

Университет ИТМО

Практическая работа №2
по дисциплине «Визуализация и моделирование»

Автор: Горбатовский Алексей Валерьевич

Поток: ВИМ 1.1

Группа: К3220

Факультет: ИКТ

Преподаватель: Чернышева А.В.

Санкт-Петербург, 2021 г.

Ссылка на датасет: <https://www.kaggle.com/sameep98/housing-prices-in-mumbai>

Краткое описание датасета:

Датасет содержит в себе информацию о домах в Мумбае, их ценой, площадью и другими характеристиками.

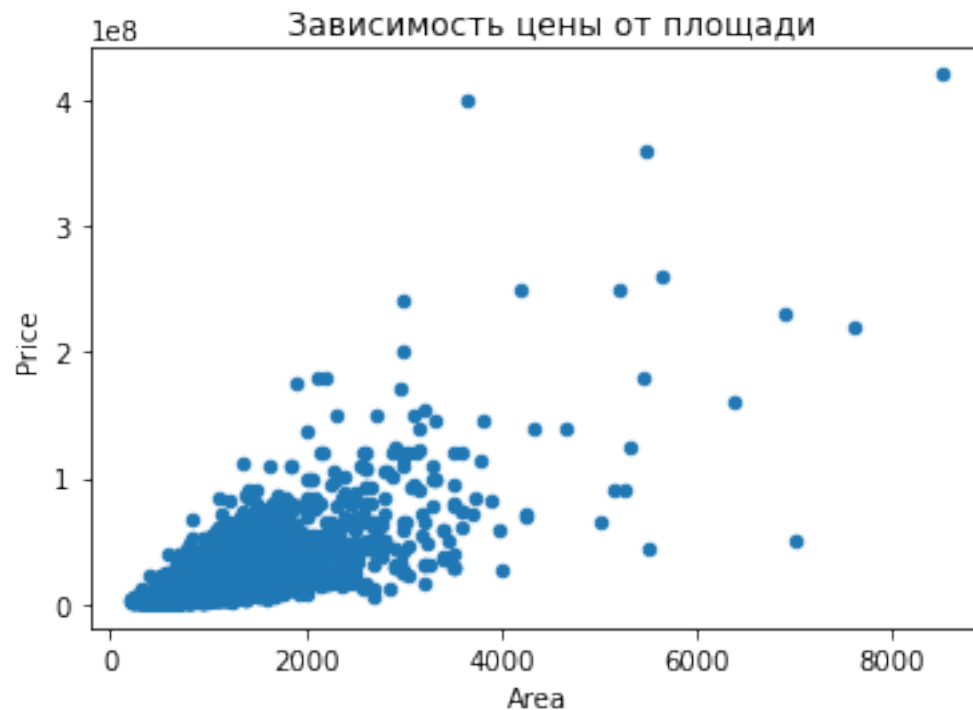
Название столбца	Данные, хранящиеся в столбце	Тип данных	Шкала
ID	Id дома	Целое число	Номинальная
Price	Цена дома (USD)	Целое число	Относительная
Area	Площадь дома (кв. метры)	Целое число	Относительная
Location	Расположение	Строка	Номинальная
No. of Bedrooms	Количество спален	Целое число	Относительная
New/Resale	Новый/Перепроданный (1/0)	Бинарный	Дихотомическая
Gymnasium	Отсутствие/Наличие (0/1) школы рядом	Бинарный	Дихотомическая
Lift Available	Отсутствие/Наличие (0/1) лифта	Бинарный	Дихотомическая
Car Parking	Отсутствие/Наличие (0/1) парковки	Бинарный	Дихотомическая
Maintenance Staff	Отсутствие/Наличие (0/1) обслуживающего персонала	Бинарный	Дихотомическая
24x7 Security	Отсутствие/Наличие (0/1) охраны 24/7	Бинарный	Дихотомическая
Children's Play Area	Отсутствие/Наличие (0/1) детской площадки	Бинарный	Дихотомическая
Clubhouse	Отсутствие/Наличие (0/1) Ночного клуба	Бинарный	Дихотомическая
Intercom	Отсутствие/Наличие (0/1) внутренней системы коммуникаций	Бинарный	Дихотомическая
Landscaped Gardens	Отсутствие/Наличие (0/1) ландшафтных садов	Бинарный	Дихотомическая
Indoor Games	Отсутствие/Наличие (0/1) спортивной площадки	Бинарный	Дихотомическая
Gas Connection	Отсутствие/Наличие (0/1) газа	Бинарный	Дихотомическая
Jogging Track	Отсутствие/Наличие (0/1) пешеходной тропинки	Бинарный	Дихотомическая
Swimming Pool	Отсутствие/Наличие (0/1) бассейна	Бинарный	Дихотомическая

Задачи, которые можно решить на данном датасете:

1. Нахождение оптимального варианта дома в зависимости от всех характеристик
2. Предсказывание стоимости дома, в зависимости от соответствующих характеристик

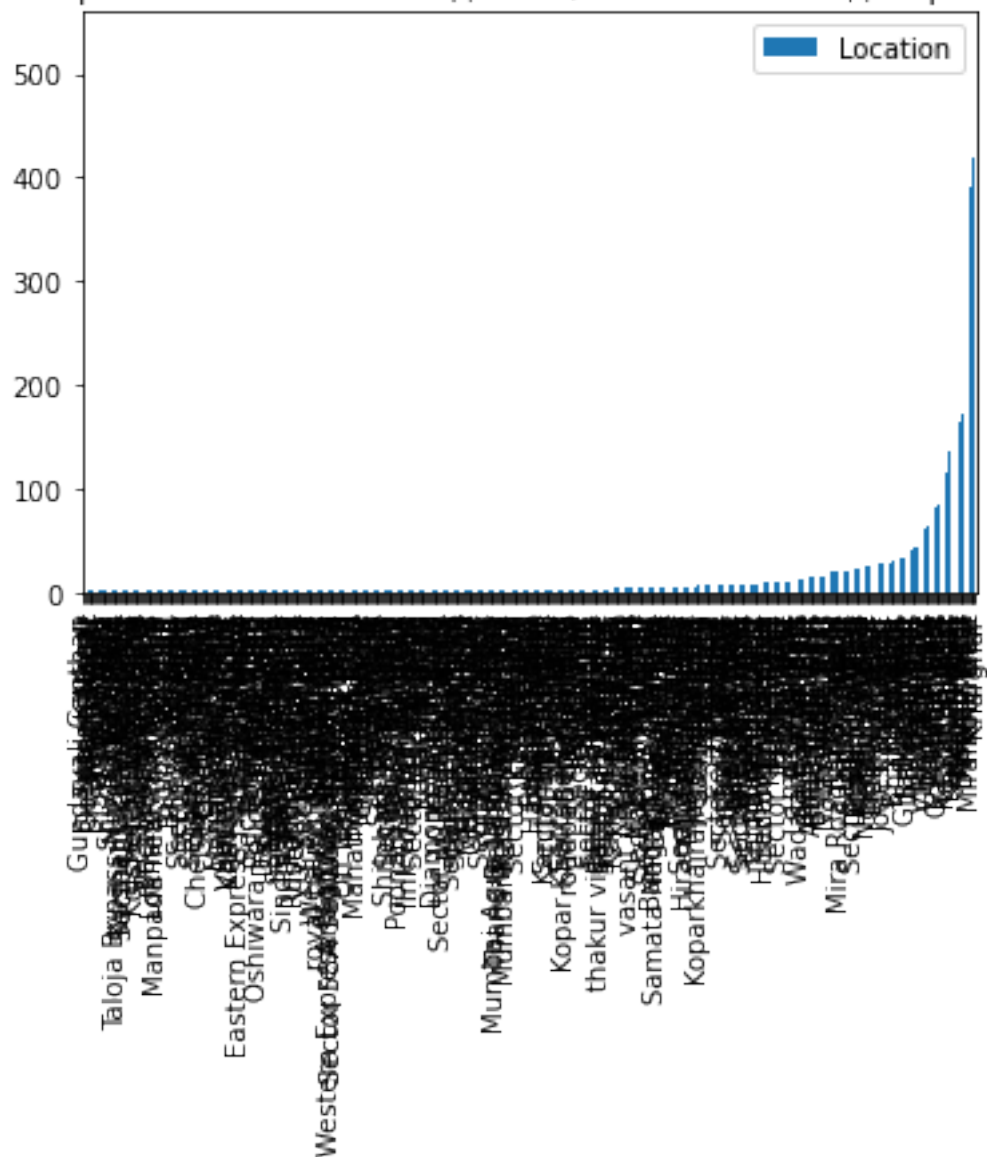
Гипотезы, проверенные во время работы (выводы в скобках):

1. Зависимость цены от площади во всей таблице данных (Вывод: чем больше площадь, тем выше цены, но выборка слишком большая, нужно сократить кол-во объектов)



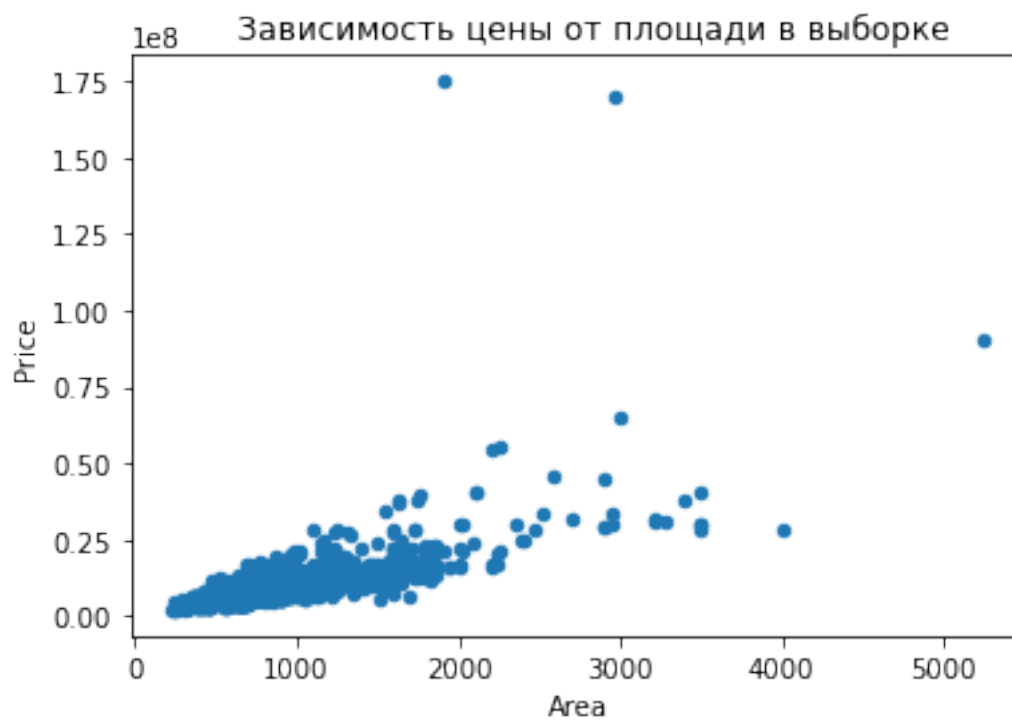
2. Возможность рассмотрения всех локаций датасета (Вывод: Все районы разом рассматривать бессмысленно)

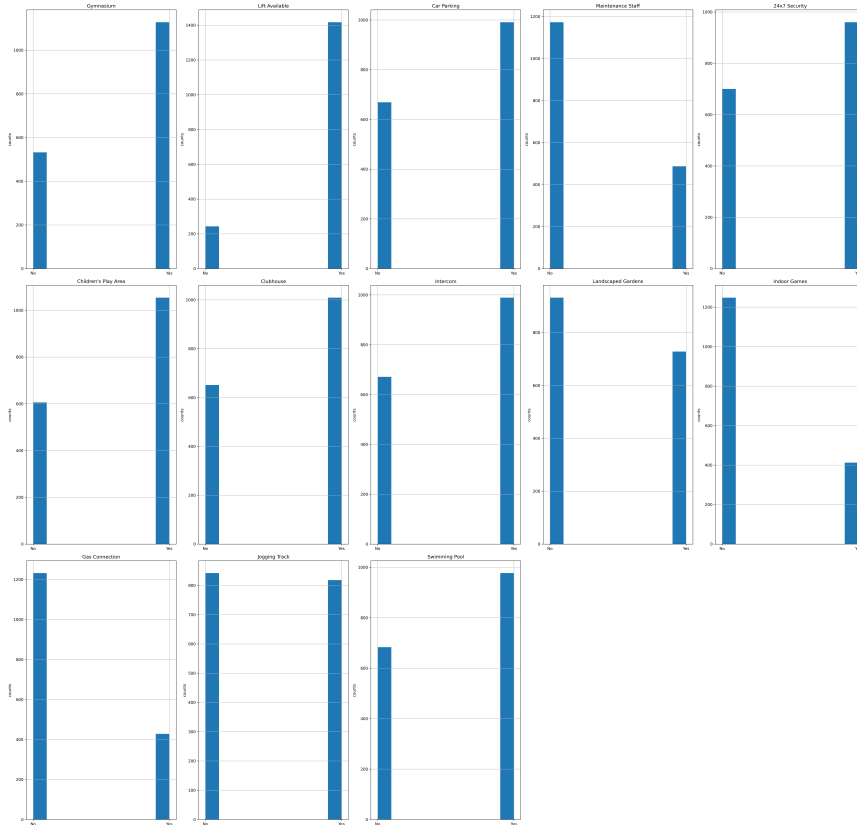
Все районы и количество домов (нечитабельная диаграмма)





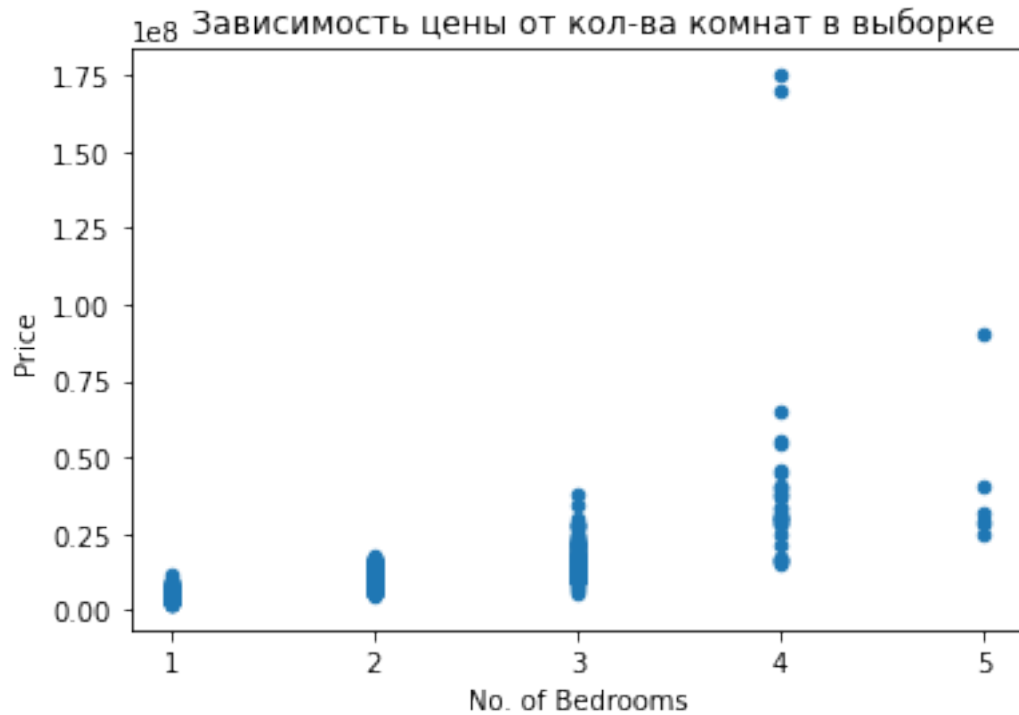
3. Чем больше площадь тем больше цена в топ 4 районах Мумбаи (Вывод: Чем больше площадь, тем больше цена дома, линейная зависимость)





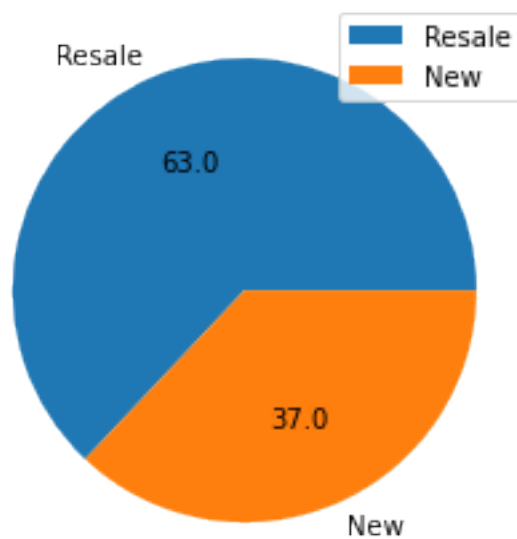
4. Домов со школой в окрестностях меньше чем домой без школы (Вывод: домов со школой больше, чем без)
5. Домов с лифтом меньше чем домой без лифта (Вывод: домов без лифта меньше, чем домов с лифтом)
6. Домов с парковкой меньше чем домой без парковки (Вывод: домов с парковкой больше)
7. Домов с персоналом меньше чем домой без персонала (Вывод: домов без персонала меньше)
8. Домов с охраной меньше чем домой без охраны (Вывод: домов без охраны площадки меньше)
9. Домов с детской площадкой меньше чем домой без детской площадки (Вывод: домов без детской площадки меньше)
10. Домов с клубом меньше чем домой без клуба (Вывод: домов без клуба меньше)
11. Домов с Intercom меньше чем домой (Вывод: домов без Intercom меньше)
12. Домов с зеленым двориком меньше (Вывод: домов без зеленого дворика больше)
13. Домов с Indoor Games меньше (Вывод: домов без Indoor Games больше)
14. Домов с газом меньше (Вывод: домов без газа больше)

15. Домов с пешеходной дорожкой меньше (Вывод: домов с пешеходной дорожкой меньше)
16. Домов с бассейном меньше (Вывод: домов без бассейна меньше)
17. Зависимость цены от кол-ва комнат (Вывод: чем больше комнат, тем выше цена дома)

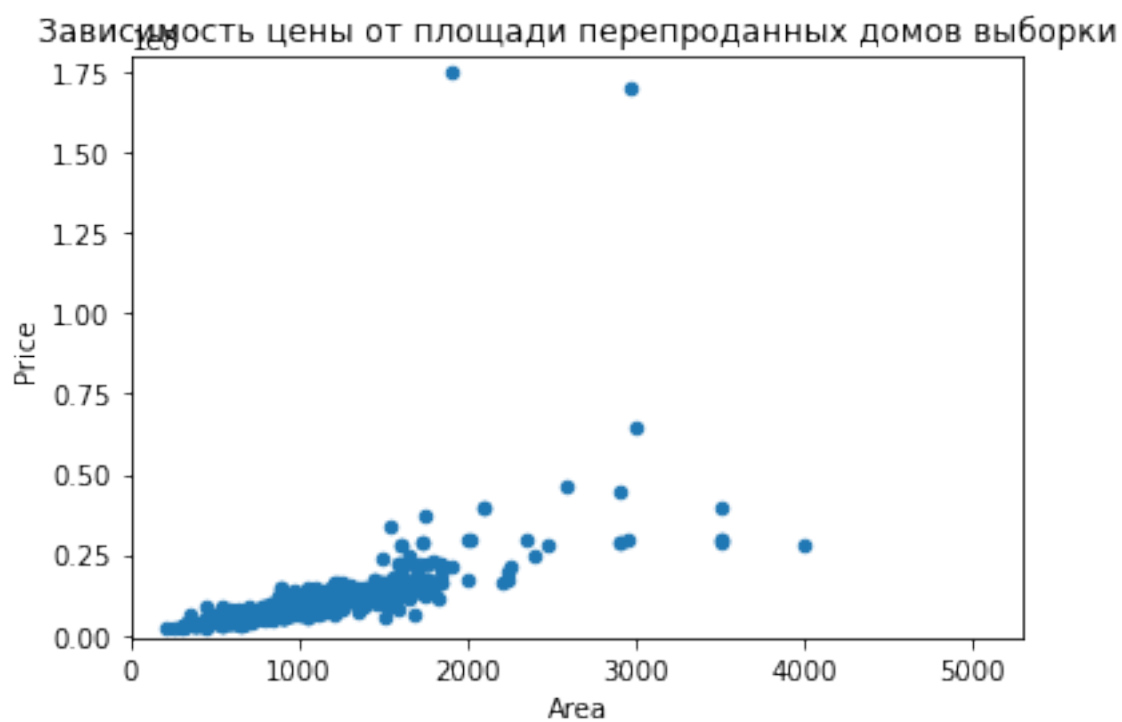


18. Соотношение старых и новых домов в выборке (Вывод: старых домов больше чем новых)

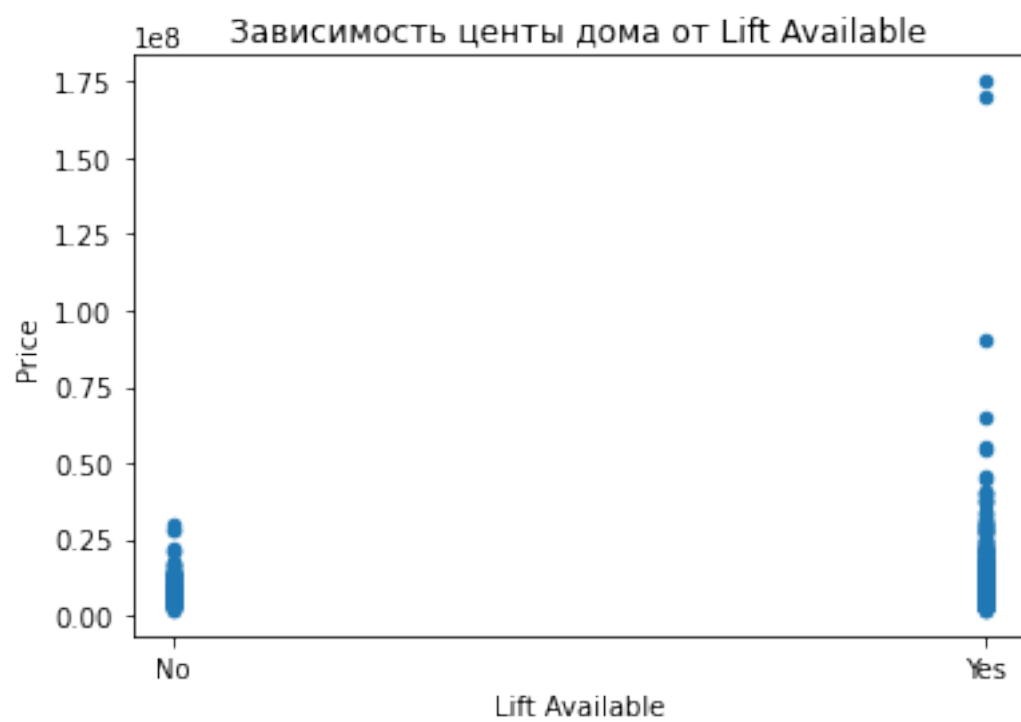
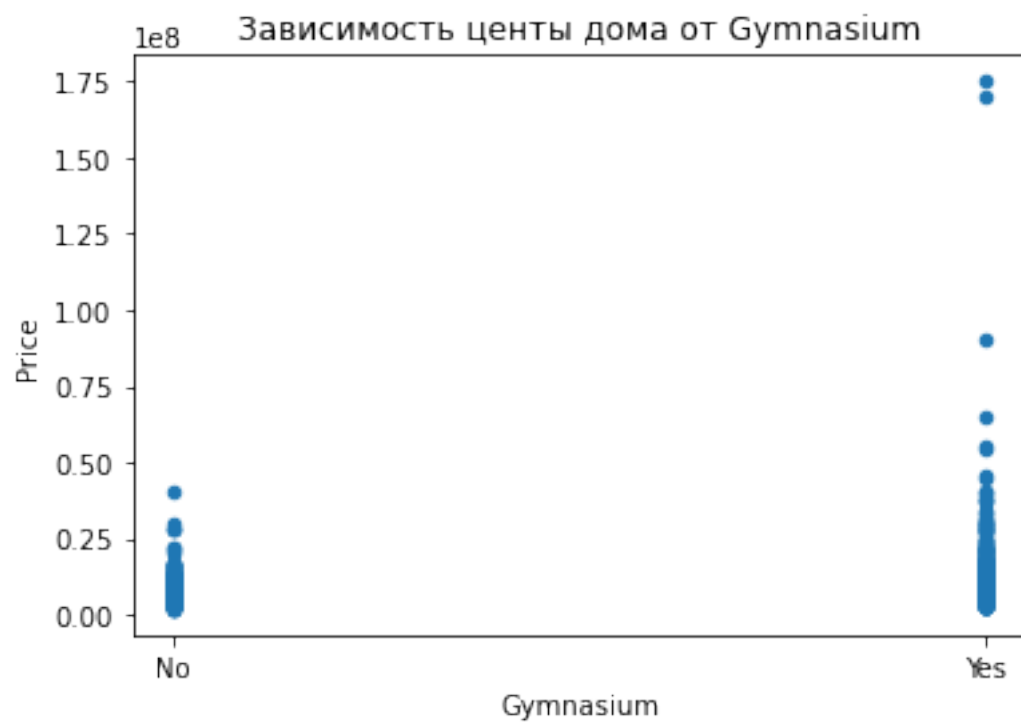
Соотношение новых и перепроданных домов



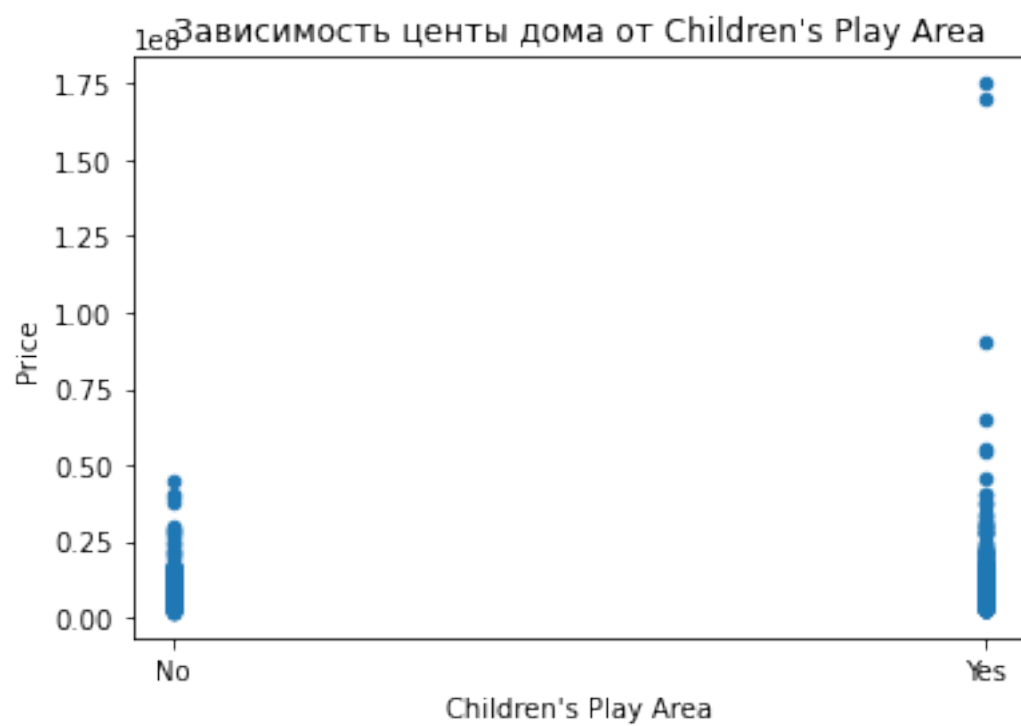
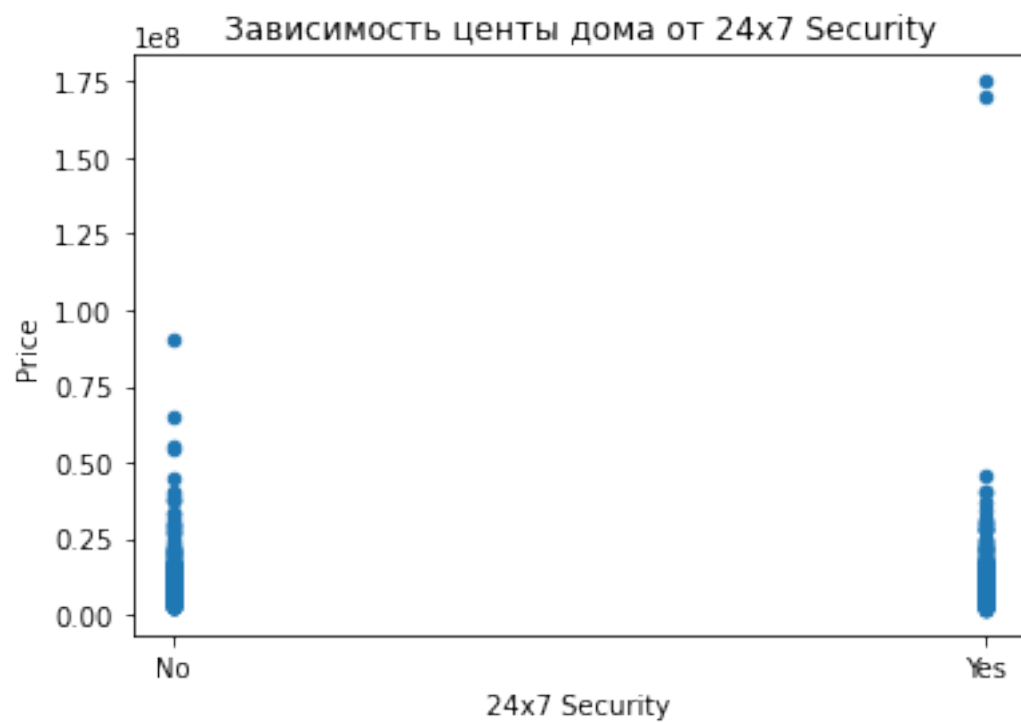
19. Динамика цен на старые и новые дома (Вывод: в основном, новые дома с той же площадью дороже старых)

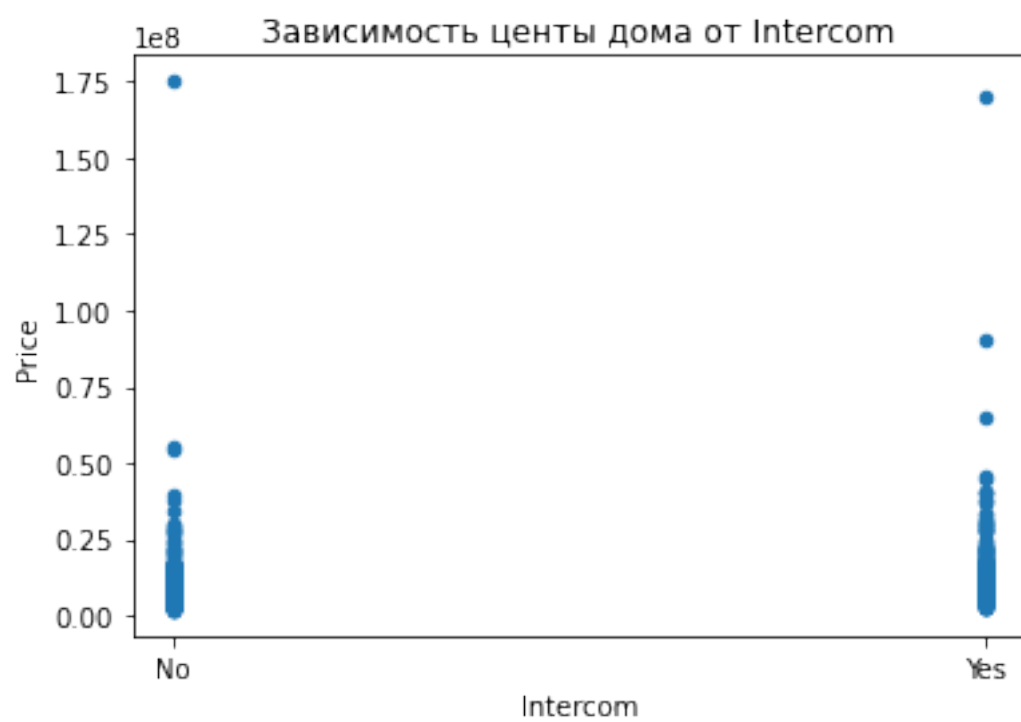


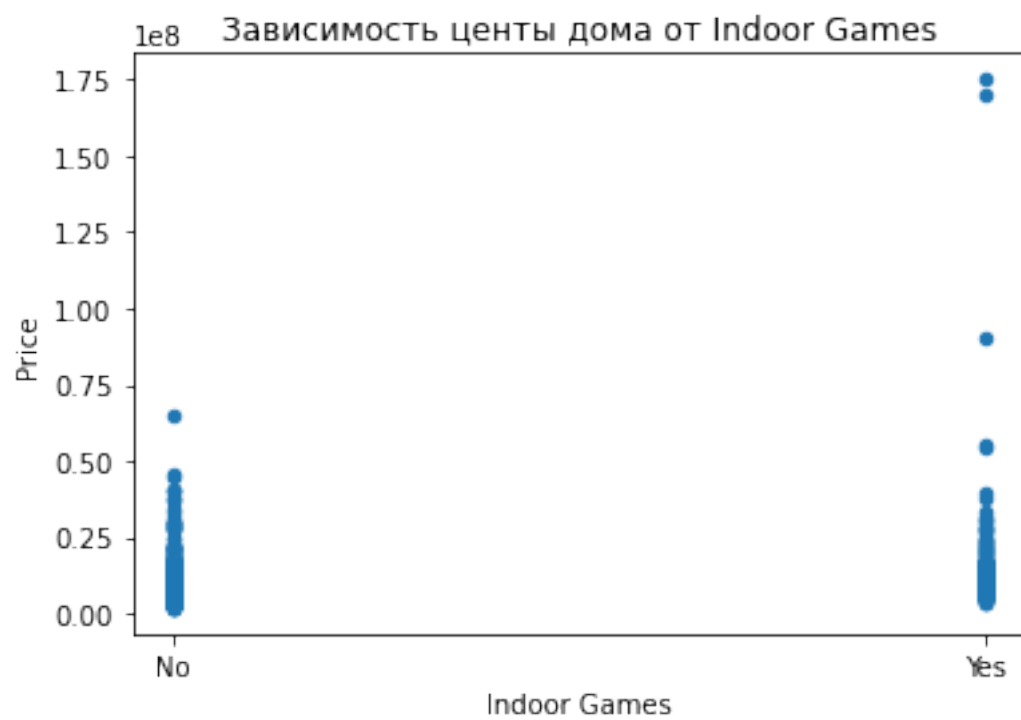
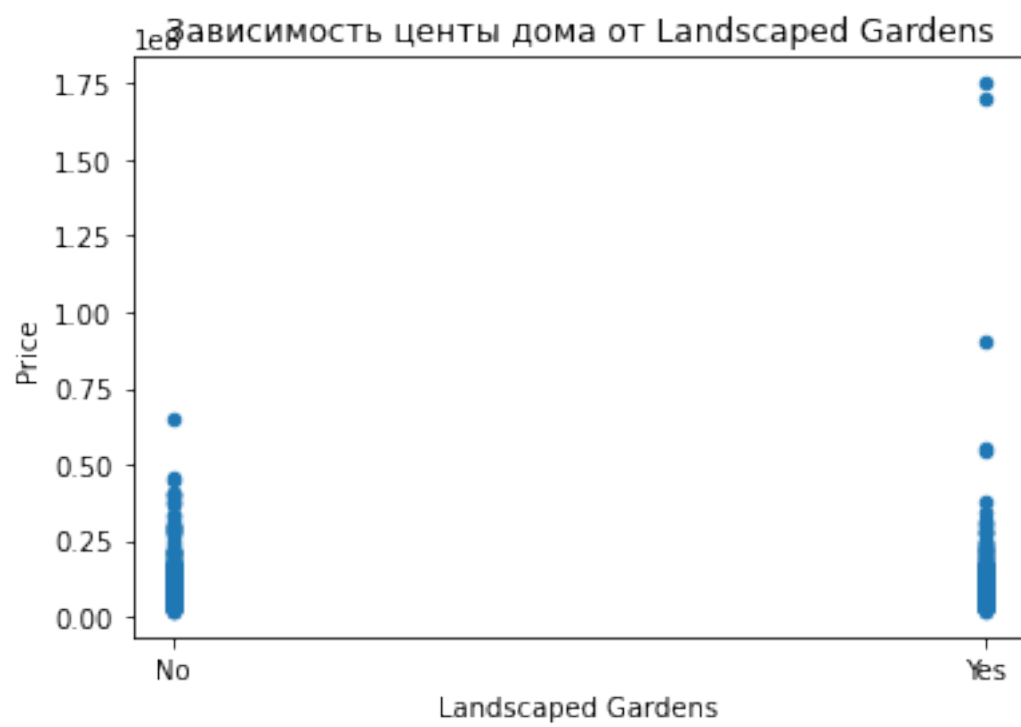
20. Зависимость цены дома от различных характеристик (Вывод: Цена дома зависит от наличия школы рядом (если, есть, то цены выше);
 Зависит от наличия лифта (если есть лифт, то цены выше);
 Зависит от наличия клуба;
 Зависит от наличия бассейна;
 Зависит от наличия газа, дома с газов дешевле
 От остальных характеристик цены зависят не так явно)

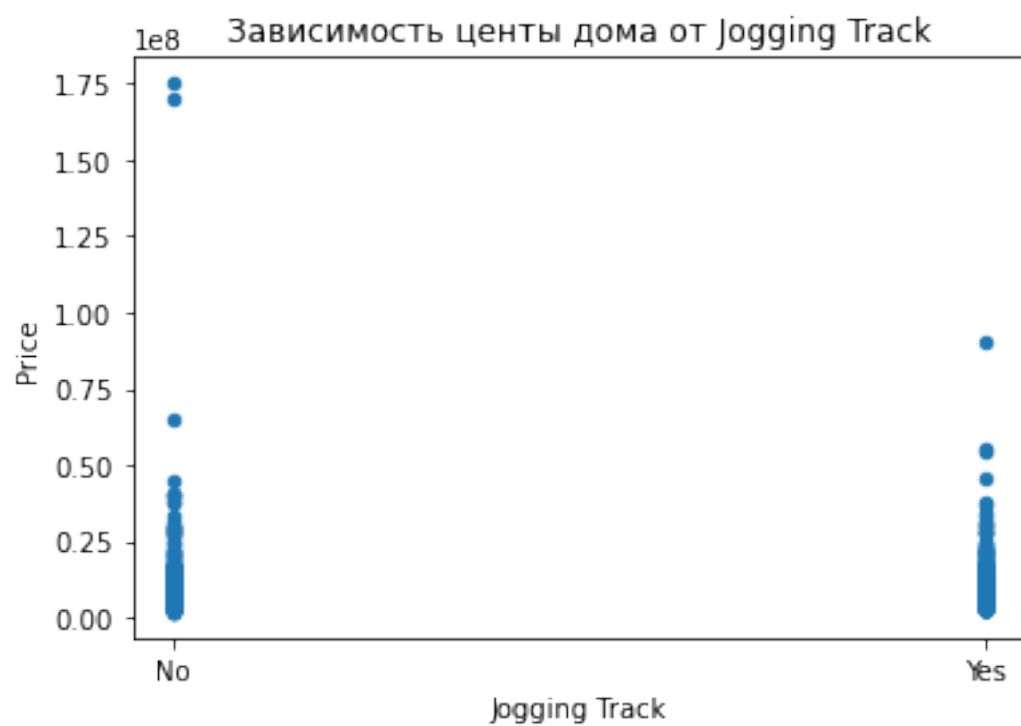
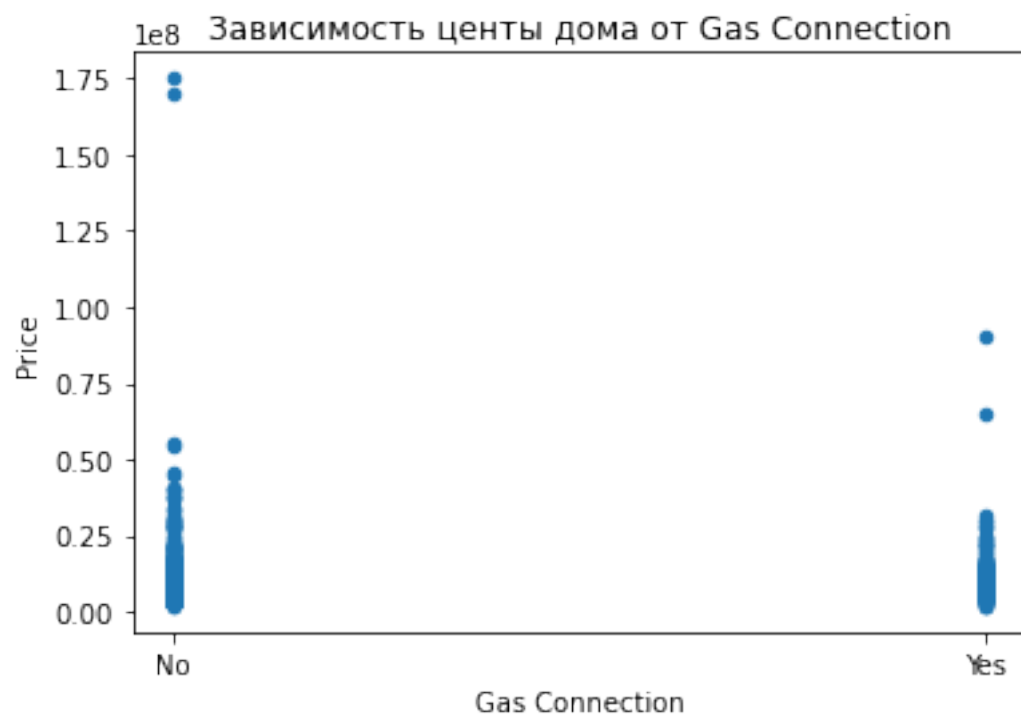


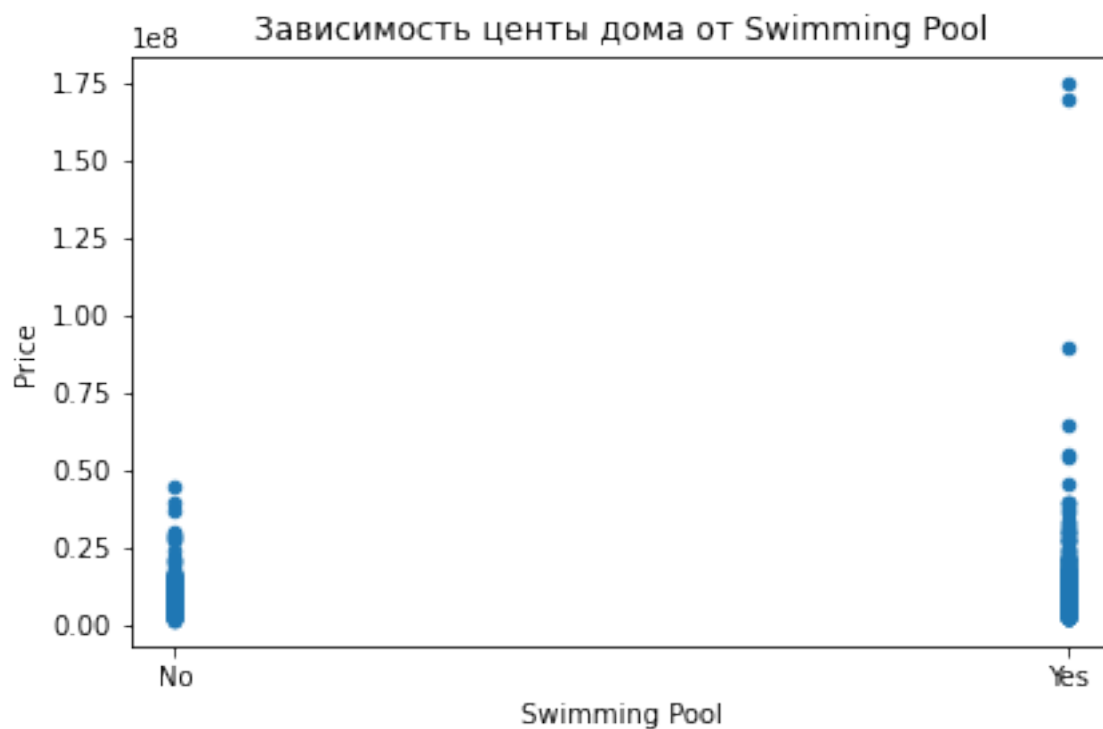






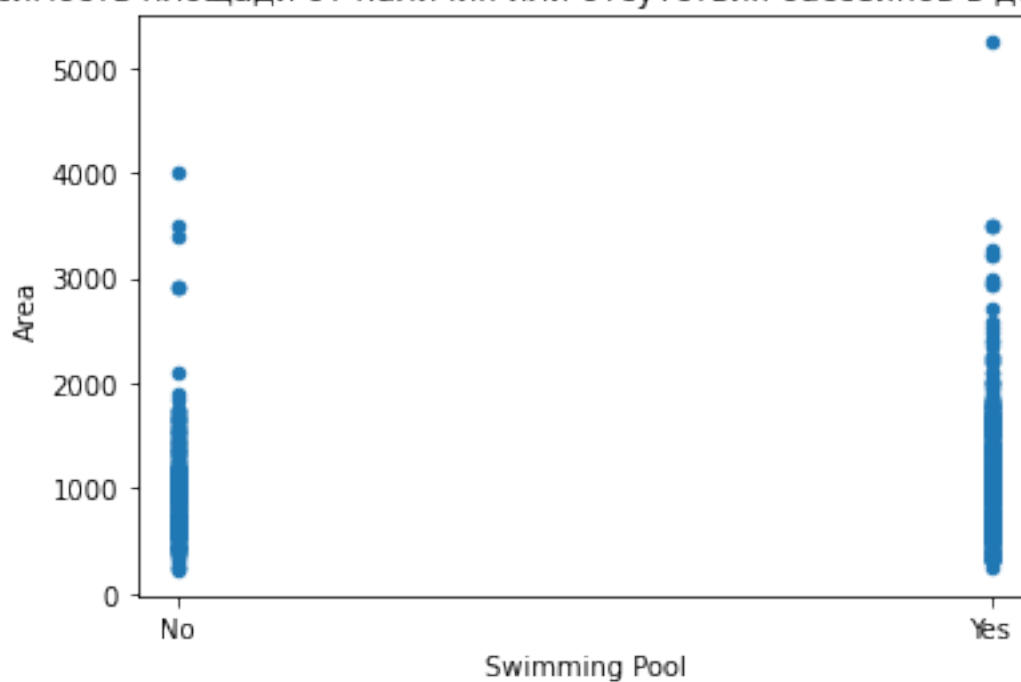






21. Разница в площади домов с бассейном и без (Вывод: дома с бассейном больше по площади)

Зависимость площади от наличия или отсутствия бассейнов в домах выборки



22. Рассматривались дома определенной ценовой категории (Вывод: вероятно, т.к. не проследивается никаких зависимостей, имеющиеся данные отвечают релевантным данным по ценам домов)

