# Digital Arch. Final Project Report: Rent Pricing Analysis

Braxton Haynie
University of Tennessee
Dept. of Electrical Eng and Computer Science
bhaynie@vols.utk.edu

Karan Patel
University of Tennessee
Dept. of Electrical Eng and Computer Science
kpatel68@vols.utk.edu

Jacob Hawkins
University of Tennessee
Dept. of Electrical Eng and Computer Science
jhawki41@vols.utk.edu

Matthew Rosenbalm
University of Tennessee
Dept. of Electrical Eng and Computer Science
rrosenb4@vols.utk.edu

## I. Objective

Housing is a crucial cornerstone of our society and greatly impacts the standard of living. Whether it is a house or apartment, having somewhere to live that is affordable and in a good location is pivotal to the rest of ones life. Housing is commonly an individuals highest expense. Therefore, it is crucial to be aware of the trends the market is taking to help individuals make better decisions regarding the apartment they plan to rent with their hard-earned money.

The housing market is constantly evolving: styles, sizes, locations all play a part in how an apartment is priced. Understanding how the market is changing, allows consumers to make better decisions regarding renting an apartment, understanding the true value of an apartment, as well as what to expect in the near future. The data we collect will hopefully help us answer important questions such as what apartments are within my budget? What locations best align with my needs and budget? What styles and sizes impact the price the most? Our aim is to help answer these questions through a data-driven approach to provide users with all the information they need to best prepare them for one of the most important investments of their lives.

The objective of our final project is to analyze and extract trends about rent prices for various areas. Additionally, we hope to develop a model which could be used to estimate the rental price for a given home in a specific area to make sure that it is a fair value. We plan on analyzing real and current data from Zillow by obtaining currently available listings and utilizing that data to discern relevant trends that impact current rent prices. We plan to explore the impact of location, house size, number of bathrooms, etc. to draw conclusions about rent pricing. For example, looking into how the location of an apartment has impacted the price, if the size of an apartment is more or less important, etc. Being able to log the various trends in the industry would provide useful insights moving forward into the future.

## II. Methodology

In this project, we used web scraping as the primary form of data collection. We were able to pull from Zillow information about current listings in user-specified zip codes. We gathered features such as longitude and latitude, square footage, number of bedrooms and bathrooms, address, amount of time listed on Zillow, and price. Because we are using web scraping to collect the data, we had to create a proxy rotator to prevent the scrapper from being blocked by Zillow.

Once all the data was found and collected, we began analyzing it. We used the data to identify key trends in rent prices in relation to location, square footage, and features (such as number of bedrooms or bathrooms). We also created a machine learning model (neural network), trained on the data scraped within a zip code, to be able to predict the price of an apartment/house based on size, location, and features (i.e. number of bedrooms and bathrooms). The model is able to predict the price within a three-hundred dollar range.

Lastly, we used Python to perform the web scraping for the needed data. Python also helped us better understand the data through visualization by various graphing techniques.

### A. Data

In order to get the data required for us to perform the analysis of rent pricing in the country, we looked at a wide range of data sources. Our initial plan was to utilize APIs to retrieve the data based on zip codes and dynamically analyze them so we didn't have to store them within a database. However, we soon discovered these resources were locked behind a paywall that prevented us from utilizing this method. We next looked at curated datasets that were available free of charge. However, the issue with this method is that one, they are not updated for recent markets, and two, there aren't enough sample points for each zip code (some zip codes aren't even included) for us to make meaningful conclusions. Therefore, we decided to create a web scraper for Zillow, a popular real estate website that provides millions of for-sale and rental listings. The appropriate web URL is curated based on the zip codes the end user is interested in which is then used to extract the longitudinal and latitudinal coordinates from Zillow for each of the respective zip codes. The next step is to take these coordinates and create another appropriate Zillow URL that provides all the listings within the area. The HTML source code is scraped to pull the necessary information for each listing such as the price, address, number of bedrooms, number of bathrooms, area, and time on Zillow to list a few. In order to prevent our bot from being blocked by Zillow, we created a proxy rotator that periodically changes our IP to make it more difficult for the bot to be located. Finally, the listings for each of the zip codes are stored in a database.

This makes it easier for us to pull the respective data in the visualization step and training of the machine-learning model in the future without having to make unnecessary calls to the Zillow website and risk getting blocked.

### *B. Models/Algorithms*

A key deliverable we set out to meet for this project was to develop a machine-learning model that predicts the rental pricing of a particular zip code based on parameters such as the number of bedrooms/bathrooms and the area of the apartment. This will allow users to get an idea of what to expect to pay for an apartment meeting their requirements in a particular zip code they are interested in living in. This will help users in the decision-making process of determining their living situation. This deliverable can be viewed as a regression model that takes the aforementioned input parameters to produce the output – rent pricing. In order to create this model, we employed the deep-learning approach that utilizes layers of artificial neurons that can estimate any non-linear function. We used the TensorFlow framework to create a neural network comprised of 5 layers (not including the input layer) with 256, 128, 64, 32, and 1 neuron(s) in each layer respectively. Additionally, the ReLu activation function was utilized with the RMSE loss function for the back-propagation algorithm. Depending on the zip codes the user is interested in, we take all the listings for the zip codes and separate the inputs (number of bedrooms/bathrooms, area, zipcode) and the output (pricing). The inputs are then normalized and the zip code input is encoded using the one-hot-encoding method in order to evenly weigh each of the parameters to allow for stable training. Finally, the network is trained for 150 epochs to get the end model.

## III. Results

We began our investigation of the data by examining a zip code in our local area, "37920". Utilizing Python, we looked at correlations between the area, number of bedrooms, number of bathrooms, and the price of apartment listings for this zip code as seen in figure 1. Expected correlations were observed, such as the correlation between the price of an apartment and its area and number of bedrooms. To build on this observation and to continue our search for more meaningful trends in our data set, we began to examine the time feature of our data objects. Once we converted our time data from milliseconds to days, we were able to observe a small correlation between the time an apartment's listing was active on Zillow, and the price of the listing. Looking at figure 3, we can observe a noticeable decline in apartments listed at a price above $3000 after around 60 days being an active listing. At this point in our analysis, we became intrigued by the occurrence of these outlier priced listings and wanted to better understand the distribution of rent pricing geographically for the zip code "37920". Figure 2 shows the distribution of rent prices geographically, categorizing price ranges by color. We expected to see a strong decrease in rent prices as apartments moved away from the city center, but this was not observed. Here, we decided that the zip code "37920" did not have enough apartment data. Consequently, we began to look at prices across multiple zip codes in areas with greater abundance of apartment listings, leading us to New York City. Here, we began our rent pricing analysis by examining the average rental price for each zip code in New York City.

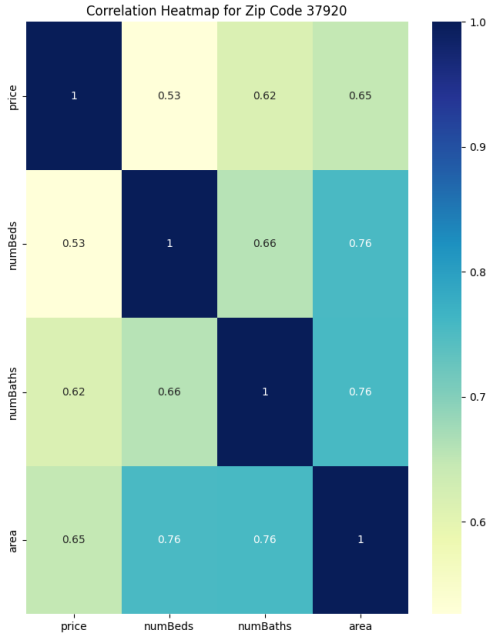


Fig. 1. Correlation heat map of zip code "37920".

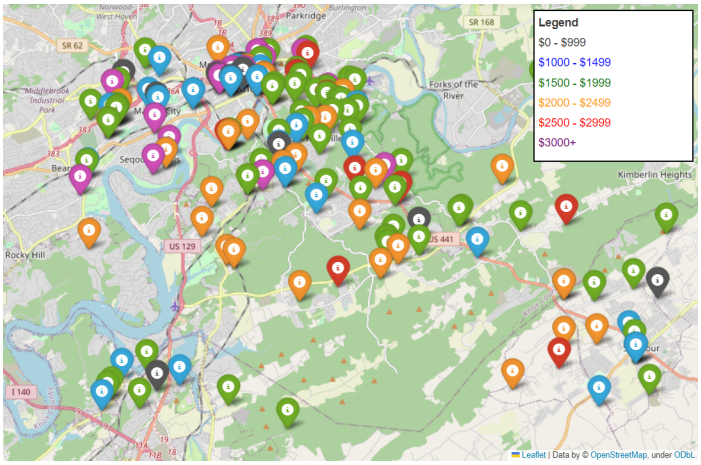


Fig. 2. Geo-spatial map of apartment listings in zip code "37920".

Figure 4 shows the results of the average rental price for each zip code in the New York City area. Note that prices shown for each zip code in the map was found by taking the average of all the rentals across the area. This map is not entirely surprising as far as what area is the most expensive, but it does still have some interesting insights into the area. The most expensive area is lower Manhattan which is notorious for being tremendously expensive. What is interesting to see is how rapidly the average rent price drops

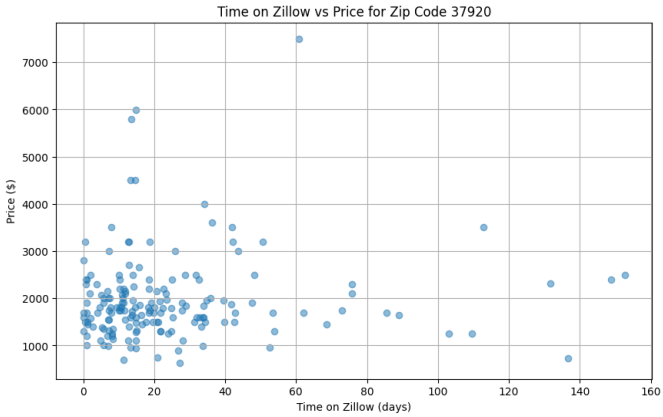


Fig. 3. Scatter plot of time(days) on Zillow vs. price for zip code "37920".

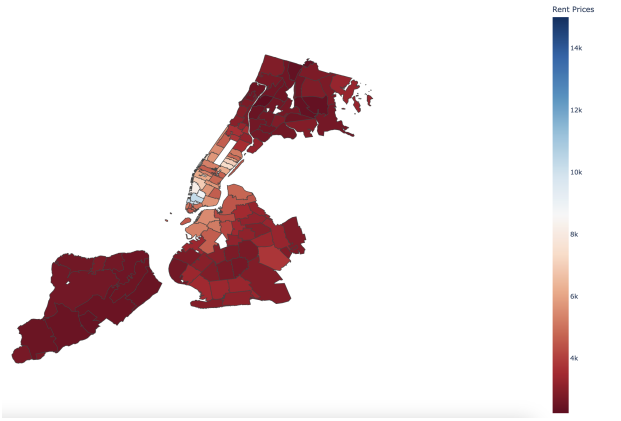


Fig. 4. Map shows a choropleth map of the rent prices for all the zip codes in New York City.

from being around 11k/month to being around 4k/month within just a few zip codes of each other. Another interesting thing to note is how the average rent price drastically drops when looking at the entirety of Manhattan compared to Staten Island or the Bronx(North most area) of the city. The average rents in the Bronx and Staten Island appear to be relatively similar in average price sitting around 3k/month.

When comparing Figure 4 to some crime statistic of felony assault, robbery, and murder found in Figures 5, 6, and 7 respectively, it is interesting to see how the rent prices match compare. Assault and robbery are a bit more difficult to pull information from as it seems like the locations might be more tied to police precincts rather than the true location of the crime possibly, but it is still interesting to see that the number of robberies seems to be highest in lower Manhattan where rent is the highest in the city on average. Figure 7 is the most interesting as it shows reports of murder across the city. This figure seems to align relatively close to the average pricing map, Figure 4, by showing that areas where murder is more prevalent the rent prices tend to be on the lowerside of the pricing spectrum. There are a few exceptions such as Staten Island where both rent and the murder rate are low as well as an area of Brooklyn where the murder rate is high and the rent is also fairly high, around 6k/month.

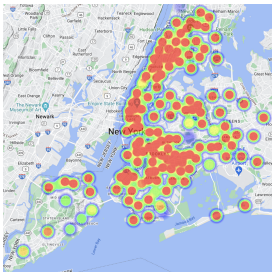

Fig. 5. Figure shows a heatmap of felony assaults.

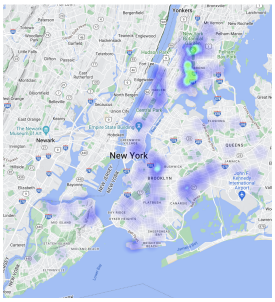


Fig. 7. Figure shows a heatmap of murders.

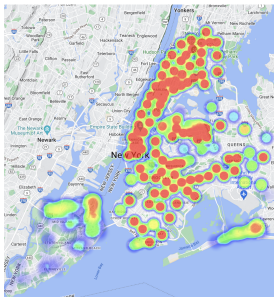

Fig. 6. Figure shows a heatmap of robberies.

## IV. Issues

There were a few issues that were encountered throughout the development of this project mostly relating to web requests. Zillow has established certain guidelines regarding web requests to their website and greatly discourages the use of web scrapers on their platform. Zillow seeks to block any usage of webscraping through the use of CAPTCHAs and other methods such as throttling and blocking IP addresses. Throttling and IP blocking was solved with the proxy rotator, but we were not able to get around CAPTCHAs. Another issue with Zillow was when trying to get historical pricing for homes. It appears that this information is not sent directly to the user but rather somehow rendered on the page through the usage of JS and security keys of some kind. We were unable to circumvent this issue in the time given unfortunately. The lack of access to this data greatly hindered our ability to extract meaningful correlations from our analysis.

We also ran into a small issue of running our scrapper and model on the DA machines. We were not able to connect to our own database as it was blocked from the docker container. We were able to overcome this issue by running the system on our personal machines.

## V. Future Work

In the future, this work could be expanded to perform price projection to predict what rent pricing will be in the future. In addition, this work could be expanded to include historical rent pricing of apartments/houses. This would allow a more accurate projection and would allow additional analysis to be done.

We can also work to create a more accurate predictive model that takes size, location, and features (i.e. number of

bedrooms and bathrooms) of an apartment/house. Currently, the model can predict the price within three hundred dollars of the actual price. This could be improved by restructuring the neural network by adding more layers and changing the ratio of the number of layers and the widths of each layer.

## VI. Timeline and Responsiblities

### A. Braxton Haynie

NetID: bhaynie
Responsibilities:

- Developed proxy rotator
- Helped with data visualization

### B. Karan Patel

NetID: kpatel68
Responsibilities:

- Developed Zillow scraper
- Developed ML model for price prediction

### C. Jacob Hawkins

NetID: jhawki41
Responsibilities:

- Integrated data into database
- Helped clean up the data

### D. Matthew Rosenbalm

NetID: rrosenb4
Responsibilities:

- Developed visualization module
- Modeled data to help identify key trends

## VII. Timeline

| | |
|---|---|
| Week 1 | Develop project idea and proposal. |
| Week 2 | Start work on data collection methods. |
| Week 3 | Keep working on web scraper and other tools. |
| Week 4 | Collect data and finish web scraper. |
| Week 5 | Begin writing scripts to analyze data. |
| Week 6 | Continue writing scripts. |
| Week 6 | Begin testing and training models. |
| Week 7 | Complete testing. |
| Week 8 | Present findings. |

TABLE I
Table showing the project timeline

## References


[1] Zillow, "Listings API - Data & APIs," Listings API, https://www.zillowgroup.com/developers/api/mls-broker-data/listing-api/

[2] "Crime map!," NYC Open Data, https://data.cityofnewyork.us/Public-Safety/Crime-Map-/5jvd-shfj (accessed Nov. 22, 2023).