COVID-19: relation between cases and vaccination

Team name: Sigma

Team leader: احمد هانى

Team leader email: ahmed.tawfik011@eng-st.cu.edu.eg

Department: computer engineer department

Year: 2nd year (Junior 1)

Name	Section	Bench number	Student code	Email
عبدالرحمن محمد سالم حسن	1	38	9202794	abdosalm555@gmail.com
أاحمد حسني عبد الرازق عبد الغني	1	2	9202077	eng.ahmedhosny2024@gmail.com
أحمد فوزى محمد إبراهيم	1	7	9202153	Ahmed.Ibrahim011@eng-st.cu.edu.eg
احمد هانى فاروق توفيق	1	11	9202213	ahmed.tawfik011@eng-st.cu.edu.eg
احمد سيد سيد مدبولي	1	3	9202111	ahmedmadbouly186@gmail.com
يوسف سعيد ابراهيم ربيع	2	38	9203767	youssef.ibrahim01@eng-st.cu.edu.eg

Abstract

This paper proposes a statistical analysis of two important issues relating to the COVID-19 pandemic, namely: the relation of the vaccinations rate to the new cases rate in multiple countries, and an estimation for the end date of the fourth wave in Egypt. For the first objective, the factors of economic and social variations were neutralized, as much as feasible, through the selection of two dissimilar groups of countries. Inter-group countries had to fall within specified ranges of criteria. As for the second objective, data on the three preceding waves were juxtaposed against the available data for the current wave, and a prediction is made for the end date using linear regression.

The results pertaining to the first objective are inconclusive, where even seemingly homogenous *-culturally, economically, and socially-* showed conflicting behaviors, and to the second objective do make sense, only considering what is expected.

Problem definition

The problems we are solving

We all can't deny the effect of the covid-19 on the whole world. So, the problems we are trying to solve are (when the fourth wave of covid will end in Egypt and the effect of vaccination on the new cases in multiple countries).

when the fourth wave of covid will end in Egypt

Egypt is currently in the fourth wave which started 2 months ago. It affected our lifestyle so we were concerned about when it will end. It's so important as we want to know when our old lifestyles will return and it will help the government to consider the rate of spread of the virus to be able to evaluate the current situation whether it's good or bad.

the effect of vaccination of the new cases

It's essential to know whether the vaccination affects the new cases or not.it is important to be able to determine whether is effective with this country or not. To know if there is a relation between the number of people getting the vaccination and the people getting the disease also helps to determine the efficiency of different types of vaccines.

why it’s important

Its importance comes from the effect it will do as for the first problem, it will help to determine when the fourth wave will end in Egypt, and it may help to determine whether there is a fifth wave or not.

The second problem will help to determine the efficiency of different types of vaccines and determine the suitable one for each country to be able to face the pandemic.

What makes it challenging

There are a lot of challenges that we faced during this project. From collecting the data to analyzing it to prediction to the hypothesis.

If we took it one by one. So, firstly we will talk about the data collection and how it was challenging to collect the data and to choose a trusty source to get the data form.

The second was analyzing the data to determine the suitable methods to analyze it to be able to make the excel dashboard.

The third was prediction to be able to put a suitable criterion and choose a suitable algorithm for prediction and choosing the language we used to implement the code for the prediction.

And last the hypotheses to determine the old assumption we want to reject.

Data Description

Data used to test our methods against are the countries' official reports of new cases, and vaccination. Luckily, such data was readily available at Kaggle which was our primary source. Following is the specification of what data is utilized for our purposes.

Vaccinations and new cases

The constraining criteria

First, the constraining criteria, briefly mentioned in the abstract, governing the selection of inter-group countries are stringency index[*], aged_65_older, and hospital beds per thousand. We are somewhat aware that this may not be a comprehensive bias neutralization, but it should help significantly.

- **The First Group**

  The reference is taken to be USA, with stringency index around 60, aged_65_older around 15, and hospital beds per thousand per 2.5.

  So, the first group will consist of the following countries: USA, Australia, Netherlands, Italy, Canada, UK, Belgium, France, and Germany.

  Aside to this, vaccines are also somewhat similar. Only Johnson & Johnson, Pfizer, AstraZeneca, and Moderna were prevalently used. This is shown in the following table

Country	Vaccines
USA	Johnson & Johnson, Moderna, Pfizer/BioNTech
Australia	Moderna, Oxford/AstraZeneca, Pfizer/BioNTech
Netherlands	Johnson & Johnson, Moderna, Oxford/AstraZeneca, Pfizer/BioNTech
Italy	Johnson & Johnson, Moderna, Oxford/AstraZeneca, Pfizer/BioNTech
Canada	Moderna, Oxford/AstraZeneca, Pfizer/BioNTech
UK	Moderna, Oxford/AstraZeneca, Pfizer/BioNTech
Belgium	Johnson & Johnson, Moderna, Oxford/AstraZeneca, Pfizer/BioNTech
France	Johnson & Johnson, Moderna, Oxford/AstraZeneca, Pfizer/BioNTech
Germany	Johnson & Johnson, Moderna, Oxford/AstraZeneca, Pfizer/BioNTech

- **The Second Group**

  The reference is taken to be Egypt, with stringency index around 50, aged_65_older around 5, and hospital beds per thousand around 1.5

  The following countries were selected for this group: Egypt, Cape Verde, Djibouti, Guatemala, Paraguay, Guyana, Iran, and Philippines.

  Here, we find additional different vaccines than the first group, including: Sinovac, Sputnik V, Sinopharm, and less prevalently Covaxin. This is shown in the following table.

Country	Vaccines
Egypt	Johnson & Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing, Sinovac, Sputnik V
Cape Verde	Moderna, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing
Djibouti	Johnson & Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing, Sinovac, Sputnik V
Guatemala	Moderna, Oxford/AstraZeneca
Paraguay	Covaxin, Moderna, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing, Sinovac, Sputnik V
Guyana	Oxford/AstraZeneca, Sputnik V
Iran	COVIran Barekat, Covaxin, Oxford/AstraZeneca, Sinopharm/Beijing, Soberana02, Sputnik V
Philippines	Johnson & Johnson, Moderna, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing, Sinovac, Sputnik V

If you could observe, the countries

- (Netherlands, Italy) takes the same kinds of vaccines
- (Canada, UK) takes the same kinds of vaccines
- (Belgium, France, Germany) takes the same kinds of vaccines
- (Egypt, Djibouti) takes the same kinds of vaccines

So, we should see same results among each group.

The data used to plot the relationships

- Reported numbers:
  - Daily new cases.
  - Daily number of total people fully vaccinated (always increasing because it's accumulating).
  - Daily number of new people fully vaccinated.
- Rates:
  - Rate of new cases per month.
  - Rate of people fully vaccinated per month.
  - The rate of increase in people fully vaccinated.

Graphs

- Clustered column graph: actual number of new cases per day.
- Clustered column graph: actual number of total people fully vaccinated per day.

- Line graph: rate of new cases per month with rate of people fully vaccinated per month.
- Dot graph: rate of new cases per day with rate of people fully vaccinated per day.

- 100% Stacked area graph: actual number of new cases per day, actual number of new people fully vaccinated per day (100% truncation).
- Stacked area graph: actual number of new cases per day, actual number of new people fully vaccinated per day (full image)

Indicators

- A cell to represent the relation between the last day new cases and the average of new cases of all the time
- A cell to represent correlation coefficient between rate of new cases per day and rate of people fully vaccinated per day
- A cell to represent the average of new cases in 2020.
- A cell to represent the average of new cases in 2021.
- A hypothesis Z-Table test to indicate how much we are confident that the vaccination affected the rate of the new cases.

The fourth wave in Egypt

The data used to plot the relationships

- Reported numbers of cases throughout the three previous waves were utilized. A prediction is made using linear regression on the progression of the current fourth wave.

From the graph we made by linear regression on the progression of the current fourth wave on Egypt, we predicted that the peak date of the 4^{th} wave is : 29/11/2021 and the 4^{th} wave will end at date : 13/2/2022.

Methods of Solution

In the statistics part we needed to make prediction about the fourth wave of the covid19 in Egypt.

the algorithm that we used to make the prediction is the polynomial regression.

we used the polynomial regression features from library SKlearn using python.

How Polynomial Regression works?

Polynomial Regression tries to protect the relationship between dependent variable y according to another independent variable (or vector of variables) x using Previously collected valid data set and using this data set to find the parameters of the relationship between x and y.

the first step is to find the degree of the Polynomial relation between x and y and put a parameter as unknowns then using the data set, we must calculate this parameter.

so, the general polynomial relation of degree (n) has the formula:

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0,$$

the Polynomial Regression method help us to find a approx. value to each parameter (a_n, a_{n-1}.... a_1, a_0)

using at least n+1 points of the data and of course the larger data is the more accurate values of parameters (a_n, a_{n-1}.... a_1, a_0).

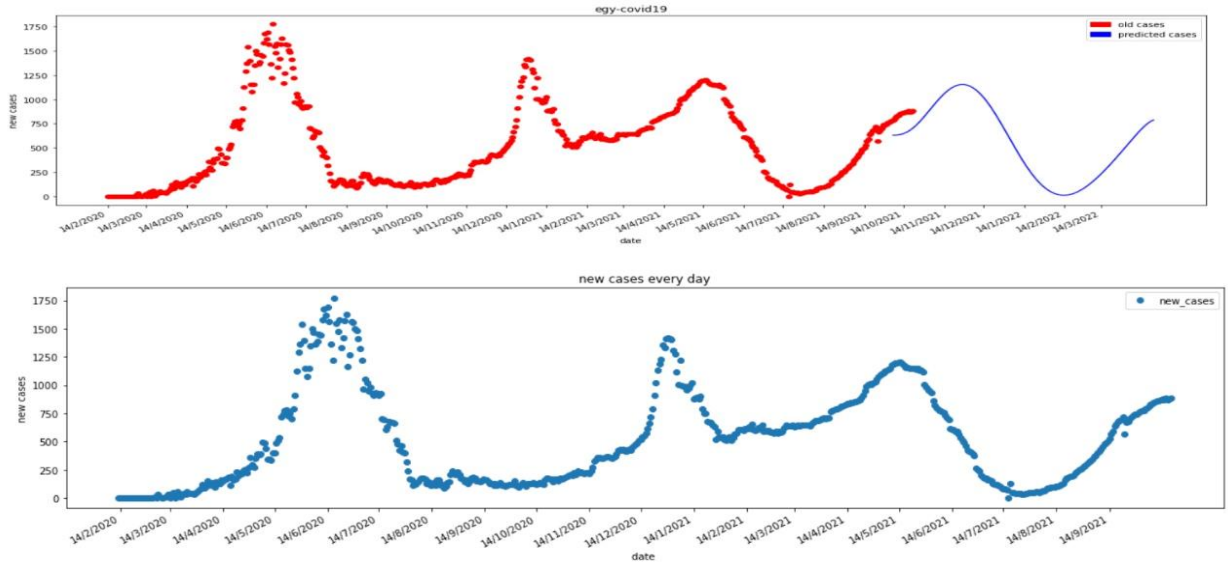
so, the polynomial regression at general using the next formula to expect Relation between y, x:

$$\begin{pmatrix} n & \sum x_i & \sum x_i^2 & \cdots & \sum x_i^n \\ \sum x_i & \sum x_i^2 & \sum x_i^3 & \cdots & \sum x_i^{n+1} \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 & \cdots & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ \sum x_i^n & \sum x_i^{n+1} & \sum x_i^{n+2} & \cdots & \sum x_i^{2n} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \\ \vdots \\ \vdots \\ \sum x_i^n y_i \end{pmatrix}$$

How did we make the expectation?

we use the polynomial regression features from library SKlearn using python but when we work with the data, that we get from WHO we have a problem to solve:

- First is the *how* we should put some meaningful indicating and predicting data in relation to the number of new cases. And thus it is solved: what's predicted is the number of cases, plotted at the y axis, in relation to the day's sequence in the whole period of the wave, plotted at the x-axis. Days are sequenced in each wave from 1 till its length.
- Of course, each wave should be an independent dataset of its own, to give the rise and descend of the new cases in each single one of them some meaning. The 600 days are then divided by the three waves, each of an approximate length of 200.



Analysis of Results

We observed the following from graphing the rate of new cases with the rate of increase in vaccinations.

For the first group:

- **USA**: Inversely proportional through first 3 months then directly proportional.
- **Australia**: Directly proportional all the way down
- **Netherlands**: No relation observed.
- **Italy:** Inversely proportional all the way down
- **Canada:** Inversely proportional between months 6 to 8 in 2021, other than that no relation.
- **UK:** Inversely proportional between months 1 to 5 in 2021 then directly proportional all the way down
- **Belgium:** No relation but we can say it decreased the peaks of the waves
- **France:** No relation but it decreased the peaks of the waves
- **Germany:** Inversely proportional between months 5 to 8 in 2021, then it decreased the values of the peaks of waves.

For the second group

- **Egypt:** inversely proportional between months 1 to 7 then no relation, also it didn't decrease the peak of the waves
- **Djibouti:** no relation as not enough data
- **Cape Verde:** no relation, but it decreased the peaks of the waves
- **Guatemala:** no relation, also it didn't decrease the peak of the waves
- **Paraguay:** inversely proportional all the way down
- **Guyana:** no relation
- **Iran:** hard to tell but we can say inversely proportional between months 8 to 10
- **Philippines:** no effect, even the peaks of the waves become higher

Conclusion

In some countries, the vaccination doesn't affect new cases rate, but it decreases the peaks of the next waves of COVID-19, while in others, the vaccination doesn't have any effect on the rate of the new cases. Even the effect of a single specific vaccine differs from one country to another like (Netherlands and Italy) where in Italy there was inversely relationship while in Netherlands there was no relation . In Some other countries the vaccination did decrease the rate of new cases, also just a few countries have a proportional relationship between vaccination and rate of new cases.

So, to sum up, we have the following three cases:

- Vaccination doesn't affect the rate of new cases, but it decreases the peaks of the next waves of COVID-19.
- Vaccination doesn't have any effect on the rate of the new cases.
- Vaccination did decrease the rate of new cases.

We also put an indicator for each country to indicate the current state of the country regarding the new cases, where if the new cases estimate of today is larger than the average of the new cases across the whole period, then it indicates that new cases are increasing in these countries and that probably vaccination didn't have a major effect in this country.

Also, we can interpret the result of some countries as the appearance of new COVID-19 mutated versions so that will affect the effectiveness of the vaccines.

Resources:

- https://github.com/owid/covid-19-data/tree/master/public/data
- https://ourworldindata.org/covid-vaccinations?country=
- https://www.kaggle.com/gpreda/covid-world-vaccination-progress
- https://news.google.com/covid19/map?hl=en-US&mid=%2Fm%2F02k54&gl=US&ceid=US%3Aen&state=7
- https://www.kaggle.com/account/login?returnUrl=%2Ftherealcyberlord%2Fcoronavirus-covid-19-visualization-prediction&message=voteLogin
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8481107/
- https://www.cdc.gov/coronavirus/2019-ncov/vaccines/effectiveness/index.html