

Университет ИТМО

Практическая работа №4  
по дисциплине «Визуализация и моделирование»

**Автор:** Горбатовский Алексей Валерьевич

**Поток:** ВИМ 1.1

**Группа:** К3220

**Факультет:** ИКТ

**Преподаватель:** Чернышева А.В.

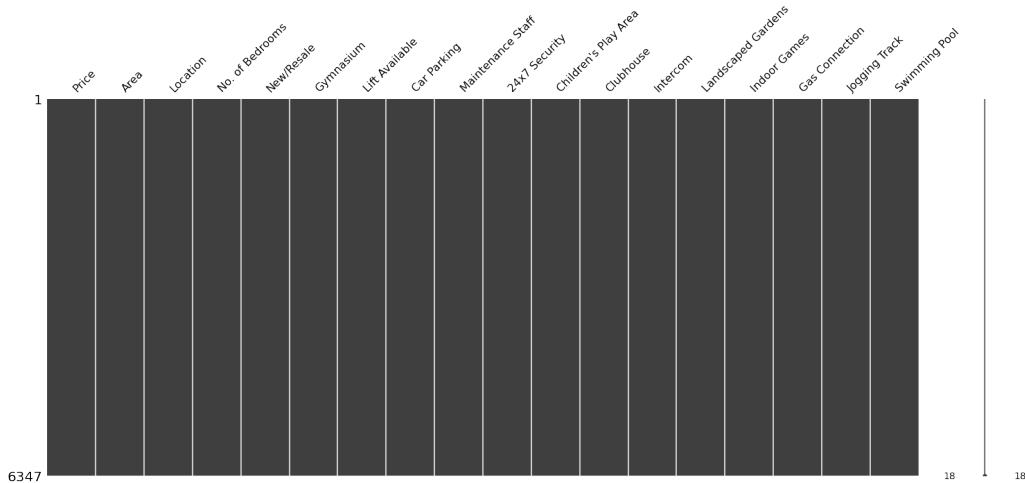
Санкт-Петербург, 2021 г.

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.

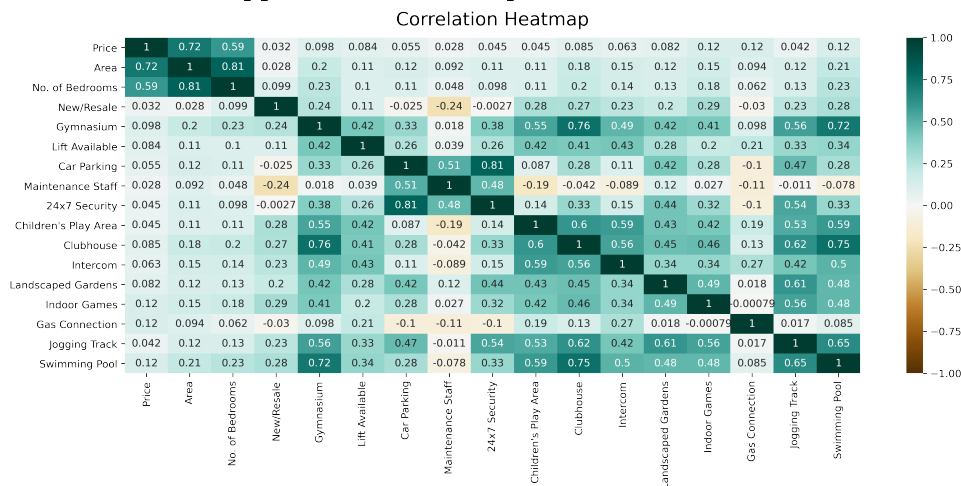
Датасет был проверен на наличие пустых значений, которых не оказалось.



Также были проверены гипотезы:

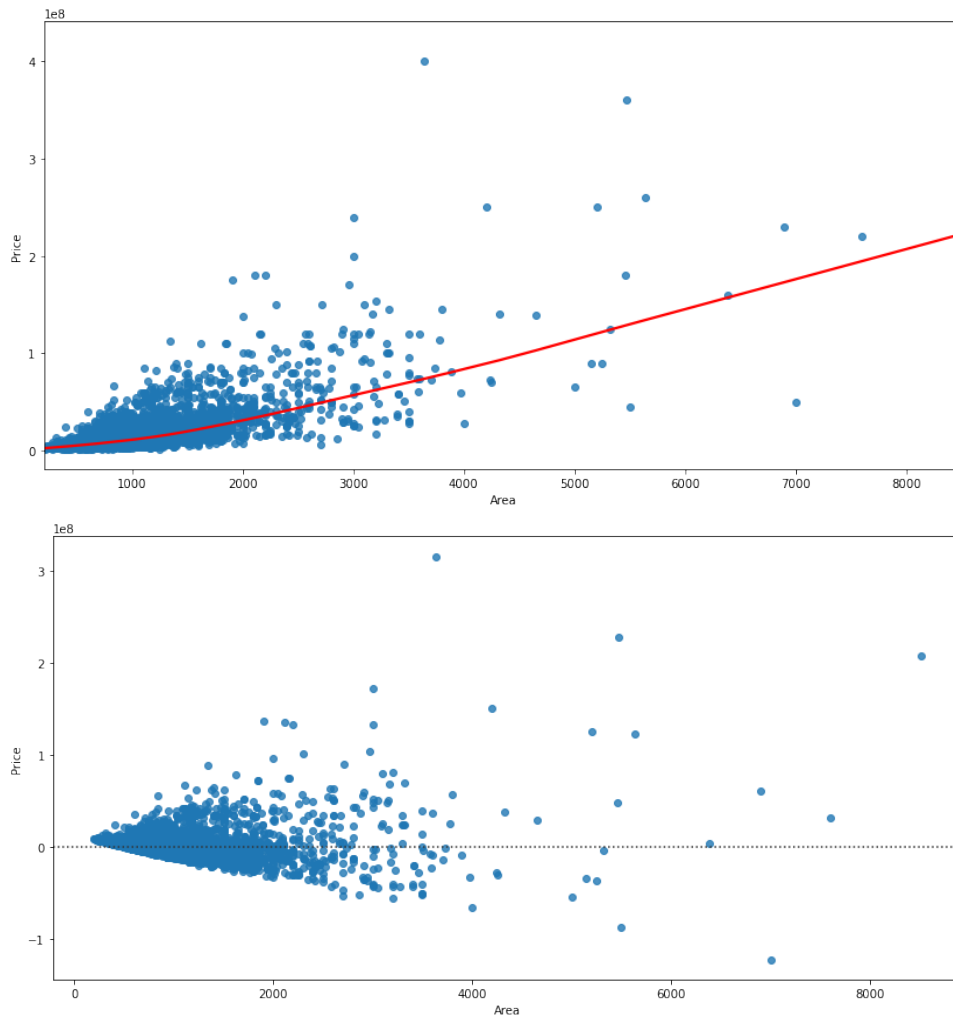
1. Чем лучше уровень инфраструктуры около дома, тем выше цена, причем зависимость цены от площади линейная

Была сделана корреляционная матрица



По корреляционной матрице видно, что цена наибольшим образом зависит от площади и кол-ва комнат

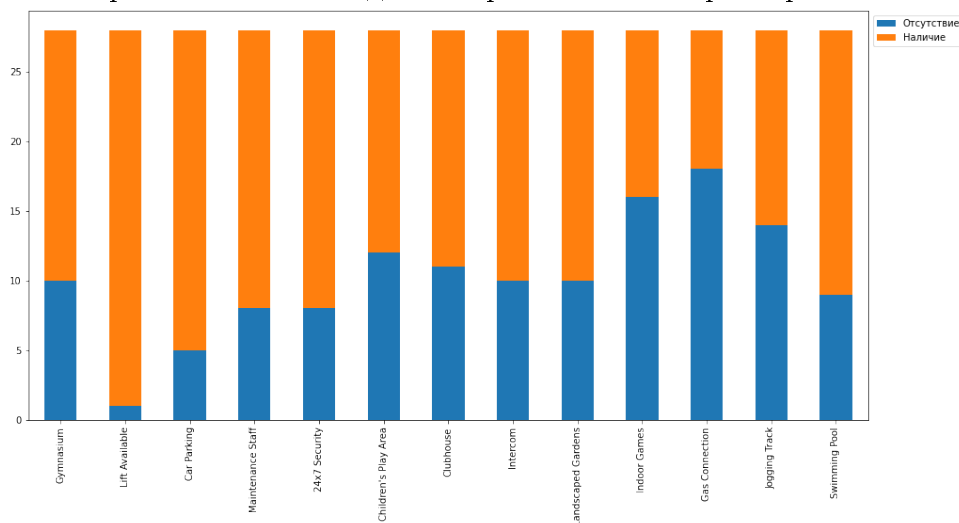
Посмотрим зависимость Price от Area, предположительно она линейная.



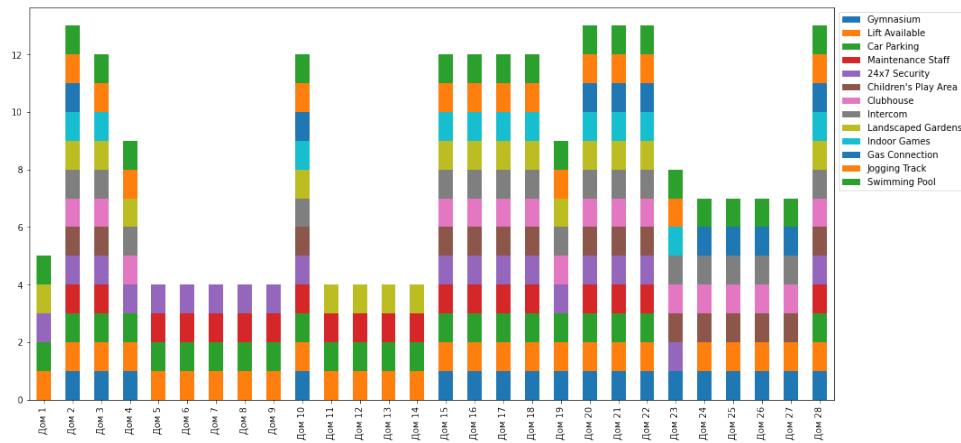
По приведенным графикам можно сделать, что цена имеет линейную зависимость от площади.

2. Район Worli является элитным районом, т.е. большинство домов имеют рядом школы, постоянную охрану и пр.

Посмотрим соотношение домов с различными характеристиками



Видно, что большинство домов с бассейнами, паркингом и др. преимуществами. Также рассмотрим каждый дом по отдельности, т.к. их относительно немного



Делаем вывод, что район Worli можно считать элитным.

### 3. Оптимальный дом находится в локации Kharghar

Оптимальным домом называется дом с наименьшей ценой при наибольшей площади, т.е. отношение Area/Price должно быть максимальным. Также дом должен содержать наибольшее кол-во характеристик, не считая кол-во комнат и газопровод

Посмотрим отношения в двух датасетах: обычном, исследуемом изначально, и с домами, имеющими все характеристики.

```
In [18]: df["Attitude"] = df["Area"]/df["Price"]
df["Attitude"].max()
```

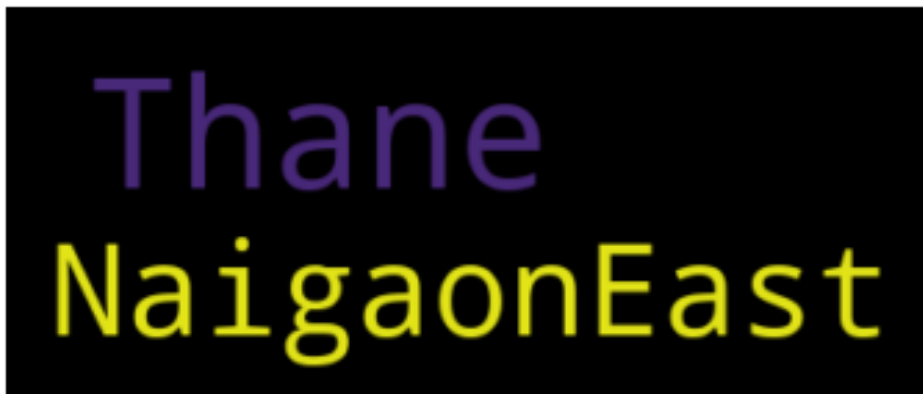
```
Out[18]: 0.000626
```

```
In [19]: best = best_df.copy()
best["Attitude"] = best_df["Area"]/best_df["Price"]
best["Attitude"].max()
```

```
Out[19]: 0.00025714285714285715
```

Видно, что наиболее оптимальный дом только по соотношению Area/Price не совпадает с наиболее оптимальным домом, имеющим все характеристики.

Поэтому просто возьмем и определим две локации, в которых расположены эти дома.



Причем в Thane расположен дом со всеми преимуществами.

### 4. В районе Vusulla наибольшее количество перепродаваемых домов, потому что это

криминальный район.

```
byc_df = df.loc[df["Location"] == "Byculla"]
```

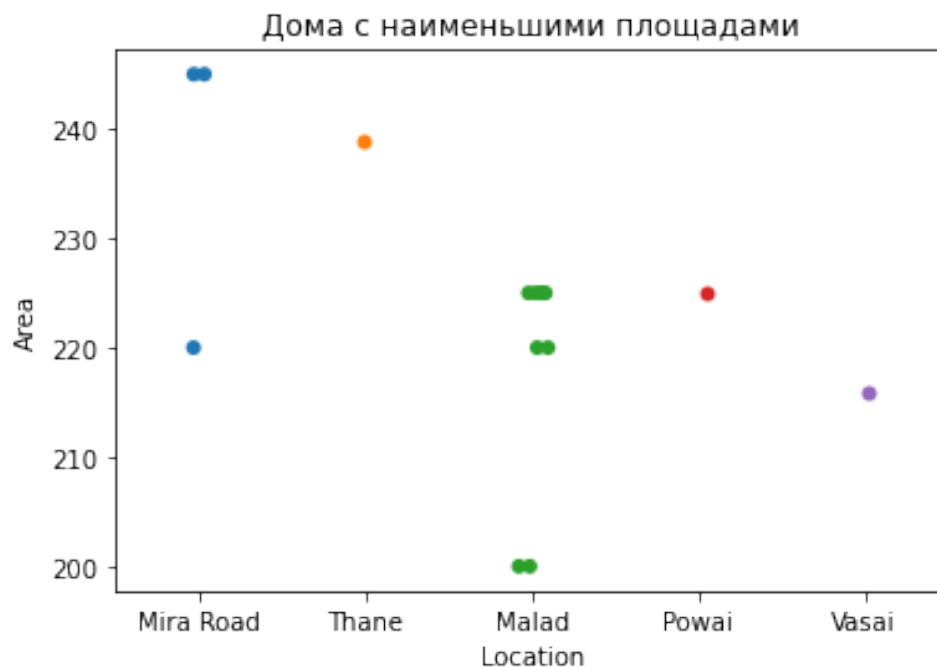
```
len(byc_df)
```

1

На этом датасете эту информацию проверить невозможно, т.к. дом изэтой локации только один, то есть можно сказать, что гипотеза опровержена.

5. Дома с наименьшей территорией расположены в Khaghar

Отсортируем датасет по столбцу Area и посмотрим 15 домов с наименьшей площадью.



Гипотеза опровержена, т.к. дом с наименьшей территорией расположен в Malad.