

Homework 2 (EDA part 1)

EDA - NYC property sales (100 points)

Due Date: Tuesday Sep 14 at 11:59 pm

Instruction:

- Highly recommended: Watch the YouTube video for class 3 first!
- This is a group-work assignment!
- You are expected to submit the **.ipynb** file and the exported **.html** Only one member in each group needs to submit the assignment. It will be automatically submitted for the rest of group members.
- As we discussed it in class, you can find the answer key [here on Kaggle website](#)! I highly encourage you to first try it yourself without looking at the answer key.
- This is a long assignment, start early!

Question 1 As your first exercise I need you to do a complete EDA (Explanatory Data Analysis) on the NYC property sales data set. Once you are done with cleaning and restructuring the variables, you can save it as a new csv file and use it for your future regression exercises. We will revisit this data set later in the course.

Hint 1: make sure you explore other people's codes for this Kaggle competition. There are couple of good EDAs written for this data set on Kaggle website including mine :)

Hint 2: There is not necessarily a right or wrong answer for this exercise but I expect you to deliver something like what we did in class for the wage data set!

Where can you find the data set?

Please find the **.csv** file on the Kaggle website! scroll down to the **Data Explorer** section and you should find a csv file named as "**nyc-rolling-sales.csv**" 12.99MB

Have fun with Python and GOOD LUCK!