

Assignment 4 ECON7890

Instruction

You are being given a file hw.csv which contain the Property Transaction price of Kowloon between Nov 2014 to Nov 2020. The data was extracted from the "District Latest Deals" from <https://data.28hse.com/en/>

The header of the data file is quite strange. You may have to refer to the 28hse website and some other research. (e.g. GFA vs SFA)

You are free to use other data set, but it's optional. Using it will help you getting some bonus score. If you do it please upload the data file on moodle. Pls note that there's a limit on file upload of 200MB.

Question 1 - Import and cleaning data - 15%

Import the data to a pandas dataframe from the data file, clean up the data, check and handle missing values, rename column to a more meaningful name, etc.

Question 2 - Merge Data - 15%

Get some other dataset and merge them with this data set. e.g. unemployment rate, money supply, inflation, any other interesting data.

Question 3 - Exploratory Data Analysis - 25%

Do some exploratory data analysis and do visualization.

Question 4 - Learning model - 35%

Apply multiple statistical learning / machine learning to determine what kind of factors impact the property price (district? area? floor?), which should include training vs testing data split, training model. Evaluate and compare the model.

Question 5 - Findings and Conclusion - 10%

Show me the summary of interesting findings and conclusion of your analysis