

Презентации дипломной работы

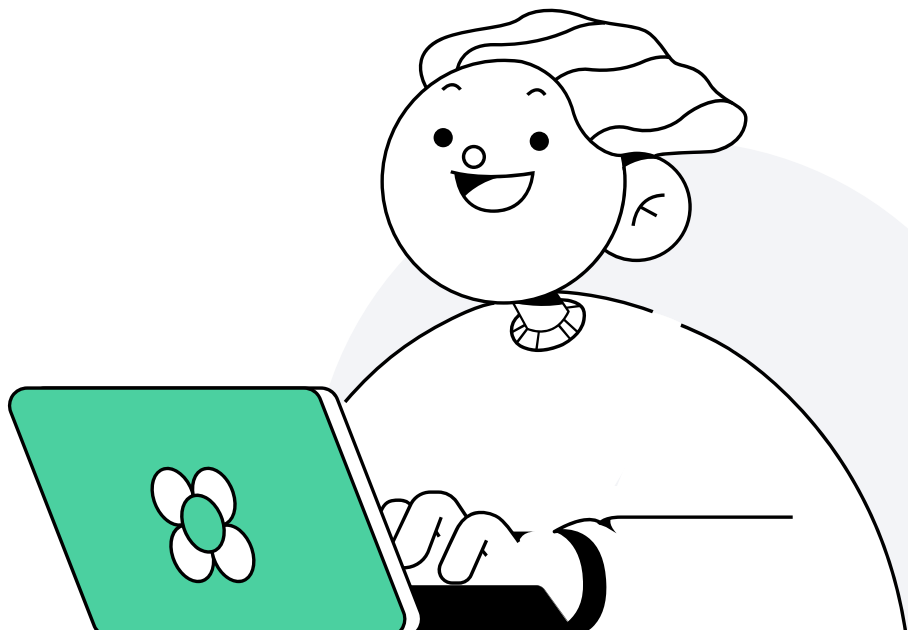
Дипломная работа по программе «Data Science»: Рекомендательные системы

Соломин Н. Н.
Группа: DS-CP-1



Структура презентации

- 1 Описание данных
- 2 Описание проведенного исследования
- 3 Результаты проверки гипотез
- 4 Выводы и рекомендации



Описание данных



1

Описание данных

Анализ данных крупного агрегатора такси (поиск инсайтов, проверка гипотезы, составление рекомендаций стейкхолдерам).

Датасет содержит информацию о поездках на такси одного из известных агрегаторов:

- `order_gk` — идентификатор поездки;
- `weekday_key` — день недели совершения поездки;
- `hour_key` — час совершения поездки;
- `distance_km` — дистанция поездки;
- `offer_class_group` — класс поездки (Economy, Comfort, Premium, Delivery);
- `driver_response` — статус поездки (1 — поездка совершена; 0 — отмена).

	order_gk	weekday_key	hour_key	distance_km	offer_class_group	driver_response
0	412	7	23	53.114	Premium	1
1	436	3	10	5.781	Delivery	1
2	1327	6	19	16.200	Delivery	0
3	2013	2	2	13.732	Delivery	0
4	2241	6	3	20.436	Premium	1
5	2601	4	8	46.301	Premium	0
6	2710	4	19	2.142	Comfort	1
7	2911	7	11	42.588	Premium	1
8	2948	6	23	67.540	Premium	1
9	3013	2	16	33.417	Delivery	1



Описание проведённого исследования



2

Описание проведенного исследования

Рассчитали основные описательные статистики, из результатов можно сделать вывод, что данные чистые, не содержат пропущенных и нулевых значений

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2000 entries, 0 to 1999
Data columns (total 6 columns):
#   Column                Non-Null Count  Dtype  
---  -
0   order_gk               2000 non-null  int64  
1   weekday_key            2000 non-null  int64  
2   hour_key               2000 non-null  int64  
3   distance_km            2000 non-null  float64
4   offer_class_group      2000 non-null  object  
5   driver_response        2000 non-null  int64  
dtypes: float64(1), int64(4), object(1)
memory usage: 93.9+ KB
```

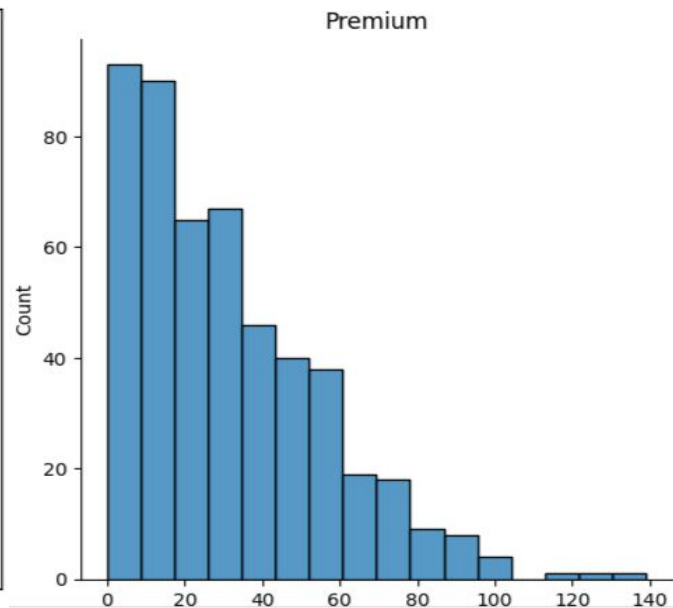
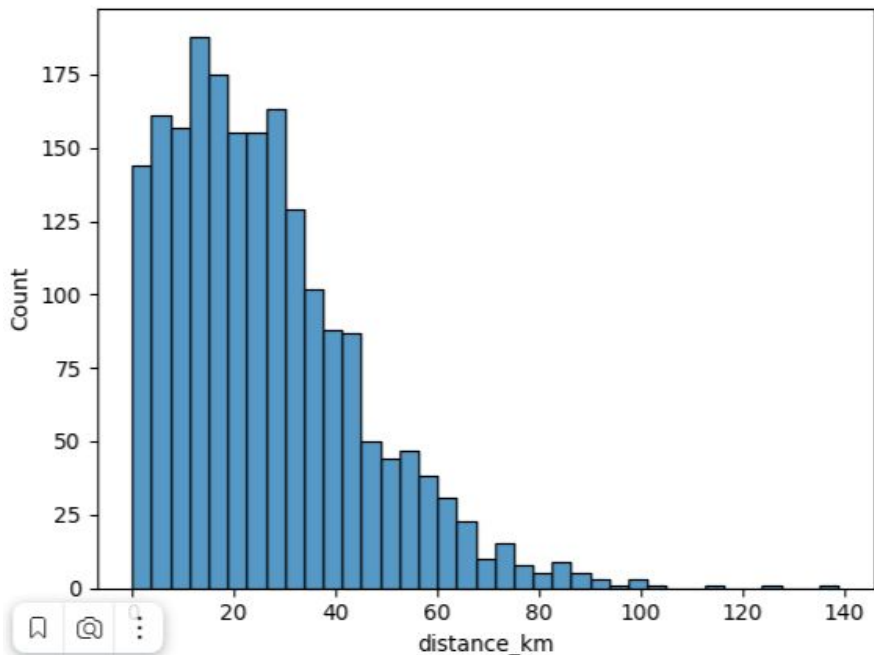
	order_gk	weekday_key	hour_key	distance_km	driver_response
count	2000.000000	2000.000000	2000.000000	2000.000000	2000.000000
mean	462807.384000	4.044000	11.608500	26.275848	0.587000
std	270556.211847	2.007256	6.918465	18.876336	0.492496
min	412.000000	1.000000	0.000000	0.024000	0.000000
25%	227126.500000	2.000000	6.000000	12.177000	0.000000
50%	458637.000000	4.000000	12.000000	22.828500	1.000000
75%	709420.250000	6.000000	18.000000	36.434250	1.000000
max	919196.000000	7.000000	23.000000	138.950000	1.000000



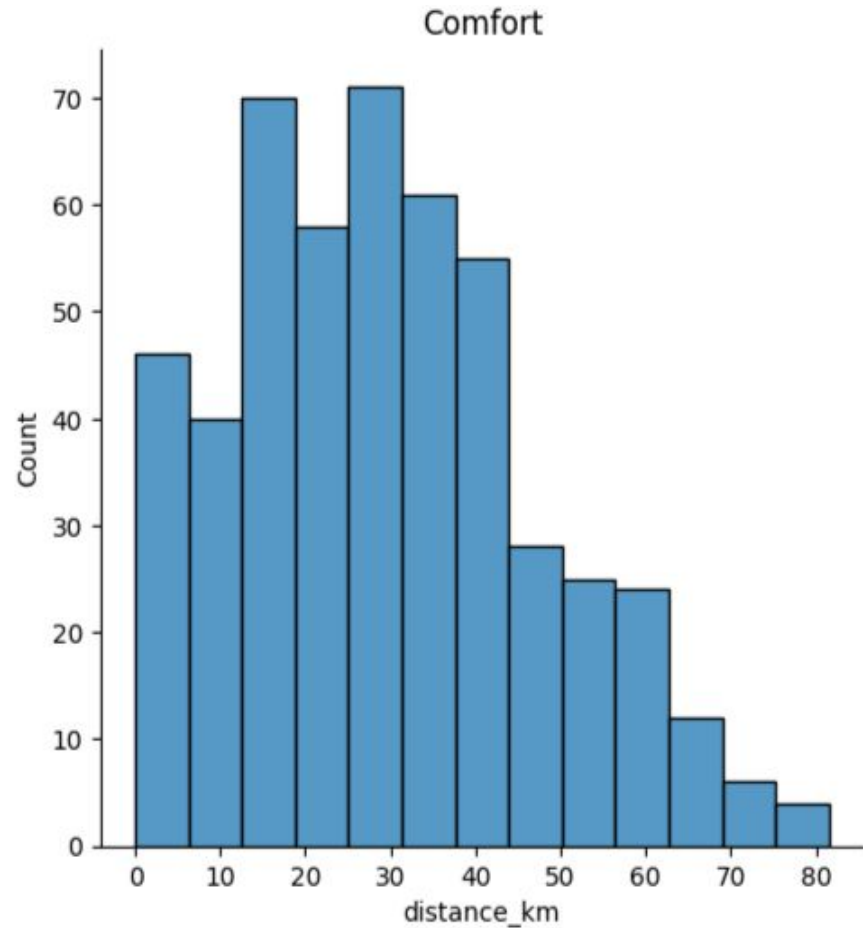
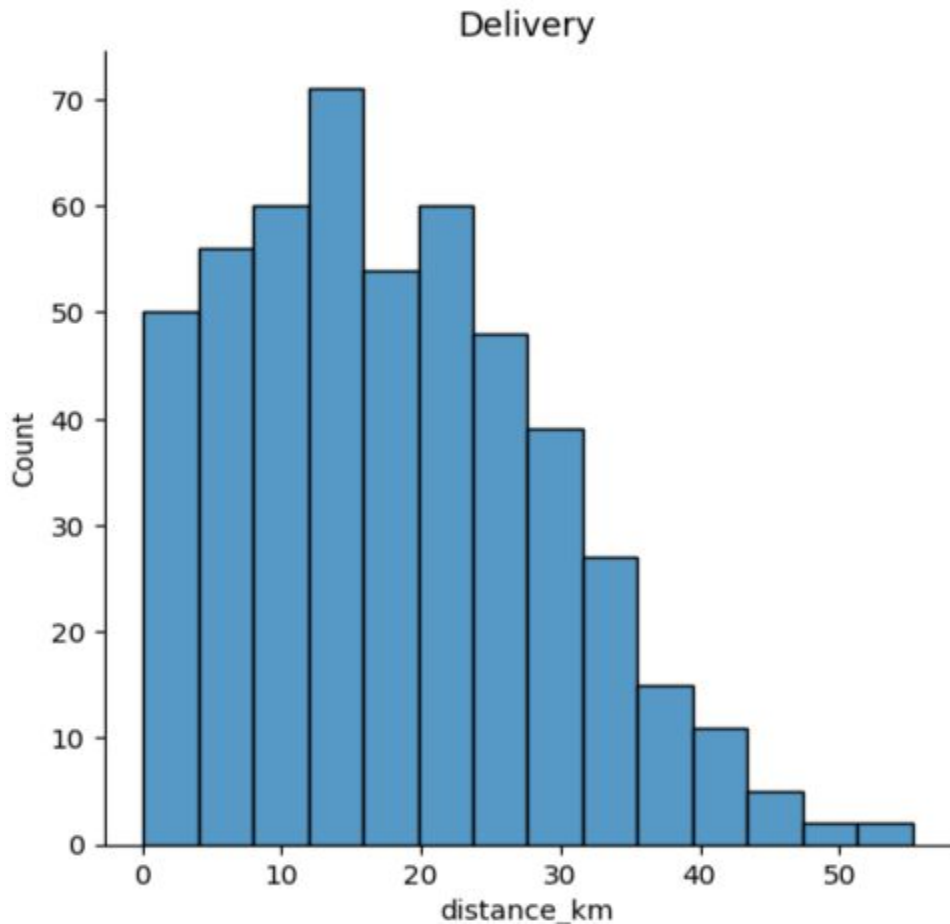
Описание проведенного исследования

Сформировали графики распределения дистанций поездок, общий, а также по каждому классу поездок отдельный.

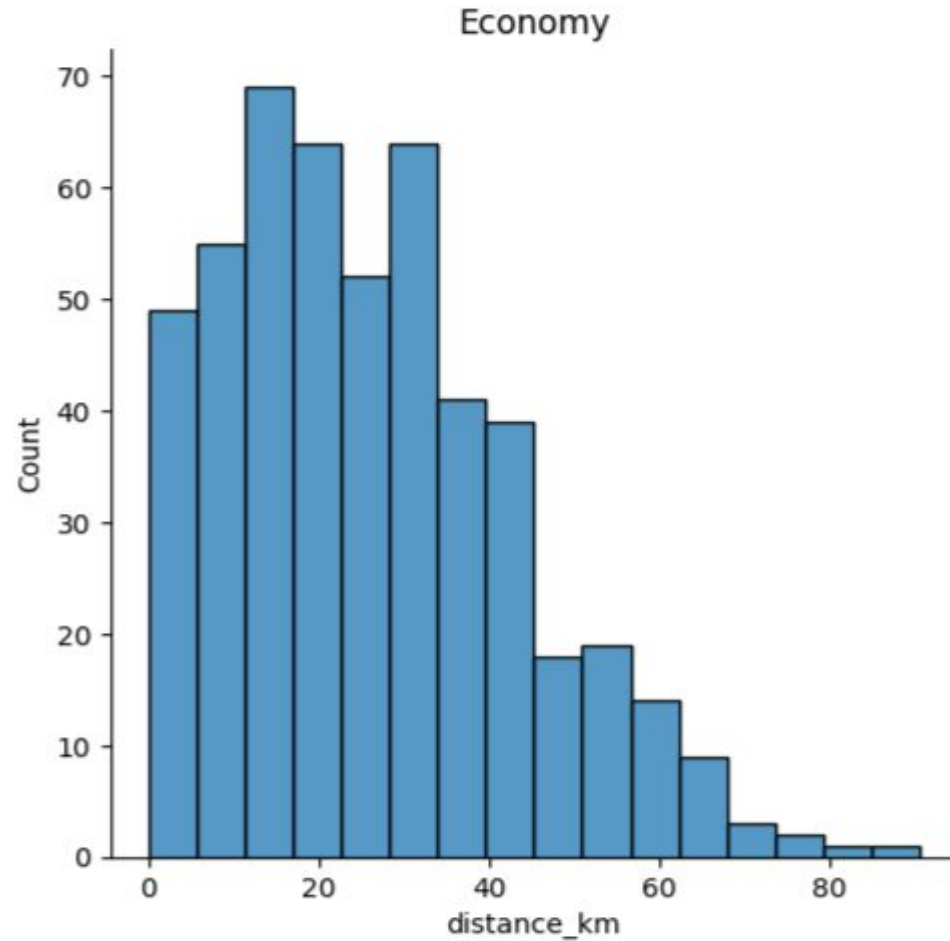
Согласно графикам, можно сделать вывод, что для комфорт-класса - распределение нормальное, а для эконом-класса - логнормальное



Описание проведенного исследования

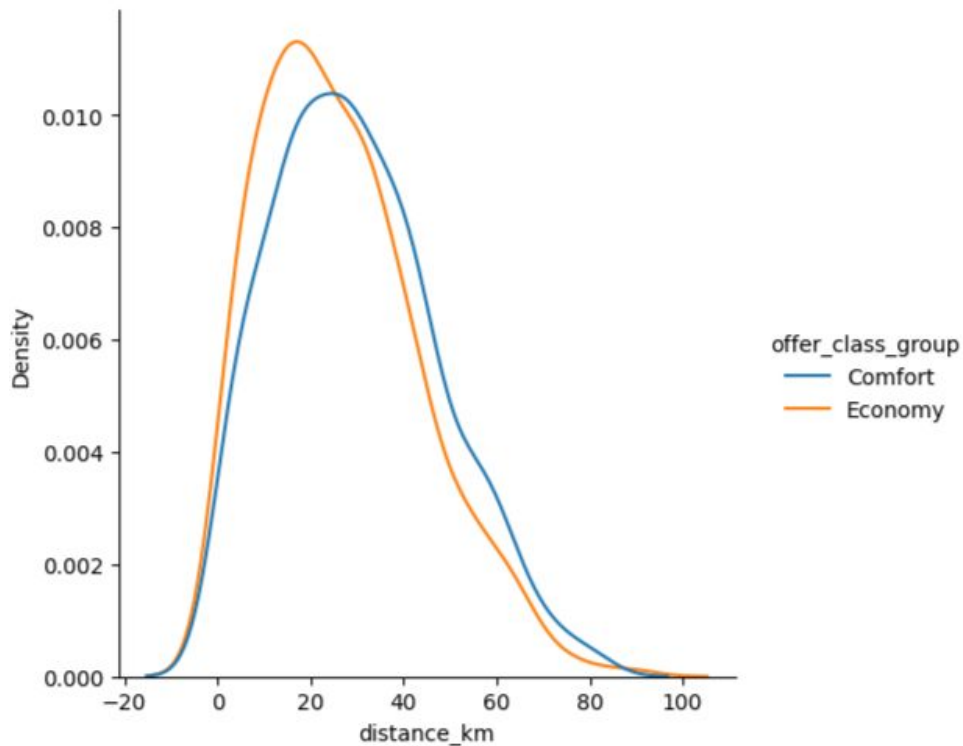


Описание проведенного исследования



Описание проведенного исследования

Сформировали график пересечений интервалов дистанций для эконом и комфорт - классам.



Описание проверки гипотез



3

Описание проверки гипотез

1. Сформировали выборки для эконом и комфорт - классов
2. Формулируем гипотезы:
H0 - Выборки не отличаются
H1 - Выборки отличаются
3. Проводим статистический тест для проверки гипотезы

```
1 t, p_value = ttest_ind(econ['distance_km'], comf['distance_km'])  
2 t, p_value  
  
(-3.1706426042389815, 0.0015673809987275479)
```

По результатам теста, мы видим, что значение вероятности равно 0.00157, что менее 5%.

Согласно этой информации, мы имеем основания отклонить нулевую гипотезу и принять альтернативную гипотезу.



Описание проверки гипотез

Вычисляем средние выборки исследуемых классов, с учётом полученного результата проверки гипотез.

Среднее значение в выборке Эконом - класса = **26.283098**

Среднее значение в выборке Комфорт - класса = **29.7589160000000003**

Из полученных результатов, следует сделать вывод, что на дальние поездки предпочтительно используют комфорт-класс



Выводы и рекомендации



4

Выводы и рекомендации

Выводы:

В среднем, дистанции поездок для комфорт и эконом - классов различаются

В поездках на дальние расстояния, предпочтительно используют комфорт-класс

Рекомендации:

Согласно проведенному исследованию, можно рекомендовать уделить повышенное внимание автопарку комфорт - класса, а также, в “часы-пик” увеличить количество автомобилей данного класса.



Построение модели заболевания COVID-19

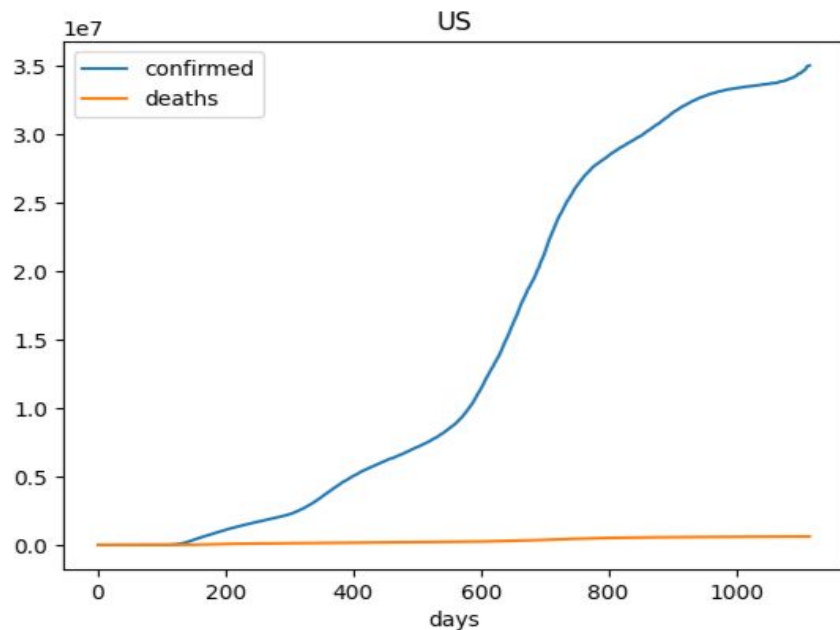
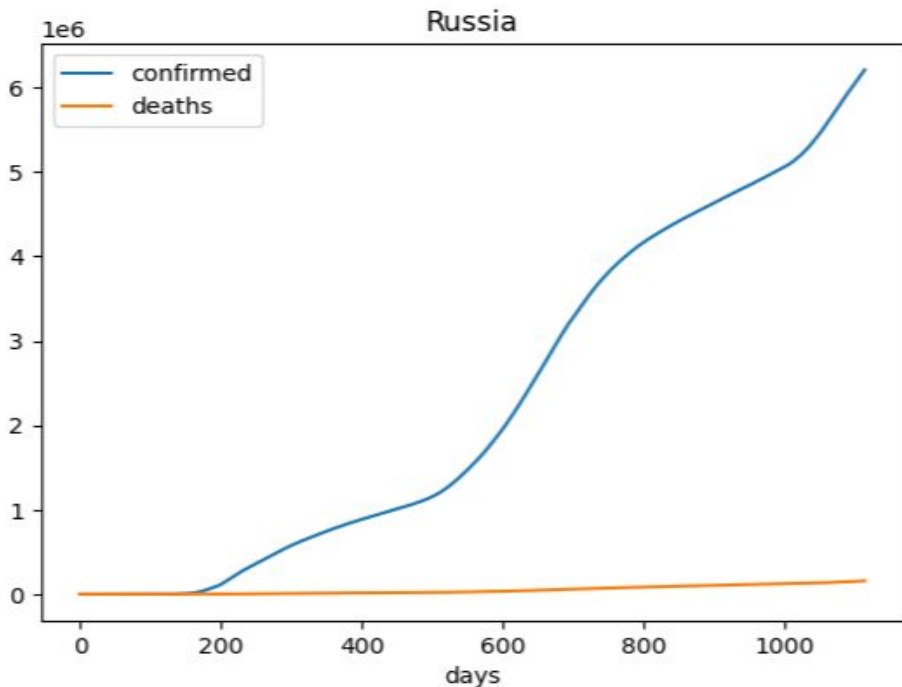


5

Описание проведенного исследования

Посмотрим графики заболевания и смертности по разным странам

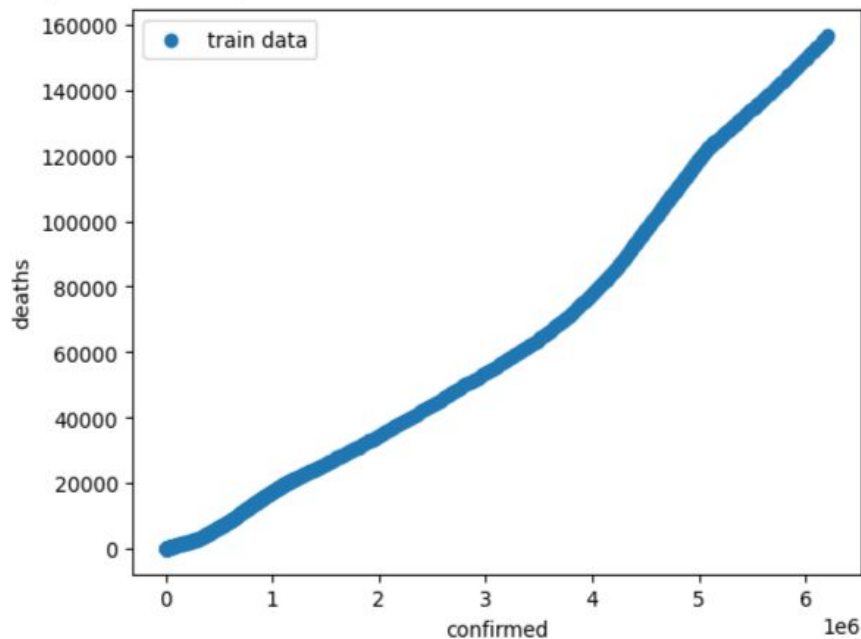
По графикам видим что заболевания растут с каждым днём интенсивно, а смертность почти не возрастает



Описание проведенного исследования

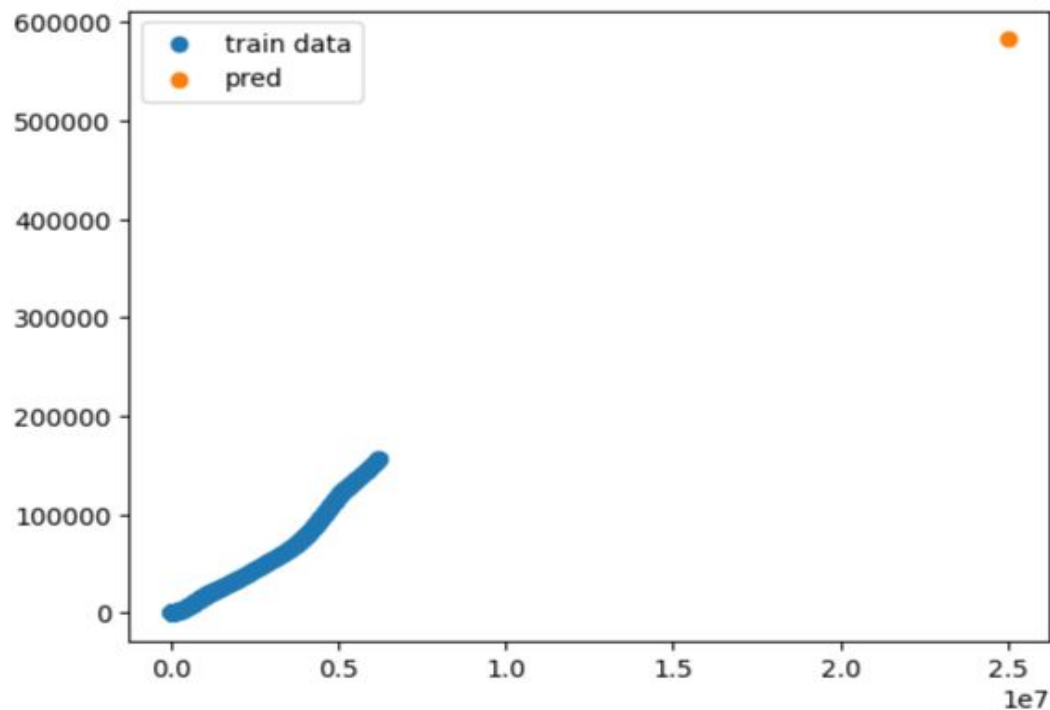
Построили модель линейной регрессии по смертям от заболеваемости и вычислили прогноз количества смертей в России для 25 000 000 случаев заражений, построили прогноз смертей для 25 000 000 заболевших.

Посмотрим на линейную связь между заболеваниями и смертностью



Описание проведенного исследования

По предсказанию модели на 2.5 миллионов заболевших будет всего 583 тысячи смертельных исходов, что составляет чуть более 2 % от числа заболевших



Рекомендации для бизнеса



6

Рекомендации для бизнеса

В условиях роста заболевания COVID-19 для агрегатора такси рекомендуем придерживаться предписанию врачей РФ. А также увеличить кол-во транспорта на доставку, т.к. люди следуя рекомендациям врачей будут больше находиться дома.



Спасибо за внимание!

