**Damion Brown**

**CIS 2300  
Professor Feng  
12/17/2023**

***Final Report***

***Motivations***

Most places in the US are in a massive housing price crisis currently, especially large cities like NYC. Buying a house, or even renting, is out of reach of more and more hard-working people. I wanted to see if the NYC government was taking steps to help address this by looking at affordable housing data.

***Questions***

Has the amount of affordable housing being built been increasing or decreasing? Has the time to build it been increasing or decreasing? Where is this housing located? Which borough is the most family and/or senior friendly?

***Methodology***

First, I calculated the percentage of housing units that were unfinished. The amount of affordable housing in NYC is 41.45%. I cleaned the data by dropping null values from the project completion date column to get only finished housing projects. I wanted to look at just finished housing to get a complete picture of the breakdown of housing units that were already built. Then I converted the project start and end date columns to datetimes instead of strings so I could use python to do time calculations based on them. Using those date columns I added a new column of the amount of days it took to finish each project. This allowed us to calculate an average amount of days to finish projects based on the start year. I also utilized Geopandas to visualize the different postcodes, boroughs, census tract and NTA-neighborhoods using shape files through NYC open databases.

***Findings***

I discovered several things, the first being that the time it takes for affordable housing projects to finish has been slowly decreasing. It hovers between roughly 242-424 days between 2014 and 2021 but plummets to 16.06 days in 2023. However, Projects that take 0 days to finish have not increased wildly and are only 50-90 projects per year while projects that take more than 0 days to finish have been increasing and will spike in 2019 and 2021. Still, when I removed projects with a duration of 0 days, I saw an almost perfect decrease in the amount of time it takes on average to finish a project, even including 2015. This points to the city becoming more efficient in constructing affordable housing

The second is that the amount of units being built has been increasing. Starting in 2019 the amount of units coming online per year jumped from roughly 2,560 to almost 8,8498, peaking in 2022 with roughly 15,581 units being finished. I hypothesize this is due to the economic downturn caused by the COVID-19 pandemic.

I generated a correlation table and discovered a strong positive correlation between Counted Homeownership Units and Very Low Income Units. I also discovered that as the number of bedrooms increased the less units contributed to the overall unit count.

Damion discovered that extremely low income, very low income, and low income units were concentrated in the Bronx, Manhattan, and Brooklyn. Moderate income units are concentrated in Manhattan with an almost even amount in Brooklyn, Queens, and the Bronx, and middle income units are concentrated heavily in Manhattan and Brooklyn. He also added columns to total all the low income units and find what percentage of each project was low income and display them by the highest percentages, however, the projects with the highest percentages were only 1 unit.

Jack built code to see what areas and projects were the most family and senior friendly. He discovered that the Bronx, Brooklyn, and Manhattan are almost tied for most “family friendly” borough and Brooklyn had the most senior units. Family friendly units are buildings that have 2 or more bedrooms and adding all those units together and grouping the results by borough is fascinating. I was surprised that the Bronx, Brooklyn and Manhattan were closely distributed in a number of family friendly units. I also discovered that family friendly units accounted for 34.25% of all affordable housing, and that senior housing accounted for 5.23% of total housing. By filtering the dataframe to only have completed projects with the word “SENIOR'',“Senior”, “Elderly”, or “ELDERLY” in the name I found 14 completed projects dedicated completely to senior housing, however they were not the projects that had the most senior housing by quantity or percentage (Though most are close to 100% senior housing).

I also utilize the Python’s Geopandas package to visualize the different populations, zip codes and Council Districts.I found that Staten Island was the least populated borough because the original dataset had the lowest entries.

***Challenges***

I faced three main challenges. Getting code assembled into a single place, getting the same code to run on everyone’s computer, and finding a dataset with everything I needed. I used a google colab notebook to assemble our code into one place. The geopandas code only works in one place, Jack’s computer. I have included it in the submission separately with the hope that it will work for other users, however it temporarily broke Wylie’s pandas installation, so try with caution. The data of housing by project had the data about senior units I needed, while the data of housing by building had everything else, so I simply had to import the two data sets to solve this. I had difficulty with GeoPandas and displaying the correct Legend. If the datatype is a string, then the legend will display all the data points because it treats them as categorical data. However, if the data type is float/int, the legend will display a color gradient. I solve this issue by converting some data into integers to make it easier to visualize.