BCU COVID 19 Hackathon

AIHackTeam4

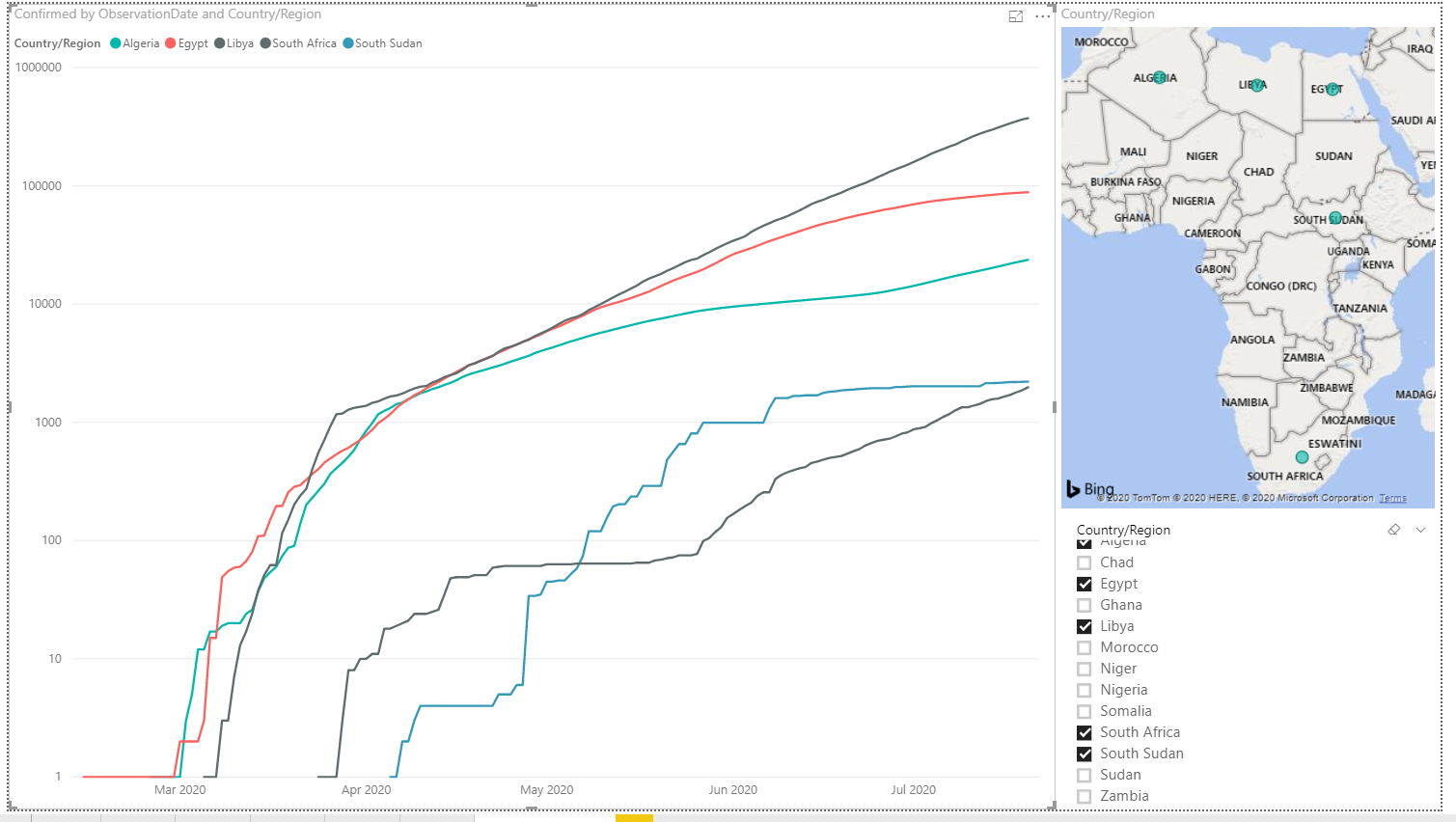
Zakrea Almansouri

Challenge 2

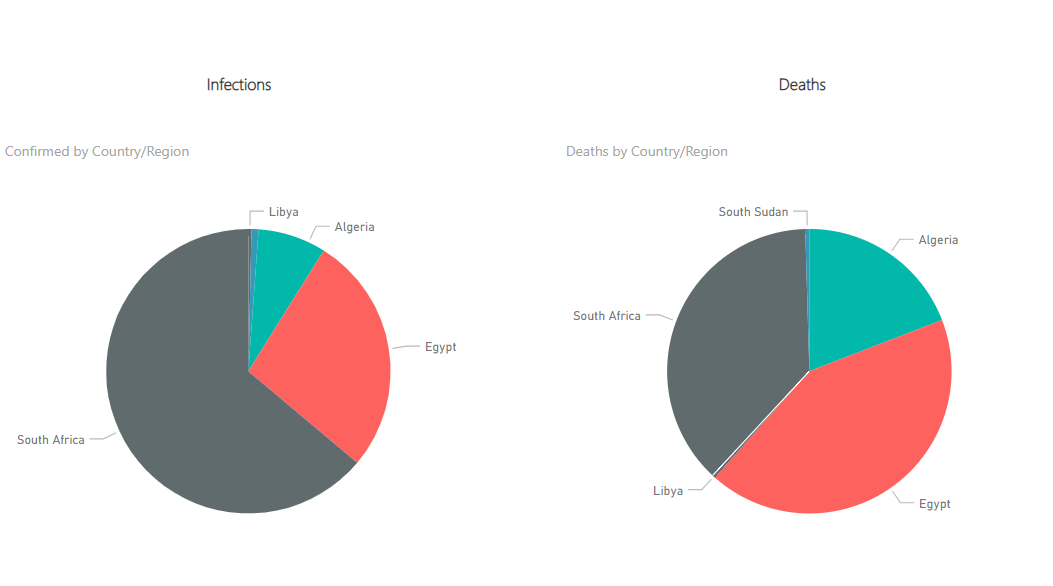
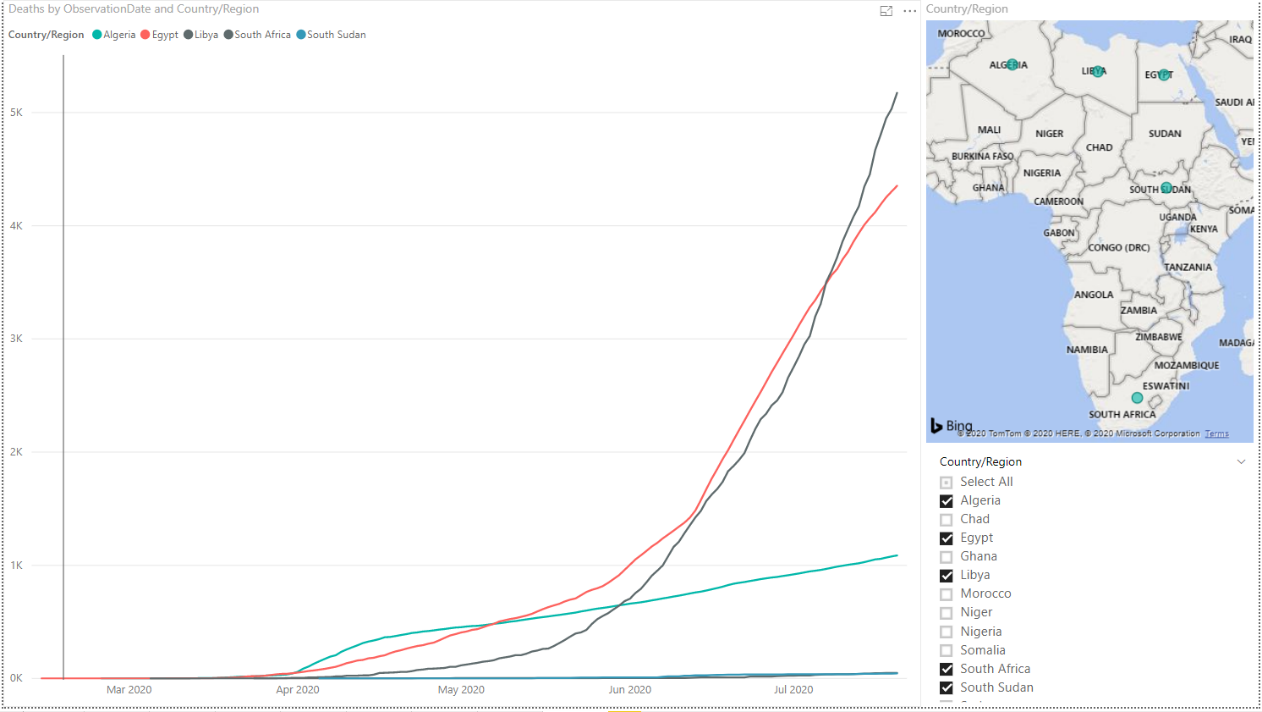
Challenge definition:

Predicting: Effective resource planning and strategising relies on knowledge of the velocity and veracity of viral spread. Subtle differences between nations have been evident in the changes between the death rate rises and falls over time allowing for targeted approaches to maximise lives saved. Effective planning is most important where resources are naturally the scarcest. **Predict the future cases and death rates for 5 countries in Africa**. Create a visualisation and a short blog explaining your modelling steps and results.

Solution:

We have chosen challenge 2 as our task and picked these 5 countries to perform the prediction on : ['South Africa' , 'Egypt' , 'South Sudan' , 'Libya' , 'Algeria']. These provided a nice variance in the data as some of these countries have higher levels of covid then others and are at different stages of the covid cycle. Here are a few of the graphs we first looked at to make this decision.

Graph showing cases for these countries in log scale

Pie chart showing split of deaths and cases between the countries. 

Graph showing deaths for these countries

The I decided to use a clustering algorithm to try and get countries that where similar to each of our chosen countries so that we can then build a model which has access to more data. after making the model and tweaking it I found that it was not the best avenue to go down as it required a lot of sophisticated pre-processing and modelling to complete and as this was only a side step we simply did have the time so it was scraped. The useful thing out of that was I gained a far more in depth understanding of the data and it made the next step easier and quicker 😋

So, after moving from the clustering idea I went on to tackling the main problem directly so using modelling to predict the future cases but this time by just using the single curve for the country. This made it that making long term predictions was going to be practically impossible but that is fine as this was never my intention. So, we ended making a predication of 14 days into the future. The interesting thing was how much data should i use from the past to make this prediction, now an obvious answer would be all the data i had but that doesn’t work as i need that to be a constant number, so i decided to use a 2 week time span which turned out to be quite good.

So after reshaping the data into a single row per 14 day time span prediction I wanted to create a model for the prediction so I first went with random forest regressor (a bit lazy as that doesn’t need scaling ;P ) and that failed miserably and then I decided to scale the data. The problem with scaling the data is that i want it to be between 0 and 1 and if I simply scale on the training data then the highest the model could predict would be the highest value in the training data and because our data is mostly increasing i was going to face a problem 🙁 so instead I decided to add a new record and make it around 10x-15x larger than the largest training value and that worked better then I thought 😁 .

So then i tried a different model with this new scaled data and that is the SVR (based on svm) and then failed just as bad as Random forest regressor

..... 🙁

Finally, I decided to bring out the heavy tools 😋 and use a deep neural network. so i used the same single row per 14-day prediction and scaled data. The neural network has 100 nodes in the first layer with relu activation function and 42 inputs, followed by 2 hidden layers with 100 nodes and relu , and finally an output layer with 1 node and linear activation function. loss function was mainly mean squared error and Adam as the optimizer, using a batch size of 3 i ran the model….

The results we are actually very very good. i was getting mse values as low as x10^-7 and the graphs we very nice 😁. this was just for case confirmed. So, I had to do all of that over again but just for deaths. I actually tried to make the model output both deaths and cases but suffered significantly on the accuracy so it was a no go. in general, the neural network found it more difficult to predict deaths then cases but I got some good results there too with a bit of a tweak in the nn architecture. Now after doing all this for one country (South Africa) i need to these FOUR more times for the other countries. So now I have 10 neural networks all with slightly different models but in general can do quite well

Here are the 10 graphs each representing the models predicting the cases for deaths and cases separately for each country.

As you can see the results speak for them selves with very strong predicting power and if these kinds of tools are used by these counties, I believe they will greatly improve their performance in dealing with Covid-19

I would like to Thank BCU for this excellent hackathon and look forward to future ones.