

## Исследуем столбцы

### Columns withOUT nan:

['Unnamed: 0', 'ID объявления', 'Тип', 'Адрес', 'Площадь, м2', 'Дом', 'Цена', 'Телефоны', 'Описание', 'Ссылка на объявление']

### Columns with nan:

['Количество комнат', 'Метро', 'Парковка', 'Ремонт', 'Площадь комнат, м2', 'Балкон', 'Окна', 'Санузел', 'Можно с детьми/животными', 'Дополнительно', 'Название ЖК', 'Серия дома', 'Высота потолков, м', 'Лифт', 'Мусоропровод']

Column name Red - has Nan Green- no Nan	Format / possible values	What to do with nan  What formatting can be done	NEW COLUMN	Owner [Osana, Vanya, Masha, Anya]
'Unnamed: 0',		установить индексом (set_index)		
'ID объявления',		все ок, все int	id	
'Количество комнат',	['4' '4, Оба варианта' '2' '3' '5, Оба варианта' '3, Изолированная' '5' '5, Изолированная' '2, Оба варианта' '3, Оба варианта' '6' '1' '4, Изолированная' '6, Оба варианта' '2, Изолированная' nan '2,	<b>Сделал столбец “Кол-во комнат” (пока float), и “Тип комнаты”.</b> Из Площади комнат значения в Nan-ы не перенес  NaN заменить инфой из Площади комнат, м2 (разделители разные) Если NaN в двух столбцах, то среднее значение по площади квартиры.  Создать доп. столбец с значениями только кол-во комнат int Создать столбец с типом	rooms_count, rooms_type  Кол-во комнат, тип комнат	Ваня

	Смежная' '4, Смежная' '3, Смежная' '6, Изолированная' '1, Изолированная' '1, Оба варианта' '5, Смежная' '6, Смежная' '1, Смежная']	rooms из этой колонки, из площади комнат room_type		
'Тип',	['Квартира']	Все значения одинаковые => колонку удаляем		Осана
'Метро',	Пример формата:: м. Смоленская (9 мин пешком	<p><b>Разбил на колонки “Станция Метро” и “Доступность”.</b></p> <p><b>Разбивку пешком/на машине не сделал. Добавил столбец “Округ” в соответствии со станцией метро (но есть пустые значения, т.к. базу станций нашел лишь 5-летней давности).</b></p> <p><b>Возможно найду более новую</b></p> <p>Если мы поделим по районам, то NaN по адресу можно будет проставить.</p> <p>Столбец - пешком или на машине</p> <p>Разделить и создать дополнительный столбец (расстояние до метро).</p> <p><b>Может быть создать доп. столбец по зонам/районам</b></p> <p>В центре метро расположено не так далеко друг от друга. Например, внутри кольцевой не учитывать расстояние до метро, скомпоновать метро по районам, а за кольцом учитывать тоже район, но если до метро больше 10 мин, то помечать это.</p>	<p>metro, district, time_to_metro, by_feet_car</p> <p>Станция метро, Район, Время до метро, На чем</p>	Ваня
'Адрес',	['Москва' , 'Санкт-Петербург', 'Краснодарский край']	Удалить строки с Краснодарским краем и Санкт Петербургом	<p>address</p> <p>Написать для всех команду, чтобы удалить строки не с мск</p>	Осана
'Площадь, м2',	Может быть 1, 2, 3 элемента	Для дальнейшего анализа берем только общую площадь квартиры (первое число в данных колонки), так как неизвестно	square_total	Аnya

	198.0/95.0/18.0 52.5/10.0 43.1	что обозначают остальные числа. Преобразовываем тип данных столбца во float.		
'Дом',	<b>Пример:</b> 5/16, Монолитный  <b>Возможные значения:</b> ['Монолитный' 'Монолитно-кирпичный' пан 'Панельный' 'Сталинский' 'Кирпичный' 'старый фонд' 'Блочный' 'Деревянный' 'Щитовой']	Выделить: этаж, этаж всего дома, тип дома  Можно использовать для заполнения NaN в столбцах: Лифт, Мусоропровод, Площадь комнат, м2 (Когда нет Названия ЖК)	floor_flat, floors_total, building_type	Осана
'Парковка',	['подземная', пан, 'наземная', 'открытая', 'многоуровневая', 'на крыше']	Заменяем значения NaN на 'Нет парковки', так как в приведенных данных есть информация о наличии парковки, значит вариант с NaN является ее отсутствием.	parking	Аnya
'Цена',		Из данного столбца выделяем дополнительные колонки с информацией: Плата за месяц, Объем залога, Включены ли коммунальные услуги, Срок аренды, Предоплата (кол-во месяцев)  Пример: '500000.0 руб./ За месяц, Залог - 500000 руб., Коммунальные услуги включены, Срок аренды - Длительный, Предоплата 1 мес'	rent_price, deposit, utility, lease, prepayment	Аnya
'Телефоны',		Можно убрать, как будто не нужны		Осана
'Описание',		Можно попробовать выделить категории по следующим значениям: "Рассматриваются квартиранты только славяне и только на длительный срок" и "Без животных" (как пример, можно поискать что-то еще)	description	

		Мне кажется можно убрать колонку, насчет животных и детей есть столбец 'Можно с детьми/животными'		
'Ремонт',	['Дизайнерский', 'Евроремонт', 'Косметический', nan, 'Без ремонта']	Данные NaN заполняем значениями "Без ремонта", так как пропущено всего 2 755 значений, а это всего 11%.	renovation	Anya
'Площадь комнат, м2',		Восстановить количество комнат по этому столбцу	area_rooms	Ваня
'Балкон',	nan, 'Балкон (N)' 'Лоджия (M)' N: 1, 2, 3, 4 M: 1, 2, 3, 4	Если Nan, то 0 балконов, 0 лоджий	balcony, loggia  Кол-во балконов Кол-во лоджий	Маша (готово)
'Окна',	[nan, 'На улицу и двор', 'Во двор', 'На улицу']	Так как невозможно определить на какую сторону выходят окна у квартиры, то значения NaN мы заполняем как 'Нет данных'.	windows	Anya
'Санузел',	Nan, 'Совмещенный (N)', 'Раздельный (M)', N, M 1 – 4	Nan - по типу дома заполнить 1 Сталинский - 1 раздельный	bathroom	Маша (готово) Пока тут есть 0 - значения
'Можно с детьми/животными',	['Можно с детьми, Можно с животными', 'Можно с детьми', 'Можно с животными', nan]	Создаем столбцы pets_allowed и kids_allowed и проставляем значени 1 или 0 в зависимости от условий проживания. NaN значения будут равняться 0 в двух столбцах.	pets_allowed, kids_allowed  Животные, Дети	Anya
'Дополнительно',	['Мебель в комнатах', 'Мебель на кухне', 'Ванна', 'Душевая кабина',	Мебелерованная или нет, должно влиять  Можно создать доп столбцы Есть Мебель (1/0), Посудомойка (0/1) и тд.	furniture, dishwasher, fridge, etc	Осана

	'Стиральная машина', 'Кондиционер', 'Посудомоечная машина', 'Телевизор', 'Холодильник', 'Интернет' 'Телефон']			
'Название ЖК',	1134 вариантов	По названию ЖК много NaN можно заполнить (этажи, район, лифты и тд). Данных очень мало.	building_name	
'Серия дома',		Не учитывать, очень мало данных, да и корреляция маленькая скорее всего Можно использовать для заполнения NaN (Когда нет Названия ЖК) (Серий дома меньше, чем ЖК)  Пока игнорируем	building_series	
'Высота потолков, м',	От 2. до 6., Но есть ещё 12.8 , 25. , 26. , 27. , 28. , 260. , 264. , 265. , 280. , 310. , 320. , nan	<b>Этот столбец особо не трогал, лишь построил гистограмму в отчет</b> 260. , 264. , 265. , 280. , 310. , 320. Это явно см  А вот 12.8 , 25. , 26. , 27. , 28. ?  Что делать с 1,2?  Мне кажется средняя высота от 2,5 до 6 метров 2,5 минимум, чтобы жить, а 6 это студии/двойной свет, второй этаж не в счет	Ceiling_height	Ваня
'Лифт',	Пасс (N), Груз( M)	Надо проверить корреляцию со стоимостью квартиры, если она маленькая, то можно преобразовать столбик на Есть/Нет  Обновленными государственными строительными нормами (ДБН) по проектированию жилых домов введены новые требования: <ul style="list-style-type: none"><li>• рекомендуется устраивать лифты в новых жилых домах от 3-х этажей;</li><li>• в новом жилье высотой 4 этажа и более лифты</li></ul>	service_lift, passenger_lift  Кол-во груз, кол-во пасс	Маша (готово)

		<p>предусматриваются обязательно.</p> <p>По старым нормам лифты в жилых домах должны проектироваться лишь от 5 этажей.</p>		
'Мусоропрово д',	['Да', 'Нет', nan]	<p>NaN =&gt; нет</p> <p>Насколько мусоропровод влияет на стоимость? Половины данных нет, может стоит не учитывать столбик.</p>	garbage_chute	Осана
Ссылка на объявление'		Можно убрать, инфо не нужно	link	Осана