Housing Data Guide

**What are you looking at?** [A list of almost 3,000 affordable housing properties](https://drive.google.com/open?id=1skUQMViBnpkgx4G2391HZ1CZj5hWTmuGsJxrRRPvNZk) in Manhattan, along with statistics about their buildings, subsidies and the neighborhood.

**Where we got this data:** [coredata.nyc](http://app.coredata.nyc/?mlb=false&ntii=&ntr=&mz=14&vtl=https%3A%2F%2Fthefurmancenter.carto.com%2Fu%2Fnyufc%2Fapi%2Fv2%2Fviz%2F691a2b7c-94d7-46ac-ac4d-9a589cb2c6ed%2Fviz.json&mln=false&mlp=true&mlat=40.718&ptsb=&nty=&mb=roadmap&pf=%7B%22subsidies%22%3Atrue%7D&md=map&mlv=false&mlng=-73.996&btl=Borough&atp=properties), New York City’s housing and neighborhoods data hub, presented by the NYU Furman Center. This subsidized housing data pulls from multiple sources (various city and state agencies and HUD), and the data isn’t formatted the same for each source. Furman tries to clean it up and match entries, but this data is a little messy.

**NOTE: We have altered many of these datasets for instructional purposes, so please download the data from the original source if you want to use it in your reporting.**

***Before you begin*:**

Take a look at the different categories in this table (always a good idea with any dataset).

* Each row shows a **project name** (if there is one) in column B, a number of **buildings** (shown in column C) and the total number of **units** (column D).
* Some projects involve **multiple tax lots** identified by BBLs (borough, block and lot numbers). So the same project name could have more than one entry.
* There are 3 fields with dates:
  + The year the building was built (year\_built)
  + The year the housing subsidy started (end\_date)
  + The year the housing subsidy ends (start\_date)
* If the “end\_date” is blank, it means there is no foreseeable expiration date for the subsidy.

Without doing any analysis yet, what are three questions you might try to answer using this dataset?

1. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Some questions to answer:**

1. What are some things you notice about this dataset that shows it isn’t perfect?
2. What are the total number of units in the Upper East Side?
3. When were most of these buildings built?
   1. Now make a chart of that data.
   2. Generally, how long after construction did the subsidies kick in?

Hint: keep in mind some of the gaps in this dataset.

1. Which neighborhood has the most units?
   1. Compare the number of units in each neighborhood to the neighborhood’s population, median income and 2016 racial diversity. Is the neighborhood with the most units exceptional in any of these categories?
2. Which neighborhood has the most units that will expire this year?
   1. The majority of units expiring this year are affected by the expiration of which subsidy program?
   2. Which neighborhoods will be most affected (in terms of number of units) in each of the next 10 years by expiring subsidies and which one is the most secure?
3. Which owner has the most *buildings*? Order your table so the list of owners is descending based on the number of buildings owned.
4. Who are the biggest owners in each neighborhood based on the number of units? Limit your table to owners who have more than 1,000 units.
5. Which owner has the most tax delinquent projects? Which one has the most serious housing code violations?
6. Compare the average value *per unit* for different neighborhoods. Which is the most expensive and which is the cheapest?