**Logo

Description automatically generated with low confidence**

Covid 19 prediction for South African

analyzing and predicting the coronavirus(COVID-19) cases and death cases of South Africa using "Power Bi" for visualizations and linear regression in python for the prediction model.

II. INTRODUCTION

The coronavirus (COVID-19) that was first reported at the end of 2019 has impacted

almost every aspect of life as we know it. Since then, researchers are still working on discovering

the trends about the pandemic and when it will end. In our project, we are enthusiastic about

analyzing and visualizing South Africa provinces' cases, deaths, and recoveries. Also, we are

interested to know more about the prediction cases and deaths in the

country. We will approach this prediction using the linear regression model

III. PROBLEM DESCRIPTION

In this project, we are willing to explore the timeline of cases, deaths, and recoveries in

each province of South Africa. Also, we want to know which group age and gender are most

affected. So that can help the country put more effort to low the cases for the high rate cases or

deaths provinces. In addition, knowing the expected cases and deaths for the future could avoid

such a problem that the country offers more vaccination or takes a step to make the vaccination

mandatory for work or go out in public.

DATASET DESCRIPTION

We are using the Coronavirus (COVID-19) Data Repository for South Africa, created,

maintained, and hosted by the Data Science for Social Impact research group, led by Dr. Vukosi

Marivate, at the University of Pretoria. The data is as accurate as possible to collate the COVID

19 reporting data from NICD and DoH[1]. They update that data almost every day once there is

an official report or statement. We used three CSV files(timeline cumulative confirmed cases,

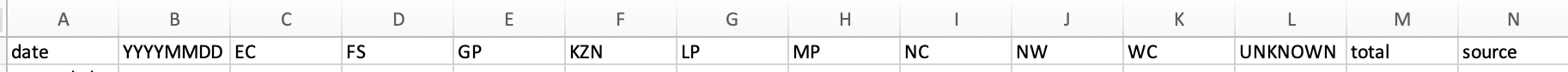
deaths, and recoveries). Each file has a date column from March 2020 up to date. Also, they

contain the cumulative province timeline for each province plus the total. The province's

columns, as Figure-1, are Western Cape, Northern Cape, Eastern Cape, Gauteng,

KwaZulu-Natal, Mpumalanga, Free State, Limpopo, and North-West. Also, for analyzing the

gender and age count, we used another data for 2021.



Graphical user interface, application, table, Excel

Description automatically generated

Figure-1

METHODOLOGY

We used python language on Jupyter notebook and Power Bi business analytics service

by Microsoft for creating a dashboard that can be published to the public. We start off using

python by exploring the data to get a high-level insight of how our data is structured. We

extracted and cleaned the data and then plotted with histograms for visualizations. We dropped

the rows that contained the null values. Then we used a linear regression model using the total as

the target and date as the feature to predict the timeline cases and deaths. Then we extract the

prediction for visualization. After that, we used Power Bi to make the dashboard interface for the

users.

DATA ANALYSIS

We created line charts that display information as a series of data points for each month

(the end of each month). For the confirmed cases, as shown in Figure-2, GP province has the

highest confirmed cases, which are 919636 cases, and NC has the lowest, which is 92513 cases.

For the line chart, we can see all the provinces are stable until December 30, and then they

gradually get high, but GP province cases are suddenly getting very high from May 30, 2021, to

July 30, 2021. Then the cases are stable again till November 2, 2021.

Graphical user interface

Description automatically generated

Figure-2

For the death cases, as shown in Figure-3, there are a total of 89K deaths in all the provinces. The first death recorded was in April 2020. WC has 20147 deaths by November 2, 2021, which is the highest. However, MP has the lowest, which is 2004 deaths. Unlike the confirmed cases, the cases are getting higher gradually then get stable for a while. The deaths start to rise between June and July 2020. EC, GP, WC, and KZN increased dramatically between November 2020 and February 2021. Then they slowed down noticeably and remained steady till around March and April 20201. However, the other provinces have increased gradually without sudden increase.

Graphical user interface, chart

Description automatically generated

Figure-3

For the recovery cases, as shown in Figure-4, there are a total of 2.8M. AS GP gets the most confirmed cases, it also has the most recovered cases. The same with the NC that has the few recovered cases. The exact peak happens for GP that has a massive increase from May to August 2021. KZN has a growth rate of December 2020 and then grown slowly till October 2021.

Graphical user interface, chart

Description automatically generated

Figure-4

Figure-5 shows the age group's count of confirmed cases for a few months (2020 data only). As we can notice that the age group 30-39 are the most infected. Groupage 85 and above are the least infected group.

Graphical user interface

Description automatically generated

Figure-5

Figure-6 shows the Female and male counts. We used the same data for the age group (2020 data). We dropped the “not specified” value to get more insight. We can see males are more infected.

A screenshot of a computer

Description automatically generated

Figure-6

VII. CONCLUSIONS

This analysis was conducted to increase people's awareness to limit the spread of the

disease in the country; half of the solution is to educate the citizens and work with the

government to reduce the number of cases and limit their spread. This analysis showed an

increase in the number of deaths for 29/10/2021 to 10/10/2021, so it is necessary to take the required

vaccinations and limit the gathering of people in events and workplaces to reduce the spread of

this disease so that the government can control the disease in each region to return to normal life

and restore the economy and tourism to the country.

VII.REFERENCES

[1] <https://github.com/dsfsi/covid19za>