

# Exploratory data analysis

База данных о стоимости аренды квартир в Москве

В файле 24 столбца, 23368 строк

Уникальных значений по столбцам:

ID объявления	23368
Количество комнат	24
Тип	1
Метро	5866
Адрес	12850
Площадь, м2	11255
Дом	2565
Парковка	5
Цена	3922
Телефоны	13172
Описание	23078
Ремонт	4
Площадь комнат, м2	3990
Балкон	18
Окна	3
Санузел	20
Можно с детьми/животными	3
Дополнительно	528
Название ЖК	1798
Серия дома	393
Высота потолков, м	95
Лифт	34
Мусоропровод	2
Ссылка на объявление	23368
dtype: int64	

Всего пропущенных значений: 118063

Пропущенных значений по столбцам:

ID объявления	0
Количество комнат	1041
Тип	0
Метро	1315
Адрес	0
Площадь, м2	0
Дом	0
Парковка	13417
Цена	0
Телефоны	0
Описание	0
Ремонт	2755
Площадь комнат, м2	8910
Балкон	7978
Окна	6613
Санузел	2672
Можно с детьми/животными	6096
Дополнительно	357
Название ЖК	17520
Серия дома	21205
Высота потолков, м	12162
Лифт	5500
Мусоропровод	10522
Ссылка на объявление	0
dtype: int64	

## Саммари по признакам и их наполнению:

| Признак | Количество Nan | Тип данных | Предполагаемый тип данных |  
 Комментарий | ----- | ----- | ----- | ----- | ----- | ID объявления | 0 | int64  
 | int64 | идентификатор | | Количество комнат | 1041 | object | int64 | | Тип | 0 | object |  
 cat | только 1 вариант | | Метро | 1315 | object | cat | в представленном виде не имеет  
 статистической значимости, нужно сокращать | | Адрес | 0 | object | cat | в  
 представленном виде не имеет статистической значимости, нужно сокращать | |  
 Площадь | 0 | object | int64 | неясный формат | | Дом | 0 | object | object | нужно  
 выделить отдельные признаки | | Парковка | 13417 | object | cat | большое количество  
 пропущенных | | Цена | 0 | object | float64 | неподходящий формат | | Телефоны | 0 |  
 object | int64 | не имеет статистической значимости | | Описание | 0 | object | object |  
 не имеет статистической значимости | | Ремонт | 2755 | object | cat | | | Площадь  
 комнат | 8910 | object | int64 | формат зависит от показателя количество комнат | |  
 Балкон | 7978 | object | cat | большое количество пропущенных | | Окна | 6613 | object |  
 cat | большое количество пропущенных | | Санузел | 2672 | object | cat | | | Можно с  
 детьми/животными | 6096 | object | cat | | | Дополнительно | 357 | object | cat | можно  
 вывести признак касательно мебели | | Название ЖК | 17520 | object | object | не  
 имеет статистической значимости | | Серия дома | 21205 | object | object | не имеет  
 статистической значимости | | Высота потолков | 12162 | int64 | int64 | большое  
 количество пропущенных | | Лифт | 5500 | object | cat | большое количество  
 пропущенных | | Мусоропровод | 10522 | object | cat | большое количество

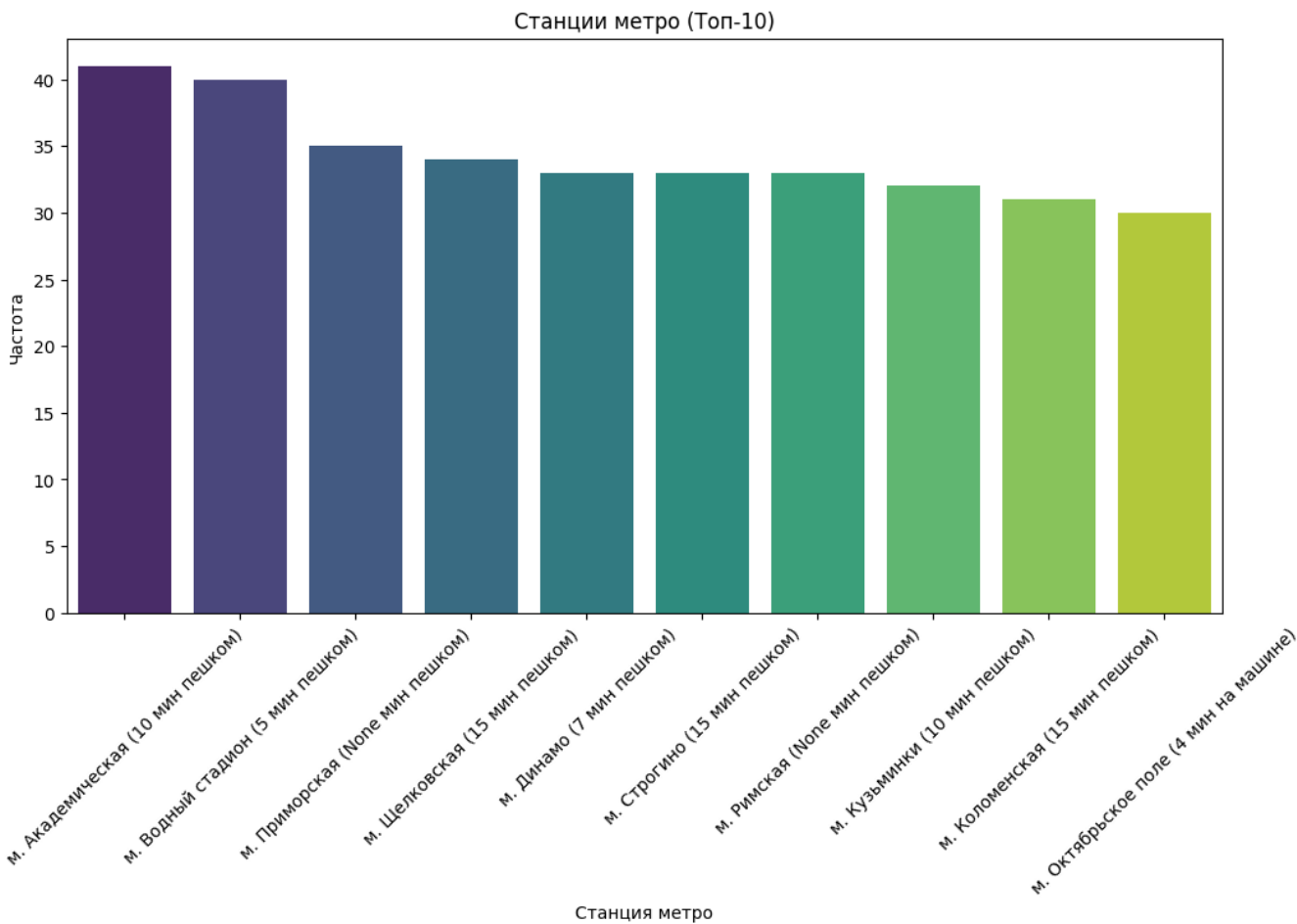
пропущенных | | Ссылка на объявление | 0 | object | object | не имеет статистической значимости |

## Гистограмма стоимости месячной аренды

Приведена для цен в рублях. В датасете присутствуют цены в иностранной валюте, однако таких записей всего 24.

## Самые популярные записи в столбце "Станция метро"

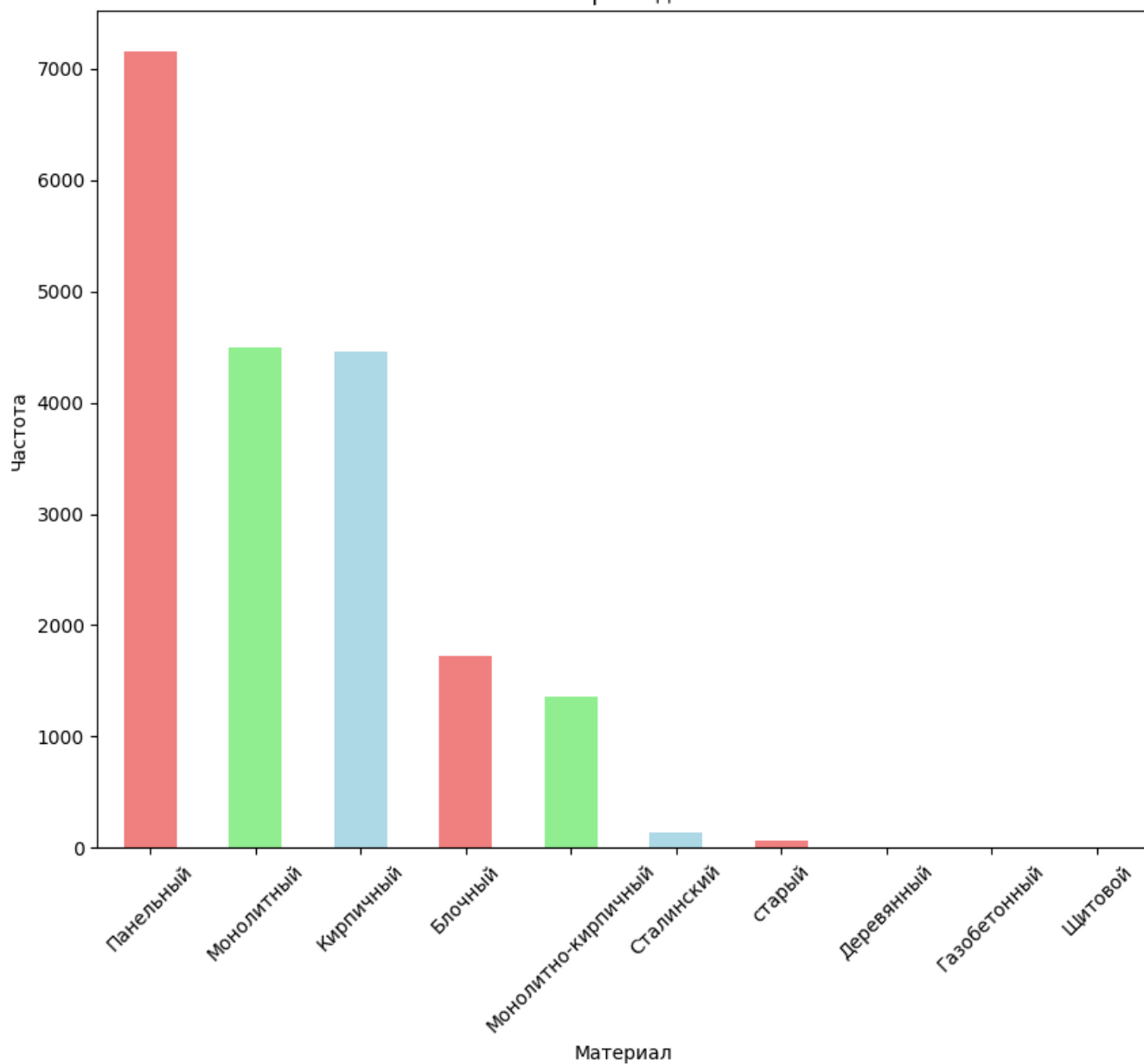
Предлагается разделить признак на 2: расстояние до метро и станция

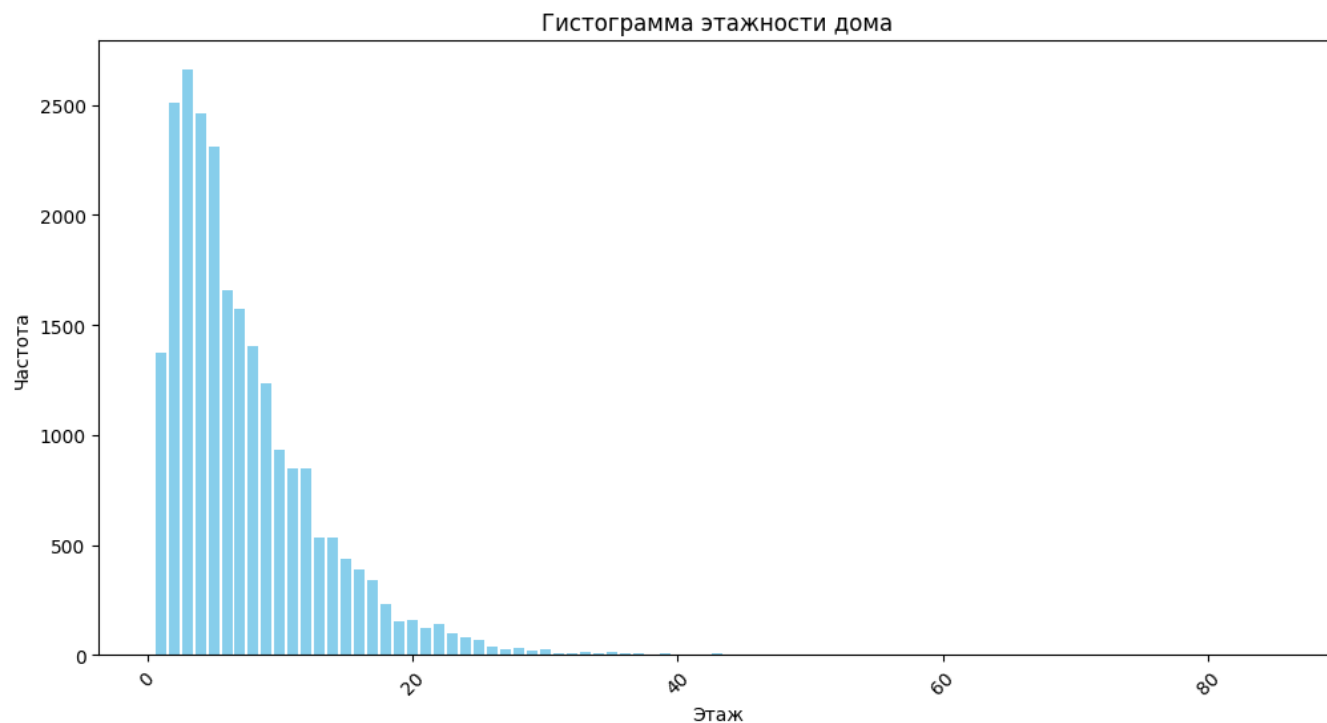


## Информация о домах

Извлечение статистики об этажности и материале здания

Материал дома

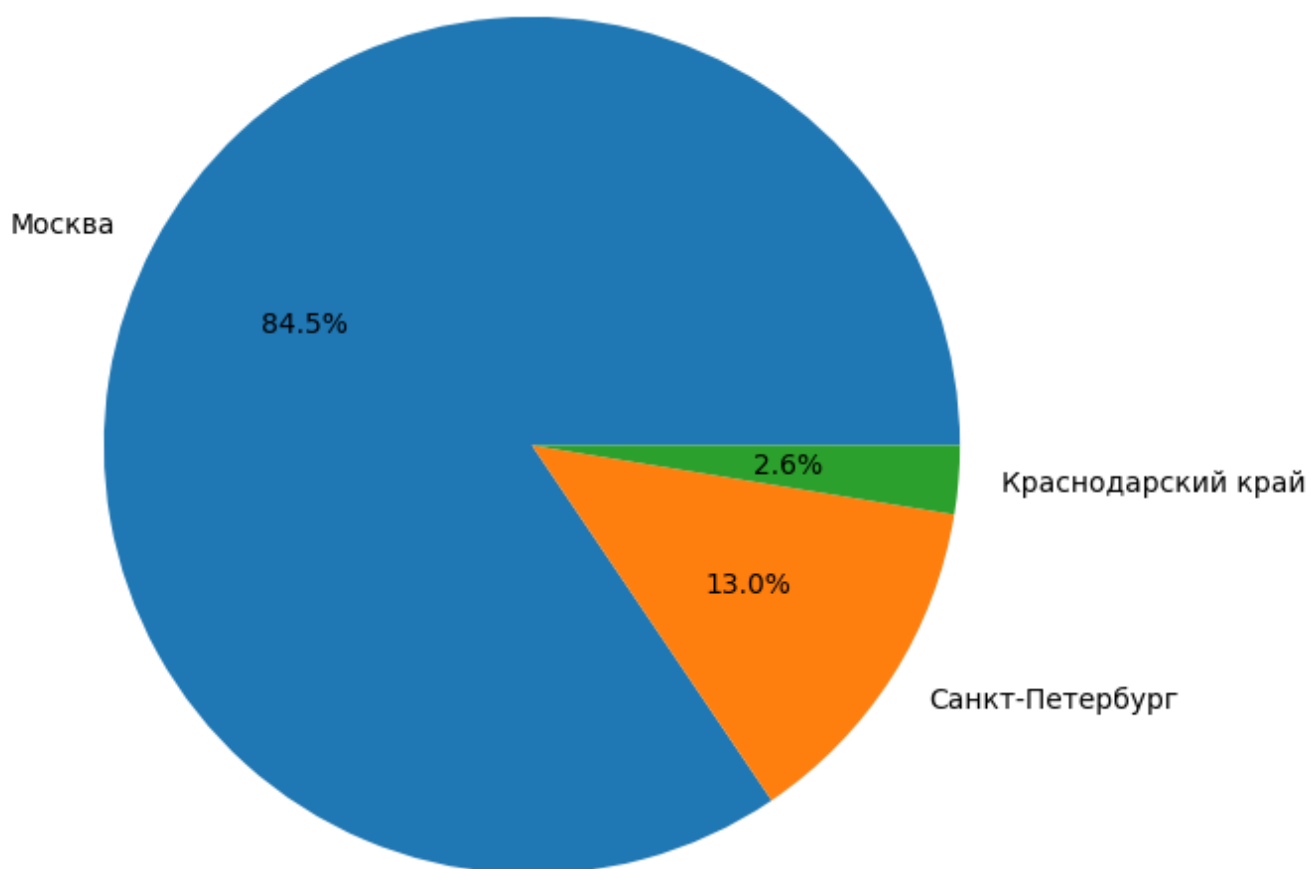




## Проверка адреса на город

Исследование касается квартир в Москве. В датасете были обнаружены другие города:

## Распределение городов



Исследование столбца "Парковка"

Распределение парковок

