

Анализ характеристик спроса на квартиры и
апартаменты в новостройках Москвы, Московской
области, Санкт-Петербурга и Ленинградской области в
2019-2020 г.г.

Информационно-аналитическая записка.

Содержание

Постановка задачи.....	3
Анализ.....	5
Методика решения.....	7
Основные результаты.....	10
Выводы и заключение.....	24
Приложения.....	25

Постановка задачи

Описание бизнес-задачи, ее актуальность для заказчика.

Компании-застройщику для реализации нового проекта жилой недвижимости требуется оценить размер, динамику и структуру спроса на территориях своей деятельности. Свои проекты она осуществляет на территории Москвы и области, а также Санкт-Петербурга и Ленинградской области. Эти данные необходимы для выбора параметров проекта, разработки его технико-экономического обоснования и, в частности, его финансовой модели. Они непосредственно влияют на квартирографию проекта, планировочные и другие решения.

Предполагаемые источники информации.

1. Федеральная Служба государственной регистрации, кадастра и картографии (Росреестр)
В соответствии с законом «О государственной регистрации недвижимости» все сделки с квартирами подлежат обязательной государственной регистрации. Данные о переходе права собственности в результате сделки вносятся в единую по всей России базу данных прав на недвижимость Росреестра (ЕГРН).

Это электронная база данных, в которую вносится вся информация о существующих и прекращенных правах на объекты недвижимости, информация о правообладателях, существующих обременениях прав собственности. В ней также хранятся и административно-технические данные о самих объектах недвижимости - их адреса, назначение (жилое/нежилое), размер площади и т.п.

Эти данные, дополненные оценочной ценой сделки, позволяют проанализировать продажи по территориям, застройщикам и отдельным проектам с учетом различных параметров. Например, можно узнать, сколько квартир конкретной площади в проектах различного класса, на каких этажах было продано, какова доля ипотечных сделок и т. д.

2. Открытые источники данных (Яндекс Недвижимость, Яндекс Карты или им подобные, Центральный банк РФ и другие доступные источники информации)

Предполагается оценить возможность и, при их доступности, использовать данные из других источников, чтобы дополнить имеющиеся данные Росреестра характеристиками объектов недвижимости, включая оценки по отзывам потенциальных покупателей, их местоположения, информацией о застройщиках. Это позволит проанализировать дополнительные детерминанты спроса.

Из данных Центрального банка РФ интерес представляют динамика валютного курса российского рубля, позволяющая учесть влияние на спрос макроэкономической ситуации в стране, а также динамика процентной ставки по предоставленным ипотечным жилищным кредитам.

Стейкхолдеры. Сбор бизнес-требований по задаче.

Основными стейкхолдерами по данной задаче выступают представители топ-менеджмента застройщика: финансовый и коммерческий директора.

Для них результаты анализа данных предполагается представить в виде презентации или интерактивного дашборда для каждой из анализируемых территорий, который будет отражать основные характеристики спроса на квартиры и апартаменты за указанный период.

Вторичной аудиторией данной работы станут сотрудники коммерческой и маркетинговой служб, которым результаты анализа данных могут быть предоставлены в форме аналитической записки более подробно.

В целом потребности в информации обеих категорий стейкхолдеров можно свести к данным об объемах продаж в натуральном и стоимостном измерении для следующих срезов:

- . Календарный период (год, месяц, возможно, квартал)
- . Территория (регион, административно-территориальные единицы, муниципальные образования, направление, а также условные зоны)
- . Класс жилья
- . Застройщик

- . Жилой комплекс
- . Тип помещения
- . Площадь
- . Этаж
- . Количество комнат
- . Наличие и тип обременений (прежде всего, ипотеки)
- . Залогодержатель
- . Тип дома (по технологии строительства)

Дополнительно для анализа могут быть использованы средние значения:

- . Стоимости одного квадратного метра площади
- . Стоимости проданных объектов (квартир, апартаментов)
- . Площади проданных объектов

Основные гипотезы.

В рамках поставленных задач требуется проверить следующие гипотезы, характеризующие изменение спроса на первичном рынке в 2020 году по сравнению с 2019 годом:

- . Увеличились продажи объектов жилой недвижимости (включая апартаменты).
- . Увеличилось количество сделок с ипотекой.
- . Спрос на объекты комфорт и бизнес-класса увеличился в большей степени, чем на объекты других классов.
- . Средняя стоимость квадратного метра площади возросла.
- . Значимость (веса) переменных, влияющих на стоимость единицы площади, в 2020 году изменилась.

Для проверки гипотез требуется провести анализ следующих данных за 2019-2020 г.г.:

- . Средняя стоимость объекта недвижимости с разбивкой по типам помещений и территориям.
- . Средние показатели стоимости объекта и единицы площади в разных классах жилья.
- . Топ застройщиков по объему продаж. Средние значения показателей площади проданных помещений, стоимости сделок, стоимости единицы площади.
- . Количество и сумма сделок с ипотекой.
- . Топ залогодержателей с разбивкой по территориям и классу жилья.
- . Факторы, влияющие на спрос (количество проданных помещений) и на среднюю стоимость единицы площади (провести регрессионный анализ).
- . Наличие разницы в предпочтениях покупателей – физических лиц и юридических лиц.

Метрики. Ограничения и допущения.

Для анализа данных и проверки указанных гипотез предполагается использовать следующие количественные показатели:

- . Объем продаж в абсолютном выражении, шт.
- . Объем продаж в стоимостном выражении, руб.

Как уже отмечалось, в данных Росреестра отсутствует информация о стоимости сделки. Поэтому она добавлена поставщиком данных (dataflat.ru) и является оценочной, на основании прайс-листов, без учета возможных скидок и рассрочки платежей. В случае отсутствия в прайс-листах, она рассчитывается или указывается другое значение по одному из алгоритмов, описанных в спецификации (см. Приложение).

Анализ

Исследование аналогичных решений.

Аналогичных решений, позволяющих в полном объеме провести требуемый анализ рынка, в открытом доступе, не обнаружено.

Наиболее близким решением можно считать Единую информационную систему жилищного строительства (ЕИСЖС) Минстроя России, доступную на сайте наш.дом.рф (оператором является АО «Дом.РФ»). Она позволяет получить аналитическую информацию о рынке жилищного строительства за период от месяца до года как для России в целом, так и для отдельных субъектов и городов в составе субъектов, а также аналитическую информацию по отдельным девелоперам.

В частности, по проданным квартирам доступны:

. Основные характеристики реализованных квартир:

Стоимость, млн. руб.

Количество, штук

Площадь, тыс. кв. м

Цена за 1 кв. м, тыс. руб.

Площадь квартиры, кв. м

Стоимость квартиры, млн. руб.

. Структура сделок (по интервалам значений):

По стоимости жилья, млн. руб.

По цене 1 кв. м, тыс. руб.

По площади жилья, кв. м

. Топ девелоперов (-10, -20, -50), с указанием по каждому из них следующих показателей:

Суммарная стоимость проданных ДДУ, млн. руб.

Количество проданных квартир, штук

Площадь проданных квартир, тыс. кв. м

Медианная площадь квартиры, кв. м

Средняя площадь квартиры, кв. м

Медианная цена 1 кв. м, тыс. руб.

Медианная цена квартиры, млн. руб.

Средняя цена за 1 кв. м, тыс. руб.

Средняя цена квартиры, млн. руб.

Но, на сайте представлены не полные данные по первичному рынку недвижимости, а только по ДДУ, заключенным на приобретение жилых помещений (ресурс создан в рамках постановления Правительства РФ от 26.03.2019 № 319 «О единой информационной системе жилищного строительства»). Это, а также фиксированный набор срезов данных, делает ЕИСЖС неприменимой для целей нашего анализа.

Сайт можно использовать как источник дополнительных данных. На нем доступен сервис API «Сведения о жилищном строительстве», предназначенный для предоставления сведений о группах компаний, застройщиках, жилых комплексах и объектах строительства. Спецификация и презентация API доступны по ссылке: наш.дом.рф/интеграция

Предварительный анализ данных.

Для решения поставленных задач были получены данные в виде файла с расширением .csv и спецификация с названиями полей, их описанием и комментариями, поясняющими содержание данных.

Анализ данных проведен в Jupyter Notebook с использованием языка и библиотек Python.

Исходный массив представляет собой случайную выборку (30-35%) из базы Росреестра за 2019-2020 г.г. и содержит 369 845 записей в 57 колонках (полях). В 28 полях имеются пропуски данных. Кроме того, как выяснилось на этапе устранения пропусков данных, есть ошибочные значения и опечатки в текстовых полях.

В соответствии с постановкой задачи из этого массива отобрали квартиры и апартаменты с датой регистрации в интервале с 1.01.2019 по 31.12.2020. Получилось 295 020 записей о продажах в 1024 жилых комплексах от 432 застройщиков.

Этому предшествовало устранение пропущенных значений в поле «Тип помещения», обнаруженных в 16-ти строках. Для этого были просмотрены описания помещений (поле «Описание помещения») и для предположительно жилых помещений был восстановлен тип и отсутствующие значения в других полях (чаще всего, комнатность) по объектам с аналогичными характеристиками (ЖК, корпус, площадь).

Далее выводим основные статистические показатели по выборке данных с помощью метода describe() для того, чтобы получить более полное представление о данных и их качестве.

Основные статистические показатели по выборке данных:

- . Средняя площадь проданного жилья: 47,1 кв.м
- . Медианное значение площади: 40,6 кв.м (меньше среднего, что свидетельствует о положительной асимметрии распределения)
- . Не менее половины проданных объектов - однокомнатные
- . Стоимость сделки:
 - среднее значение: 6,79 млн. рублей
 - медиана: 4,96 млн. рублей
- . Стоимость 1 кв. м:
 - среднее значение: 135 тыс. рублей
 - медиана: 118 млн. рублей
- . С привлечением ипотеки совершено не менее половины сделок (фактически, 50,5%)
- . Количество жилых комплексов: 1024
- . Количество застройщиков: 432
- . Наиболее представленный регион объекта: Санкт-Петербург (следующий: Москва + Новая Москва)
- . Жилой комплекс с наибольшими продажами: Чистое небо
- . Лидер среди застройщиков по продажам (шт.): ПИК
- . Больше всего объектов относится к классу комфорт: 157 388 (53%)
- . Наиболее часто встречающийся залогодержатель: Сбербанк
- . Чаще всего договор долевого участия (ДДУ) заключался на стадии котлована
- . Полнота данных по полю Стадия строительства в дату ДДУ: 59%

Одновременно, выявлены экстремальные значения переменных, которые выглядят ошибочными и требуют проверки. Это, например:

Минимальное значение площади помещения: 2,7 кв. м

Максимальное значение площади помещения: 3135,3 кв. м

Минимальное значение этажа: -2

Максимальное значение этажа: 4563

Следующий этап работы был посвящен выявлению, анализу и устранению пропусков в значениях наиболее важных с точки зрения решения поставленной задачи переменных, исправлению неверных значений и анализу значений, значительно отличающихся от остальных. Модели линейной регрессии, которые планируется использовать, чувствительны к выбросам.

Методика решения

К числу наиболее значимых переменных были отнесены: площадь помещения, количество комнат, этаж, корпус и оценка стоимости. По каждой из них очистка данных сводилась к следующим действиям:

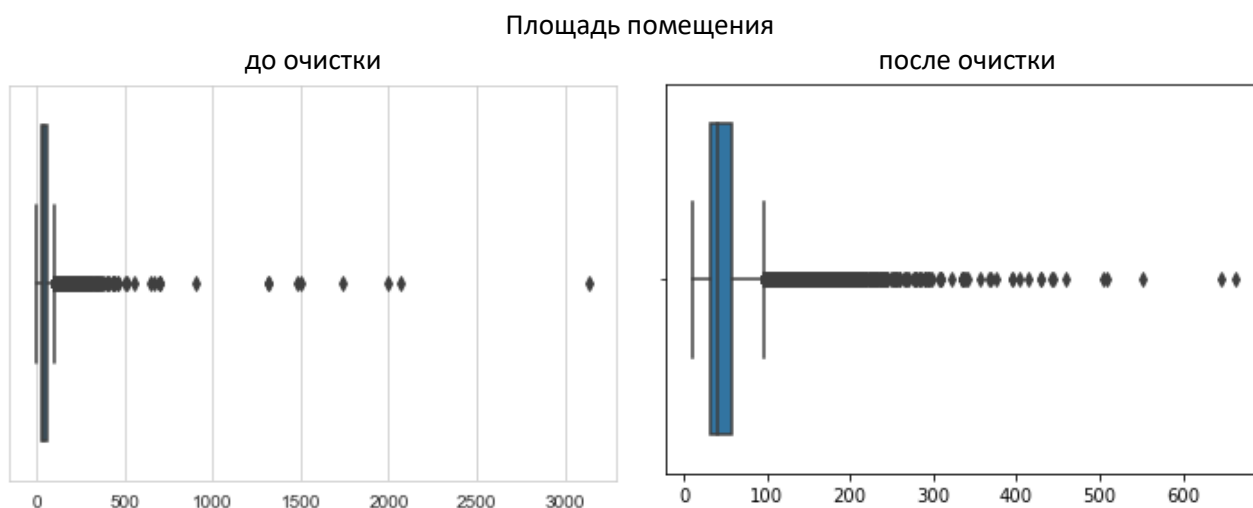
1. Отсутствующие значения по возможности заполнялись, в случае отсутствия такой возможности – исключались из набора данных. Для заполнения отсутствующих значений использовалась информация из описания помещения, расчеты на основе данных из других строк (например, среднее значение площади помещения по квартирам с известной площадью в том же корпусе ЖК и с тем же количеством комнат), а также значения из похожих записей.
2. Значения, значительно отличающиеся от остальных, проверялись на корректность заполнения остальных полей, по возможности исправлялись, а некорректные и при отсутствии возможности исправления, недостатка информации для принятия решения – исключались из набора данных.
3. Ошибки в записях, опечатки, по возможности исправлялись, иначе – удалялись. Так, например, были выявлены и удалены записи, содержащие значения нескольких ID ЖК в одной строке.

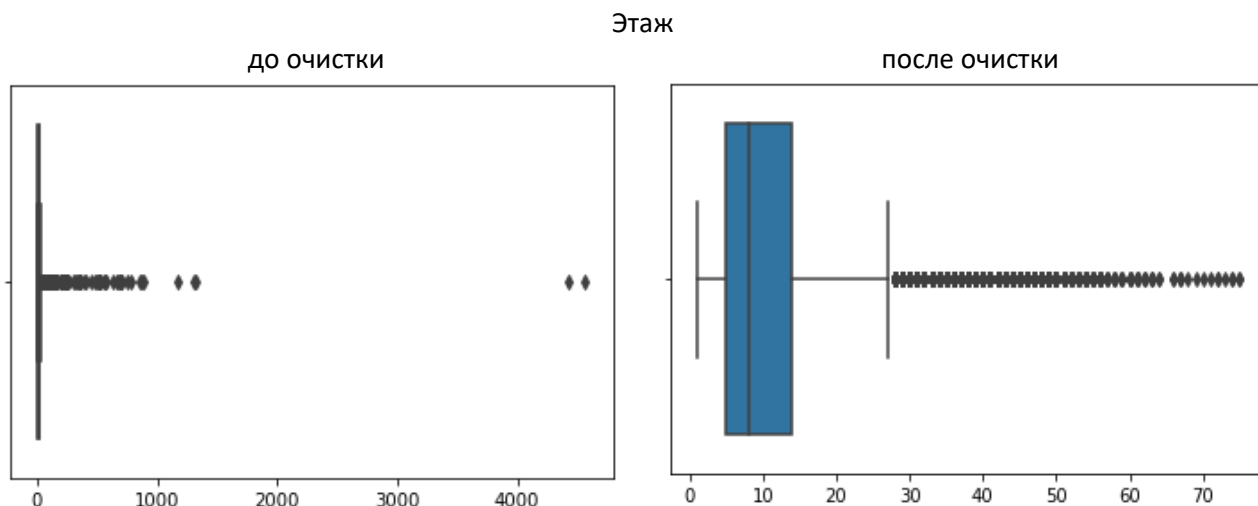
Основные решения, принятые по указанным полям, указаны в Приложении (с. 28).

Наряду с этим осуществлялась проверка между полями, которые связаны определенными условиями заполнения. Например, проверялось соблюдение условия, что для всех помещений, имеющих обременение (поле «Тип обременения») должен быть указан залогодержатель (поле «Залогодержатель»).

В результате было выявлено несоответствие в заполнении поля с типом покупателей («is_buyer_ul») и полей с указанием наименований (для юридических лиц) и идентификаторов покупателей (для физических лиц). В 135 строках для юридических лиц были указаны наименования и идентификаторы обоих типов покупателей. Для дальнейшего анализа было принято решение руководствоваться значениями типа покупателей.

В результате очистки данных их качество повышено. Об этом могут свидетельствовать распределения значений переменных до и после очистки:





Далее был проведен анализ данных и осуществлена проверка выдвинутых гипотез. Для этого использовались два основных количественных показателя:

- . Объем продаж в абсолютном выражении, шт.
- . Объем продаж в стоимостном выражении, руб.

Дополнительно были использованы средние значения таких показателей как:

- . Стоимость одного квадратного метра площади
- . Стоимость проданных помещений (квартир, апартаментов)
- . Площадь проданных помещений

А также:

- . Количество покупателей помещений
- . Доля ведущих застройщиков в продажах помещений

В ходе анализа были получены и визуализированы данные для следующих срезов:

- . Объем продаж в натуральном и стоимостном измерении по годам 2019-2020 г.г.
- . Количество покупателей помещений по типу – физические и юридические лица по годам 2019-2020 г.г.
- . Средняя стоимость 1 кв. м для проданных квартир и апартаментов по годам 2019-2020 г.г.
- . Динамика объема продаж в натуральном и стоимостном измерении по месяцам за исследуемый период в целом и по классам помещений (эконом, эконом (панель), комфорт, бизнес, премиум, делюкс).
- . Динамика средней стоимости 1 кв. м проданных помещений по месяцам за исследуемый период в целом и по классам помещений (эконом, эконом (панель), комфорт, бизнес, премиум, делюкс).
- . Динамика средней стоимости проданных помещений по месяцам за исследуемый период в целом и по классам помещений (эконом, эконом (панель), комфорт, бизнес, премиум, делюкс).
- . Динамика средней площади проданных помещений по месяцам за исследуемый период.
- . Объем продаж по типам помещений в натуральном и стоимостном измерении по годам 2019-2020 г.г.
- . Объем продаж по классам помещений в натуральном и стоимостном измерении по годам 2019-2020 г.г.
- . Объем продаж помещений по количеству комнат в натуральном и стоимостном измерении по годам 2019-2020 г.г. (студии, одно-, двух-, трех-, четырех- и более комнатные)
- . Объем продаж помещений в зависимости от этажа в натуральном и стоимостном измерении по годам 2019-2020 г.г. (1-2 этажи, 3-7 этажи, 8-15 этажи, 16-25 этажи, 26 этаж и выше)
- . Объем продаж помещений в зависимости от площади в натуральном и стоимостном измерении по годам 2019-2020 г.г. (до 25 кв.м, 26-40 кв.м, 41-55 кв.м, 56-70 кв.м, 71-85 кв.м, 86-100 кв.м, свыше 100 кв.м)
- . Объем продаж помещений по наличию и виду обременений в натуральном и стоимостном измерении по годам 2019-2020 г.г. (ипотека, военная ипотека, залог, отсутствие обременений)

- . Рейтинг ведущих застройщиков по продажам в натуральном и стоимостном измерении по годам 2019-2020 г.г. (топ-10)
- . Рейтинг ведущих залогодержателей по продажам в натуральном и стоимостном измерении по годам 2019-2020 г.г. (топ-10)
- . Рейтинг ведущих залогодержателей по выручке в 2020 г. (топ-10) по регионам и классам помещений
- . Рейтинг жилых комплексов по объему продаж с указанием количества проданных помещений по годам 2019-2020 г.г. и классам.
- . Объем продаж помещений в зависимости от региона в натуральном и стоимостном измерении по годам 2019-2020 г.г. (Москва, Московская обл., Новая Москва, Санкт-Петербург, Ленинградская обл.)
- . Объем продаж помещений в зависимости от территории в натуральном и стоимостном измерении по годам 2019-2020 г.г. (по административным районам, муниципальным образованиям, направлениям и выделенным зонам регионов)
- . Средняя стоимость квартир и апартаментов в зависимости от территории по годам 2019-2020 г.г. (по административным районам, муниципальным образованиям, направлениям и выделенным зонам регионов)
- . Средняя стоимость проданных помещений у ведущих застройщиков 2020 г. в сравнении с аналогичным показателем по рынку.
- . Средняя стоимость 1 кв. метра помещений у ведущих застройщиков 2020 г. в сравнении с аналогичным показателем по рынку.
- . Средняя площадь проданных помещений у ведущих застройщиков 2020 г. в сравнении с аналогичным показателем по рынку.
- . Средняя стоимость помещений по классам по годам 2019-2020 г.г.
- . Средняя стоимость 1 кв. м помещений по классам по годам 2019-2020 г.г.

Кроме того, с целью анализа факторов, влияющих на спрос (количество проданных помещений) и на среднюю стоимость единицы площади были изучены корреляции переменных и проведен регрессионный анализ с помощью пакета Python Scikit-Learn для машинного обучения.

Решение задачи регрессии позволяет установить математическую зависимость между целевой (зависимой) переменной и независимыми переменными, оценить их веса (значения коэффициентов регрессии) и спрогнозировать значение целевой переменной. Выбрана линейная модель в предположении наличия линейной зависимости между переменными и как наиболее простая в реализации и интерпретации результатов.

На этом этапе имеющийся массив данных был дополнен новыми переменными, предположительно влияющими на целевые переменные.

Это данные из открытых источников:

- . Средневзвешенная ставка по ипотечным кредитам на первичном рынке (%) по месяцам 2019-2020 г.г. (источник: Центральный банк РФ)
- . Рейтинг ЕРЗ – показатель соблюдения застройщиком декларируемого срока ввода жилья (источник: портал «Единый ресурс застройщиков» (ЕРЗ.РФ), что такое рейтинг ЕРЗ и как он рассчитывается: <https://erzrf.ru/news/chto-takoye-reyting-yerz-i-kak-on-rasschityvayetsya>)
- . Оценка ЕРЗ - оценка потребительских качеств жилых комплексов на основе 100+ показателей:
 - транспортная доступность
 - парковка
 - социальная инфраструктура
 - объекты спорта
 - дворовые пространства, безопасность
 - и т.д.

А также данные, полученные преобразованием имеющихся в датасете:

- . Категории застройщиков по показателю среднегодовой (за исследуемый период) выручки (выделены на основании расчета статистических показателей выручки и визуального анализа распределения застройщиков по среднегодовой выручке):
- low: до 0,37 млрд. рублей (208 застройщиков)

- lower-middle: 0,37 – 2,4 млрд. рублей (132 застройщика)
- middle: 2,4 – 10 млрд. рублей (53 застройщика)
- upper-middle: 10,1 – 23 млрд. рублей (13 застройщиков)
- high: свыше 23 млрд. рублей (10 застройщиков)

Основные результаты анализа данных приведены ниже, полные результаты и анализ представлены в Jupyter Notebook.

Основные результаты.

В рамках поставленных задач были проверены гипотезы, характеризующие изменение спроса на первичном рынке в 2020 году по сравнению с 2019 годом.

1. Продажи квартир и апартментов в 2020 году по сравнению с 2019 годом в стоимостном выражении увеличились.

Эта гипотеза подтвердилась. Результаты её проверки представлены на следующей диаграмме:



В 2020 году выручка от реализации квартир и апартментов составила 1,02 трлн. рублей, тогда как годом ранее – 0,93 трлн. рублей. Рост продаж составил 9,3%.

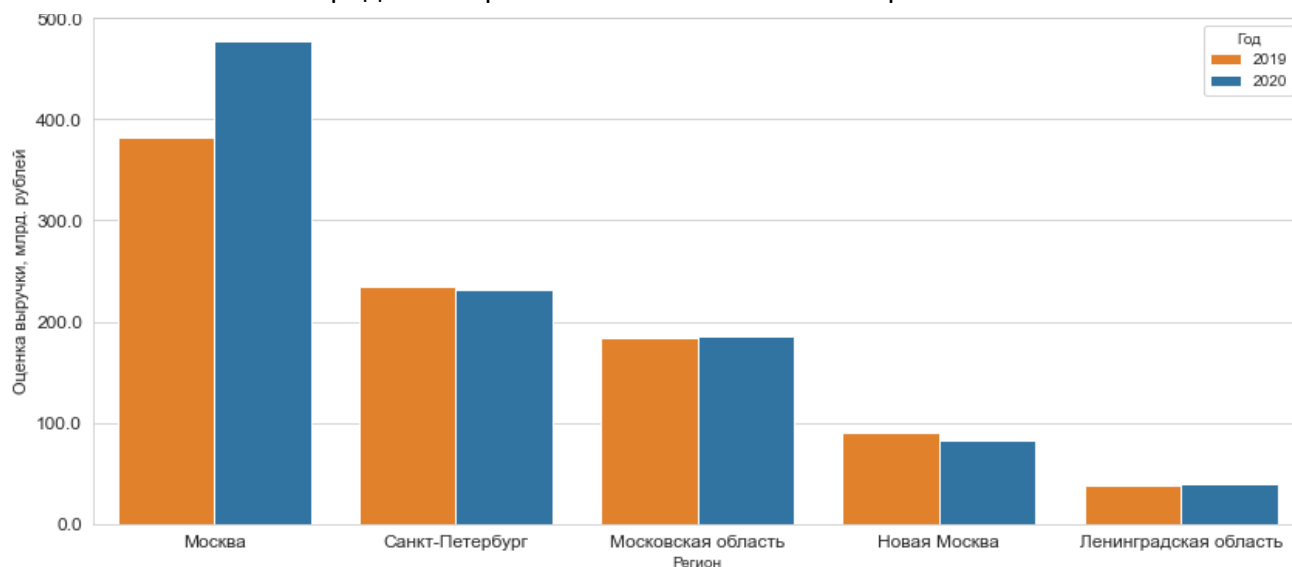
При этом в 2020 году было реализовано на 7,4% помещений меньше: около 140 тыс. против 151 тыс. квартир и апартментов в 2019 году:



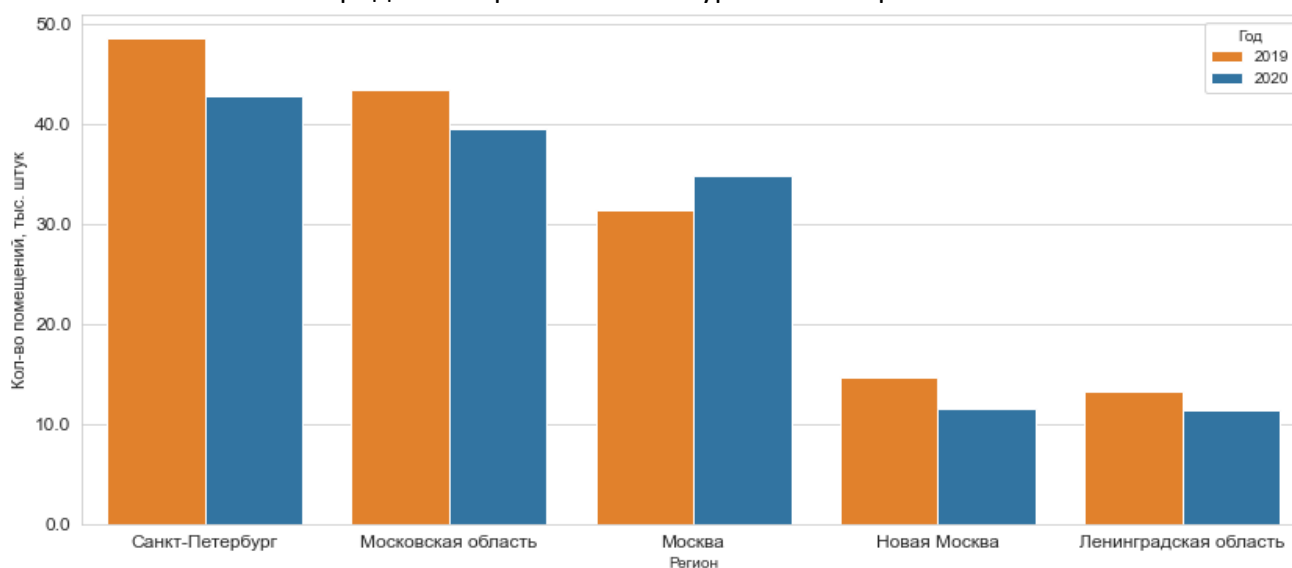
Это свидетельствует о том, что средняя стоимость проданных помещений в 2020 году выше. Причиной этого может быть как изменение стоимости недвижимости, так и изменение структуры спроса в сторону помещений более высокого класса. Об этом – ниже.

Несколько иная ситуация с динамикой продаж по регионам:

Продажи по регионам в стоимостном выражении



Продажи по регионам в натуральном выражении

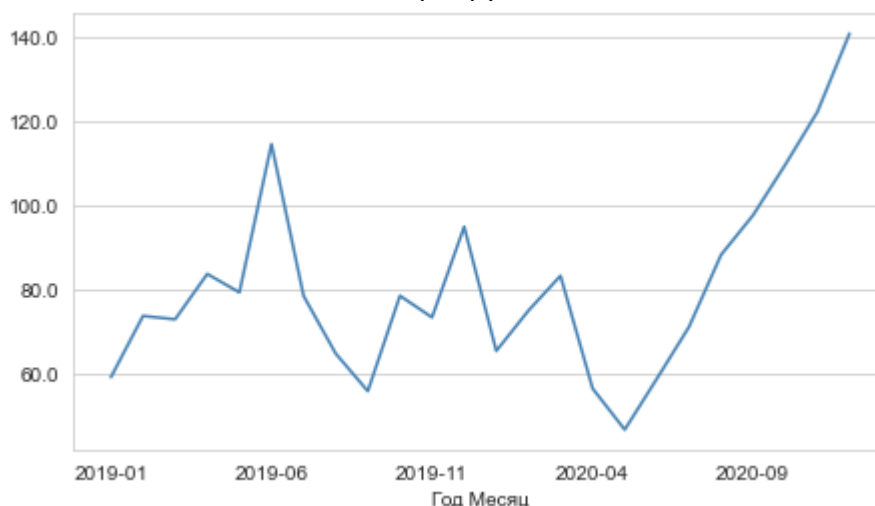


Среди регионов выделяется, прежде всего, Москва: и количество реализованных помещений, и сумма продаж в 2020 году увеличились. Выручка превысила показатель 2019 года на 25%, а количество квартир и апартаментов – на 11%.

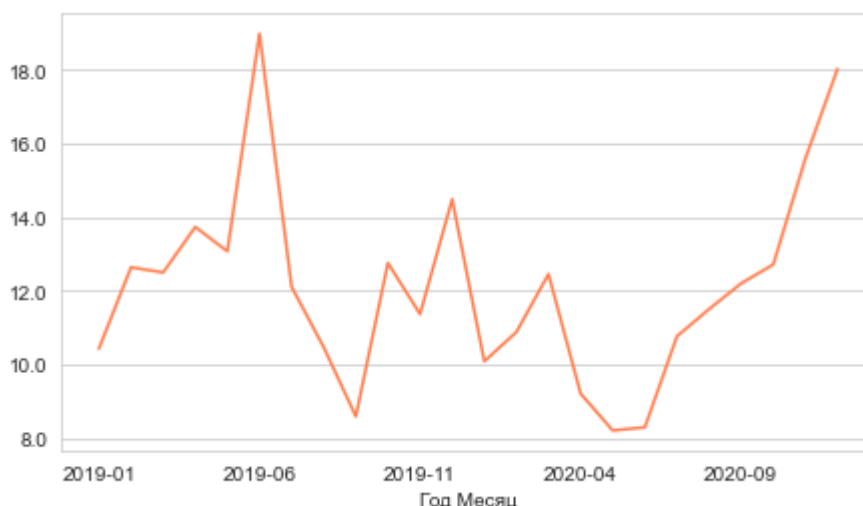
Отрицательная динамика по обоим показателям наблюдается на территории Новой Москвы и в Санкт-Петербурге. Новой Москве в 2020 году принадлежит первенство по размеру снижения продаж: реализовано почти на 22% меньше помещений, что соответствует меньшей на 9% выручке по сравнению с 2019 годом.

Следующая диаграмма показывает динамику продаж по месяцам двухгодичного периода и наглядно демонстрирует влияние на спрос наиболее значимых событий на первичном рынке недвижимости и в стране:

Динамика продаж помещений по месяцам 2019-2020 г.г.,
млрд. рублей



Динамика продаж помещений по месяцам 2019-2020 г.г.,
тыс. штук



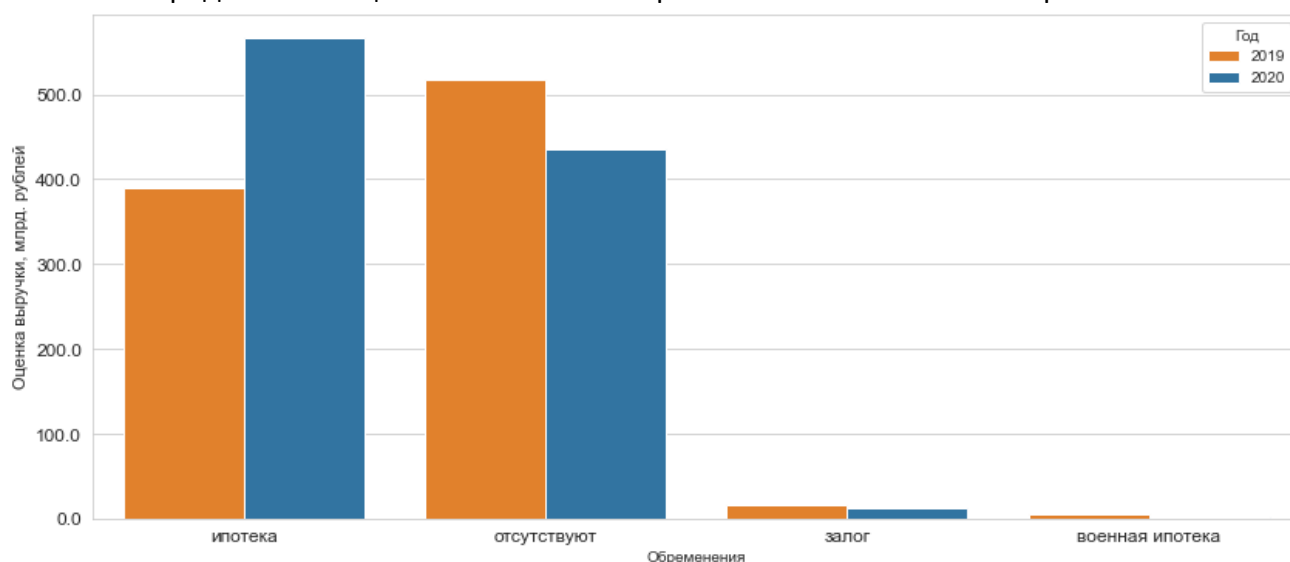
На обоих графиках есть два ярко выраженных пика. Первый приходится на июнь 2019 года и отражает влияние на спрос новостей о переходе отрасли на эскроу-счета. До июня спрос рос, поскольку покупатели стремились совершить покупку на понятных для себя условиях, до вступления в силу новых правил. Затем спрос снижался, пока не стало известно, что большая часть жилых комплексов будет реализована по старым правилам.

С начала 2020 года на рынок оказывала влияние ситуация с заболеваемостью Covid-19 и сопутствующими ограничениями. Спрос достиг минимума в апреле-мае. Последовавший далее стремительный рост продаж во многом обусловлен действием программы льготной ипотеки для покупки жилья в новостройках. По количеству проданных квартир и апартаментов продажи приблизились к показателям июня 2019 года, а по сумме выручки заметно превзошли их.

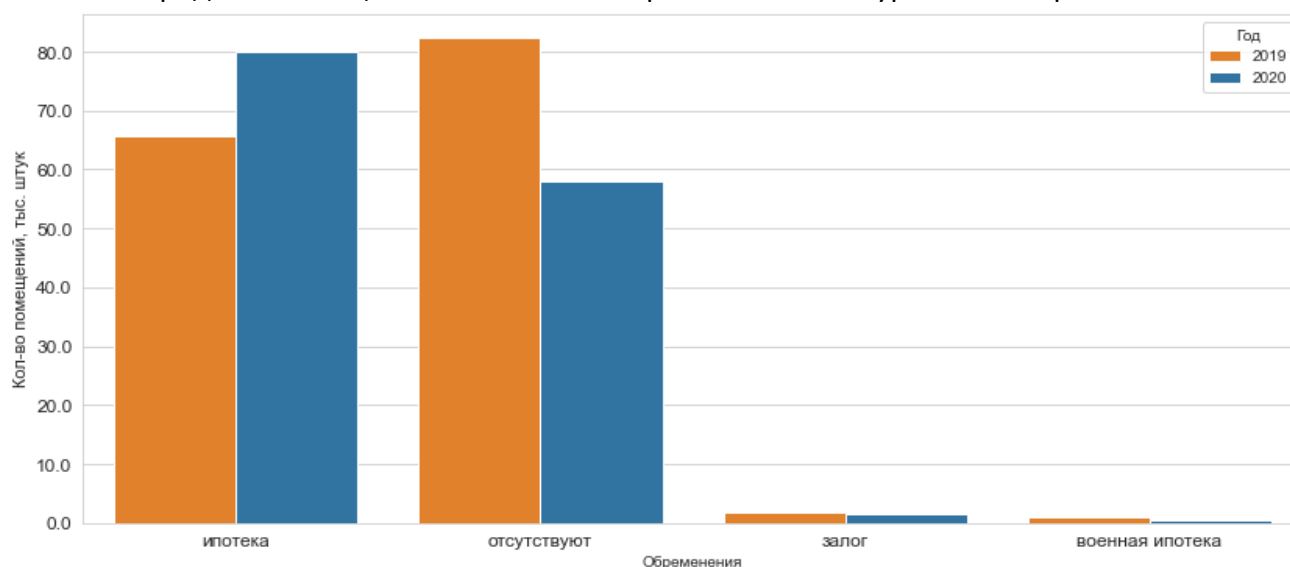
2. Увеличилось количество сделок с ипотекой.

С участием ипотеки в 2020 году продано 57,5% помещений, а в 2019 году – 44,2%. Прирост составил 13,3 процентных пункта (п.п.). В стоимостном выражении доля сделок с использованием ипотечных кредитов увеличилась в той же мере. Выручка от сделок с ипотекой увеличилась в 2020 году на 43,8%, количество реализованных помещений – на 20,3%:

Продажи помещений по наличию обременений в стоимостном выражении

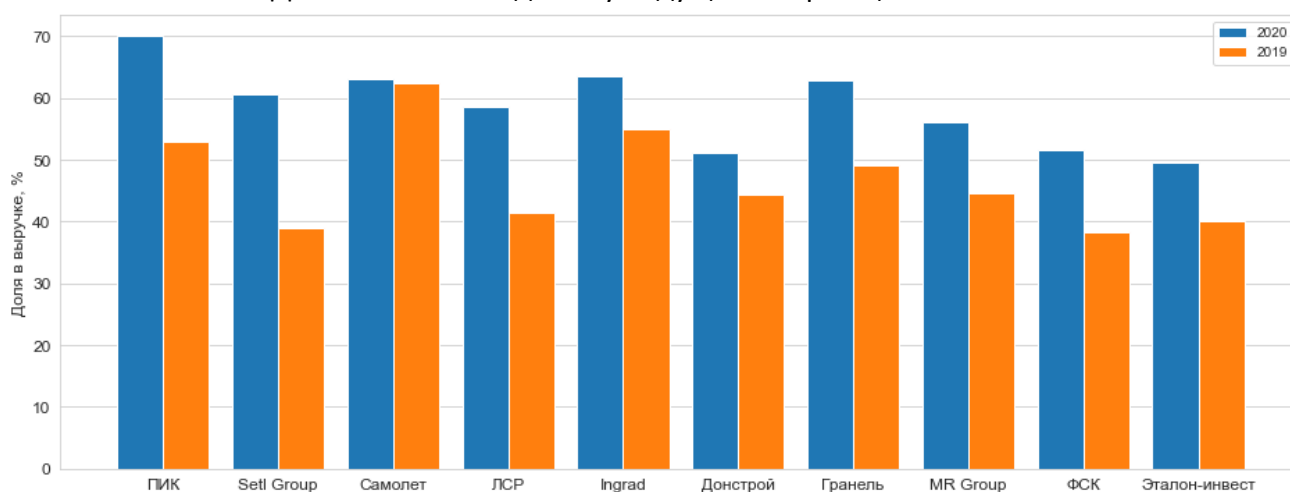


Продажи помещений по наличию обременений в натуральном выражении



Популярность ипотечных сделок хорошо демонстрируют данные по ведущим застройщикам 2020 года:

Доля ипотечных сделок у ведущих застройщиков 2020 г.



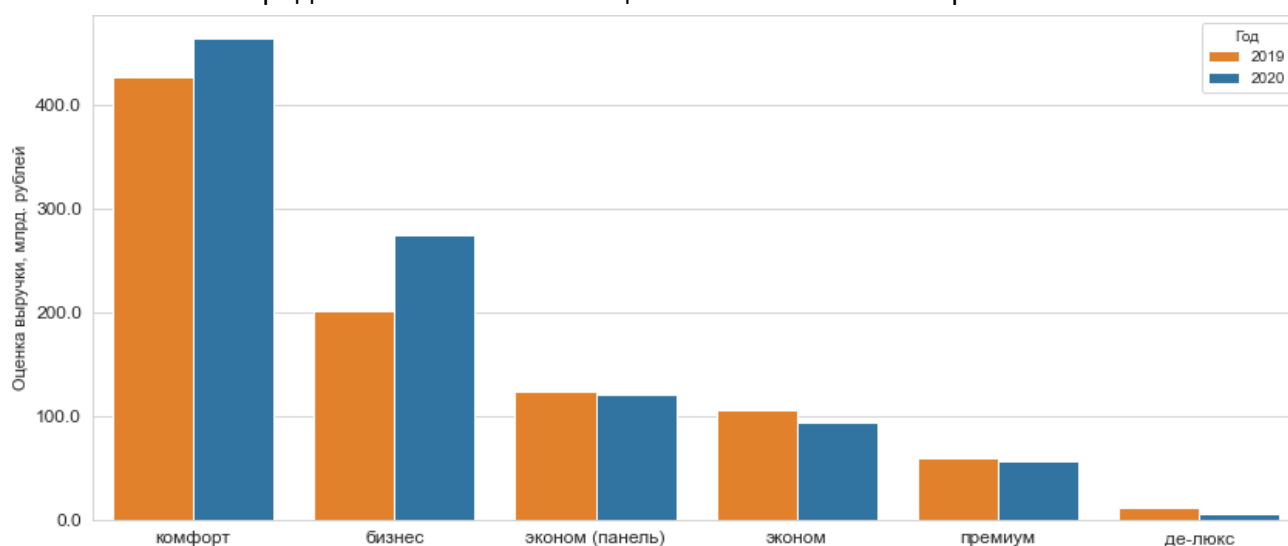
Практически у всех застройщиков, являющихся лидерами по сумме продаж квартир и апартаментов, доля сделок с ипотекой в 2020 году составила не менее 50%. У ПИКа, возглавляющего рейтинг по продажам, доля таких сделок в объеме реализации достигла 70%.

По сравнению с 2019 годом по доле ипотечных сделок в выручке наибольший прирост наблюдался у Setl Group (+21,6 п.п.), а также у ПИК и ЛСР (+17,6 п.п.). Лишь у ГК Самолёт доля сделок с ипотекой сохранилась на уровне 2019 года (63%), но и это значение позволяет входить застройщику в тройку лидеров по этому показателю.

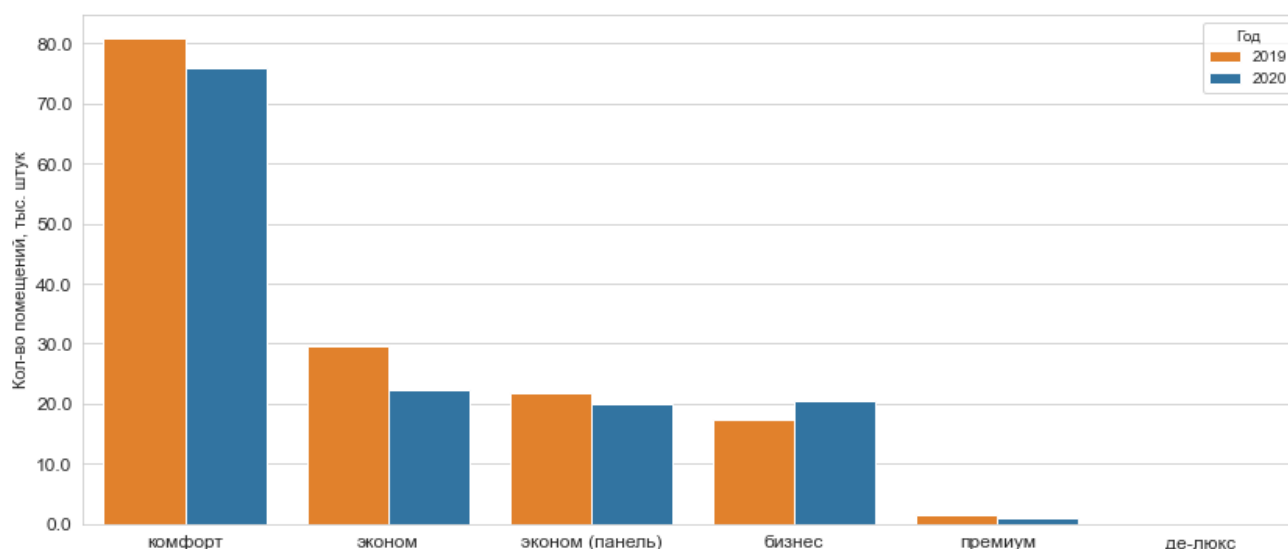
3. Спрос на объекты комфорт и бизнес-класса увеличился в большей степени, чем на объекты других классов.

Спрос на помещения указанных двух классов, действительно, в 2020 году вырос в отличие от других классов:

Продажи по классам помещений в стоимостном выражении



Продажи по классам помещений в натуральном выражении



Наиболее высокий прирост продаж наблюдается в бизнес-классе как по количеству реализованных помещений (+18,1%), так и по выручке (+36,4%). И если в натуральном выражении продажи помещений бизнес-класса уступают эконо- и комфорт-классу, то в стоимостном выражении они уступают только лидеру.

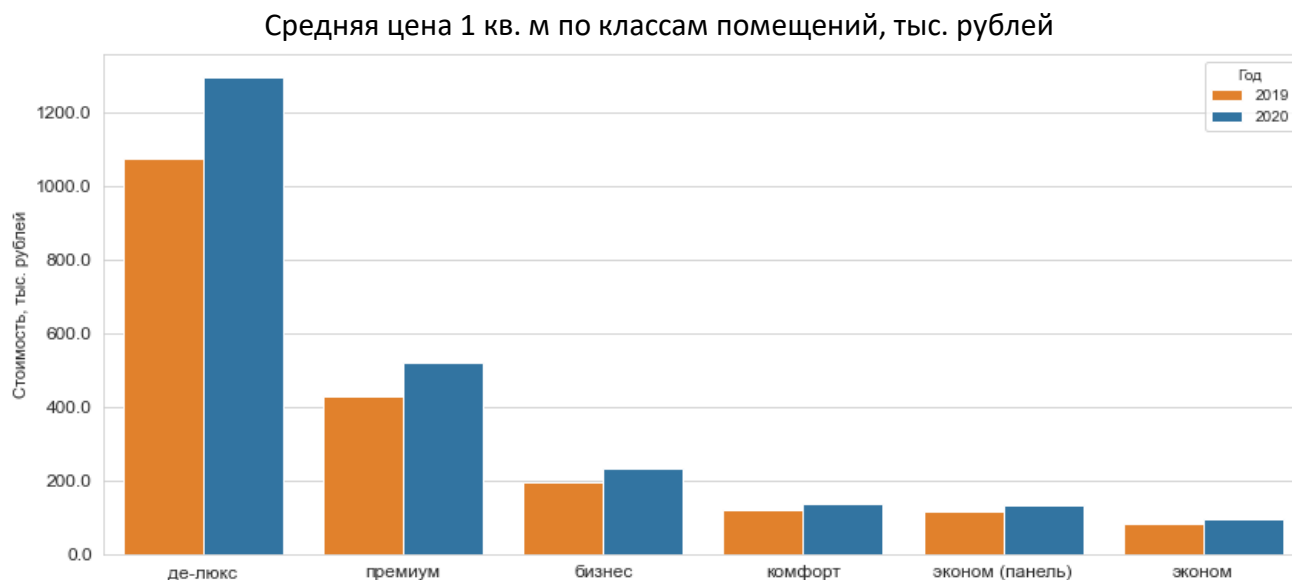
Лидером же по объему продаж на протяжении двух лет остается комфорт-класс. В 2020 году было реализовано на 6,1% меньше помещений данного класса, а выручка от их продаж увеличилась на 8,6%.

В остальных классах продажи в 2020 году сократились. Наиболее значительное снижение продаж произошло в классе делюкс (-49,3% по выручке), наименьшее (-2,3%) – в классе эконом, для квартир и апартаментов в домах, построенных по панельной технологии.

4. Средняя стоимость квадратного метра площади помещений увеличилась.

Данная гипотеза также подтвердилась. Средняя стоимость единицы площади в 2020 году увеличилась на 17,9%, с 124 тыс. рублей до 146 тыс. рублей.

Причем, рост характерен для всех классов помещений:



Данная диаграмма демонстрирует еще одну особенность: рост стоимости тем выше, чем выше класс помещений. Средняя стоимость 1 кв. м площади помещений класса делюкс и премиум увеличилась в 2020 году на 21%, а класса эконом – на 14%.

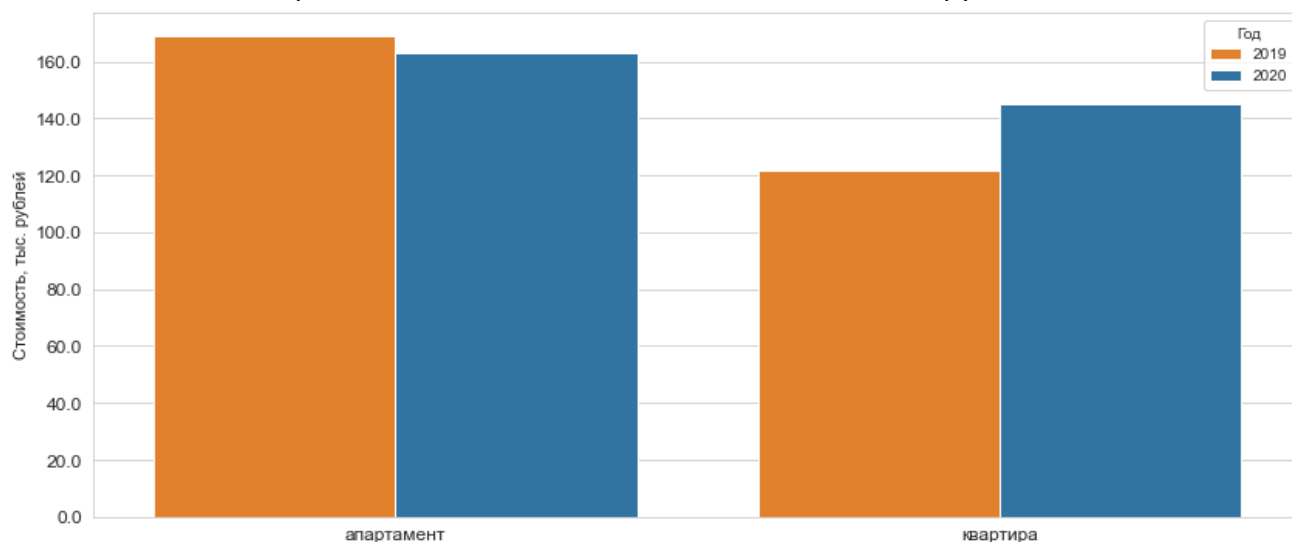
Таким образом, рост стоимости единицы площади в 2020 году для помещений эконом- и комфорт класса был ниже, чем по рынку, для помещений бизнес-класса соответствовал рыночным показателям, а для помещений более высоких классов превышал их.

Если обратиться к графику, характеризующему изменение этого показателя по месяцам анализируемого двухлетнего периода, то обнаружим, что во второй половине 2020 года характер изменения кривой согласуется с динамикой изменения спроса (см. выше). Рост продаж с началом действия программы льготной ипотеки повлек за собой и рост стоимости 1 кв. метра, который наблюдался с июня по октябрь 2020 года.



В то же время динамика средней стоимости единицы площади для квартир и апартаментов различается:

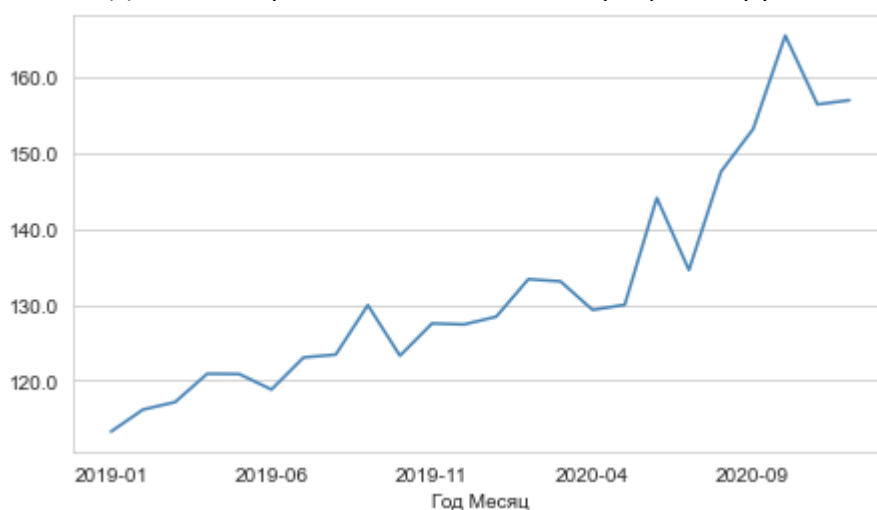
Средняя цена 1 кв. м по типам помещений, тыс. рублей



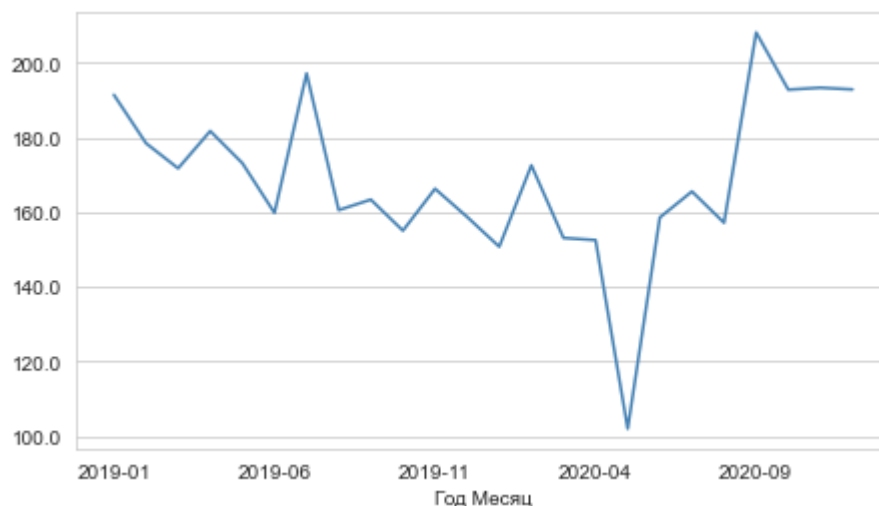
В 2020 году по сравнению с предыдущим годом стоимость 1 кв. м площади квартир увеличилась на 19,2%, с 122 тыс. рублей до 145 тыс. рублей. Аналогичный показатель для апартаментов уменьшился на 3,4%, с 169 тыс. рублей до 163 тыс. рублей.

На протяжении всего анализируемого периода стоимость единицы площади квартир росла, тогда как стоимость 1 кв. метра площади апартаментов до мая 2020 года снижалась и, видимо, только действие программы льготного кредитования изменило характер динамики этого показателя:

Динамика средней цены 1 кв. м квартир, тыс. рублей



Динамика средней цены 1 кв. м апартаментов, тыс. рублей



5. Значимость (веса) переменных, влияющих на стоимость единицы площади, в 2020 году изменилась.

В ходе проверки данной гипотезы сначала для количественных переменных были визуализированы возможные корреляции целевого признака с остальными. Самая выраженная взаимосвязь со стоимостью 1 кв. метра оказалась у площади помещения, менее выраженная - с этажом, оценкой потребительских характеристик ЖК (ЕРЗ).

Для того, чтобы оценить эти взаимосвязи количественно, были построены распределения переменных. Поскольку для всех переменных распределение оказалось отличным от нормального, для количественной оценки взаимосвязи были рассчитаны коэффициенты Спирмена.

Согласно этим расчетам, максимально коррелирующие с ценой 1 кв. метра площади переменные:

- Оценка ЕРЗ ЖК (значение коэффициента корреляции 0,4) - чем выше оцениваются потребительские качества ЖК, тем выше стоимость 1 кв.м
- Широта (-0,23) - в Москве (lat = 55) стоимость выше, чем в С.-Петербурге (lat = 59,57)
- Ставка по ипотеке для первичного рынка (-0,21) - меньшим значениям ставки соответствует большее значение стоимости 1 кв. м
- Год (0,19) - в 2020 г. стоимость 1 кв. м выше, чем в 2019 г.

Остальные:

- Месяц, кол-во дней с начала года (0,15)
- Этаж (0,14)
- Площадь (0,12)

Далее массив данных был подготовлен для машинного обучения:

- . категориальные переменные были преобразованы в числовые (pd.get_dummies())
- . разделены целевая переменная и остальные
- . выделены обучающая и отложенная части массива
- . данные масштабированы (StandardScaler())

И была решена задача множественной регрессии с использованием двух моделей (DecisionTreeRegressor для сравнения). Модели, а также значения метрик оценивания моделей приведены в таблице:

Модель	Среднеквадратическая ошибка (RMSE), рублей	Средняя абсолютная погрешность регрессии MAE, рублей	Коэффициент детерминации (R^2 score)
LinearRegression	27 737	15 579	0,85
DecisionTreeRegressor	16 735	6 537	0,94

Для модели LinearRegression были получены интересные нас коэффициенты модели (коэффициенты регрессии), представляющие собой веса анализируемых факторов. Каждый коэффициент представляет собой среднее изменение стоимости 1 кв. метра площади помещения при изменении соответствующего показателя на единицу с условием, что все остальные показатели остаются неизменными. Так, например, см. ниже фрагмент изображения со списком коэффициентов модели, при увеличении оценки ЕРЗ ЖК на единицу стоимость 1 кв. метра помещения увеличится на 2347,7 рублей:

Дней с начала года	7703.9
Год	7399.6
Ing	7297.8
Оценка ЕРЗ ЖК	2347.7
Площадь	2028.9
Этаж	1897.8
Рейтинг застройщика ЕРЗ	-649.9
is_buyer_ul	-1827.4
Ипотечная ставка, %	-1942.5
Месяц	-2285.8
Уступка	-2871.1
Кол-во комнат	-8173.6

Далее, чтобы проверить гипотезу разного веса факторов стоимости 1 кв. метра в 2019 и в 2020 годах массив был разделен по годам и для каждой из частей повторена процедура регрессионного анализа. В результате были получены коэффициенты модели, сравнение которых показывает, что веса факторов, действительно, разные (гипотеза подтвердилась).

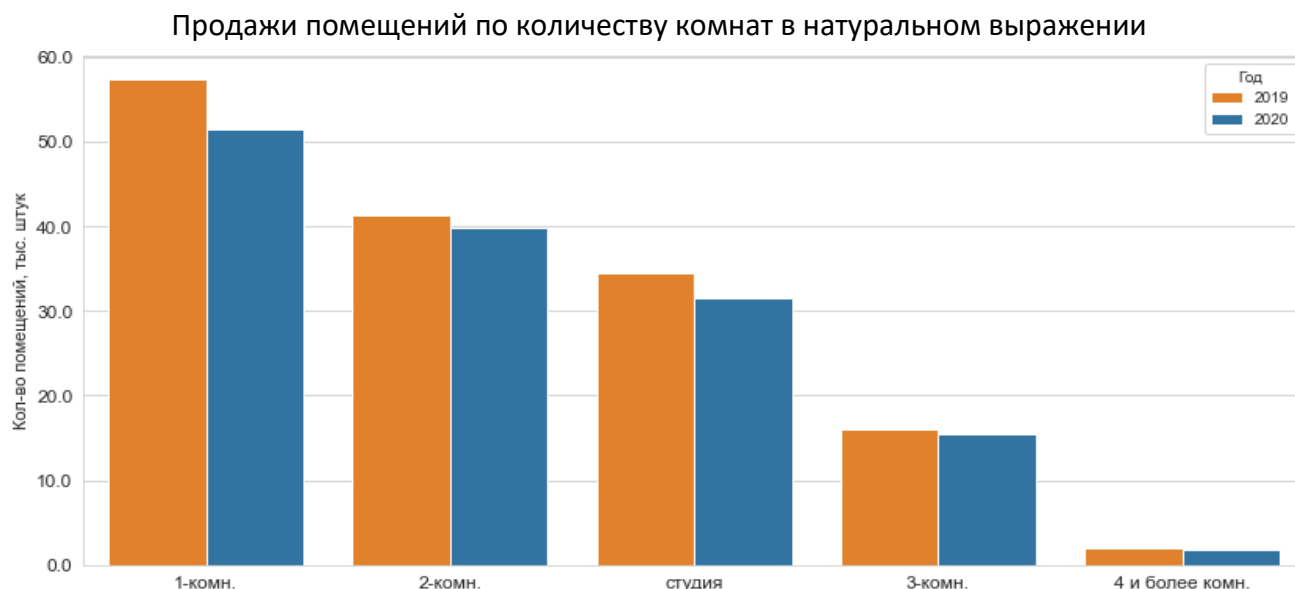
Для примера ниже приведены фрагменты списков коэффициентов модели, которые упорядочены по убыванию (первые 15 значений):

2020 год		2019 год	
Регион_Московская область	91889690173334880.0	Регион_Новая Москва	79031827668029296.0
Обременения_ипотека	83929508355877216.0	Тип помещения_апартамент	62361510168591352.0
Обременения_отсутствуют	83564254337838640.0	Тип помещения_квартира	62361510168589992.0
Категория застройщика_high	50248180721540880.0	Зона_3) от ~ТТК до МКАД	53792070228465144.0
Зона_5) Н. Москва ближ.	48595667974558296.0	Зона_2) СПб - КАД	45868885862712904.0
Категория застройщика_middle	43326408582809024.0	класс_комфорт	41484266010607816.0
Тип помещения_квартира	41192759119285512.0	класс_эконом	32988439287567876.0
Тип помещения_апартамент	41192759119285104.0	Тип оценки_Оценка по среднему	31222169370446164.0
Категория застройщика_lower-middle	36361538292094336.0	Тип оценки_Оценка по номеру	31167109246055380.0
Категория застройщика_upper-middle	36124720047280992.0	класс_эконом (панель)	29263175849483204.0
Зона_6) Н. Москва дальн.	26067038836318800.0	класс_бизнес	26519428119409788.0
Регион_Москва	18806596892777436.0	Зона_4) Москва за МКАД	25739322256140332.0
Обременения_зalog	17343906940503296.0	Зона_3) СПб - КАД+	24784716124617976.0
Категория застройщика_low	16578607270055150.0	Зона_2) от СК до ~ТТК	23970300199289208.0
Обременения_военная ипотека	9181637323958720.0	Зона_1) СПб - центр	18225926456104448.0

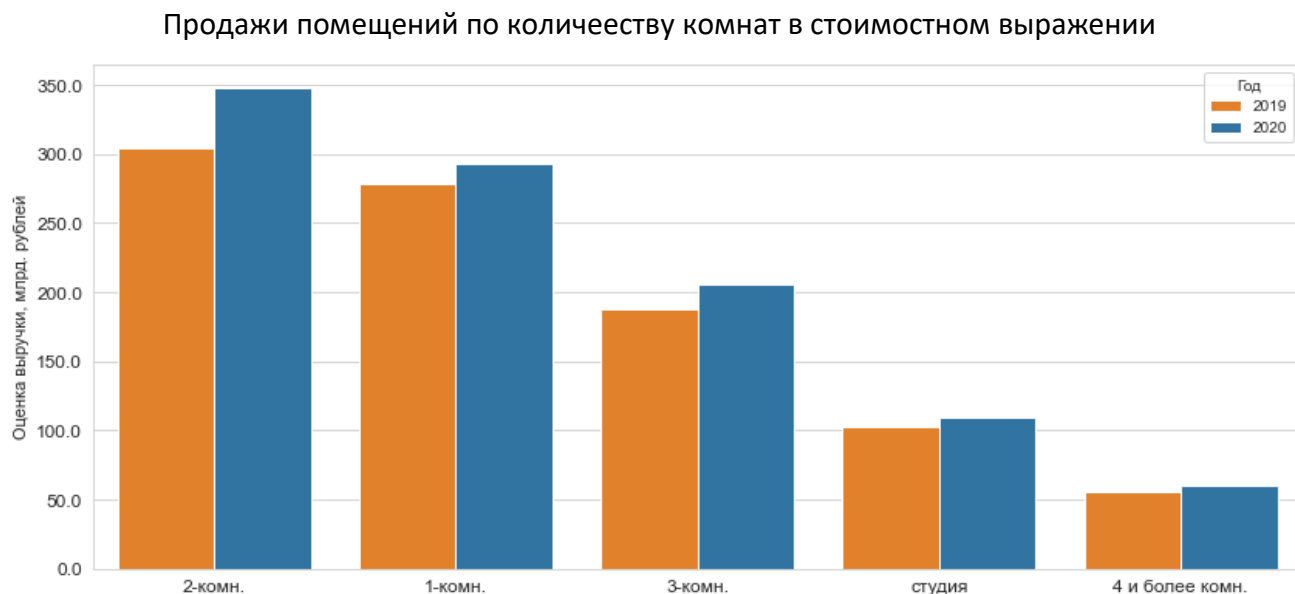
Из остальных результатов анализа данных остановимся только на характеристиках реализуемых помещений (количество комнат, площадь, этаж) и данных по участникам рынка – застройщикам и залогодержателям.

6. Характеристики помещений.

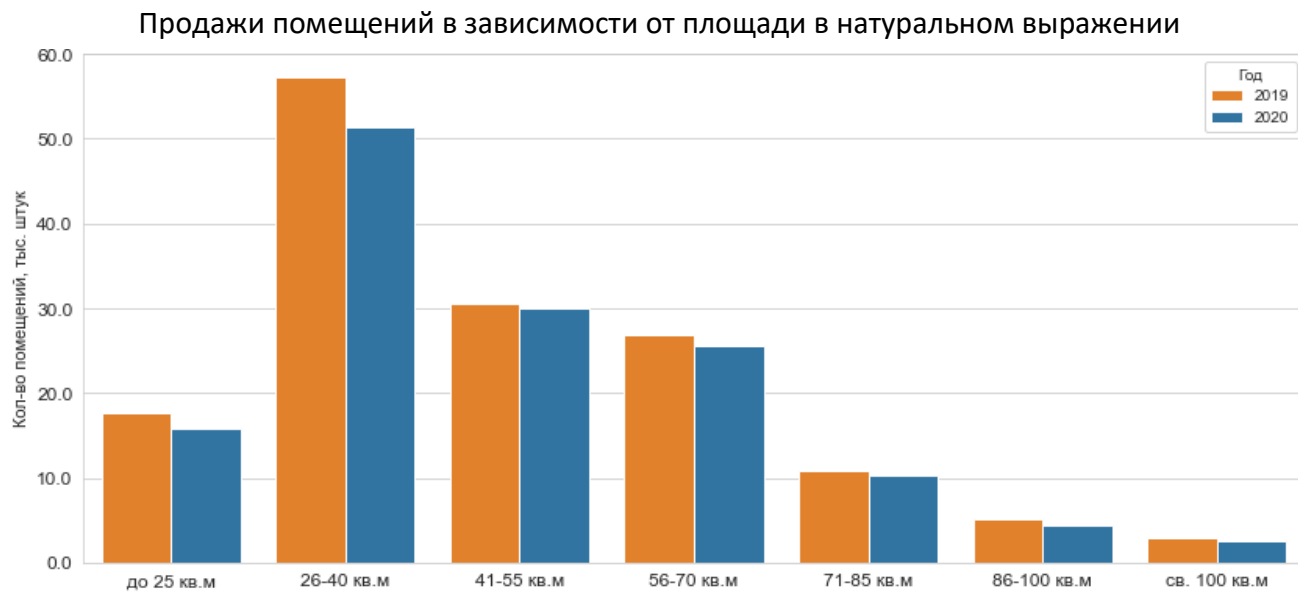
На протяжении анализируемого периода предпочтения покупателей квартир и апартаментов не меняются: больше всего продается однокомнатных помещений. В 2020 году всех типов помещений было продано меньше, чем годом ранее:



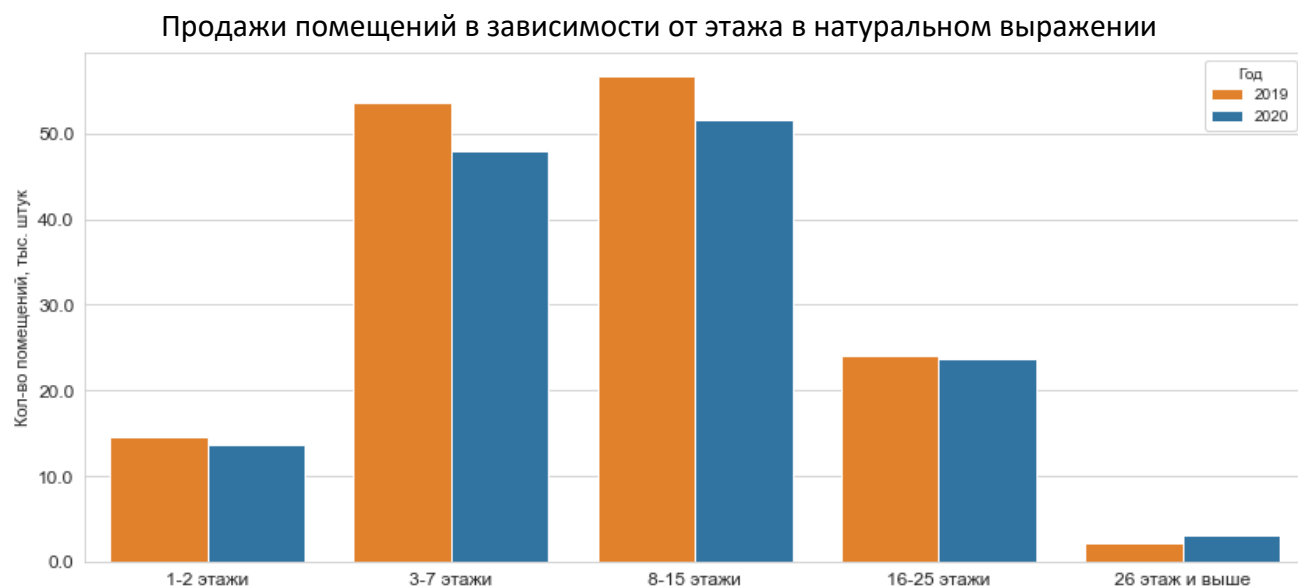
По выручке от продаж первенство принадлежит 2-ух комнатным помещениям, прирост этого показателя для них также наибольший:



Наибольшим спросом пользуются помещения площадью от 26 до 45 кв. м:



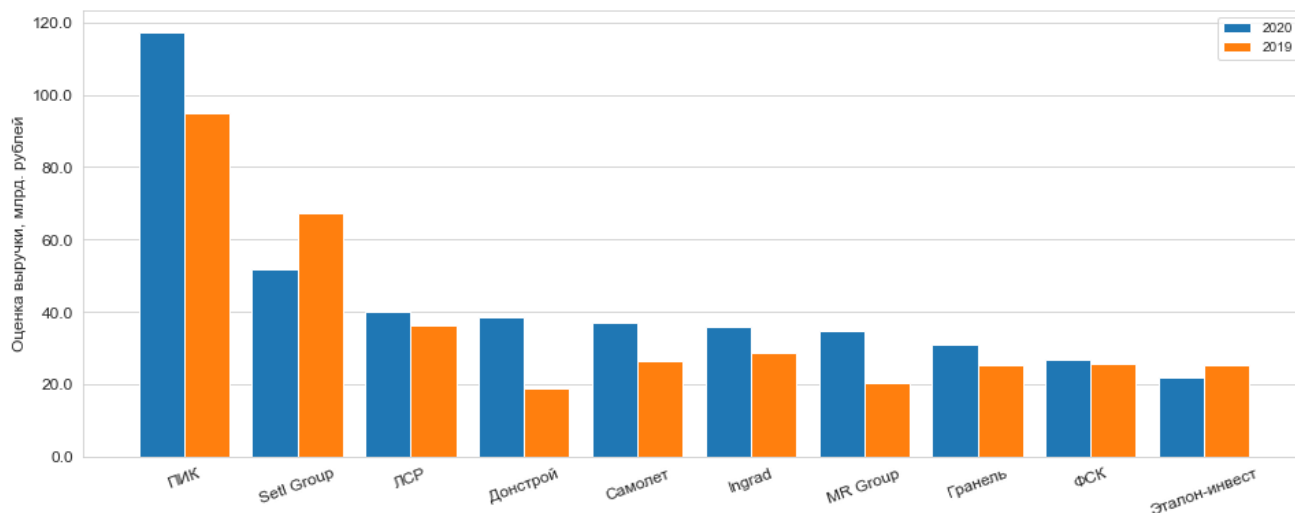
Предпочтения по этажам в течение 2019-2020 г.г. не меняются – это 8-15 и 3-7 этажи:



7. Топ застройщиков по объему продаж. Средние значения показателей площади проданных помещений, стоимости сделок, стоимости единицы площади.

В топ-10 по показателю выручки в 2020 году входят следующие застройщики:

Рейтинг ведущих застройщиков 2020 г. по выручке

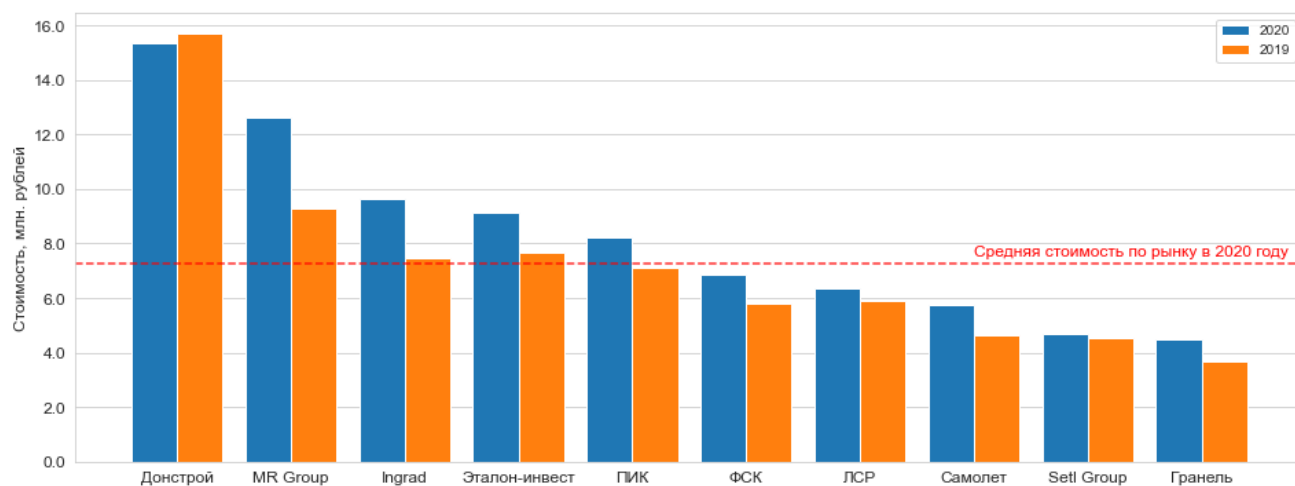


С большим преимуществом, которое в 2020 году увеличилось, лидирует ПИК. По сравнению с 2019 годом в топ-10 вошли два новых участника рынка строительства: MR Group и Донстрой. На протяжении двух лет в число ведущих застройщиков входят: ПИК, Setl Group, ЛСР, Самолет, Ingrad, Гранель, ФСК и Эталон-инвест.

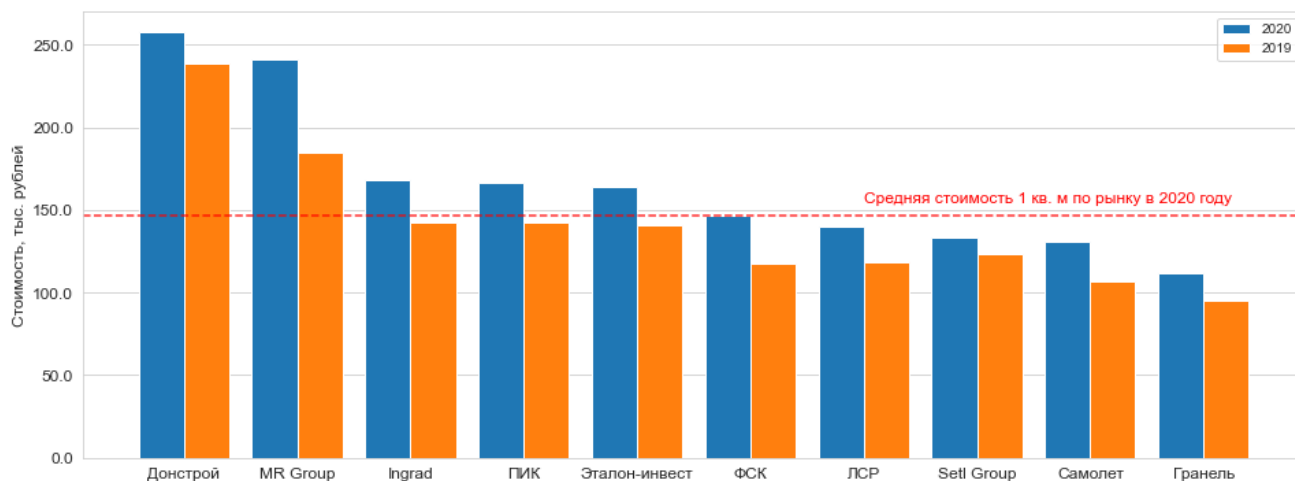
Доля 10-ти ведущих застройщиков в продажах за год в 2020 году увеличилась с 41% до 43%.

Ниже представлены диаграммы с показателями деятельности ведущих застройщиков 2020 года. Для сравнения указаны средние значения.

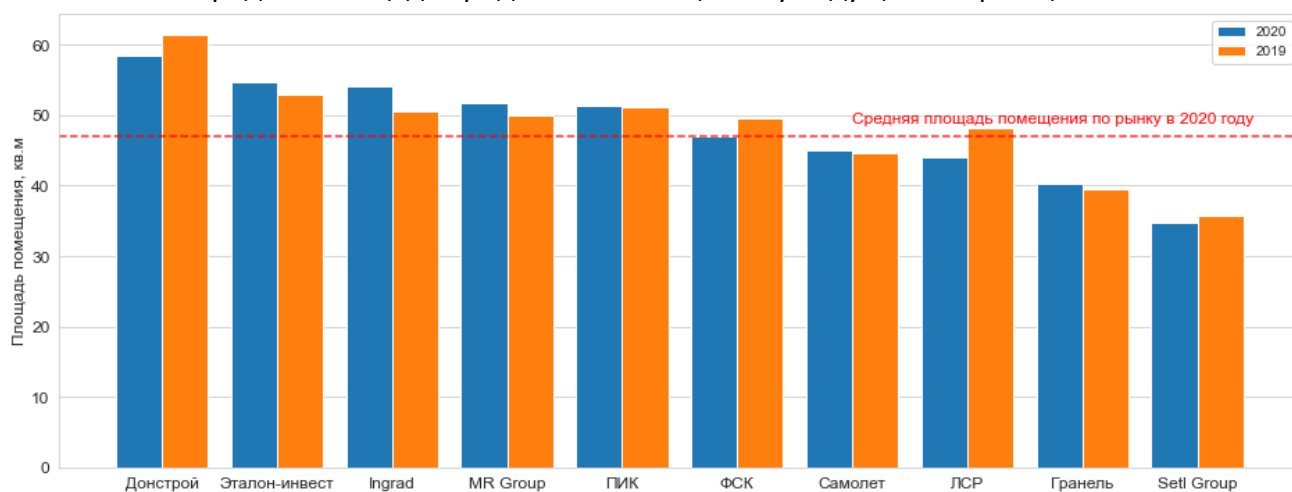
Средняя стоимость проданных помещений у ведущих застройщиков



Средняя стоимость 1 кв. метра помещений у ведущих застройщиков



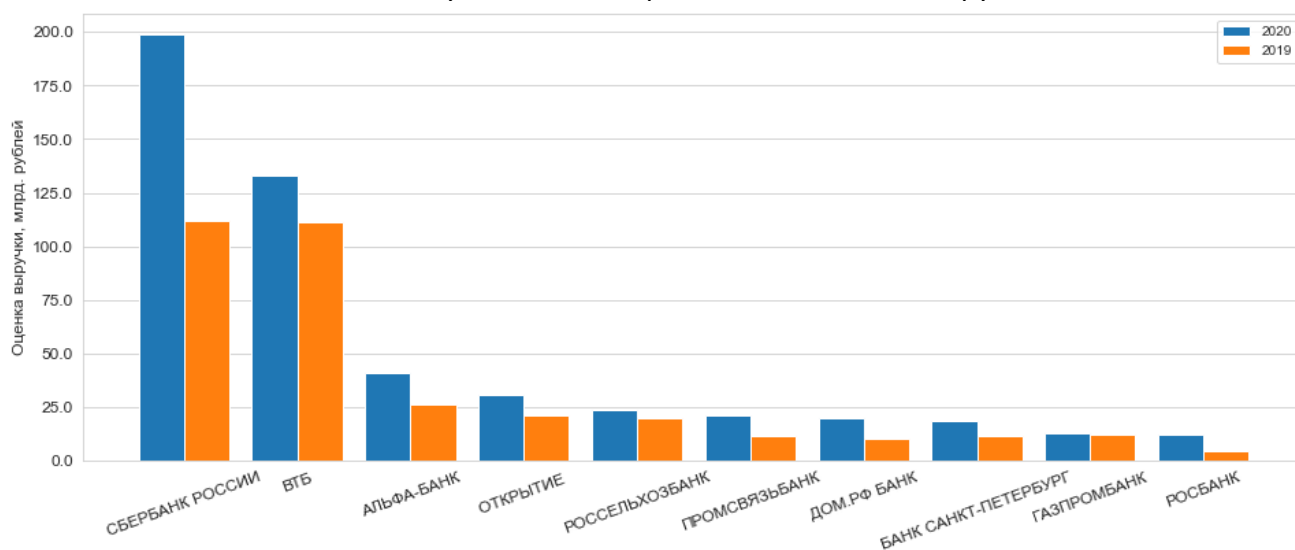
Средняя площадь проданных помещений у ведущих застройщиков



8. Топ залогодержателей.

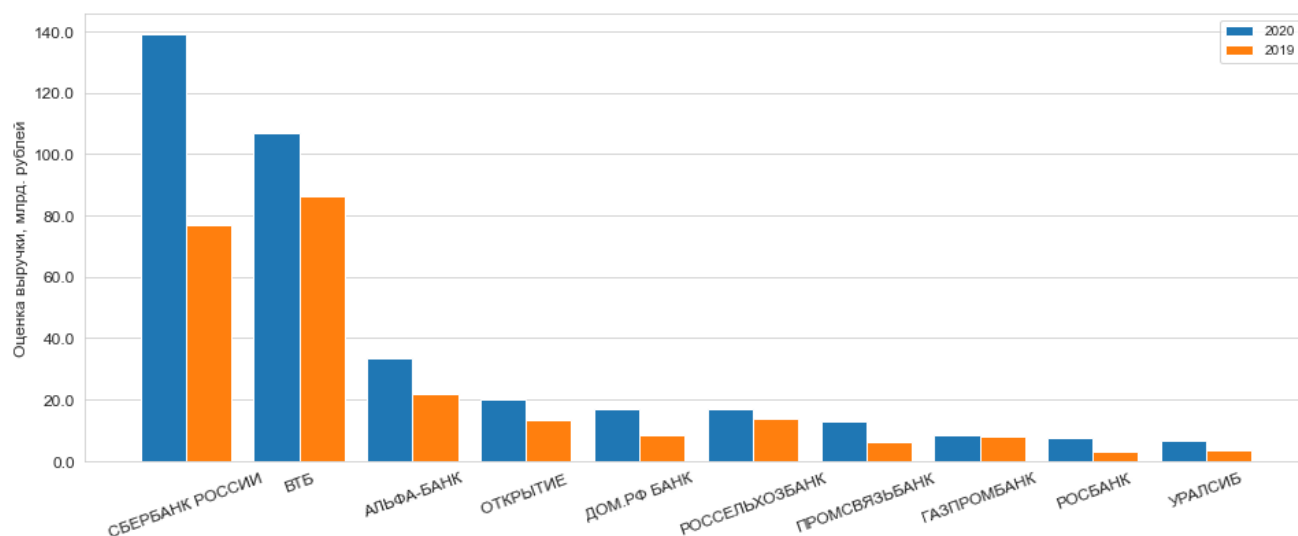
В топ-10 по показателю выручки в 2020 году входят следующие залогодержатели:

Рейтинг ведущих залогодержателей 2020 г. по выручке

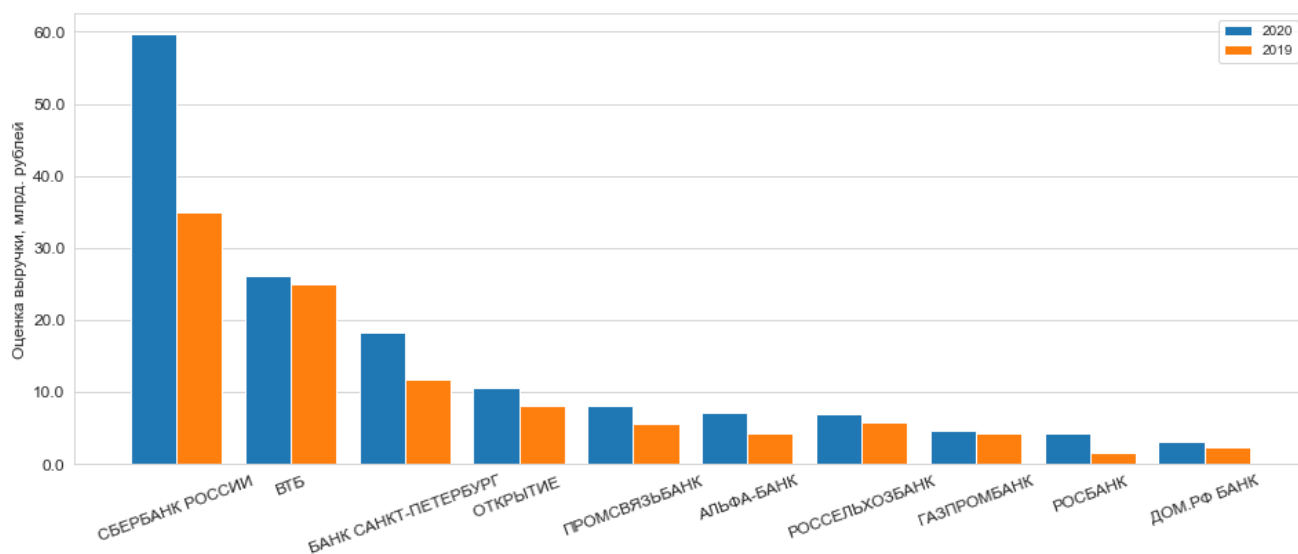


Их список практически не меняется в зависимости от региона:

Рейтинг залогодержателей по выручке в 2020 г.
Москва и Московская область



Рейтинг залогодержателей по выручке в 2020 г.
Санкт-Петербург и Ленинградская область



Выводы и заключение

По результатам выполнения данной работы можно сделать следующие выводы:

1. Используемый источник исходных данных позволяет решать поставленные задачи по анализу первичного рынка объектов недвижимости (по крайней мере, по квартирам и апартаментам).
2. Данные, полученные из источника, по отдельным полям содержат неверные значения и пропуски, но их количество относительно небольшое и при очистке данных их качество может быть улучшено.
3. Имеющиеся данные позволяют осуществлять не только ретроспективный анализ, но и решать задачи прогнозирования основных показателей рынка (выручки, количества реализуемых помещений, стоимости помещений, стоимости единицы площади).
4. Возможности обогащения данных новыми признаками в данной работе реализованы не в полной мере. Для этого требуется более основательный поиск новых источников и решение задачи объединения данных по таким полям как наименование застройщика (или другим его идентификаторам), наименование ЖК, географические координаты местоположения, наименование залогодержателя.
5. Решение задачи регрессионного анализа, вероятно, предпочтительнее осуществлять отдельно для каждого региона и с учетом класса помещений.
6. Результаты подобного анализа данных по первичному рынку недвижимости могут быть востребованы широким кругом заинтересованных лиц.

Приложения

Спецификация.

Наименование поля	Содержание
Файл	Файл - источник данных
ID ЖК	Идентификатор ЖК
ЖК рус	Название ЖК
ЖК англ	Альтернативное название ЖК
Район Город	Наименование внутригородского муниципального образования
Округ Направление	Наименование округа или направления Актуально для Московской области, для остальных регионов совпадает с АТД
Регион	Москва, Новая Москва (Троицкий и Новомосковский округа), Московская область, Санкт-Петербург, Ленинградская область
АТД	Наименование единицы административно-территориального деления
Застройщик ЖК	Девелопер (брендовое название, если несколько участников - то один из (в т.ч., если проект реализуется по схеме фи-девелопмента)
Описание помещения	Описание помещения (как в выписке ЕГРП\ЕГРН)
Площадь	Площадь помещения. 0 или пусто - площадь неизвестна.
Комнатность	Число комнат в помещении как указано в Описании помещения в выписке (если параметр комнатность присутствует в описании в ЕГРП\ЕГРН). 0 или пусто - комнатность неизвестна (не указана)
Тип Комнатности	Комнатность, определенная по алгоритму, в том числе если не указана в регистрационной записи: ст - студии, 1- одно-, 2- двух-, 3- трех-, 4 - многокомнатные 0 или пусто - комнатность не удалось определить алгоритмом или не соответствует минимальным и максимальным ограничениям по площади
Этаж	Этаж помещения (если параметр номера этажа присутствует в описании). Если ниже уровня земли, то этаж = 0
Номер	Номер помещения (если параметр номера помещения присутствует в описании), распознавание из текстового описания по алгоритму
Тип помещения	Квартира, машиноместо, апартаменты, кладовки или нежилое помещение - распознавание по алгоритму Апартаменты - нежилые помещения в ЖК, корпусах с апартаментами, выше 2-го этажа и более 15 кв. м
Корпус	Название корпуса
Покупатель ЮЛ	Наименование юр. лица
Покупатель ФЛ	Идентификатор физ. лица
Номер регистрации	Согласно выписке ЕГРП\ЕГРН На одном номере регистрации может быть как одно помещение, так и множество (оптовые сделки, мультисделки - например квартира + машиноместо)
Дата регистрации	Дата регистрации ДДУ или уступки ДДУ
Условия обременения	Условия обременения
Номер обременения	Номер обременения согласно выписке ЕГРП\ЕГРН
Дата обременения	Дата регистрации обременения согласно выписке ЕГРП\ЕГРН
Залогодержатель	Залогодержатель (банк или др. организация)
Длительность обременения	Длительность обременения (месяцев).

Наименование поля	Содержание
	0 или пусто - рассчитать длительность не удалось или обременения нет.
Тип обременения	Ипотека отсутствует (пусто или 0), ипотека, военная ипотека, залог (распознавание по алгоритму) Залог – обременения, наложенные не банком или ипотечным агентством, а иными организациями, а также при сделках между юр. лицами.
Оценка цены (Цена по прайс-листу)	Цена квартиры, апартаментов из прайс-листа - поиск по уникальному номеру квартиры, апартаментов в базе предложения на дату, близкую к дате сделки по ДДУ. Если не найдено, то рассчитывается оценочная стоимость помещения по аналогам на основе архива цен. Если нет цены по прайс-листу и оценка не посчитана, то может применяться Цена ДДУ. 0 или пусто - нет информации по ценам.
Дата ДДУ	Максимальная дата в реквизитах ДДУ/УДДУ (при уступках в записи может быть несколько дат)
Уступка	Наличие признака уступки в реквизитах договоров либо смены стороны сделки
Купил лотов в ЖК	Количество помещений любого типа оформленных на физ. лицо или юрлицо в конкретном ЖК
ID Корпус	Присваивается в случае, если распознан корпус
Пол	Пол покупателя - физического лица
класс	Класс (экспертная оценка). При классификации учитывается площадь квартир, квартирография, емкость паркинга, местоположение, число квартир, цена и другие параметры в совокупности, а также позиционирование объекта застройщиком. В одном ЖК могут быть корпуса разных классов (к примеру "эконом (панель)" и "комфорт").
Срок сдачи	Срок сдачи корпуса
Стадия строительства	Стадия строительства корпуса на текущий момент (не на момент продажи)
lat	Широта ЖК-корпуса
lng	Долгота ЖК-корпуса
Ипотека	0, если залог или нет ипотеки, 1 - если ипотека или военная ипотека
Секция	Секция, в которой находится помещение
Отделка	Есть - отделка обязательная во всем корпусе - квартира или апартамент были проданы с отделкой, Нет - отделки нет во всем корпусе - квартира была продана без отделки, Неизвестно - может быть, а может и нет (отделка не во всем корпусе или по желанию покупателя, то есть квартира или апартамент могли быть куплены без отделки)
Старт продаж К	Квартал в котором стартовали продажи корпуса
Продавец ЮЛ	Название юр. лица от которого было помещение по уступке
Продавец ФЛ ID	ID физ. лица от которого было помещение по уступке
Зона	Экспертное деление на зоны доступности
Купил кв и ап в ЖК	Количество квартир и апартаментов оформленных на физ. лицо или юрлицо в конкретном ЖК
Стадия строительства в дату ДДУ	Стадия строительства корпуса на текущий момент (на дату заключения ДДУ)
Цена ДДУ	Цена указанная в договоре ДДУ. Только для договоров застрахованных в фонде дольщиков в 2018-2019 годах. Данные после сентября 2019 года недоступны.

Наименование поля	Содержание
Цена ДДУ Есть/Нет	Наличие цены ДДУ
Тип сделки	Принимает значения ЗФ - сделка между застройщиком и физ. лицом (ДДУ) ЗЮ - сделка между застройщиком и юр. лицом (ДДУ) ФФ – сделка (уступка) между физ. лицами ЮФ – сделка (уступка) между юр. лицом и физ. лицом ФЮ – сделка (уступка) между физ. лицом и юр. лицом ЮЮ - сделка (уступка) между юр. лицами НФ – сделка (уступка) с физ. лицом - вторая сторона сделки неизвестна НЮ – сделка (уступка) с юр. лицом - вторая сторона сделки неизвестна ЗП - сделка между застройщиком и ПИФ (ДДУ) ПП – сделка (уступка) между ПИФ и ПИФ ПФ - уступка от ПИФ на физ. лицо ПЮ - уступка от ПИФ на юр. лицо ЮП - уступка от юр. лица на ПИФ ФП - уступка от физ. лица на ПИФ
Цена со скидкой	Стоимость помещения на основе архива цен предложения (по данным, где есть информация о скидке, заданной явным образом в привязке к помещению)
Тип оценки	Алгоритм, использованный для расчета оценки цены
Разница дат	Разница дат договора ДДУ и даты регистрации в выписке
Дата регистрации модель	Корректировка "бракованных" дат регистрации Равна дате регистрации, кроме случаев, когда при вторичной и т.д. сделке по помещению дата регистрации в выписке осталась от первичной сделки. В этом случае дата регистрации модель = дата ДДУ + 14 дней.
version	Тип исходных данных: 1 - данные из выписок нового формата (с лета 2020 г.), 0 - иное
Участок	Кадастровый номер участка по которому получены выписка
is_buyer_ul	признак покупателя: 1 - покупатель ЮЛ, 0 - покупатель ФЛ

Основные решения, связанные с очисткой данных.

Наименование переменной	Количество и доля отсутствующих значений	Решения по восстановлению недостающих значений	Решения после анализа экстремальных значений
Площадь помещения	2 (0%)	Средняя площадь квартир в том же корпусе ЖК и с тем же количеством комнат	Минимальное значение (2,7) указано неверно, заменено на правильное, из описания помещения. Максимальное значение (3135,3) соответствует коммерческому помещению, тип апартаменты был указан неверно. Исключено. Объекты с "подозрительно" большой площадью (>300 кв. м), в описаниях которых не указано, что это квартиры или апартаменты. После проверки исключены. Остальные подобные помещения, в основном, относятся к классу комфорт и выше.
Количество комнат Тип комнатности Комнатность	2951 (1%) 160161 (54%)	Использование данных из двух столбцов.	Значения количества комнат от 5 до 8 соответствуют описаниям помещений.
Этаж	-	-	Отрицательные значения исключены (описание помещений не позволили принять другое решение). Нулевые значения соответствуют, как правило, мансардным этажам в описаниях помещений. Для таких помещений значения заменены на максимальные значения этажей в этих ЖК. Остальные записи (в описании нет указания на мансарду) исключены. Для очень больших значений этажей в каждом регионе, по данным Единого ресурса застройщиков (ЕРЗ, erzrf.ru), определена этажность самых высоких домов с жилыми единицами в 2020 г. Строки, содержащие значения этажей, превышающие их, исключены.
Оценка цены	1652 (0,6%)	Записи с отсутствующими значениями исключены.	-