Real Estate Analysis

Description:

In this data analysis project, the mission is to gather and process real estate data from Otodom Real Estate Agency using the Bright Data platform. The primary goal is to create reports tailored to answer users' questions about properties, including price, area, and related details. Otodom plays a pivotal role in property listings, while Bright Data specializes in data collection. The project involves data collection, transformation for analysis, and the provision of insights into price trends, property types, and more. The resulting reports will empower users, including property buyers, sellers, and investors, with data-driven information for informed real estate decisions presented in a user-friendly format.

Working with Snowflake:

-- Storing the Otodom Real Estate data in snowflake database

-- Creating database

```
create or replace database realestate;
```

-- Creating data warehouse

```
select or replace warehouse realestate_wh;
```

-- Creating table to store the data

```
create table otodom_data_dump(json_data text);
```

-- Creating file format for working with csv files

```
create or replace file format csv_format
type = csvfield_delimiter = ','
field_optionally_enclosed_by="";
```

-- Creating stating area to store csv files in staging area

```
create or replace stage real_estate_stage
file_format=csv_format;
```

-- Upload the data(csv file) in staging area via snowsql command line or through snowflake UI

```
-- Copying the data from files in stage area to table
copy into otodom_data_dump
from @real_estate_stage
on_error = "skip_file";

-- No of records in table
select count(1) from otodom_data_dump;

-- Retreiving price data
select parse_json(json_data):price from otodom_data_dump limit 5;

select * from otodom_data_dump limit 5;

-- Flattening the data
CREATE OR REPLACE table otodom_data_flatten
as
select row_number() over(order by title) as rn
, x.*
from (
select replace(parse_json(json_data):advertiser_type,'')::string as advertiser_type
, replace(parse_json(json_data):balcony_garden_terrace,'')::string as balcony_garden_terrace
, regexp_replace(replace(parse_json(json_data):description,''), '<[^>]+>')::string as description
, replace(parse_json(json_data):heating,'')::string as heating
, replace(parse_json(json_data):is_for_sale,'')::string as is_for_sale
, replace(parse_json(json_data):lighting,'')::string as lighting
, replace(parse_json(json_data):location,'')::string as location
, replace(parse_json(json_data):price,'')::string as price
, replace(parse_json(json_data):remote_support,'')::string as remote_support
, replace(parse_json(json_data):rent_sale,'')::string as rent_sale
, replace(parse_json(json_data):surface,'')::string as surface
, replace(parse_json(json_data):timestamp,'')::date as timestamp
, replace(parse_json(json_data):title,'')::string as title
, replace(parse_json(json_data):url,'')::string as url
, replace(parse_json(json_data):form_of_property,'')::string as form_of_property
```

```
, replace(parse_json(json_data):no_of_rooms,""):string as no_of_rooms
, replace(parse_json(json_data):parking_space,""):string as parking_space
from otodom_data_dump
) x;
```

--Retrieving records from table containing flattened data

```
select * from otodom_data_flatten limit 5;

select count(*) from otodom_data_flatten_translate;
```

-- Transforming the data (Using Python)

Using python to connect to snowflake and transform the following:

-- Returing Address from the coordinates in the Location column

-- Translating Title column from Polish Language to English language

We can use the python notebook to transform and store the data directly to snowflake OR

We can manually upload the files as follows:

For Address

-- Creating table for fetched address data

```
create table otodom_data_transformed_address
(
   rn   int,
   location  text,
   address  text
);
```

-- Creating a stage for uploading the transformed address data

```
create or replace stage real_estate_stage_address
file_format=csv_format;
```

-- Upload the data in staging area via snowsql command line or through snowflake UI

-- Copying the data from files in stage area to table

```
copy into otodom_data_transformed_address
from @real_estate_stage_address
on_error = "skip_file";
```

For Title

-- Creating table for translated title column

```
create table otodom_data_transformed_title
(
   rn   int,
   title  text,
   title_eng  text
);
```

-- Creating a stage for uploading the transformed title data

```
create or replace stage real_estate_stage_title
file_format=csv_format;
```

-- Upload the data in staging area via snowsql command line or through snowflake UI

-- Copying the data from files in stage area to table

```
copy into otodom_data_transformed_title
from @real_estate_stage_title
on_error = "skip_file";
```

Transforming the data (using sql query)

-- Transforming price column to remove currency notation,conversion of currency values and modify number format

-- Transforming surface column to remove metrics notation and modify number format

-- Adding appartment flag column to identify if a property is an apartment or not.

```
-- Table for keeping the record of transformation
create or replace table otodom_data_transformed
as
with cte as
   (select ot.*
   , case when price like 'PLN%' then try_to_number(replace(price,'PLN ',''),'999,999,999.99')
         when price like '€%' then try_to_number(replace(price,'€',''),'999,999,999.99') * 4.43 -- Conversion to PLN
     end as price_new
   , try_to_double(replace(replace(replace(replace(surface,'m²',''),'м²',''),' ',''),',','.'),'9999.99') as surface_new
   , replace(parse_json(addr.address):suburb,'"', '') as suburb
   , replace(parse_json(addr.address):city,'"', '') as city
   , replace(parse_json(addr.address):country,'"', '') as country
   , title.title_eng as title_eng
   from otodom_data_flatten ot
   left join otodom_data_transformed_address addr on ot.rn=addr.rn
   left join otodom_data_transformed_title title on ot.rn=title.rn)
select *
, case when lower(title_eng) like '%commercial%' or lower(title_eng) like '%office%' or lower(title_eng) like '%shop%' then 'non apartment'
      when is_for_sale = 'false' and surface_new <=330 and price_new <=55000 then 'apartment'
      when is_for_sale = 'false' then 'non apartment'
      when is_for_sale = 'true'  and surface_new <=600 and price_new <=20000000 then 'apartment'
      when is_for_sale = 'true'  then 'non apartment'
  end as apartment_flag
from cte;

Select * from otodom_data_transformed;
```

Reports to solve problems related to the property market in Poland based on Otodom Real Estate Agency data.

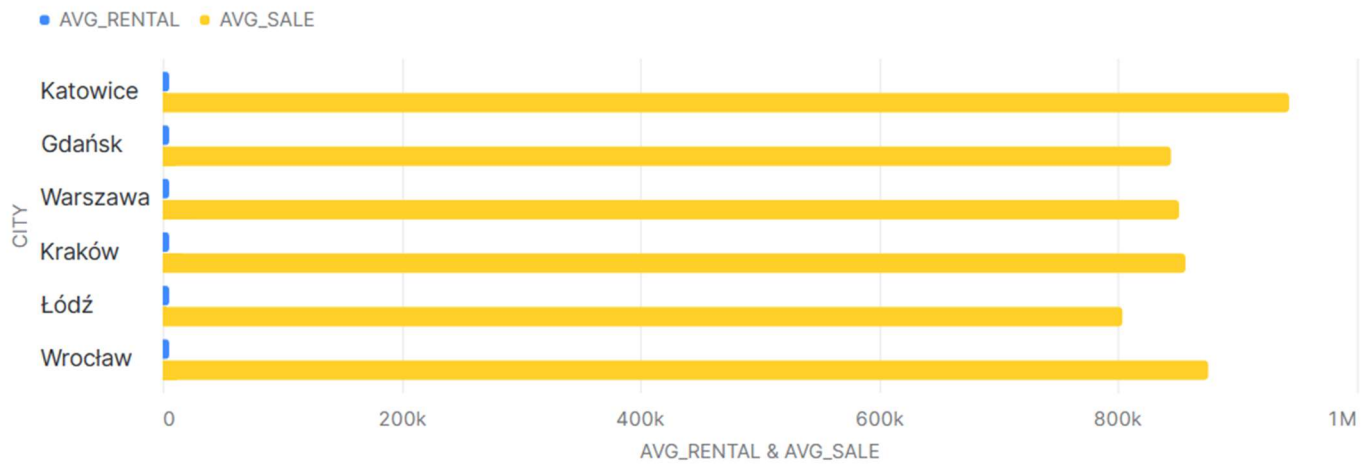
Problem 1:What is the average rental price and sale price in some of the major cities in Poland?

-- Query:

```
with cte as
(select city
, (case when is_for_sale='false' then round(avg(price_new),2) end) as avg_rental
, (case when is_for_sale='true' then round(avg(price_new),2) end) as avg_sale
from otodom_data_transformed
where city in ('Warszawa', 'Wrocław', 'Kraków', 'Gdańsk', 'Katowice', 'Łódź')
and apartment_flag = 'apartment'
group by city, is_for_sale)
select city, max(avg_rental) as avg_rental, max(avg_sale) as avg_sale
from cte
group by city
order by avg_rental desc ;
```

--Results:

	CITY	AVG_RENTAL	AVG_SALE
1	Wrocław	5,924.03	875,795.07
2	Łódź	5,832.47	803416.40
3	Kraków	5,687.03	856,620.95
4	Warszawa	5,599.43	850,279.66
5	Gdańsk	5,400.22	844,101.16
6	Katowice	5,121.88	942,990.17



Problem 2: Which suburb in warsaw has the most and least no of private ads?

-- Query:

```
select distinct
first_value(suburb||' - '||count(1)) over(order by count(1)) as least_private_ads
, last_value(suburb||' - '||count(1)) over(order by count(1)) as most_private_ads
from otodom_data_transformed
where city = 'Warszawa'
and advertiser_type = 'private'
and suburb is not null
group by suburb;
```

--Results:

	LEAST_PRIVATE_ADS	MOST_PRIVATE_ADS ...
1	Wesoła - 1	Wola - 83

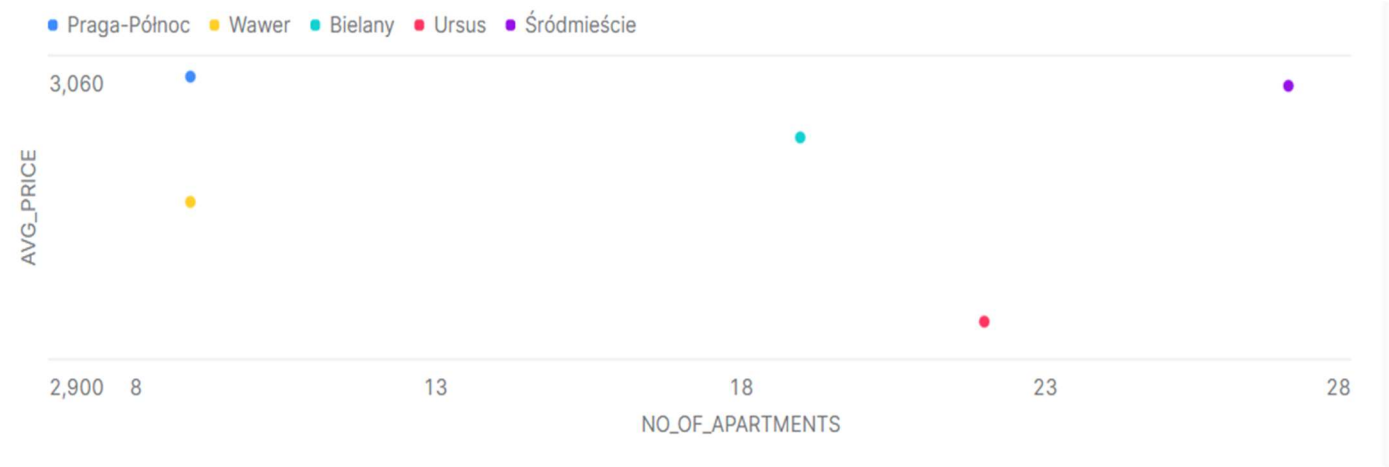
Problem 3: What are the top 5 most affordable neighborhoods in warsaw? Affordable neighbourhoods are suburbs with an aprtment size of about 40-60 m2

-- Query:

```
select suburb, avg_price, no_of_apartments
from (
select suburb, round(avg(price_new),2) avg_price, count(1) as no_of_apartments
, rank() over(order by avg_price ) as rn
from otodom_data_transformed
where city = 'Warszawa'
and apartment_flag = 'apartment'
and is_for_sale = 'false'
and surface_new between 40 and 60
and suburb is not null
group by suburb) x
where x.rn <= 5;
```

--Results:

	SUBURB	AVG_PRICE	···	NO_OF_APARTMENTS
1	Ursus	2,919.18		22
2	Wawer	2,982.22		9
3	Bielany	3,016.26		19
4	Śródmieście	3,042.96		27
5	Praga-Północ	3,047.96		9



Problem 4: Which are the top 3 most luxurious neighborhoods in Warsaw? Luxurious neighbourhoods are the suburbs which has the most no of of apartments costing over 2M in cost.

-- Query:

```
select suburb, luxurious_apartments
from (
select suburb, count(1) luxurious_apartments
, rank() over(order by luxurious_apartments desc ) as rn
from otodom_data_transformed
where city = 'Warszawa'
and apartment_flag = 'apartment'
and is_for_sale = 'true'
and price_new > 2000000
and suburb is not null
group by suburb) x
where x.rn <= 3;
```

--Results:

	SUBURB ···	LUXURIOUS_APARTMENTS
1	Mokotów	32
2	Wola	28
3	Włochy	24

Problem 5: What is the average rental price for apartments in warsaw in different suburbs? Also categorize the result based on surface area 0-50, 50-100 and over 100

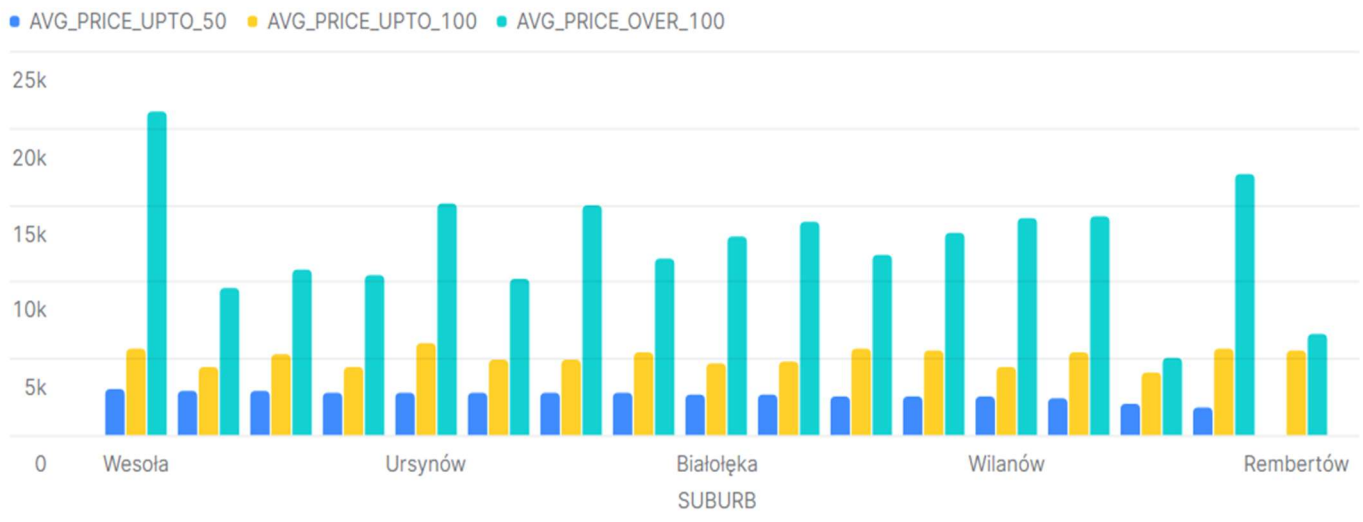
-- Query:

```
with cte1 as
(select a.*
, case when surface_new between 0 and 50 then '0-50'
when surface_new between 50 and 100 then '50-100'
when surface_new > 100 then '>100'
end as area_category
from otodom_data_transformed a
where city = 'Warszawa'
and apartment_flag = 'apartment'
and is_for_sale = 'false'
and suburb is not null ),
cte2 as
(select suburb
, case when area_category = '0-50' then avg(price_new) end as avg_price_upto50
, case when area_category = '50-100' then avg(price_new) end as avg_price_upto100
, case when area_category = '>100' then avg(price_new) end as avg_price_over100
from cte1
group by suburb,area_category)
select suburb
```

```
, round(max(avg_price_upto50),2) as avg_price_upto_50
, round(max(avg_price_upto100),2) as avg_price_upto_100
, round(max(avg_price_over100),2) as avg_price_over_100
from cte2
group by suburb
order by suburb;
```

--Results:

	SUBURB	AVG_PRICE_UPTO_50	···	AVG_PRICE_UPTO_100	AVG_PRICE_OVER_100
1	Bemowo	2410.70		5,403.08	14,277.41
2	Białołęka	2,632.96		4,722.05	12,899.06
3	Bielany	2,882.48		5,281.36	10,748.55
4	Mokotów	2,511.91		5,546.47	13,109.43
5	Praga-Południe	2,576.81		4,834.38	13,921.44
6	Praga-Północ	2,802.78		4,460.63	10352.70
7	Rembertów	null		5460.00	6565.00
8	Targówek	2,553.16		5,657.36	11753.60
9	Ursus	2,744.94		4,860.99	14,905.79
10	Ursynów	2,800.94		5,923.32	15,071.52
11	Wawer	1825.00		5,640.77	16994.50
12	Wesoła	3050.00		5628.50	21000.00
13	Wilanów	2,509.38		4,475.75	14,057.71
14	Wola	2,768.87		4,951.66	10,171.62
15	Włochy	2,725.55		5,327.12	11,522.11
16	Śródmieście	2,912.97		4,440.14	9,606.25
17	Żoliborz	2087.40		4,116.67	5000.00



Problem 6: What is the avg sale price for apartments within 50-70 m2 area in major cities of Poland?

-- Query:

```
select city, round(avg(price_new),2) as avg_sale_price
from otodom_data_transformed
where city in ('Warszawa', 'Wrocław', 'Kraków', 'Gdańsk', 'Katowice', 'Łódź')
and apartment_flag = 'apartment'
and is_for_sale = 'true'
and no_of_rooms = 3
and surface_new between 50 and 70
group by city
order by avg_sale_price desc;
```

--Results:

	CITY	AVG_SALE_PRICE
1	Gdańsk	656,124.61
2	Warszawa	653,508.87
3	Wrocław	650,865.71
4	Katowice	647,147.07
5	Łódź	642,795.62
6	Kraków	637,627.71

Problem 7: What is the percentage of private & business ads on otodom?

-- Query:

```
with all_ads as
(select count(1) tot_ads from otodom_data_transformed),
```

```
ads_type as
(select advertiser_type
, sum(case when advertiser_type='business' then 1 end) as business_ads
, sum(case when advertiser_type='private' then 1 end) as private_ads
from otodom_data_transformed
group by advertiser_type)
select concat(round((max(business_ads) * 100)/max(tot_ads),2),'%') as business_ads_perc
, concat(round((max(private_ads) * 100)/max(tot_ads),2),'%') as private_ads_perc
from ads_type ty
cross join all_ads al ;
```

--Results:

	BUSINESS_ADS_PERC ···	PRIVATE_ADS_PERC
1	89.99%	10.01%

Problem 8: What are the most expensive apartments in major cities of Poland? Also display suburb, cost, size.

-- Query:

```
with cte as
(select city, max(price_new) max_price, min(price_new) min_price
from otodom_data_transformed
where city in ('Warszawa', 'Wrocław', 'Kraków', 'Gdańsk', 'Katowice', 'Łódź')
and apartment_flag = 'apartment'
and is_for_sale = 'true'
group by city)
select x.rn, x.title_eng, x.city, x.suburb, x.price_new, x.surface_new, x.url
from otodom_data_transformed x
join cte on cte.city=x.city and cte.max_price=x.price_new
where apartment_flag = 'apartment'
and is_for_sale = 'true'
order by x.city,x.price_new;
```

--Results:

	RN	TITLE_ENG	CITY	SUBURB	PRICE_NEW	SURFACE_NEW	URL
1	17,554	2-room apartment 43m2 + balcony without commission	Gdańsk	Ujeścisko - Łos	11000000.00	274	https://www.otodom.pl/pl/oferta/apartament-r
2	17,151	2-room apartment 42m2 + loggia without commission	Katowice	Koszutka	17720000.00	226	https://www.otodom.pl/pl/oferta/apartament-f
3	55,968	Apartment in a representative tenement house!	Kraków	Grzegórzki	16000000.00	560	https://www.otodom.pl/pl/oferta/ultra-nowocz
4	55,351	Delux apartment in the heart of Powiśle	Warszawa	Wesoła	20000000.00	577	https://www.otodom.pl/pl/oferta/topowa-lokal
5	32,257	3-room apartment 55m2 + 2 gardens	Wrocław	Przedmieście O	9450000.00	349	https://www.otodom.pl/pl/oferta/ks-witolda-ca
6	25,793	3 rooms 2 20m2 balconies ideal for investments.	Łódź	Łódź-Polesie	7500000.00	570	https://www.otodom.pl/pl/oferta/ekskluzywna

Problem 9: What size of an apartment can I expect with a monthly rent of 3000 to 4000 PLN in different major cities of Poland?

-- Query:

```
select city, avg(surface_new) avg_area
from otodom_data_transformed
where city in ('Warszawa', 'Wrocław', 'Kraków', 'Gdańsk', 'Katowice', 'Łódź')
and apartment_flag = 'apartment'
and is_for_sale = 'false'
and price_new between 3000 and 4000
group by city
order by avg_area;
```

--Results:

	CITY	AVG_AREA
1	Kraków	52.472058824
2	Wrocław	52.832840909
3	Gdańsk	53.408571429
4	Łódź	53.815285714
5	Katowice	54.363484848

Barchart:

Problem 10: Display the suburbs in warsaw where one can find apartments which is around 90-100 m2 and within a range of 800,000 to 1M PLN

-- Query:

```
select suburb, count(1), avg(price_new) avg_price
from otodom_data_transformed
where city in ('Warszawa')
and apartment_flag = 'apartment'
and is_for_sale = 'true'
and surface_new between 90 and 100
and price_new between 800000 and 1000000
group by suburb
order by count(1) desc;
```

--Results:

	SUBURB	COUNT(1)	AVG_PRICE
1	Praga-Południe	6	888,633
2	Ursus	5	895,247
3	Mokotów	4	916,088
4	Bielany	4	868,750
5	Bemowo	3	932,667
6	Włochy	3	917,530
7	Białołęka	3	929,333
8	Wola	2	949,500
9	Ursynów	2	877,688
10	Wilanów	2	895,000
11	Targówek	2	907,000
12	Praga-Północ	1	997,663
13	Wawer	1	928,000
14	null	1	864,000

		SUBURB				
	COUNT	Bemowo	Białołęka	Bielany	Mokotów	Praga-Południe	Praga-Północ
1							997,663
2							
3		932,666.66666667	929,333.33333333				
4				868,750	916,087.5		
5							
6						888,633.16666667	

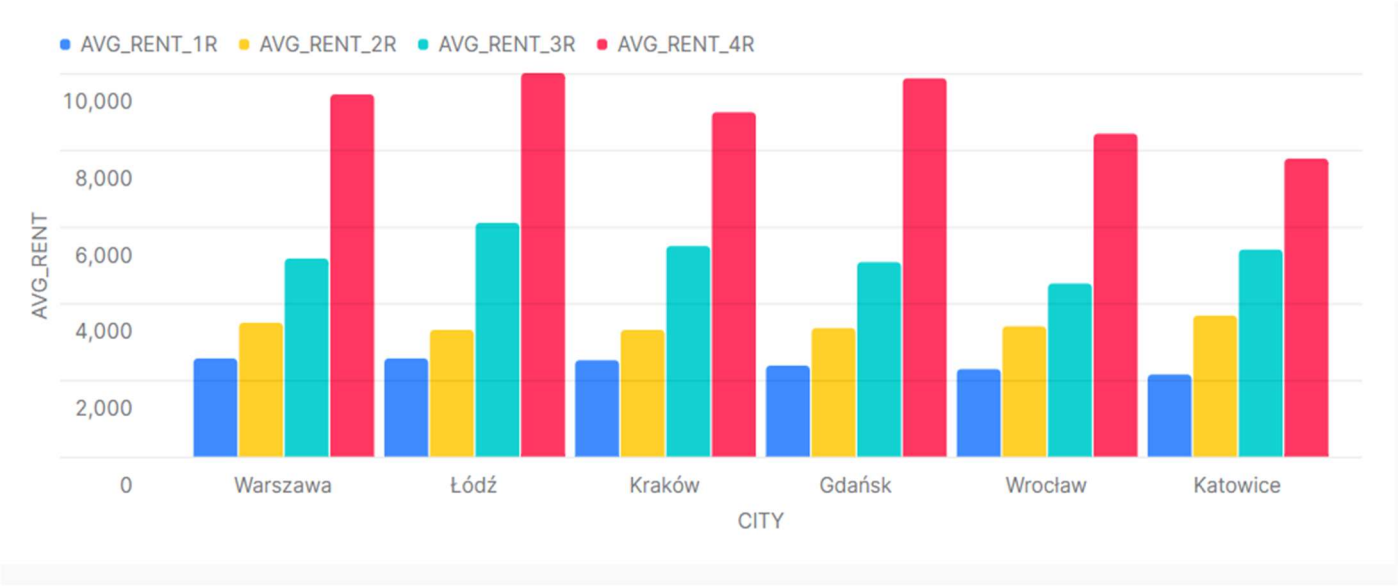
Problem 11: What is average rental price of 1 room, 2 room, 3 room and 4 room apartments in some of the major cities in Poland? Also arrange the result such that avg rent for each type fo room is shown in seperate column.

-- Query:

```
select city, round(avg_rent_1R,2) as avg_rent_1R
, round(avg_rent_2R,2) as avg_rent_2R, round(avg_rent_3R,2) as avg_rent_3R
, round(avg_rent_4R,2) as avg_rent_4R
from (
select city,no_of_rooms,price_new
from otodom_data_transformed
where city in ('Warszawa', 'Wrocław', 'Kraków', 'Gdańsk', 'Katowice', 'Łódź')
and apartment_flag = 'apartment'
and is_for_sale='false'
and no_of_rooms in (1,2,3,4)) x
pivot
(
avg(price_new)
for no_of_rooms in ('1','2','3','4')
)
as p(city,avg_rent_1R, avg_rent_2R, avg_rent_3R, avg_rent_4R)
order by avg_rent_4R desc;
```

--Results:

	CITY	AVG_RENT_1R	···	AVG_RENT_2R	AVG_RENT_3R	AVG_RENT_4R
1	Łódź	2,562.76		3303.90	6,080.58	9,984.38
2	Gdańsk	2,349.73		3,370.19	5,088.26	9,851.79
3	Warszawa	2,571.56		3,476.04	5,139.82	9,454.67
4	Kraków	2,523.97		3,306.27	5,510.42	8980.80
5	Wrocław	2,271.78		3,398.93	4,529.12	8,409.57
6	Katowice	2,142.38		3,656.16	5,396.53	7,768.75


AVG_RENT_1R
AVG_RENT_2R
AVG_RENT_3R
AVG_RENT_4R
10,000
8,000
6,000
4,000
2,000
0
AVG_RENT
Warszawa
Łódź
Kraków
Gdańsk
Wrocław
Katowice
CITY