

*

Описание данных

Таблица advertisement

Содержит информацию об объявлениях:

id — идентификатор объявления (первичный ключ).

first_day_exposition — дата подачи объявления.

days_exposition — длительность нахождения объявления на сайте (в днях).

last_price — стоимость квартиры в объявлении, в руб.

Таблица flats

Содержит информацию о квартирах:

id — идентификатор квартиры (первичный ключ, связан с первичным ключом id таблицы advertisement).

city_id — идентификатор города (внешний ключ, связан с city_id таблицы city).

type_id — идентификатор типа населённого пункта (внешний ключ, связан с type_id таблицы type).

total_area — общая площадь квартиры, в кв. метрах.

rooms — число комнат.

ceiling_height — высота потолка, в метрах.

floors_total — этажность дома, в котором находится квартира.

living_area — жилая площадь, в кв. метрах.

floor — этаж квартиры.

is_apartment — указатель, является ли квартира апартаментами (1 — является, 0 — не является).

open_plan — указатель, имеется ли в квартире открытая планировка (1 — открытая планировка квартиры, 0 — открытая планировка отсутствует).

kitchen_area — площадь кухни, в кв. метрах.

balcony — количество балконов в квартире.

airports_nearest — расстояние до ближайшего аэропорта, в метрах.

parks_around3000 — число парков в радиусе трёх километров.

ponds_around3000 — число водоёмов в радиусе трёх километров.

Таблица city

Содержит информацию о городах:

city_id — идентификатор населённого пункта (первичный ключ).

city — название населённого пункта.

Таблица type

Содержит информацию о городах:

type_id — идентификатор типа населённого пункта (первичный ключ).

type — название типа населённого пункта.

Зависимости данных вы можете увидеть на ER-диаграмме:

Таблица с объявлениями advertisement связана с данными таблицы с квартирами flats, а та — с данными таблиц с информацией о городах: city и type

*/

-- 1. Время активности объявлений

WITH *limits* **AS** (-- Определим граничные значения аномальных данных:

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

),

filtered_id **AS** (-- Найдем id объявлений, которые не содержат выбросы:

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**)

),

categorized_statistics **AS** (

SELECT

*

,

ROUND(last_price::NUMERIC / total_area::NUMERIC, 2) **AS** sq_met_price,

```

CASE
    WHEN days_exposition <= 30 THEN '-до месяца'
    WHEN days_exposition <= 90 THEN '--до 3х месяцев'
    WHEN days_exposition <= 180 THEN '----до полугода'
    WHEN days_exposition > 180 THEN '----более полугода'
END AS periods,
CASE
    WHEN city = 'Санкт-Петербург' THEN 'Санкт-Петербург'
    ELSE 'ЛО'
END AS regions
FROM real_estate.flats AS f
JOIN real_estate.advertisement AS a USING(id)
JOIN real_estate.city AS c USING(city_id)
JOIN real_estate.type AS t USING(type_id)
WHERE id IN (SELECT * FROM filtered_id)
AND type = 'род' --- Выборка только по городам
AND days_exposition IS NOT NULL
),
grouperd_stat AS (
    SELECT
        regions AS "Регион",
        periods AS "Активность объявлений",
        ROUND(AVG(sq_met_price), 2) AS "Средняя стоимость кв.метра",
        ROUND(AVG(total_area)::NUMERIC, 2) AS "Средняя площадь",
        PERCENTILE_DISC(0.5) WITHIN GROUP (ORDER BY rooms) AS "Медиана количества комнат",
        PERCENTILE_DISC(0.5) WITHIN GROUP (ORDER BY balcony) AS "Медиана количества балконов",
        PERCENTILE_DISC(0.5) WITHIN GROUP (ORDER BY floor) AS "Медиана этажности",
        COUNT(id) AS "Кол-во объявлений"
    FROM categorized_statistics
    GROUP BY regions, periods
)
SELECT
    *,

```

```

ROUND("Кол-во объявлений"::NUMERIC / SUM("Кол-во объявлений") OVER(PARTITION BY Регион), 2) AS "Доля от объявлений в
регионе"
FROM grouperd_stat
ORDER BY "Активность объявлений" DESC, "Регион" DESC;

```

Регион	Активность объявлений	Средняя стоимость кв.метра	Средняя площадь	Медиана количества комнат	Медиана количества балконов	Медиана этажности	Кол-во объявлений	Доля от объявлений в регионе
Санкт-Петербург	-до месяца	110 568,88	54,38	2	1	5	2 168	0,19
ЛО	-до месяца	73 275,25	48,72	2	1	4	397	0,14
Санкт-Петербург	--до 3х месяцев	111 573,24	56,71	2	1	5	3 236	0,29
ЛО	--до 3х месяцев	67 573,43	50,88	2	1	3	917	0,33
Санкт-Петербург	---до полугода	111 938,92	60,55	2	1	5	2 254	0,2
ЛО	---до полугода	69 846,39	51,83	2	1	3	556	0,2
Санкт-Петербург	----более полугода	115 457,22	66,15	2	1	5	3 581	0,32
ЛО	----более полугода	68 297,22	55,41	2	1	3	890	0,32

-- 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
),
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)
),
cleaned_flats AS (
SELECT
*,
ROUND(last_price::NUMERIC / total_area::NUMERIC, 2) AS sq_met_price
FROM real_estate.flats AS f
JOIN real_estate.advertisement AS a USING(id)
JOIN real_estate.city AS c USING(city_id)
JOIN real_estate.type AS t USING(type_id)
WHERE id IN (SELECT * FROM filtered_id)
AND type = 'ropод'
AND first_day_exposition BETWEEN '2015-01-01' AND '2018-12-31' -- фильтр на полные годы
AND first_day_exposition + INTERVAL '1 day' * days_exposition BETWEEN '2015-01-01' AND '2018-12-31'
),
exposition AS (
SELECT
EXTRACT(MONTH FROM first_day_exposition) AS "Месяц",
COUNT(id) AS "Новые объекты",
ROUND(COUNT(id) * 100.0 / SUM(COUNT(id)) OVER(), 2) AS "Доля от числа публикаций (%)",
ROUND(AVG(sq_met_price)::NUMERIC, 2) AS "Средняя цена за метр (публик.)",
ROUND(AVG(total_area)::NUMERIC, 2) AS "Средняя площадь объекта (публик.)"
FROM cleaned_flats
GROUP BY EXTRACT(MONTH FROM first_day_exposition)
),
sale AS (
SELECT
EXTRACT(MONTH FROM first_day_exposition + INTERVAL '1 day' * days_exposition) AS "Месяц",
COUNT(id) AS "Проданные объекты",
ROUND(COUNT(id) * 100.0 / SUM(COUNT(id)) OVER(), 2) AS "Доля от числа продаж (%)",
ROUND(AVG(sq_met_price)::NUMERIC, 2) AS "Средняя цена за метр (продажа)",
ROUND(AVG(total_area)::NUMERIC, 2) AS "Средняя площадь объекта (продажа)"

```

```

FROM cleaned_flats
GROUP BY EXTRACT(MONTH FROM first_day_exposition + INTERVAL '1 day' * days_exposition)
)
SELECT
    *,
    DENSE_RANK() OVER(ORDER BY "Новые объекты" DESC) AS "Ранг по публикациям",
    DENSE_RANK() OVER(ORDER BY "Проданные объекты" DESC) AS "Ранг по продажам"
FROM exposition AS e
JOIN sale AS s USING("Месяц")
ORDER BY "Ранг по продажам";

```

Месяц	Новые объекты	Доля от числа публикаций (%)	Средняя цена за метр (публик.)	Средняя площадь объекта (публик.)	Проданные объекты	Доля от числа продаж (%)	Средняя цена за метр (продажа)	Средняя площадь объекта (продажа)	Ранг по публикациям	Ранг по продажам
10	1 113	9,28	101 233,64	57,3	1 360	11,34	104 317,33	58,86	5	1
11	1 181	9,84	102 030,18	56,99	1 301	10,84	103 791,36	56,71	2	2
9	1 140	9,5	106 684,56	59,05	1 238	10,32	104 070,07	57,49	3	3
12	766	6,38	102 060,52	57,25	1 175	9,79	105 504,52	59,26	11	4
8	998	8,32	104 438	56,82	1 137	9,48	100 036,51	56,83	7	5
7	984	8,2	103 100,6	57,79	1 108	9,23	102 290,72	58,54	8	6
1	674	5,62	104 266,11	57,67	870	7,25	103 814,62	57,33	12	7
3	1 010	8,42	101 429,57	58,8	818	6,82	105 165,05	58,4	6	8
6	1 125	9,38	103 618,57	57,83	771	6,43	101 863,69	59,82	4	9
4	934	7,78	101 468,25	59,58	765	6,38	100 187,56	56,56	9	10
2	1 246	10,39	101 789,46	58,75	740	6,17	100 820,1	59,62	1	11
5	827	6,89	102 255,15	58,78	715	5,96	99 558,57	57,82	10	12

-- 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
),
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)
),
cleaned_flats AS (
    SELECT
        *,
        ROUND(last_price::NUMERIC / total_area::NUMERIC, 2) AS sq_met_price
    FROM real_estate.flats AS f
    JOIN real_estate.advertisement AS a USING(id)
    JOIN real_estate.city AS c USING(city_id)
    JOIN real_estate.type AS t USING(type_id)
    WHERE id IN (SELECT * FROM filtered_id) AND city != 'Санкт-Петербург' ---только ЛО
)
SELECT
    city "Г.п.",
    COUNT(id) AS "Количество публикаций",
    COUNT(days_exposition) AS "Количество продаж",
    ROUND(COUNT(days_exposition)::NUMERIC / COUNT(id), 2) AS "Доля продаж от публикаций",
    ROUND(AVG(sq_met_price)::NUMERIC, 2) AS "Средняя стоимость кв.метра",
    ROUND(AVG(total_area)::NUMERIC, 2) AS "Средняя площадь",
    PERCENTILE_DISC(0.5) WITHIN GROUP (ORDER BY days_exposition) AS "Медиана длительности продажи",
    DENSE_RANK() OVER(ORDER BY COUNT(days_exposition) DESC) AS "Ранг по продажам"

```

FROM *cleaned_flats*
GROUP BY city
ORDER BY "Ранг по продажам"
LIMIT 15;

Месяц	Новые объекты	Доля от числа публикаций (%)	Средняя цена за метр (публик.)	Средняя площадь объекта (публик.)	Проданные объекты	Доля от числа продаж (%)	Средняя цена за метр (продажа)	Средняя площадь объекта (продажа)	Ранг по публикациям	Ранг по продажам
10	1 113	9,28	101 233,64	57,3	1 360	11,34	104 317,33	58,86	5	1
11	1 181	9,84	102 030,18	56,99	1 301	10,84	103 791,36	56,71	2	2
9	1 140	9,5	106 684,56	59,05	1 238	10,32	104 070,07	57,49	3	3
12	766	6,38	102 060,52	57,25	1 175	9,79	105 504,52	59,26	11	4
8	998	8,32	104 438	56,82	1 137	9,48	100 036,51	56,83	7	5
7	984	8,2	103 100,6	57,79	1 108	9,23	102 290,72	58,54	8	6
1	674	5,62	104 266,11	57,67	870	7,25	103 814,62	57,33	12	7
3	1 010	8,42	101 429,57	58,8	818	6,82	105 165,05	58,4	6	8
6	1 125	9,38	103 618,57	57,83	771	6,43	101 863,69	59,82	4	9
4	934	7,78	101 468,25	59,58	765	6,38	100 187,56	56,56	9	10
2	1 246	10,39	101 789,46	58,75	740	6,17	100 820,1	59,62	1	11
5	827	6,89	102 255,15	58,78	715	5,96	99 558,57	57,82	10	12