# Williamsburg Landlord Database – Final Writeup

**Nico Rapallo**
Data Science, William & Mary
`NFRapallo@wm.edu`

## 1    Introduction and Background

The goal of this project is to create a relational database of all Williamsburg landlords and the properties they own. Additionally, this project allows individuals to provide information about and a rating of their rental unit and landlord, which is also stored in the database. Users interact with the database through a dash web app.

This project was motivated by my difficulty with Williamsburg landlords and conversations with other renters around campus. As having trouble with landlords is not uncommon in Williamsburg or in general, and landlords do not typically have review sections on their websites that allow tenants to express their honest feelings, I thought it would be useful to me and to others for there to be a place where tenants can leave feedback and where prospective tenants can see that feedback.

## 2    Data and Methods

### 2.0.1    Data Overview

The data for this project come from three sources: (1) Williamsburg City Government Parcel Data, (2) State Corporation Commission Clerk's Information System, and (3) Survey data. All final cleaned data will be available on my GitHub repository.

### 2.0.2    Williamsburg City Government Parcel Data

Williamsburg City government publishes parcel data that includes property ID (PID), owner name, address, value, land use, and other information for each parcel. This can be found in the Real Estate Assessment section of the city's website under 'Fiscal year evaluation results.'

The city also keeps property information cards for every parcel in pdf forms. These cards contain additional information such as the number of bedrooms, number of bathrooms, gas or electric heating, air conditioning, a picture of the property, as well as other information. The cards can be found on the city's interactive GIS page. They can also be found using the PID of a parcel by inserting the PID number into the link: "https://gis.williamsburgva.gov/vision/PID.pdf." I used the pypdf and Spire library to scrape the number of rental units, bathrooms, and an image of the house from each property card.

### 2.0.3    State Corporation Commission Clerk's Information System

One problem with the city data for the purposes of my project is that the owner of many rental properties was often a company that only existed to own that property. For example, the single-family rental property at 427 Roycroft Street is owned by 427 Roycroft LLC. To solve this issue, I use the State Corporation Commission (SCC) clerk's information system to look up the registered agent (RA) of any company that owns a property in Williamsburg. I called the SCC clerk's office and they could not provide me with a dataset with all the companies in Virginia. The easiest way to get this data was to download the information for all companies based near Williamsburg, Norfolk, Richmond,

Virginia Beach, Herndon, and Arlington, as well as any company in Virginia registered as 'General.' Then, I manually searched for the remaining few properties and created a CSV with their information. I could not figure out a way to easily automate this. After merging all of these CSVs, I had a list of companies and corresponding RAs that contained nearly every company that owned a house in Williamsburg.

### 2.0.4 Survey data

Finally, I obtained data through a form given to William and Mary students who live off-campus in Williamsburg. The form gathered data about how residents feel about their landlord, how they feel about their rental, and how much they pay in rent. The form also asks residents who manages their property / who they pay their rent to. This information could be used in the future to help connect properties through property managers in case different owners rent their houses through the same property manager. This survey data is used so there is some initial data in the database, but all future responses will be recorded through the dash interface.

### 2.0.5 Data Cleaning

The data required a significant amount of cleaning. This was the most significant time expense in the project. Williamsburg landlords do not practice standardized naming conventions across all the forms where their names and companies appear. There are often unpredictable capitalization, punctuation, middle initials, prefixes, and suffixes. I used the Levenshtein library to assist with this data cleaning project. I specifically used Levenshtein ratio, which is a measure of string similarity calculated by checking the number of edits required to make two strings equal. I wrote a function that ran through every property owner name, and prompted me with two names if the Levenshtein ratio was ever greater than 0.7. If the names should be the same I entered a 1, and the function replaced all instances of the less used name with the more used name, and if they were very similar names but different people, I entered 0 and the function passed on to the next set of names. In this way I was able to partially automate the manual data cleaning, and saved both the before and after data-frames for the sake of reproducibility. All the final cleaned data are available on my GitHub repository. The original and intermediary files that include the company information are available upon request but are too large for GitHub. To clean and manipulate the data, I used the pandas and RegEx libraries.

### 2.0.6 Data Storage

To store the data after cleaning, I am used an SQLite database. I decided to use a relational database because it is an efficient way to store the data and maintain the integrity of the data. I also wanted more practice creating an SQLite database after taking Databases last semester. Additionally, both SCC and the City of Williamsburg appear to use an SQL database to store their information, so it would make sense for me to also use one for the sake of consistency. To create the database, I am using the Python SQLite3 module. I chose to store image paths in the database and the images themselves in the local file system.

## 3 Dash Web App

To interact with the database, I created a simple dash web app. This web app allows users to select an address, and see a picture of the property along with property details taken from the city data, as well as ratings and reviews from other respondents.

It also allows users to click an "I live here" button, which reveals a form for the user to leave a review, which will be inserted into the database.

### 3.1 Limitations

This project currently faces a few limitations. While I was able to connect many different companies to the same people, I believe there are some that have been missed. Additionally, the database only includes addresses officially listed as rentals, multi-family, or apartments, so any property that is incorrectly labeled as not being a rental is not included in the database. For apartments, mailing addresses are not always the same as the address listed on the City of Williamsburg property card,

**Williamsburg Landlord Dashboard Address Lookup**

Enter Address   215 MONTICELLO AVE     × ▾   [ I live here! ]

**Property Details**

Owner: INCORP SERVICES, INC.

Owner Rating: 3.67 / 5 Stars (3 reviews)

Property Rating: 3.67 / 5 Stars (3 reviews)

Bedrooms: 0

Half Baths: 0

Full Baths: 0

Rental Units: 240

Reported Total Rent: 2,500 - 3,500

Reported Individual Rent: 1000 - 1,500

Utilities Included: All

Free Parking: No

Most Recent Property Review: Fairly new, can't really complain

Most Recent Owner Review: Fine experiences so far

Figure 1: Image / Details View

which is a potential source of confusion. For example, Midtown is stored in the database under 215 Monticello Ave instead of 221. Sometimes, multiple houses or apartments will be listed under the same property ID if they are adjacent and have the same owner. This leads to a situation where a house might be listed under the address of the mixed-use/residential building next door. The database also includes properties located in the City of Williamsburg, which does not include the entire Williamsburg housing market, as many students live in James City and York counties. Bedroom and bathroom data also seems to be incorrect on some multi-unit buildings and apartment buildings. Finally, as of right now, the dash website and database are locally hosted, so any response will only be recorded locally.

# 4 Future Work

Hosting the website publicly is the clear next step for this project. This would make the project a real tool that tenants could use as opposed to an academic exercise for the purposes of a final class project. Making the website more visually appealing and robust to vandalism are two actions that would need to be taken before the website is hosted. Better company data is another avenue of future work for this project. A program to query specific company names from the SCC instead of downloading a huge amount of property information and manually searching the remaining companies would allow the database to be updated completely automatically when it needs to be updated with the next year's information.