**Fort Worth, TX Neighborhood Comparison**

Robert Crosby

October 1, 2021

**1. Introduction**

**1.1 Background**

Many people are currently leaving major west coast US cities to move to Texas. My family is among those planning such a move from Portland, OR to Texas. The primary drivers for us and many others are the rising housing costs, increased cost of living, rising crime, job opportunities, and political unrest. The purpose of this Data Science project is to provide a framework for exploring neighborhoods in Texas to determine which areas to consider moving to and will explore factors including crime rates, percent of population near grocery stores/libraries/parks, and availability of other amenities. I will specifically be exploring neighborhoods in Fort Worth as this is where we are considering moving but the code will be applicable to other cities/neighborhoods and will only need to be updated with specific datasets for whatever area is being considered.

**1.2 Problem**

Moving to a new city is a big endevor with many variables that need to be considered. Every person or family considering a big move has their own priorities so for this project I decided to focus on the issues that are most important to my family. We have a large family with 6 children who we homeschool so the features that matter most to us when considering a move are crime rates, the availability of public utilities such as libraries and parks, and the availability of family friendly venues and grocery stores.

**1.3 Interest**

While this project is being used for my own personal neighborhood exploration it would be of interest to anyone considering a move to a new area. The dataset I used is for Fort Worth, TX but the basic coding could be used with data from a different city or region by someone wanting to perform a similar exploration.

**2. Data acquisition and cleaning**

**2.1 Data Sources**

Many cities and counties provide open data sources that can be used for free by anyone interested in learning more about the area. The dataset I found to use for this project is from the Fort Worth government website and is available in multiple formats. I chose to use a json file <https://data.fortworthtexas.gov/resource/ruhd-2sjc.json> for this as I was planning on reading it into a Pandas dataframe for my exploration. This dataset included zipcodes, neighborhood names, crime statistics, and percent of population near grocery stores, parks, and libraries along with other data I wasn’t interested in and didn’t include in my dataframe.

In addition to the json I used from the Fort Worth website I also used the Foursquare API to explore venues in close proximity to each neighborhood.

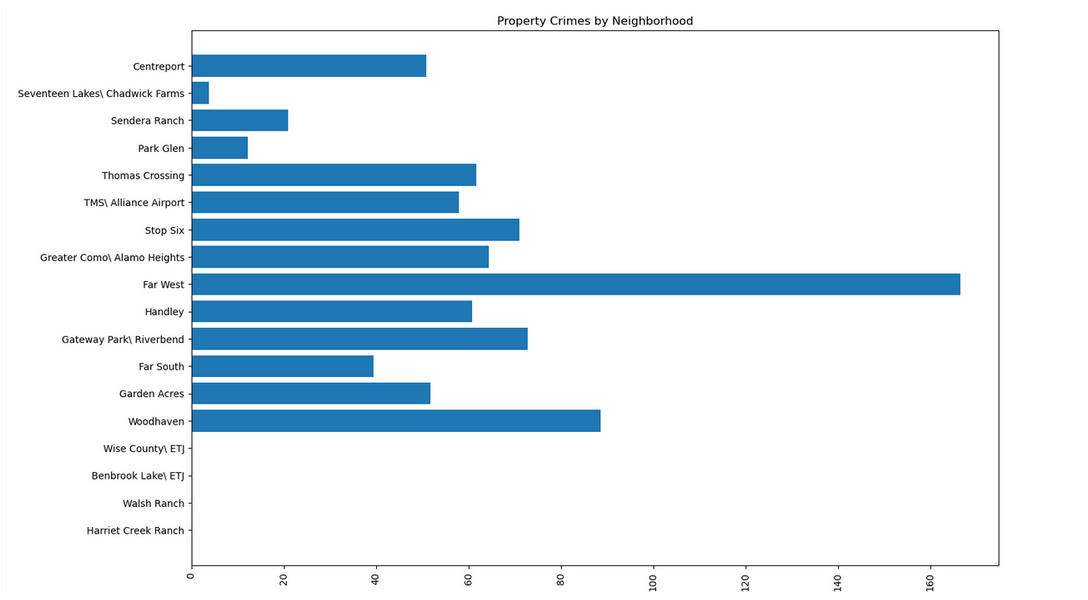
**2.2 Data Cleaning**

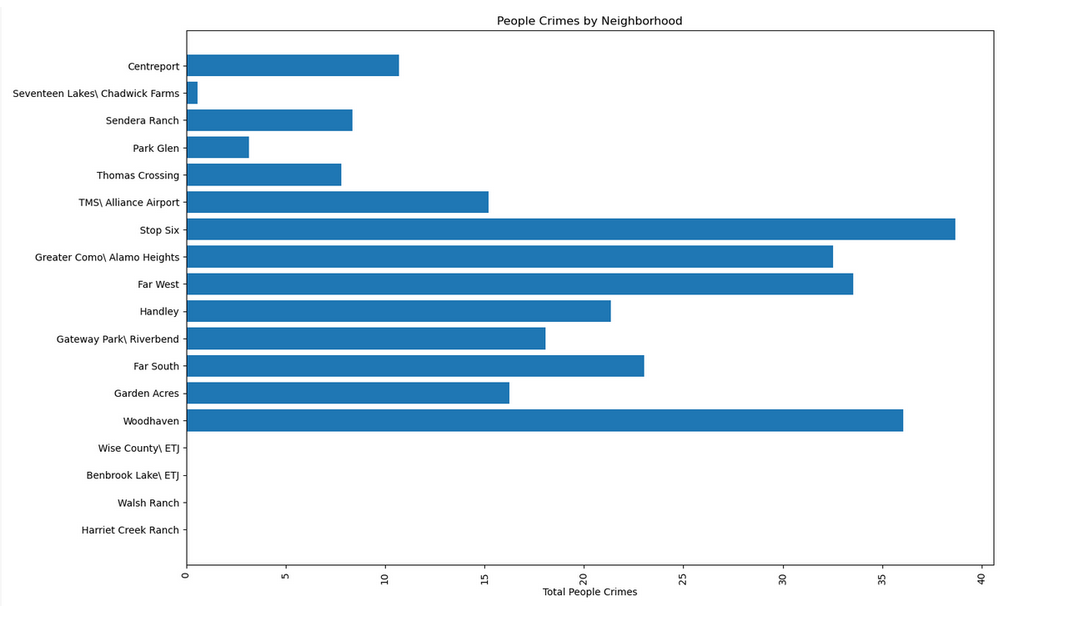
The data I downloaded from the Fort Worth website included over a million addresses rather than just data clustered by neighborhood. To simplify my exploration I decided to consolidate my data down to only use the first instances of each unique neighborhood rather than every address within a neighborhood. There also ended up being two neighborhoods with no Foursquare values returned so I removed those as well when I started exploring venues with Foursquare.

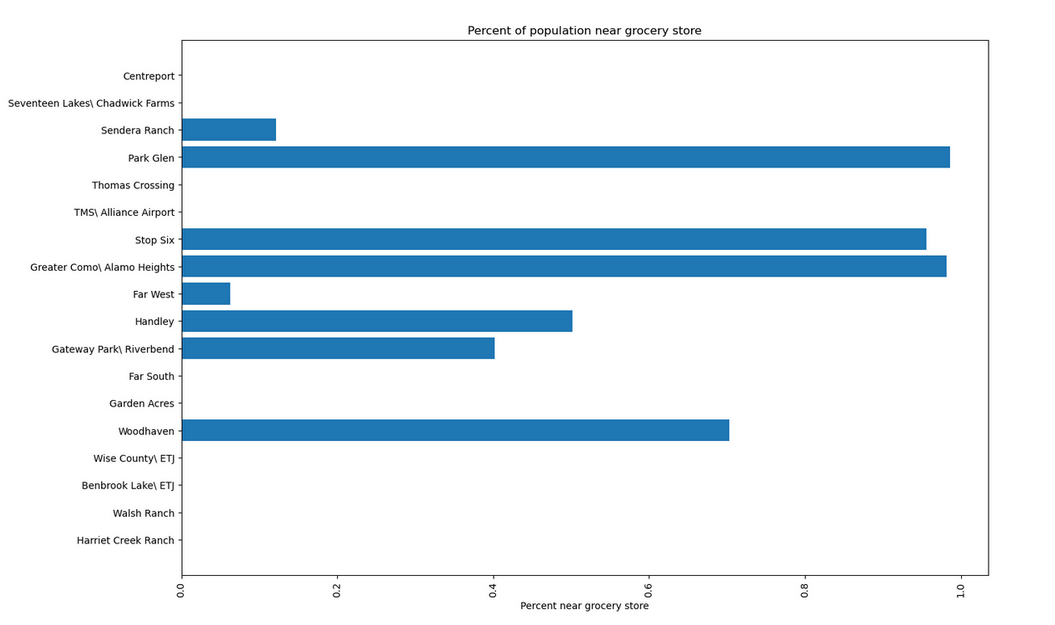
**3. Exploratory Data Analysis**

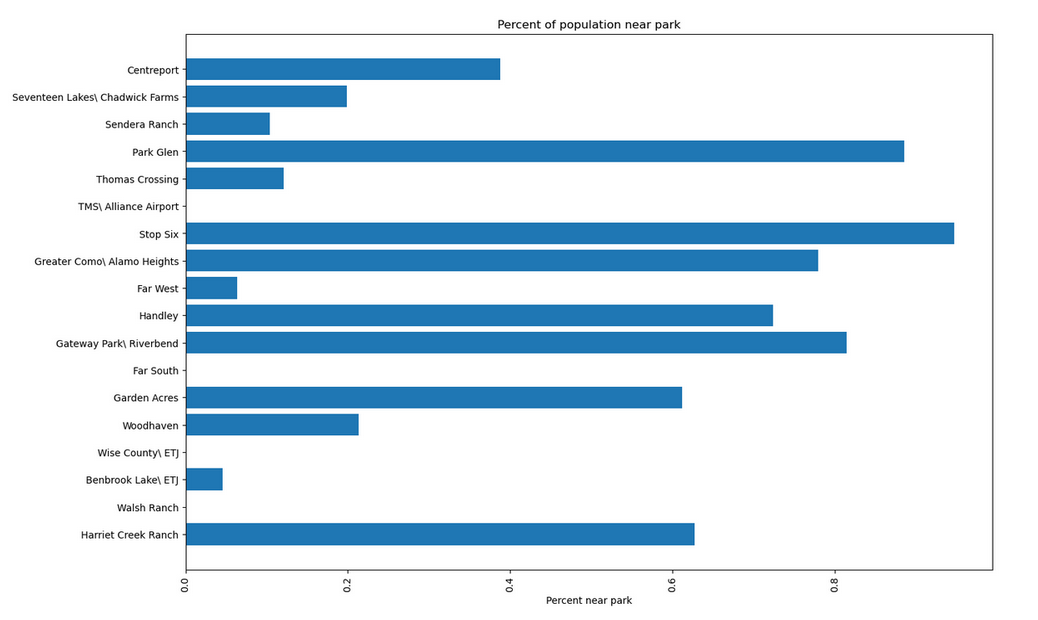
**3.1 Comparison of Neighborhoods**

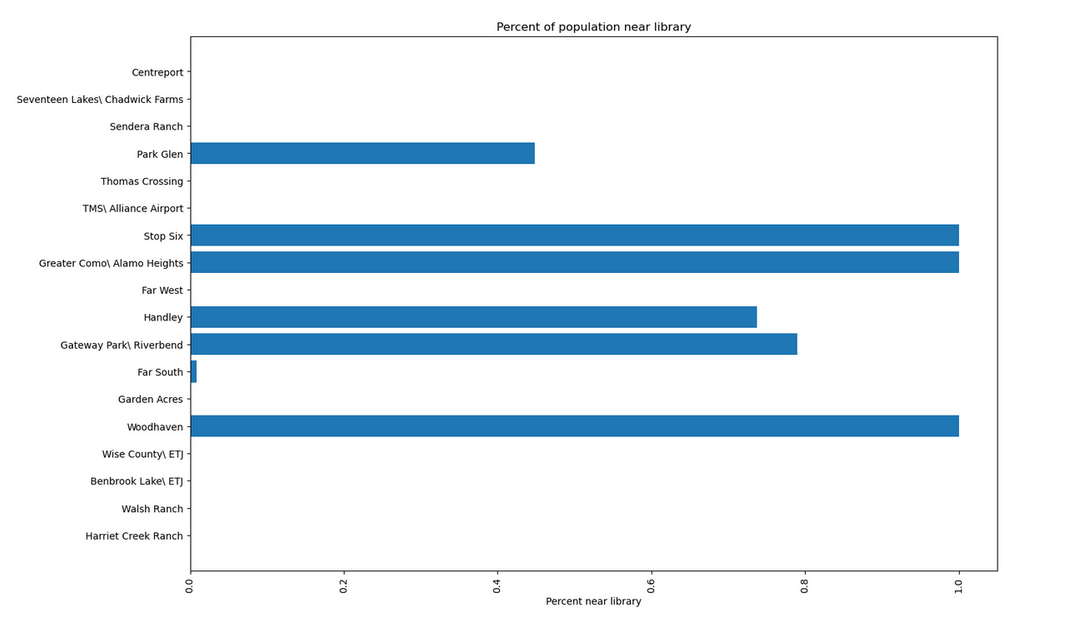
I began my comparison of the neighborhoods by using bar graphs to show how each neighborhood stacks up based on property crimes, people crimes, and percent of population near grocery stores, parks and libraries.











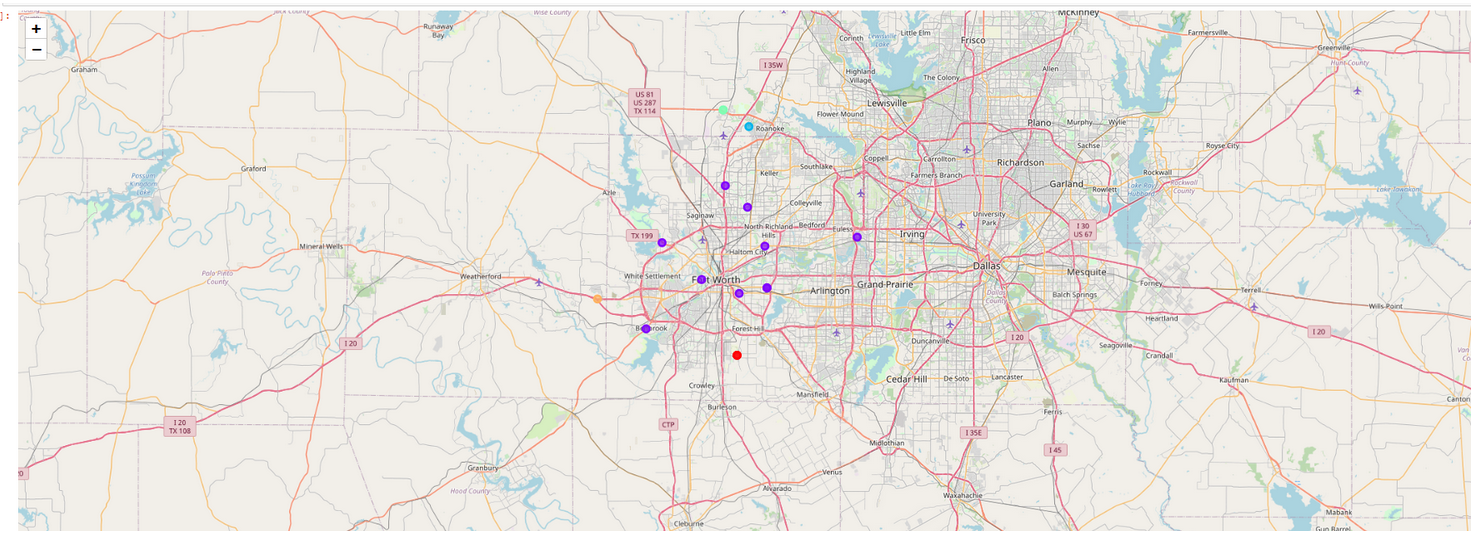
In these graphs the lower the number of each type of crime and the higher percent near grocery stores, parks, and libraries are preferred.

**3.2 Exploration of neighborhood Foursquare data**

Once I had finished comparing the previous data from the Fort Worth dataset I wanted to explore what venues were in each area through the Foursquare API. I processed a request to locate the top 10 most common venues in each neighborhood and create 5 clusters so I could see which neighborhoods were most similar.



After I had a dataframe showing the 10 most common venues I plotted the neighborhood clusters on a map of Fort Worth showing where they are and which clusters they are a part of.



The largest cluster is around downtown Fort Worth and the surrounding areas. This wasn’t surprising as I expected there to be more venues in this area than in outlying suburbs.

I explored what each cluster included to see what kinds of venues were common. Since Cluster 1 around downtown was the largest I wanted to see what was popular and available in these neighborhoods.



**4. Conclusion**

Since I am primarily looking for a balance of low crime and high availability of public and commercial amenities, the Park Glen neighborhood is the best fit for what is important to my family and possibly to other families with similar priorities.

**5. Future Directions**

The Park Glen Neighborhood appears to be the best fit for my family but there are always other variables that need to be considered. While preparing to move to Texas I will be using this data in conjunction with real estate apps while looking at houses to make sure they are in suitable areas. This will allow me to narrow down my search parameters and ignore areas that don’t have the features my family is looking for.