Danny Cabrera

Data 205

CRN 202530

Prof. Perine

5/11/25

Rental/Homebuyer Affordability, Trends, and Impacts

In my project I decided to take a look at the rental/homebuyer market. I wanted to look at the trend of many people my age (gen Z) saying they can’t afford to buy a home or rent an apartment. For many they see it as a crisis others say to not worry about it and as for me prior to this project I did not know which side to lean towards. I researched and got my answer. The data I used for this project is compiled from many different sources (US Census Bureau, Zillow, Realtor.com, Data Montgomery, DC.gov). All of these sources are government sources bureau in charge of providing data for many different economic and socioeconomic factors expect Zillow and realtor.com these are not affiliated with the government. These sources were used to create visuals for rental vacancies, income, home prices, monthly rents, construction permits. I only excluded labor force participation and homeownership rates, I planned to due something with them but I ran out of time.

The importance of doing all is to try and show what exactly is going on with the housing market. More specifically what’s the problem with affordability. As stated earlier many in gen z and millennials feel that housing is a big issue in their lives. Everyday it feels like the dream of owning a home gets further away from those of us who are younger. For the generations before us owning or even renting a home was still a challenge but not the same challenge it is today. Overtime these things change and nothing stays the same but to say these challenges don’t have solutions would be an understatement. With many factors leading to those outcomes including the ones I’m mentioning in this project. Also it would many in the older generations better understand what the current generation is facing.

Some of the tools I used for this project was using python for data cleaning and manipulation. Using libraries like pandas and numPy helped me tremendously with all the mess that was the data. Plus, libraries like matplotlib and seaborn helped a good chunk with all the math involved. Included in that is the statistical analysis which without sci.py stats and pandas again saving me when doing these analyses. I did a correlation analysis with t test to test which columns in the data had the most correlation and showed to be the least likely hood of being random. This showed a positive correlation in the categories mention above (rental vacancies, income, home prices, monthly rents, construction permits).

The most annoying part of any coding project is the data cleaning. There was a lot this was many due the same data source having different departments for different topics and that said department formatting there data different than the other departments. What I mainly focus on was removing all symbols from all numbers no matter what. I kept getting errors on data points because 546,000 had a dollar sign before it or 7.6 had a percentage sign in front of it. For this reason I split the data into 3 separate categories all national data, Rockville data, and DC data. This eased the transition to create calculations. For example I decided for the fun of it to create a rent to income and a income to rent to show how much of a person income is going to rent in a given year and whether it is better to buy a home or rent a home. In this example in Rockville you pay 37% of your income in rent and price to rent ratio came back at 19.85 meaning it’s barely better to buy a home than rent. While in DC 42% of your yearly income is going to rent and the price to rent ratio is 23.3 meaning it’s better to rent than buy.

Before I go any further the most important thing to understand from this project is what are the categories/columns I’m using. Rental vacancies are when a unit is completely empty aka no one lives there or aka vacancy. For many landlords this may last months at a time or be as short as a few days. Income is pretty self explanatory. Home prices are the average prices homes are going for in that area. Monthly rents how much is the average rent in that area. And lastly construction permits this is the amount of construction permits being approved for new private residential units.

After taking all of this into account we find that yes the idea that rents/home prices are unaffordable for anyone really doesn’t have to be just gen z. How did I get there by looking at many different variables. To start I looked at the data as a whole I looked at the rental vacancy rates overtime and found that is not something that stays stagnate it fluctuates a lot over the course of the last 70 years. From there I knew that and decided to take a closer look at which time frames most make sense and can be presentable to show a trend. I found that although a history of this would be great I was limited time to present so I decided to move it to today and show the residential construction permits being approved nationally for 2024 to find a cause for this of how vacancy rates are affected by other factors. In order to show this data I had to rename columns, remove the extra spaces, and any extra lines that they had around the data. Relabel all the months because they weren’t displaying correctly. Plot each series so the graph would look appropriate. After all that the graph worked just not as intended so I manually input the data instead. Then I broke it down even further It wasn’t good to just look at the country as a whole the united states is just too big. The northeast, Midwest, south, and west were the four regions I chose to make for this graph. It worked well without to much data cleaning from me I would say it more of a data addition by adding the regions to each state. But I wanted more so I compared the regional data of each state vacancy rate into quarters. It created this cool visualization that compares the housing permits approved and the vacancy rates by regions. I found that the Midwest out of all places could not meet demand (not enough permits are being approved or lack of permits) and has a high vacancy rate. Knowing that I wanted to keep searching further and potentially that income and rents had something to do with that. Again I broke it into regions and looked at the average income after taxes and average rents in that area. For this graph I made an assumption that everyone in the country has a tax rate of 25%. Which highly varies on many factors (married, children, homeowner, w-2 form, 1099 form, etc.). Overall the country is spending more than the recommended amount on housing which is a worrying factor as prices only seem to go up. I could have gone further looking at different socioeconomic factors I decided to take it closer to home and compare two cities closest to us.

Rockville and DC fit this criteria perfectly. One’s a major city and the other bigger city in the state. I wanted to explore the different price points for both median sales and rental prices to show what really is the difference between buying and renting as the graph shows it is a pretty big difference for buying a home in Rockville or dc. Whereas for renting they are similar unfortunately Rockville is catching up to dc in that regard dc is still more expensive but by a lesser margin. With everything being so expensive I thought it might be the prices making many not want not want to live in these cities but surprisingly the vacancy rates for both cities is normal and that would suggest that many still want to live in these cities and that the demand for housing is being met. To better show what an individual feels in there pocket every month. There’s a lot of money being spent on housing here. At least over 35% of an individual income is going towards rent every single month. And still trying to buy home we’ll forget about it. At the national rates for both first time down payments (9%) and the average rate of saving (3.9%) it would take you at least 17 years to afford the down payment for a place.

With all that doom and gloom I did manage to create a prediction model for rental vacancy rates for the year of 2024. Thankfully out of everything this has gone down. Not much only about .4% for dc and .3% for Rockville but this does mean that home prices have increased. For the prediction model I used the categories of median rent for a 1 bed apartment (because it is the most common type of building being constructed over 70% of new constructions are 1 bedroom apartments ) and the median sale price of a home in these two areas. I used sklearn.ensemble package to use the random forest regressor to train the models. I found that the importance of both models came back to 90 meaning that there is correlation and there is relevant information being shown. Plus I added in 3% increase to the home purchase price because it was the historical trend of the data over the last 5 years. The results again are that the rental vacancy rates are lower than the year before continuing trend ever since Covid of rental vacancy rates going down year over year. Potentially had I removed the increase in home sales the outcome could have been different.

Before all this I did create a heatmap showing a correlation matrix with all the categories in the data for both cities. I found that some categories had very low correlation, surprisingly rental vacancies rates was one of them it didn’t correlate well with the majority of categories expect one homeownership rates. I did try to create a graph showing the differences between the two although it did make for an interesting graph I decided to cut it out as I thought I would have enough time to explain it in the final presentation. Also it wasn’t what I was really going for. With that said every other column have a high correlation with each other (median sale price, avg income, rent for 1 bedroom, years to down payment).

Interestingly doing a t test on this dataset gave me a different type of result. It only found a significant differences in the home sales price, homeownership rate, and rent to income ratio. One reason is a home in dc is $150k more than in Rockville no wonder the the homeownership rate is 40 compared to Rockville 64. Although it is true that these have the biggest difference they aren’t the only ones. The income is a difference of $4000 a year between Cities, and a difference of rent of $150 a month.

Finally, after all this work put into this project I can remember when this started only as an idea and nothing else. This wasn’t my initial goal, but it turned into it. At first, I remember looking at just national data sets because I didn’t wanna focus on anything to local but as it evolved it turned into that. It was fun to look at many different aspects of the housing market. I just wish I had more time to explore these these categories further to create an analysis better suited to the market as a whole for both national and local data. I’m glad I accomplished what I accomplished because it helped me submit a better understanding for data analysis projects. Plus, I found that it wasn’t so boring looking at governmental data normally in the other classes I have taken for these types of projects. I would’ve done something sports related because it’s what I’m really passionate about and I found all types of other data like healthcare or something related to education or governmental data, pretty boring and not fun to analyze because it felt more like a chore than it felt to be intrigued by the actual results that you found from the data, but this project was very different in that way. I just wanted to thank the whole data science (prof. Saidi, prof. Alaree, prof. Perine) and math department (Dr k. , prof. Muhammad) for being there at every step of the way.

**Acknowledgements**

**Data Sources**

U.S. Census Bureau

* [American Community Survey (ACS)](https://data.census.gov/)
* <https://www.census.gov/construction/nrc/index.html>
* <https://www.census.gov/housing/hvs/index.html>
* <https://www.census.gov/housing/hvs/index.html>
* <https://www.census.gov/programs-surveys/acs/data.html>

Zillow Research

* [Zillow Observed Rent Index (ZORI)](https://www.zillow.com/research/)
* [Home Value Index (ZHVI)](https://www.zillow.com/research/zhvi-methodology-2019-high-26244/)

Realtor.com Market Data

* [Monthly Housing Trends](https://www.realtor.com/research/)

Rockville Data

* <https://data.montgomerycountymd.gov/>

DC Data

* <https://opendata.dc.gov/>

**Professors**

* Prof. Saidi
* Prof. Alaree
* Prof. Perine
* Prof. Mohamed
* Prof. Mohammad
* Prof. Katiraie