EDA

Информация о данных исходного датасета

Датасет **_data.csv** представляет собой информацию по объявлениям об аренде квартир в Москве. Набор данных содержит сведения о типе недвижимости, расположении объекта, его основных характеристиках и условиях аренды.

Сводная информация о наборе данных и типах значений

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 23368 entries, 0 to 23367
Data columns (total 25 columns):
 #   Column                 Non-Null Count  Dtype  
---  ------                 --------------  -----  
 0   Unnamed: 0             23368 non-null  int64  
 1   ID  объявления         23368 non-null  int64  
 2   Количество комнат      22327 non-null  object 
 3   Тип                    23368 non-null  object 
 4   Метро                  22053 non-null  object 
 5   Адрес                  23368 non-null  object 
 6   Площадь, м2            23368 non-null  object 
 7   Дом                    23368 non-null  object 
 8   Парковка               9951 non-null   object 
 9   Цена                   23368 non-null  object 
 10  Телефоны               23368 non-null  object 
 11  Описание               23368 non-null  object 
 12  Ремонт                 20613 non-null  object 
 13  Площадь комнат, м2     14458 non-null  object 
 14  Балкон                 15390 non-null  object 
 15  Окна                   16755 non-null  object 
 16  Санузел                20696 non-null  object 
 17  Можно с детьми/животными  17272 non-null  object 
 18  Дополнительно          23011 non-null  object 
 19  Название ЖК            5848 non-null   object 
 20  Серия дома             2163 non-null   object 
 21  Высота потолков, м     11206 non-null  float64
 22  Лифт                   17868 non-null  object 
 23  Мусоропровод           12846 non-null  object 
 24  Ссылка на объявление   23368 non-null  object 
dtypes: float64(1), int64(2), object(22)
memory usage: 4.5+ MB
```

Размер набора данных

Количество строк в датасете: 23368

Количество столбцов: 25

Столбцы датасета содержат следующую информацию:

Unnamed: 0
ID объявления
Количество комнат
Тип
Метро
Адрес
Площадь, м2
Дом
Парковка
Цена
Телефоны
Описание
Ремонт
Площадь комнат, м2
Балкон
Окна
Санузел
Можно с детьми/животными
Дополнительно
Название ЖК
Серия дома
Высота потолков, м
Лифт
Мусоропровод
Ссылка на объявление

Количество пропущенных значений

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

Принимая во внимание количество пропущенных значений и цели нашего исследования, было решено, что информация в следующих колонках не представляет для нас практического интереса:

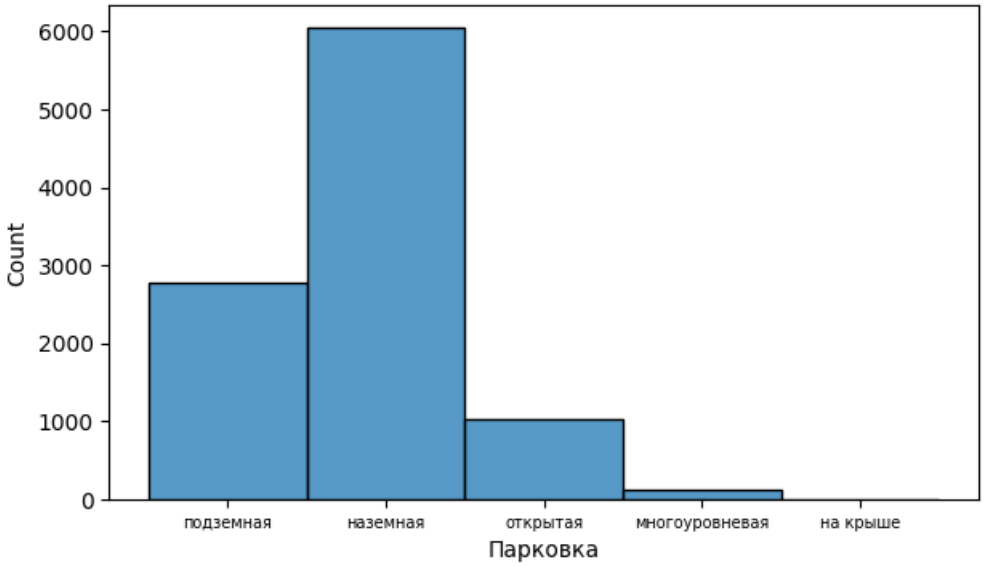
- **Телефоны**
- **Описание**
- **Unnamed: 0**
- **Серия дома**
- **Ссылка на объявление**
- **Название ЖК**
- **Площадь комнат, м2**
- **Тип**

Категориальные метки

Перечислить колонки, графики зависимости цены от каждой категории, количество значений каждой категории

Колонка "Парковка"

	Парковка
NaN	13417
наземная	6043
подземная	2772
открытая	1017
многоуровневая	118
на крыше	1



Поменяем NaN

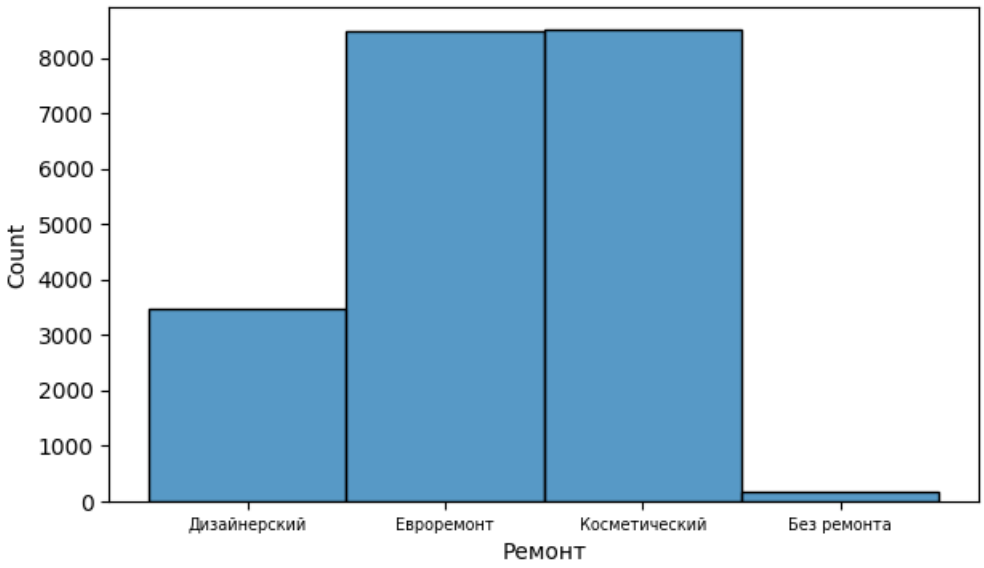
Колонка "Балкон"

	Балкон
NaN	7978
Балкон (1)	7428
Лоджия (1)	6007
Балкон (1), Лоджия (1)	716
Лоджия (2)	568
Балкон (2)	474
Балкон (3)	55
Лоджия (3)	45
Балкон (2), Лоджия (2)	25
Балкон (1), Лоджия (2)	24
Балкон (2), Лоджия (1)	20
Балкон (4)	6
Балкон (1), Лоджия (3)	5
Лоджия (4)	5
Балкон (3), Лоджия (1)	5
Балкон (2), Лоджия (3)	3
Балкон (1), Лоджия (4)	2
Балкон (3), Лоджия (3)	1
Балкон (4), Лоджия (4)	1

Делим на две колонки "Балкон" и "Лоджия", NaN заполняем 0

Колонка "Ремонт"

	Ремонт
Косметический	8499
Евроремонт	8470
Дизайнерский	3474
NaN	2755
Без ремонта	170

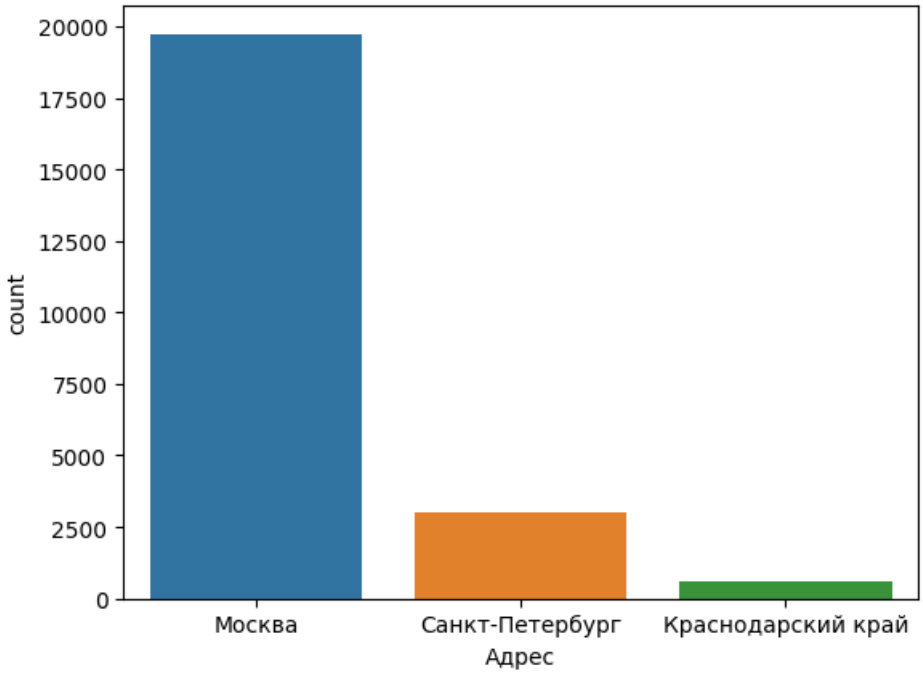


Заменим NaN

Колонка "Адрес"

Поскольку для пилотного проекта выбрана Москва, то необходимо будет отфильтровать значения только для этого города.

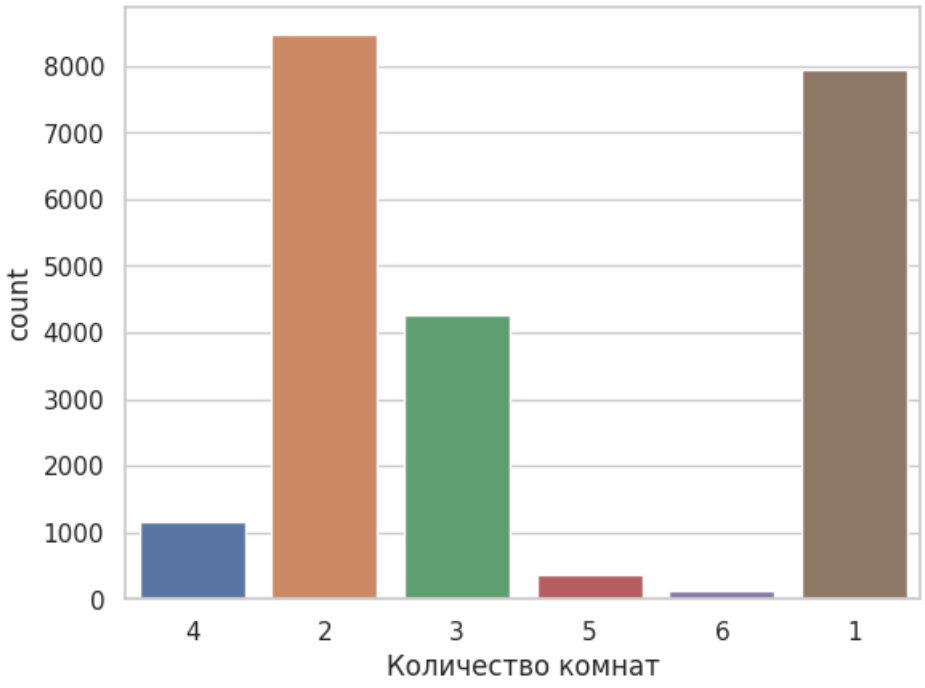
	Адрес
Москва	19737
Санкт-Петербург	3028
Краснодарский край	603



Колонка "Количество комнат"

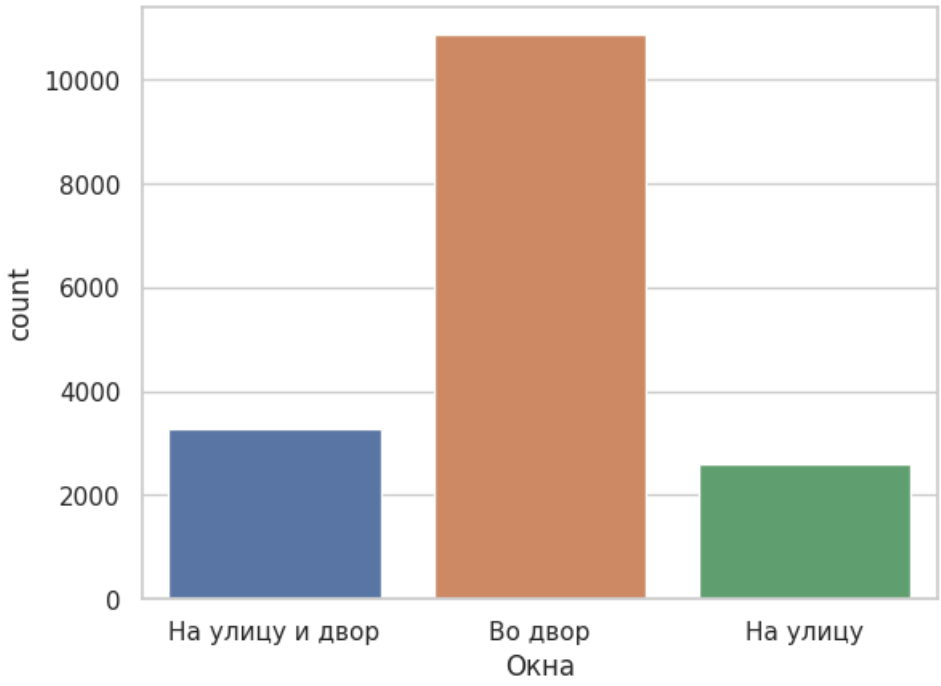
	Количество комнат
1	7917
2, Изолированная	4623
2	2591
3	1717
3, Изолированная	1583
3, Оба варианта	875
4	674
2, Смежная	637
2, Оба варианта	615
4, Оба варианта	253
5	235
4, Изолированная	223
6	87
3, Смежная	87
5, Оба варианта	81
5, Изолированная	47
6, Оба варианта	31
6, Изолированная	17
4, Смежная	13
1, Изолированная	8
1, Оба варианта	4
5, Смежная	4
6, Смежная	3
1, Смежная	2

Данная колонка содержить в себе большое количество значений, которые можно поделить на две колонки с количеством комнат и вариантом размещения. Колонку с количеством комнат, заполнить пропущенные значения и перевести в тип int.¶



Колонка "Окна"

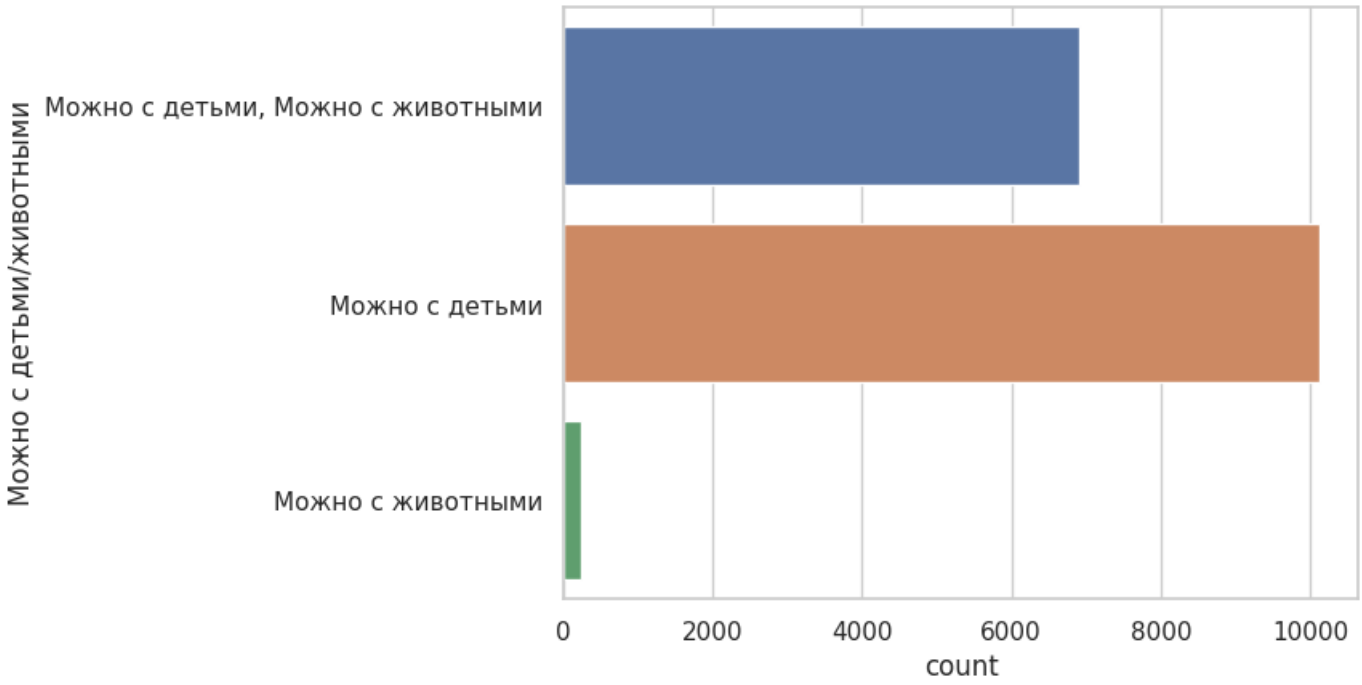
Окна	
Во двор	10870
NaN	6613
На улицу и двор	3295
На улицу	2590



Заменим NaN

Колонка "Можно с детьми/с животными"

Можно с детьми/животными	
Можно с детьми	10134
Можно с детьми, Можно с животными	6899
NaN	6096
Можно с животными	239



Заменим NaN

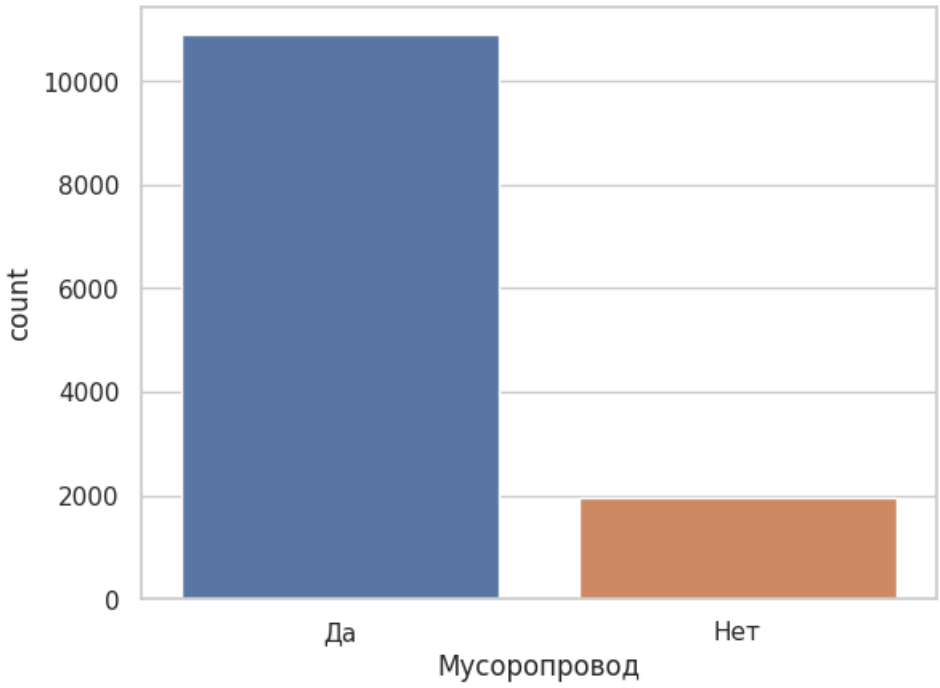
Колонка "Лифт"

	Лифт
Пасс (1)	5911
NaN	5500
Пасс (2)	4326
Пасс (1), Груз (1)	3962
Пасс (2), Груз (1)	1224
Пасс (2), Груз (2)	654
Пасс (3)	636
Пасс (4)	346
Пасс (3), Груз (1)	187
Пасс (1), Груз (2)	185
Груз (1)	95
Пасс (4), Груз (2)	65
Пасс (4), Груз (1)	64
Пасс (3), Груз (3)	45
Пасс (3), Груз (2)	44
Пасс (4), Груз (4)	28
Груз (4)	25
Груз (2)	20
Груз (3)	15
Пасс (4), Груз (3)	7
Пасс (6)	6
Пасс (1), Груз (3)	6
Пасс (2), Груз (3)	3
Пасс (60)	2
Пасс (50)	2
Пасс (1), Груз (12)	1
Пасс (5), Груз (1)	1
Пасс (8), Груз (8)	1
Пасс (7)	1
Пасс (1), Груз (4)	1
Пасс (5), Груз (3)	1
Пасс (2), Груз (4)	1
Груз (6)	1
Груз (8)	1
Пасс (5)	1

Нужно заменить NaN, также поделить на тип лифта, а также посчитать общее количество лифтов

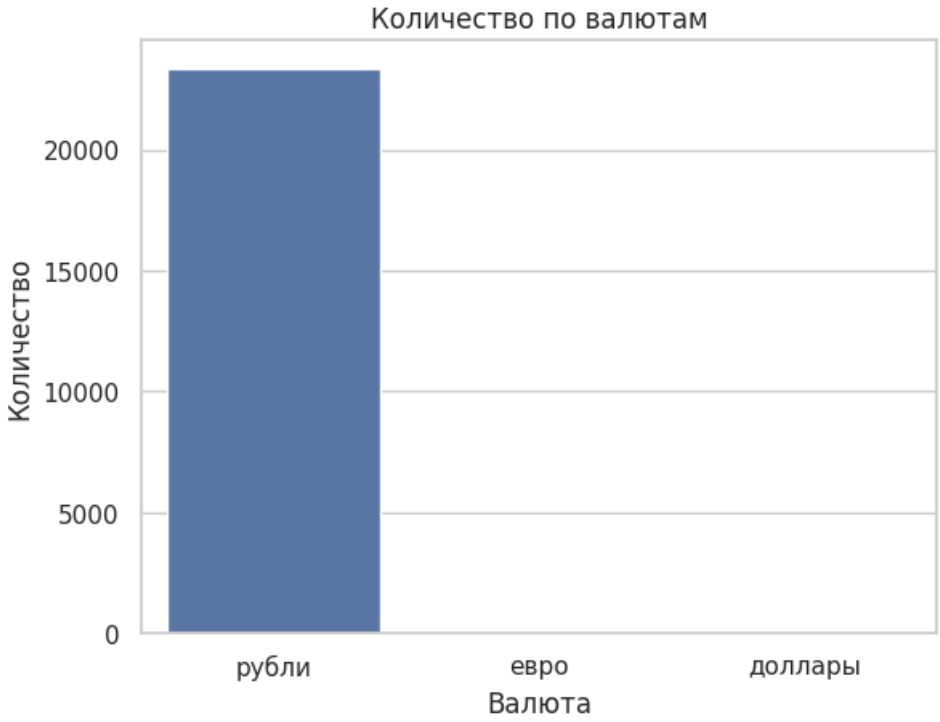
Колонка "Мусоропровод"

	Мусоропровод
Да	10897
NaN	10522
Нет	1949



Колонка имеет два значения (не включая NaN), которые нужно заменить на 0 или 1¶

Цена

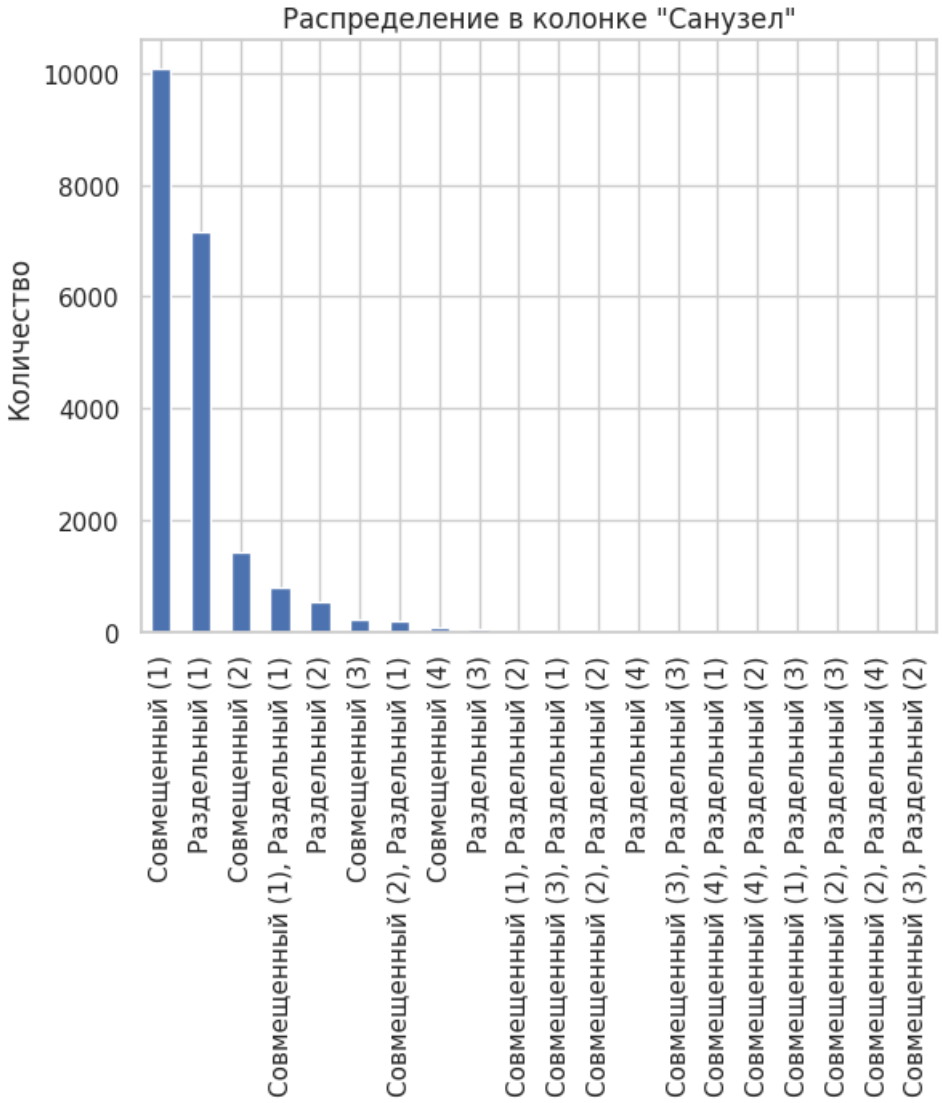


RUB: 23344
EUR: 10
DOL: 14

Из графика видно, что оснвная часть цен предвсвалена в рублях.

- Количество объявлений в рублях 23344
- Количество объявлений в евро 10
- Количество объявлений в долларах 14

Санузел


Распределение в колонке "Санузел"
Количество
10000
8000
6000
4000
2000
0
Совмещенный (1)
Раздельный (1)
Совмещенный (2)
Совмещенный (1), Раздельный (1)
Раздельный (2)
Совмещенный (3)
Совмещенный (2), Раздельный (1)
Совмещенный (4)
Раздельный (3)
Совмещенный (1), Раздельный (2)
Совмещенный (3), Раздельный (1)
Совмещенный (2), Раздельный (2)
Раздельный (4)
Совмещенный (3), Раздельный (3)
Совмещенный (4), Раздельный (1)
Совмещенный (4), Раздельный (2)
Совмещенный (1), Раздельный (3)
Совмещенный (2), Раздельный (3)
Совмещенный (2), Раздельный (4)
Совмещенный (3), Раздельный (2)

	Санузел
Совмещенный (1)	10078
Раздельный (1)	7158
NaN	2672
Совмещенный (2)	1437
Совмещенный (1), Раздельный (1)	812
Раздельный (2)	534
Совмещенный (3)	241
Совмещенный (2), Раздельный (1)	188
Совмещенный (4)	77
Раздельный (3)	52
Совмещенный (1), Раздельный (2)	30
Совмещенный (3), Раздельный (1)	27
Совмещенный (2), Раздельный (2)	25
Раздельный (4)	15
Совмещенный (3), Раздельный (3)	6
Совмещенный (4), Раздельный (1)	6
Совмещенный (4), Раздельный (2)	4
Совмещенный (1), Раздельный (3)	2
Совмещенный (2), Раздельный (3)	2
Совмещенный (2), Раздельный (4)	1
Совмещенный (3), Раздельный (2)	1

Нужно заменить NaN, также поделить на тип санузла, а также посчитать общее количество санузлов

Метро

	Метро
NaN	1315
м. Академическая (10 мин пешком)	41
м. Водный стадион (5 мин пешком)	40
м. Приморская (None мин пешком)	35
м. Щелковская (15 мин пешком)	34
...	...
м. Печатники (38 мин пешком)	1
м. Печатники (35 мин пешком)	1
м. Авиамоторная (1 мин пешком)	1
м. Фонвизинская (16 мин пешком)	1
м. Солнцево (5 мин на машине)	1

5867 rows × 1 columns

Делим это на две колнки название метро и время пешком до метро

Высота потолков

	Высота потолков, м
NaN	12162
2.64	4467
3.00	1322
2.70	1040
2.48	676
...	...
265.00	1
4.15	1
9.00	1
260.00	1
3.02	1

96 rows × 1 columns

Нужно заменить NaN, также перевести в едину СИ.

Дом

	Дом
3/5, Кирпичный	322
4/5, Кирпичный	296
2/5, Кирпичный	255
1/5, Кирпичный	232
5/5, Кирпичный	231
...	...
19/35, Монолитный	1
8/35, Монолитно-кирпичный	1
12/49, Монолитный	1
13/34, Монолитный	1
12/25, Блочный	1

2565 rows × 1 columns

Нужно заменить NaN, также поделить на тип дома, этажность дома и этаж квартиры