

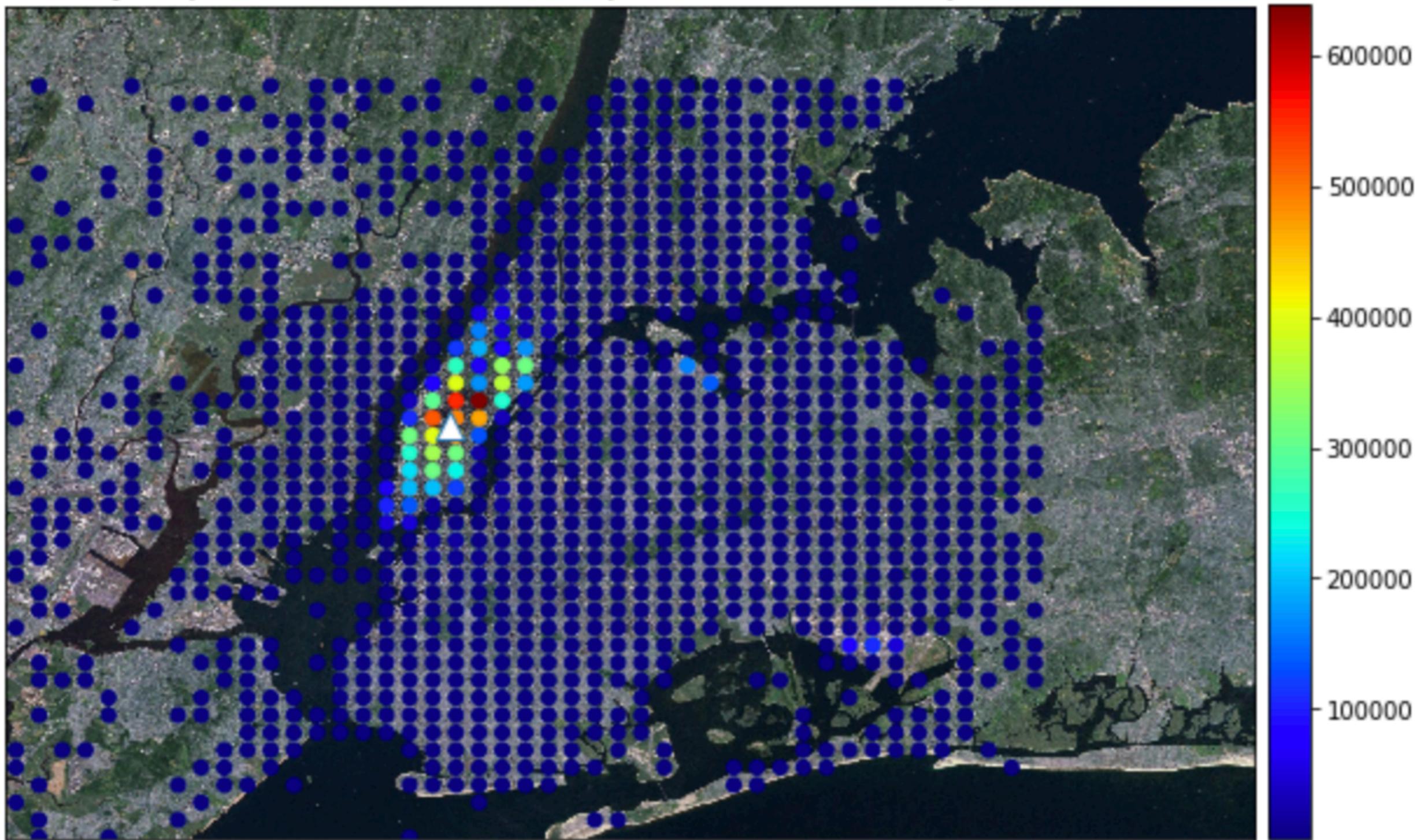
ПРОГНОЗИРОВАНИЕ ПОКАЗАТЕЛЕЙ В БУДУЩЕМ ПЕРИОДЕ

---

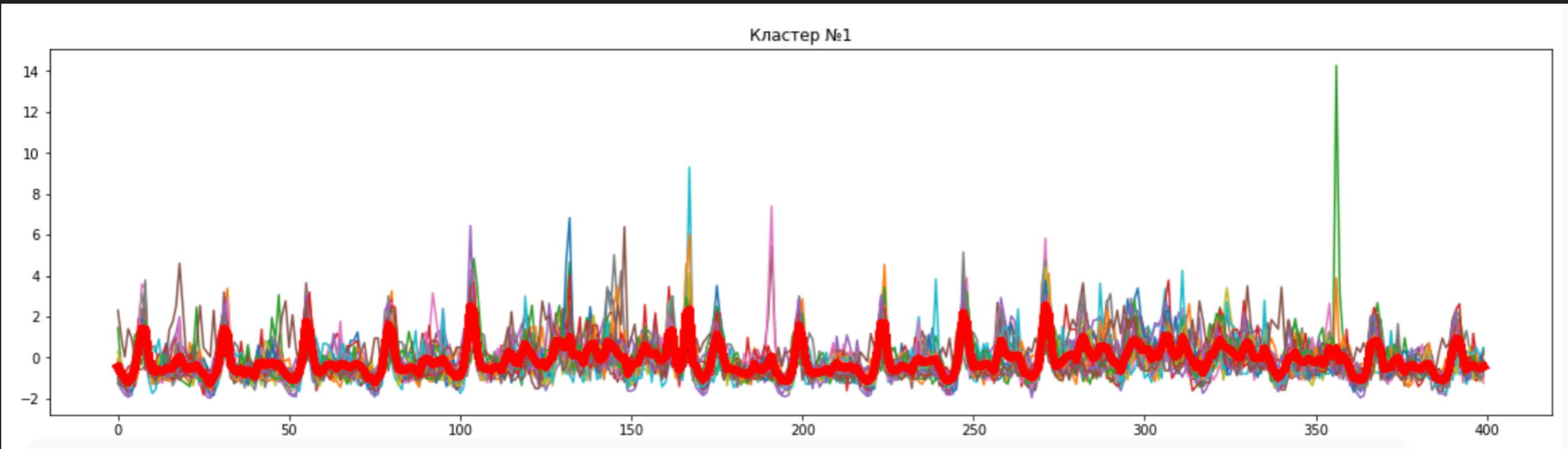
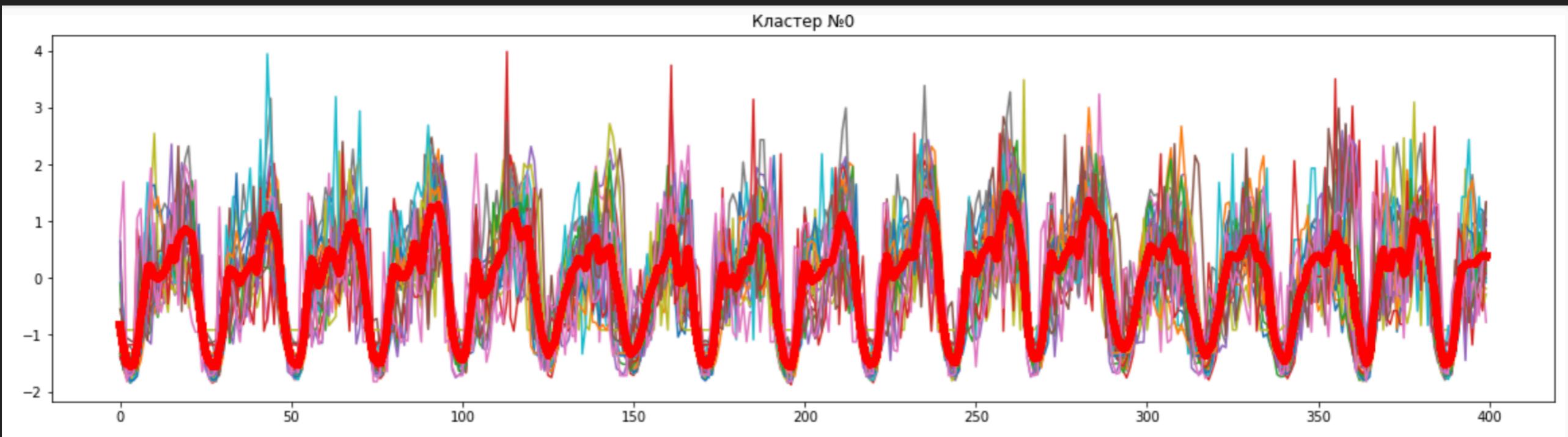
# НА ПРИМЕРЕ ТАКСИ НЬЮ-ЙОРКА

## ОБЩАЯ КАРТИНА ПО НЬЮ-ЙОРКУ

Суммарное кол-во вызов такси из различных зон Нью-Йорка (май 2016)

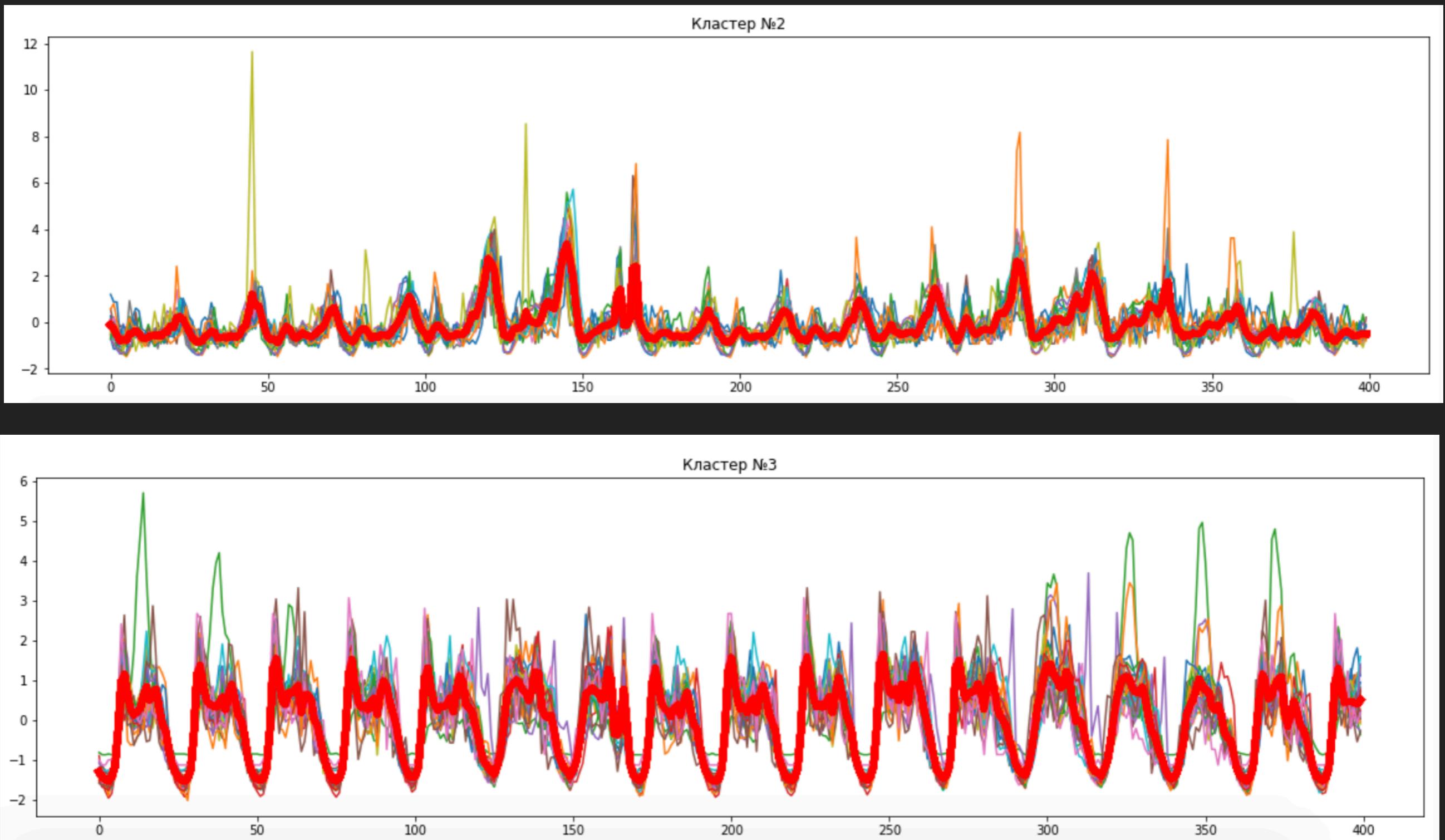


# НАХОДИМ РЕГИОНЫ С ПОХОЖЕЙ ДИНАМИКОЙ



Жирной линией выделен центр кластера, показывающий общую динамику всех регионов, относящихся к этому кластеру.

## НАХОДИМ РЕГИОНЫ С ПОХОЖЕЙ ДИНАМИКОЙ



Для каждого кластера строим модель ARIMA и прогнозируем временной ряд для региона.

МОДЕЛЬ ARIMA:

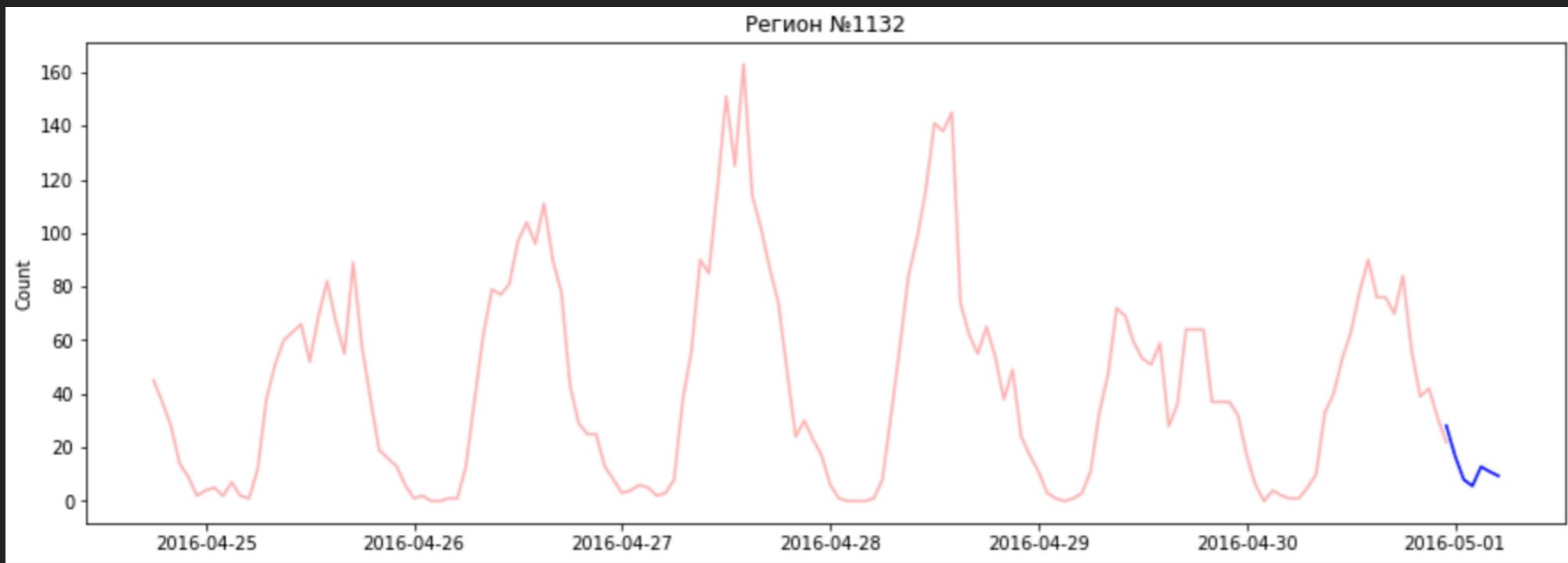
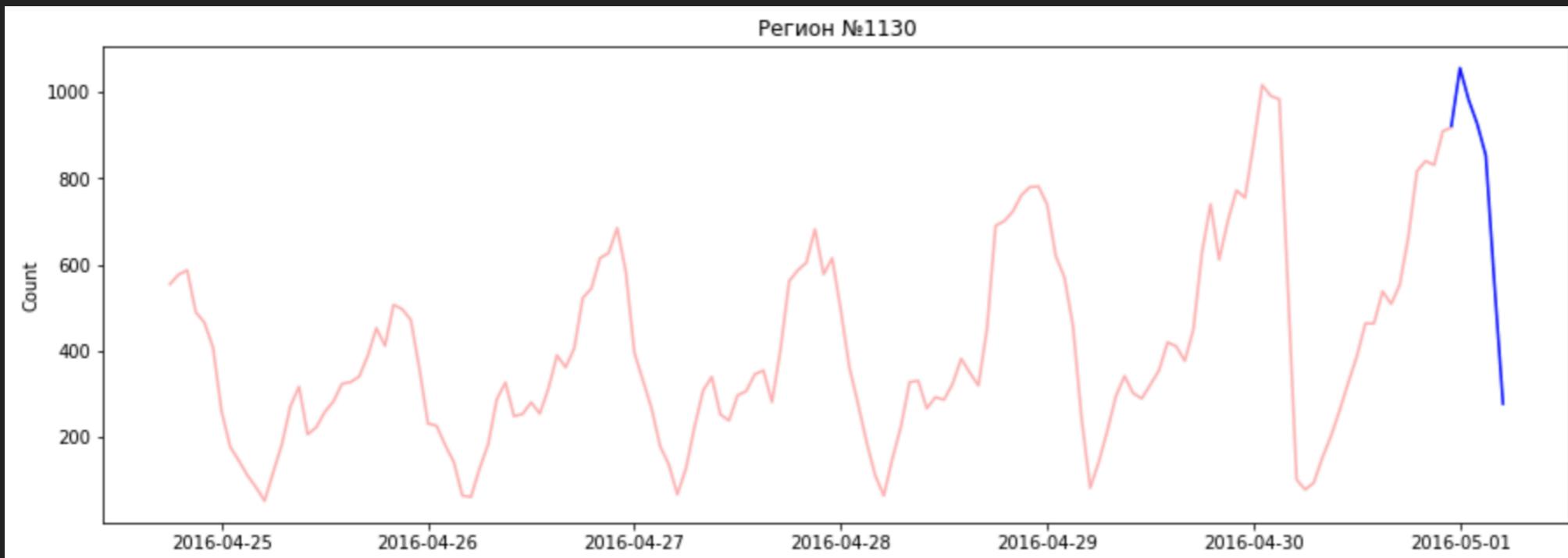
---

## ДЛЯ ПОСТРОЕНИЯ МОДЕЛИ ARIMA:

- ▶ Дифференцируем, чтобы добиться стационарности ряда;
- ▶ Подбираем оптимальные параметры **автокорреляционных** признаков ( $p, q, P, Q, d$ ), описывающие дневную сезонность;
- ▶ Добавляем **регрессионные** признаки, описывающие недельную сезонность.

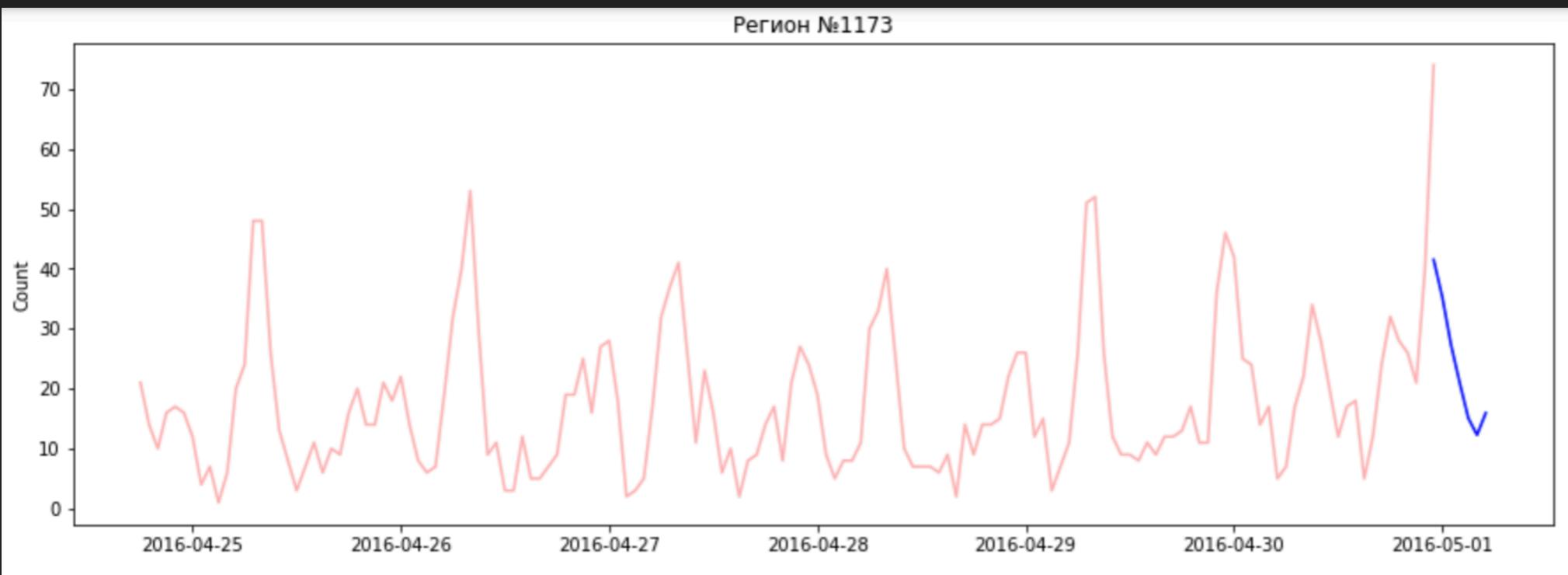
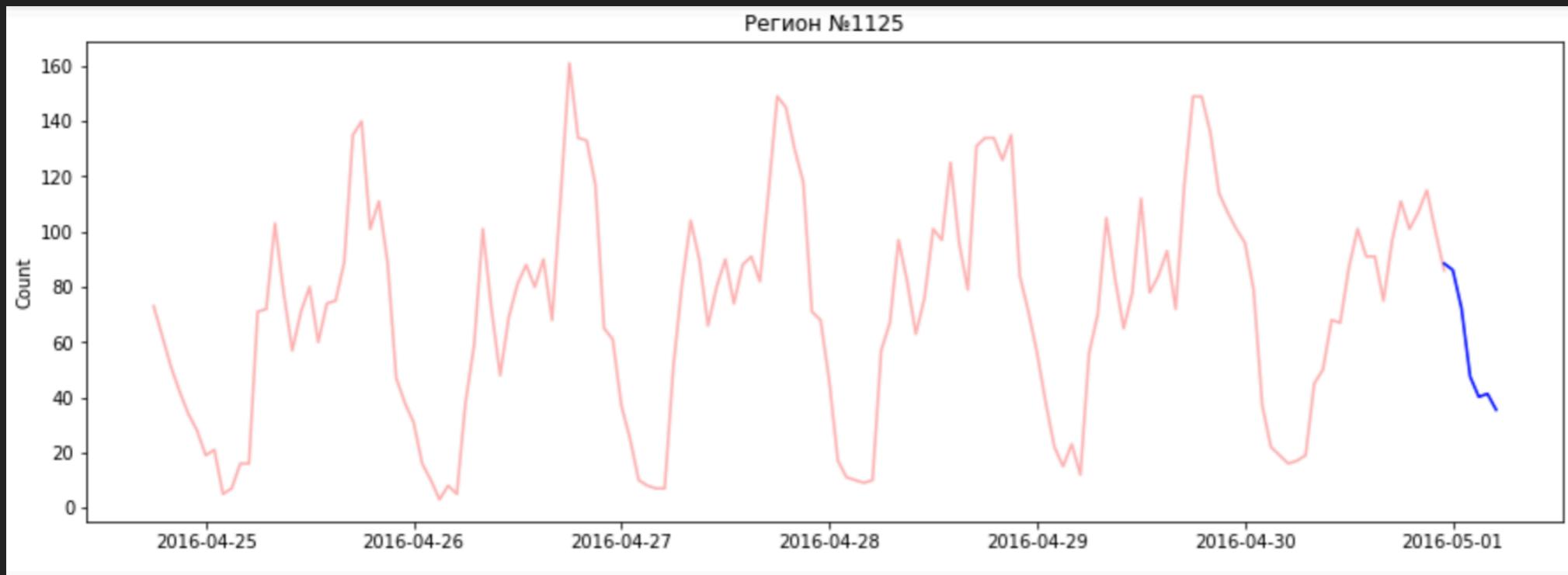
ОБУЧАЕМ МОДЕЛЬ ARIMA ДЛЯ КАЖДОГО КЛАСТЕРА

ПРОГНОЗИРУЕМ РЕГИОНЫ ▶ МАЕ = 38 (поездок)



## ОБУЧАЕМ МОДЕЛЬ ARIMA ДЛЯ КАЖДОГО КЛАСТЕРА

ПРОГНОЗИРУЕМ РЕГИОНЫ ▶ МАЕ = 38 (поездок)



СТРОИМ 6 РЕГРЕССИЙ ДЛЯ ПРОГНОЗА НА КАЖДЫЙ ИЗ 6 ЧАСОВ ВПЕРЕД

---

## В КАЧЕСТВЕ ПРИЗНАКОВ ИСПОЛЬЗУЕМ:

- ▶ 24 признака: количество поездок в предыдущие часы
- ▶ 7 признаков: количество поездок в аналогичный час за последние 7 дней
- ▶ Суммарное кол-во поездок за последние 24 часа
- ▶ Гармоники (вида: `np.sin(x*2*np.pi*n/24))`, где n от 2 до 6

СТРОИМ 6 РЕГРЕССИЙ ДЛЯ ПРОГНОЗА НА КАЖДЫЙ ИЗ 6 ЧАСОВ ВПЕРЕД

---

## РЕЗУЛЬТАТЫ

- ▶ RandomForest
- ▶ МАЕ = 18,7 (поездок)