



Федеральное государственное автономное образовательное учреждение  
высшего образования

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ  
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

Факультет экономических наук, образовательная программа «Экономика»

**Домашняя работа по курсу «Эконометрика – 1».**

**Анализ факторов, влияющих на стоимость недвижимости в Бельгии на  
ноябрь 2023 года**

Выполнили:

Рыбин Сергей Васильевич, БЭК217

Золотухина Евгения Олеговна, БЭК217

Преподаватель:

Станкевич Иван Павлович

Москва

2023

# Введение

Современный рынок недвижимости олицетворяет собой важный сегмент экономики, в котором взаимодействуют множество факторов, оказывающих влияние на ценообразование жилья. Так, в статье *“Analysis of factors influencing the price of real estate based on interpretative structural model”* (2018)[1] исследуются факторы ценообразования недвижимости Китая, путем построения интерпретативной структурной модели (ISM). Также проводились эконометрические исследования влияния года постройки, местоположения объекта и других факторов на стоимость на турецком рынке недвижимости [4]. Бельгия в свою очередь характеризуется разнообразием культурных и социальных аспектов, поэтому исследование цен на недвижимость в данной стране вызывает особый интерес.

Цель настоящего эконометрического исследования заключается в построении модели линейной регрессии для анализа факторов, влияющих на цены на недвижимость в Бельгии. Основной задачей исследования является выявление и количественная оценка влияния различных переменных на ценообразование жилья в Бельгии.

Актуальность исследования подтверждается существующим разнообразием исследований рынка недвижимости разных стран и обуславливается тем, что полученные результаты не только предоставят более глубокое понимание механизмов формирования цен на жилье, но и могут служить основой для разработки эффективных стратегий управления недвижимостью и принятия решений в контексте динамичного рынка жилья в Бельгии

## Экономическая модель

Источником данных служит датасет [Belgian Property Prices 2023](#), взятый с Kaggle. [5]

Этот набор данных содержит информацию с рекламных сайтов недвижимости в Бельгии за ноябрь 2023 года, включая различные характеристики недвижимости и соответствующие им цены. Он предлагает всесторонний обзор бельгийского рынка жилья, позволяя пользователям изучать и анализировать характеристики недвижимости и тенденции ценообразования. Данные охватывают широкий спектр характеристик недвижимости, от количества спален и ванных комнат до оценок энергоэффективности.

Изначально датасет содержал 4770 записей, после очистки null значений получаем 541 строку, с которыми и будем работать дальше.

## Определение переменных

Обоснование выбора переменных для исследования цен на недвижимость в Бельгии основывается на нескольких ключевых факторах, которые подтверждаются научной литературой в области эконометрики и исследований рынка недвижимости:

1. **Год постройки** (construction year, в годах) – эта переменная может служить индикатором общего состояния дома и его технической современности. Новые дома могут обладать современными технологиями, лучшей энергоэффективностью и, возможно, находиться в лучших районах с более развитой инфраструктурой. Мы предполагаем, что, чем новее дом, тем выше будет его стоимость в силу лучшего состояния, использования более технологичных материалов и наличия более развитой инфраструктуры. [2, 3]

2. **Площадь парковочного места** (outdoor\_parking\_spaces, метры) - увеличение площади парковочного места может рассматриваться как дополнительное удобство для владельцев автомобилей, особенно если у них есть несколько автомобилей. Это может повысить комфортность проживания и, следовательно, влиять на цену недвижимости. Мы предполагаем, что, чем больше площадь парковочного места, тем выше стоимость дома. Так, если парковка вмещает 2 и более мест, то у потенциальных владельцев может быть 2+ автомобиля. [4]

3. **Площадь участка** (garden surface, м<sup>2</sup>) – большая площадь участка может быть привлекательной для покупателей, которые ценят простор и могут использовать землю для садоводства или отдыха. Логично предполагать, что чем больше земли во владении, тем выше стоимость дома. [4]

4. **Столичный/нестоличный регион** (1 – столичный, 0 – нестоличный) – расположение в столичном регионе может оказывать влияние на цену из-за более высокого спроса, престижа, развитой инфраструктуры и доступности рабочих мест. Эта переменная учитывает географическую особенность и экономическую активность региона. Мы предполагаем, что в столичном регионе стоимость домов должна быть выше, так как выше экономическая активность и престиж жизни в столице или рядом, больше возможностей для трудоустройства, более развитая инфраструктура. [2, 3, 4]

5. **Газовое/негазовое отопление дома** (1 – газовое, 0 – негазовое) – Газовое отопление может рассматриваться как более современный и эффективный способ обогрева, что может повысить ценность дома. Экологические и экономические преимущества газового отопления могут влиять на решение о покупке дома. Газовое отопление обычно энергоэффективнее, надежнее и экологичнее чем другие способы связанные со сжиганием топлива, а также

доступнее чем электрообогреватели, поэтому дома с газовым отоплением вероятнее будут стоить дороже. [4]

6. **Объем выбросов CO<sub>2</sub>** (CO<sub>2</sub> emission, kg CO<sub>2</sub>/m<sup>2</sup>/year) – объем выбросов CO<sub>2</sub> жилого дома напрямую связан с энергоэффективностью систем отопления и кондиционирования внутреннего пространства, уровнем теплоизоляции, а также использованием бытовых приборов. Логично предположить, что более экологичные и экономичные дома будут стоить дороже в силу использования более технологичных материалов и потенциальной экономии на стоимости энергии в будущем. [1, 2, 4]

## Гипотезы

В нашей работе мы опирались на экономические исследования, в которых изучались цены на недвижимость в других странах и факторы, влияющие на неё. В статье Models of factors influencing the real estate price (2011) [2] авторы пришли к выводу о том, что на цену в основном влияет местоположение и его престиж; люди хотят, чтобы окружающая среда соответствовала их социальному статусу и подчеркивала его; в зависимости от престижности района цена может отличаться примерно на 25%. К аналогичному выводу приходят авторы другого исследования - Canals vs. Streams: To What Extent Do Water Quality and Proximity Affect Real Estate Values? [3]. Исходя из этого, мы хотим проверить гипотезу о влиянии местоположения дома на цену: недвижимость в столичных регионах стоит дороже, то есть, если дом находится в любом другом регионе, его стоимость будет ниже, поэтому коэффициент при фиктивной переменной, показывающей, находится ли дом в столичном регионе Бельгии должен быть положительным.

Также в первой упомянутой статье сделан вывод о том, что следующим важным фактором цены на жилье является год постройки здания. Мы хотим проверить гипотезу о влиянии года постройки на цену: чем новее дом, тем выше его стоимость, и наоборот, следовательно, коэффициент при переменной “год постройки” должен быть положительным.

В работе Models Of Factors Influencing The Real Estate Price [4] также упоминается влияние расположения в центре на цену, при этом отмечается, что у данного вида недвижимости отсутствуют парковки и парки, но в сравнении с расположением они не значимы. Таким образом, мы решили проверить гипотезу о влиянии площади сада на цену: недвижимость с большей площадью сада имеет более высокую стоимость, и наоборот, следовательно, коэффициент при переменной “площадь участка вокруг дома” должен был положительным.

# Предварительный анализ данных

Предварительно данные были очищены от null значений по всем переменным признакам для возможности построения множественной регрессии.

Далее из признаков “Объем выбросов CO<sub>2</sub>” и “Площадь” были удалены несколько совсем не реалистичных значений, которые могут негативно влиять на качество модели. По ящикам с усами (Приложение 4) видим, что выбросы остаются, но их количество небольшое и значения реальные, поэтому оставляем в датасете для дальнейшего анализа.

Анализируя корреляционную матрицу (Приложение 3) делаем вывод, что среди наших переменных не наблюдается сильной корреляции с ценой (сильная статистическая зависимость, если коэффициент корреляции  $r > 0,7$ ), поэтому все параметры будут в дальнейшем включены в модель регрессии. При этом все регрессоры, кроме объема выбросов CO<sub>2</sub> имеют положительную корреляцию с ценой, что можно также наблюдать на диаграммах рассеяния (Приложение 6)

В Таблице 1 приведены основные описательные статистики: минимум, максимум, медиана, среднее значение, стандартное отклонение для переменных модели. Можно сделать вывод о том, что данные адекватны: в них нет ошибок, контринтуитивных значений и аномалий.

В нашем случае, из-за присутствия больших значений некоторых признаков, гистограммы являются недостаточно информативными для визуального определения нормальности распределения, но так как выборка достаточно большая, то считаем, что распределение факторов асимптотически стремится к нормальному. Также по круговым диаграммам (Приложение 5) видим отсутствие дисбаланса классов в переменных столичный регион и газовое отопление.

## Оценка модели:

В данной работе применяется модель многомерной линейной регрессии. В качестве регрессанта используется стоимость объекта недвижимости (`price`), в роли регрессоров выступают год постройки (`construction_year = con`), площадь парковочного места (`outdoor_parking_spaces = par`), площадь участка (`garden_surface = gar`), столичный/нестоличный регион (`is_capital_region = cap`), газовое/негазовое отопление дома (`heated_with_gas = gas`), объем выбросов co<sub>2</sub> (`co2_emission = co2`).

В начале проведём тест Чоу для оценки нашей модели. Для теста Чоу мы использовали модель без бинарных переменных. Мы также разбили данные на две разные подвыборки, одну для недвижимости с газовым отопление, а

другую - без газового отопления, поскольку классы переменной `heated_with_gas` достаточно сбалансированы. Нулевая гипотеза теста Чоу заключается в том, что в двух подвыборках нет структурных изменений. То есть индивидуальные коэффициенты регрессии двух подвыборок равны.

В результате мы получили  $F$  - расчётное = 8,2,  $F$ -критическое на уровне значимости 5% = 2,23, то есть  $F$  - расчётное ( $F$ -статистика теста Чоу) >  $F$  - критическое (было рассчитано на уровнях значимости 1, 5 и 10%), следовательно, нулевая гипотеза отвергается при любом разумном уровне значимости, для недвижимости с газовым отоплением и без имеет место разная зависимость. То есть не следует оценивать объединённую выборку. В таком случае обычно вводятся dummy переменные, позволяющие не задавать заранее какой-либо коэффициент одинаковым для обеих подвыборок, именно это мы и сделали. Добавление дополнительного фактора/фиктивной переменной для указания того, находится ли недвижимость в столичном регионе, может помочь более точно описать взаимосвязи в каждой из подвыборок, так как между данными признаками наблюдается слабая корреляция.

Исходя из теоретических и практических соображений, мы приняли решение строить модель используя все перечисленные переменные датасета.

Оценка модели была осуществлена с помощью МНК, ее результаты представлены в Таблице 2. Поскольку значение  $F$ -statistic составляет 18.46, можно считать, что выбранная модель адекватна и релевантна, значение  $R$ -squared для данной модели(0,172) больше, чем для модели без фиктивных переменных, которая использовалась при проведении теста Чоу(0,065), соответственно, это может указывать на лучшее соответствие модели данным. Выводы о значимости независимых переменных были сделаны на уровне значимости 5% по значениям  $p$ -value: если значение  $p$ -value меньше 0.05, то коэффициент считается значимым на уровне 5%. Таким образом, в нашей модели все независимые переменные модели, кроме объема выбросов CO2 значимы.

Рост значений таких показателей как **год постройки** (`construction_year = con`), **площадь парковочного места** (`outdoor_parking_spaces = par`), **площадь участка вокруг дома** (`garden_surface = gar`), **столичный/нестоличный регион** (`is_capital_region = cap`), **газовое/негазовое отопление дома** (`heated_with_gas = gas`) приводят к росту стоимости объектов недвижимости, поскольку коэффициенты  $\beta > 0$ . Такая закономерность достаточно логична, так как данные факторы указывают на привлекательность недвижимости с точки зрения комфорта, удобств, что может повышать её стоимость на рынке. Эти факторы отражают потребности и предпочтения покупателей, а также качество и уникальные характеристики недвижимости. Рост значения показателя **объем выбросов co2** приводит к снижению стоимости объектов недвижимости, поскольку коэффициенты  $\beta < 0$ . Это логично, поскольку высокий объем выбросов CO2 жилого дома связан низкой

энергоэффективностью отопления и кондиционирования помещения, загрязнением окружающей среды и негативным воздействием на экосистему. Покупатели недвижимости могут предпочитать места с более чистым воздухом, что может влиять на их готовность платить за жилье, также высокие уровни выбросов CO<sub>2</sub> могут быть индикатором неэффективности систем отопления, вентиляции и кондиционирования в доме. Потенциальные покупатели могут опасаться дополнительных расходов на модернизацию систем для снижения выбросов и повышения энергоэффективности.

## Выводы

Целью исследования было изучение влияния характеристик домов в Бельгии на их стоимость с использованием модели линейной регрессии. Все выдвинутые нами гипотезы о влиянии на цену подтвердились. Рассмотрим их ещё раз по порядку: 1) факт расположения дома в столичном и не столичном регионе действительно влияет (положительно) на цену недвижимости (также присутствует слабая положительная корреляция). Подтверждение гипотезы логично, поскольку столичные регионы часто ассоциируются с высоким статусом и престижем. Жилье в столичных районах может считаться более привлекательным, что может увеличивать спрос и, следовательно, цены. 2) год постройки также оказывает влияние (положительное, т.е. чем новее дом, тем выше его цена) на стоимость объекта (коэффициент корреляции между ценой и годом постройки равен 0.11 - это говорит об очень слабой положительной зависимости). Это может быть связано с тем, что современные строительные материалы и инженерные решения могут обеспечивать более высокую энергоэффективность и долговечность, а также повышается уровень комфорта недвижимости. 3) Кроме того, величина площади сада влияет на цену (положительно, то есть, чем больше площадь участка, тем больше цена). Стоит отметить, что согласно коэффициенту корреляции, равному 0.21, зависимость между признаками является слабой. Логично, что этот фактор оказывает влияние на цену, так как покупатель платит за квадратный метр земли. При этом слабая связь может быть обоснована тем, что не все покупатели могут придавать большое значение величине площади сада, так как большой сад может потребовать больше ухода и обслуживания за ним.

# Литература

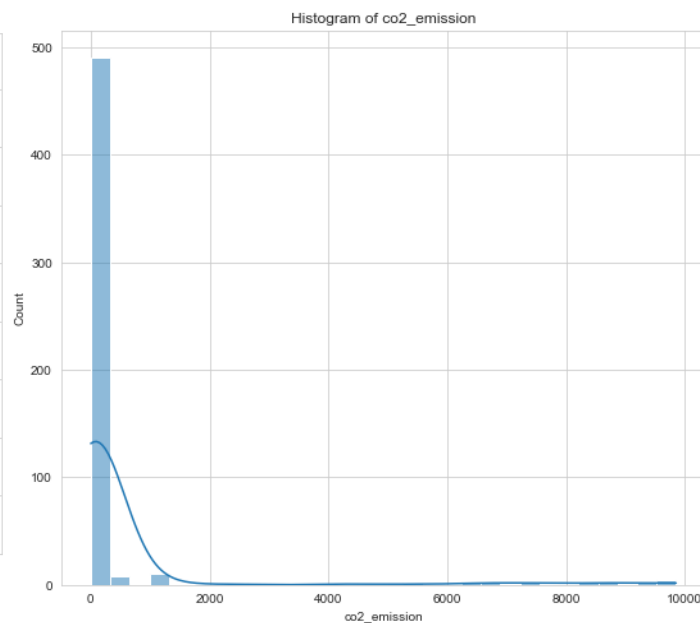
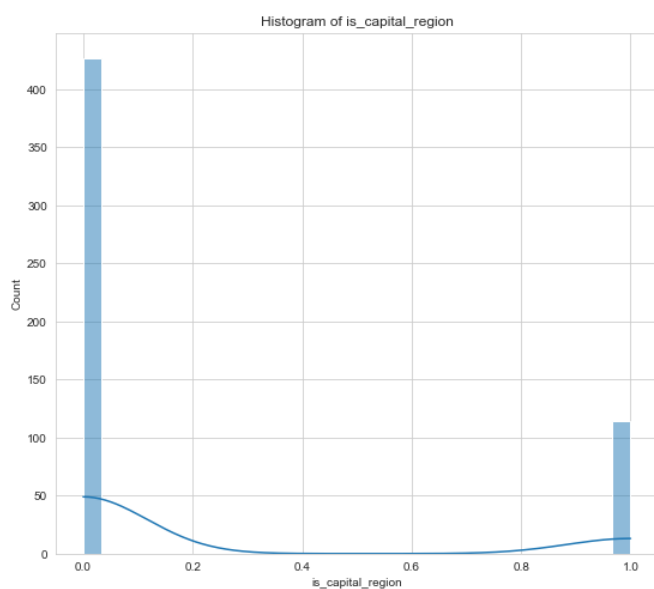
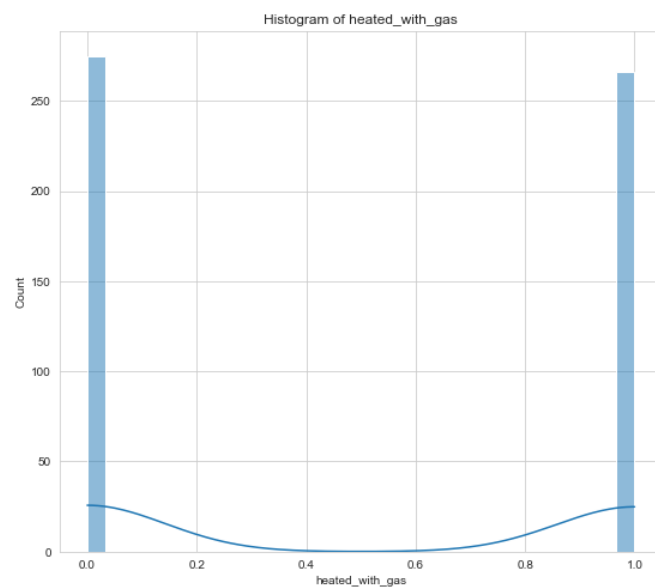
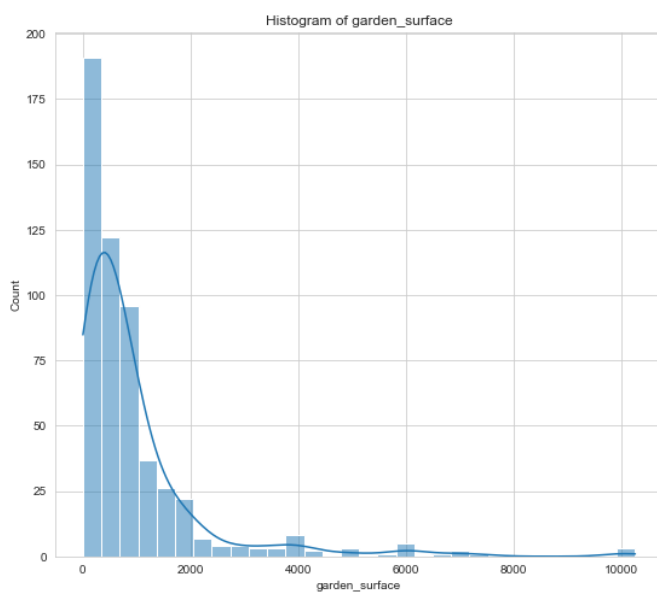
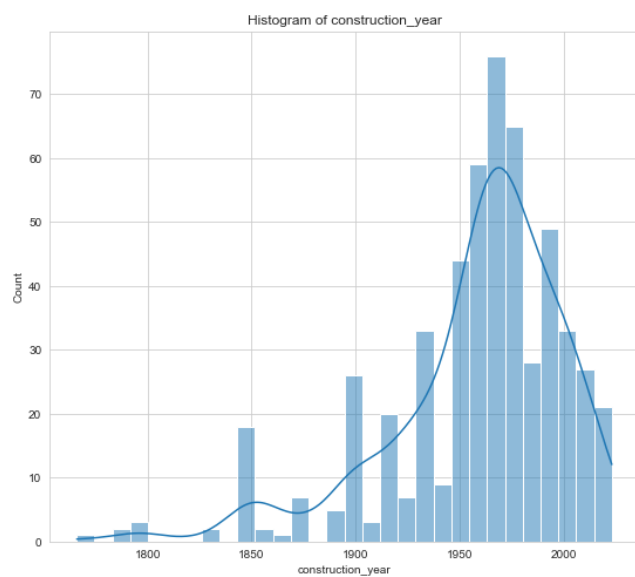
1. Analysis of factors influencing the price of real estate based on interpretative structural model Jingxin Gao , Hong Ren, Yongjie Du, January 2018, Advances in Social Science, Education and Humanities Research (ASSEHR), volume 130: 44-49 URL:  
[https://www.researchgate.net/publication/323180729\\_Analysis\\_of\\_factors\\_influencing\\_the\\_price\\_of\\_real\\_estate\\_based\\_on\\_interpretative\\_structural\\_model](https://www.researchgate.net/publication/323180729_Analysis_of_factors_influencing_the_price_of_real_estate_based_on_interpretative_structural_model)
2. Models of factors influencing the real estate price, Marija Burinskiene, Vitalija Rudzkiene, Jurate Venckauskaite, January 2011 URL:  
[https://www.researchgate.net/publication/266878986\\_Models\\_of\\_factors\\_influencing\\_the\\_real\\_estate\\_price](https://www.researchgate.net/publication/266878986_Models_of_factors_influencing_the_real_estate_price)
3. Bonetti, F.; Corsi, S.; Orsi, L.; De Noni, I. Canals vs. Streams: To What Extent Do Water Quality and Proximity Affect Real Estate Values? A Hedonic Approach Analysis. Water 2016, 8, 577. URL: <https://doi.org/10.3390/w8120577>
4. Research on Factors Affecting Real Estate Values by Data Mining, Filiz Ersoz, Taner Ersoz, Muhammet Soydan ,December 2018, Baltic Journal of Real Estate Economics and Construction Management 6(1): 220-239 URL:  
[https://www.researchgate.net/publication/329778823\\_Research\\_on\\_Factors\\_Affecting\\_Real\\_Estate\\_Values\\_by\\_Data\\_Mining](https://www.researchgate.net/publication/329778823_Research_on_Factors_Affecting_Real_Estate_Values_by_Data_Mining)
5. Датасет - Belgian Property Prices 2023  
<https://www.kaggle.com/datasets/unworried1686/belgian-property-prices-2023>



# Приложение

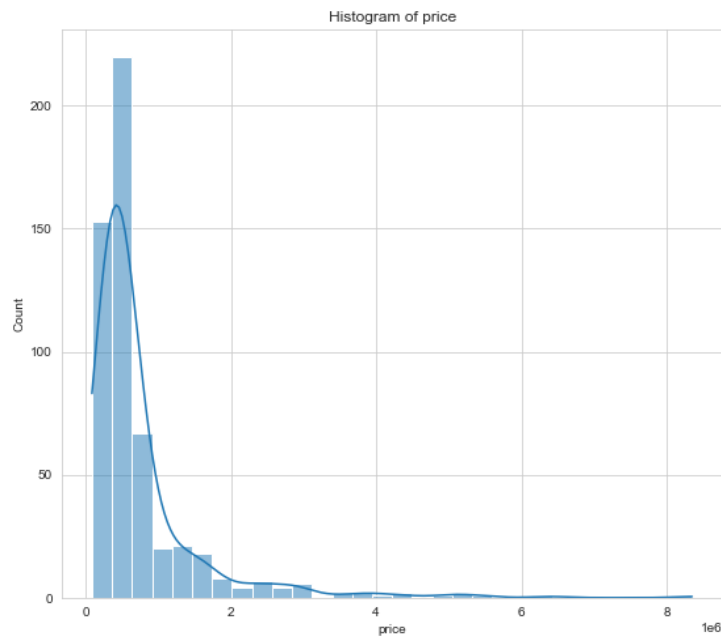
## Приложение 1

Рассмотрим индивидуальные распределения регрессоров.



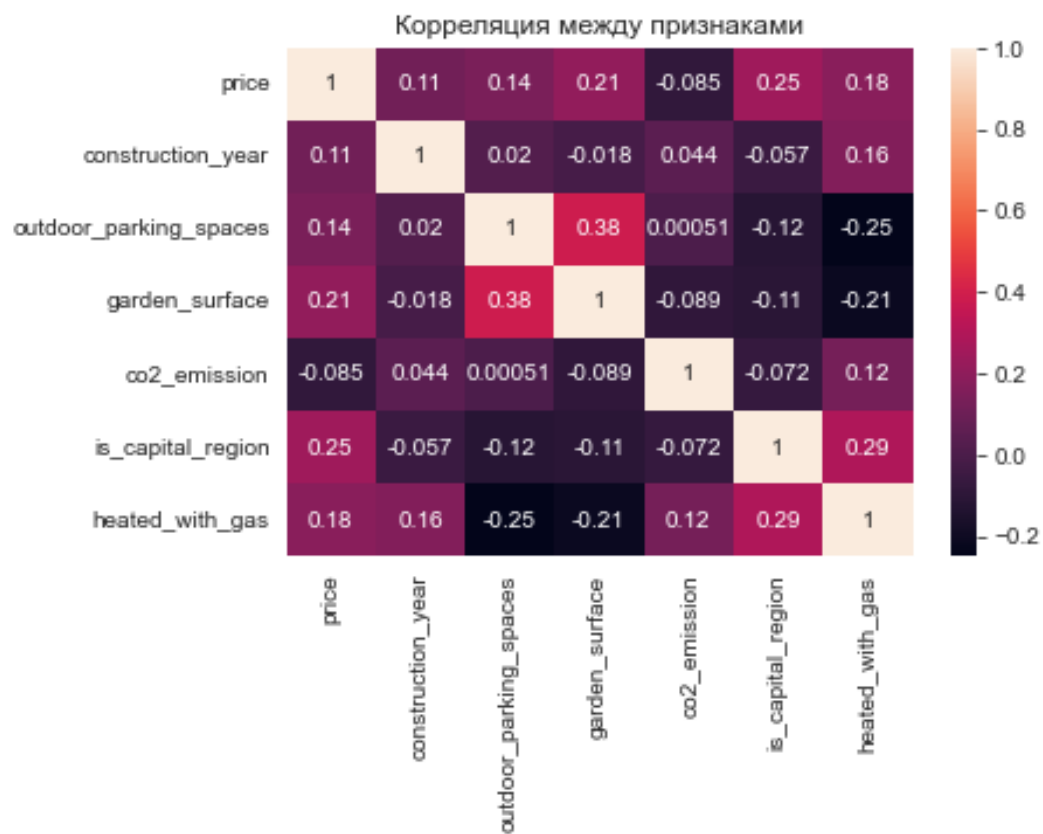
## Приложение 2

### Распределение зависимой переменной

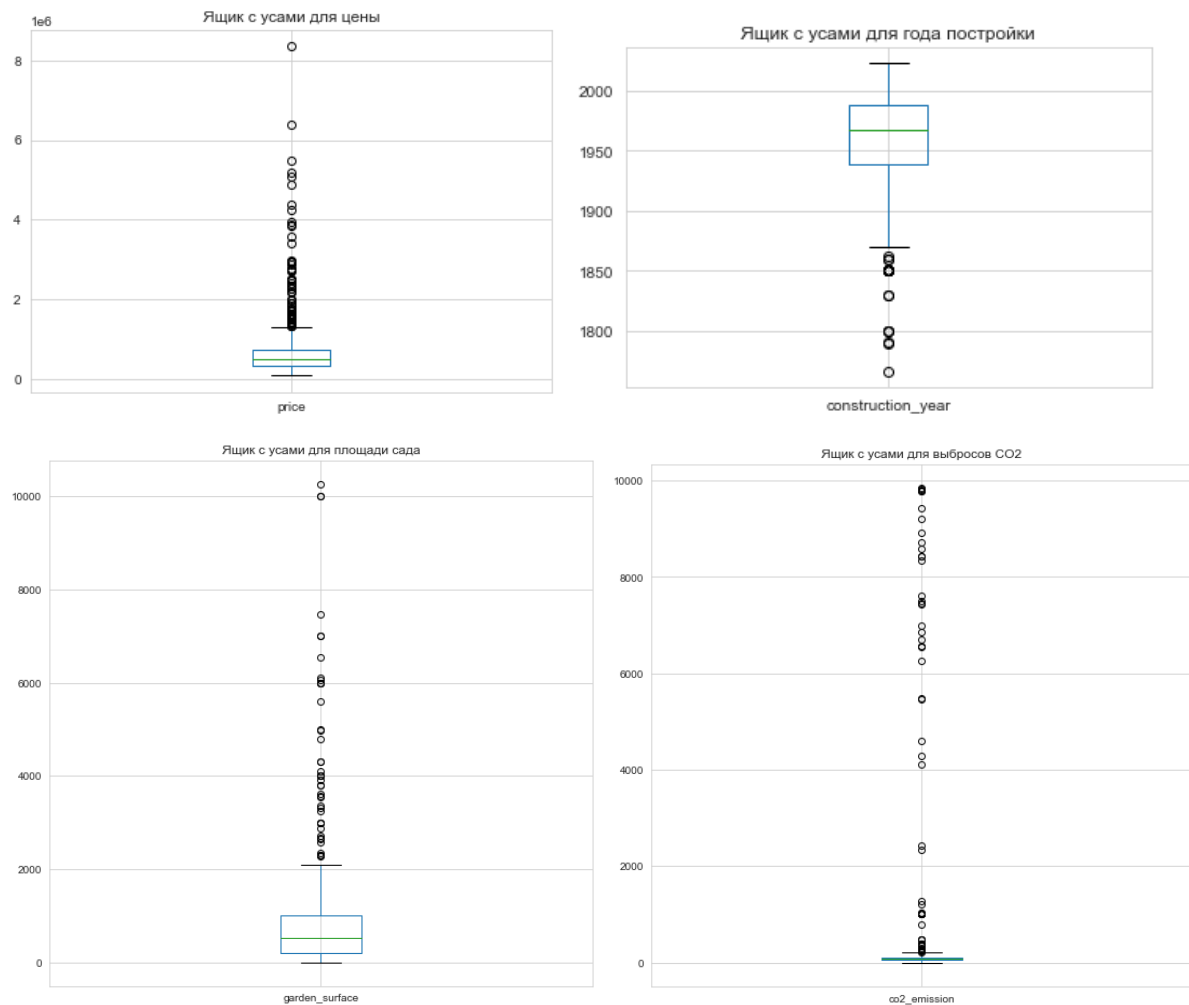


## Приложение 3.

### Корреляционная матрица регрессоров и зависимой переменной



## Приложение 4. Исследование выбросов



## Приложение 5 Баланс классов в переменной Столичный регион и Газовое отопление



## Приложение 6

### Анализ диаграмм рассеяния

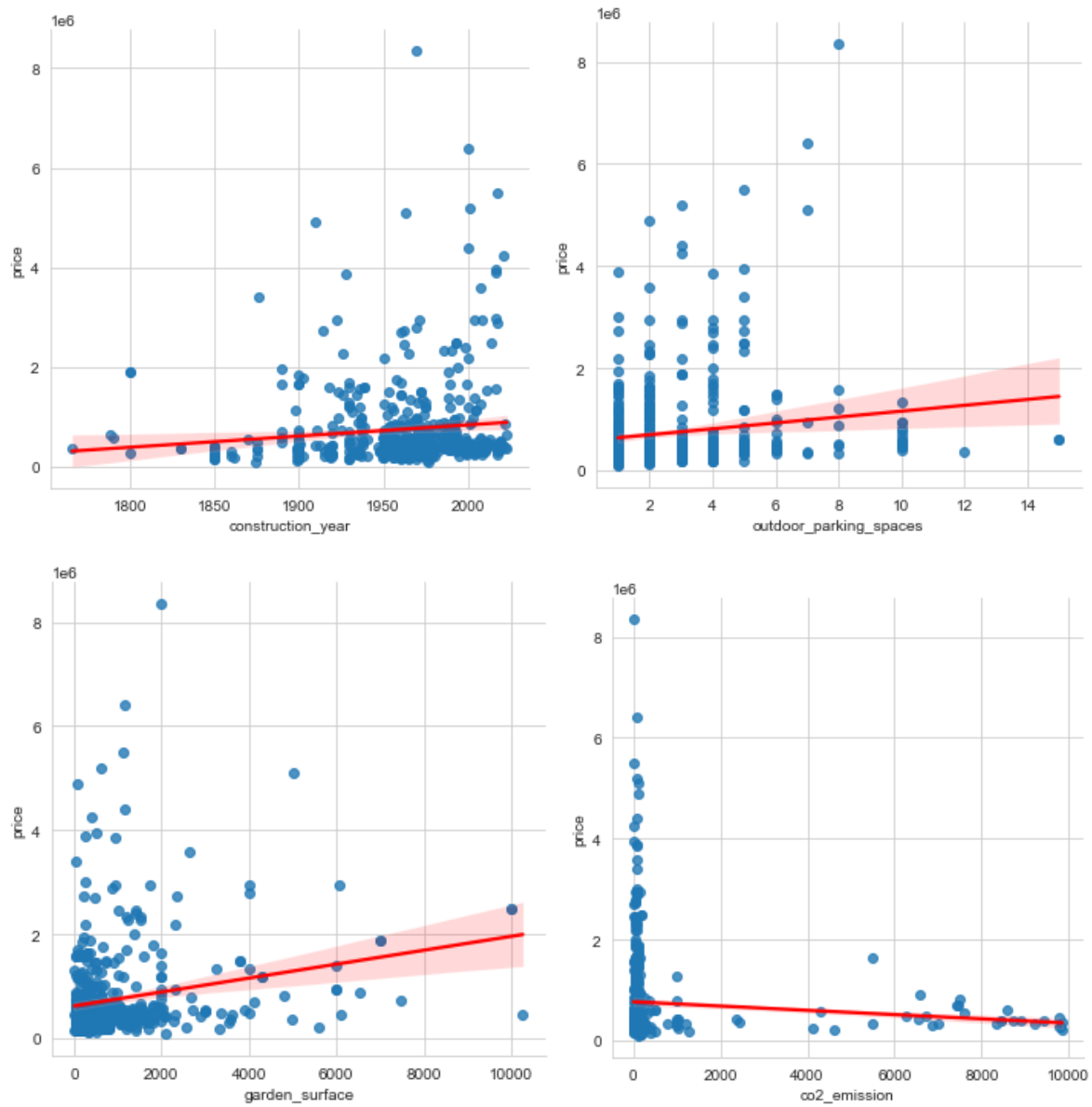


Таблица 1

	price	Constructio n year	Outdoor parking spaces	Garden surface	Co2 emission	Is capital region	Heated with gas
count	541	541	541	541	541	541	541
mean	750273,2	1958,11	2,86	931,47	500,88	0,21	0,49
std	859141,84	42,83	2,1	1327,21	1719,77	0,41	0,5
min	89000	1766	1	1	1	0	0
25%	345000	1939	1	200	49	0	0

50%	490000	1968	2	517	75	0	0
75%	745000	1988	4	1000	113	0	1
max	8350000	2023	15	10264	9854	1	1

Таблица 2

Variable	coef	std err	t	P> t	[0.025	0.975]
Intercept	-3684000	1570000	-2.34	0.02	-6780000	-591000
Con	1997.662	807.016	2.475	0.014	412.346	3582.978
Par	52880	17800	2.965	0.003	17900	87900
Gar	140.136	27.972	5.01	0	85.188	195.085
CO2	-37.678	20.034	-1.881	0.061	-77.034	1.678
Cap	491600	87800	5.597	0	319000	664000
Gas	317200	75100	4.221	0	170000	465000