STEM Fellowship Big Data Challenge 2020

Correlation between Covid-19 cases in African Countries and flights

Akrash Sharma*, Tanyaradzwa Gozhora†, and Xiang Zhang‡

University of Alberta

June 1, 2020

Abstract

The growth of the global COVID-19 pandemic has been relatively slow in many African countries, potentially due to lesser international travel to major cities compared to other continents.The goal of this paper is to determine the extent and type of the relation between international flights and cases of COVID-19 observed currently in Africa. The observations could show where the focus of countries should be between initiating lock downs and imposing border restrictions. This analysis was done through clustering and linear regression modelling for the clusters. Significant correlation was observed during the analysis.

Keywords
COVID-19, Coronavirus, Africa, Flights, Policy

1 Introduction

As of 11 March 2020, WHO had officially declared the corona-virus as a pandemic, noting the concerning rate of proliferation of the disease globally[1]. However Africa seemed to be an exception then and still seems to be one. A general consensus is that the virus is not yet in full swing in Africa, with some countries only obtaining their first case on the 13th of May[2], as the cases in Africa significantly lower than elsewhere. Predictions for the continent, however pertain that caution should be observed as WHO has stated that 190'000 people in Africa could die of the virus if respective governments do not exercise heed[3].

A fairly more optimistic viewpoint has been presented, in that unlike on other continents, in Africa the virus is likely to stay mainly within transmission hot-spots [3]. Moreover, we can also note most cases in Africa are in countries which are transportation hubs, as we would expect. This raises the question of how weighty is the relation between incoming flights in these countries and cases. Especially, given that many African countries have imposed strict lock down restrictions on their citizens and are contemplating their economic cost[4]. In addition, one also wonders how does this fare with other factors such as population density, median age, extreme poverty index and others. To analyse the effects of the different variables we used a heat map, to show if any relation lies with the other factors we found commonly used in COVID-19 research [5]. Factors such as poverty index, GDP per ca pita, hospital beds per 100, hand-washing index and others[5].

Thereafter we made our regression models for the randomly chosen countries, following which came some residual analysis to assess our models. Anticipating that the existence of a strong linear correlation, which was then found, may discourage the removal of the border restrictions of these countries.

2 Materials & Methods

Firstly using the python sea-born library, a heat map was made of all the factors so as to visually gauge average correlation of COVID-19 cases and factors commonly thought to have an influence such as tests taken, poverty index and so forth across Africa description of figure (Fig. 1)[5] . Having collected data for corona virus cases, data was collected for the flights coming into each of 10 randomly selected countries in Africa[6]. For each of these countries data was collected from every airports in the nation on all international arrivals for a period of a month

*akrash@ualberta.ca
†gozhora@ualberta.ca
‡xzhang23@ualberta.ca

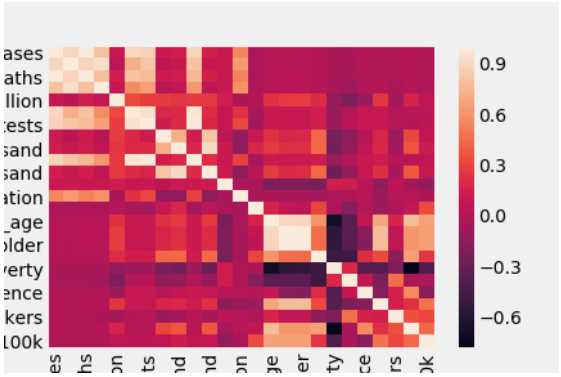

Figure 1: Correlation between other factors visually displayed

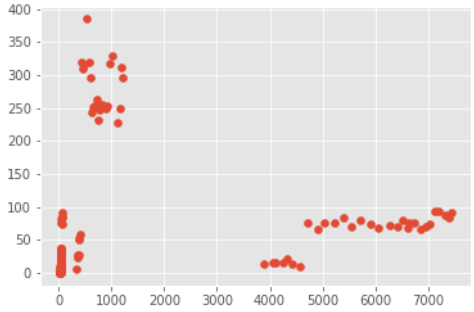

Figure 2: Clusters initial countries fell in

(26/04/2018 - 25/04/2018)[3] . Thereafter having plotted our scatter graph it became evident the two clusters the data fell in (Fig 2).

After which linear regressions were calculated for the two countries in each of these two groups in python using pandas and the plot-lib library. To then confirm validity of our models, we then got a standard residual for our regression models.

3 Results

Initially the heat map showed insignificant relations between most factors except between deaths and cases, as expected "description of figure"(Fig 1). The linear regressions, yielded strong correlations. For the four random countries from the two clusters "description of figure"(Fig 2) an R^2 of 0.9562 was found for Kenya "description of figure"(Fig 3) and for Morocco "description of figure"(Fig 4)we had R^2 of 0.9420. We had R^2 of 0.9538 for SA "description of figure"(Fig 5). Also an R^2 of 0.8337 for Angola "description of figure"(Fig 6). Our residual analysis had a standard deviation of 0.983 "description of figure"(Fig 7).

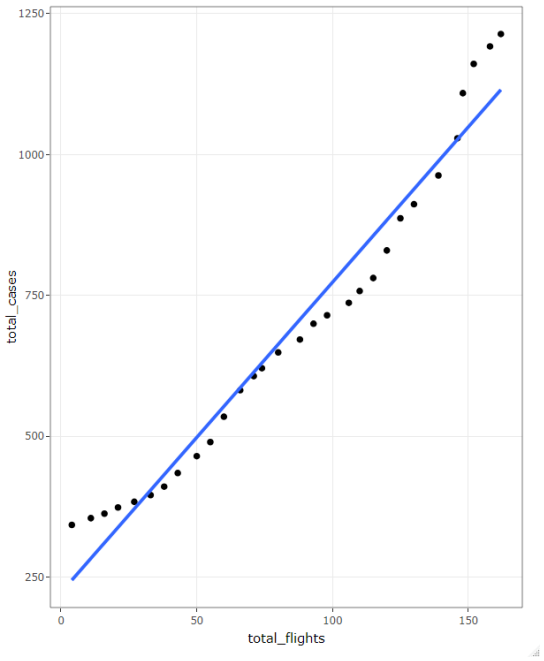

Figure 3: Correlation between Kenyan cumulative cases and flights.

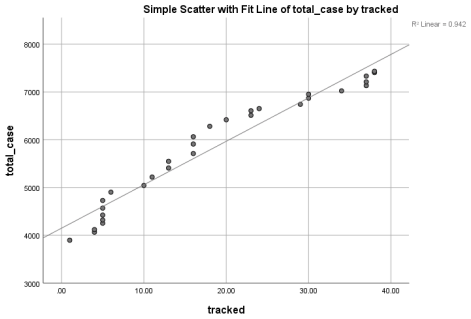

Figure 4: Correlation between Moroccan cumulative cases and flights.

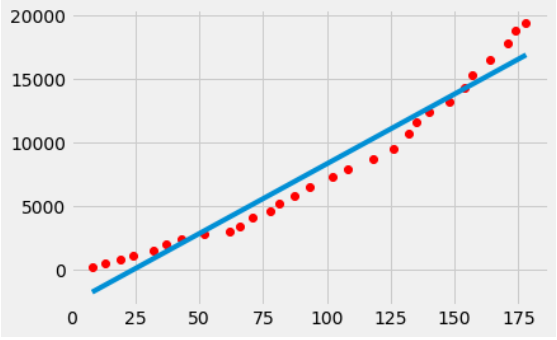

Figure 5: Correlation between South African cumulative cases and flights

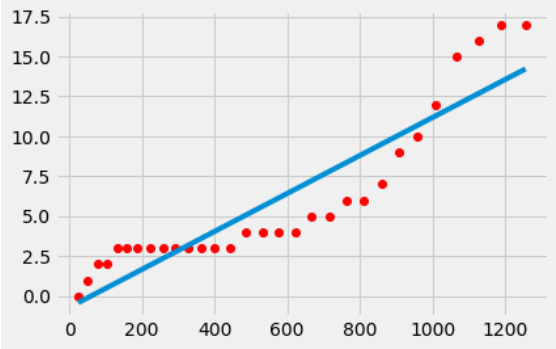

Figure 6: Correlation between Angolan cumulative cases and flights

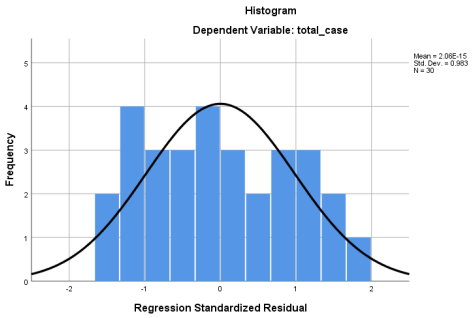

Figure 7: Standardized residual analysis of the Moroccan regression

4 Discussion

Our initial thought that flights do have a strong correlation with cases at the relatively early stage that countries in Africa seem to be in, does seem hold up to our analysis. With regard to the statistical tools used, a linear regression after a scatter plots showed it unlikely to be a quadratic relation. Higher order and trigonometric relations were avoided to prevent over-fitting. This was further justified this through our standardized residual in our regression analysis (fig 7).

As for potential sources of error. Firstly, we have that African countries have a lot of movement on the ground (car and by foot) so it is plausible although highly unlikely, in this time, movement across multiple countries could occur on the ground and note be accounted for and thus our countries although distant could have movement from one to the other.Secondly, with a least squares linear regression we have an assumption that our independent variable have been measured with no error, and our source [6] does not promise that results are error free as they are subject to change within a month of recording. Also, all the data accumulated had it's limitations. For example, the data on flights could only be acquired for a period of up to a month prior and the COVID-19 cases data differed in that testing began on different date for each country. Hence the peculiar choice of timeline (26/04 - 25/05). Finally, an important source of error is that of under testing which has been commonly critiqued by the international community towards most but not all African countries.

5 Conclusions

Our findings do suggest that African countries should maintain the border restrictions they've placed on international travels for the time being. As across the continent international arrival flight seem to be greatly correlated to the cases in their respective countries. However, much more research is necessary before any conclusive statements are made. It would also be wise to look at factors that influence the two (flights and cases). For example, flights seem to spike every Friday even in quarantine, and social events also generally have a recurring trend with a peak on Fridays despite quarantine measures, wedding, funerals, celebrations and the like. Hence one cannot be too conclusive

To be kept in mind is the fact that most African countries have a large informal sector, i.e employment not of a formal nature, such as street vendors and the like, 67.3 per cent in Northern Africa and 89.2 per cent in sub-Saharan Africa, which has been greatly hampered by lock-downs within their countries[7]. As such, further research and guidance for African states, on the possible severity of their actions is needed. As most African governments are torn between the economic well being of their nations and the health of their people. According more conclusive research on the factors greatly influencing spread on the continent would be better now than later. Consequently leading to wiser decisions, in these countries which cannot afford to be in quarantine.

Acknowledgements

Mike Dorrington

References

[1] World Health Organization. WHO timeline-Covid-19. https://www.who.int/news-room/detail/27-04-2020-who-timeline---covid-19, 2020. Online; accessed 26 May 2020.

[2] World Health Organization. WHO African Region Covid-19. https:

//apps.who.int/iris/bitstream/handle/10665/332199/SITREP_COVID-19_WHOAFRO_20200527-eng.pdf, 2020. Online; accessed 26 May 2020.

[3] World Health Organization(regional office for Africa). New WHO estimates: Up to 190 000 people could die of COVID-19 in Africa if not controlled. https://www.afro.who.int/news/new-who-estimates-190-000-people-could-die-covid-19-africa-if-not-controlled, 2020. Online; accessed 26 May 2020.

[4] BBC. Coronavirus: How African countries are lifting lockdowns. https://www.bbc.com/news/world-africa-52395976, 2020. Online; accessed 31 May 2020.

[5] Our world in Data. Coronavirus Pandemic. https://ourworldindata.org/coronavirus, 2020. Online; accessed 26 May 2020.

[6] flighttradar24. Airports by Country. https://www.flightradar24.com/data/airports, 2020. Online; accessed 26 May 2020.

[7] International Labour Organization. Impact of Covid-19 on the informal economy. https://www.ilo.org/wcmsp5/groups/public/---africa/---ro-abidjan/documents/briefingnote/wcms_741864.pdf, 2020. Online; accessed 26 May 2020.