Государственное бюджетное профессиональное

образовательное учреждение Московской области

«Физико-технический колледж»

**Аналитический отчет**

Работу выполнил:

Студент группы ИСП-21

Минин Матвей

Долгопрудный, 2024

**ВВЕДЕНИЕ**

**Цель отчета**: собрать и проанализировать данные используя различные инструменты, преобразовать их для дальнейшей работы.

**Важность**: cектор недвижимости играет значительную роль в экономике, и осознание факторов, которые влияют на стоимость, способствует более точной оценке объектов и предсказанию рыночных тенденций.

**Задачи:**

* Используя различные инструменты из интернета собрать данные.
* После сбора данных отредактировать и отфильтровать лишнее, подготовить данные для дальнейших действий
* Произвести анализ данных, найти различные взаимосвязи параметров, построить графики.

### СБОР И ПОДГОТОВКА ДАННЫХ

1. **Сбор данных**: для сбора данных была предложена библиотека cianparser, я попробовал собрать данные при помощи неё, но был неудовлетворен её скоростью работы, поэтому из также предложенного варианта был выбран domclick-parser и в последствии модифицирован для записи данных в таблицу

Изображение выглядит как текст, снимок экрана, Шрифт

Автоматически созданное описание

1. **Просмотр полученных данных:**

**Изображение выглядит как текст, снимок экрана, Шрифт, информация

Автоматически созданное описание**

# выводим количество данных

print(df.shape)

(11273, 1)

# выводим типы данных

print(df.dtypes)

location,floor,floor\_count,rooms\_count,square\_meter\_price,price,year\_of\_construction,ipoteka\_rate,area,discount,distance,monthly\_payment object dtype: object

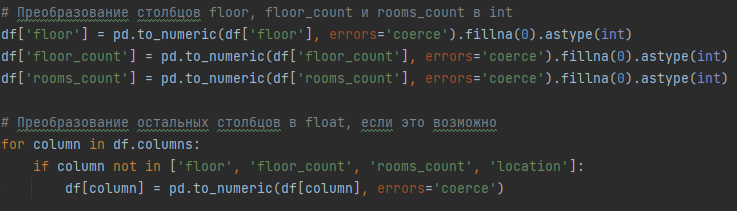
1. **Предобработка данных в Python**:
2. В столбце location были удалены начальные значения (Россия, Московская область), так как во всех столбцах они одинаковы.

# Очищаем данные в столбце "location"  
df['location'] = df['location'].str.replace(r'^Россия, Московская область,\s\*', '', regex=True)

1. Удалены столбцы “ipoteka\_rate” и “discount” информация содержащаяся в них не несет полезной нагрузки, также удален столбец с годом постройки так как его заполненность ниже 20%.

# Удаляем столбцы 'ipoteka\_rate' и 'discount' и 'year\_of\_construction'  
df.drop(columns=['ipoteka\_rate', 'discount', 'year\_of\_construction'], inplace=True)

1. Изменен тип переменных “floor”, “floors\_count”, “rooms\_count” на int, остальные переменные так как у них большие значения были конвертированы в float.



1. Удаляем строки в которых есть подозрительные значения

в floor\_count это одноэтажные и двухэтажные дома, так как нам нужны квартиры, а не дачи, а также в room удаляем строки с значением 0 так как в квартире не может быть 0 комнат.

Изображение выглядит как текст, снимок экрана, Шрифт

Автоматически созданное описание

1. Заполнил недостающие значения в столбце monthly\_payment (ежемесячная плата по ипотеке) модой, а также столбец distance (расстояние до метро) заполнил средним значением.

Изображение выглядит как текст, снимок экрана, Шрифт, программное обеспечение

Автоматически созданное описание

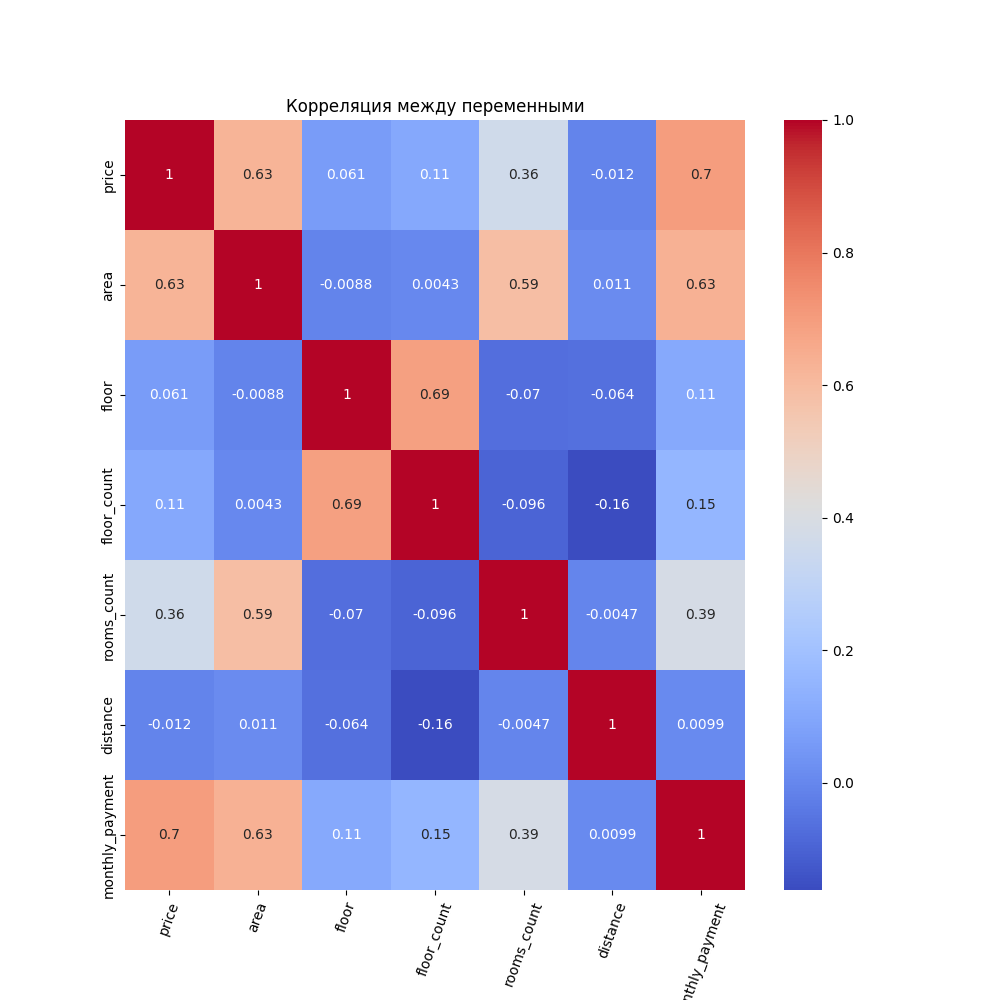
1. И конечно же не стоит забывать удалить дубликаты.

Изображение выглядит как текст, Шрифт, снимок экрана, Графика

Автоматически созданное описание

**Анализ данных**

1. Корреляционный анализ:
   * С помощью Python и библиотеки pandas будет рассчитан коэффициент корреляции между переменными, такими как общая площадь (area), этажность (floor, floor\_count), количество комнат (rooms\_count), дистанция до метро (distance) и цена ипотечной платы в месяц (monthly\_payment) чтобы определить, какие факторы имеют наибольшее влияние на стоимость недвижимости.



На основе этой корреляционной матрицы можно сделать несколько выводов:

1. **Общая площадь (area):**
   * Это наиболее значимый фактор, влияющий на цену корреляция(0,69)
2. **Этажность (floor):**
   * Имеет очень слабую корреляцию всего (0.061) что странно но может быть скорее всего на это повлияло малое количество квартир на первых этажах по отношению к высоким этажам.
3. **Количество этажей (floor\_count):**

* Более высокий показатель корреляции (0.11) что логично так как многоэтажных домов намного больше.

1. **Количество комнат (rooms\_count):**
   * Корреляция с ценой за квадратный метр составляет (0.36), что указывает на слабую положительную связь. Количество комнат важно учитывать, так как потенциально влияет на спрос.
2. **Площадь кухни (distance):**
   * Опять же как не странно дистанция до ближайшего метро не сыграла особой роли показатель корреляции отрицательный

(-0.012)

1. **Ежемесячная плата за ипотеку (monthly\_payment):**
   * Корреляция (0.7) вполне логично что здесь самая большая корреляция так как этот параметр напрямую зависит от цены

**Общий вывод**: среди всех факторов на цену заметное влияние оказывает **общая площадь(area)** (коэффициент 0.63). Остальные факторы, такие как этажность и количество комнат, также оказывают влияние, но они не так ярко выражено.

1. Анализ зависимостей цены от различных параметров

Изображение выглядит как текст, снимок экрана, дисплей, программное обеспечение

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, График, линия

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, График, линия

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, Шрифт, График

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, линия, число

Автоматически созданное описание

Из графиков в целом можно сделать логичный вывод что основные параметры от которых зависит цена это локация и площадь, а также количество комнат, интересно то что на некоторых графиках есть пики, мы видим в одинцовском районе самые дорогие квартиры, еще есть пик на графике с площадью он обусловлен наличием в данных семикомнатных квартир так же как и на графике rooms\_count.

### ЗАКЛЮЧЕНИЕ

Основываясь на проведённом анализе, можно выделить следующие ключевые выводы:

1. **Количество комнат и площадь**: общая цена квартиры напрямую зависит от количества комнат и площади и прямо пропорциональна им.
2. **Этажность**: анализ по общему количеству этажей и по конкретному этажу на котором расположена квартира показал средние значения однако есть выраженные пики высокой стоимости обусловленные единичными объявлениями с особыми предложениями, в общем цена средне-стабильная.
3. **Локация**: локация играет важную роль как мы можем увидеть в некоторых районах недвижимость имеет очень большую стоимость нежели в других.
4. **Дистанция до ближайшего метро**: имеет плохо выраженные границы и неточности что было понятно еще из матрицы корреляции.
5. **Другие факторы**: несмотря на выявленные зависимости, такие факторы, как наличие инфраструктуры, год постройки, вид из окна, качество отделки, также могут оказывать влияние на цену, однако их необходимо исследовать отдельно, так как их вклад может быть неоднозначным и индивидуально различаться по объектам. К тому же эти данные могли бы участвовать в анализе, однако инструмент сбора данных не позволял сохранить значения в большинстве случаев.

**Итог**: Проведённый анализ показывает, что ключевые факторы, такие как транспортная доступность, количество комнат и локация, играют основную роль в формировании цены на недвижимость.