ASSIGNMENT 3: REGRESSION MODELS
20 POINTS

• This assignment will be solved in groups of **two** students. You must upload your solution (**html file**) at Aula Digital before the deadline. Be sure to included both student's name at the top of the document and upload only one solution per group.

• **Due date**: October 25th, 2022 at 23.55h

• **NO late assignments will be allowed.**

The goal of this assignment is to create a regression model that would accurately predict housing prices in Kings County, Washington.
To complete this assignment a years worth of data spanning from 5/2/2014 to 5/27/2015 will be used. You can find the king.csv file in the DATASETS folder. The variables are:

- id: Unique ID for a house
- date: Date day house was sold
- price Selling price of the house
- bedrooms: Number of bedrooms
- bathrooms: Number of bathrooms
- sqft_living: square footage of the home
- sqft_lot: square footage of the lot
- floors: Total floors (levels) in house
- waterfront: Whether house has a view to a waterfront
- view: Number of times house has been viewed
- condition: How good the condition is (overall)
- grade: overall grade given to the housing unit, based on King County grading system
- sqft_above: square footage of house (apart from basement)
- sqft_basement: square footage of the basement
- yr_built: Year when house was built
- yr_renovated: Year when house was renovated
- zipcode: zip code in which house is located
- lat: Latitude coordinate
- long: Longitude coordinate
- sqft_living15: The square footage of interior housing living space for the nearest 15 neighbors
- sqft_lot15: The square footage of the land lots of the nearest 15 neighbors

1.- THE DATA (6 points):
Get a feeling of the data using descriptive statistics and visualizations; study linearity assumptions and data distributions. What attributes contribute to higher housing prices? Decide which variables should be considered in a regression model. Reflect on the transformation of different features in order to obtain better models. Justify your answers.

2.- THE MODELS (10 points):
Perform a minimum of 5 different regression models (iterations) for price prediction and comment their validity. You should provide an iterative approach to modeling, give details on parameters and results of the model at each iteration. Whenever necessary, briefly explain the changes made from one iteration to the next, and why you made these choices.

3.- CONCLUSION (4 points):
Provide at least 1 paragraph explaining your final model. Pick at least 3 coefficients from your final model and explain their impact on the price of a house in this dataset. Give concrete recommendations for how to improve the selling price of a home.