**MARKET TRENDS**

The task is to make real estate market analysis (based on macro trends and sellers behaviour in promoting real estate ads).

Please prepare a complete analysis using the available data (provided in CSV files)

The task itself is divided into 2 parts:

**Technical part:**

1. SQL queries that allow analyzing market data and seller data - it can be in the form of a separate file (.sql) or as part of a script in python / R. Configuring a database connection (mysql, postgesql, etc.) would be a big plus, but it’s optional

2. Please save cleaned dataset (“data\_market\_prices”) into new CSV file and keep the same format as in initial dataset

**Analytical part:**

1. Please prepare a complete analysis of the data that was sent, along with the answers to the following questions

Using “data\_market\_prices” dataset:

* Describe data structure.
* What differences do you see between the advertisement types in the data available? (what are the mean prices among real estate categories?)
* What changes do you see in the market trends?
* Can you make some hypotheses for reasons for observed market changes?

Using “data\_vas\_purchases” and “data\_categories” datasets:

* Do you see any specific patterns of purchasing promotional products (which real estate category is the most promoted one etc)?
* Do you see anything in common / correlations between “data\_market\_prices” and “data\_vas\_purchases” datasets

**Form:**

* SQL codes used for cleaning and joining the datasets provided in the analysis (Python and Jupyter notebook would be a big plus, but it’s optional)
* The scripts can be in separate files or as part of the notebook depending on the methods chosen
* Please present the final results and the most important conclusions in the form of a presentation with visualization (e.g. Google slides)

The data we have at your disposal:

* data\_market\_prices - information about historical prices of properties splitted by category from February and March 2021 and 2022
* data\_vas\_purchases - information about all promotional products purchases made by sellers
* data\_categories - mapping to category\_id available in datasets mentioned above (in order to get the names of categories)

Column names for data\_market\_prices

* ad\_id
* city\_id
* category\_id
* market
* date\_posted
* date\_expired
* price\_per\_sqm
* price
* rooms\_num

Column names for data\_vas\_purchases

* date\_day
* ad\_id
* city\_id
* category\_id
* amount\_spend\_to\_promote\_ads

Column names for data\_categories

* category\_id
* category\_name
* subcategory\_name