## ***Recap & Conclusions:***

As mentioned in the introduction, the main goal of our project was to try and predict housing prices, with as much precession as possible.

Our predictions are based on using simple and complicated ML models on a dataset of households which contain several features with regards to each observation, such as the area (size) of the house, number of parking spots, furnishing status and so on.

The first stage of our project was Data Exploration. Our purpose was to understand the dataset better (mainly the features that were the most relevant to us for later stages of the project).
In that part we've shown how the dataset looks like (distributions of some features) and some key visualization because we believe that a picture is worth a thousand words (and a good visualization can be worth even more).

The second stage was all about Data Pre-Processing. The idea here was to prepare our dataset in such a way that the ML models we chose can preform on it, without losing the meaning of any feature. It's also important to mention that we've decided to predict the log of the housing price instead of the actual price, so we can see how a change in quantities of different features affect the housing price by percentages.

The third stage was focused on running and evaluating our ML models. We chose to perform Th models Random Forest, Multiple Linear Regression and Multilayer Perceptron. We performed them on the train dataset and evaluating them by the results we got from using the MSE scoring method, which fits very well the type of problem we're trying to solve. The ambition here was to pick the best way we can to predict the log price of the house while seeing how a change in each feature affects it, and trying to make the bias variance tradeoff accurate as possible.

The last stage was making a prediction. Now's the time to finally check out our results! i.e., comparing our predictions to the actual log of prices found in our test dataset. We took our best model and ran it on the test and gathered the results into a different csv file so it would be easier to make comparisons. We also believe that the addition of some more visualizations in this part can help you, the reader, to understand the strong resemblance between our predictions and the real values found in the test dataset.

A lot of effort was put into this project, but it's still not perfect. We think that trying to estimate such a complicated subject such as a price of a house based only on 12 initial features is obviously lacking in depth of analysis, and if we had more features (for example the country which generated the dataset, or the amount of crime that happens near the house) could have helped us make even better and more precise predictions.
Never the less, even though we are not real estate agents nor housing market experts, we do believe that our work can be very beneficial to any investor or potential buyer that has some basic knowledge/information on the house he's interested in.

We had a lot of fun making this project and we hope you'll find it as interesting as we did while working on it.