

## Введение

В конце 2019 года в Китайской Народной Республике (КНР) произошла вспышка новой коронавирусной инфекции с эпицентром в городе Ухань (провинция Хубэй), возбудителю которой было дано временное название 2019-nCoV.

27 января 2020 при Правительстве России был создан оперативный штаб. 19 марта 2020 в Москве произошёл первый случай летального исхода. 18 апреля Президент России **Владимир Путин** поручил правительству к 20 апреля обеспечить формирование краткосрочного прогноза количества граждан, которые могут заболеть коронавирусом. С этого момента Правительство РФ включило в координационный план меры по оперативному анализу и прогнозированию распространения 2019-nCoV в России. С ростом количества больных, прогнозирование перешло в формат оперативных действий. На основе математических моделей, которые коррелировались с реальной статистикой - осуществлялись работы по наращиванию коечного фонда и обеспечению материально-технической базы. Таким образом, методы статистического прогнозирования, в том числе с использованием технологии искусственного интеллекта позволили заранее оценить потенциальный ущерб от эпидемии коронавируса.

В Республике Тыва республиканский оперативный штаб по предотвращению распространения новой коронавирусной инфекции начал свою работу в начале марта. Прогнозирование эпидемиологической ситуации в Туве осуществлялось на основе методических данных представленных Роспотребнадзором и Минздравом Российской Федерации. 14 мая на Заседании Правительства, Глава Республики Кара-оол Ш.В. акцентировал внимание на методы статистического прогнозирования эпидситуации в республике. 15 мая 2020 года **НИИ медико-социальных проблем и управления Республики Тыва** выпустили публичную аналитическую справку об эпидемической ситуации в Республике Тыва. Исследования носят краткосрочный характер методом экстраполяции полиномиального тренда глубиной 1-3 суток. Это очень важное событие, так как подчеркивает потребность Тувы в инновационных инструментах анализа, которые будут способствовать эффективному управлению и планированию.

Данный аналитический отчёт подготовлен независимым исследователем в области обработки данных, с использованием различных методов статистического анализа. Цель: оценить проблему 2019-nCoV в Туве, через собственные изыскания, посредством новых методов обработки данных.

## Открытые данные

### Источники данных

Для осуществления качественного анализа, необходимо иметь **унифицированные данные с расширенным набором метрик**. К сожалению, большинство публичных данных об эпидситуации в Тыве носят новостной характер без указания клинических деталей, которые очень важны для оценки эпидситуации. В качестве требований к источнику данных была произведена оценка текстовых данных (новостей) для последующей машинной обработки. Сайт Министерства здравоохранения РТ, показал наилучшие результаты, ввиду устойчивой семантики текста. Таким образом, в основе данного исследования лежат новости об эпидситуации с сайта Министерства здравоохранения Республики Тыва <https://minzdravtuva.ru/>

*Таблица №1 Технические возможности источников*

Источник	Оперативность	Дашборд	Наличие размеченных данных (мт)	Наличие расширенных данных	Возможность прямого парсинга	Api
Сайт Правительства РТ	-	-	-	-	+	-
Сайт Министерства здравоохранения РТ	+	-	-	-	+	-
Сайт Роспотребнадзора по РТ	-	-	-	-	+	-
Группа Оперативного штаба в ВК	+	-	-	-	+	+-

Из-за слабых технических компетенций в части масштабирования информационных систем - представленные источники не способны в полной мере раскрыть полновесную картину эпидситуации для машинной обработки. **Таким образом, в Тыве отсутствует источник данных, в формате, обеспечивающем ее автоматическую обработку без предварительного изменения (машиночитаемый формат csv, json, xml)** Отсутствие доступных и правильно сформированных данных сильно ограничивает исследовательскую работу, делаем невозможным дальнейшую интеграцию и внедрение инновационных методов анализа.

### Рекомендации от автора:

1. Рекомендовать Министерству здравоохранения РТ, НИИ медико-социальных проблем и управления Республики Тыва -

разработку единого информационного ресурса (дашборд), с расширенной выгрузкой данных в машиночитаемом формате.

2. Рекомендовать органам исполнительной власти, при формировании технических заданий на разработку информационных систем, включать требования на обязательное включение функции выгрузки данных в машиночитаемом формате за указанный период.

Для проведения анализа 2019-nCoV, я разработал программный парсер на языке python, который позволяет получать новостную ленту с сайта Министерства здравоохранения РТ и конвертировать данные в формат json и csv. Программа доступна для свободного использования в репозитории [https://github.com/tarbagan/covid\\_tuva\\_parser](https://github.com/tarbagan/covid_tuva_parser)

Программа извлекает базовые метрики **date-data**, **infection** - кол-во инфицированных, **lab** - лаб. исследований, **recovered** - выздоровевших, **dead** - умерших, **news** - текст новости

Фрагмент выходных данных программы covid\_tuva\_parser

**date,infection,lab,recovered,dead,news**

**2020-05-21,530,29921,63,1,**По данным Управления Роспотребнадзора по РТ и Минздрава РТ в Республике Тыва за весь период проведено 29 921 лабораторных исследований. По результатам активных противоэпидемических действий выявлен 530 случай заболевания COVID-19 в Республике Тыва. Выписано в связи с выздоровлением 63 пациента. В активной форме заболевания находится 466 человек. Зарегистрирован один летальный случай за весь период.

**2020-05-20,451,29342,62,1,**По данным Управления Роспотребнадзора по РТ и Минздрава РТ в Республике Тыва за весь период проведено 29 342 лабораторных исследований. По результатам активных противоэпидемических действий выявлен 451 случай заболевания COVID-19 в Республике Тыва. Выписано в связи с выздоровлением 62 пациента. В активной форме заболевания находится 388 человек. Зарегистрирован один летальный случай за весь период.

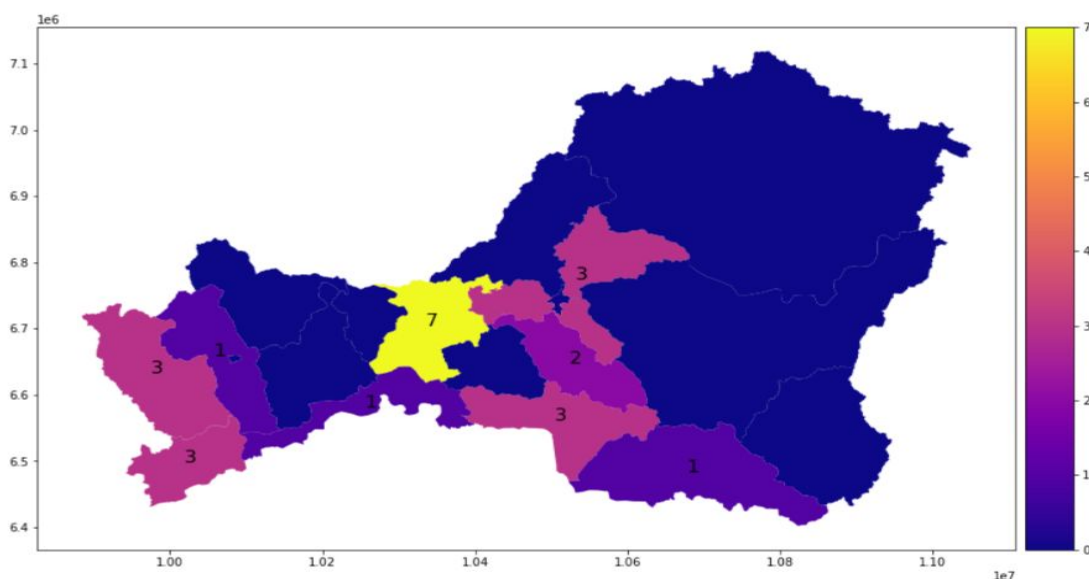
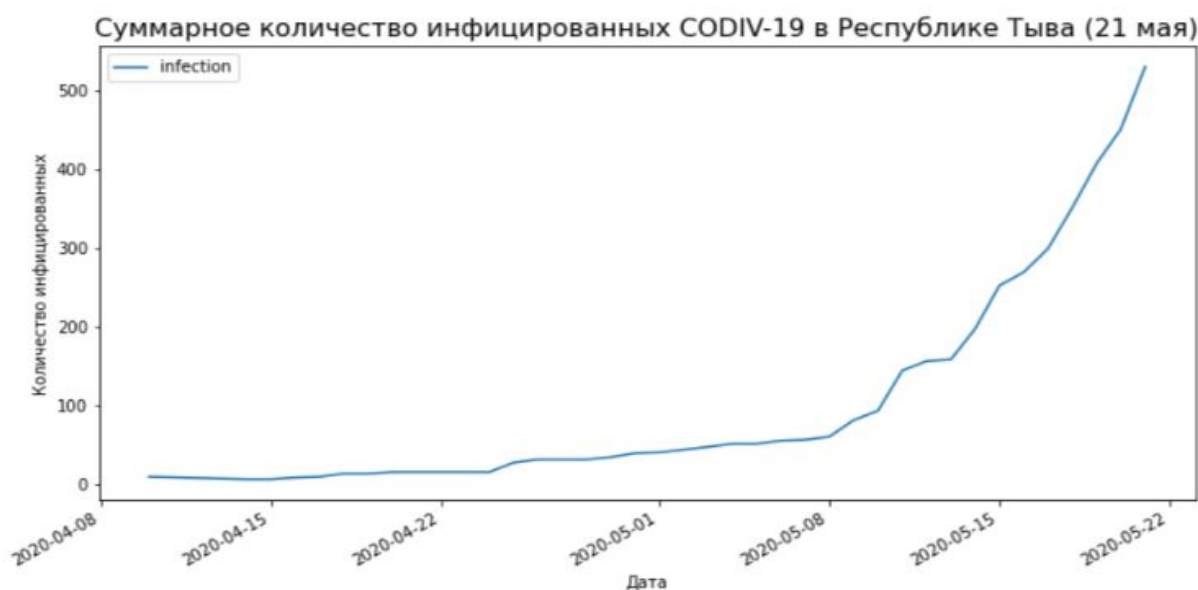
На 21 мая, для анализа было извлечено и обработано в автоматическом режиме 52 новости, что соответствует 52 дням наблюдения. Для получения некоторых видов данных, использовалась ручная выборка.

	date	infection	lab	recovered	dead	news
date						
<b>2020-05-21</b>	2020-05-21	530.0	29921.0	63.0	1.0	По данным Управления Роспотребнадзора по РТ и ...
<b>2020-05-20</b>	2020-05-20	451.0	29342.0	62.0	1.0	По данным Управления Роспотребнадзора по РТ и ...
<b>2020-05-19</b>	2020-05-19	407.0	28636.0	62.0	1.0	По данным Управления Роспотребнадзора по РТ и ...
<b>2020-05-18</b>	2020-05-18	351.0	27943.0	56.0	1.0	По данным Управления Роспотребнадзора по РТ и ...
<b>2020-05-17</b>	2020-05-17	299.0	20972.0	54.0	1.0	По данным Управления Роспотребнадзора по РТ и ...

## Анализ

### Текущая динамика

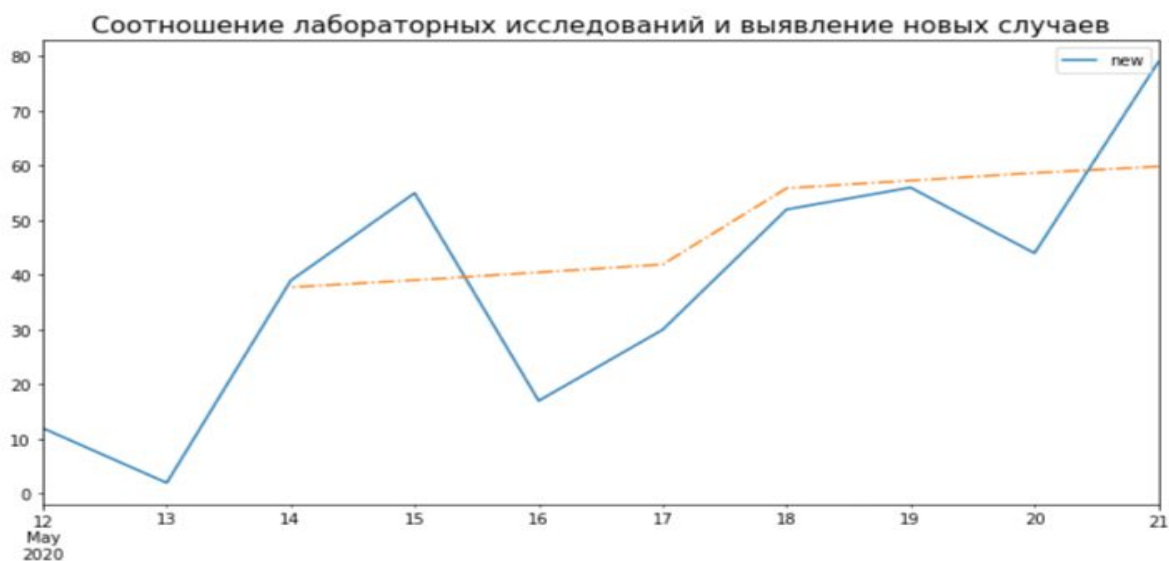
Суммарное количество инфицированных CODIV-19 в Республике Тыва отражает абсолютное значение количества инфицированных на дату. График представляет из себя логистическую кривую на росте. По всей видимости, точка перелома кривой и переход в состояние плато находится в неопределённых сроках, а мы находимся на “точке стремительного роста”. **На 21 мая заболеваемость составила 530 чел. или 0.16% от общей численности населения (по РФ 0.21%).**



## Ежедневный прирост



На 21 мая суточный прирост **79 (56.0% ранее 44) человек**. График имеет ярко выраженную ломаную кривую, с провалами через каждые сутки. По всей видимости, обработка анализов и публикация результатов происходит с двухдневной периодичностью. Таким образом, через каждые сутки наблюдается “резкий рост”.

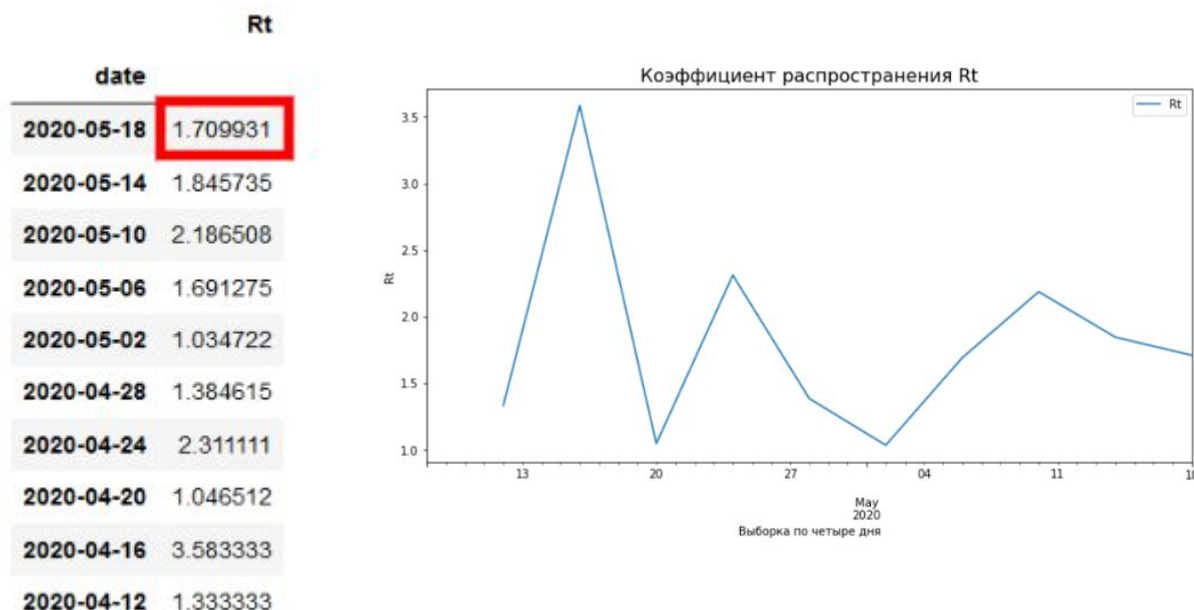


Сравнение данных по количеству проведенных лабораторных исследований и количеству выявленных случаев не дал никаких результатов, из-за небольшого количества входных данных.

**Для оценки причин роста, необходимы данные по эпидемиологическим расследованиям, а также векторы передвижений заболевших.**

## Коэффициент Rt (официальная метрика Роспотребнадзора РФ)

Коэффициент распространения Rt рассчитывается по формуле, приведенной в рекомендациях Роспотребнадзора для регионов по ослаблению ограничительных мер, введенных из-за распространения вируса. Количество (сумма) инфицированных за четыре дня, делится на сумму за предыдущие четыре дня.



Из графиков видно\*, что коэффициент Rt имеет широкий диапазон значений на протяжении двух месяцев. Связанно это с тем, что Республика Тыва находится на линии неопределенного роста, проще говоря: коэффициент распространения Rt следует учитывать лишь в краткосрочной перспективе, сроками не более семи дней. **Коэффициент Rt в Тыве на 21 мая ↑1,7 (высокий), по РФ 0,89↓** Прогноз Rt на следующий период можно оценить как неблагоприятный.

**Учитывая крайне высокий коэффициент распространения Rt, частичное снятие ограничений в Тыве может негативно повлиять на эпидситуацию**

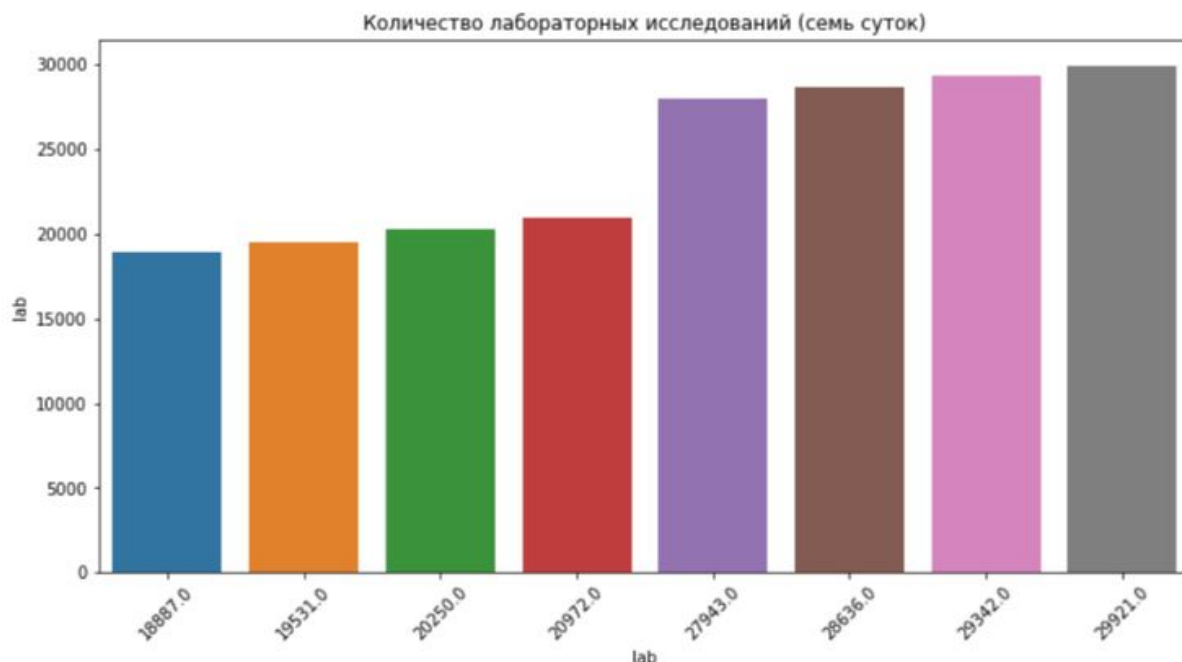
**Справочно:** Поэтапное снятие ограничений принятого Правительством РФ рекомендуется начинать не ранее достижения коэффициента Rt = 1.0

\* В виду технической особенности, последнее значение 0,7 не учитывает последнее суточное значение.



## Охват тестирования населения методом ПЦР

ПЦР метод - метод выявления вируса с помощью полимеразной цепной реакции. Для осуществления тестирования используются специальные ПЦР наборы.



Охват тестированием населения методом ПЦР рассчитывается как среднее значение за предыдущие 7 дней по формуле:

$$OT = ((TCx - TC \times 7) \times 100\,000 / HC) / 7$$

Охват тестированием населения методом ПЦР на 21 мая составил 453.0 на 100 тыс. населения. За семь дней осуществлено 10390.0 лабораторных тестов, что в среднем составляет в 1484.0 в день. **Охват на 100 тыс. населения 453.0**

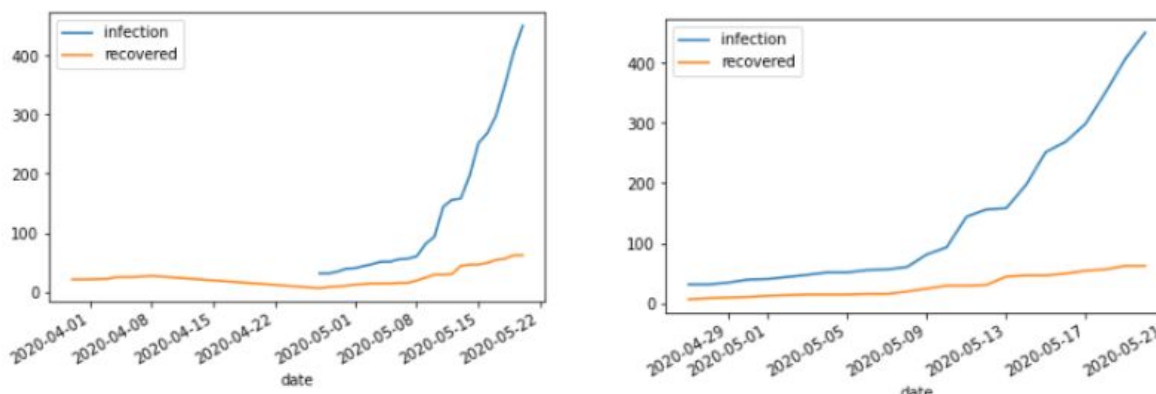
	lab	percent	percent
count	8.000000	8.000000	8.000000
mean	24435.250000	1.384150	1.384150
std	4905.760805	0.214163	0.214163
min	18887.000000	1.043045	1.043045
25%	20070.250000	1.281725	1.281725
50%	24457.500000	1.374841	1.374841
75%	28812.500000	1.453544	1.453544
max	29921.000000	1.771331	1.771331

Соотношение числа тестов и количества выявленных случаев nCoV дают возможность оценить масштаб заболеваний среди групп “риска” охваченных скринингом (контактные, въезжающие, медперсонал, госслужащие, и тд.)

**На 21 мая, средний процент заболеваемости среди этой группы составил 1,4% (всего населения 0.12%)**

## Выздоровевшие

Учёт выздоровевших (recovered) в Туве начался с момента начала работы оперативного штаба, за месяц до появления больных. Для точной оценки я исключаю эти данные, и возьму во внимание лишь период с 2020-04-27, когда появились первые подтвержденные случаи 2019-nCoV.



На 21 мая суммарное количество выздоровевших **922** человек, с момента первых инфицированных **730** чел. Средний процент выздоровевших от числа инфицированных составляет 23.8%. Следует обратить внимание, что эти данные кумулятивные, то есть накопительного характера по отношению к карантинным срокам.



Процентное соотношение выздоровевших к больным (per) снижается. Связано это с тем, что с начала точки пересечения (“проценты” + “выздоровевшие”) прошло всего 7 дней, что в контексте карантинных сроков недостаточно для перевода людей в разряд “выздоровевших”



## Прогнозы

### Отказ от ответственности

Как показывает практика, прогнозная оценка 2019-nCoV сопряжена с рядом трудностей обусловленных клинической спецификой вируса, чувствительного к целому ряду региональных особенностей, порождающих большое количество неопределенностей. На сегодняшний день, не существует методик, способных с высокой долей достоверности предсказать эпидситуацию в дальнесрочной перспективе. Компроментальные математические модели SAR используемые для расчета распространения эпидемий, могут дать хорошие результаты только в краткосрочной перспективе.

**Внимание! Прогнозы и методы представленные в данном отчете не могут являться источником достоверных данных!** Автор лишь пытается определить наиболее эффективные методы анализа. Кроме того, отсутствие необходимого набора данных могут сильно исказить картину расчётов.

### Входящие данные

В качестве входящих данных используется новостные сводки с извлеченными данными: **date**-дата, **infection** - кол-во инфицированных, **lab** - лаб. исследований, **recovered** - выздоровевших, **dead** - умерших, **news** - текст новости. Кроме этого, будут использоваться ранее осуществленные расчёты, в контексте текущей ситуации, а именно на 19 мая 2020 года.

date		infection	lab	recovered	dead
2020-05-21 00:00:00	2020-05-21 00:00:00	530.000000	29921.000000	63.000000	1.000000
2020-05-20 00:00:00	2020-05-20 00:00:00	451.000000	29342.000000	62.000000	1.000000
2020-05-19 00:00:00	2020-05-19 00:00:00	407.000000	28636.000000	62.000000	1.000000
2020-05-18 00:00:00	2020-05-18 00:00:00	351.000000	27943.000000	56.000000	1.000000
2020-05-17 00:00:00	2020-05-17 00:00:00	299.000000	20972.000000	54.000000	1.000000
2020-05-16 00:00:00	2020-05-16 00:00:00	269.000000	20250.000000	49.000000	1.000000
2020-05-15 00:00:00	2020-05-15 00:00:00	252.000000	19531.000000	46.000000	1.000000
2020-05-14 00:00:00	2020-05-14 00:00:00	197.000000	18887.000000	46.000000	1.000000
2020-05-13 00:00:00	2020-05-13 00:00:00	158.000000	nan	44.000000	1.000000
2020-05-12 00:00:00	2020-05-12 00:00:00	156.000000	nan	30.000000	nan

Для того, чтобы сформировать набор моделей, необходимо получить ряд ключевых значений.

### Пояснения

**R0 - базовое число репродукции** - это число показывает нам о том скольких людей примерно заражает один инфицированный, является константой и

специфична для каждого конкретного региона в зависимости от плотности населения и других особенностей региона. 2019-nCoV имеет градацию  $R_0$  от 1.94–5.7. **Средний показатель для России  $R_0$  2-2.5** Данная цифра может заметно отличаться как территориально, так и в конкретный момент времени.

Специалисты рекомендуют отслеживать данный показатель, чтобы корректировать национальные координационные планы по борьбе с коронавирусом. В Тыве данный показатель не отслеживается, так как Роспотребнадзор РФ не предъявил необходимых рекомендаций и методов расчета. Получить значение максимально точное значение  $R_0$  - крайне сложная задача, так-как следует принимать во внимание большое количество входящих данных, в том числе связанных с карантинными мерами, а также метриками неподдающихся учету. **SIR модель** - компроментальная модель развития эпидемиологических ситуаций

## Анализ

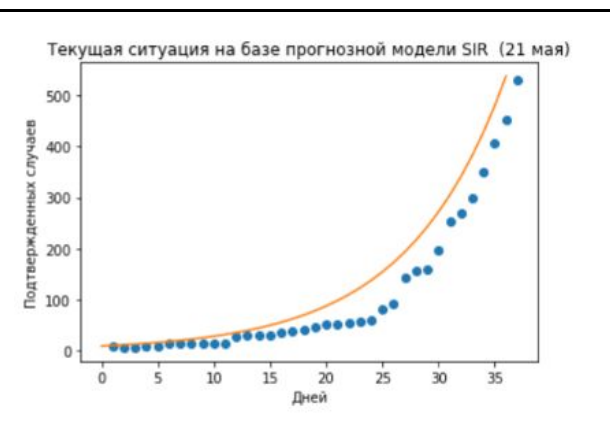
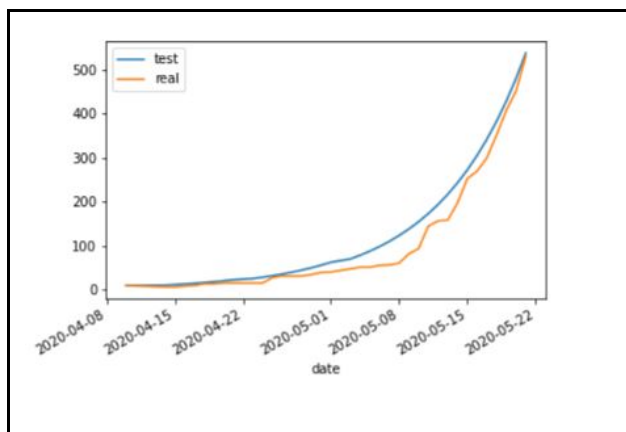
Поиск параметров beta, gamma



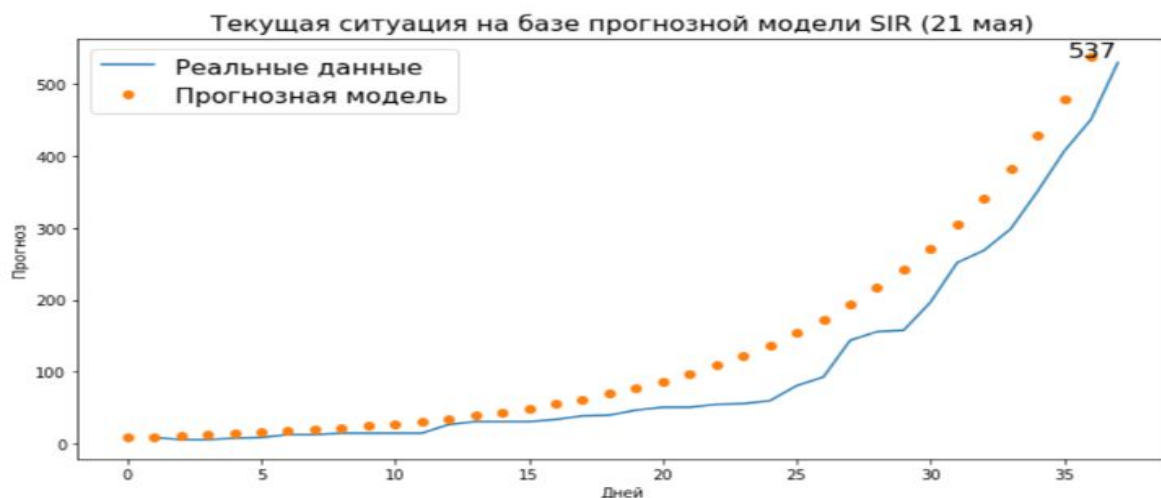
Тестирование данных на модели SIR методом *scipy.optimize.curve\_fit*

Оптимальные параметры получены (реверс) : **beta (скорость перехода S->I) = -1.11** and **gamma (скорость перехода I->R) = -1.25**

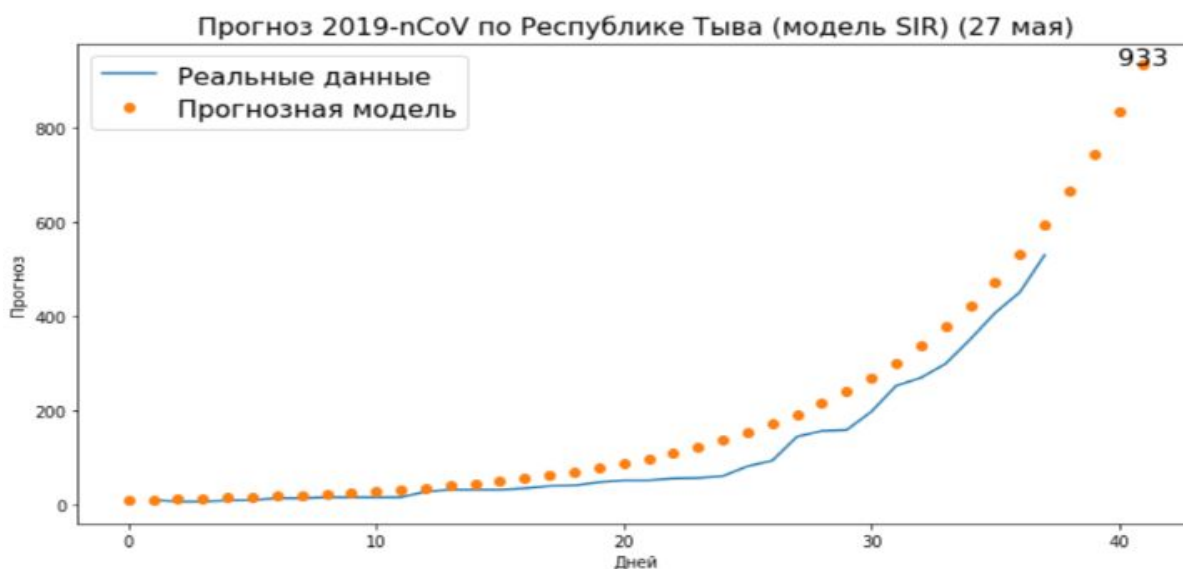
В результате многократного тестирования модели, были внесены поправки  $\beta$  +-, изучено их влияние на точность результата модели. **Таким образом, была достигнута приемлемая точность в краткосрочных прогнозах 98.7%**



### Текущая ситуация на базе прогнозной модели SIR (21 мая)



### Прогнозная модель SIR на 27 мая 933 заболевших



### Прогнозная модель SIR на 1 июня 1622 заболевших

