В данном проекте я хочу проанализировать датасет Airbnb

Для начала считаем наш датасет.

Проведём проверку датасета и выведем первые 5 колонок

Перейдём к этапу очистки данных, удалим ненужные, заполним пропуск

Рассчитаем среднюю и медиану рейтинга и цены на жильё

**Получившиеся данные:**

**Анализ цен:**

Средняя цена: $138.07

Медиана цены: $90.00

Минимальная цена: $0.00

Максимальная цена: $999.00

Распределение цен по процентам:

0.25 55.0

0.50 90.0

0.75 150.0

Name: Price, dtype: float64

**Анализ рейтингов:**

Средний рейтинг: 92.91

Минимальный рейтинг: 20.00

Максимальный рейтинг: 100.00

50% объявлений в диапазоне рейтинга:

От 90.00 до 100.00

**Выводы**

1. Цены на жилье имеют широкий разброс, с преобладанием предложений в среднем ценовом диапазоне.
2. Рейтинги в целом высокие, что указывает на удовлетворенность гостей качеством услуг.
3. Наличие выбросов в ценах требует дополнительного анализа для понимания факторов, влияющих на формирование премиум-сегмента.
4. Высокая концентрация рейтингов в верхнем диапазоне может свидетельствовать о конкурентной среде, где хозяева стремятся поддерживать высокое качество услуг.

**Распределение цен**

График распределения цен показывает, что:

* Большинство предложений сконцентрировано в нижнем ценовом диапазоне (до $200).
* Распределение имеет длинный "хвост" вправо, что указывает на наличие премиум-предложений с высокими ценами.

**Цены по типам жилья**

Боксплот цен по типам жилья демонстрирует:

* "Entire home/apt" обычно дороже других типов, что логично, учитывая больший размер и приватность.
* "Private room" занимает среднюю ценовую категорию.
* "Shared room" является самым доступным вариантом.
* Во всех категориях присутствуют выбросы, особенно в сегменте "Entire home/apt".

**Географический анализ**

Карта с маркерами позволяет увидеть:

* Концентрацию предложений в определенных районах города.
* Возможные ценовые кластеры в разных частях города.

**Зависимость цены от рейтинга**

Диаграмма рассеяния цены и рейтинга показывает:

* Слабую положительную корреляцию между ценой и рейтингом.
* Большинство предложений с высокими рейтингами сконцентрировано в среднем ценовом диапазоне.
* Наличие дорогих предложений с различными рейтингами, что может указывать на то, что высокая цена не всегда гарантирует высокий рейтинг.

**Общие выводы:**

1. Рынок Airbnb в данном датасете разнообразен, с преобладанием предложений целых квартир/домов.
2. Цены сильно варьируются, но большинство предложений находится в доступном ценовом диапазоне.
3. Высокие рейтинги не всегда коррелируют с высокими ценами, что может указывать на хорошее соотношение цена/качество в среднем сегменте.
4. Географическое расположение играет важную роль в формировании цен.
5. Количество отзывов не является определяющим фактором для цены, но может влиять на популярность предложения.

**Для дальнейшего анализа я хочу:**

* Исследовать сезонность цен.