

Проект: Анализ данных агентства недвижимости

Автор: Онищенко Надежда Владимировна

Дата:

Часть 1. Решение задач

Определим аномальные значения (выбросы) по значению перцентилей:

```
WITH limits AS (  
    SELECT  
        PERCENTILE_DISC(0.99) WITHIN GROUP (ORDER BY total_area) AS  
total_area_limit,  
        PERCENTILE_DISC(0.99) WITHIN GROUP (ORDER BY rooms) AS rooms_limit,  
        PERCENTILE_DISC(0.99) WITHIN GROUP (ORDER BY balcony) AS balcony_limit,  
        PERCENTILE_DISC(0.99) WITHIN GROUP (ORDER BY ceiling_height) AS  
ceiling_height_limit_h,  
        PERCENTILE_DISC(0.01) WITHIN GROUP (ORDER BY ceiling_height) AS  
ceiling_height_limit_l  
    FROM real_estate.flats  
)  
-- Найдём id объявлений, которые не содержат выбросы:  
filtered_id AS(  
    SELECT id  
    FROM real_estate.flats  
    WHERE  
        total_area < (SELECT total_area_limit FROM limits)  
        AND (rooms < (SELECT rooms_limit FROM limits) OR rooms IS NULL)  
        AND (balcony < (SELECT balcony_limit FROM limits) OR balcony IS NULL)  
        AND ((ceiling_height < (SELECT ceiling_height_limit_h FROM limits)  
            AND ceiling_height > (SELECT ceiling_height_limit_l FROM limits)) OR  
ceiling_height IS NULL)  
    )  
-- Выведем объявления без выбросов:  
SELECT *  
FROM real_estate.flats  
WHERE id IN (SELECT * FROM filtered_id);
```

Задача 1: Время активности объявлений (проданные квартиры)

```
WITH limits AS (  
    SELECT  
        PERCENTILE_DISC(0.99) WITHIN GROUP (ORDER BY total_area) AS  
total_area_limit,  
        PERCENTILE_DISC(0.99) WITHIN GROUP (ORDER BY rooms) AS rooms_limit,  
        PERCENTILE_DISC(0.99) WITHIN GROUP (ORDER BY balcony) AS balcony_limit,  
        PERCENTILE_DISC(0.99) WITHIN GROUP (ORDER BY ceiling_height) AS  
ceiling_height_limit_h,  
        PERCENTILE_DISC(0.01) WITHIN GROUP (ORDER BY ceiling_height) AS  
ceiling_height_limit_l  
    FROM real_estate.flats  
)  
-- Найдём id объявлений, которые не содержат выбросы:
```

```

filtered_id AS(
    SELECT id
    FROM real_estate.flats
    WHERE
        total_area < (SELECT total_area_limit FROM limits)
        AND (rooms < (SELECT rooms_limit FROM limits) OR rooms IS NULL)
        AND (balcony < (SELECT balcony_limit FROM limits) OR balcony IS NULL)
        AND ((ceiling_height < (SELECT ceiling_height_limit_h FROM limits)
            AND ceiling_height > (SELECT ceiling_height_limit_l FROM limits)) OR
            ceiling_height IS NULL)
    ),
-- Выведем объявления без выбросов, только города, категоризируем их:
NEW AS (
    SELECT id, city,
        CASE
            WHEN c.city LIKE 'Санкт-Петербург' THEN 'Санкт-Петербург'
            ELSE 'ЛенОлб'
        END AS Region,
        CASE
            WHEN a.days_exposition BETWEEN 1 AND 30 THEN '1_до месяца'
            WHEN a.days_exposition BETWEEN 31 AND 90 THEN '2_до трех месяцев'
            WHEN a.days_exposition BETWEEN 91 AND 180 THEN '3_до полугода'
            WHEN a.days_exposition > 181 THEN '4_более полугода'
            ELSE 'не проданы'
        END AS Segment_activ,
        last_price/total_area AS price_for_metre,
        total_area, kitchen_area, rooms, balcony, floors_total, floor,
        ceiling_height, is_apartment, open_plan, parks_around3000,
        ponds_around3000
    FROM real_estate.flats f
    LEFT JOIN real_estate.advertisement a USING(id)
    LEFT JOIN real_estate.city c USING(city_id)
    LEFT JOIN real_estate."type" t USING(type_id)
    WHERE id IN (SELECT * FROM filtered_id) AND TYPE LIKE 'город'
    )
--Итоговая по первой задаче
SELECT Region, Segment_activ, COUNT(id) AS Count_flat,
    ROUND(COUNT(id)::NUMERIC / (SELECT COUNT(id) FROM NEW), 2) AS dolya_total,
    ROUND(COUNT(id) FILTER (WHERE city LIKE 'Санкт-Петербург')::NUMERIC /
    (SELECT COUNT(id) FROM NEW WHERE city LIKE 'Санкт-Петербург'), 2) AS dolya_Sp,
    ROUND(COUNT(id) FILTER (WHERE city NOT LIKE 'Санкт-Петербург')::NUMERIC /
    (SELECT COUNT(id) FROM NEW WHERE city NOT LIKE 'Санкт-Петербург'), 2) AS
    dolya_Lobl,
    ROUND(AVG(price_for_metre)::NUMERIC, 2) AS avg_price_metre,
    ROUND(AVG(total_area)::NUMERIC, 2) AS avg_total_area,
    ROUND(AVG(kitchen_area)::NUMERIC, 2) AS avg_kitchen_area,
    percentile_cont(0.5) WITHIN GROUP (ORDER BY rooms) AS med_count_rooms,
    percentile_cont(0.5) WITHIN GROUP (ORDER BY balcony) AS med_count_balcony,
    percentile_cont(0.5) WITHIN GROUP (ORDER BY floors_total) AS
    med_floor_total,
    percentile_cont(0.5) WITHIN GROUP (ORDER BY floor) AS med_floor,
    ROUND(AVG(ceiling_height)::NUMERIC, 2) AS avg_ch,
    COUNT(is_apartment) FILTER (WHERE is_apartment = 1) AS count_apartment,
    COUNT(open_plan) FILTER (WHERE open_plan = 1) AS count_op
FROM NEW
WHERE Segment_activ IS NOT NULL
GROUP BY Region, Segment_activ
ORDER BY Region DESC, Segment_activ

```

Задача 2: сезонность объявлений

```
WITH limits AS (  
  SELECT  
    PERCENTILE_DISC(0.99) WITHIN GROUP (ORDER BY total_area) AS  
    total_area_limit,
```

```

        PERCENTILE_DISC(0.99) WITHIN GROUP (ORDER BY rooms) AS rooms_limit,
        PERCENTILE_DISC(0.99) WITHIN GROUP (ORDER BY balcony) AS balcony_limit,
        PERCENTILE_DISC(0.99) WITHIN GROUP (ORDER BY ceiling_height) AS
ceiling_height_limit_h,
        PERCENTILE_DISC(0.01) WITHIN GROUP (ORDER BY ceiling_height) AS
ceiling_height_limit_l
    FROM real_estate.flats
),
-- Найдём id объявлений, которые не содержат выбросы:
filtered_id AS(
    SELECT id
    FROM real_estate.flats
    WHERE
        total_area < (SELECT total_area_limit FROM limits)
        AND (rooms < (SELECT rooms_limit FROM limits) OR rooms IS NULL)
        AND (balcony < (SELECT balcony_limit FROM limits) OR balcony IS NULL)
        AND ((ceiling_height < (SELECT ceiling_height_limit_h FROM limits)
            AND ceiling_height > (SELECT ceiling_height_limit_l FROM limits)) OR
ceiling_height IS NULL)
),
-- Выведем объявления без выбросов, только города, по дате публикации:
first_day AS (
    SELECT EXTRACT(MONTH FROM first_day_exposition) AS month_first,
        COUNT(id) AS count_flat_1,
        ROUND(AVG(last_price/total_area)::NUMERIC, 2) AS avg_price_metre_1,
        ROUND(AVG(total_area)::NUMERIC, 2) AS avg_total_area_1,
        percentile_cont(0.5) WITHIN GROUP (ORDER BY rooms) AS med_count_rooms_1,
        percentile_cont(0.5) WITHIN GROUP (ORDER BY balcony) AS
med_count_balcony_1,
        percentile_cont(0.5) WITHIN GROUP (ORDER BY floor) AS med_floor_1
    FROM real_estate.flats
    LEFT JOIN real_estate.advertisement a USING(id)
    LEFT JOIN real_estate.type t USING(type_id)
    WHERE id IN (SELECT * FROM filtered_id) AND EXTRACT(YEAR FROM
first_day_exposition) BETWEEN 2015 AND 2018 AND TYPE LIKE 'город'
    GROUP BY month_first
    ORDER BY month_first
),
-- Выведем объявления без выбросов, только проданные только городаб по дате
снятия:
last_day AS(
    SELECT EXTRACT(MONTH FROM (first_day_exposition + days_exposition::integer)::date)
AS month_last,
        COUNT(id) AS count_flat_2,
        ROUND(AVG(last_price/total_area)::NUMERIC, 2) AS avg_price_metre_2,
        ROUND(AVG(total_area)::NUMERIC, 2) AS avg_total_area_2,
        percentile_cont(0.5) WITHIN GROUP (ORDER BY rooms) AS med_count_rooms_2,
        percentile_cont(0.5) WITHIN GROUP (ORDER BY balcony) AS
med_count_balcony_2,
        percentile_cont(0.5) WITHIN GROUP (ORDER BY floor) AS med_floor_2
    FROM real_estate.flats
    LEFT JOIN real_estate.advertisement a USING(id)
    LEFT JOIN real_estate.type t USING(type_id)
    WHERE id IN (SELECT * FROM filtered_id) AND days_exposition IS NOT NULL AND
EXTRACT(YEAR FROM first_day_exposition) BETWEEN 2015 AND 2018 AND TYPE LIKE
'город'
    GROUP BY month_last
    ORDER BY month_last
)

```

```

-- Итоговая по 2 задаче
SELECT month_first, ROW_NUMBER() OVER(ORDER BY count_flat_1 DESC), count_flat_1,
       ROUND(count_flat_1 / (SELECT SUM(count_flat_1) FROM first_day), 2) AS
dolya,
       avg_price_metre_1, avg_total_area_1, med_count_rooms_1,
med_count_balcony_1, med_floor_1,
       month_last, ROW_NUMBER() OVER(ORDER BY count_flat_2 DESC), count_flat_2,
       ROUND(count_flat_2 / (SELECT SUM(count_flat_2) FROM last_day), 2) AS dolya,
       avg_price_metre_2, avg_total_area_2, med_count_rooms_2,
med_count_balcony_2, med_floor_2
FROM first_day AS fd
FULL JOIN last_day AS ld ON (fd.month_first = ld.month_last)
ORDER BY month_first, month_last;

```

```

WITH Total_year AS (
SELECT EXTRACT(YEAR FROM first_day_exposition) AS YEAR, EXTRACT(MONTH FROM
first_day_exposition) AS MONTH
FROM real_estate.advertisement a )

SELECT YEAR AS year, count(DISTINCT MONTH) AS count_month
FROM Total_year
GROUP BY YEAR;

```

-- 2014 и 2019 года не полные, их использование может исказить данные во 2 задаче

Задача 3: Анализ рынка недвижимости Ленобласти

```

WITH limits AS (
    SELECT
        PERCENTILE_DISC(0.99) WITHIN GROUP (ORDER BY total_area) AS
total_area_limit,
        PERCENTILE_DISC(0.99) WITHIN GROUP (ORDER BY rooms) AS rooms_limit,
        PERCENTILE_DISC(0.99) WITHIN GROUP (ORDER BY balcony) AS balcony_limit,
        PERCENTILE_DISC(0.99) WITHIN GROUP (ORDER BY ceiling_height) AS
ceiling_height_limit_h,
        PERCENTILE_DISC(0.01) WITHIN GROUP (ORDER BY ceiling_height) AS
ceiling_height_limit_l
    FROM real_estate.flats
),
-- Найдём id объявлений, которые не содержат выбросы:
filtered_id AS(
    SELECT id
    FROM real_estate.flats
    WHERE
        total_area < (SELECT total_area_limit FROM limits)
        AND (rooms < (SELECT rooms_limit FROM limits) OR rooms IS NULL)
        AND (balcony < (SELECT balcony_limit FROM limits) OR balcony IS NULL)
        AND ((ceiling_height < (SELECT ceiling_height_limit_h FROM limits)

```

```

        AND ceiling_height > (SELECT ceiling_height_limit_1 FROM limits)) OR
ceiling_height IS NULL)
    )
-- Выведем объявления без выбросов по Ленинградской области:
SELECT city, string_agg(DISTINCT type, ', ' ORDER BY type) AS type, count(id) AS
count_flat_total, count(id) FILTER (WHERE days_exposition IS NOT NULL) AS
count_flat,
    ROUND(count(id) FILTER (WHERE days_exposition IS NOT NULL) /
count(id)::NUMERIC, 2) AS dolya,
    ROUND(AVG(days_exposition)::NUMERIC, 0) AS avg_days_exposition,
    ROUND(AVG(last_price/total_area)::NUMERIC, 2) AS avg_price_metre,
    ROUND(AVG(total_area)::NUMERIC, 2) AS avg_total_area,
    percentile_cont(0.5) WITHIN GROUP (ORDER BY rooms) AS med_count_rooms,
    percentile_cont(0.5) WITHIN GROUP (ORDER BY balcony) AS med_count_balcony,
    percentile_cont(0.5) WITHIN GROUP (ORDER BY floor) AS med_floor
FROM real_estate.flats
LEFT JOIN real_estate.city c USING(city_id)
LEFT JOIN real_estate.advertisement a USING(id)
LEFT JOIN real_estate.type t USING(type_id)
WHERE id IN (SELECT * FROM filtered_id) AND city NOT LIKE 'Санкт-Петербург'
GROUP BY city
ORDER BY count_flat DESC
LIMIT 25;

```

-- примерно 75% опубликованных и проданных квартир в этих населенных пунктах

Часть 2. Аналитические выводы

Задача 1. Время активности объявлений

Чтобы спланировать эффективную бизнес-стратегию на рынке недвижимости, заказчику нужно определить — по времени активности объявления — самые привлекательные для работы сегменты недвижимости Санкт-Петербурга и городов Ленинградской области.

1. Какие сегменты рынка недвижимости Санкт-Петербурга и городов Ленинградской имеют наиболее короткие или длинные сроки активности объявлений?

- *по итогу анализа весь спектр проданных квартир можно разделить на 2 сегмента: 60% продаются в период от месяца до трех или более полугода в равных пропорциях, 40 % остальных квартир продаются за период до месяца или от трех месяцев до полугода*

- *квартир проданных в Санкт-Петербурге в 4 раза больше, чем в Ленинградской области (11 222 и 2749 соответственно)*

2. Какие характеристики недвижимости, включая площадь недвижимости, среднюю стоимость квадратного метра, количество комнат и балконов и другие параметры, влияют на время активности объявлений? Как эти зависимости варьируют между регионами?

- *средняя стоимость квадратного метра в Санкт-Петербурге 117 тыс. рублей, в Ленинградской области 70,5 тыс. рублей*

- *средняя площадь продаваемых квартир в Санкт-Петербурге 62 м², в Ленинградской области 53 м²*

- большинство проаемых квартир 2-х комнатные с 1 балконом
- в среднем потолки квартир Санкт-Петербурга выше на 10 см (2,8 и 2,7 соответственно)
- доля апартаментов и квартир свободной планировки крайне мала во всех сегментах

3. Есть ли различия между недвижимостью Санкт-Петербурга и Ленинградской области по полученным результатам?

- наблюдается противоположная зависимость между Санкт-Петербургом и Ленинградской областью:

1) Санкт-Петербург: быстрее (до месяца) продаются квартиры в среднем меньшей площадью (54 м²) по наименьшей средней стоимости за кв.метр (110,5 тыс. руб.), дольше всего (более полугода) продаются квартиры в среднем наибольшей площадью (66 м²) по наибольшей средней стоимости за кв.метр (115,5 тыс. руб.)

2) Ленинградская область: быстрее (до месяца) продаются квартиры в среднем меньшей площадью (49 м²) по наибольшей средней стоимости за кв.метр (73 тыс. руб.), дольше всего (более полугода) продаются квартиры в среднем наибольшей площадью (55 м²) по наименьшей средней стоимости за кв.метр (68 тыс. руб.)

- покупатели недвижимости Санкт-Петербурга предпочитают более высокие дома, при этом вне зависимости от региона и времени активности предпочитается середина этажности дома

- также стоит отметить схожую черту: непроданные квартиры в обоих регионах имеют наибольшую среднюю стоимость за кв.метр и наибольшую среднюю площадь (Санкт-Петербург: 135 тыс. руб., 72 м²; Ленинградская область: 74 тыс. руб., 58 м²)

Задача 2. Сезонность объявлений

Важно понять сезонные тенденции на рынке недвижимости Санкт-Петербурга и Ленинградской области — то есть для всего региона, чтобы выявить периоды с повышенной активностью продавцов и покупателей недвижимости. Это поможет спланировать маркетинговые кампании и выбрать сроки для выхода на рынок.

1. В какие месяцы наблюдается наибольшая активность в публикации объявлений о продаже недвижимости? А в какие — по снятию? Это показывает динамику активности покупателей.

- Топ три месяца публикации объявлений ноябрь, октябрь, февраль
- Топ три месяца снятия объявлений октябрь, ноябрь, сентябрь

2. Совпадают ли периоды активной публикации объявлений и периоды, когда происходит повышенная продажа недвижимости (по месяцам снятия объявлений)?

- Исходя из полученных данных, можно выделить 2 периода наибольшей активности публикаций объявлений: осенний (ноябрь, октябрь, сентябрь) и поздний зимний (февраль). Таким образом, напрашивается вывод, что периоды перед новым годом и после новогодних праздников наиболее активные как в части публикаций, так

и в части снятия объявлений. Такое поведение продавцов-покупателей скорее всего обусловлено психологическими факторами, а именно желанием изменений в новом календарном периоде.

3. Как сезонные колебания влияют на среднюю стоимость квадратного метра и среднюю площадь квартир? Что можно сказать о зависимости этих параметров от месяца?

- *колебания средней стоимости кв.метра и средней площади минимальны*
- *наибольшая стоимость кв.метра при публикации объявления (107 тыс. руб.) отмечается в августе-сентябре. Как вариант, такие результаты связаны с тем, что публикуя объявления в начале месяцев наибольшей активности продаж, продавец скорее всего делает задел на скидку. Также следует отметить месяц январь (106 тыс. руб.), здесь возможно предположить, что продавец настраивается на более длительную продажу перед более тихим периодом продаж, а также делает задел на скидку*
- *наименьшая стоимость кв.метра при публикации объявления (102,5 тыс. руб.) приходится на март, апрель, май. Можно предположить, что продавцы выставляющие объявления в эти месяцы, перед временем затишья в продажах, снижают цены для более быстрой реализации*
- *наибольшая стоимость кв.метра при снятии объявления (107 тыс. руб.) в марте. Скорее всего, покупатели в начале года больше желают улучшить свои жилищные условия и готовы платить соответственно, несмотря на низкий уровень продаж в данном месяце*
- *наименьшая стоимость кв.метра при снятии объявления (100 тыс. руб.) в мае, месяце наименьшего спроса*

2. Совпадают ли периоды активной публикации объявлений и периоды, когда происходит повышенная продажа недвижимости (по месяцам снятия объявлений)?

- *Исходя из полученных данных, можно выделить 2 периода наибольшей активности публикаций объявлений: осенний (ноябрь, октябрь, сентябрь) и поздний зимний (февраль). Таким образом, напрашивается вывод, что периоды перед новым годом и после новогодних праздников наиболее активные как в части публикаций, так и в части снятия объявлений. Такое поведение продавцов-покупателей скорее всего обусловлено психологическими факторами, а именно желанием изменений в новом календарном периоде.*

Задача 3. Анализ рынка недвижимости Ленобласти

Определить, в каких населённых пунктах Ленинградской области активнее всего продаётся недвижимость и какая именно. Так можно увидеть, где стоит поработать, и учесть особенности Ленинградской области при принятии бизнес-решений.

1. В каких населённые пункты Ленинградской области наиболее активно публикуют объявления о продаже недвижимости?

- *в ходе анализа выбрана группа из 25 населенных пунктов Ленинградской области, где заключается примерно 75% всех опубликованных (от 58 до 568*

объявлений на населенный пункт) и проданных квартир (от 50 до 532 объявлений на населенный пункт)

- топ 5 населенных пунктов: Мурино, Кудрово, Шушары, Всеволожск и Парголово

2. В каких населённых пунктах Ленинградской области — самая высокая доля снятых с публикации объявлений? Это может указывать на высокую долю продажи недвижимости.

- в среднем доля проданных квартир в отобранной группе 89%

- лидеры - Кудрово и Мурино – 94%

3. Какова средняя стоимость одного квадратного метра и средняя площадь продаваемых квартир в различных населённых пунктах? Есть ли вариация значений по этим метрикам?

- средняя стоимость кв.метра в отобранной группе – 71 тыс. руб.

- населенные пункты Ленинградской области с наибольшей средней стоимостью кв.метра – Пушкин и Сестрорецк (104 тыс. руб.), с наименьшей – Сланцы (18 тыс. руб.) и Волхов (35 тыс. руб.)

- средняя площадь продаваемых квартир в отобранной группе – 52 м²

- Среди этих населенных пунктов выделяется Сестрорецк (62,5 м²) и Никольское (42 м²) как наибольшая и наименьшая средняя площадь продаваемых квартир

-покупатели в основном предпочитают 1-2-х комнатные квартиры с 1 балконом в середине дома

4. Среди выделенных населённых пунктов какие пункты выделяются по продолжительности публикации объявлений? То есть где недвижимость продаётся быстрее, а где — медленнее.

- среднее число дней продажи в отобранной группе – 177 дней

- населенные пункты Ленинградской области с наибольшим периодом продажи – Никольское и Коммунар (236,5 дней), с наименьшим – Янино-1 (117 дней) и Кингисепп (125 дней)

Общие выводы и рекомендации

На основании анализа доступных данных можно заключить, что рынок недвижимости в Санкт-Петербурге и Ленинградской области демонстрирует высокий уровень активности. Средние цены на квартиры в Санкт-Петербурге превышают

соответствующие показатели области примерно на 60%. В отличие от Ленинградской области, в Санкт-Петербурге стоимость квадратного метра остается относительно стабильной при увеличении жилой площади, что свидетельствует о сбалансированности ценового сегмента в городе.

Общий объем сделок особенно высок в населённых пунктах, расположенных вблизи Санкт-Петербурга. В то же время, наличие аномальных ценовых отклонений в таких населённых пунктах, как Сланцы и Волхов, требует проведения дополнительного анализа для выявления причин данных аномалий.

Распределение времени реализации объектов показывает, что наибольшее число сделок совершается в течение одного-трёх месяцев с момента публикации или после полугода с момента размещения объявления. Осенний период является наиболее предпочтительным для публикации и снятия недвижимости с рынка, что согласуется с сезонными тенденциями.

Для получения более глубокой аналитической картины необходимо расширить объем исходных данных и провести дополнительное исследование причин выявленных закономерностей.

Рекомендуется учитывать вышеуказанные особенности рынка при планировании стратегий работы, а также своевременно реагировать на сезонные колебания и аномалии.