

Аналитический отчет по распространению COVID-19 в Бразилии

Выполнила: Закирова Л.Р.

Казань
Декабрь 2021

Содержание

1. Введение
2. Постановка задачи
3. Обзор моделей
4. Описание процесса решения
5. Результаты

1. 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.

Источники данных:

Общедоступные ежедневные данные о COVID-19 за период с 24 марта 2020 года по настоящее время (декабрь 2021)

Методология:

В рамках данного проекта мы проведем анализ данных о заболеваемости, предложим и настроим несколько прогностических моделей, выполним прогноз, сравним результаты с новой статистикой, выявим наиболее эффективную модель.

Краткая справка:

БРАЗИЛИЯ – государство в Южной Америке

Площадь – 8 515 767 км (5 место в мире)

Население – более 207 млн человек (6 место в мире)

Статистика коронавируса в мире*

Случаев заражения: 280 101 606, Смертей: 5 404 608										
Данные обновлены: 26.12.21 11:47										
Статистика по миру на основе данных ВОЗ (задержка ≈ 24 часа).										
Страна	Всего случаев	за сутки	Сред. за 7 дней	при- вито	ПП+Р ?	Умерло	за сутки	Сред. за 7 дней	Смерт- ность	Сред. заб. на 100 тыс. чел.
1. США	52 095 411	+128 676	184 302	72.6%	32.7%	816 463	+27	1 411	1.6%	55
2. Индия	34 786 802	+6 987	6 660	60.5%	37.8%	479 682	+162	337	1.4%	
3. Бразилия	22 238 369	+3 746	3 180	77.5%	66.4%	618 686	+31	96	2.8%	1
4. Великобритания	11 958 928	+87	87 905	77%	70.5%	148 324		99	1.2%	131
5. Россия	10 392 020	+23 721	25 319	49.8%	33.3%	304 218	+968	1 002	2.9%	17
6. Турция	9 268 486		16 044	67.9%	67.9%	81 258		145	0.9%	19
7. Франция	9 192 843	+104 611	73 025	78%	74.3%	123 531	+84	161	1.3%	108
8. Германия	6 999 476	+17 238	26 598	73.9%	71.6%	110 367	+88	304	1.6%	32
9. Иран	6 182 905	+1 121	1 985	69.6%	61.8%	131 348	+42	45	2.1%	2
10. Испания	5 718 007		37 497	83.1%	70.6%	89 019		44	1.6%	79
11. Италия	5 622 431	+54 787	36 797	81.4%	74.7%	136 530	+144	141	2.4%	62
12. Аргентина	5 452 419	+7 183	9 424	82.6%	71.7%	117 020	+12	17	2.1%	21

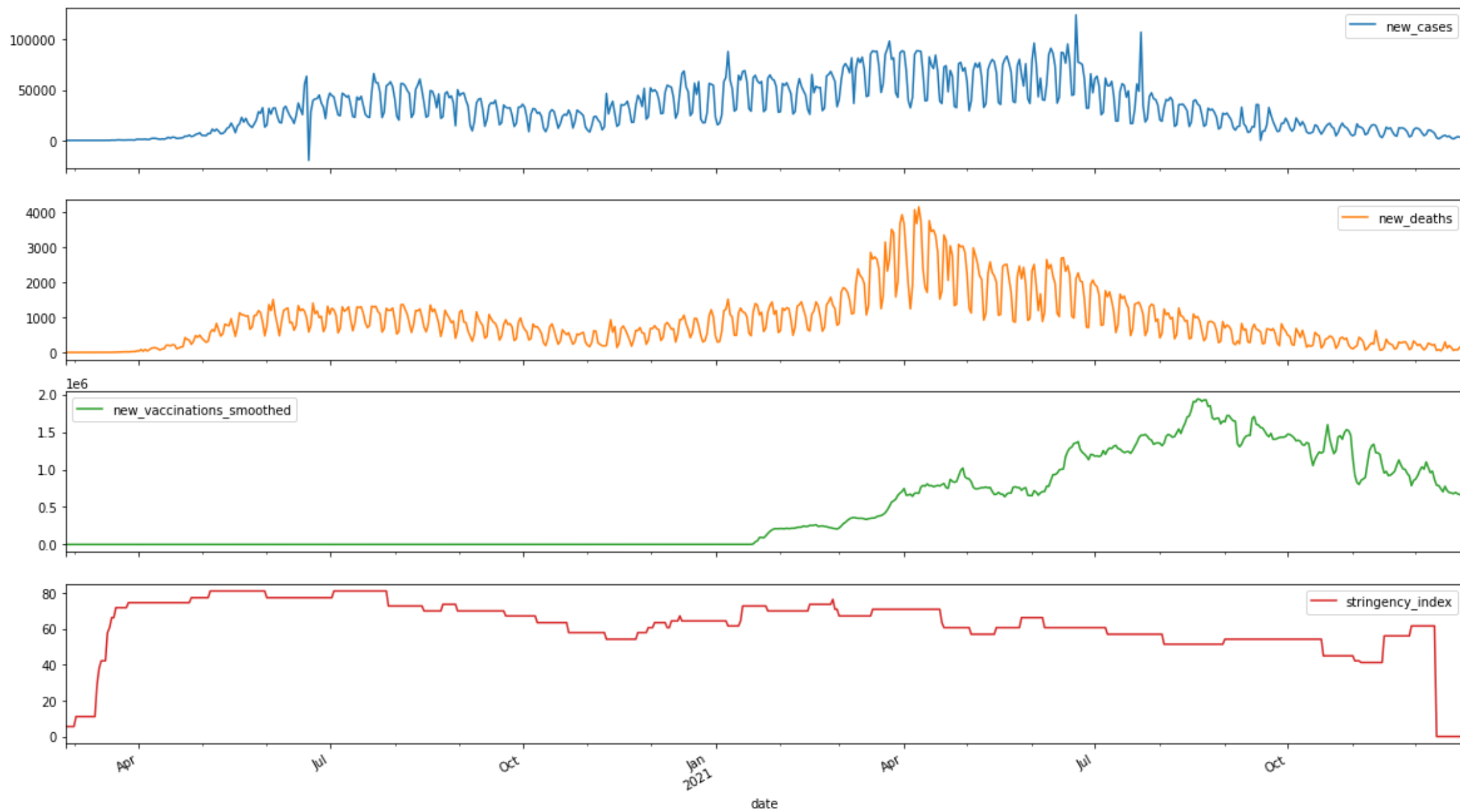
*данные ВОЗ

2. Постановка задачи.

В данном исследовании мы ставили себе 2 задачи:

- Определить влияние вакцинации на степень распространения заболевания и смертность
- Построить прогноз распространения заболевания на будущее

Основная статистика по Бразилии

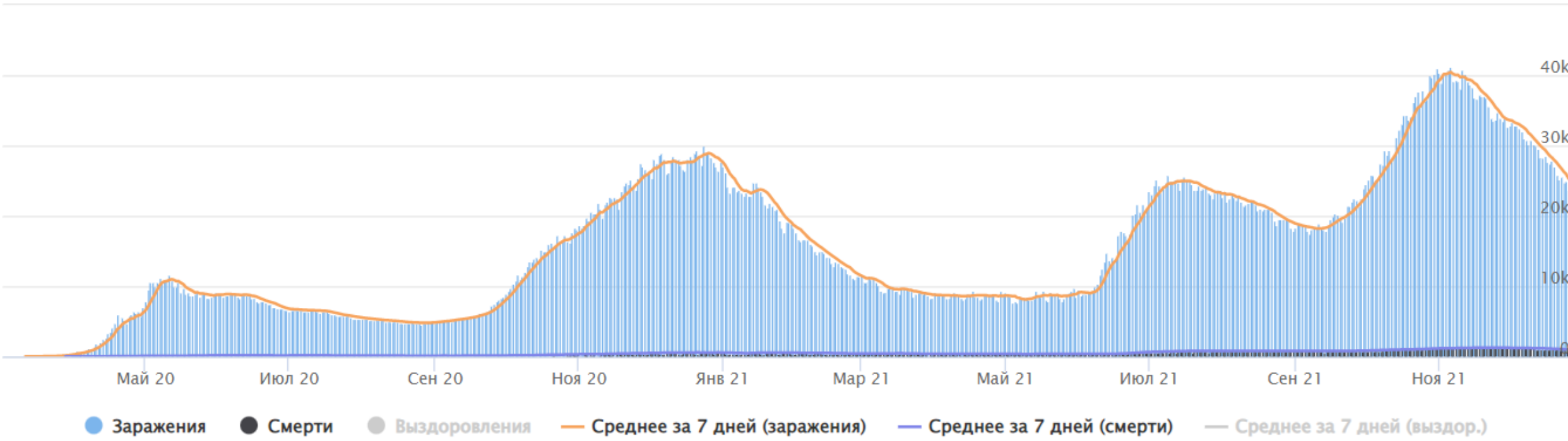


Онлайн статистика коронавируса *

в России [График](#) в Мире Карта

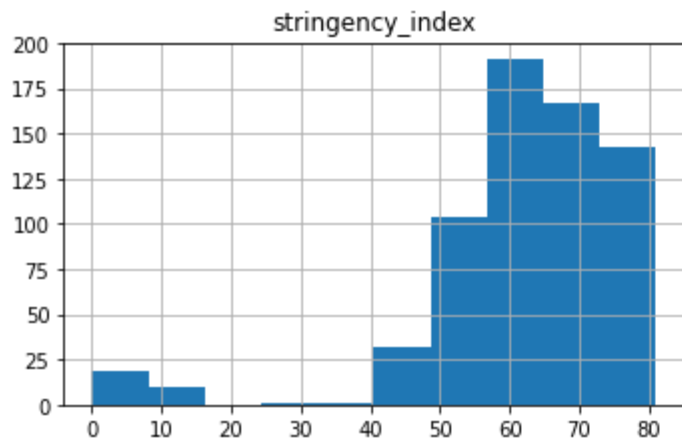
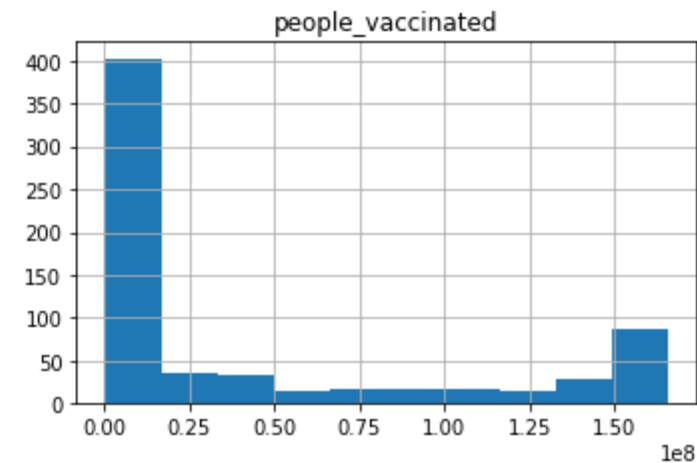
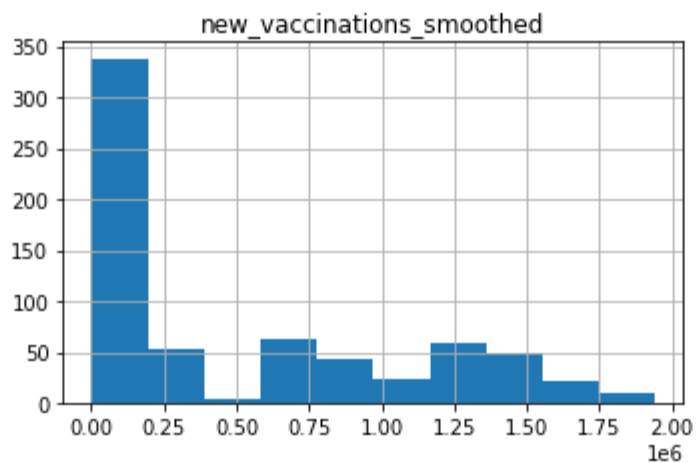
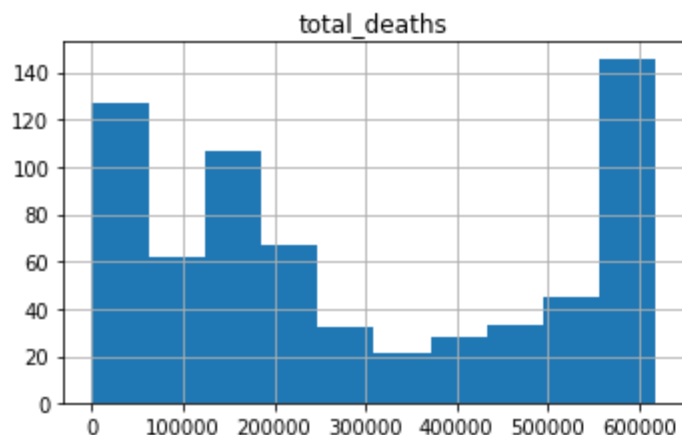
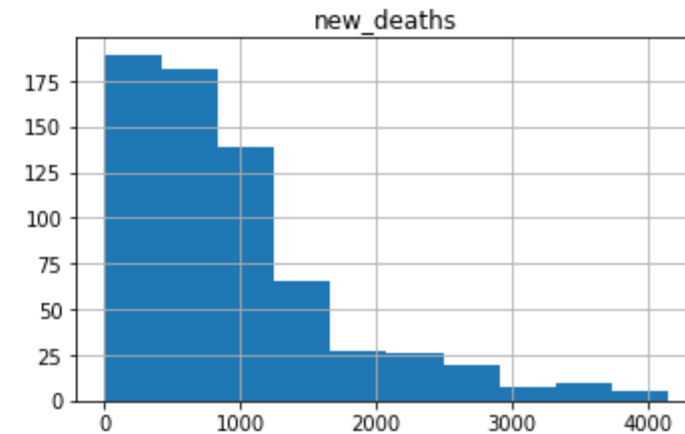
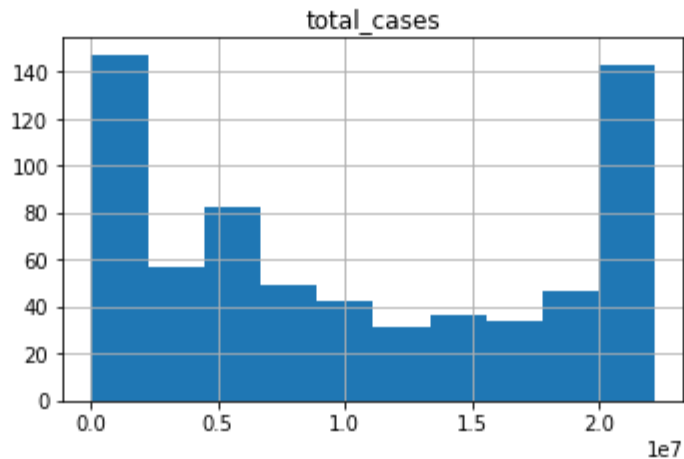
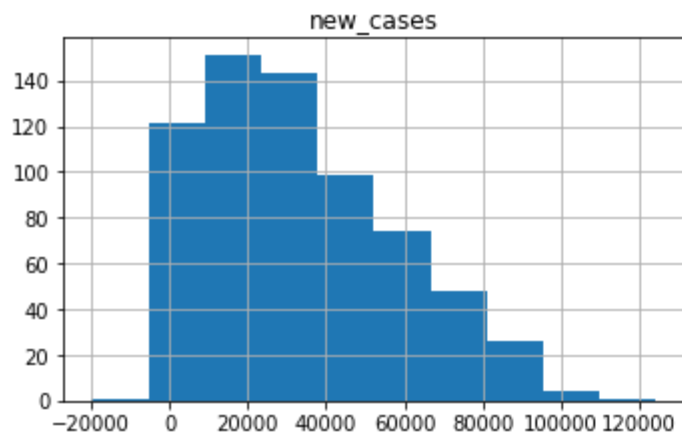
Статистика коронавируса по дням

Данные за 1 мес 3 мес 6 мес 1 год **все время**



Highcharts.com

*данные ВОЗ



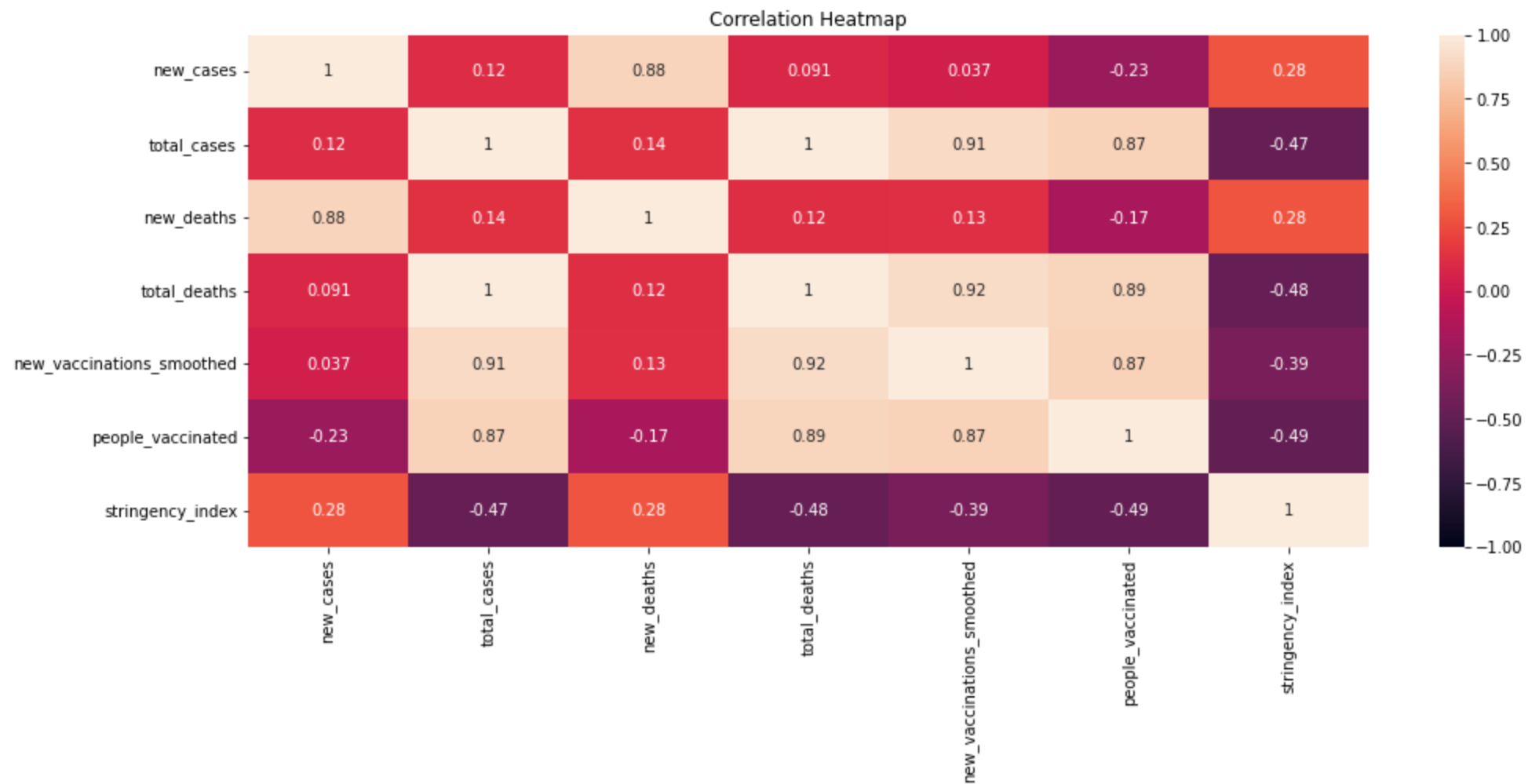
Осложняющие факторы

Главным противником жестких карантинных мер с самого начала был президент страны Жаир Болсонару.

Меры по борьбе с эпидемией вводились на уровне штатов

Новый штамм P1, впервые появившийся в штате Амазонас (ноябрь 2020)

Нехватка кислорода и оборудования ИВЛ



3. Обзор моделей

Мировой опыт показывает, что не существует единой модели, которая бы последовательно превосходила другие модели во всех ситуациях. Поэтому в данном исследовании предпринята попытка сравнить подходы, чтобы найти лучший метод, который можно использовать для прогнозирования развития пандемии в Бразилии.

Мы использовали 3 модели:

- SARIMA (Seasonal Autoregressive Integrated Moving Average) - модель сезонного авторегрессионного скользящего среднего
- Prophet
- Модель Хольта-Винтерса

Модель SARIMA (Seasonal Autoregressive Integrated Moving Average)

Это расширенная модель ARIMA с добавлением сезонности (Seasonal).

ARIMA - интегрированная модель авторегрессии — скользящего среднего.

Процесс авторегрессии — последовательная зависимость элементов временного ряда.

Процесс скользящего среднего — в процессе скользящего среднего каждый элемент ряда подвержен суммарному воздействию предыдущих ошибок.

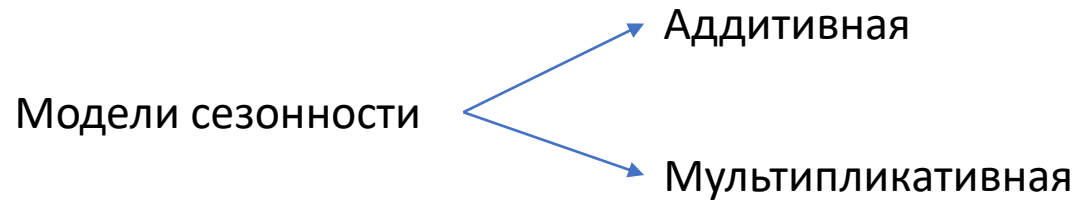
$$SARIMA \underbrace{(p, d, q)}_{non-seasonal} \underbrace{(P, D, Q)}_{seasonal}{}_m$$

p - количество членов авторегрессии;
d - количество разностей (порядок), сделанных, чтобы сделать его стационарным рядом;
q - количество членов скользящего среднего;
m – количество периодов

Таким образом, SARIMA учитывает сезонность, добавляя линейную комбинацию прошлых сезонных значений и/или прошлых ошибок прогноза.

Модель Prophet

Prophet - это процедура для прогнозирования данных временных рядов на основе аддитивной модели, в которой нелинейные тенденции корректируются с учетом годовой, еженедельной и ежедневной сезонности в дополнение к эффектам праздников. Лучше всего работает с временными рядами с сильными сезонными эффектами и несколькими сезонами исторических данных. Prophet устойчив к отсутствующим данным и изменениям в тенденциях и обычно хорошо справляется с выбросами.



На практике отличить аддитивную модель от мультипликативной можно по величине сезонной вариации. Аддитивной модели присуща практически постоянная сезонная вариация, тогда как у мультипликативной она возрастает или убывает, графически это выражается в изменении амплитуды колебания сезонного фактора.

Модель Хольта-Винтерса

Метод Хольта-Винтерса (алгоритм тройного экспоненциального сглаживания) - это трехпараметрическая модель прогноза, которая учитывает:

- * сглаженный экспоненциальный ряд;
- * тренд;
- * сезонность.

Экспоненциальное сглаживание —

один из наиболее успешных классических методов предсказаний. Прогнозы равны взвешенному среднему от старых наблюдений, и соответствующие веса убывают экспоненциально по мере хода времени.

Метод Хольта-Винтерса для расширения алгоритма позволяет добавить тренд и сезонность .

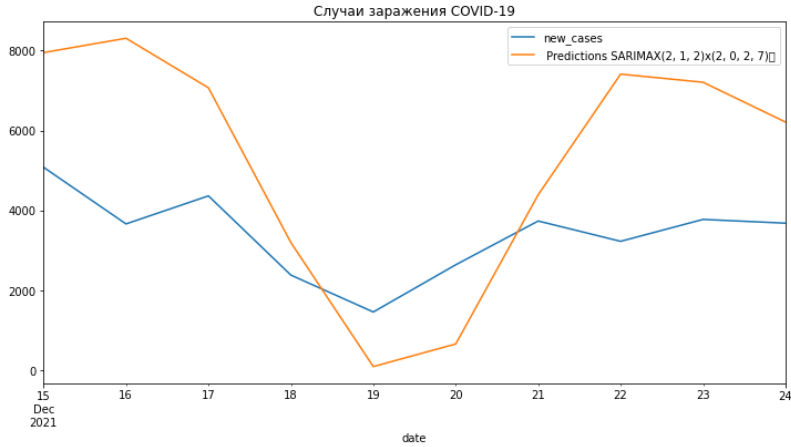
4. Описание процесса решения

Шаги:

- 1) Разделить выборку на тренировочную и контрольную
- 2) Обучить модель на тренировочной выборке
- 3) Сделать прогноз для контрольной выборки
- 4) Сравнить с реальными данными
- 5) Оценить точность модели
- 6) Сделать прогноз на будущее

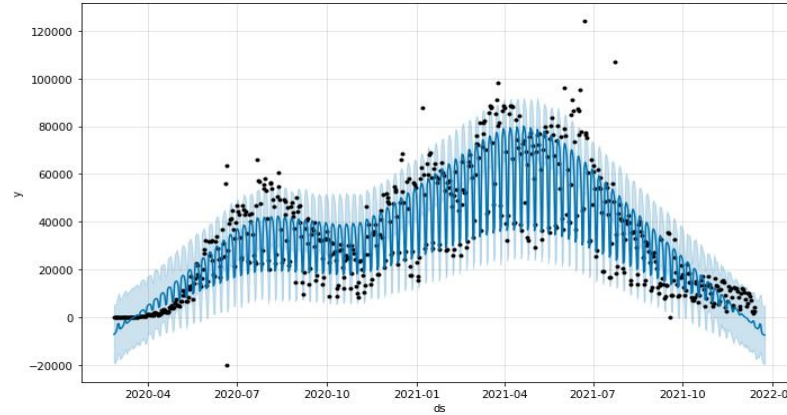
5. Результаты

SARIMA



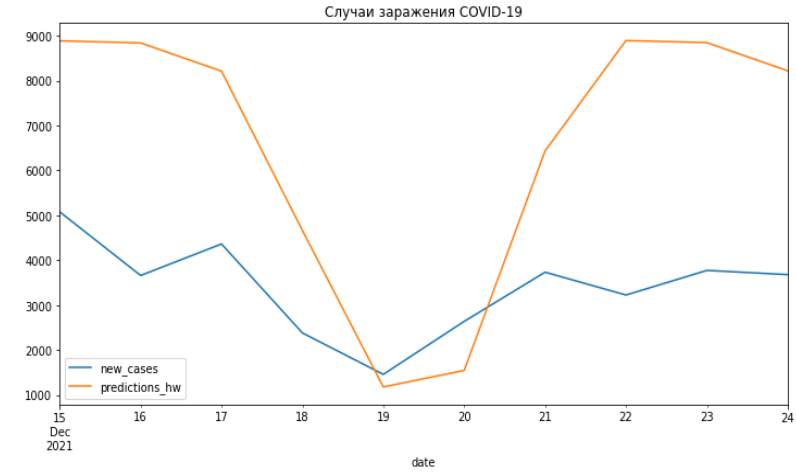
SARIMA(2, 1, 2)x(2, 0, 2, 7) MAE Error: 2518.316787
SARIMA(2, 1, 2)x(2, 0, 2, 7) MSE Error: 7969319.669
SARIMA(2, 1, 2)x(2, 0, 2, 7) RMSE Error: 2822.998347
SARIMA(2, 1, 2)x(2, 0, 2, 7) MAPE Error: 75.53439027

PROPHET

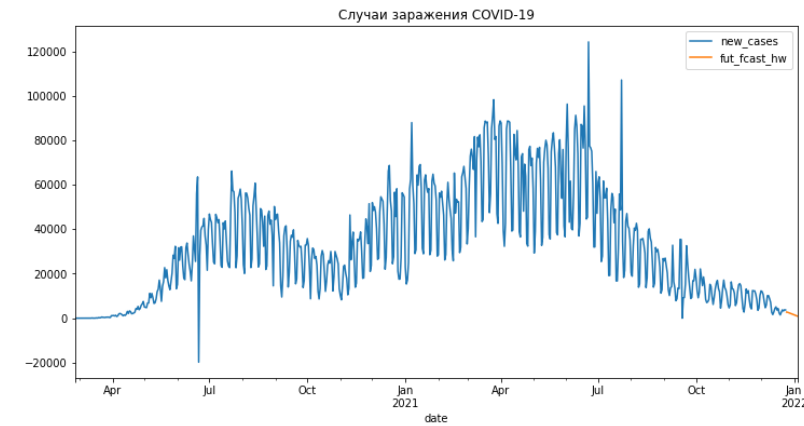
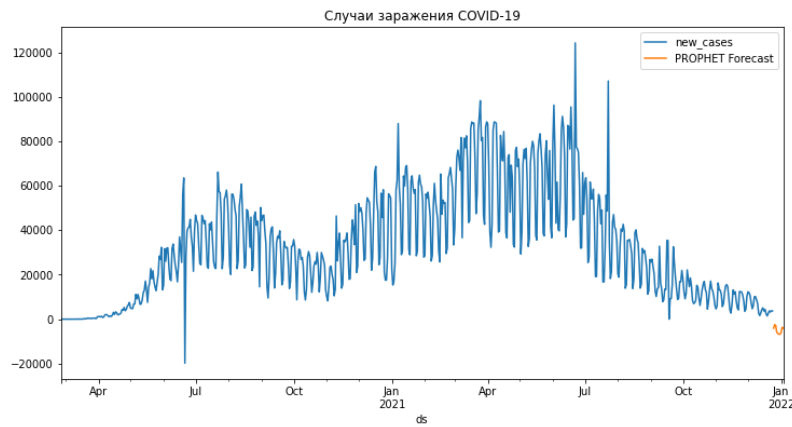
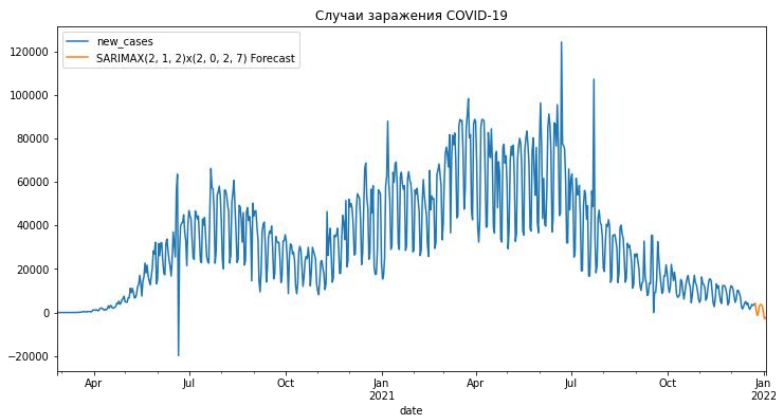


Prophet MAE Error: 8601.706071
Prophet MSE Error: 78998333.1
Prophet RMSE Error: 8888.100646
Prophet MAPE Error: 259.0947223

Holt-Winters



Holt-Winters Model MAE Error: 3447.704822
Holt-Winters Model MSE Error: 14841752.6
Holt-Winters Model RMSE Error: 3852.499526
Holt-Winters Model MAPE Error: 96.70627072



Спасибо за внимание!