

**Austin Texas – Determining Where to Move**

# Introduction

“Austin is the state capital of Texas, an inland city bordering the Hill Country region. Home to the University of Texas flagship campus, Austin is known for its eclectic live-music scene centered around country, blues and rock. Its many parks and lakes are popular for hiking, biking, swimming and boating. South of the city, Formula One's Circuit of the Americas raceway has hosted the United States Grand Prix. “ – Google

With so much to offer it’s no wonder that Austin is a hot spot to live in Texas. But in such a diverse city how does one decide where to live? This project will be of use by realtors or individuals trying to determine where in Austin to live based on zip code analyses.

# Data Utilized

Datasets from:

* data.austintexas.gov for crime data & population growth rates
* <http://zipatlas.com> for population data
* <https://public.opendatasoft.com> for zip code latitudes and longitudes
* <https://www.homecity.com> for average rental prices
* <https://api.foursquare.com> for venue information

# Methodology:

Initially I attempted this analysis using neighborhood based data but unfortunately that was much harder to find so I had to revert to a zip code based analyses. I wanted to dig further into data about socioeconomic classes, housing prices, etc but much of the data I found was already in graph form and would require me manually re-creating dataframes which is tedious and would date the project quickly. I did this for a few elements but not many.

Once I determined I had to use zip codes I spent a few days digging into what data I could find based on the zip codes. I had to merge the crime and population datasets together to get a crime rate per capita rather than a count that could be biased to more populous regions.

I also cleaned the Foursquare venue information wrapping them into larger categories of data to make it more usable as well as deleting the extraneous zip codes it initially provided.

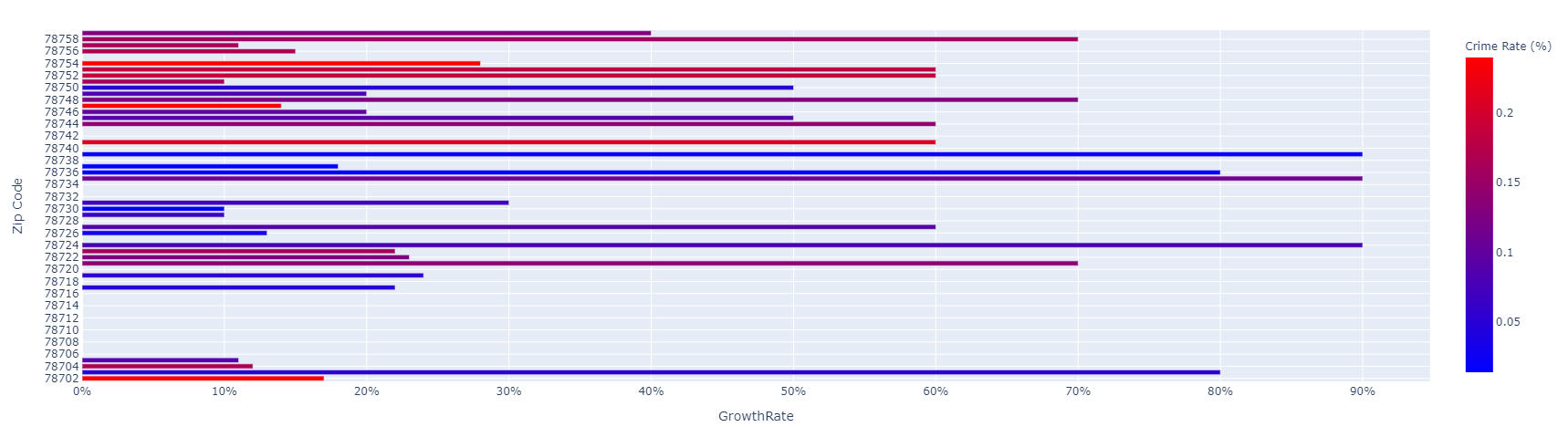
I then compared the data in a variety of ways which will be presented below. All results are descriptive statistics that could be used to determine where one would like to live based on what is most important to them.

The initial runs were over all zip codes which was hard to read due to the number of zip codes. That data was then used to narrow down choices for a theoretical couple and re-run for a subset of the zip codes for a better comparison. If an outstanding factor came into play (such as proximity from a job) the initial results could have been cleansed for a smaller number of zip codes

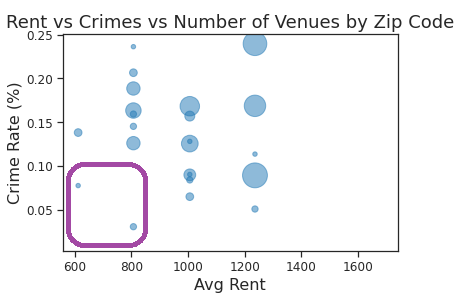
# Results:

The crime rates were low across all zip codes. This could be due to the pulling of the crime data from one source and the population data from another source. Therefore I may not take the exact percentages to heart but we could still make assumptions about which zip codes have higher rates than others.

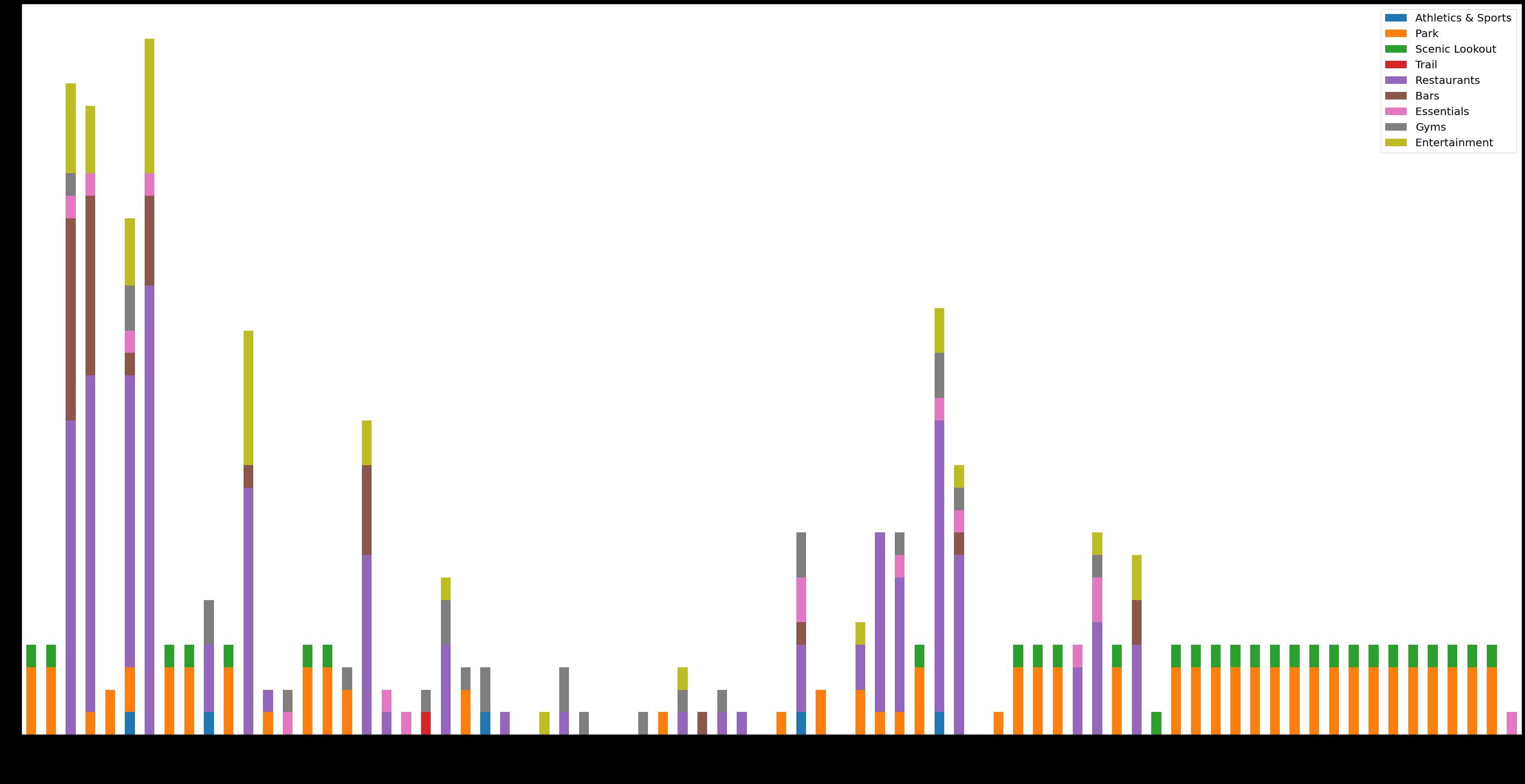
When looking at the crime rates vs the growth rates across all zip codes no discernable pattern of crime rates increasing or decreasing with growth were seen.



When comparing rent vs crime rate vs the number of venues by zip codes some predictable results popped up but some surprising ones did as well. As might be expected the number of venues was higher in areas with higher average rent. However, the crime rate did not seem to vary much based on rent nor number of venues. In fact some of the lowest crime rates were associated with the lowest rents and fewest venues (as noted in the purple square below).



It was also really interesting to see that when looking at venue information there are many zip codes with only parks and scenic lookouts yet no essentials like grocery stores.



# Discussion:

# The above can be used as a jumping off point to a deeper analysis. A person moving to Austin could use any of the analyses (or something else like proximity to work) to narrow down interesting zip codes and run the code again just on those zip codes to get more detailed / easier to compare results.

For example, if a person has a very strict budge the analysis could be filtered based on zip codes meeting their rent needs. Maybe being close to a large variety of restaurants is important so we could filter on the zip codes with the most restaurants, etc. I would suggest filtering on the 1 or 2 most important factors, then viewing the results and seeing which zip codes are most endearing to the individual.

# Conclusion:

This project was tough – finding enough data and getting Python to play nice with the data files was quite the challenge. Due to the time limitations I wasn’t able to dive as deeply as I might have liked to otherwise. Finding and learning how to use the json files was the biggest challenge for me. It also required me to change mindset as I was not able to compare neighborhoods as I initially intended.

It was a great exercise in learning how to get Python to do what I needed with the data I had, not the clean data from the examples given in the classes. It was also a very interesting dive into what data is available publicly if we just look.