



Дипломная работа Лобановой О.В.



# Введение:

*COVID-19* - это вызывающий заболевание штамм коронавируса, появившийся в декабре 2019 года и приведший к продолжающейся глобальной пандемии. Возможность предвидеть путь пандемии имеет решающее значение. Это важно для того, чтобы определить, как бороться, и отследить его распространение.

## Цель исследования:

Определить темп роста заболеваемости и смертности от инфекции в определенном временном промежутке на примере изучаемой страны - США, также спрогнозировать темп роста заболеваемости и летальности по причине COVID-19.

## Задачи:

провести анализ данных о распространении COVID19 в США;  
выявить взаимосвязи между ключевыми событиями в данных в США;  
построить прогноз распространения пандемии COVID19 в США.

## **Рассмотрение данных:**

Мы выбрали следующие данные для рассмотрения:

- 1.total\_cases - накопительный итогом кол-во новых случаев заболеваний
- 2.new\_cases - новые случаи заболеваний за день
- 3.total\_deaths - накопительный итогом кол-во смертей
- 4.new\_deaths - новые случаи смертей за день.

## **Мы выбрали США, т.к.:**

Это страна с большим числом жителей (более 100 млн.) в отличие от европейских стран и статистическая выборка больше;

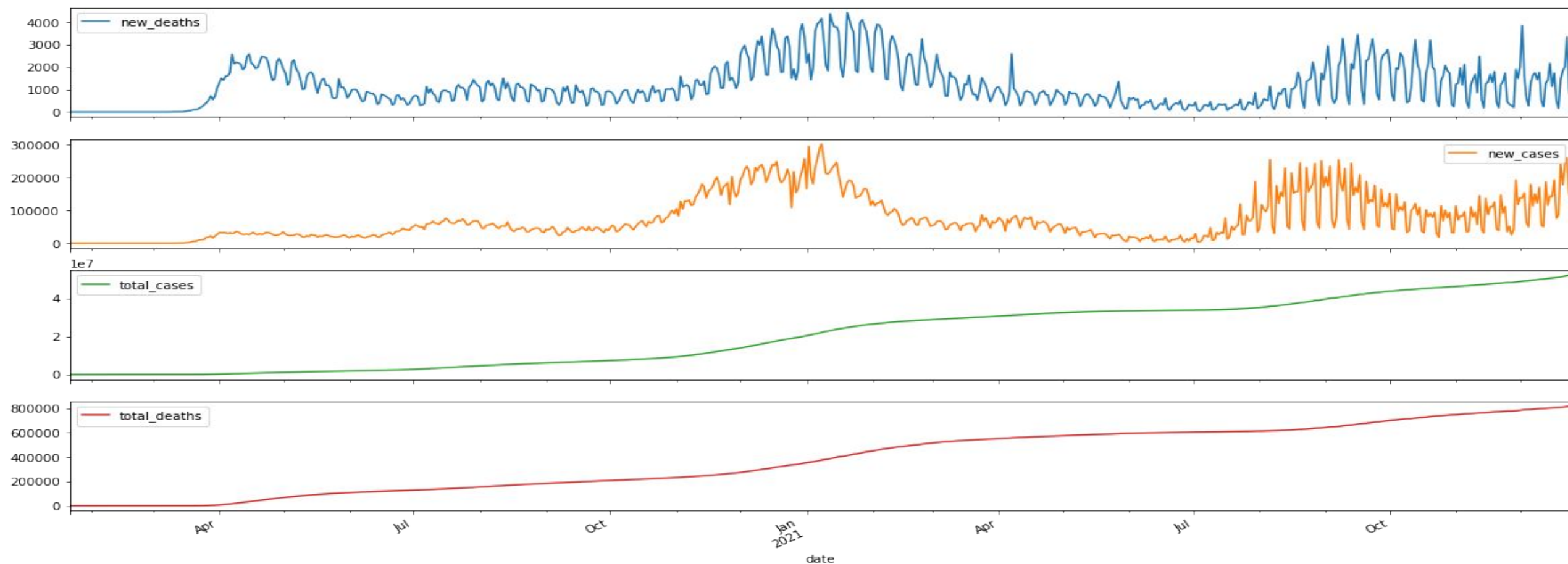
Статистика развитой страны имеет стандарты для сбора и обработки информации.

Максимальное значение новых заражений 303 461, а максимальное значение новых смертельных случаев 4442, что говорит об относительно низкой летальности заболеваний.

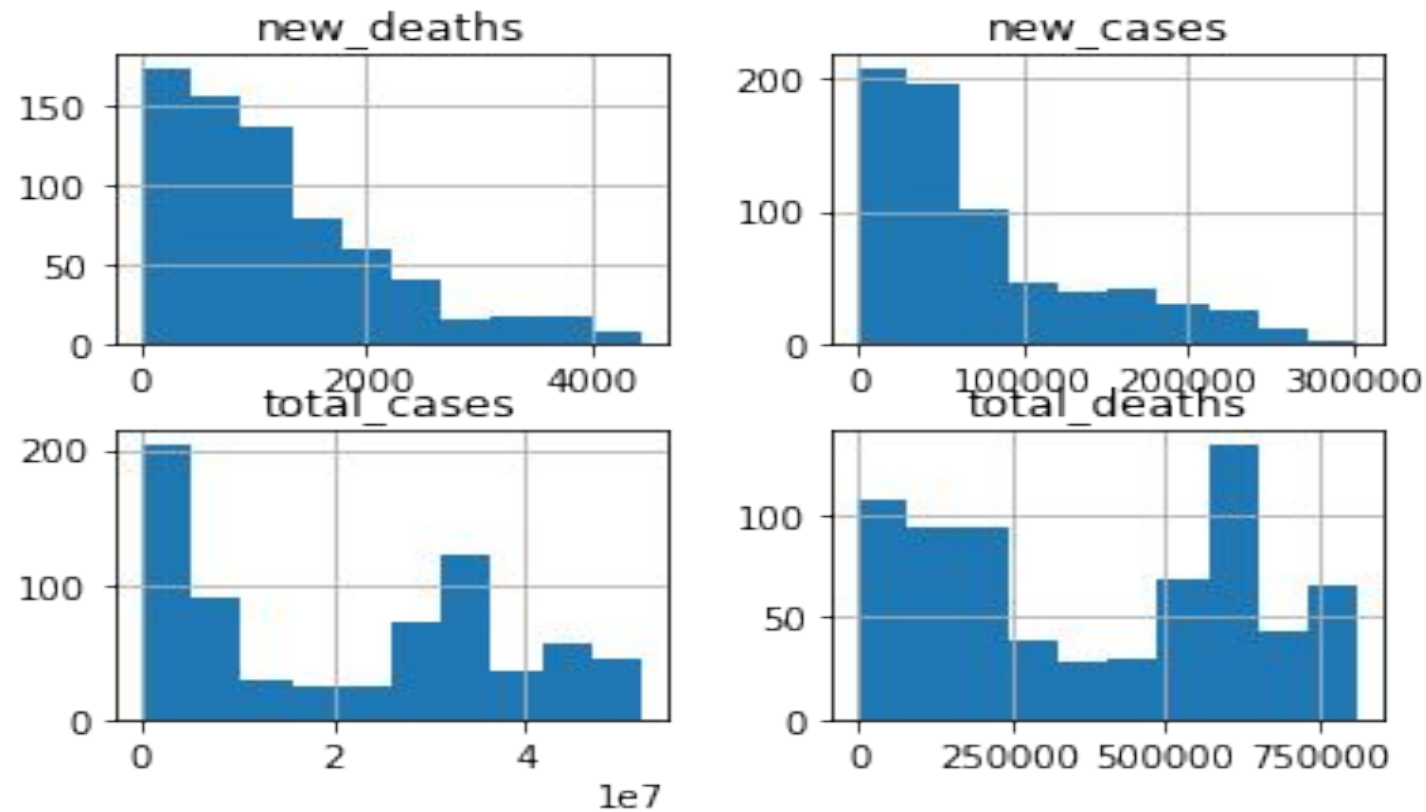
Размер выборки = 704

Мы построили общие графики и видно, что как заболеваемость, так и смертность неуклонно растет. Новые случаи заболевания и смертности имеют пики и плато, что объясняется сбором статистики в выходные дни.

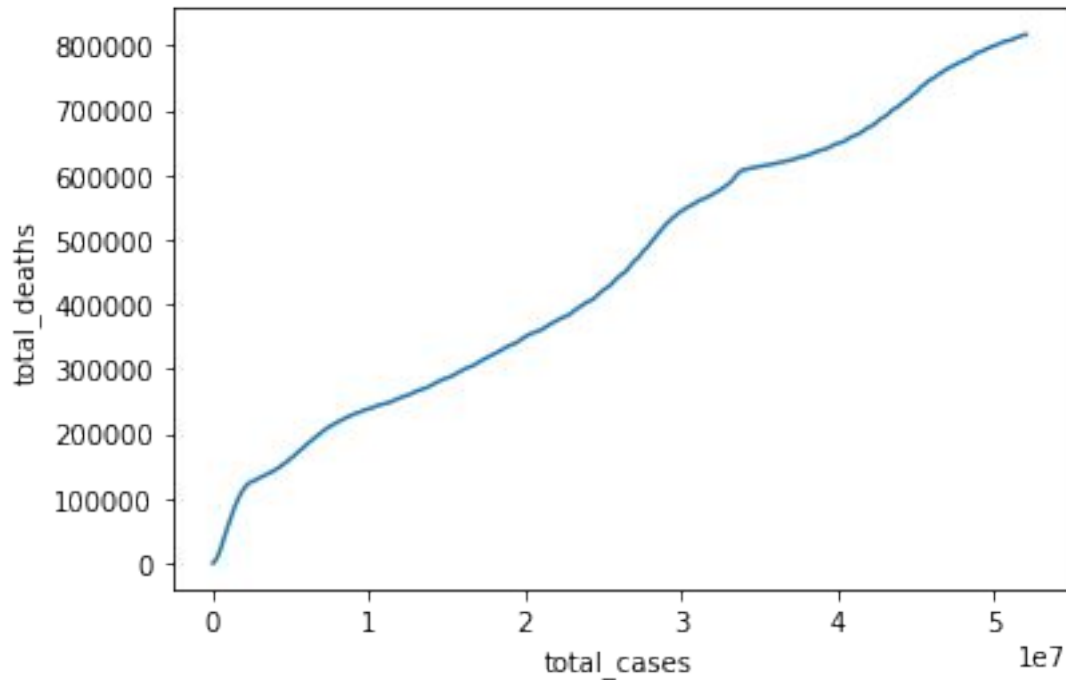
COVID-19 Соединенные Штаты



Анализ представленного ненормального распределения позволяет сделать вывод о пиковых нагрузках в период пандемии. На общем графике заболеваемости 2 основных пика, при этом увеличение смертности происходит при снижении пика заболеваемости, что говорит "сдвиге" периода заболевания и продолжительности болезни.

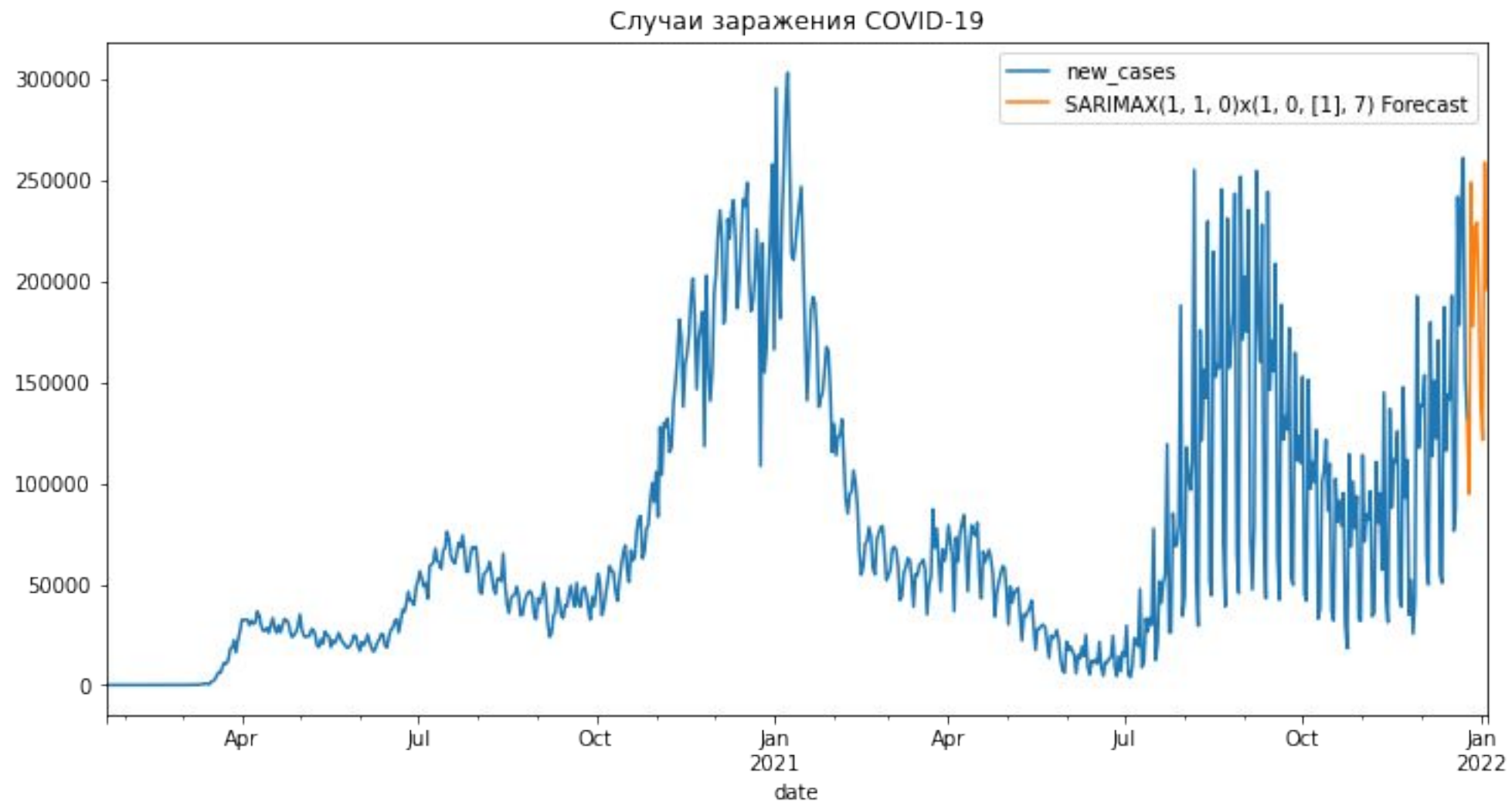


Зависимость общего количества смертности от общего количества заболеваемости

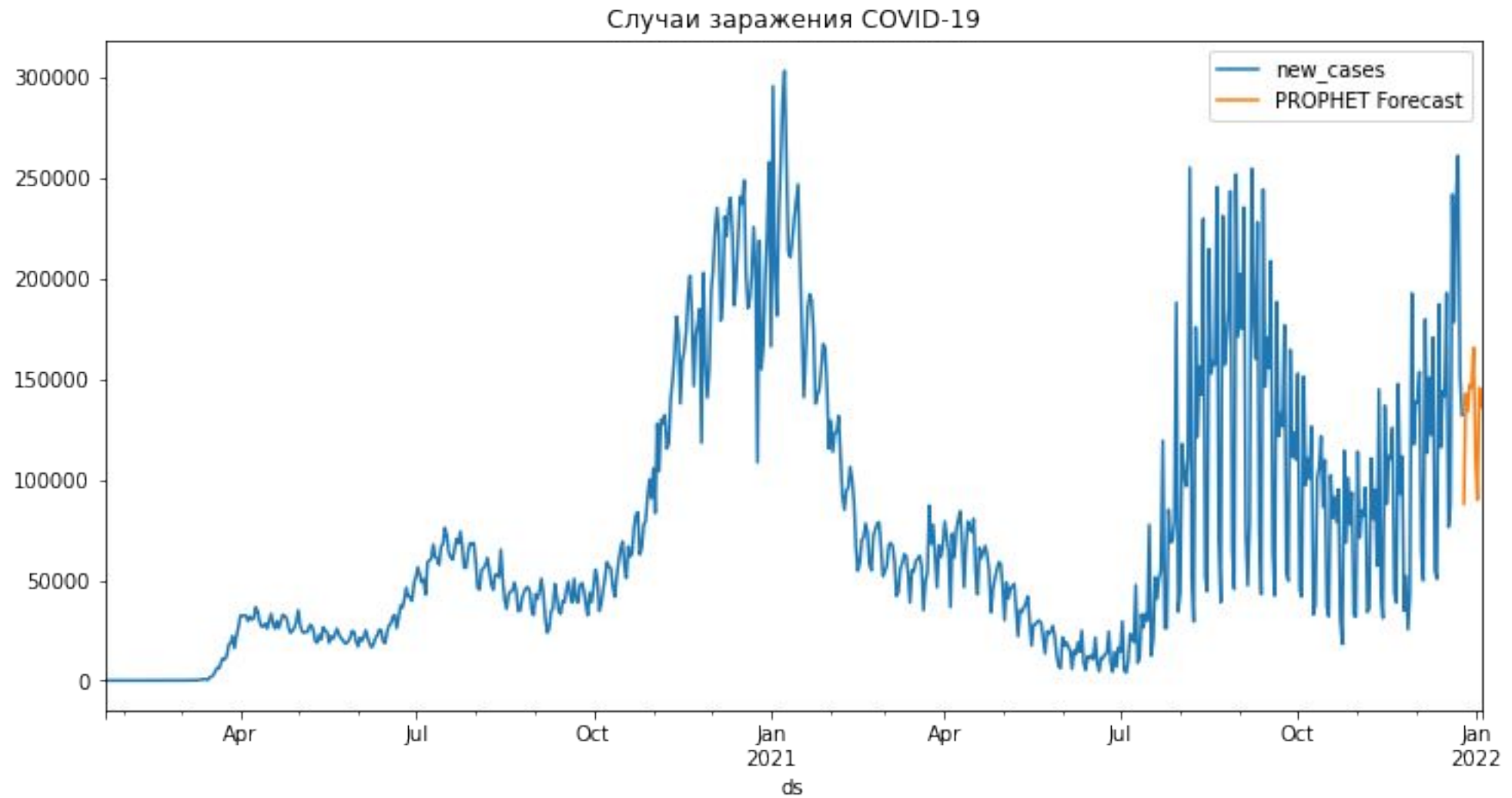


**ВИДИМ ВЫСОКИЙ**  
уровень корреляции  
между кол-вом  
заболеваний и  
смертей, что  
свидетельствует о  
воздействии  
сходных факторов  
на уровень  
заболеваемости

# Метод SARIMAX

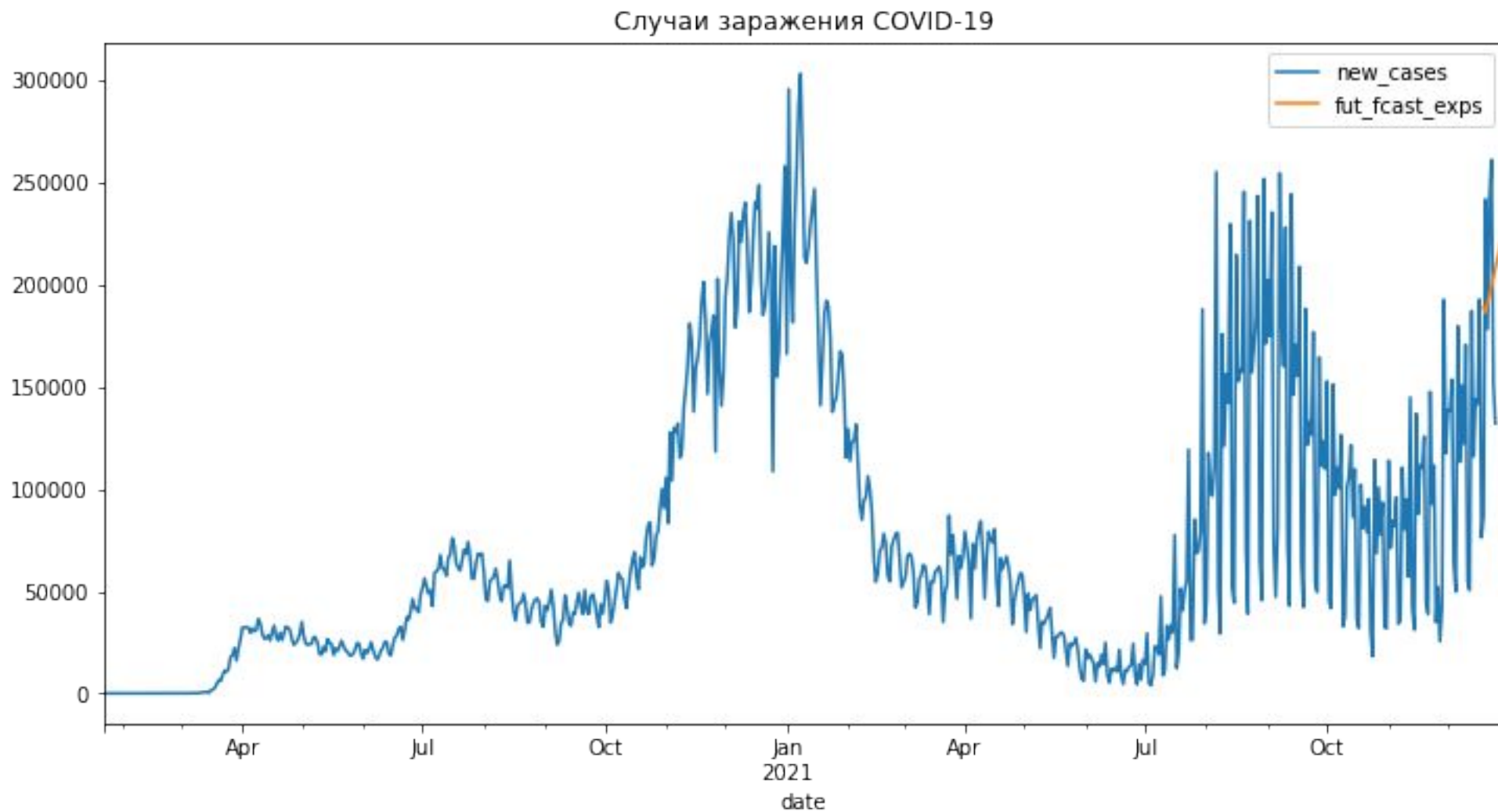


# Метод PROPHET





# Метод ХОЛЬТА-ВИНТЕРСА



## Выводы:

Мы рассмотрели самые распространенные методы прогнозирования временных рядов: Экспоненциальное сглаживание Хольта-Винтерса, SARIMA, Prophet и проведение экспериментов на временном ряде Covid-19 с ежедневными данными. Эксперимент проводился с помощью языка Python. Для прогнозирования использовалась библиотека pmdarima и fbprophet. Самой точной моделью в нашем случае будет SARIMAX. Также мы видим при прогнозировании, что рост числа заболевших растет и имеет тенденцию к росту в январе 2022 года согласно прогноза.

	SARIMAX	PROPHET	Хольта-Винтерса
MAE Error:	51076.22325	51553.94279	55279.74941
MSE Error:	4.112536579e+09	4.715242933e+09	4.432786621e+09
RMSE Error:	64129.06189	68667.62653	66579.17558
MAPE Error:	27.80707218	nan	48.43595246