



March, 2018

Data Science -- Skills Challenge

Imagine you are working on a project trying to discover how “green” areas in a neighbourhood affect housing prices and rental rates in New York City. Your Project Owner has given you an afternoon to spend on this task to come up with initial results.

Download following datasets from the NYC Open Data website:

<https://data.cityofnewyork.us/Housing-Development/DOE-Condominium-comparable-rental-income-Manhattan/ikqj-pyhc>

<https://data.cityofnewyork.us/Environment/NYC-Greenthumb-Community-Gardens/ajxm-kzmj>

Tasks:

1. Visualize the two datasets on a map. What can you observe?
2. What is the most expensive Neighborhood Tabulation Area (NTA) in New York? What else can you tell about this NTA based on the data?
3. How do the Greenthumb community gardens affect the value of the property? And what about rental income?

Notes:

Please use R or Python in this assignment. In addition to these languages, feel free to use any tools that you feel will be helpful. Please submit results as a file, preferably markdown.

You should provide the analysis showing your assumptions, workflow and results, with all required steps in a neat package (assume the target person is skilled in data science with good knowledge of the real estate sector).