



Исследование ЦИАН

Симанок Данила Андреевич,
19.Б10-мм,
st074965@student.spbu.ru,
24.03.2023



Введение

Рассматривались предложения посуточной аренды в СПб на 1 гостя.

Выбирались: тип квартиры, площадь, этаж, ближайшая станция, расстояние до нее и способ добраться, улица, цена и залог.

Дизайнерская квартира! Центр!

Студия, 27 м², 5/5 этаж

 Только на Циан

 Чернышевская • 11 минут пешком

Санкт-Петербург, р-н Центральный, Литейный, м. Чернышевская, улица Чайковского, 26

от 2 300 ₽/сутки

Посуточно, без комиссии, залог 3 000 ₽

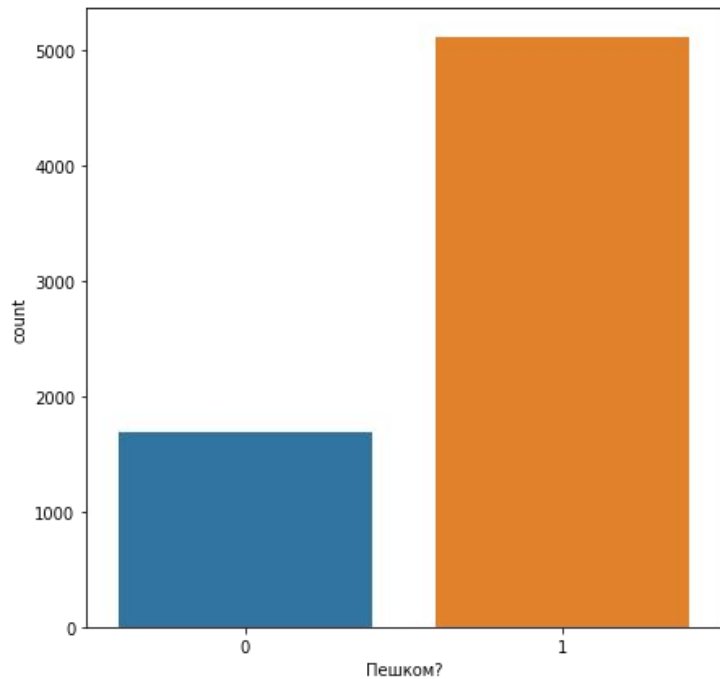
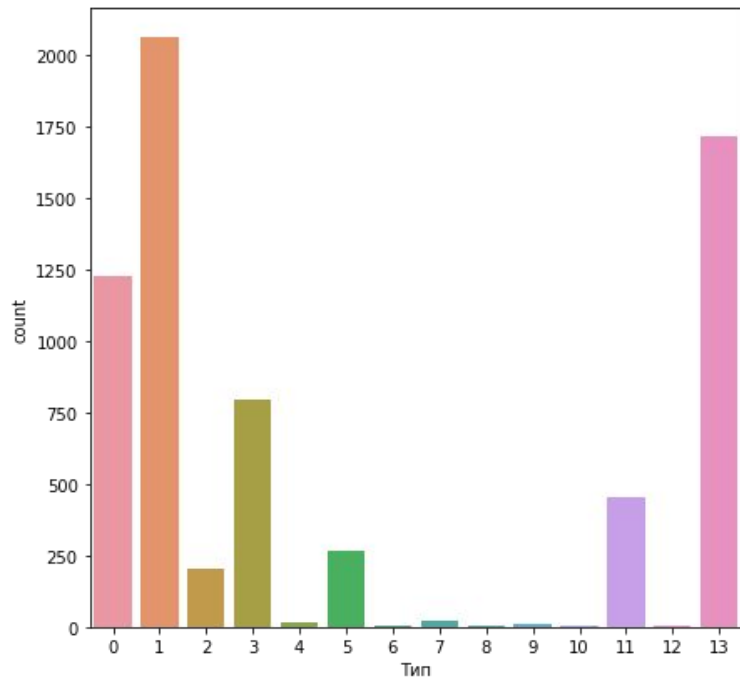
СКИДКИ ДО 40% на ближайшие 1-2 недели ! Зависит от ДЛИТЕЛЬНОСТИ ПРОЖИВАНИЯ!
Минимальная длительность проживания от 2 дней. Выходные +500 руб (пт, сб, вск). с 1/04 по 28/04+1000руб Новая дизайнерская квартира в Центральном районе. Еще 2 дизайнерские квартиры на площадке. Так же еще, интересные квартиры в профиле на ЦИАН. Обратите внимание, квартира находится на 5 этаже без лифта. Подойдет молоды...



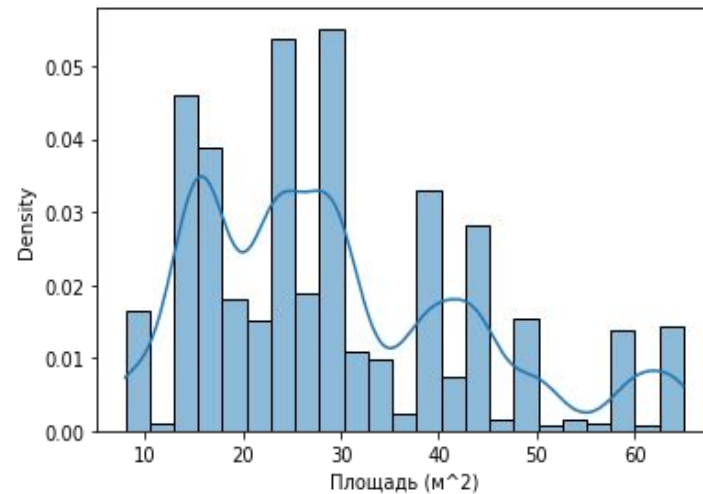
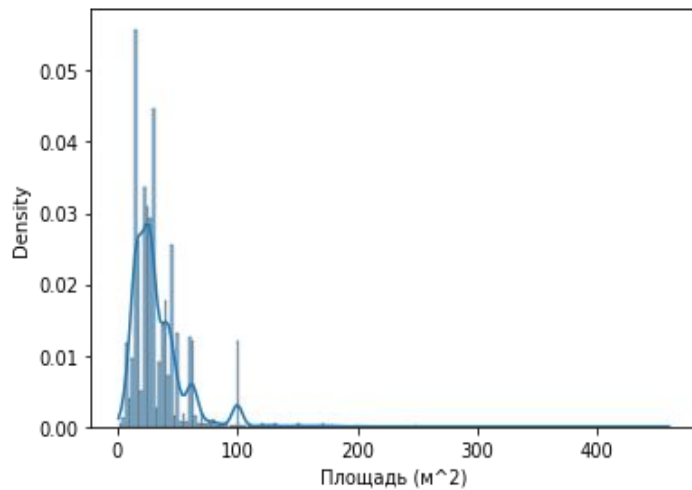
Результаты предобработки

- Только типы 0, 1, 3, 13
- Площадь в диапазоне [0; 65]
- Только станции, встречающиеся от 150 до 420 раз
- Время в диапазоне [5; 20]
- По способу добраться нужна балансировка
- Только улицы, встречающиеся от 75 до 200 раз
- Только цены в диапазоне [1200; 4000)
- Залог целесообразнее разделить на классы "<3000", "3000", ">3000"

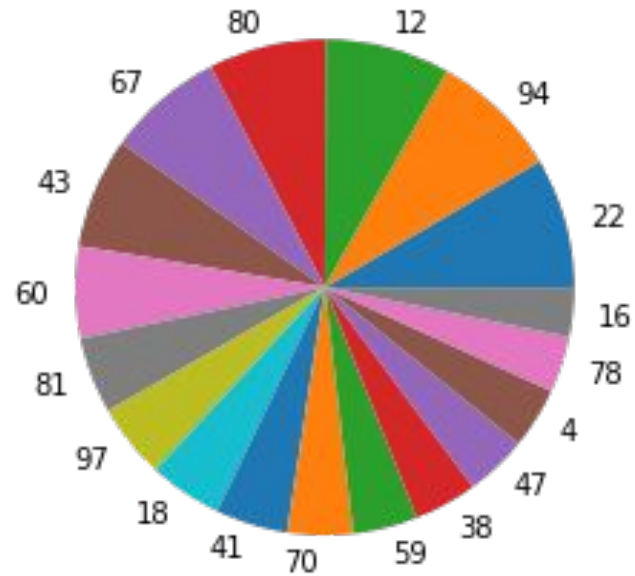
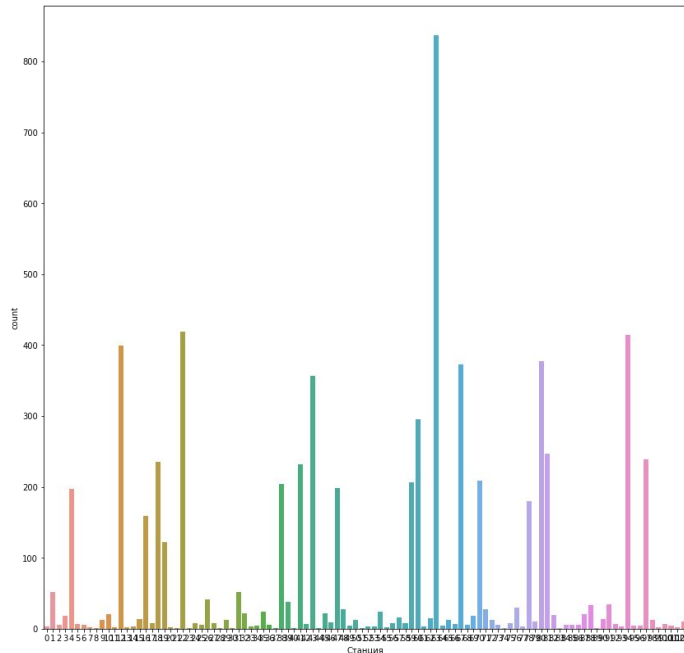
Предварительная обработка | типы, пешком



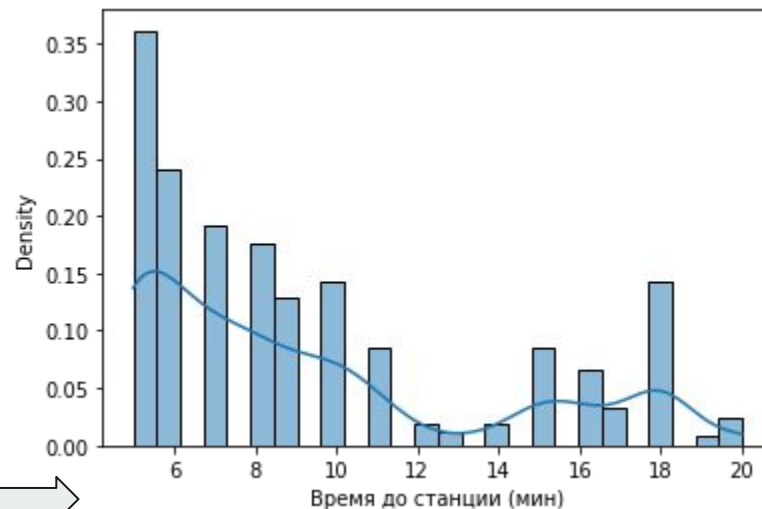
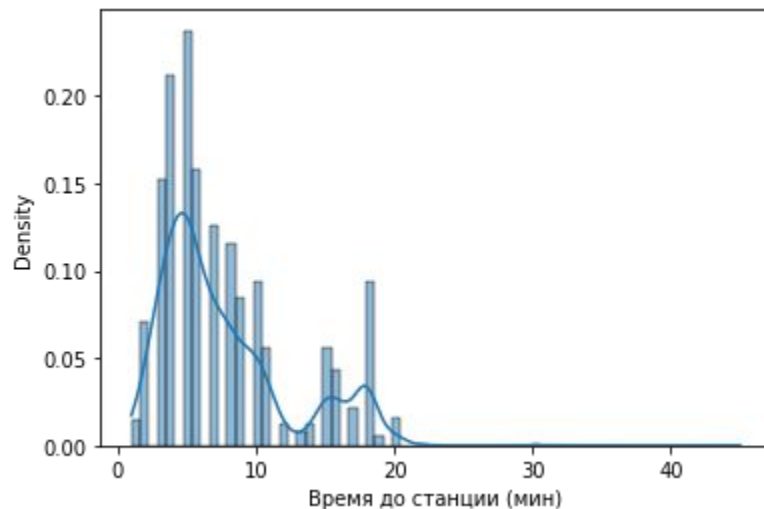
Предварительная обработка | площадь



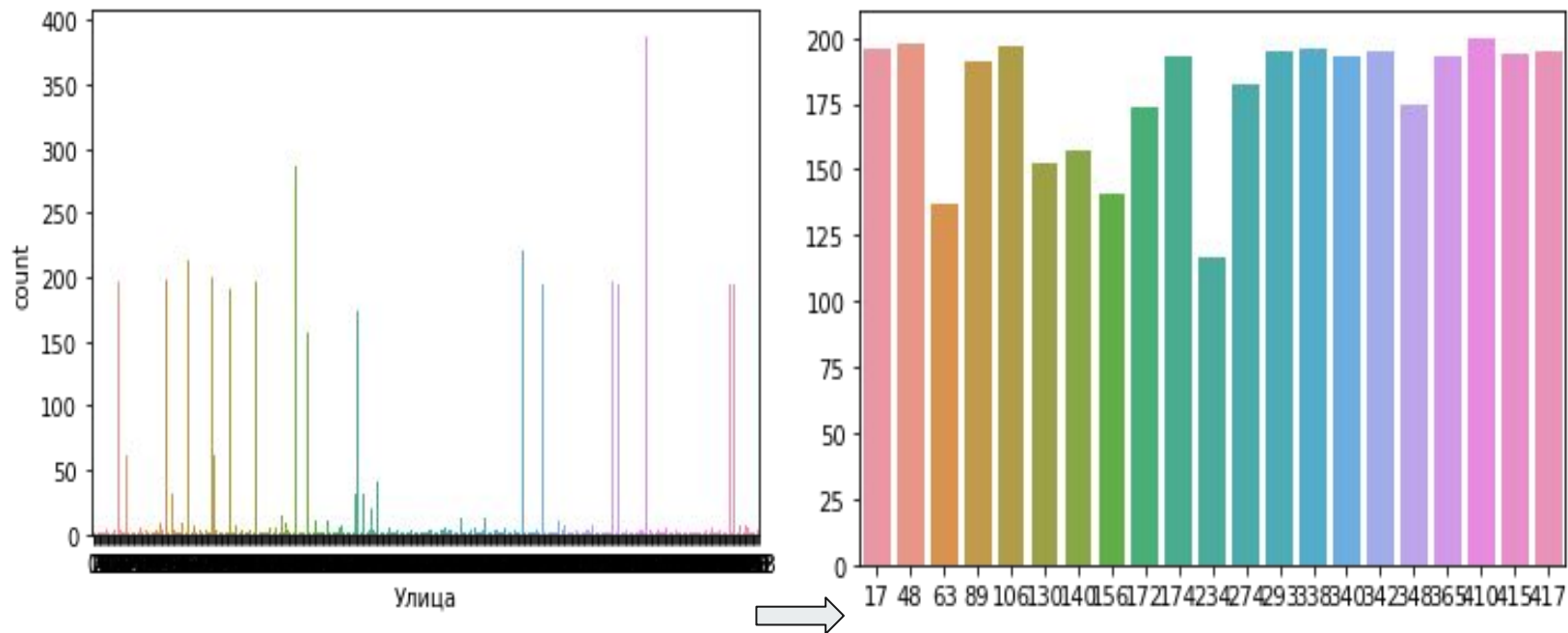
Предварительная обработка | станции



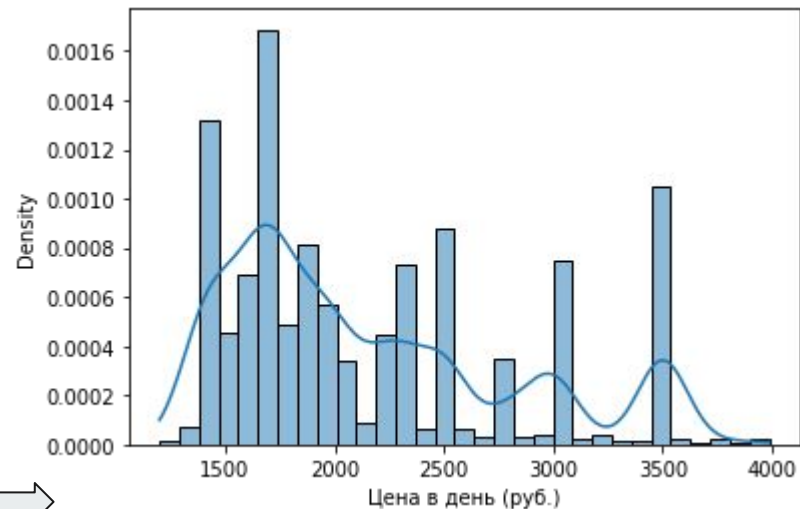
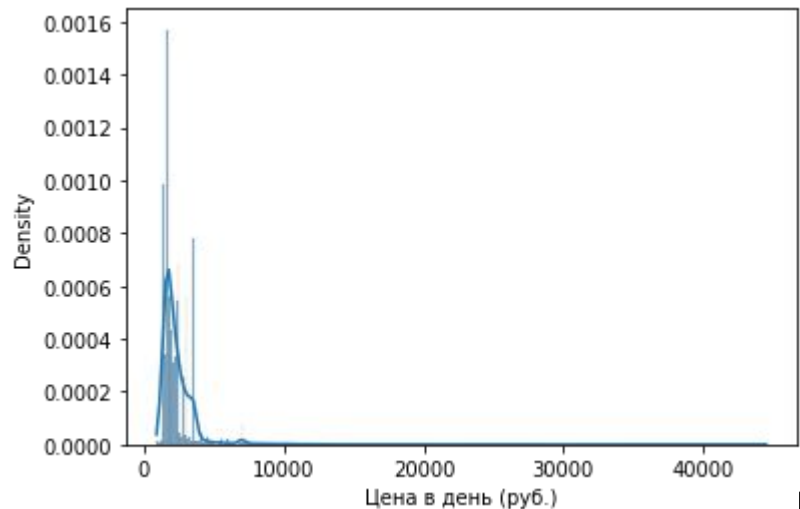
Предварительная обработка | время



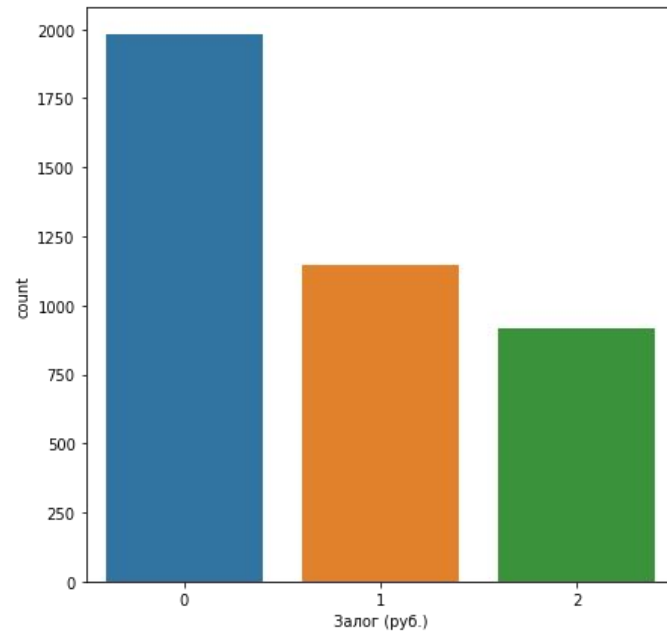
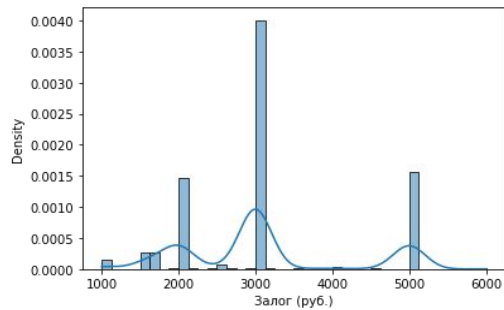
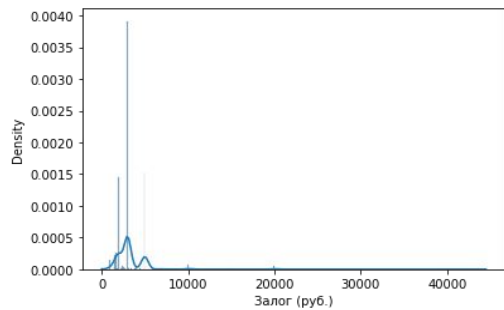
Предварительная обработка | улица



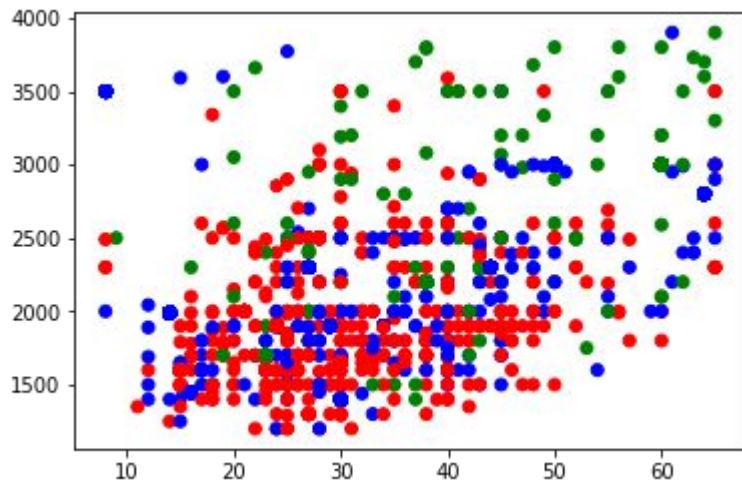
Предварительная обработка | цена



Предварительная обработка | залог

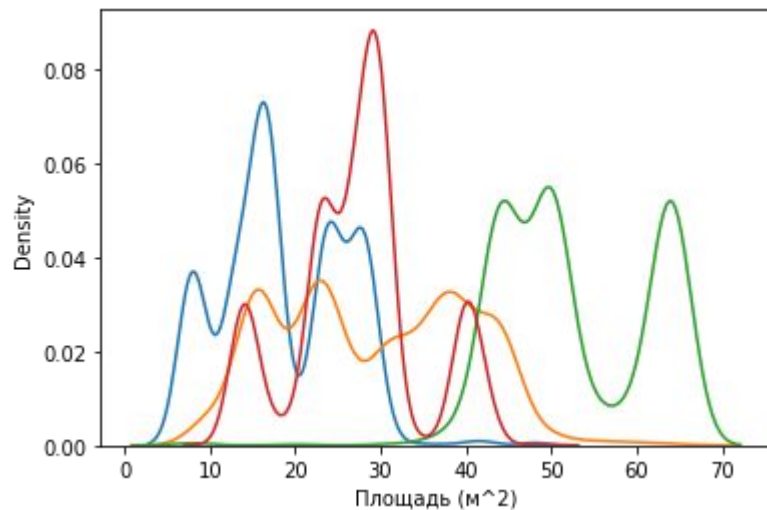


Взаимодействия | Цена (Площадь, Залог)



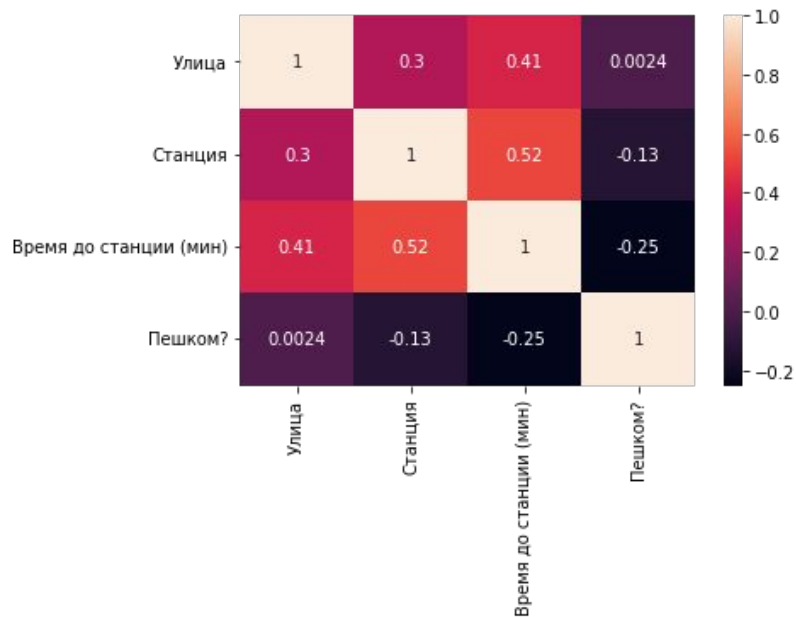
- Для предсказания цены недостаточно
- Залог зависит от цены и площади

Взаимодействия | Площадь (Тип)



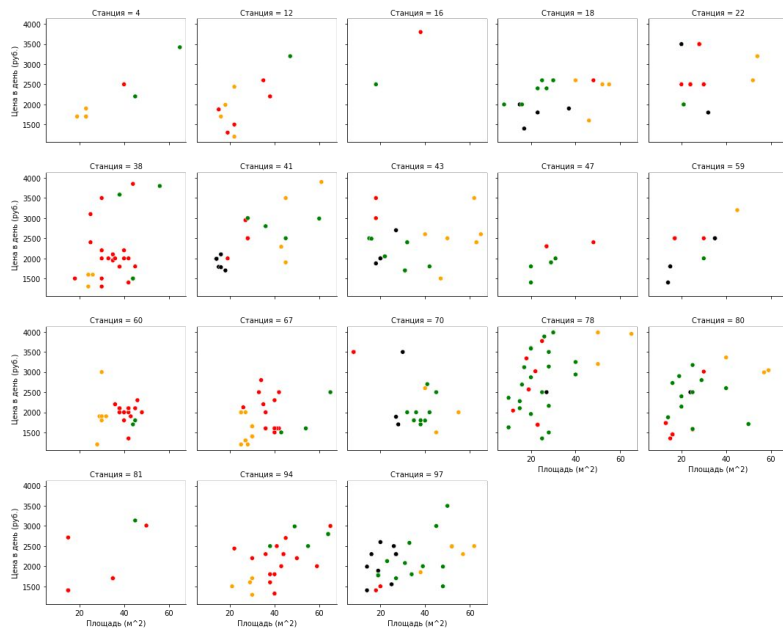
- По площади можно сделать предположение о типе

Взаимодействия | Местоположение



- Можно ожидать зависимость времени от улицы и станции

Взаимодействия | Цена (Тип, Площадь)



- Зачастую для данной станции можно построить зависимость цены от типа или площади



Тренировка моделей | площадь -> тип

- Модель: GradientBoostingClassifier
- Параметры: {'criterion': 'friedman_mse', 'learning_rate': 0.1, 'loss': 'log_loss', 'max_features': 'auto'}

	precision	recall	f1-score	support
0	0.60	0.95	0.73	207
1	0.86	0.68	0.76	622
3	0.69	0.96	0.80	128
13	0.92	0.84	0.88	467
accuracy			0.79	1424
macro avg	0.77	0.86	0.79	1424
weighted avg	0.82	0.79	0.80	1424



Тренировка моделей | (цена, площадь) -> залог

- Модель: GradientBoostingClassifier
- Параметры: {'criterion': 'friedman_mse', 'learning_rate': 1, 'loss': 'log_loss', 'max_features': 'log2'}

	precision	recall	f1-score	support
0	0.94	0.96	0.95	491
1	0.89	0.87	0.88	265
2	0.90	0.88	0.89	182
accuracy			0.92	938
macro avg	0.91	0.90	0.91	938
weighted avg	0.92	0.92	0.92	938



Тренировка моделей | (улица, станция) -> время

- Модель: DecisionTreeRegressor
- Параметры: {'criterion': 'poisson', 'max_features': 'log2', 'min_samples_leaf': 4, 'min_samples_split': 2, 'splitter': 'best'}
- Метрики
 - MSE = 0.2772601815215786
 - MAE = 0.0738775342914865
 - R2 = 0.9863327325996838



Тренировка моделей | (площадь, станция, тип) -> цена

- Модель: SVR
- Параметры: {'C': 8, 'epsilon': 0.125, 'gamma': 'auto', 'kernel': 'rbf', 'tol': 1e-05}
- Метрики
 - MSE = 42411.85531151236
 - MAE = 81.5678275983627
 - R2 = 0.8882710938522076



Исследование ЦИАН

Симанок Данила Андреевич,
19.Б10-мм,
st074965@student.spbu.ru,
24.03.2023