

ДИПЛОМ

По теме

"Анализ данных по COVID-19.

Исследование фактора вакцинации. Построение модели"

Выполнил: Студент потока DA-19
Селихов Д. Ю.

Содержание

1. Введение
2. Постановка задачи
3. Анализ
4. Методика решения
5. Результаты
6. Выводы и заключения
7. Список источников

1. Введение



В начале весны 2020 года мир захлестнула эпидемия коронавируса SARS-CoV-2. Люди заражались, заполнялись больницы, вводились локдауны по всему миру что приводило к замедлению экономики. Всех стало интересовать “когда все это кончится” и ответа по сути было два - либо все переболеют и выработается коллективный иммунитет либо на свет появится вакцина. Вакцина появилась спустя примерно год, их появилось несколько, в частности в России тоже, но не во всех странах темпы вакцинации были высокими. Попробуем в нашей исследовательской работе, **на основе данных** (никакой медицины, только цифры) ответить на вопрос - помогает ли вакцинация в борьбе с эпидемией, снижаются ли число заражений и летальных исходов.

2. Постановка задачи.

Исследовать предоставленные данные по COVID-19. Проанализировать влияние фактора вакцинации на число новых случаев заражения и смертей. Попытаться ответить на вопрос: Помогает вакцинирование или нет.

3. Анализ

Предоставленный набор данных содержит последние данные о состоянии Covid-19 в мире в деталях с 2/24/2020 по 6/29/2021. Этот набор данных отлично подходит для исследовательского анализа данных. В нем содержатся такие данные как:

число новых заражений,

всего заражений,

новых случаев смертей от коронавируса,

всего смертей от коронавируса,

число пациентов в отделении интенсивной терапии,

кол-во тестов,

тесты на тысячу населения,

количество вакцинаций за текущий день,

всего вакцинаций,

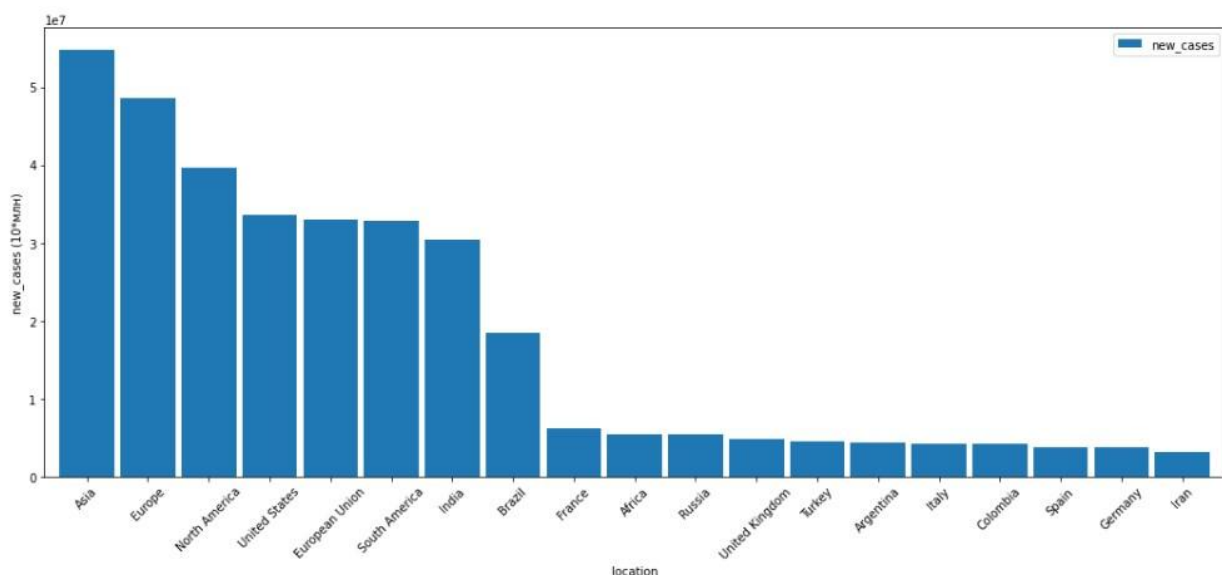
население,

средний возраст

и многое другое для каждой страны за каждую дату с 2/24/2020 по 6/29/2021

Всего датасет содержит 99 тысяч строк и 44 столбца.

ТОП стран и регионов по случаям заражения коронавируса на 29\06\2021
выглядит так:



Нам не нужны все данные по всем странам, оставим для исследования несколько стран из ТОП по заражениям, плюс Россия (наша страна), плюс Израиль (они были в первых рядах по темпам вакцинации).

Все столбцы нам также не нужны, оставим для исследования:

location - Страна

date - дата

total_cases - всего заражений на текущую дату

new_cases - новых заражений за эту дату

total_deaths - всего смертей на текущую дату

new_deaths - новые смерти за эту дату

total_vaccinations - всего вакцинаций

people_vaccinated - всего прививок первой дозой вакцины

people_fully_vaccinated - всего вакцинированных обеими дозами

new_vaccinations - новых вакцинаций за эту дату

population - Население страны

4. Методика решения

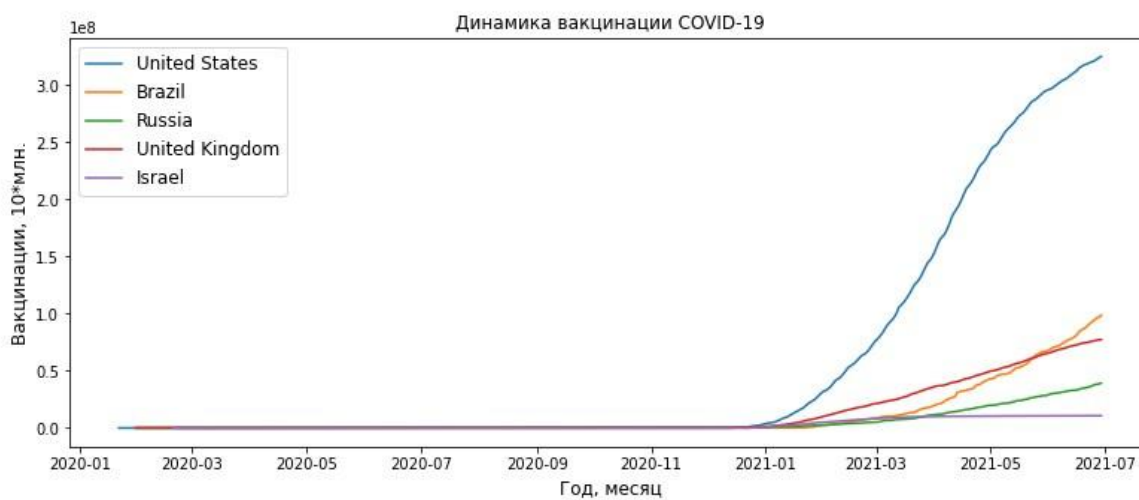
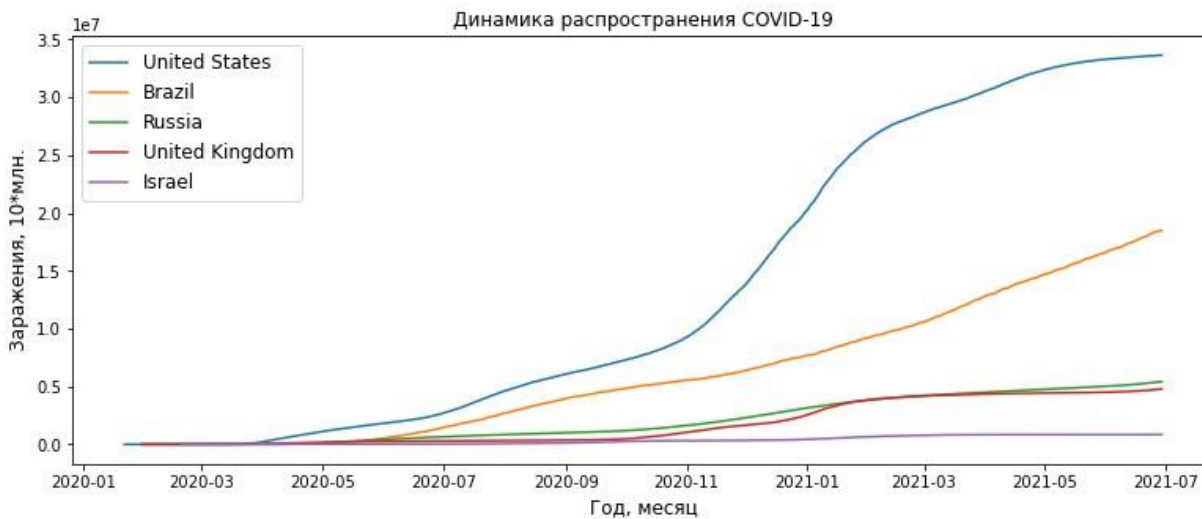
В наших данных в некоторых столбцах достаточно много пропусков, в частности в столбцах с информацией по смертям и вакцинациям. Логично предположить что данных по этим признакам сначала просто не было, т.к. Смерти фиксировались не сразу, а вакцина появилась почти спустя год после начала фиксирования случаев заражений. Таким образом количество смертей и вакцинаций до начала их фиксации, т.е. до появления данных можно заполнить нулями, НО мы не можем просто заполнить все пустые значения нулями, т.к. В данных есть пропуски и в период когда данные уже собирались. Здесь для заполнения этих пропусков применим интерполяцию, т.к. Вряд ли данные, например, по смертям или вакцинациям за день пропуска будут сильно отличаться от соседних дней. После очистки данных построим графики динамики заражений и вакцинаций, посмотрим на корреляцию между количеством вакцинированного населения и новыми случаями заражений и смертей.

Код на Python к данной работе:

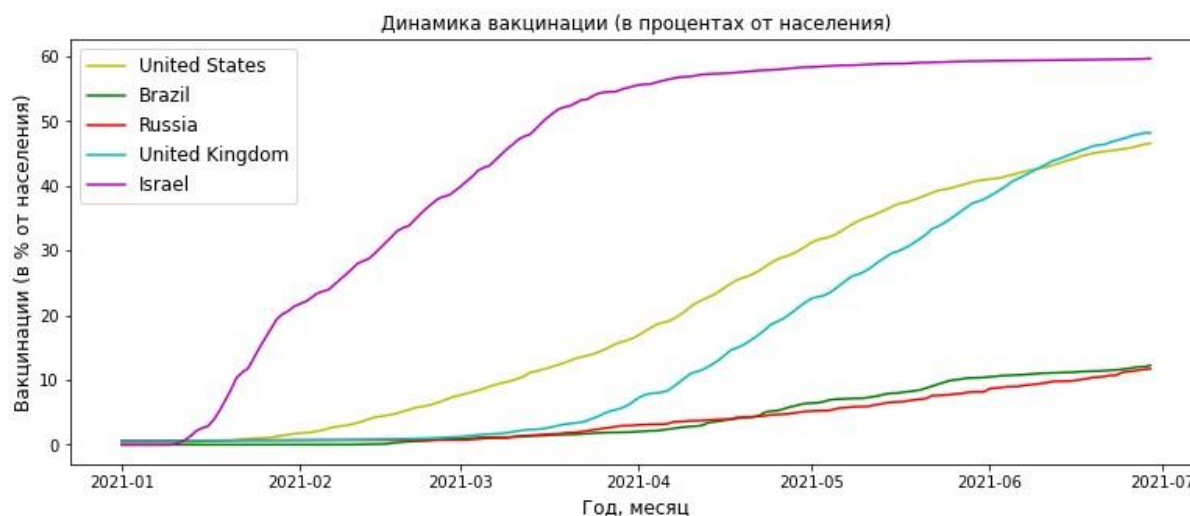
https://github.com/SelihovDmitry/pyda-19/tree/master/Diplom_Data_Analyst

5. Результаты

После предобработки и очистки данных можно построить графики динамики заражений и вакцинаций



По графику вакцинаций на первый взгляд можно сделать вывод (ошибочный ???) что США "впереди планеты всей", это так, но если считать в абсолютном числе вакцинаций, но для отдельных стран и для сравнения темпов вакцинации с другими странами будет правильнее считать кол-во вакцинаций как процент от населения. Для этого введем в наши данные новый столбец, где и будет % вакцинированного населения на каждую дату. Также оставим данные только с начала 2021 года, т.к. с этого периода началась массовая вакцинация. Посмотрим на график динамики вакцинации в процентах от населения страны:



Тут уже картина несколько другая, Израиль, практически сразу как появилась вакцина в массовом доступе (за эту отметку можно считать в среднем январь 2021) начал с высоких темпов вакцинации и довольно быстро (к марту 2021) достиг отметки близкой к 60% от населения.

США и Великобритания немного отстают, но тоже недалеко.

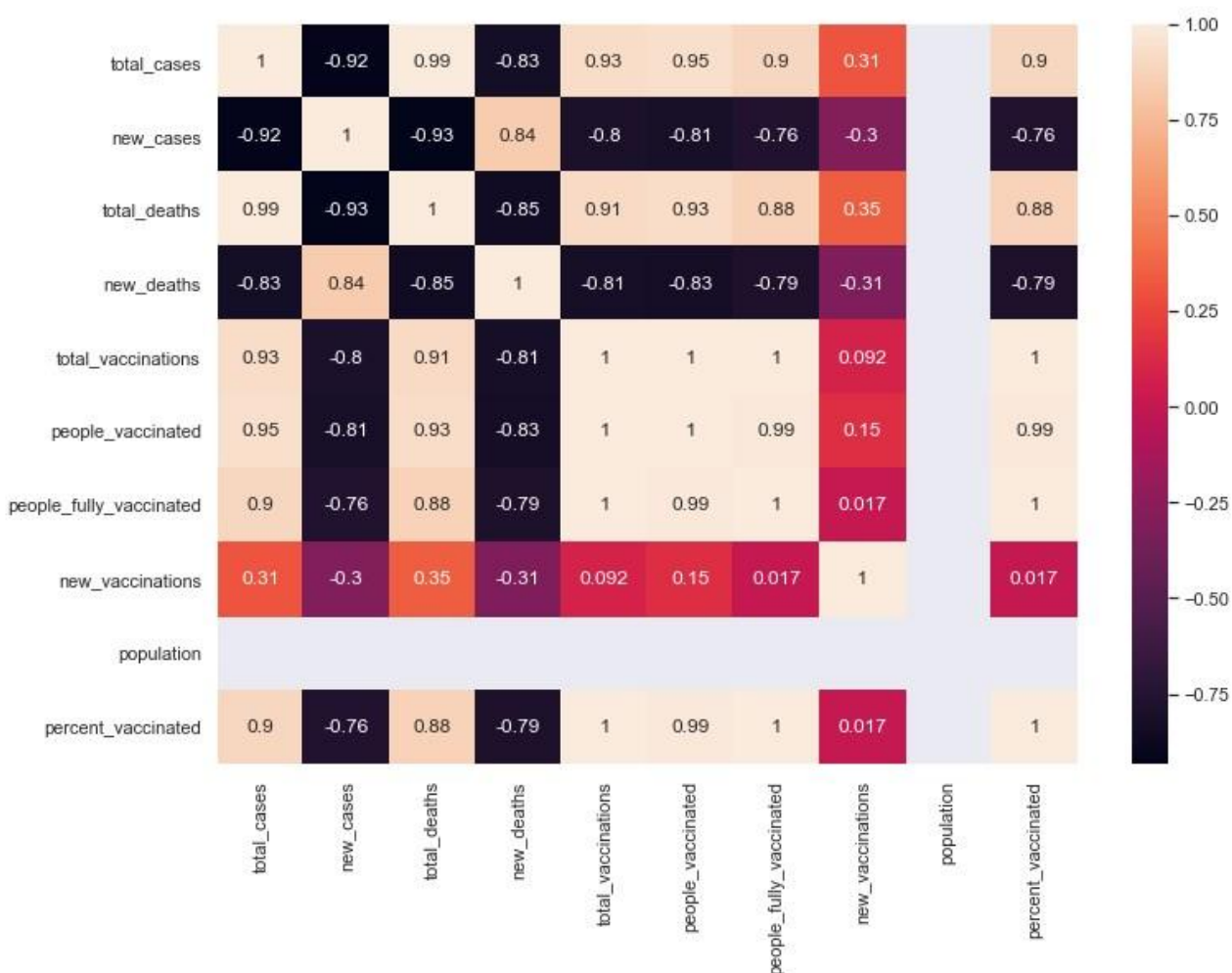
Россия и Бразилия пока в начале пути, вакцинировано не многим больше 10% населения (данные на конец июля 2021).

P.s. То, что у Израиля (да и у США и Британии наблюдается) темпы резко падают где-то после отметки 55-60% объясняется тем, что вакцинация проводится среди взрослого населения, а в популяции всё население, т.е. если из популяции убрать детей до 18-ти лет, и тех взрослых, которым вакцинироваться нельзя по медицинским показателям, то 60% от ВСЕГО населения - это будет близко к 90% от населения которому разрешено вакцинироваться.

Посмотрим корреляцию между различными показателями для США. Берем США, т.к. у них относительно плавно идет вакцинация и довольно много вакцинированных, Россия и Бразилия думаю в этом плане пока не сильно информативны...

Учитывая что мы исследуем влияние фактора вакцинации, смотреть будем также за 2021 год.

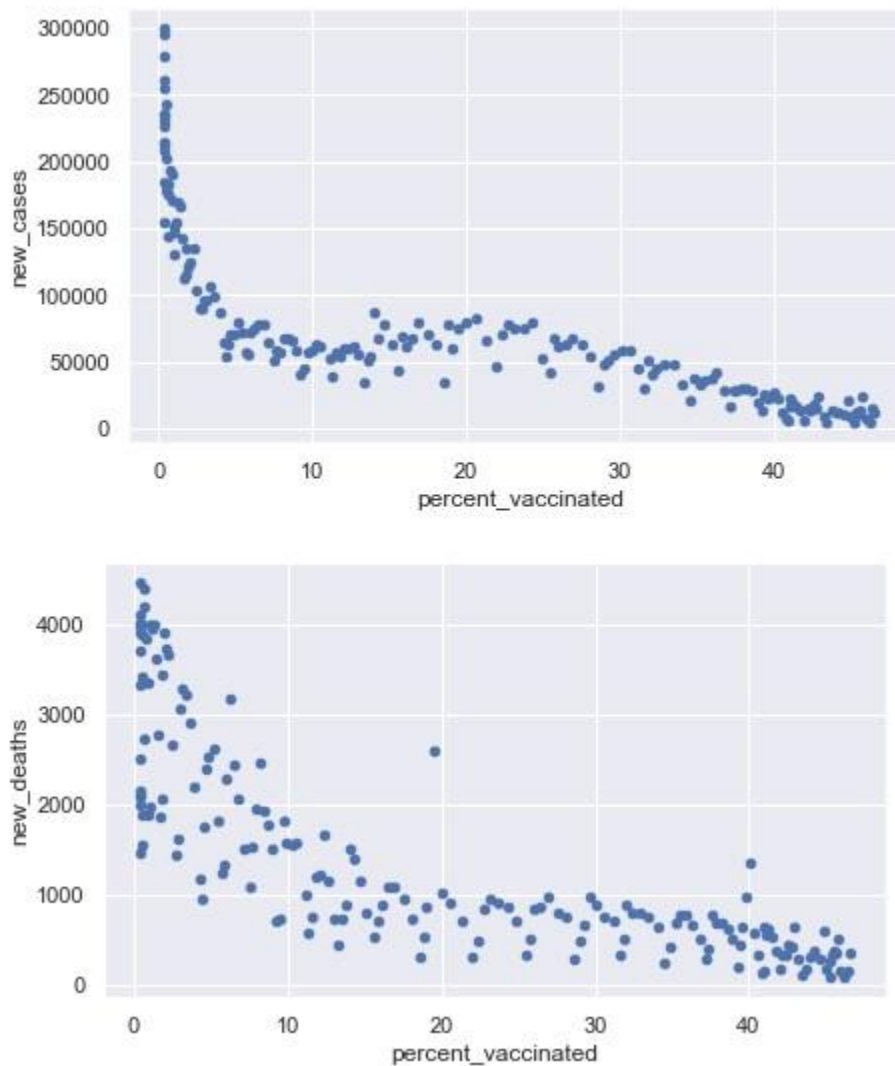
Тепловая карта зависимостей



Тут мы можем видеть сильные отрицательные зависимости заражений и смертей от фактора вакцинации, т.е. **чем больше абсолютное число и % вакцинированных - тем меньше заражений и смертей !!!**

Отсюда можно сделать вывод что **Вакцинация помогает в борьбе с эпидемией.**

Посмотрим также на графики зависимости заражений и смертей от % вакцинированного населения:



Тут также мы наглядно видим что чем больше вакцинаций, тем меньше число инфицированных и смертей.

6. Выводы и заключения

В результате исследования данных мы нашли зависимости новых случаев заражений коронавирусом и смертей от него и вакцинацией - чем больше вакцинаций, тем меньше число инфицированных и смертей.

И на основе полученных данных можем сделать вывод: **Вакцинация помогает в борьбе с эпидемией.**

7. Список источников

Данные для исследования:

<https://www.kaggle.com/hussainaliarif/largest-covid19-world-dataset>