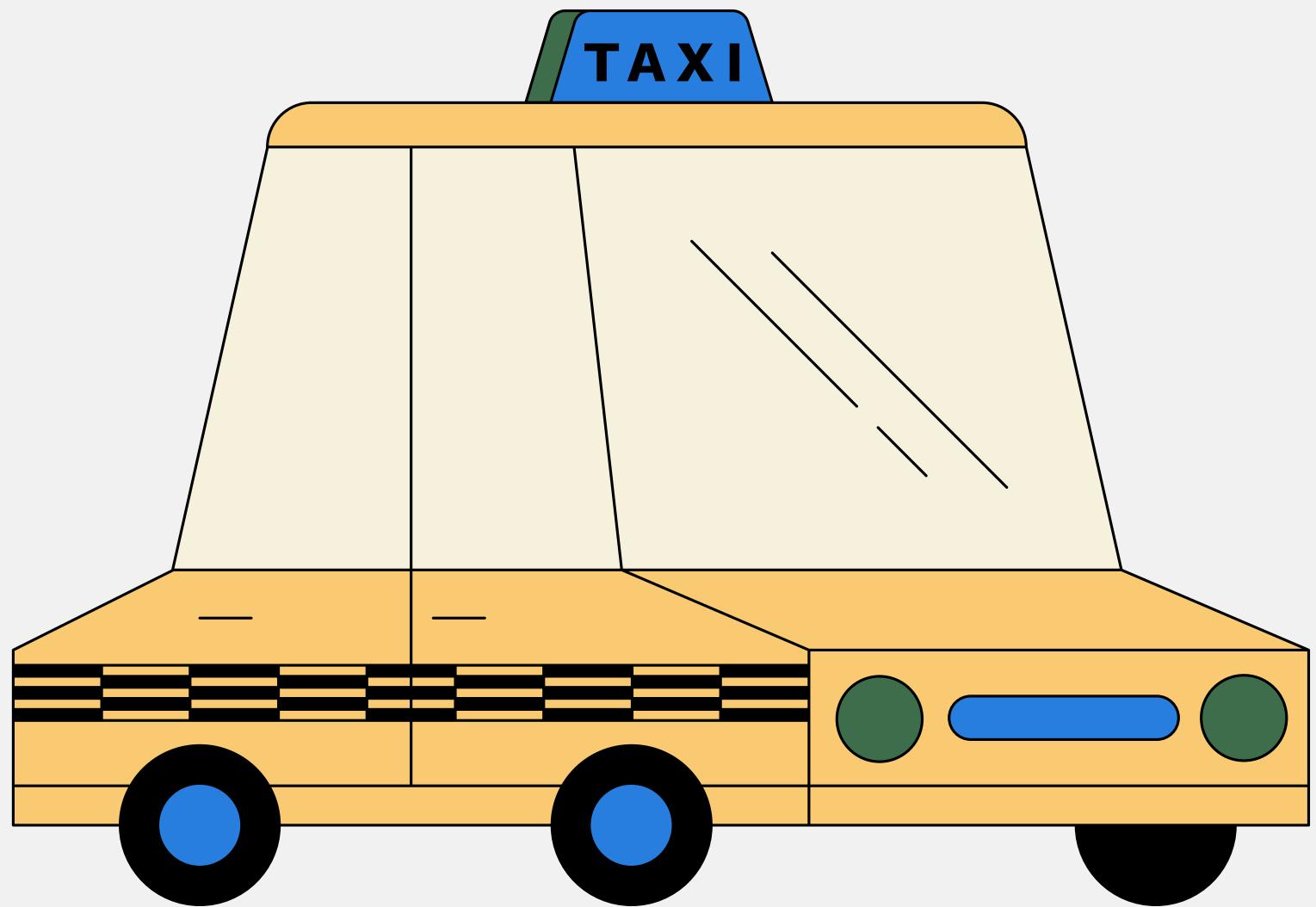


Страх и ненависть в нью-йоркском такси

- Фадеев Павел
- Райлян Данил
- Дьяков Иван



Проблематика и специфика данных



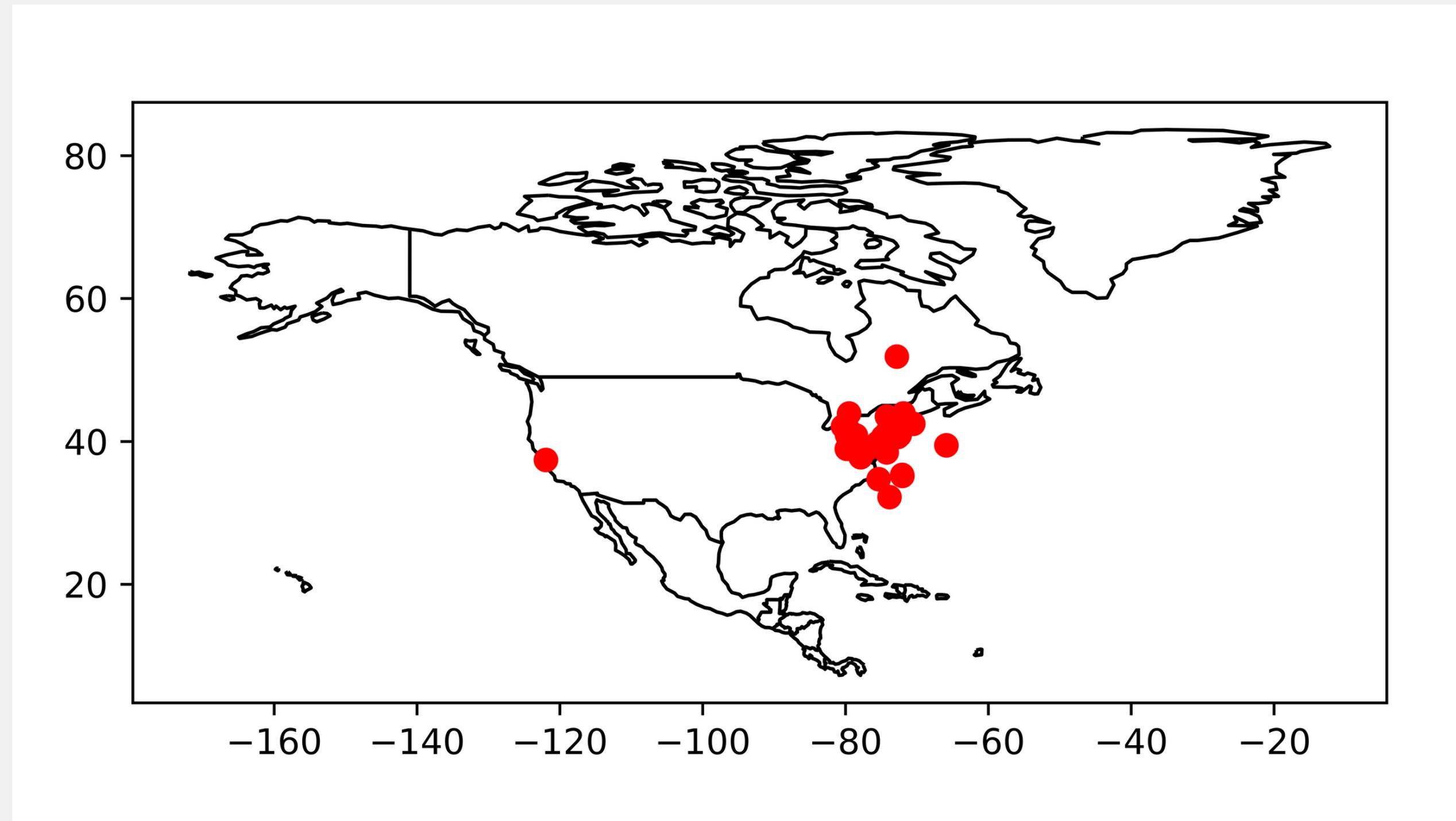
Необычные маршруты

Длинные поездки

Необычные временные интервалы

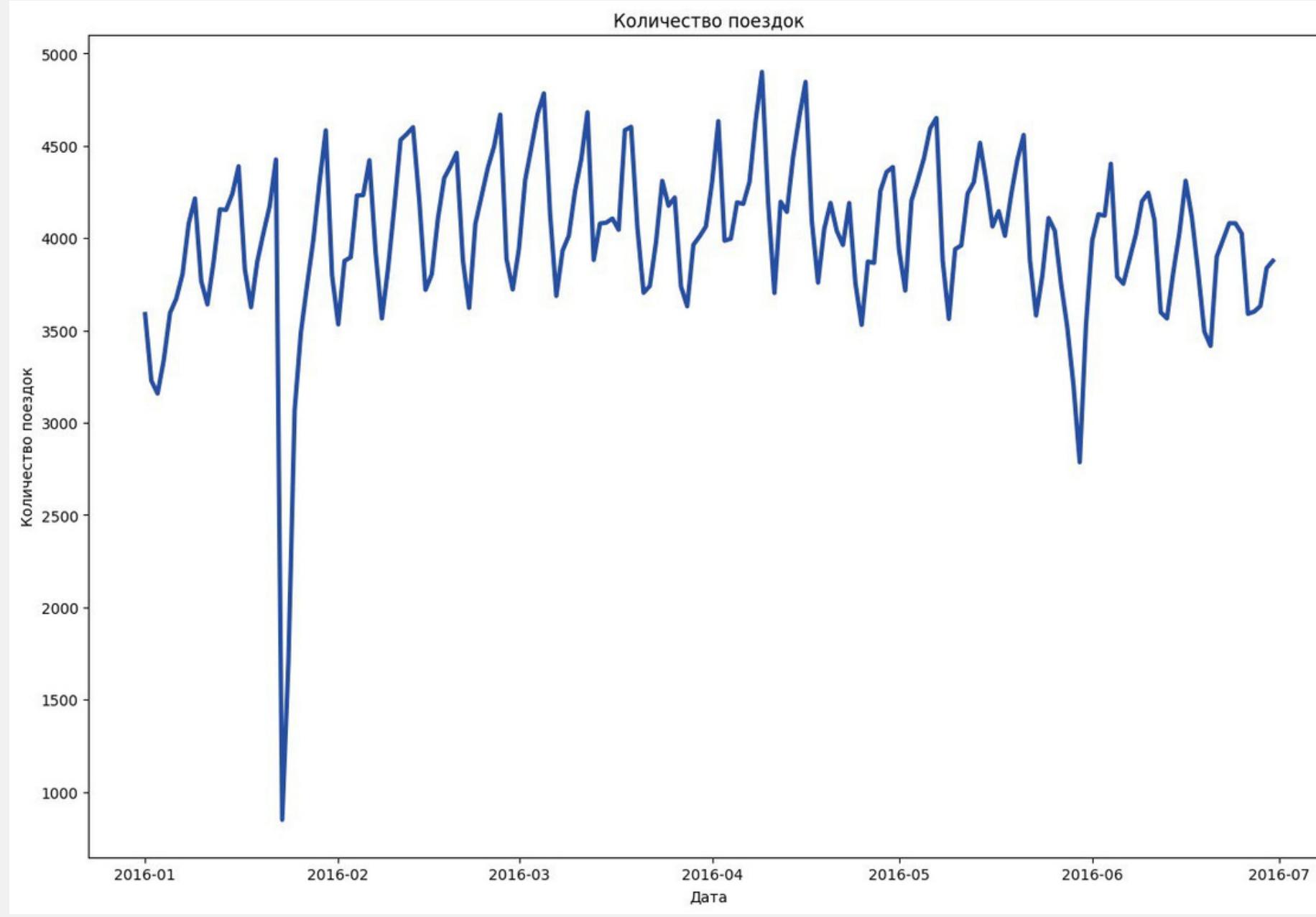
Аномальные дни

Предварительный анализ данных

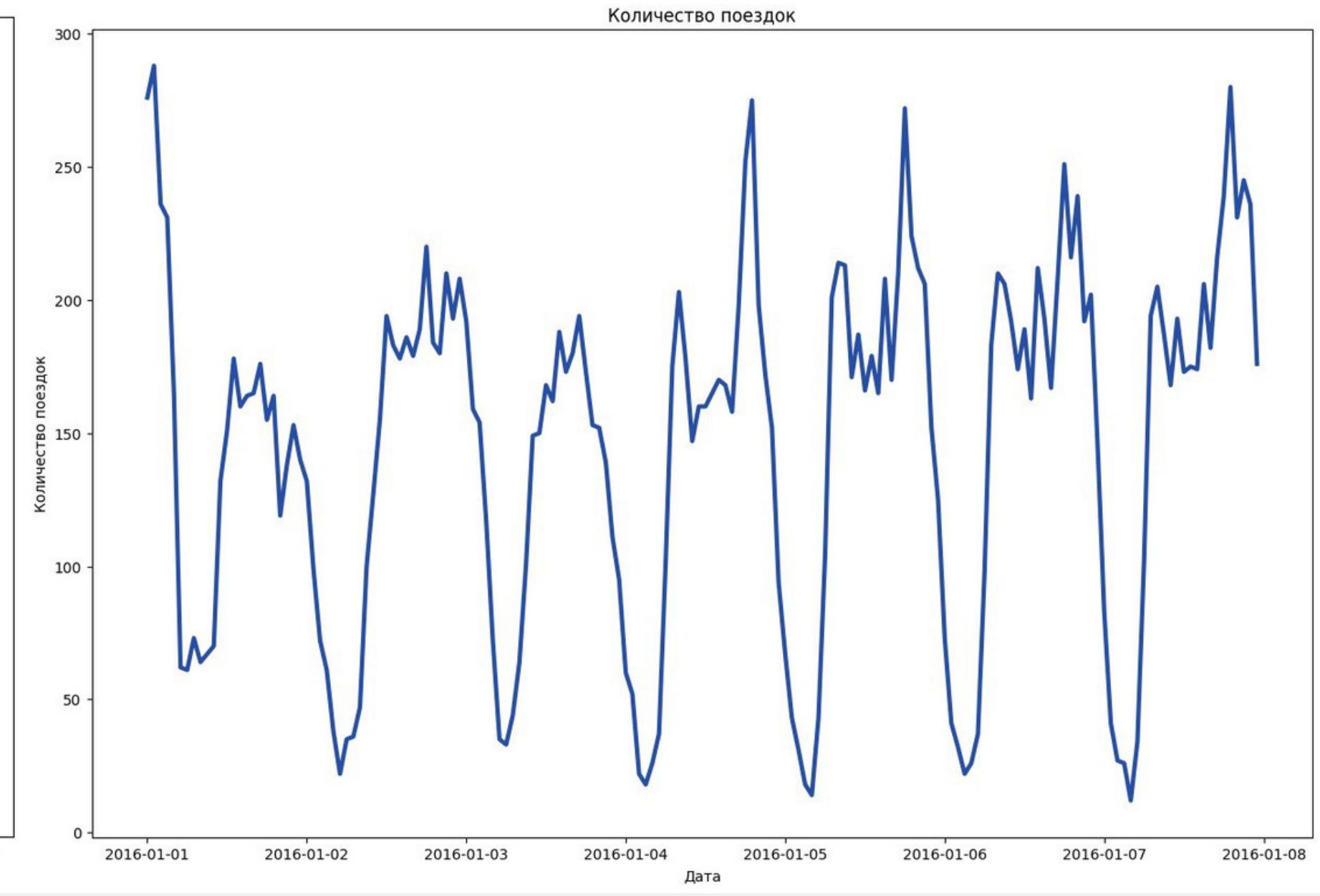


География поездок

Предварительный анализ данных

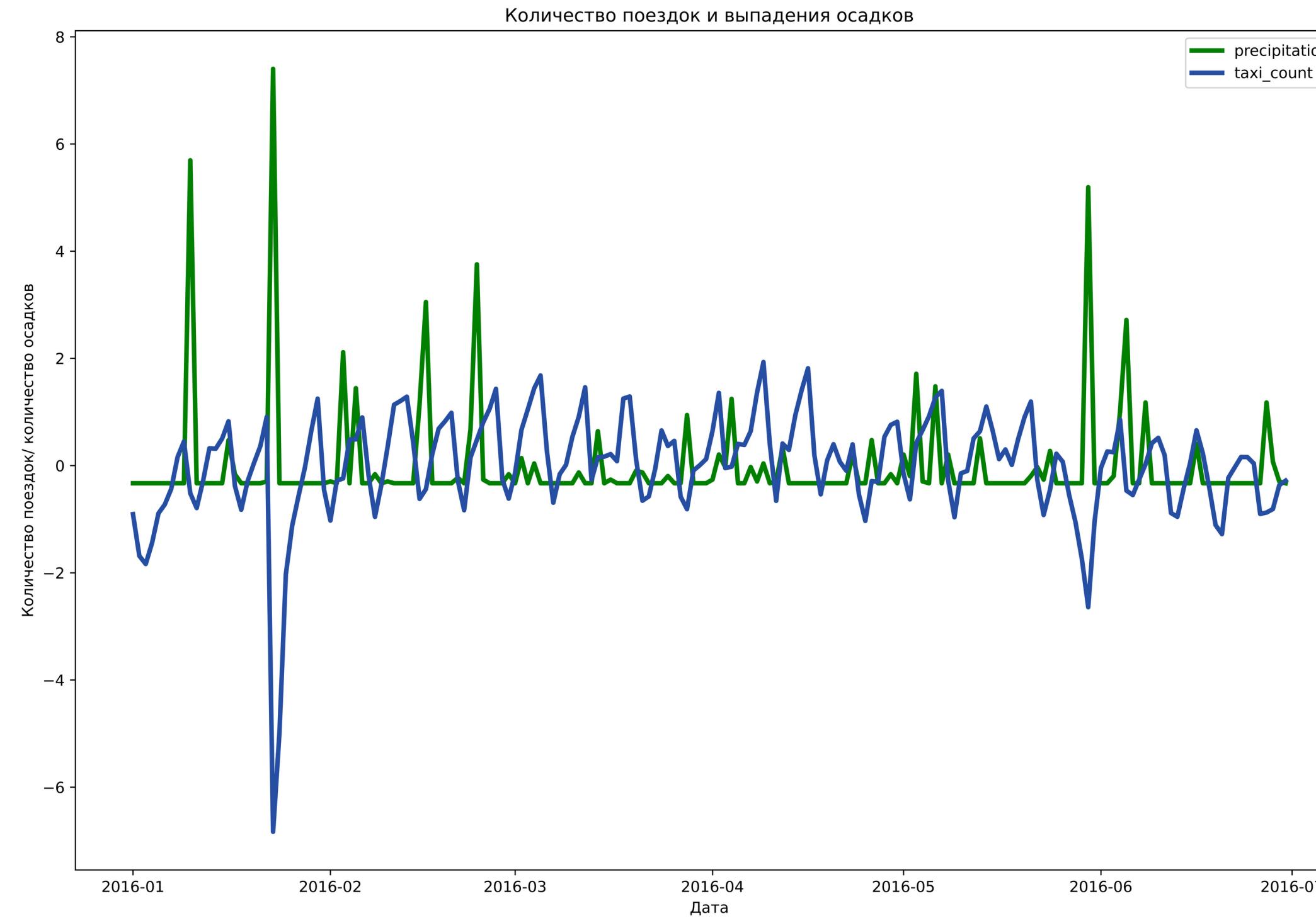


Количество поездок по дням



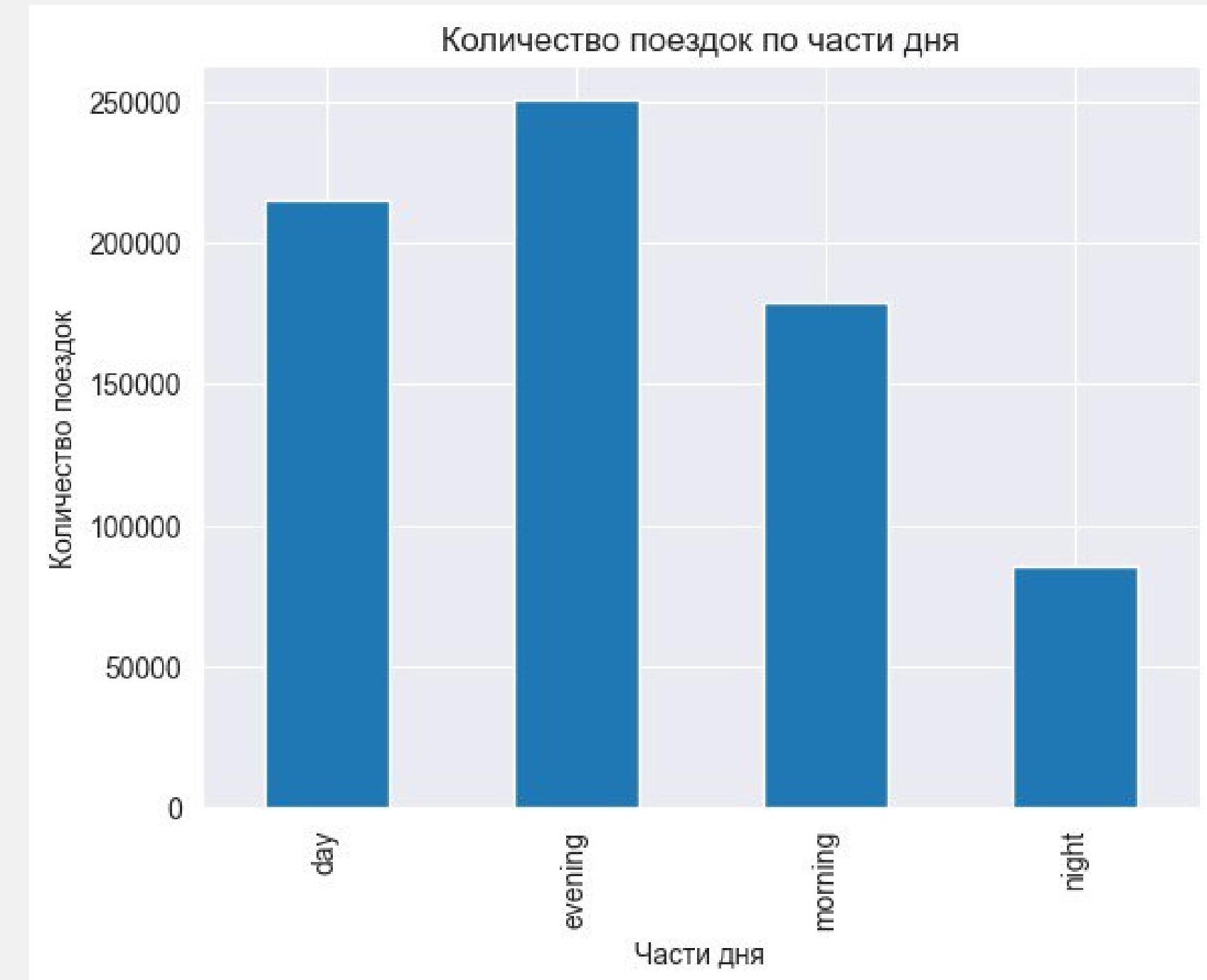
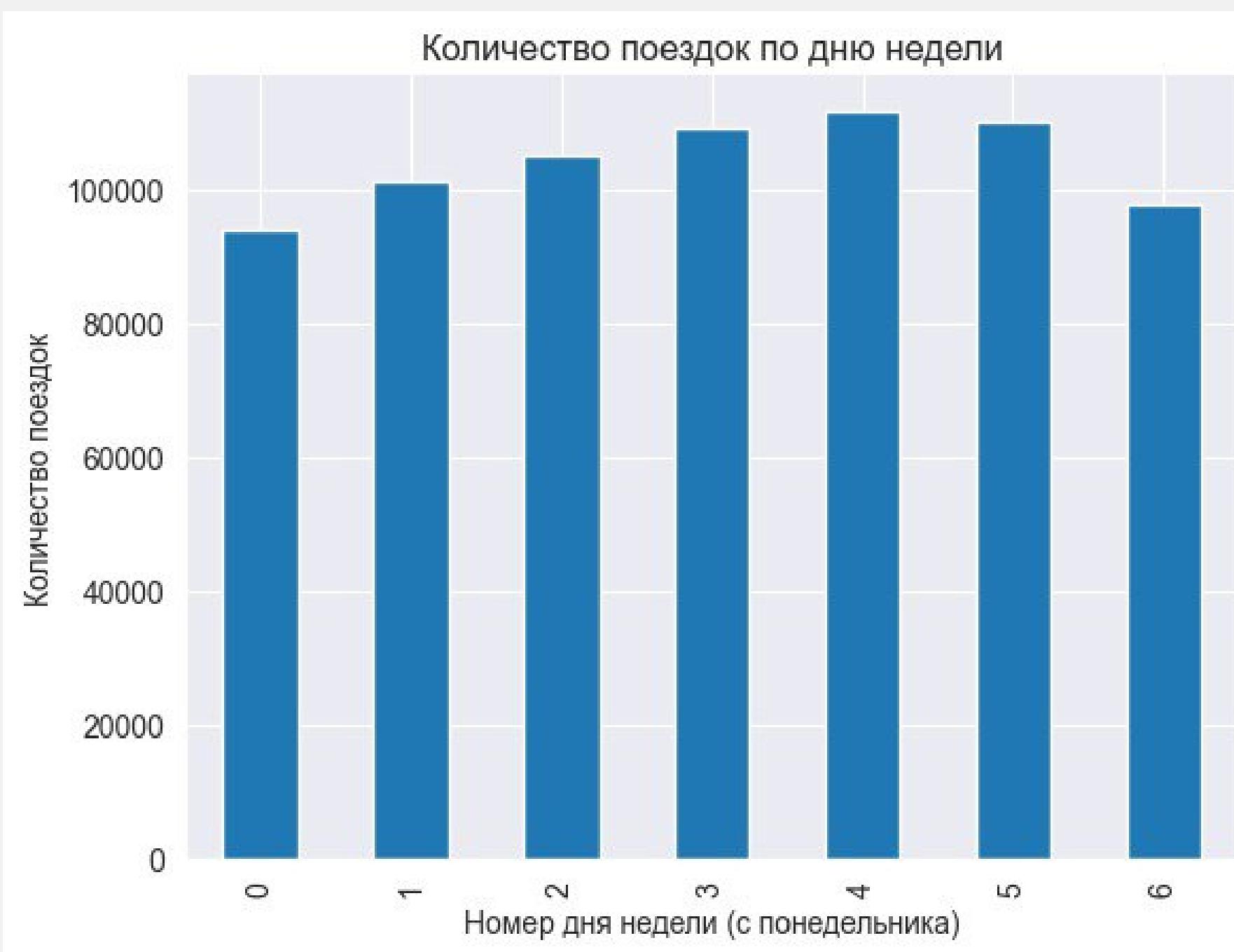
Количество поездок по часам

Предварительный анализ данных



Нормализованное количество поездок и выпадение осадков

Предварительный анализ данных



Метрики и агрегации

Метрики

Расстояние

Время поездки

Средняя скорость

Координаты

Количество пассажиров

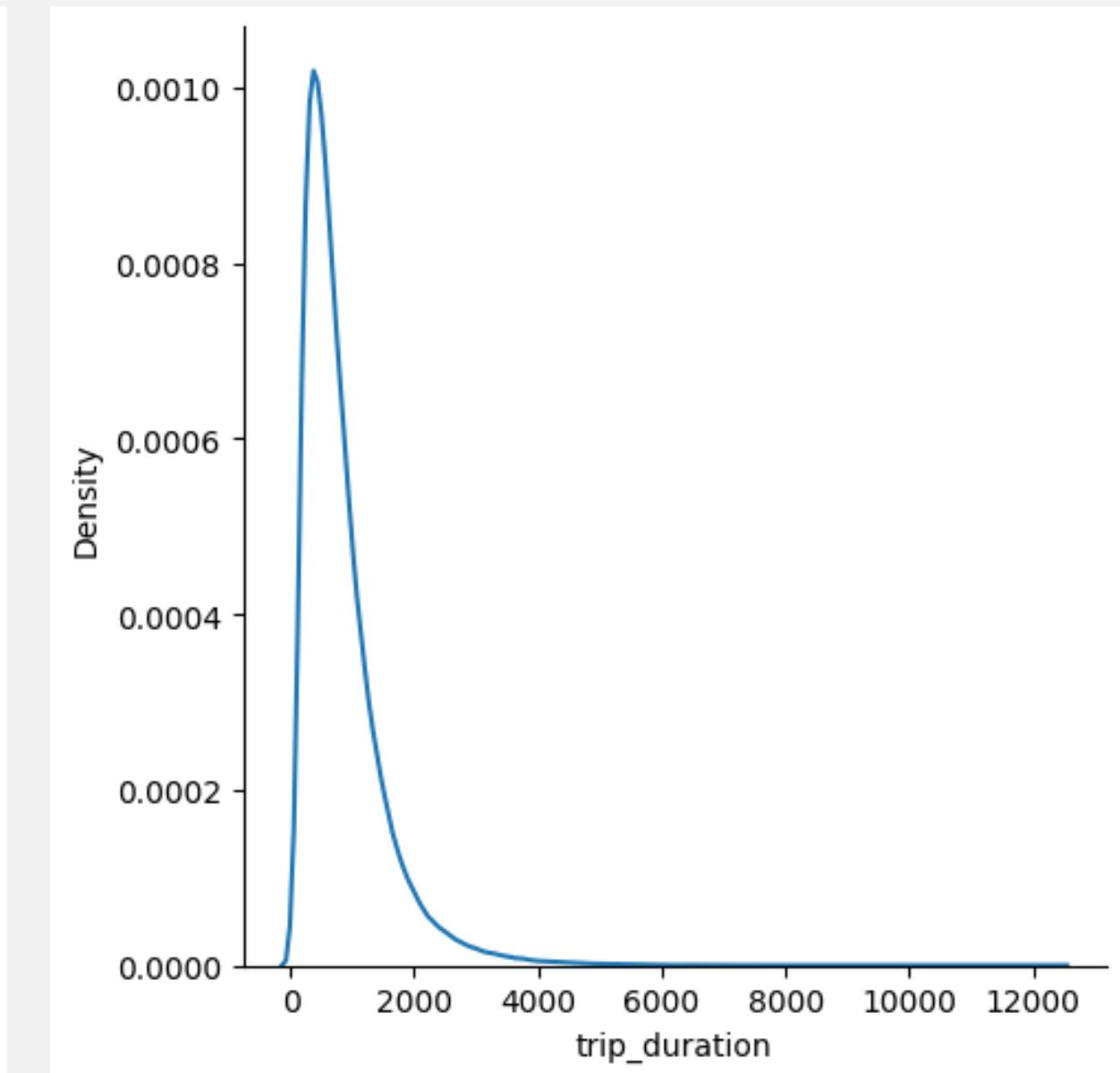
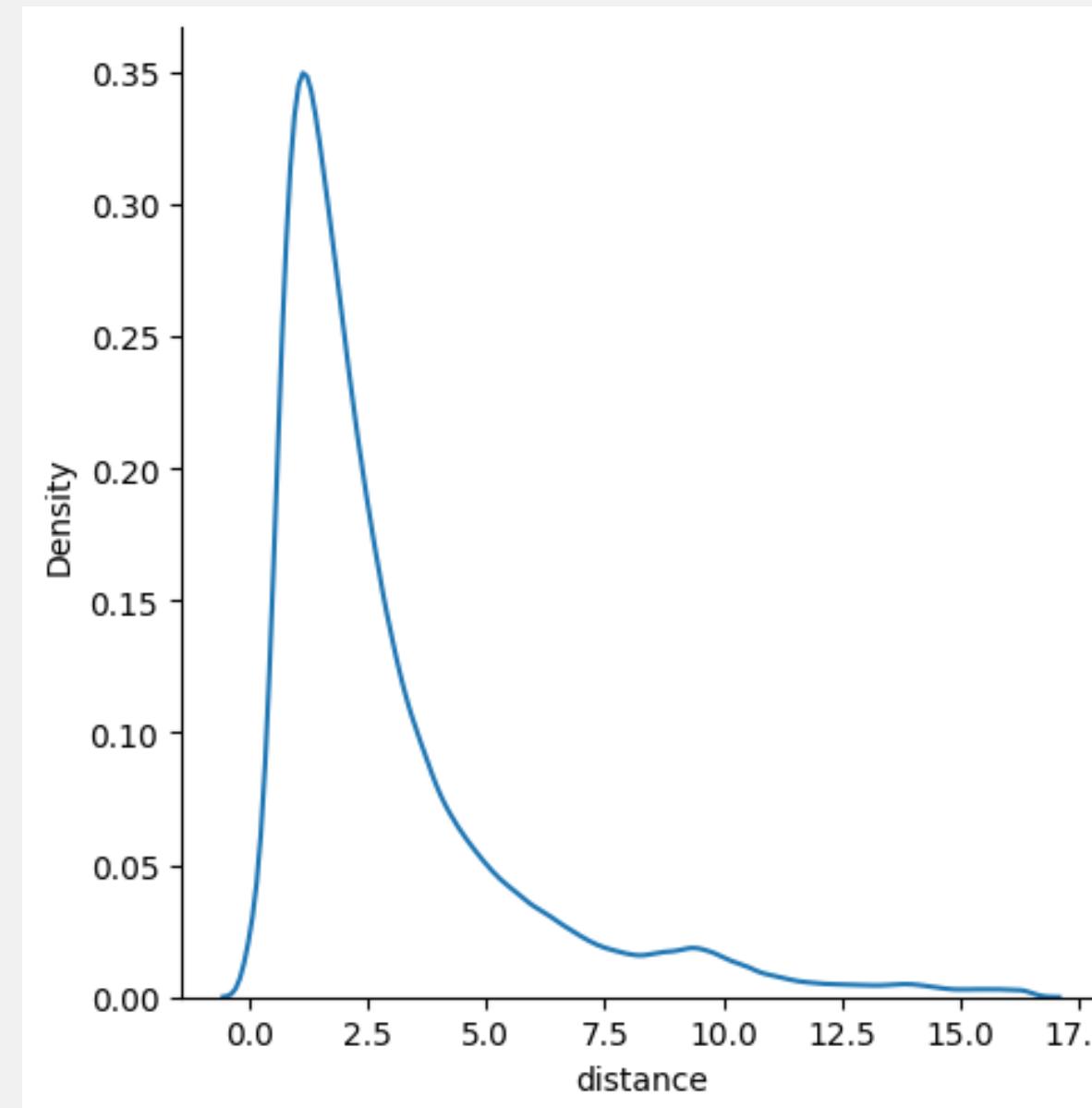
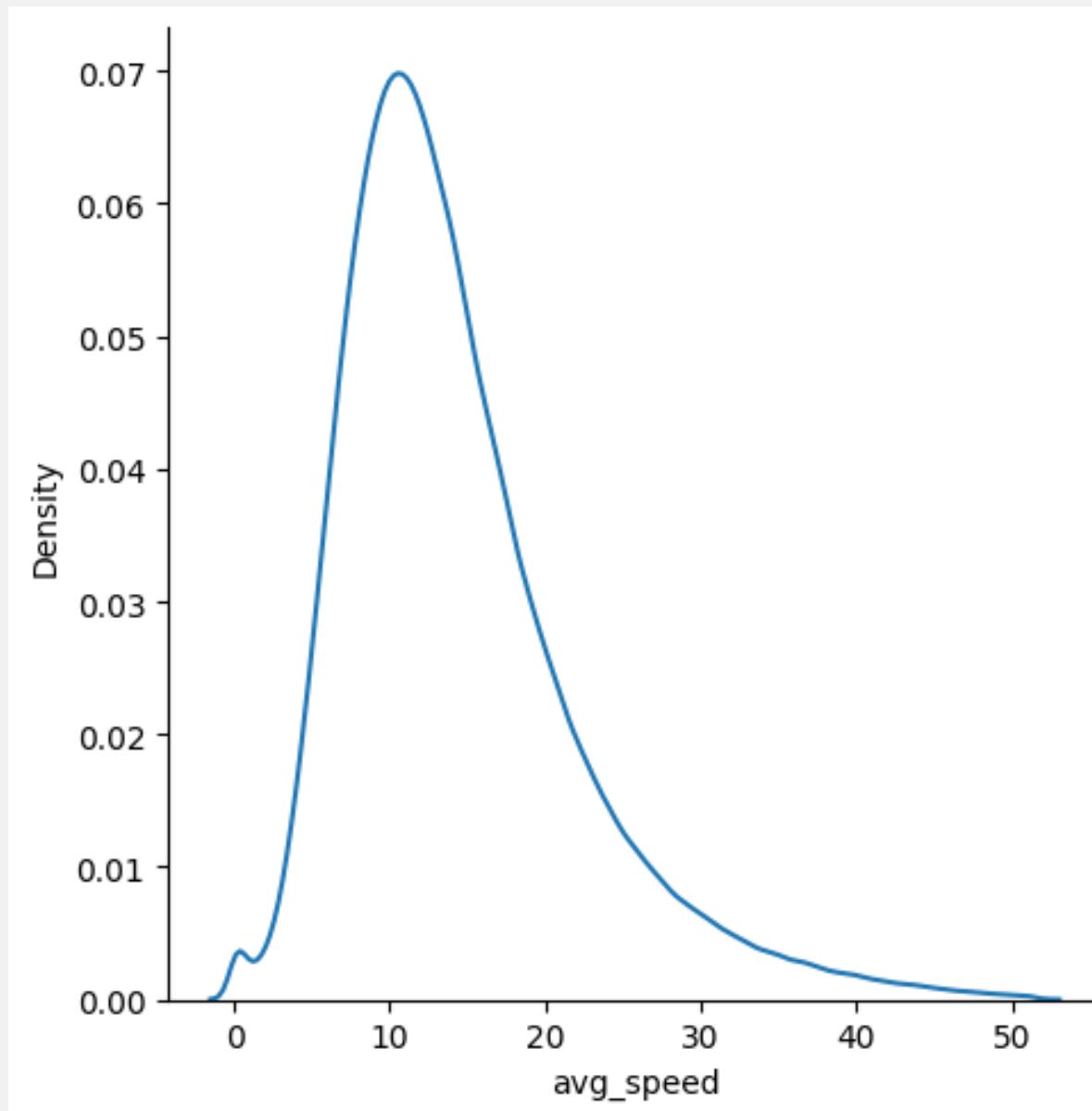
Агрегации

По дням

По дням недели

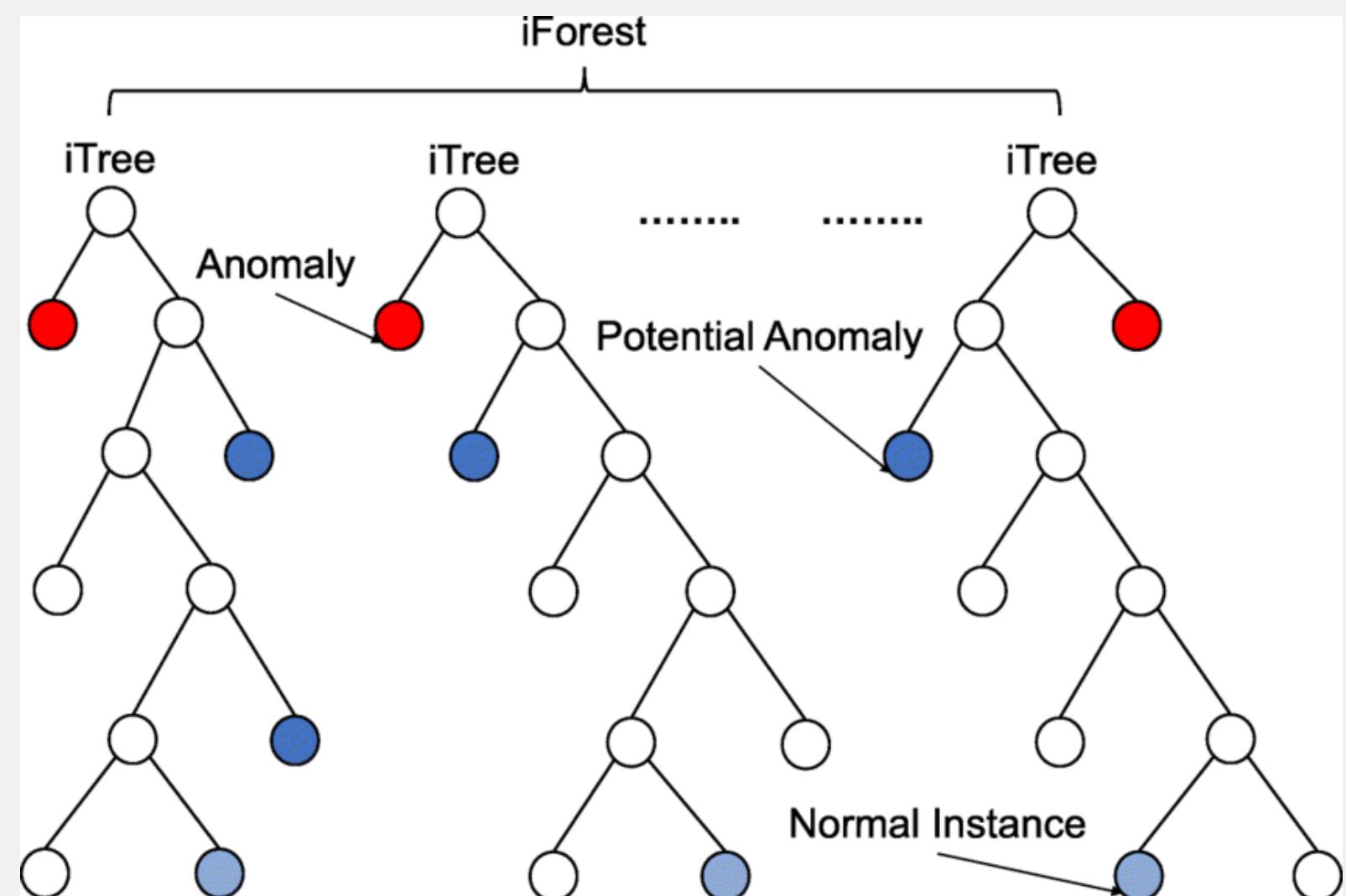
По времени суток

Распределение метрик



Инструменты

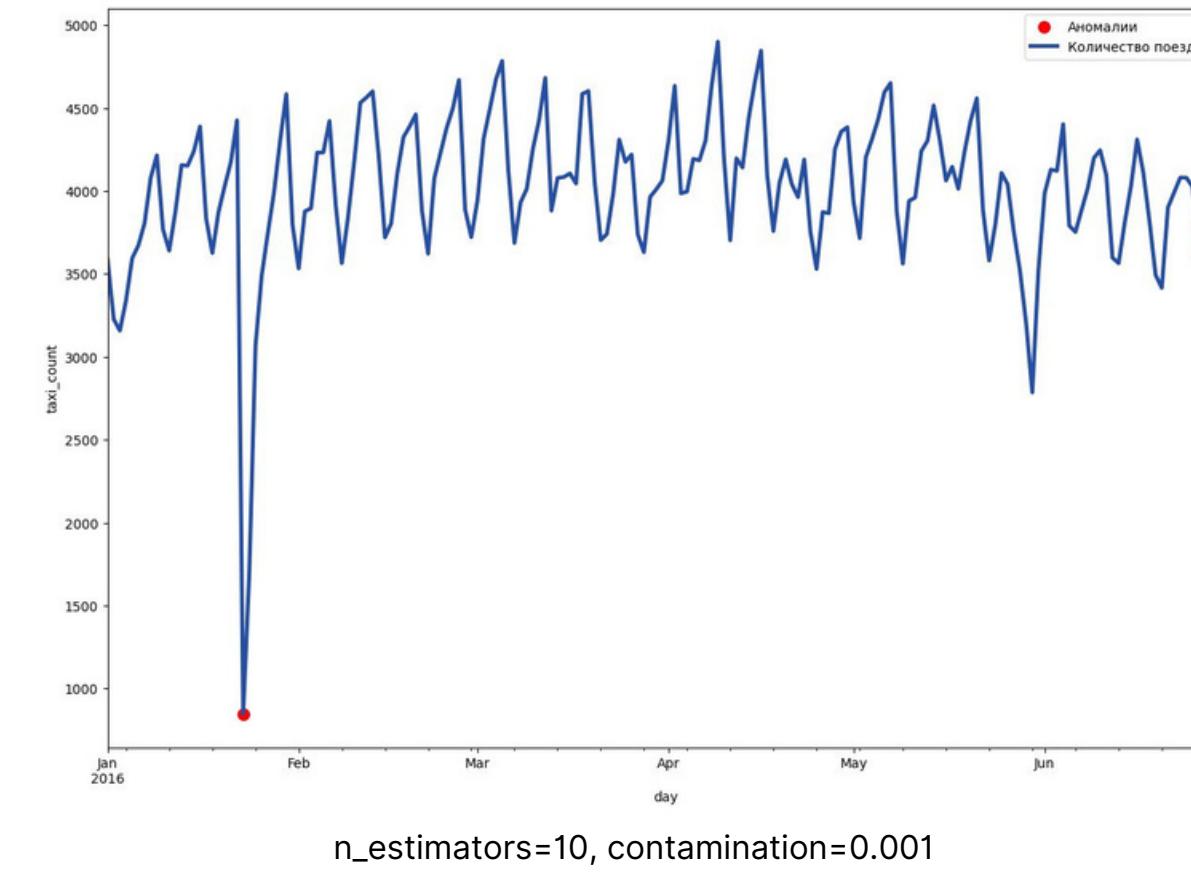
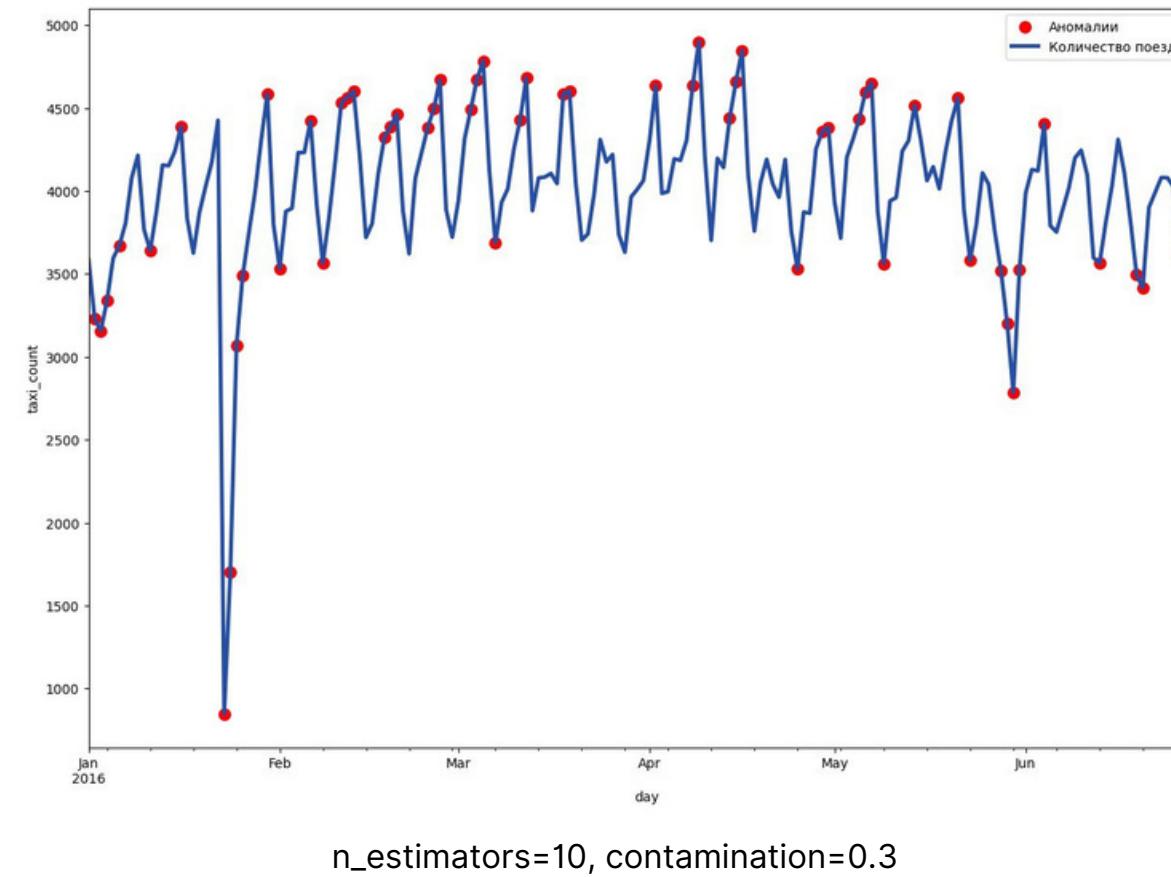
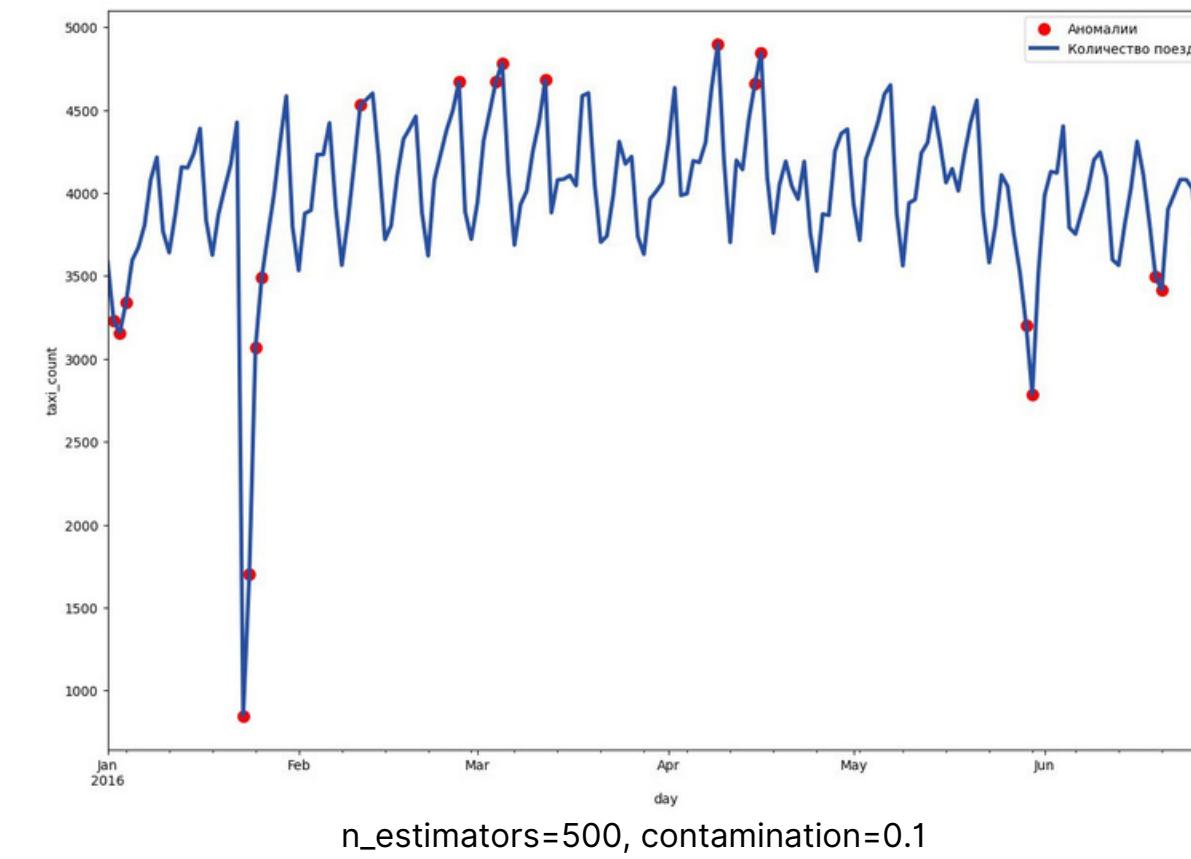
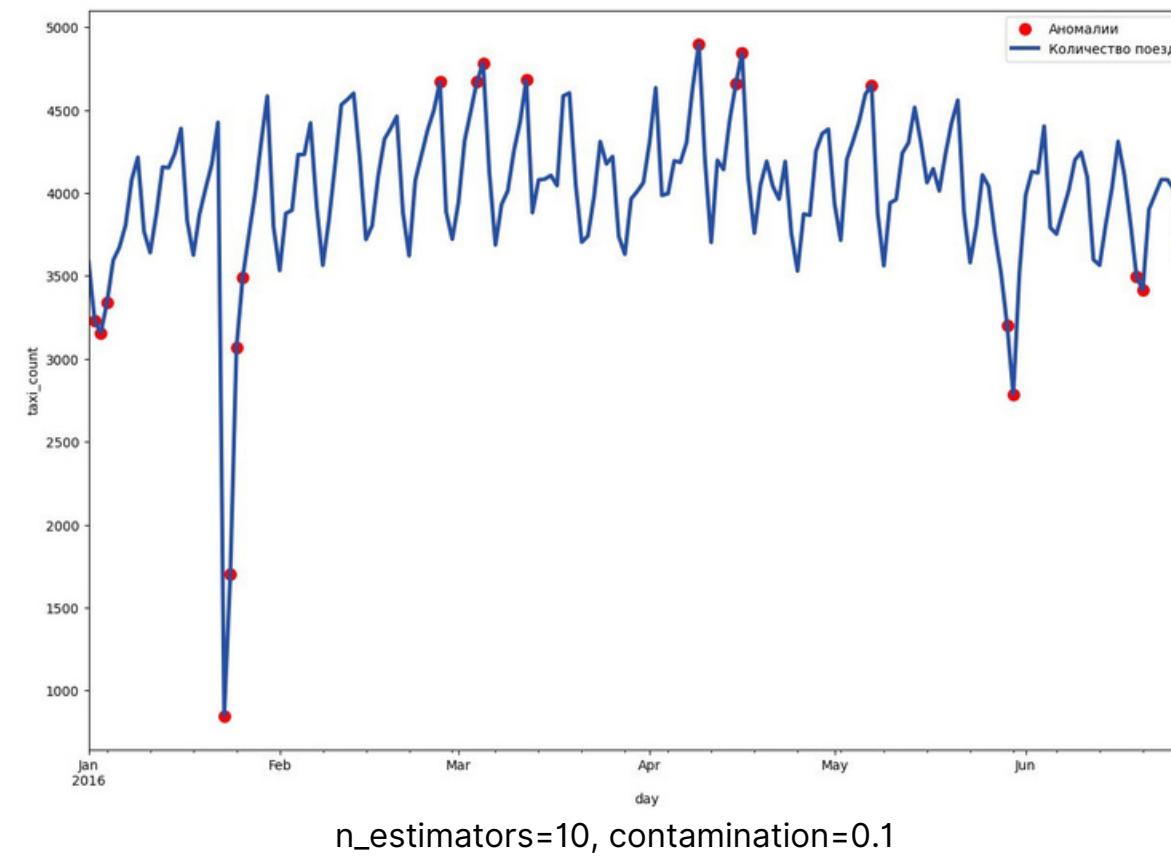
Isolation forest



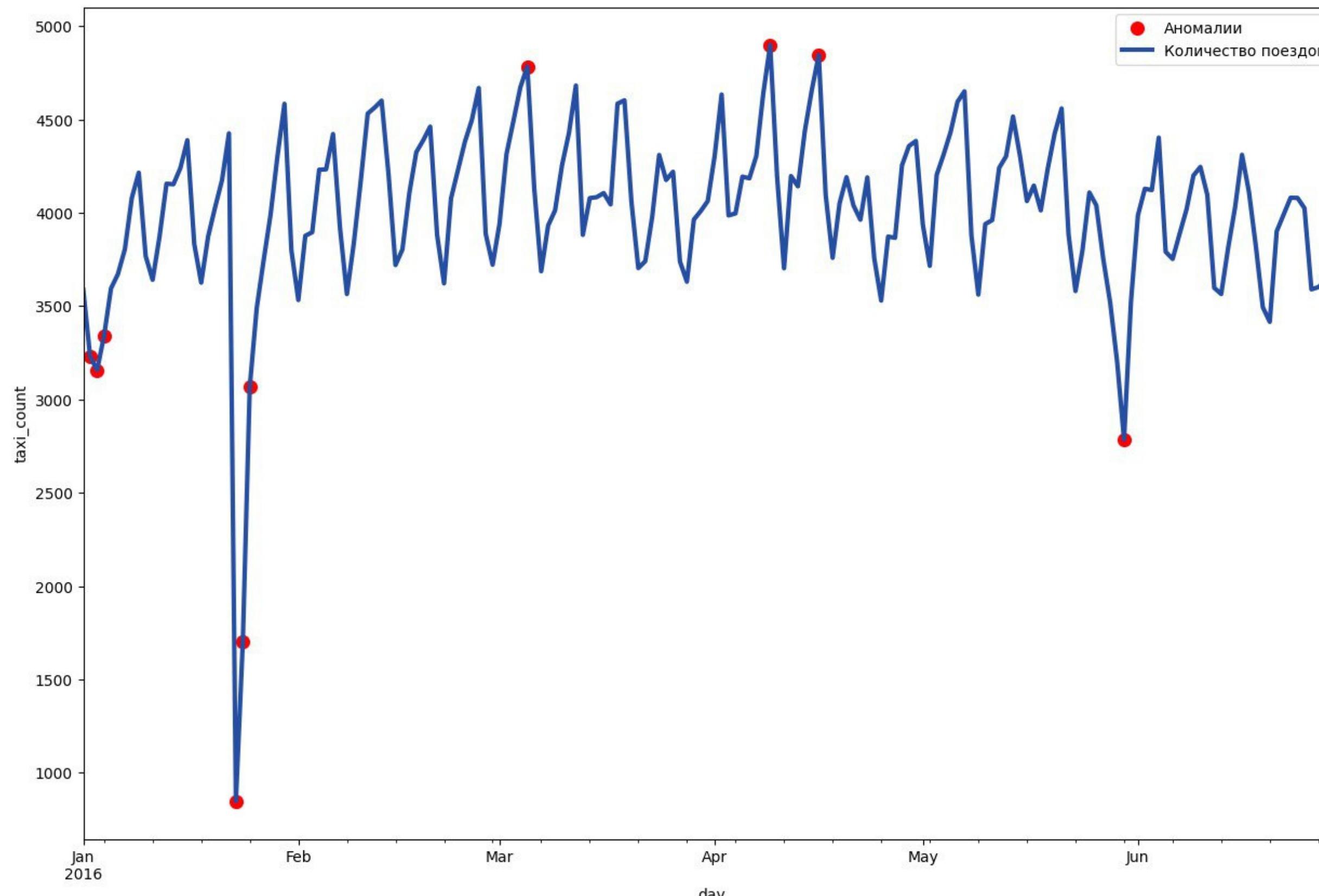
Prophet



Влияние параметров (Iso forest)

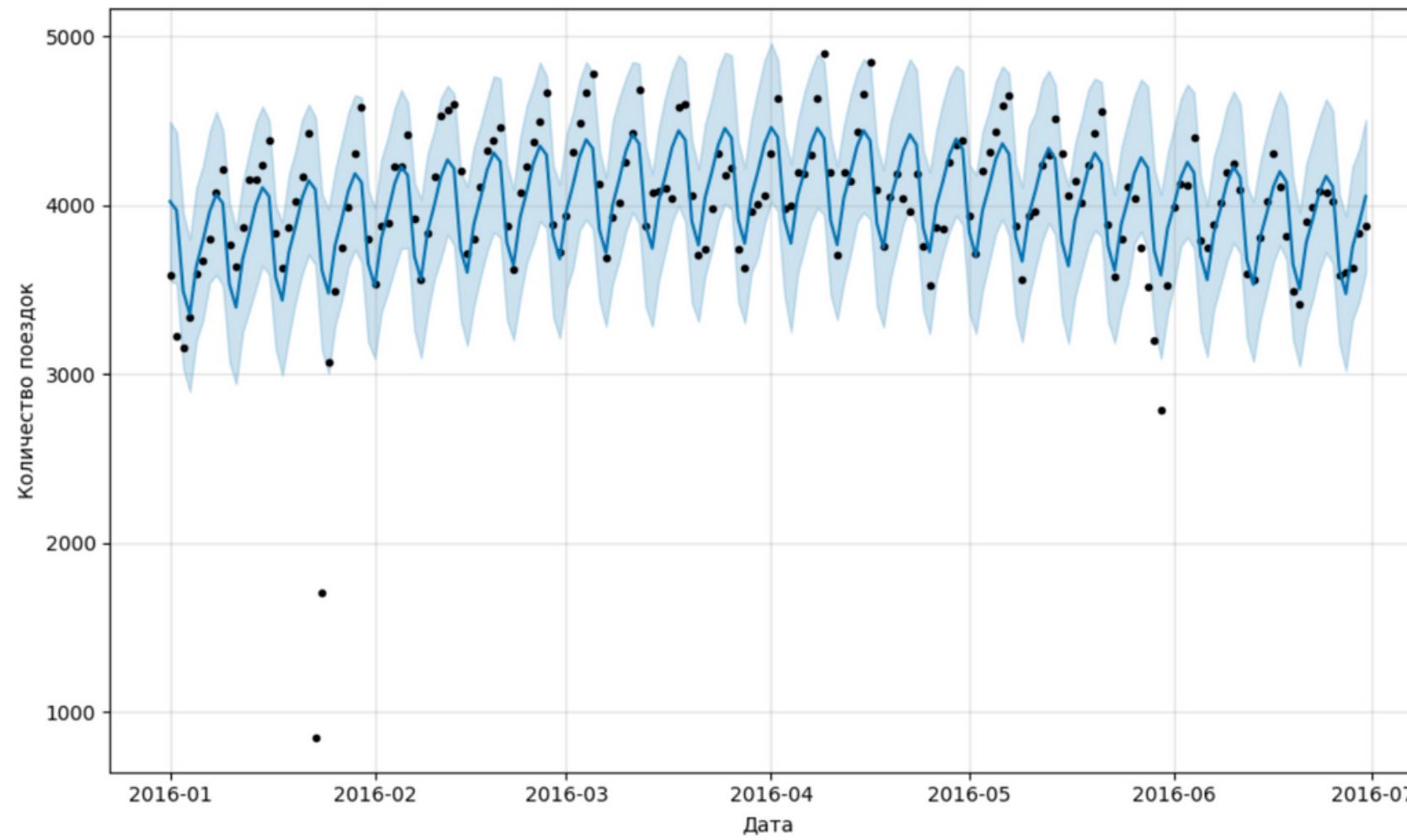


Выбранная модель (Iso forest)

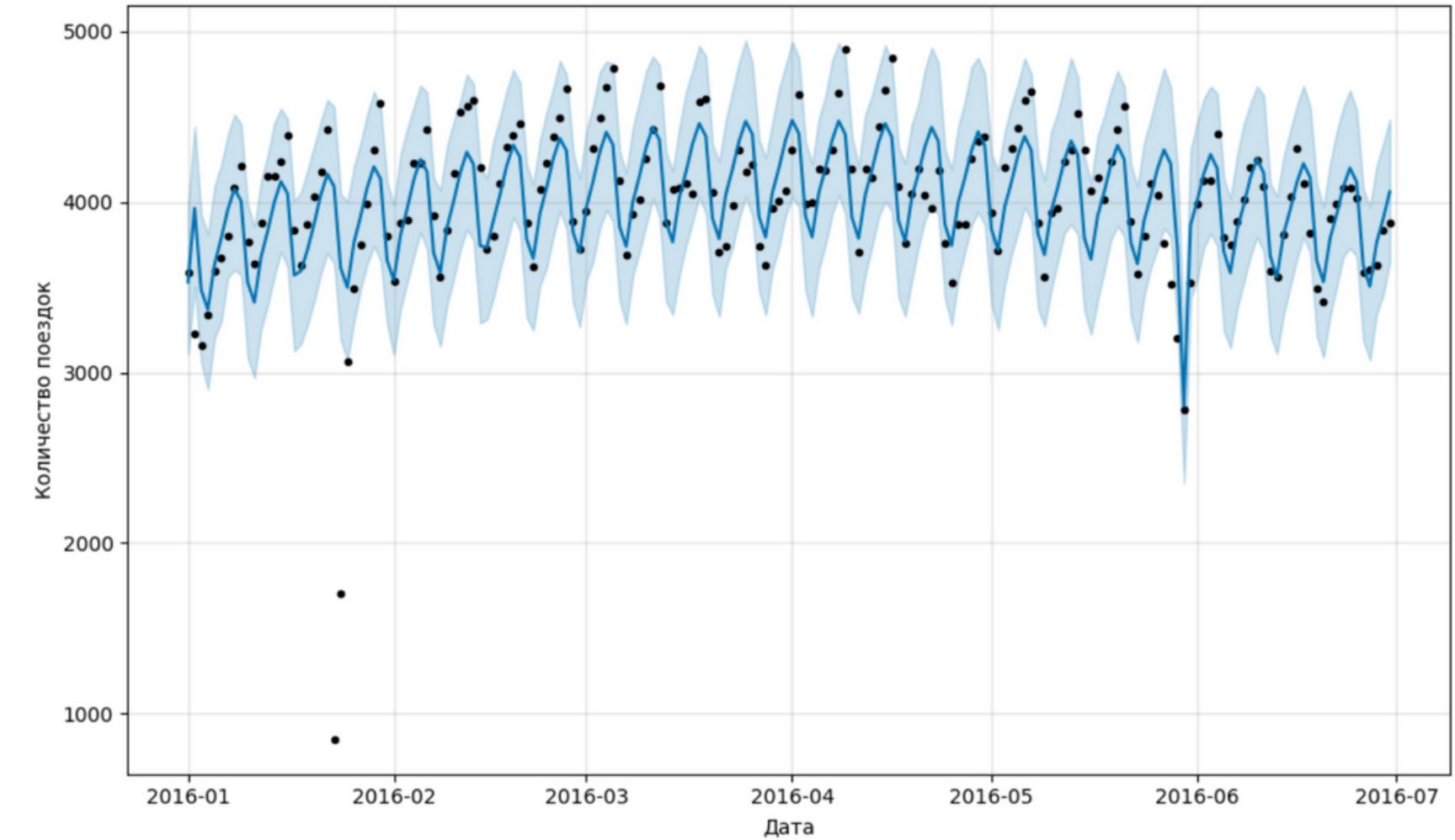


n_estimators=150, contamination=0.05

Влияние параметров (Prophet)

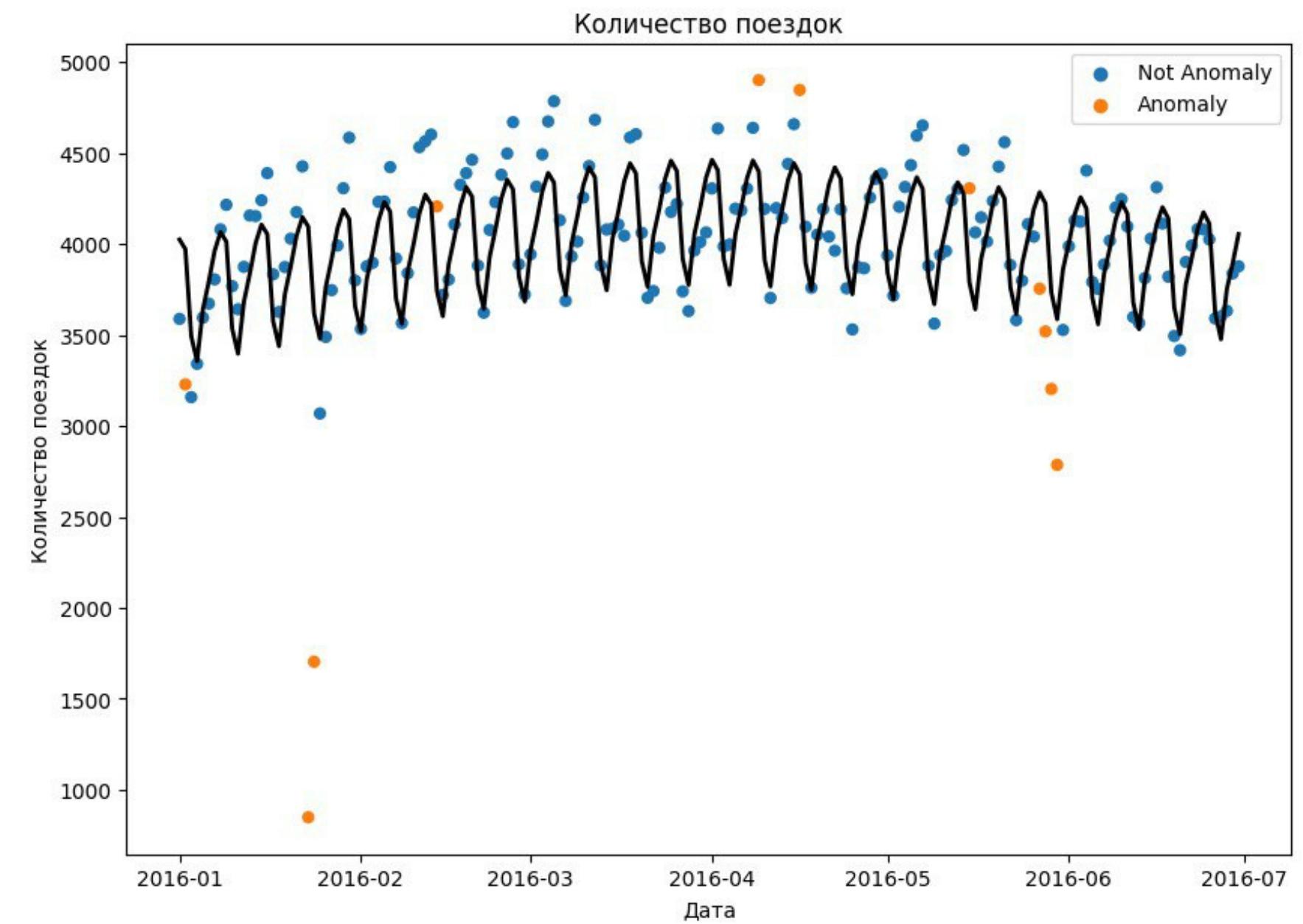
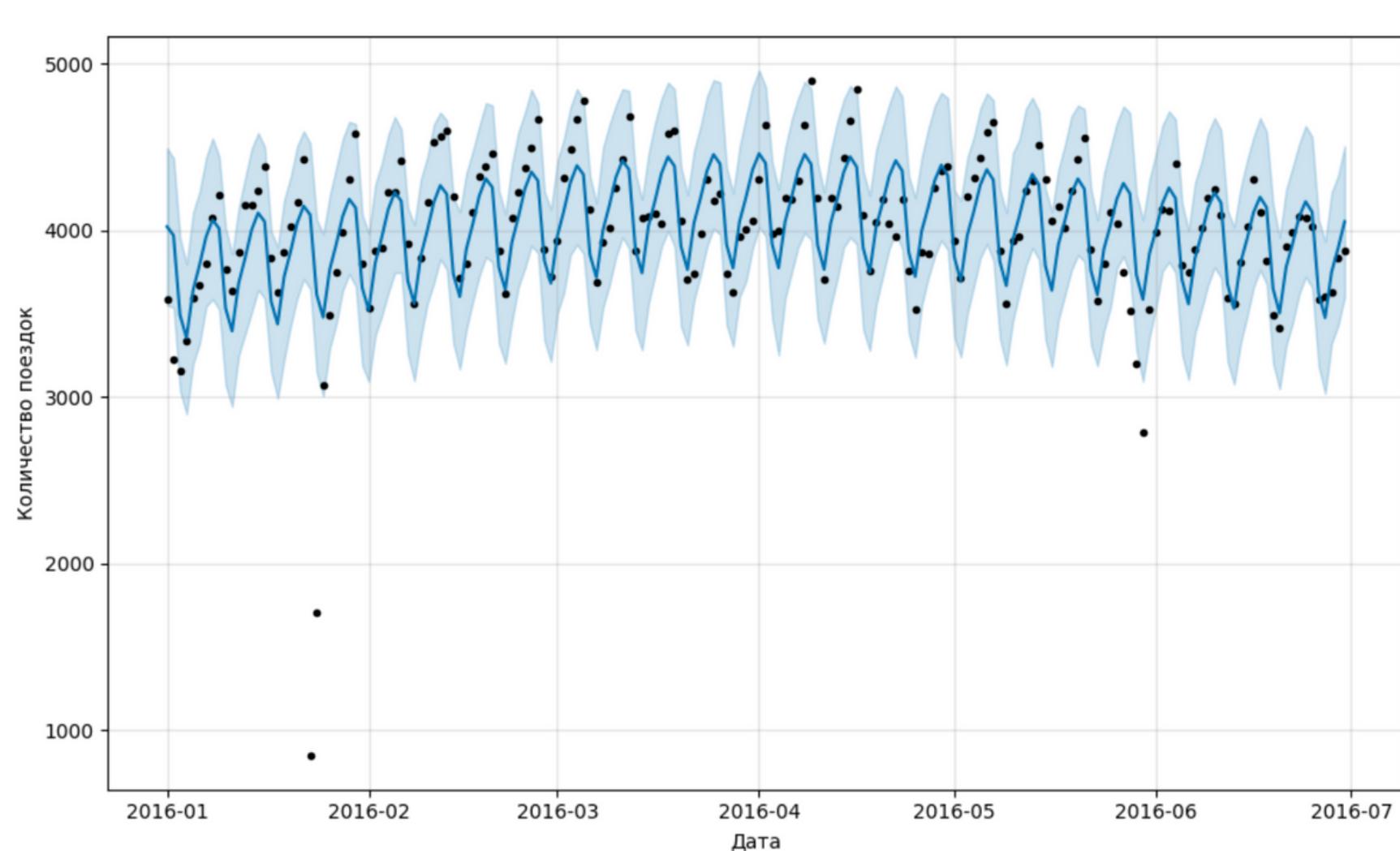


Без учета праздников



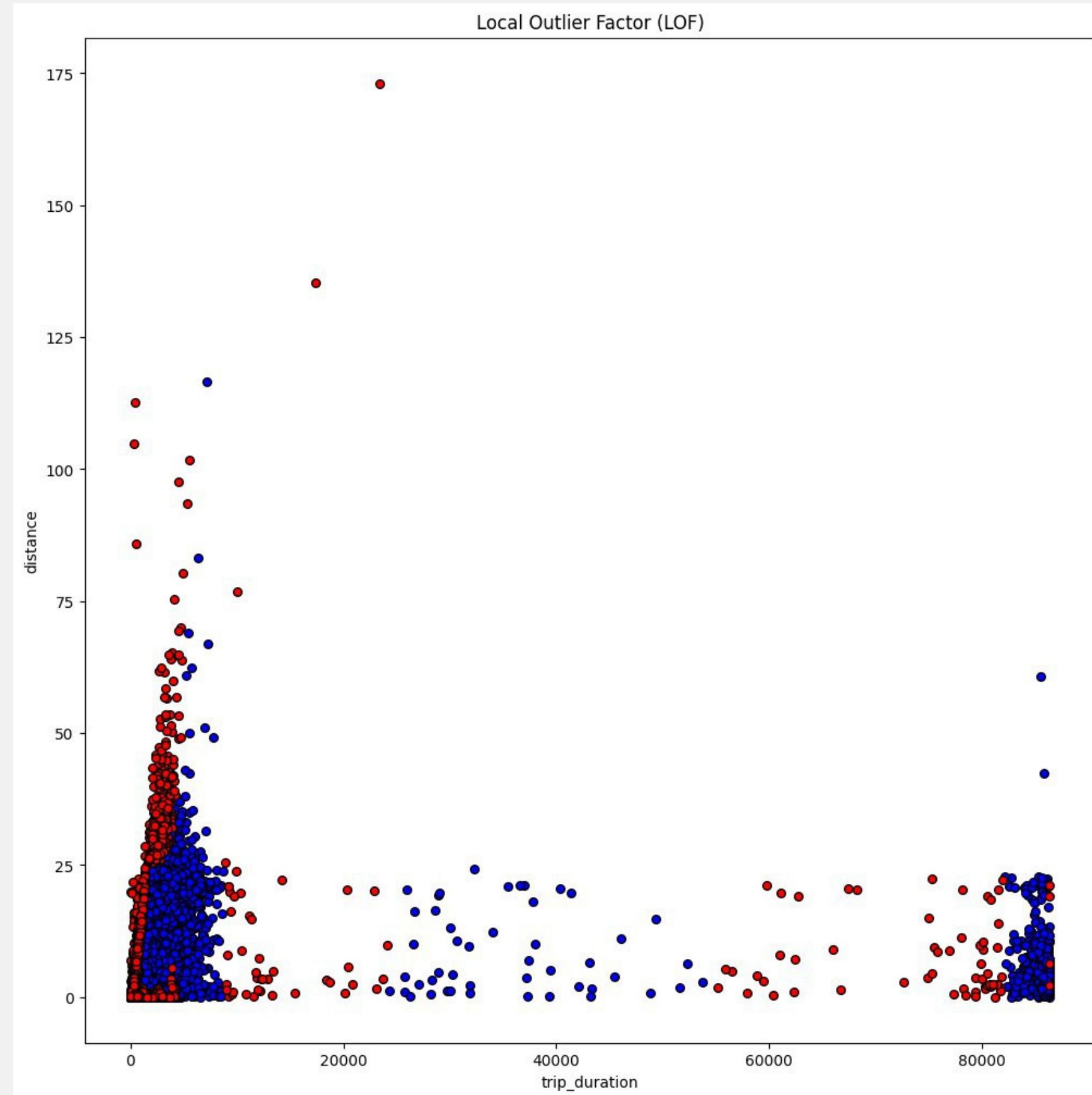
С учетом праздников

Выбранная модель (Prophet)



Дополнительный инструмент - LOF

n_neighbors=50,
contamination='auto'



Описание функции

search_geo_anomalies

добавляет колонку, где отмечает поездки, которые находятся вне Нью-Йорка и рисует большую карту Нью-Йорка

show_small_geo_map

рисует координаты поездок на карте Америки

search_count_anomalies_by_iso

возвращает аномалии в количестве поездок по дням используя isolation forest

search_count_anomalies_by_prophet

возвращает аномалии (и предсказания) в количестве поездок по дням используя prophet

search_nlo_anomalies

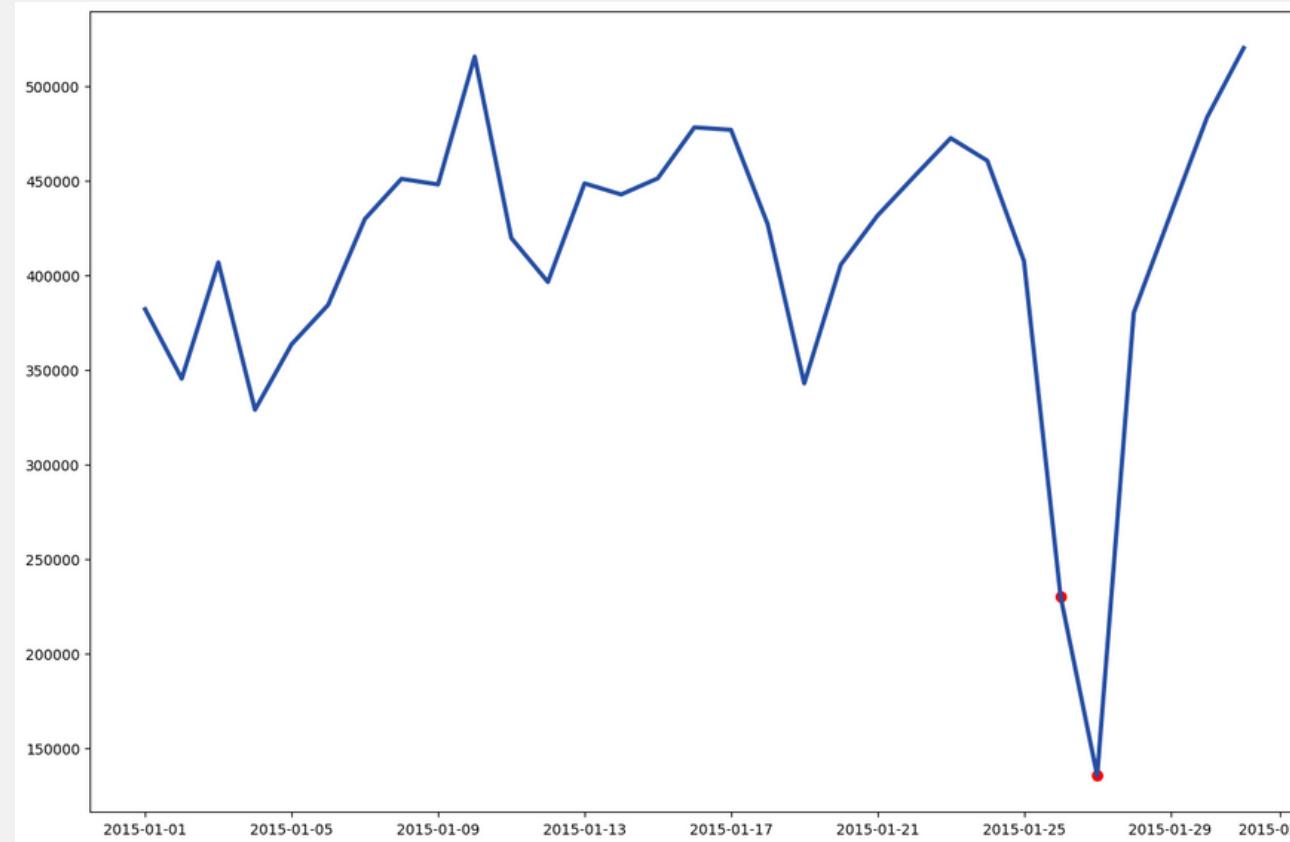
добавляет колонку аномалий, которые находят LocalOutlierFactor и рисует график

search_parameter_anomalies

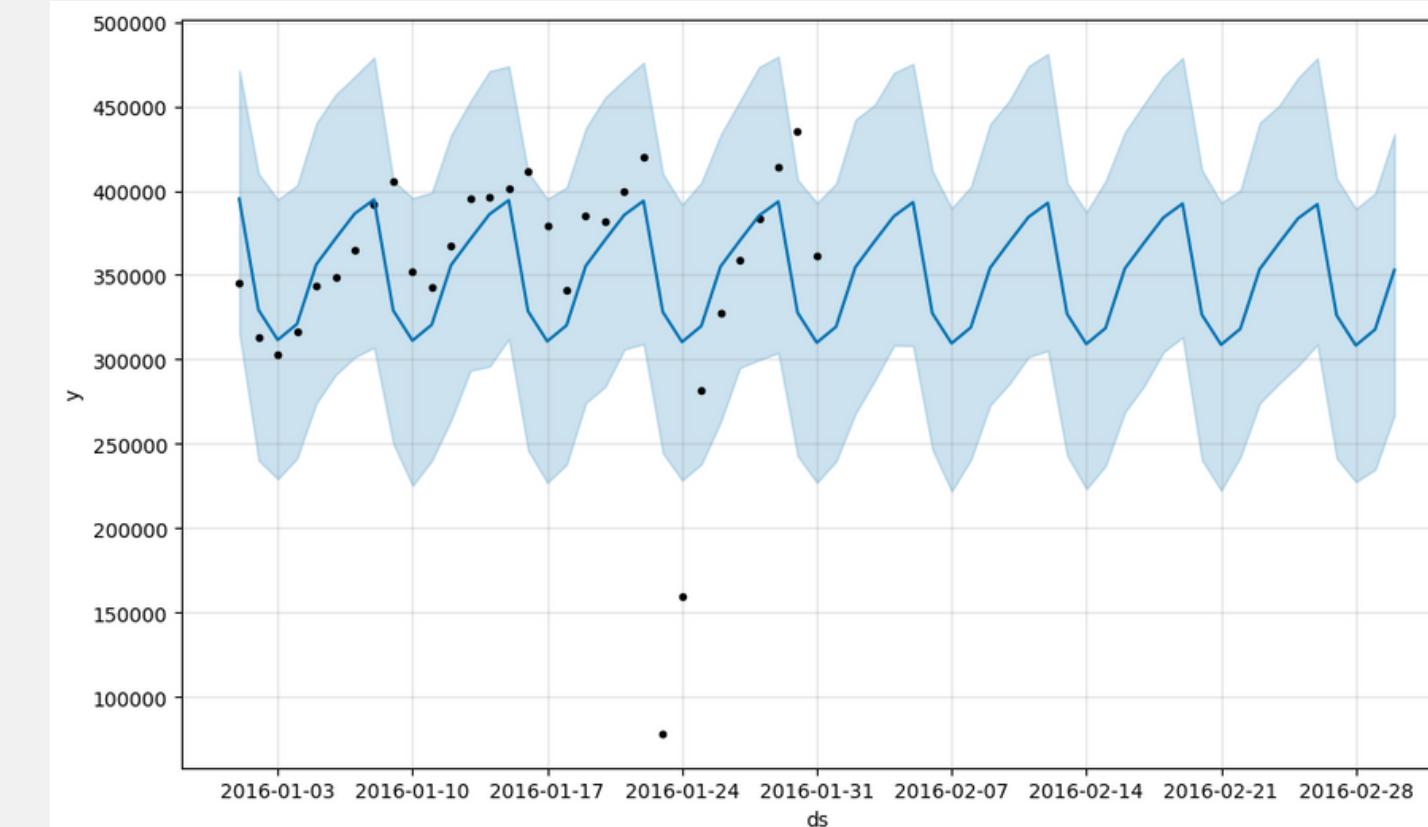
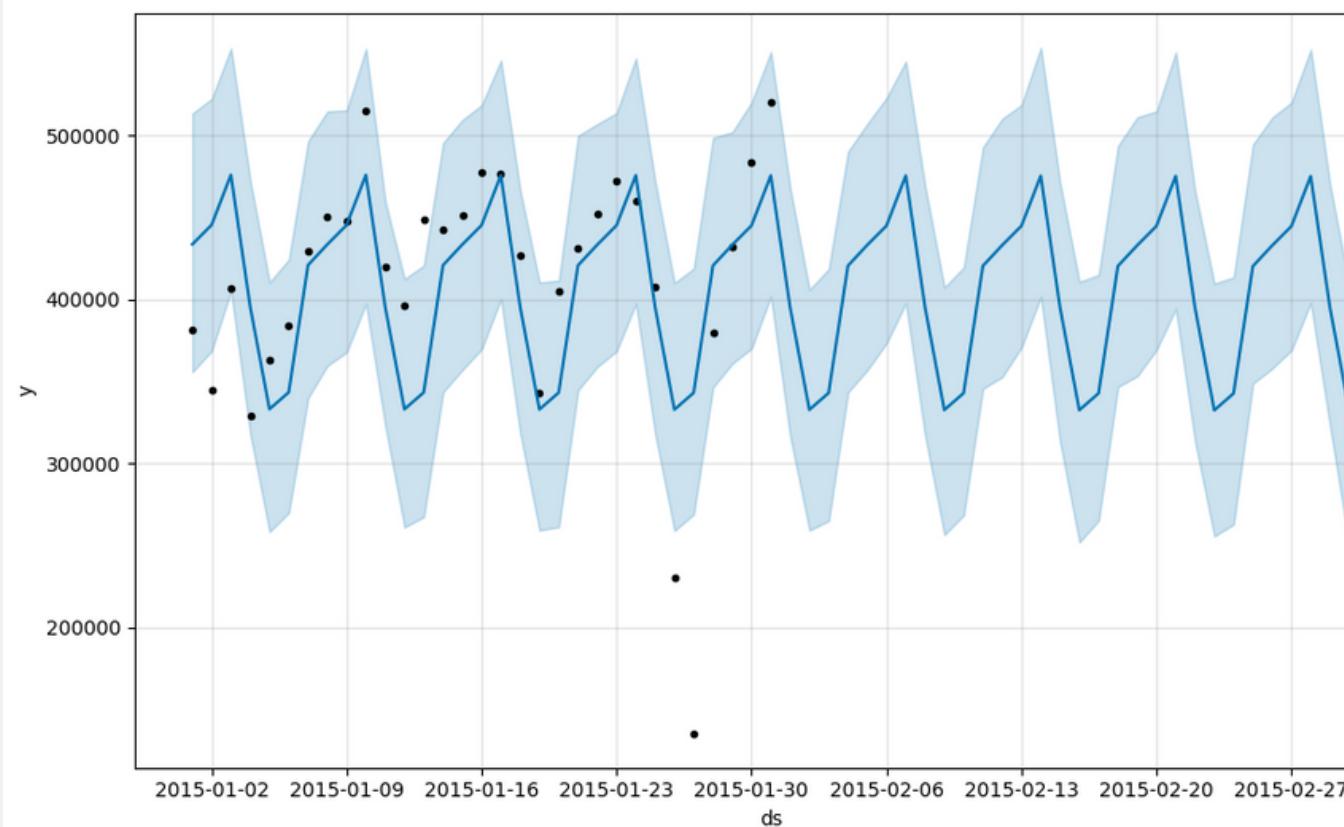
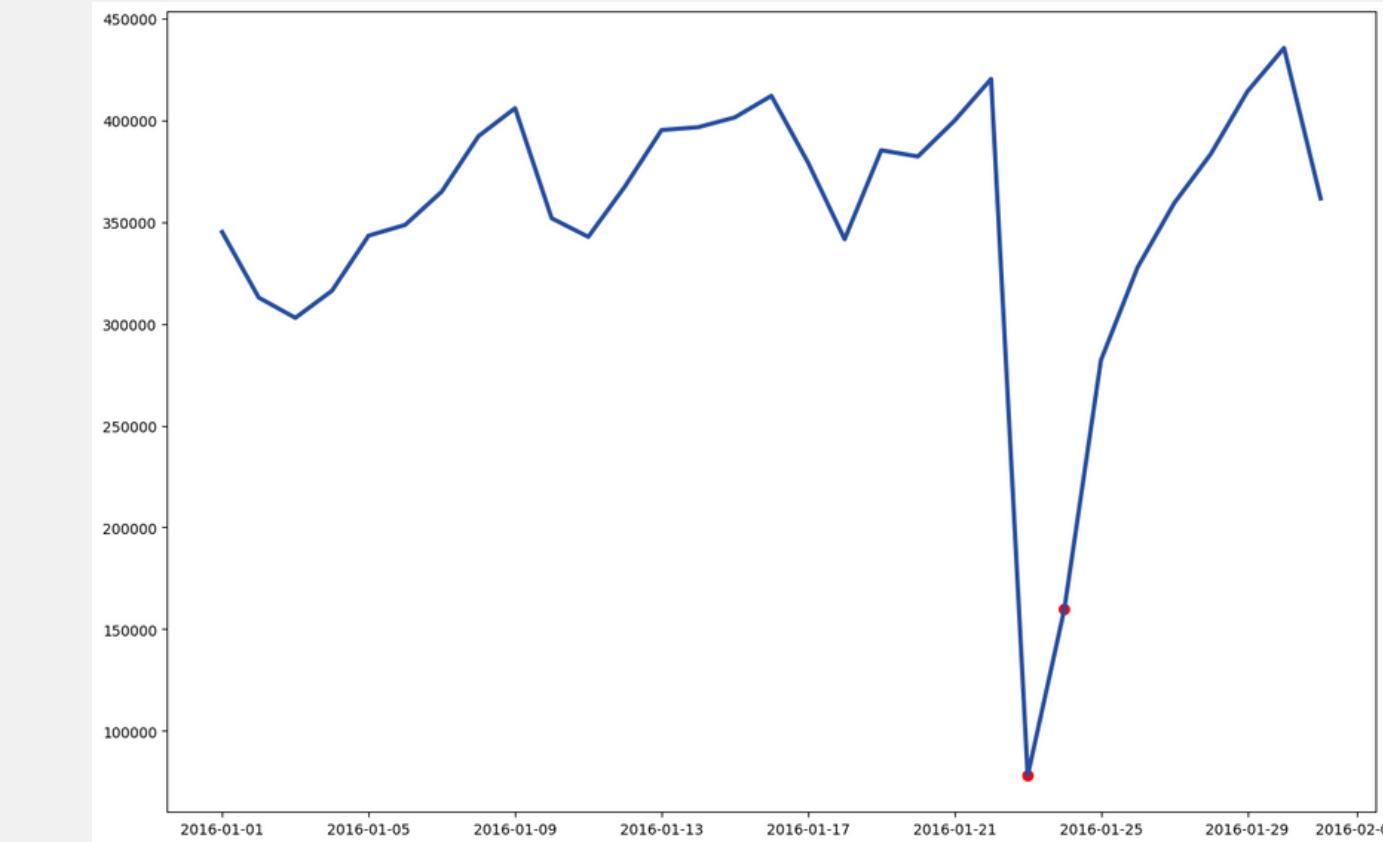
ищет логически невозможные значения

Другие датасеты

1

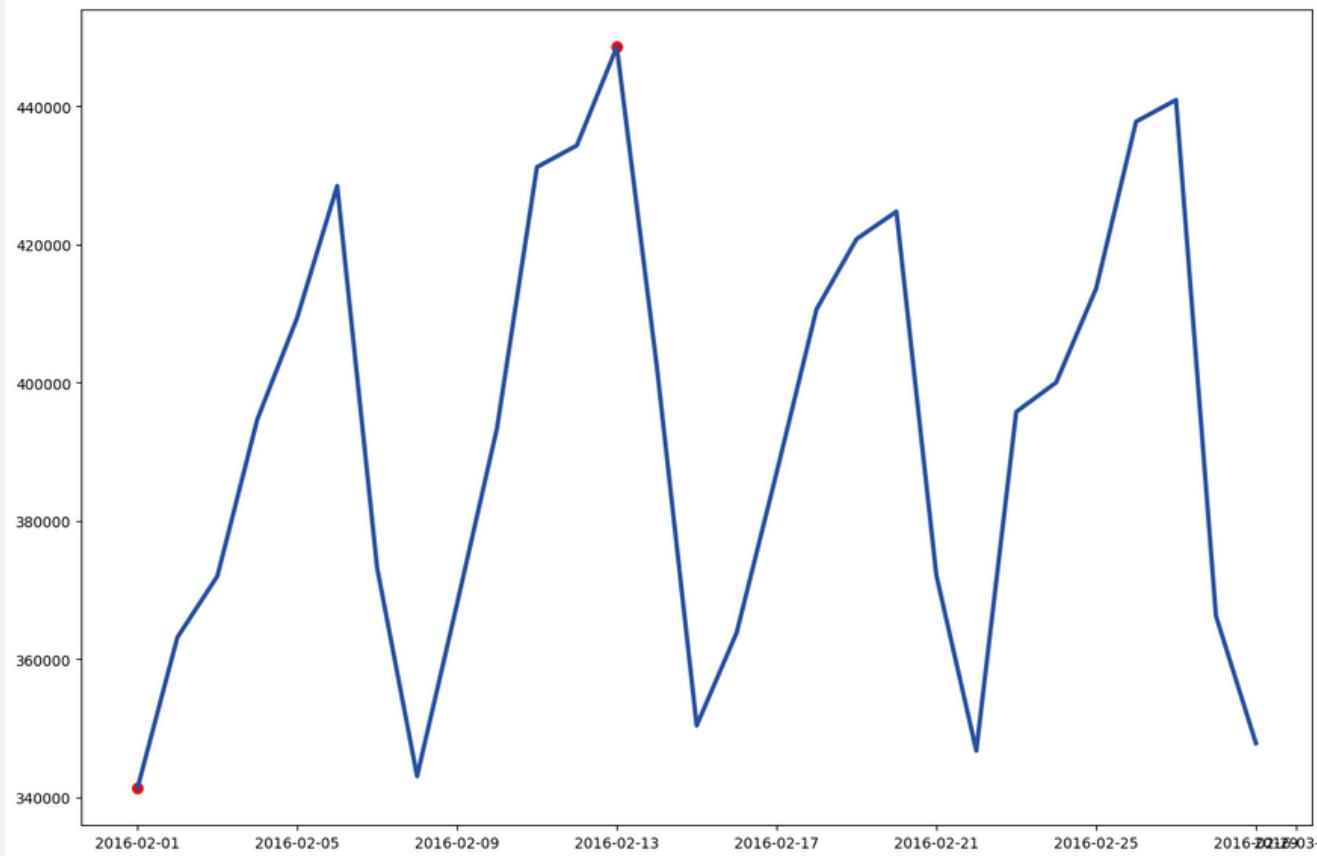


2

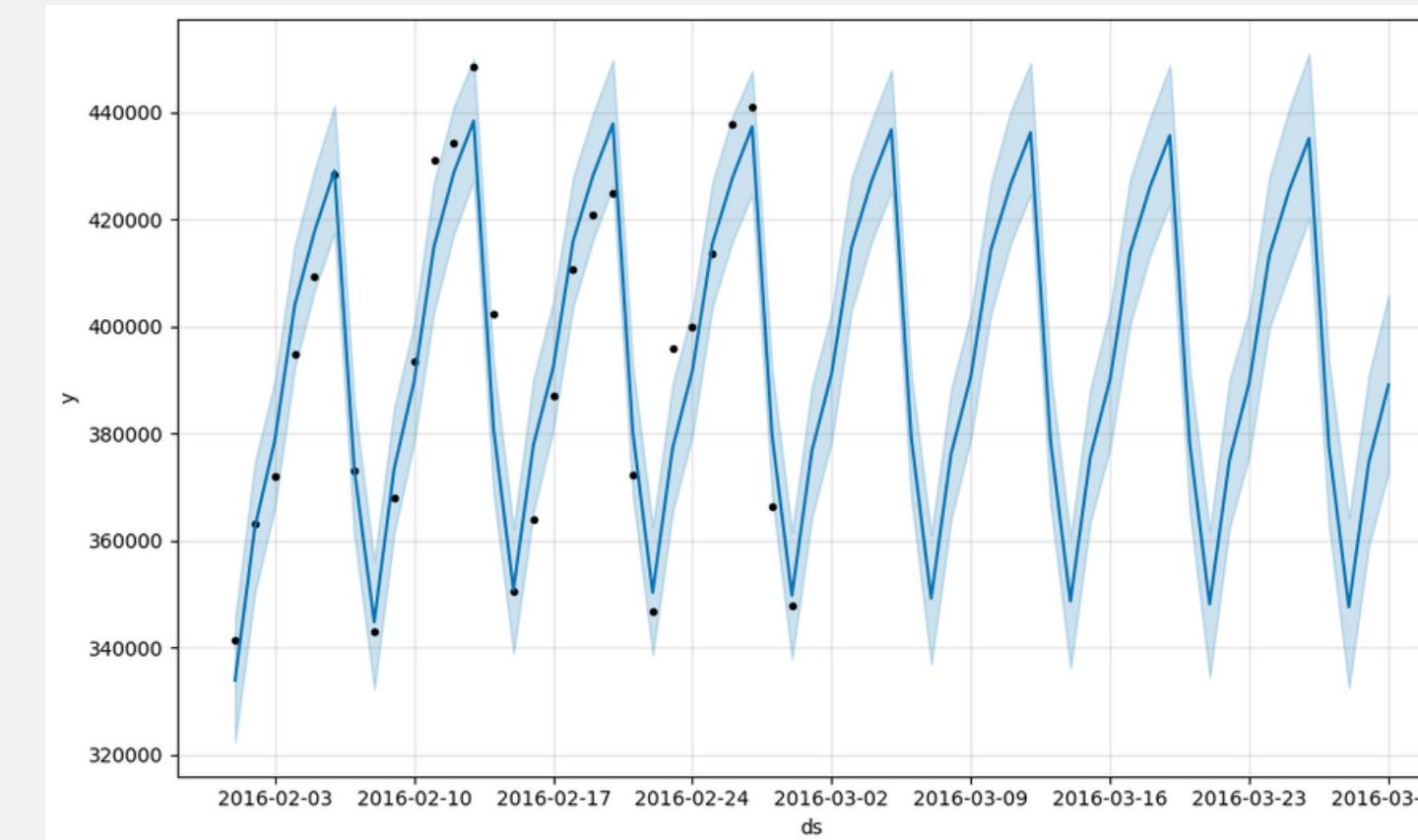
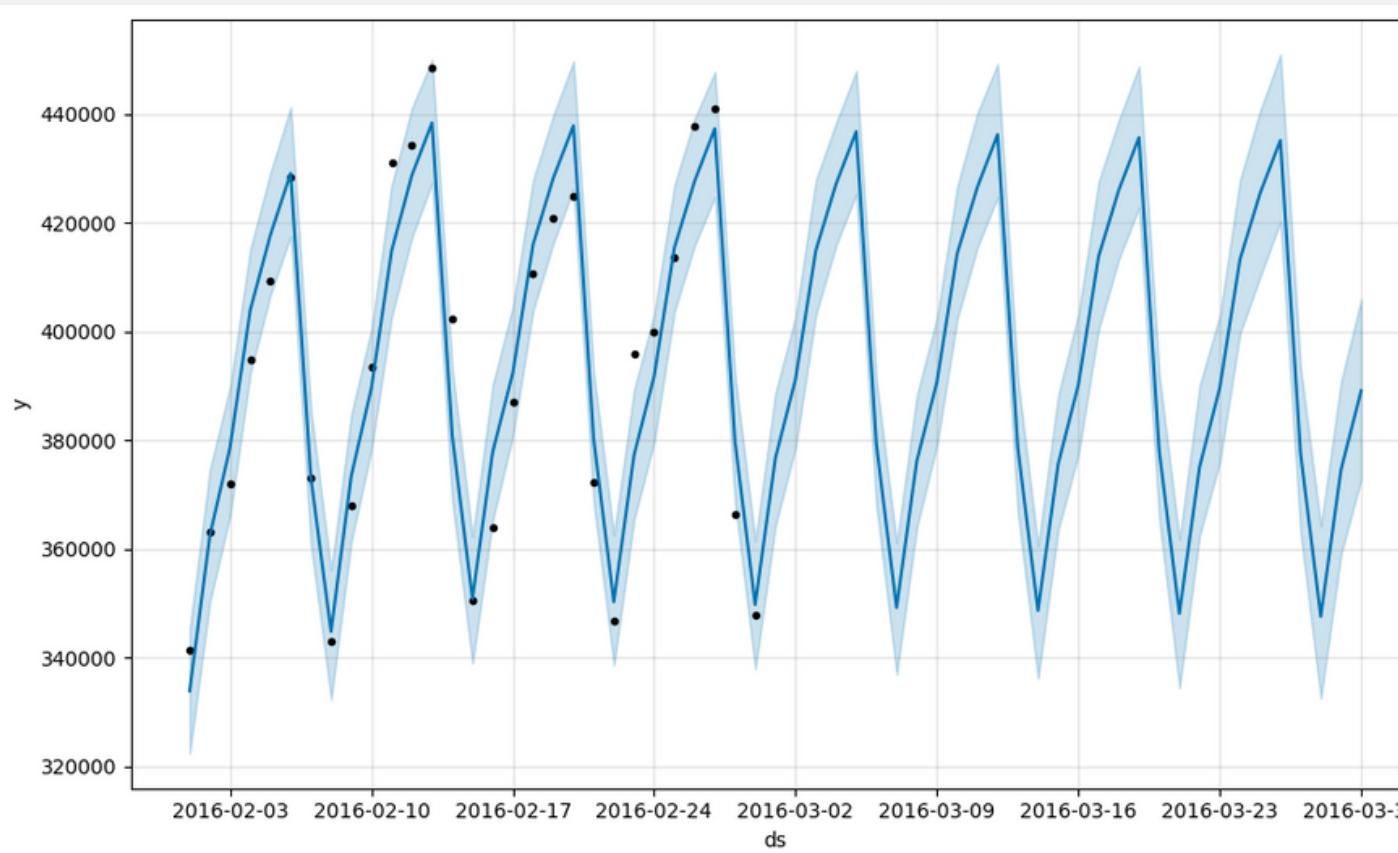
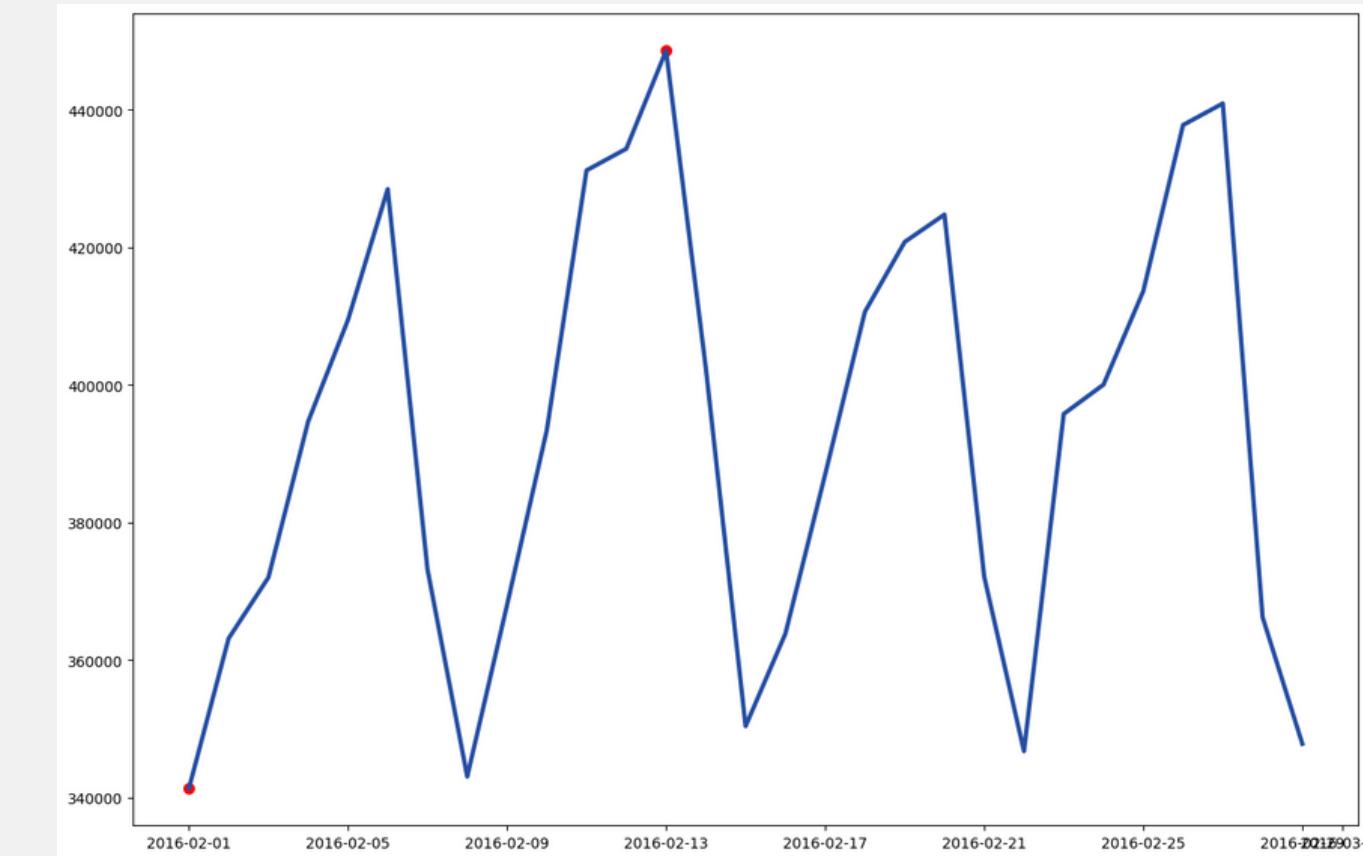


Другие датасеты

3



4



Спасибо за внимание!

Готовы выслушать
ваши вопросы

