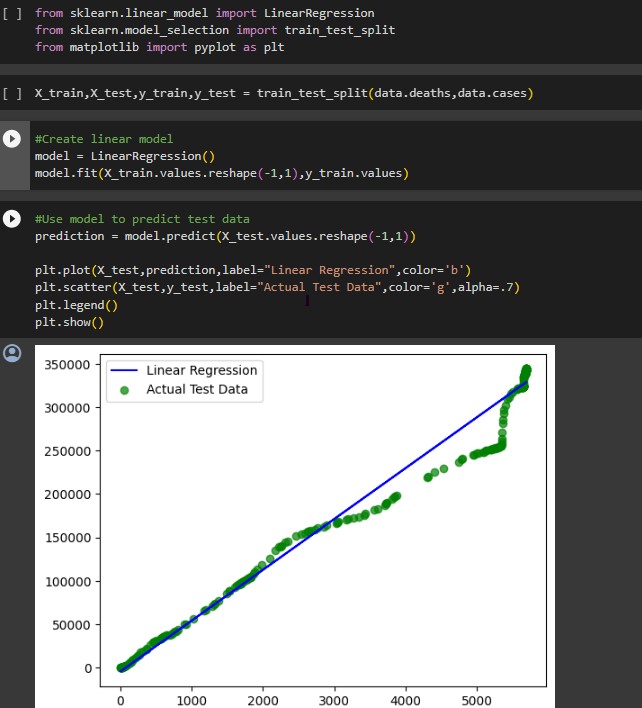
To be more specific LinearRegression and train\_test\_split

We divided the data into train\_data and test\_data using train\_test\_split

We fit the model, using the train\_data

We used the model to predict the test\_data

Visualizing the model’s prediction alongside the test data



## Data interpretation

The model was pretty on spot roughly 80%.

We can deduce that when cases hit above 150,00 deaths sky rocketed above the prediction. This is a speculation of as cases rise the pool of infection rate also rises hence leading to more deaths

This may be due to constraints such as limited resources to combat high number of viral cases thus leading to more deaths

## Data Applications

Outbreaks – For example the Ebola outbreak in West Africa

This is because the model is used to predict high population of disease cases and deaths. Thus health bodies can adequately plan on how to accommodate them

Example: Gather adequate resources, like provision of masks during Covid-19 outbreak