ECON4170 : Project Proposal
Zoé JULLIARD--BONNOUVRIEE

Project : Predicting One Family Dwellings Sale Prices

This project focuses on analyzing how the size of one-family dwellings (measured in gross square feet) influences their sale prices in New York City. We will use data from all five boroughs—Bronx, Brooklyn, Manhattan, Staten Island, and Queens—sourced from the NYC Department of Finance's official "Rolling Sales" datasets, covering the period from September 2023 to August 2024.

The goal is to build linear regression models that explore the relationship between dwelling size and sale price both across the entire city and within each borough. In addition, we will identify and address any erroneous or outlier data points that could distort the accuracy of our models.

The project will aim to answer two main questions:

1. *Citywide Analysis :* How well does the size of a one-family dwelling predict its sale price across New York City?
2. *Borough-Level Analysis :* How does the relationship between dwelling size and sale price vary across individual boroughs?

Analysis Structure :

1. *Data Import and Cleanup :* We'll begin by importing the datasets for each borough and cleaning the data, including handling missing values and standardizing variables.
2. *Bivariate Relationships Exploration :* Using scatterplots, we'll explore the relationship between sale price and gross square footage, examining the direction, linearity, and strength of the relationship. We will create plots both for all boroughs combined and for each borough individually.
3. *Outliers and Data Integrity :* We will investigate potential outliers that could distort the models, especially those caused by data entry errors. Erroneous data will be removed before modeling to ensure accuracy.
4. *Citywide Linear Regression Model :* A linear regression model will be built to analyze the relationship between dwelling size and sale price for the entire city.
5. *Borough-Level Linear Regression Models :* Separate regression models will be created for each borough to assess how the size-price relationship varies across different areas of the city.
6. *Conclusion :* Using regression results and summary statistics in R, we will conclude on the strength of the relationship between dwelling size and sale price, both citywide and for each borough individually.