

Скоринг человека. Балансировка участков.

Котик Константин, физфак МГУ
Гамов Артем, физфак МГУ
Волков Александр, МИФИ

Что интересного есть в
данных?

Общее число пациентов в данных

967 044

Число клонов

101



8 человек старше



950 лет !!!

Число пациентов, у которых число совместно
проживающих детей больше 20



222

Число пациентов, которые совместно проживают с
50 и более льготниками



123

Интересная фиксация визитов.

Среднее количество приёмов у людей, которые за 2 года ни разу не появлялись:

12

Средний возраст пациента, который за 2 года ни разу не посетил поликлинику:

47.5

Среднее количество приёмов у людей, которые за 2 года хоть раз появлялись:

15

Средний возраст пациента, который посещал за 2 года хотя бы раз поликлинику:

51



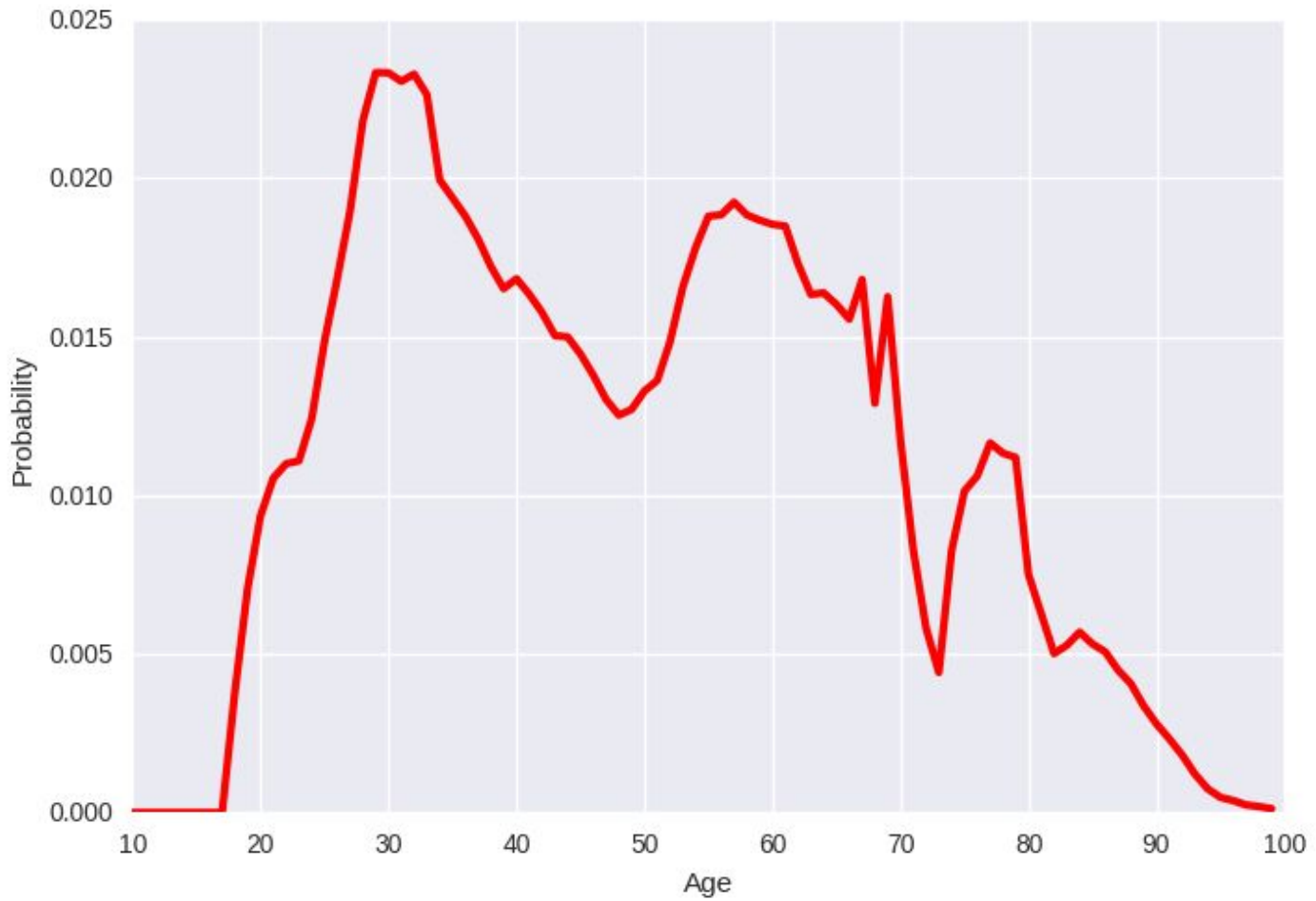
Отношение пациентов к смене участка

Готовы сменить участок



- Неизвестна реакция 755 000 пациентов о возможной смене их участков
- 196 000 пациентов готовы к смене участков
- около 16 000 пациентов не хотят менять участок

Вероятность визита к врачу в зависимости от возраста



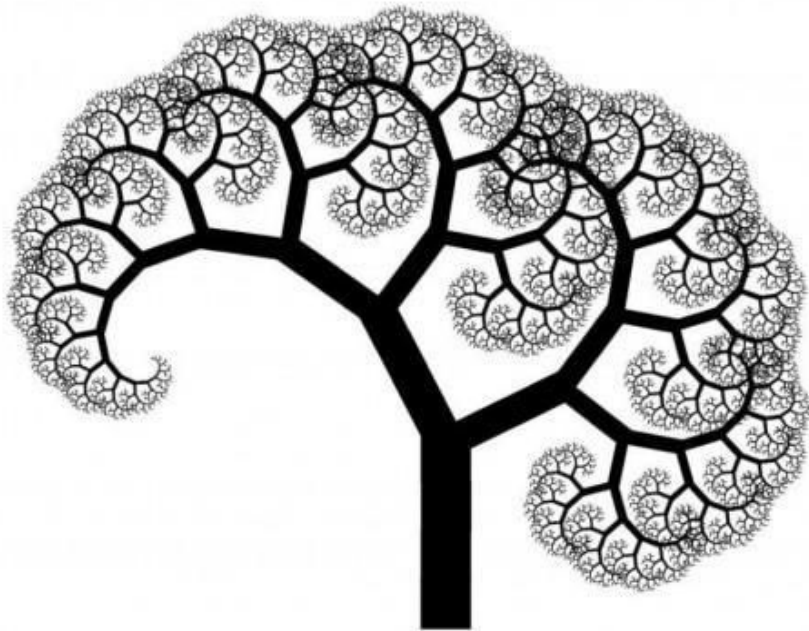
Построение модели
предсказания количества
приёмов

Метрики

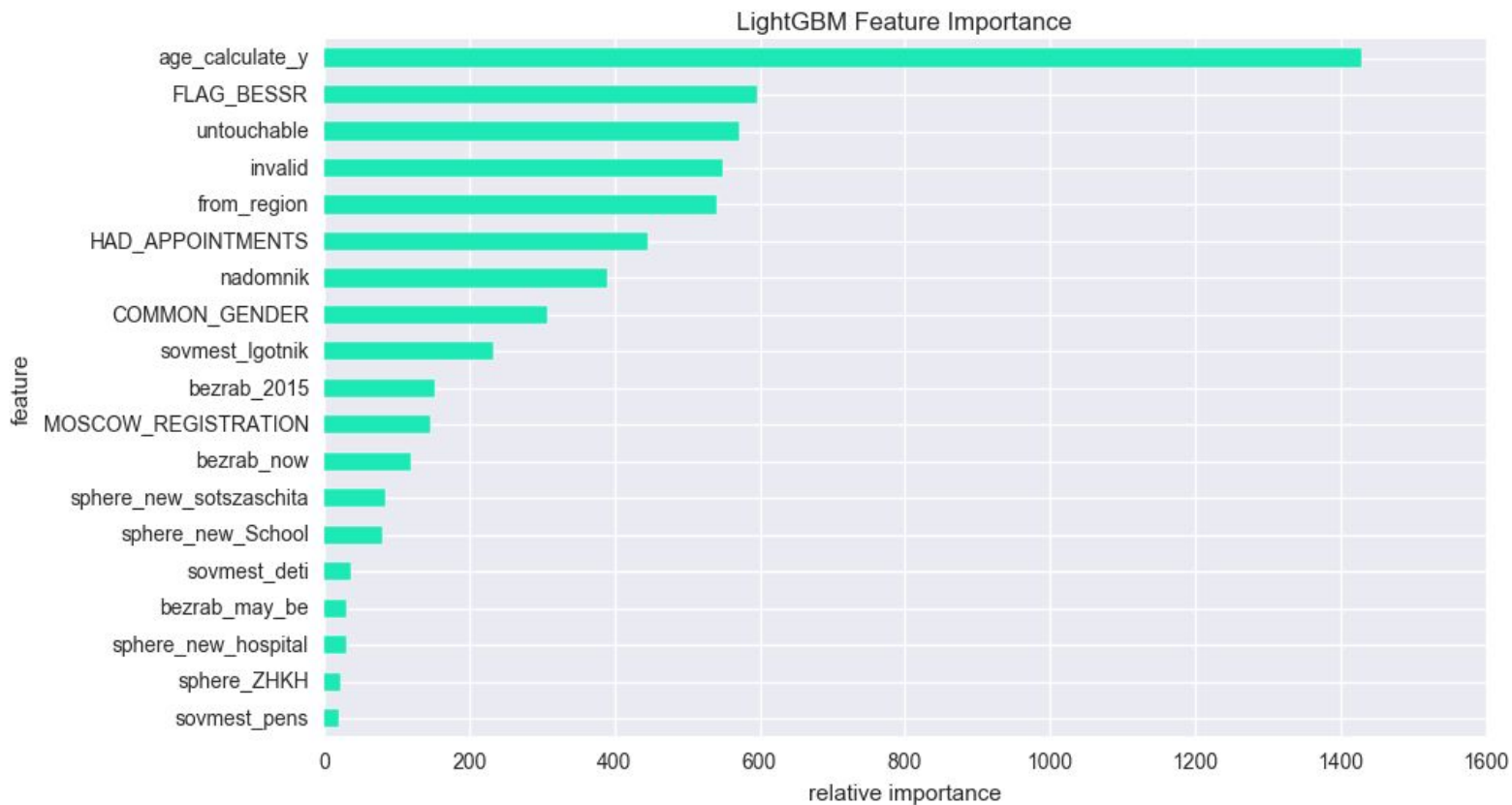
$$MAE = \frac{1}{n} \sum_{j=1}^n |y_j - y'_j|$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \tilde{y}_i)^2$$

Light Gradient Boosting Machine



Значимость признаков в построенной модели

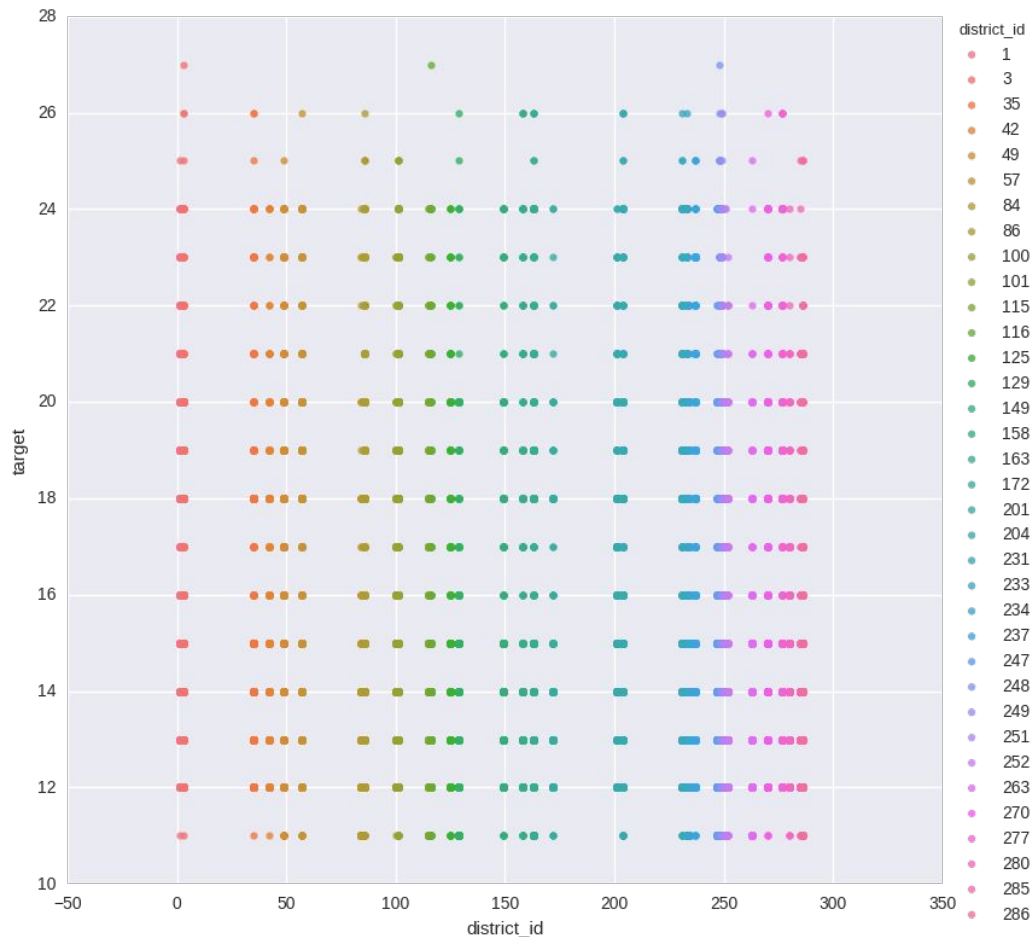


MAE = 1.42

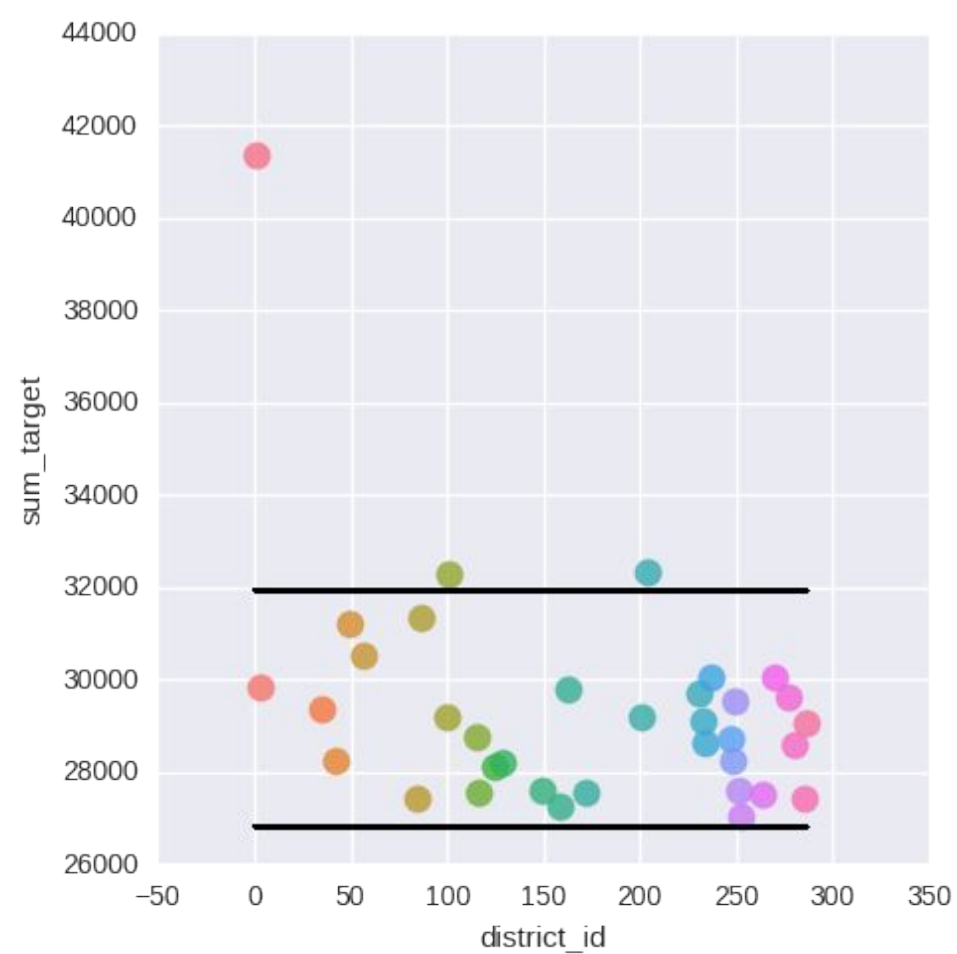
MSE = 6.72

Балансировка участков

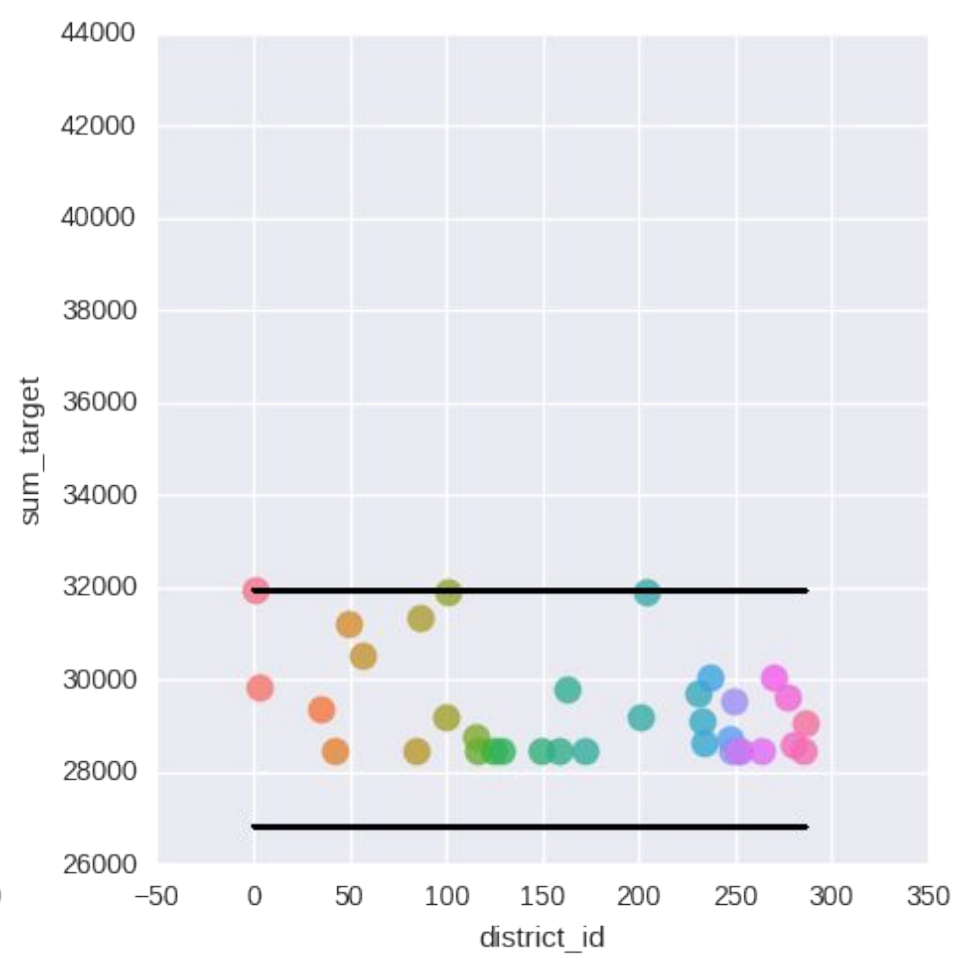
Распределение визитов каждого пациента по участкам LPU_ID=13.



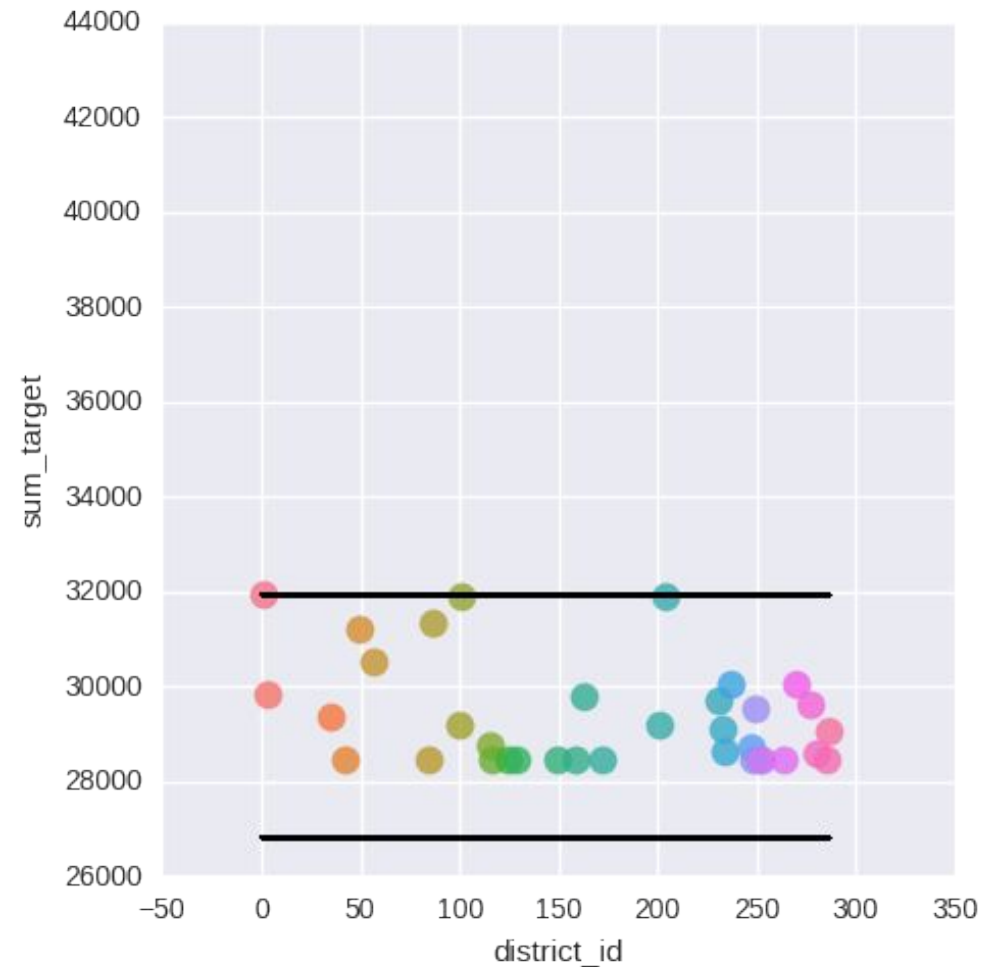
Да уж, из такого
графика не
сбалансируешь :))



Было



Стало



Стало

**Чёрные линии -- это
границы погрешности.**

**Погрешность составляет
8.7 % от целевого среднего
значения.**

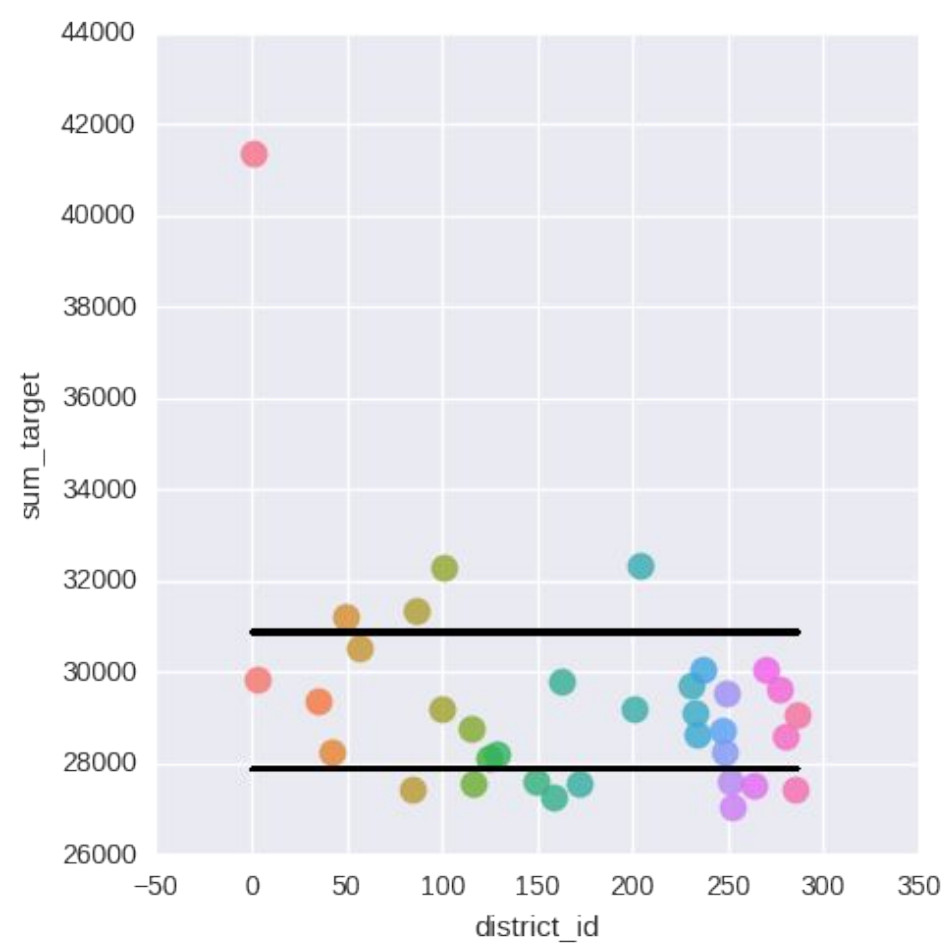
8.7 % рассчитано с учётом MAE построенной
модели по предсказанию target.

**Число перемещённых
людей: 525.**

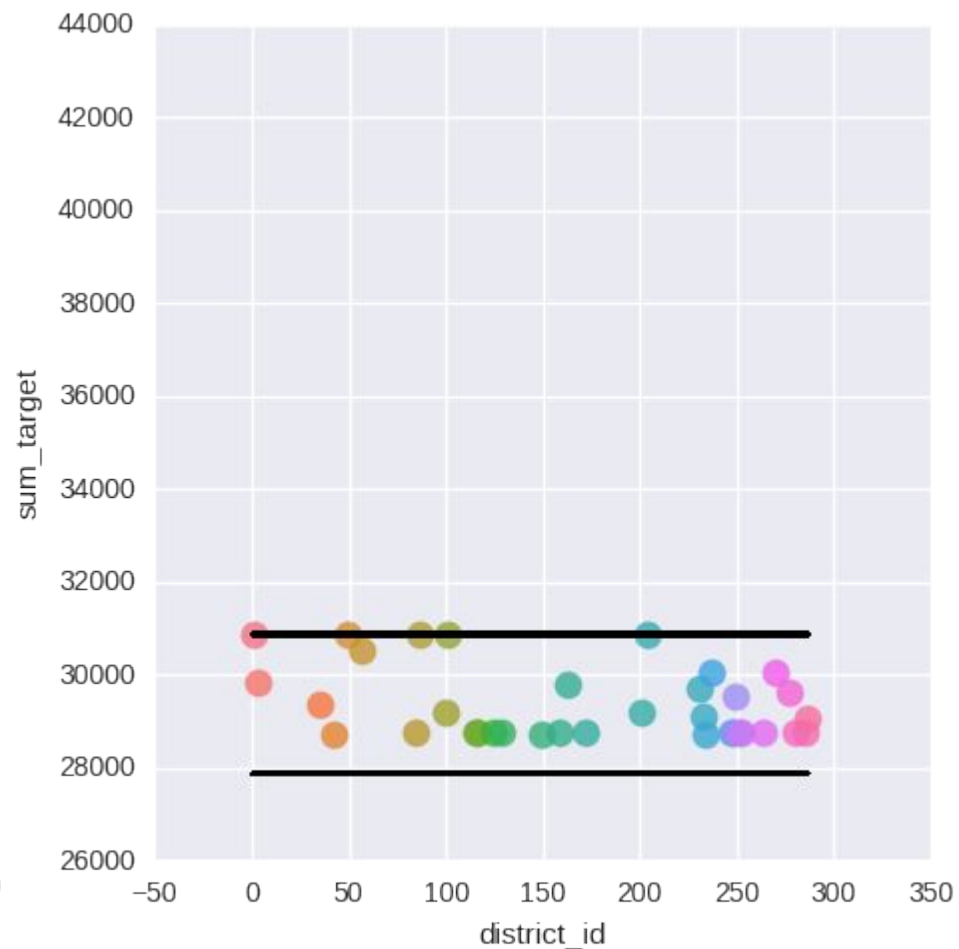
Предположим врачу захотелось погрешность в:

5 %

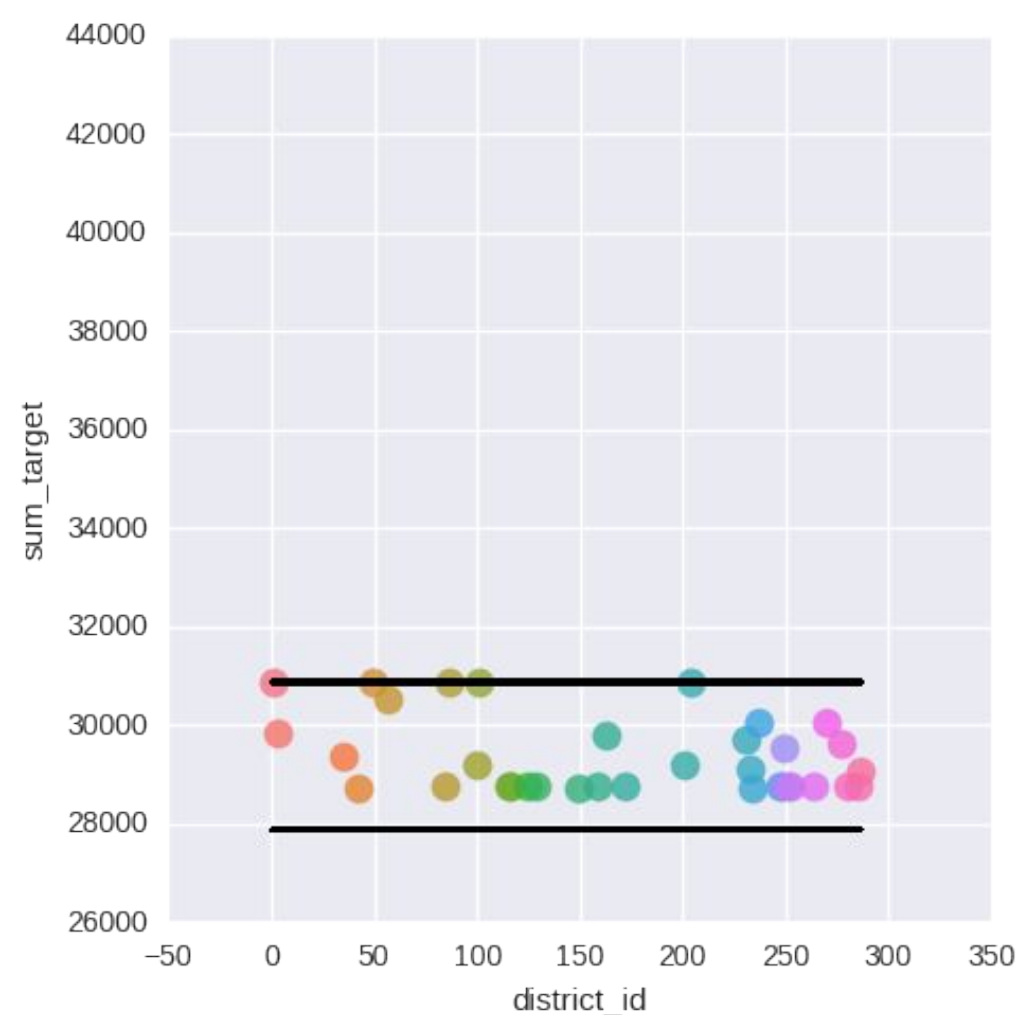
Тогда:



Было



Стало

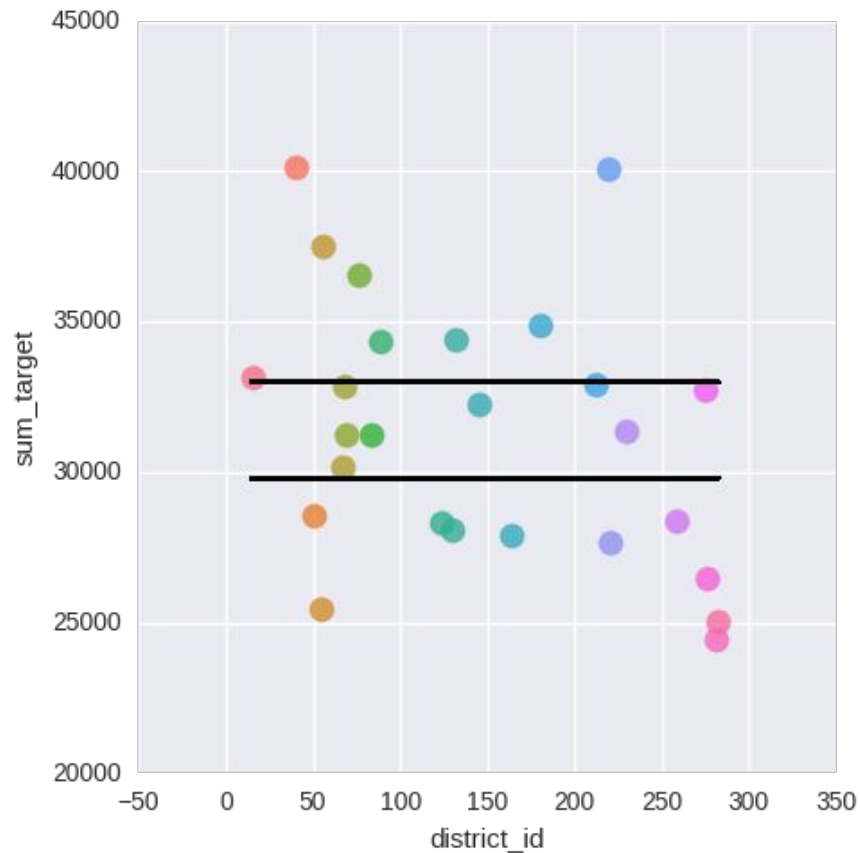


**Погрешность составляет
5 % от целевого среднего
значения.**

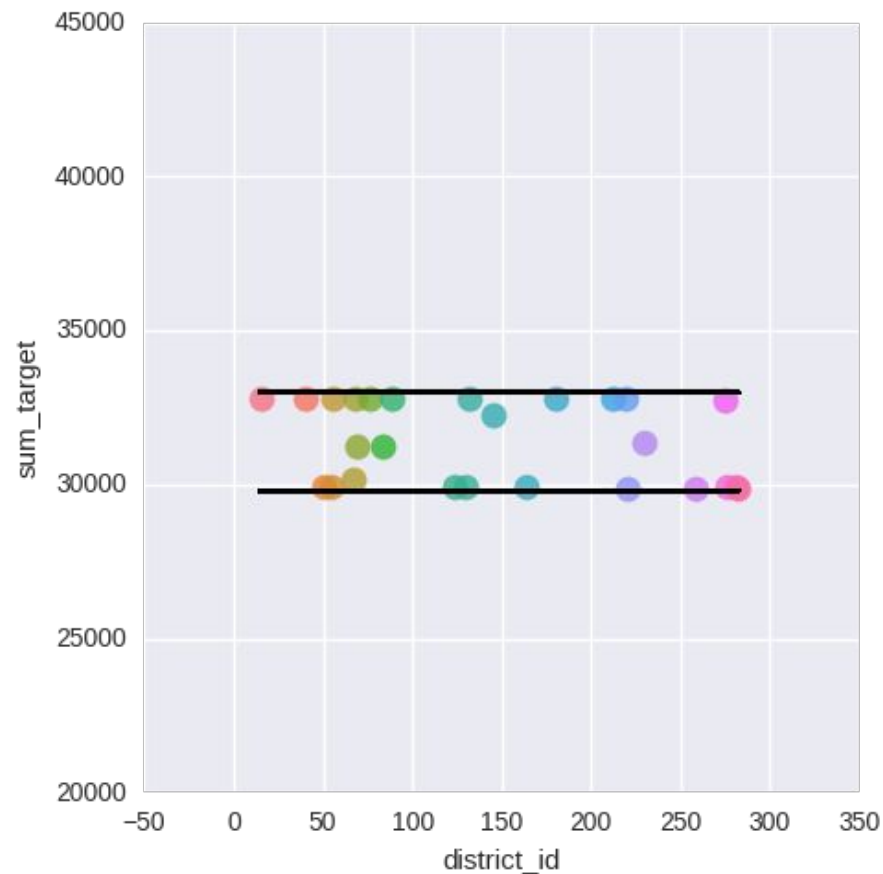
**Число перемещённых
людей: 742.**

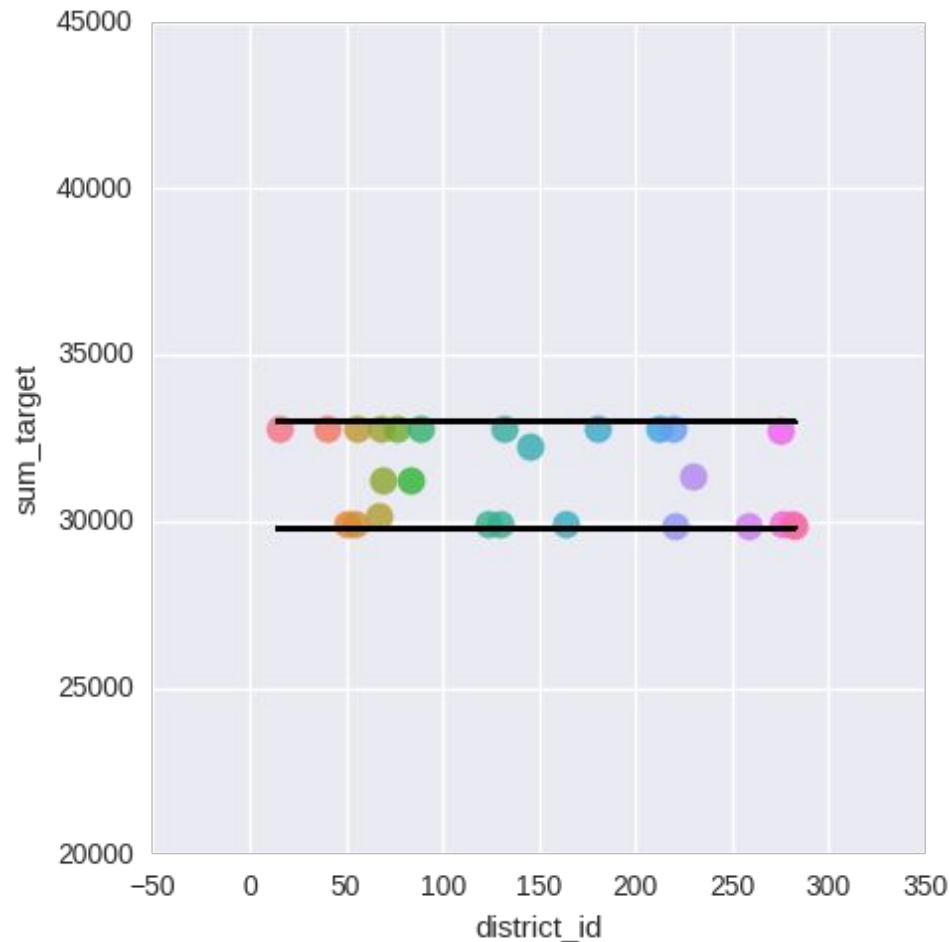
LPU_ID = 17

Было



Стало



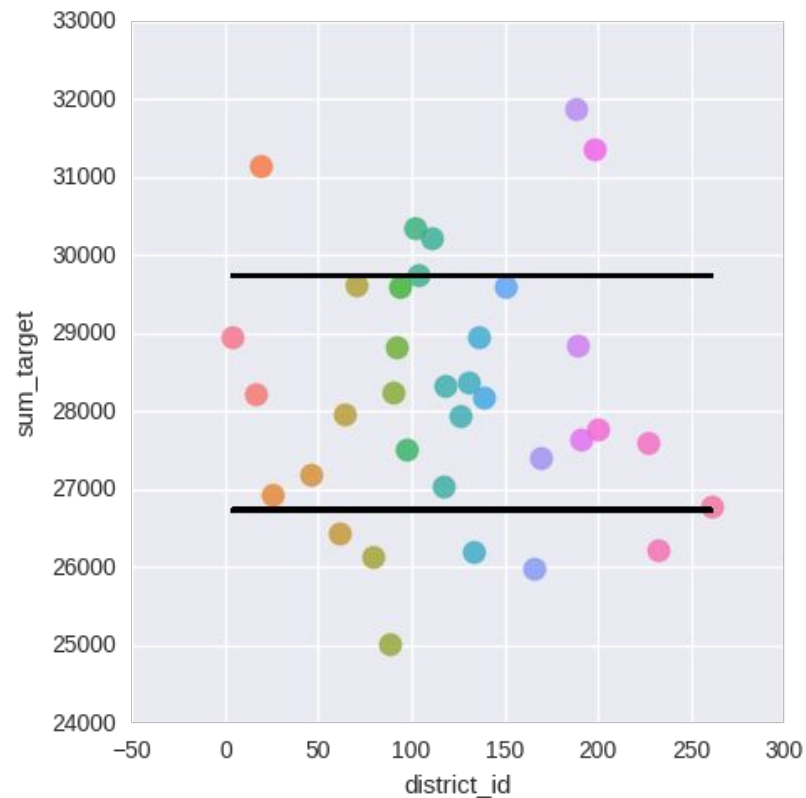


**Погрешность составляет
5 % от целевого среднего
значения.**

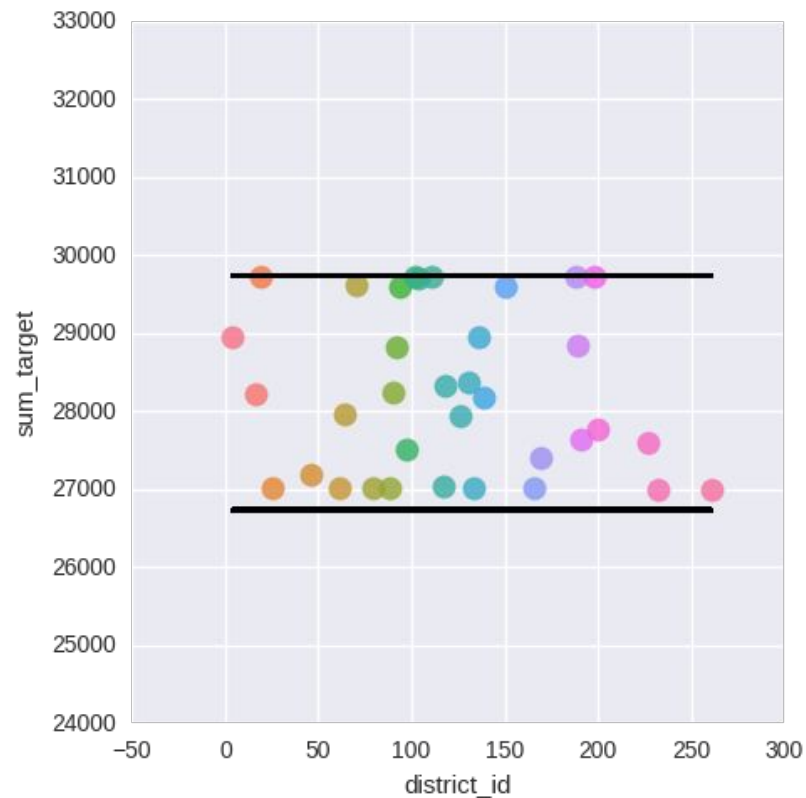
**Число перемещённых
людей: 1336.**

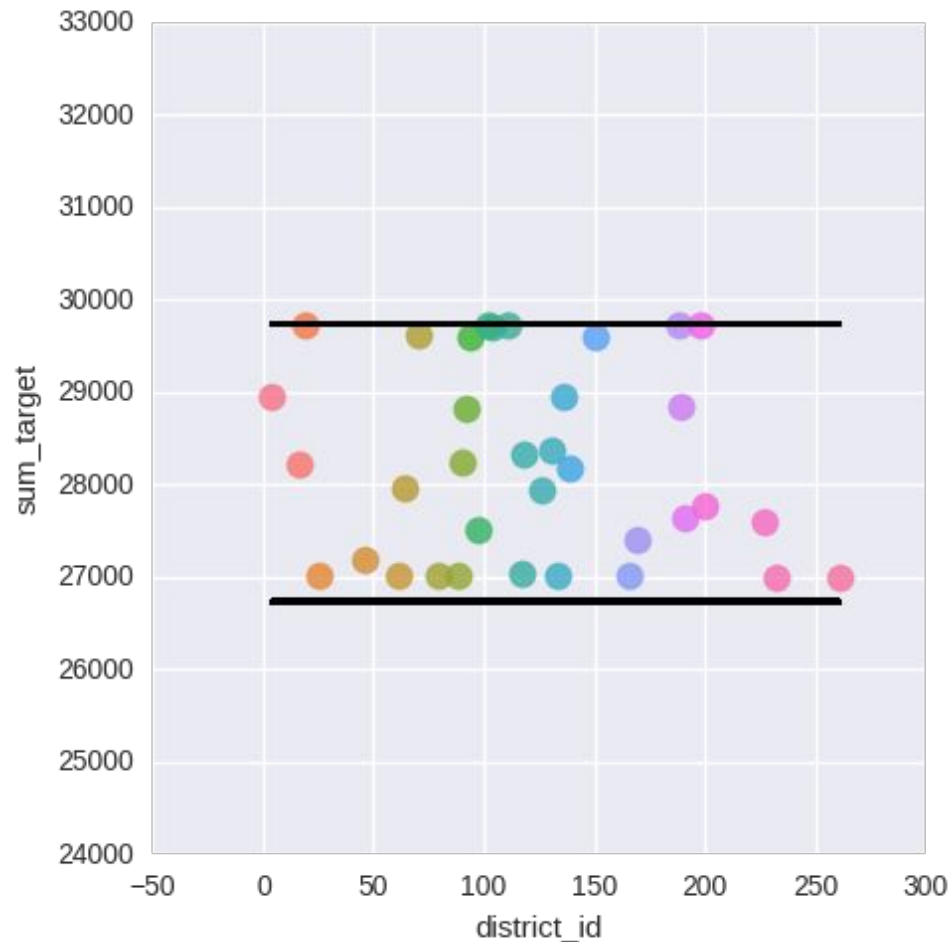
LPU_ID = 19

Было



Стало





**Погрешность составляет
5 % от целевого среднего
значения.**

**Число перемещённых
людей: 332.**

