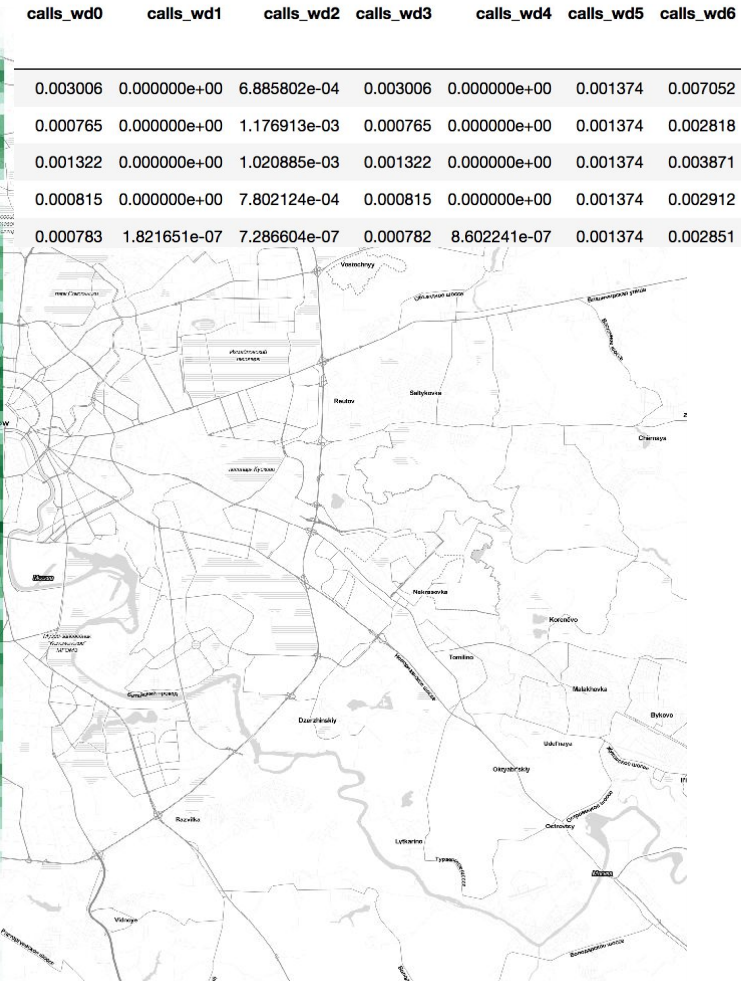
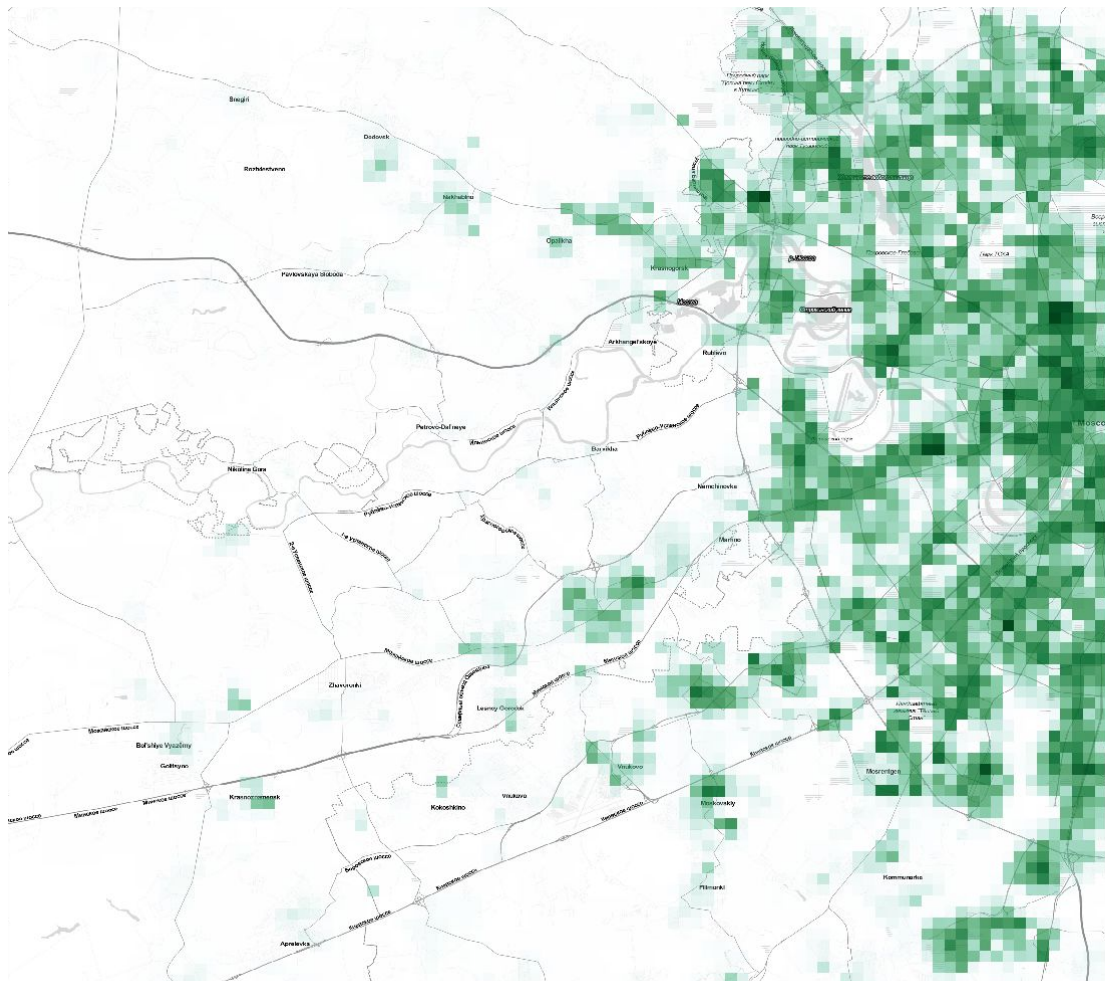


# Предсказание вызовов в 112 (MTS GeoHack.112)

—

Павел Логачев



# Feature engineering

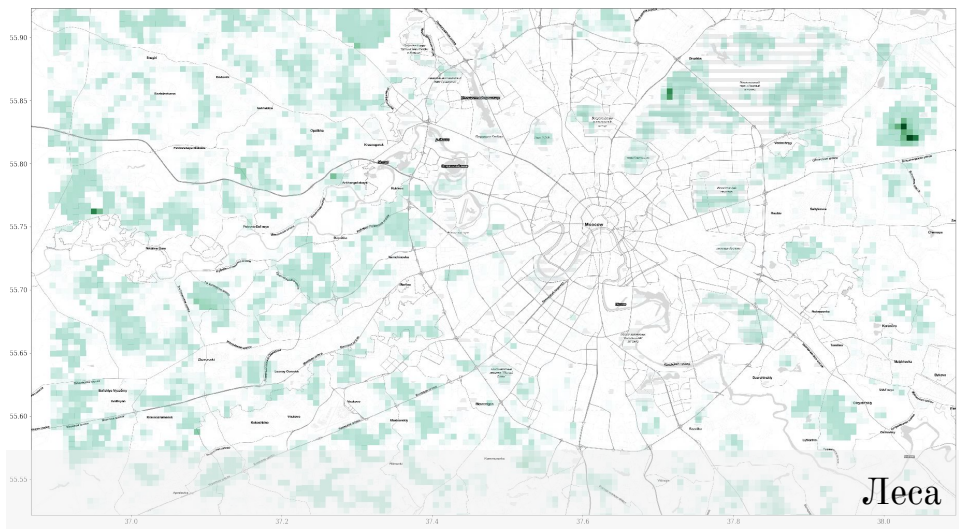
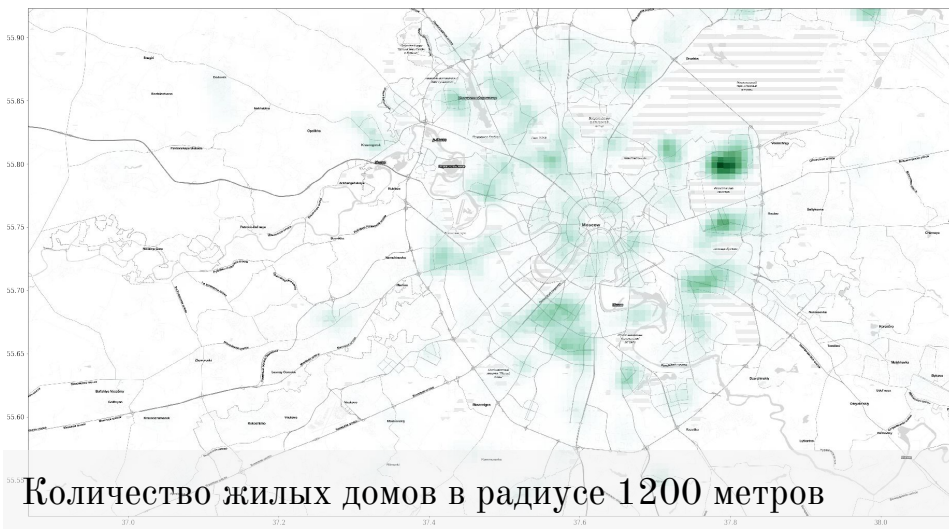
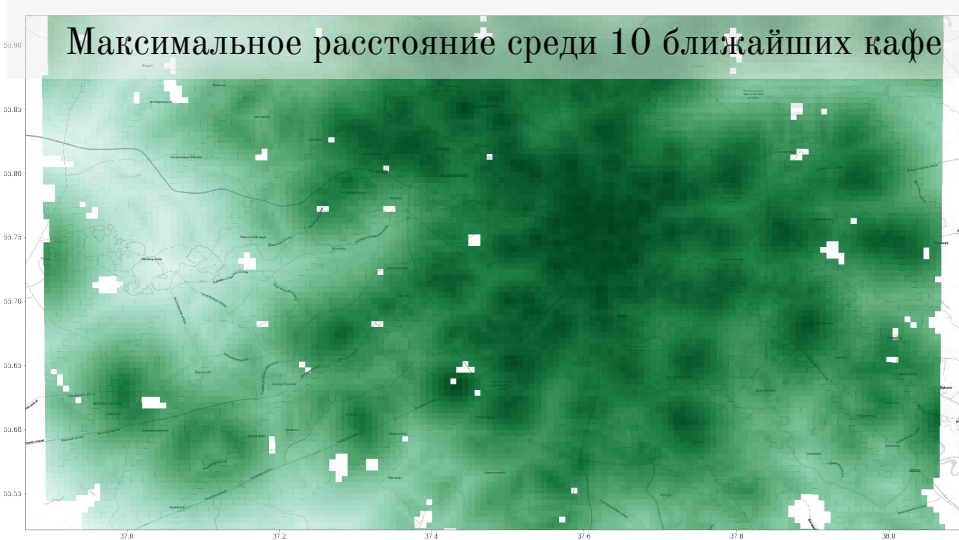
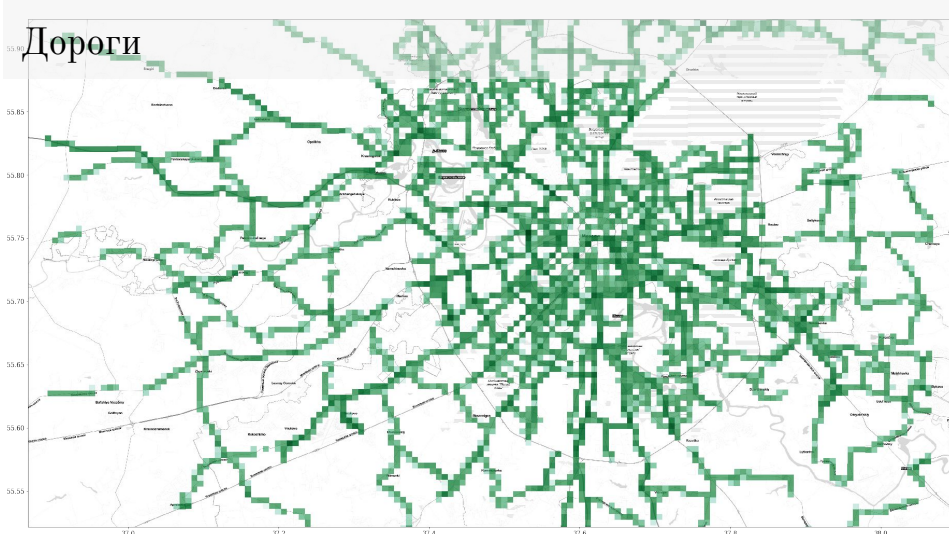
## Данные

- OSM
  - Дома и участки
  - Дороги
  - Точечные объекты
- Yandex API
  - Магазины
  - Банкоматы
  - Учебные заведения
  - Бары
  - Станции метро
  - Офисные центры
  - Остановки
  - Парикмахерские
- MTS Базовые станции

## Обработка

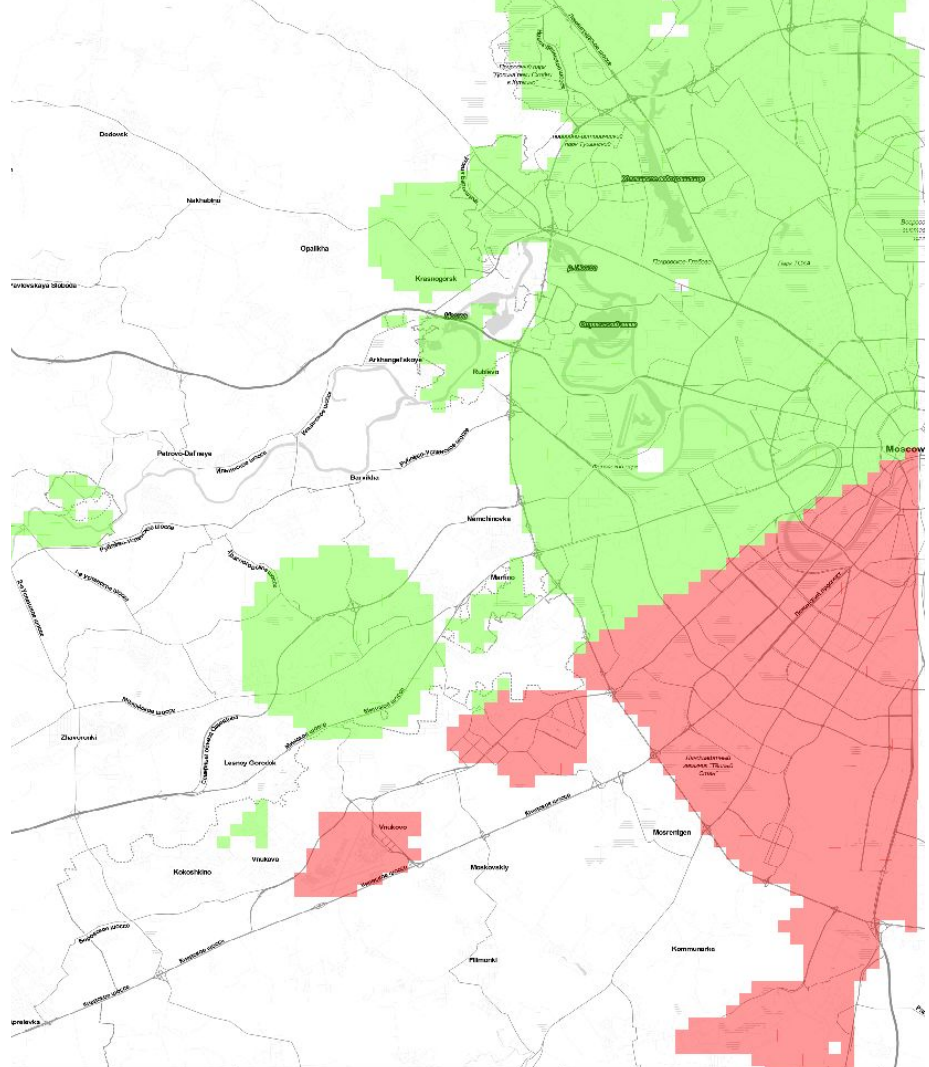
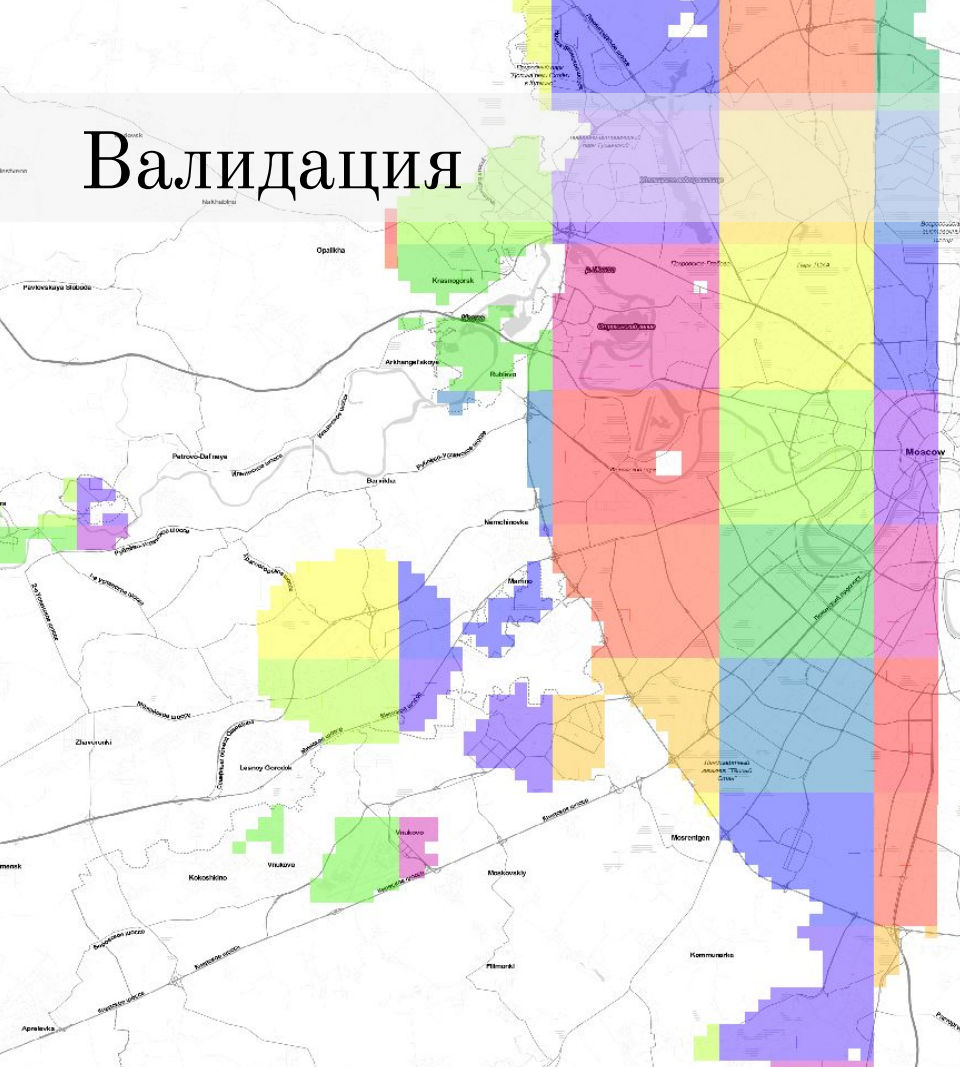
- Для полигонов суммарная площадь объектов который пересекает квадрат
- Для дорог суммарная длина объектов который пересекает квадрат
- Для точечных объектов максимальное расстояние до ближайших 1, 2, 10 объектов и объектов в радиусе 300, 500, 1200 метров. Важный момент, не забыть перейти из градусов в метры

Всего 168 features





# Валидация



# ExtraTreesRegressor



# Модель

Своя на каждый день

Быстрые  
эксперименты

Одна на все дни

Кендалл-тау нехорошо  
усреднять

---

# Kendall tau-b

$$\tau_B = \frac{n_c - n_d}{\sqrt{(n_0 - n_1)(n_0 - n_2)}}$$

where

$$n_0 = n(n - 1)/2$$

$$n_1 = \sum_i t_i(t_i - 1)/2$$

$$n_2 = \sum_j u_j(u_j - 1)/2$$

$n_c$  = Number of concordant pairs

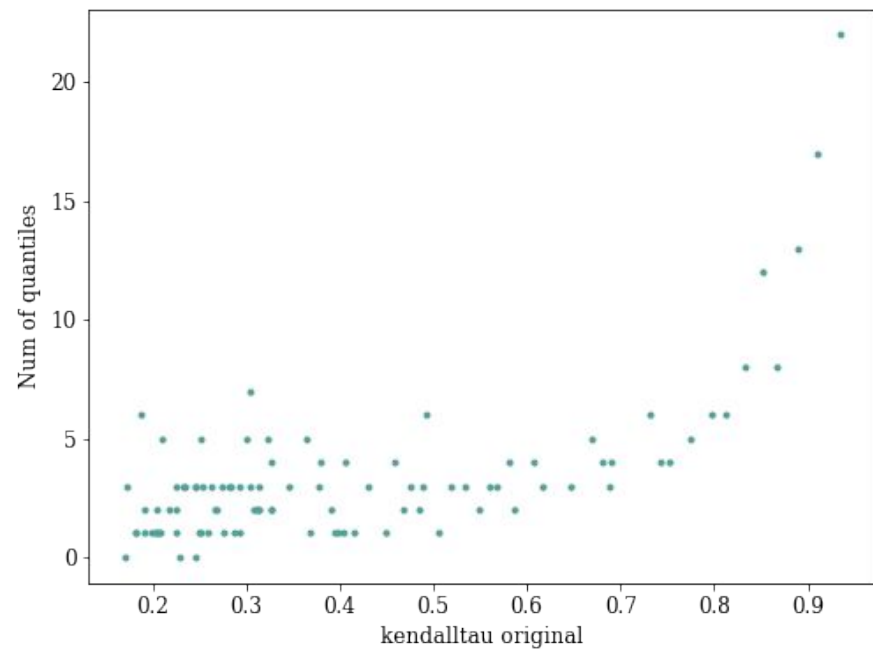
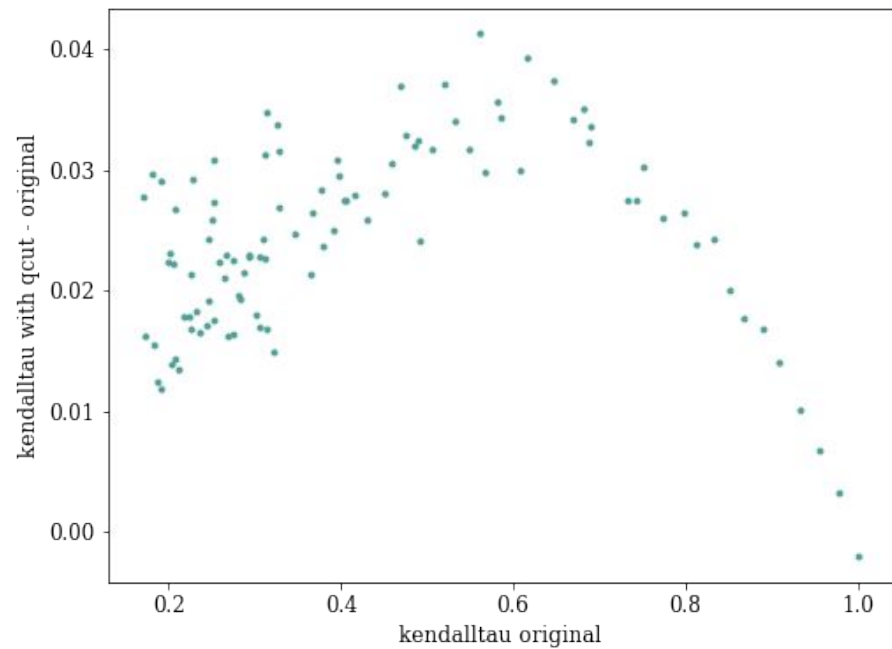
$n_d$  = Number of discordant pairs

$t_i$  = Number of tied values in the  $i^{\text{th}}$  group of ties for the first quantity

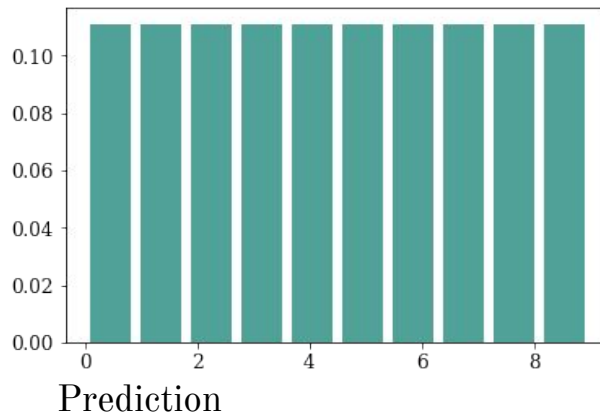
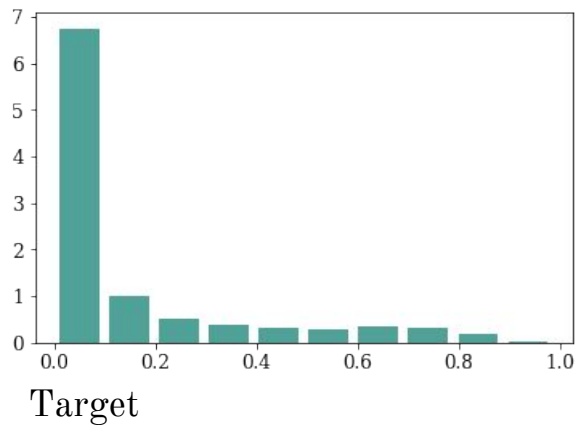
$u_j$  = Number of tied values in the  $j^{\text{th}}$  group of ties for the second quantity





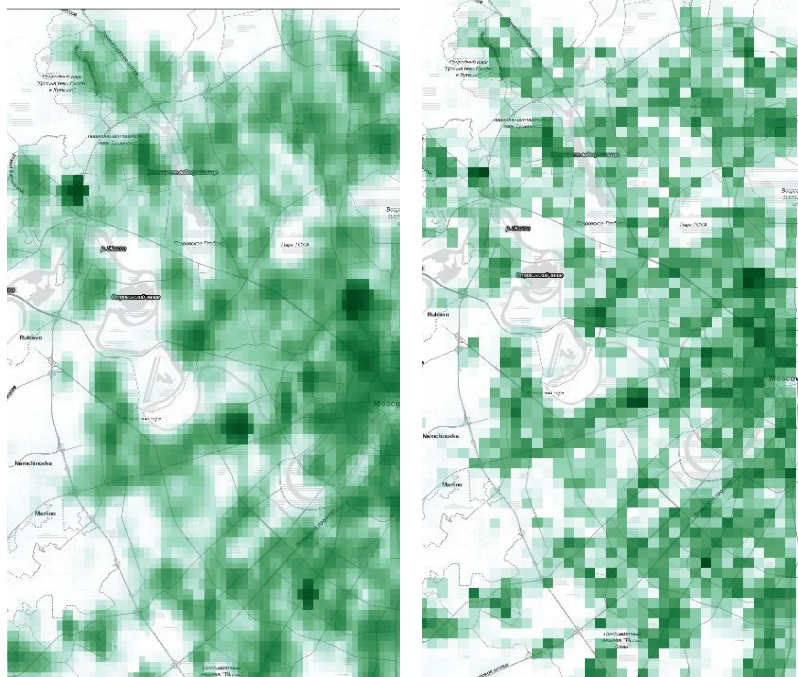


# GeoHack



10 квантилей -  
цифр от 0 до 9  
-> вес csv  
МИНИМАЛЬНО  
ВОЗМОЖНЫЙ)

# Попробовал



- Интерполяцию
- Транзакции с хакатона Райффайзенбанка
- Пожары по округам с [data.mos.ru](http://data.mos.ru)
- Все базовые станции с [opencellid.org](http://opencellid.org)

Спасибо за внимание

