The work we have done in the Working Lab serves as the basis for this installment. Starting from the housing rental data, we have found that a very important variable for predicting housing prices is the size of the housing measured in square metres. This is an approach that we knew a priori and that we knew would have an impact on the final model.

Fearing that the square meter variable may have too much weight in the model and therefore downplay the rest of the variables, we propose to create a new series of models that are capable of predicting a new target variable: rental price per square meter. .

The objective of this delivery is, using the knowledge acquired in the Working Lab, to generate a model that is capable of predicting the rental price per square meter, to generate predictions in the apartment purchase dataset.

It is requested:

**Exploratory analysis and cleanup of instances (1.5p):**
At this point we want you to understand the dataset well and to make the necessary transformations to better understand the data you have. It will be assessed by:

Descriptive statistics and general data quality
Visualizations that help understand the distribution of variables and categories
Visualizations that help understand the relationship between the attributes and the target variable rental_price_per_square_meter
Statistics that help us better understand the relationship (PCA, ANOVA, Correlation...)
Summary of the conclusions drawn at this point

**Creating and comparing models: (4.5p)**
At this point, it is requested that at least: 1 Bagging model, 1 Random Forest model and 1 Boosting model be developed to choose between CATBoost, XGBoost or LightGBM. Each model must follow the following steps:

Data preprocessing: Make the necessary transformations so that each model can be trained

- Split between training and test
- Parameter hyper-optimization using grid search
- Training of the model in training through 10 Fold Cross Validation
- Model training with all training data and validation with test
- Comparison of results obtained both in training and in test
- As conditions, the split between train and test must be 80% - 20% and the validation metric must be R2

**Shap values analysis for the chosen boosting model: (1.5p)**
At this point you are asked to develop a complete analysis of Shap on the boosting model that you have developed in the previous point.

**Generation of predictions in the sales dataset and selection of the 20 most profitable and cheapest homes to buy: (1.5p):**
At this point you are asked to select the best model of those generated previously and to generate predictions about the home sales dataset following these steps:

Preparation of the flat sales dataset: The model you have generated admits data that you have previously transformed (Change of variable for the target variable, substitution of null values, unused columns...).
Calculation of predictions on the sales dataset
Calculation of the Break-even metric, resulting from dividing the price of the house by the number of square meters and by the prediction of the rental price per month.
Selection of the 20 best most profitable homes (lowest break-even) and cheapest (lowest purchase price)

**Final results, conclusions and open pathways: (1p):**
First of all, it is requested that you include the result of R2 for the validation dataset of your best model. Later, that you develop the following points: What limitations have we had with our model? Would we use it in real life? How can we improve the results obtained?

Comments:

Two files must be delivered: 1 Notebook in ipynb format with the code already run in all the cells and the same delivery in pdf format since some of the visualizations such as those of catboost are lost.
This exercise does not have a unique solution. Many options may be valid, but it is expected that an interpretation of the data obtained at all points will be made.
Obviously, there are certain variables that must be normalized based on the number of square meters:
Rental price must be converted to price per square meter
Number of rooms must be converted to rooms per square meter
Number of bathrooms must be converted to bathrooms per square meter
Maximum delivery date: March 20