Austin, Tx, an Informational Analysis for Incoming Residents and Interested Parties

---

A. Introduction

A.1 Background on this project

This peer reviewed capstone report is the final report for the IBM Data Science Professional Certification (https://www.ibm.com/blogs/ibm-training/data-science-ibm-coursera/). This course is a 9 week course that covers the fundamentals of data science, open source tools for data science, data science methodology, Python for data science and AI, databases and SQL for data science, data analysis with Python, data visualization with Python, and machine learning with Python. All of these courses have been completed using the Jupyter Labs notebook in IBM's Cognitive Class Laboratories (https://labs.cognitiveclass.ai/) and are stored on GitHub (https://github.com/jeff-mos-def/Coursera_Capstone/tree/master/Applied%20Data%20Science%20Capstone). This report is typed using Markdown formatting.

A.2 Topic Selection Data Description

Austin, Tx is the technology centerpiece of the State of Texas. Nested north of San Antonio, Tx and south of Dallas/Fort Worth, Tx Austin serves as the capital of the state with a population of 790,390 and population density of 3,182/sq mi (https://en.wikipedia.org/wiki/Austin,_Texas). The city has seen growth within the last decade due to large technology and software companies coming into the Central Texas Region. From 2007 - 2017 alone, the percentage population growth of the City of Austin rivaled that of the State of Texas and the United States (https://www.austinchamber.com/economic-development/austin-profile/population/overview).

Population Growth from 2007 - 2017 (https://cdn1.austinchamber.com/archive/images/ed/chart-population-growth.jpg?mtime=20190620093848)
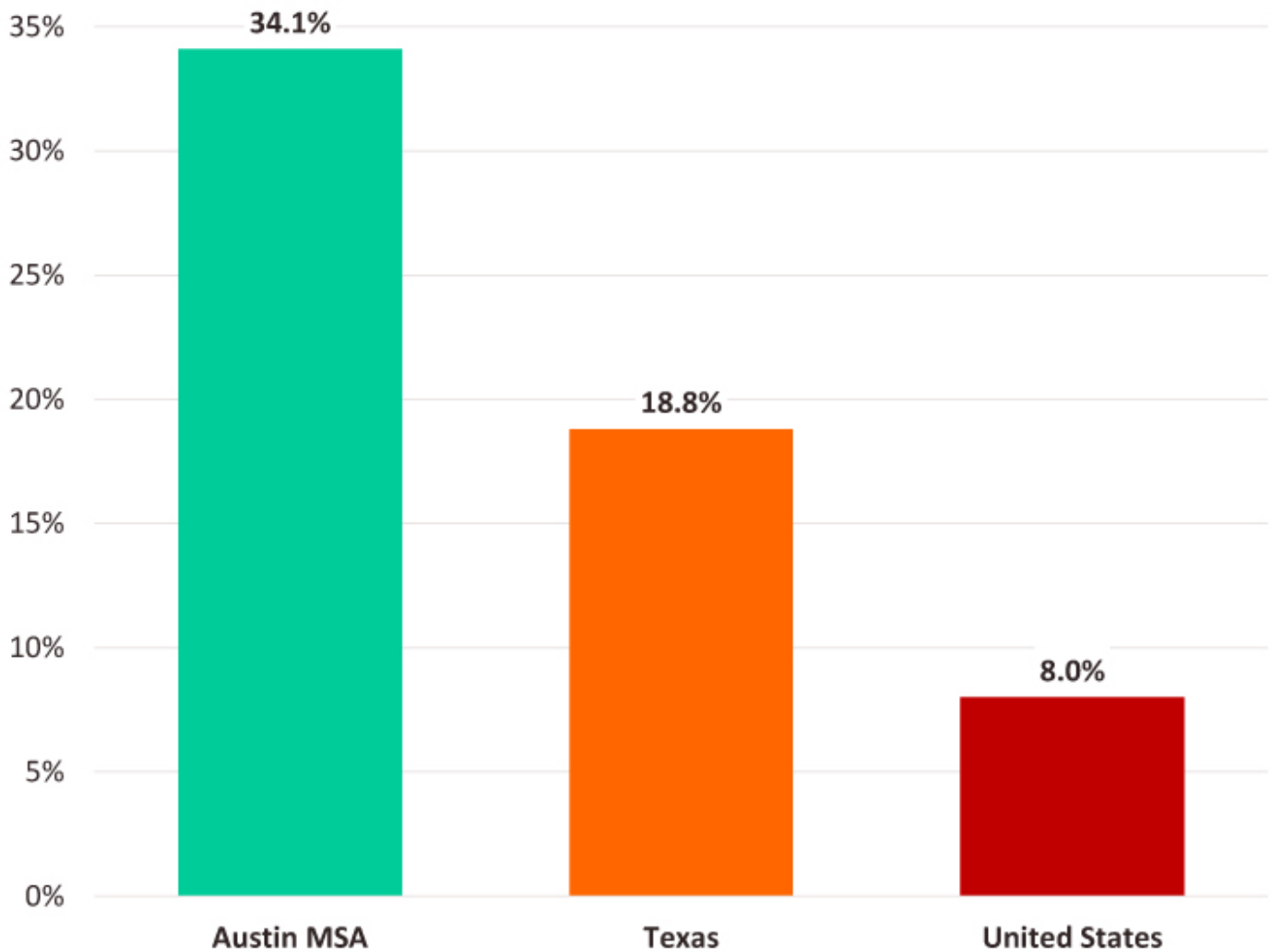
With this knowledge known, the focus here is to educate any interested party that would be coming to the City of Austin for work or leisure would like to know about some particular venues or hotspots within the city. Any business owners that may want to start a business in a given section of a city would also want to know what the current popular venues are in each respectable neighborhood. Those looking to buy a home would want to see what the median cost of a home would be in each neighborhood.

These problems will be considered and a map will be created with an informational view on clustered venues and median real estate values.

B. Methodology

In order to find all of this data, several sources will be used:

- Google Sheets scripts for quickly obtaining zip codes and latitudinal and longitudinal coordinates
- Foursquare API (https://developer.foursquare.com/) to scrape the local area for venues in each neighborhood
- GEOJson file (https://geo.nyu.edu/) for mapping of the Austin, Tx area
- A local area analysis of housing prices obtained through DemographicsNow (https://library.austintexas.gov/database/gale-business-demographicsnow)

First, the latitudinal and longitudinal coordinates of each neighborhood would be needed. The zip codes and names of these neighborhoods would be found on WIKIPEDIA and independently scraped using an external Jupyter Notebook utilizing BeautifulSoup4 (https://www.crummy.com/software/BeautifulSoup/bs4/doc/), a web scraper. Once the zip codes were identified, latitudinal and longitudinal coordinates were found via a GSheets script (https://raw.githubusercontent.com/jeff-mos-def/Coursera_Capstone/master/Applied%20Data%20Science%20Capstone/BoTN%20Final/Snapshots/ziploc.jpg). This was stored in a GSheets file and, using the scraped data previously acquired, the neighborhoods had their names concatenated with "Tx" so the zip codes were then able to be matched (https://raw.githubusercontent.com/jeff-mos-def/Coursera_Capstone/master/Applied%20Data%20Science%20Capstone/BoTN%20Final/Snapshots/geo2zip.jpg) to the scraped neighborhoods (https://raw.githubusercontent.com/jeff-mos-def/Coursera_Capstone/master/Applied%20Data%20Science%20Capstone/BoTN%20Final/Snapshots/MatchedNe) Areas containing the same zip codes were then matched and strung together.

Next, the report of median housing costs would be needed. Fortunately, the Austin Virtual Library has access to DemographicsNow (https://library.austintexas.gov/database/gale-business-demographicsnow), a tool which could be used to quickly obtain all relevant data in regards to median housing prices. Immediate housing prices were not available, so the data used was taken in 2018. The housing cost data was then added to the neighborhood listings.

All data was stored in a GitHub repository and in IBM Cognitive Class Laboratory. The data was stored as a CSV and called into a dataframe using Pandas. This resulted in the working dataframe to be used:

	ZipCode	Neighborhood	Lat	Long	HousingCost
0	78744	Dove Springs, Southeast Austin	30.1876	-97.7472	101990
1	78741	East Riverside-Oltorf, Montopolis	30.2315	-97.7223	133822
2	78758	North Burnet, Gracy Woods	30.3764	-97.7078	147457
3	78753	Windsor Hills	30.3648	-97.6827	150718
4	78745	Battle Bend Springs, East Congress, West Congr...	30.2063	-97.7956	151641

Folium (https://python-visualization.github.io/folium/) was used to map the local Austin area. Markers were overlayed to designate which areas consisted of neighborhoods, as designated by latitudinal and longitudinal coordinates.
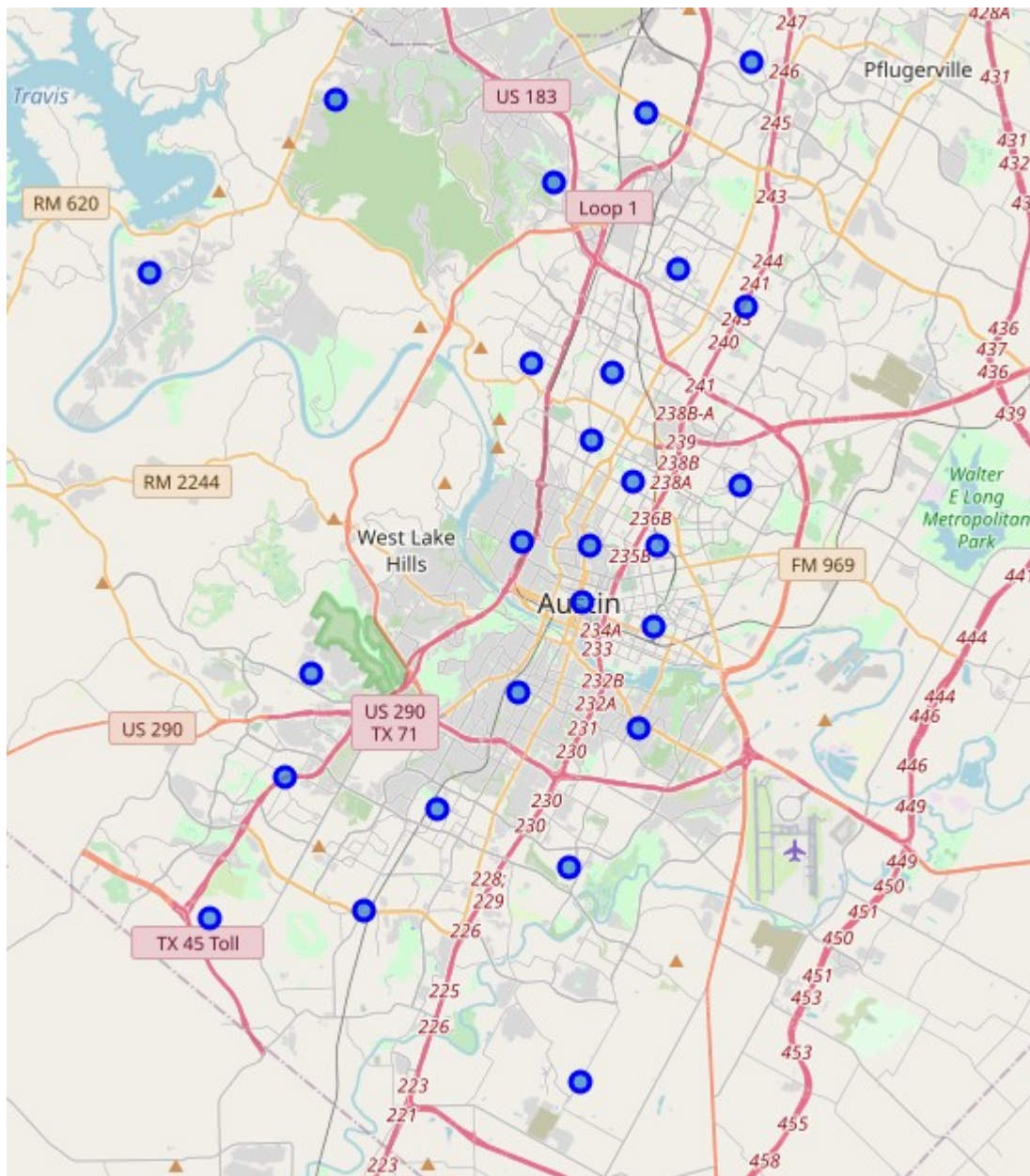
In a previous exercise it was learned that the FourSquare API could be utilized to to explore the contents of buroughs, here defined as neighborhoods, and segment them into a local area search. Due to the large area of Austin, a radius of 5500 meters (34.18 miles) was used as a visitor may want to drive across Austin to see various sights. Venues were limited at 100 for this exercise. When FourSquare was called upon, the `head()` of the call brought the following venues:

	name	categories	lat	lng
0	Cosmic Coffee + Beer Garden	Coffee Shop	30.226882	-97.762536
1	Lady Bird Lake Boardwalk Trail	Pedestrian Plaza	30.246597	-97.730338
2	El Primo Taco Truck	Food Truck	30.245016	-97.757393
3	Whole Earth Provision Company	Shoe Store	30.231258	-97.797953
4	Home Slice Pizza	Pizza Place	30.249004	-97.749758

When this data was looked at as a whole, it was discovered that FourSquare included gas stations, convenience stores, and bus stops in the API feedback. Ultimately, this was left in the whole of the data as some gas stations and convenience stores in Austin have bespoke goods or services. Bus stations and street intersections were also left in to keep the whole of the FourSquare data.

When checked, a total of 720 separate venues within the city was met. This table was then merged with the table of the neighborhoods.

	Borough	Borough Latitude	Borough Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Dove Springs, Southeast Austin	30.1876	-97.7472	Redbox	30.193995	-97.744820	Video Store
1	Dove Springs, Southeast Austin	30.1876	-97.7472	Stassney & Pleasant Valley	30.193985	-97.744629	Intersection
2	Dove Springs, Southeast Austin	30.1876	-97.7472	Fruitlandia	30.180719	-97.747339	Mexican Restaurant
3	Dove Springs, Southeast Austin	30.1876	-97.7472	Texas Reliable Services	30.184080	-97.754140	Home Service
4	Dove Springs, Southeast Austin	30.1876	-97.7472	Stassney / Nuckols Crossing	30.190723	-97.739941	Intersection
5	Dove Springs, Southeast Austin	30.1876	-97.7472	Mi Victoria Bakery	30.180493	-97.747866	Mexican Restaurant
6	East Riverside-Oltorf, Montopolis	30.2315	-97.7223	888 Vietnamese Restaurant	30.229627	-97.730117	Vietnamese Restaurant
7	East Riverside-Oltorf, Montopolis	30.2315	-97.7223	El Super Taco	30.230419	-97.730210	Food Truck
8	East Riverside-Oltorf, Montopolis	30.2315	-97.7223	Twin Liquors	30.234994	-97.720449	Liquor Store
9	East Riverside-Oltorf, Montopolis	30.2315	-97.7223	Wells Fargo Bank	30.235286	-97.721015	Bank

When the data was pooled for evaluation, there were some interesting findings. Only one neighborhood hit the limit of 100 venues per zip code, despite the fact that this is a merged zip code list. This could be due to incorrect identification by the business itself, or by not enough FourSquare reviews in certain areas.

It was also found that the zip codes of certain neighborhoods (WestLake Hills, NorthWest Austin, and RiverPlace) were producing errors in the data analysis and returning 0 venues. These zip codes were removed for venue analysis, but kept for housing cost analysis.

The original bar graph generated had shortened neighborhoods names. Full names are listed for complete viewing.
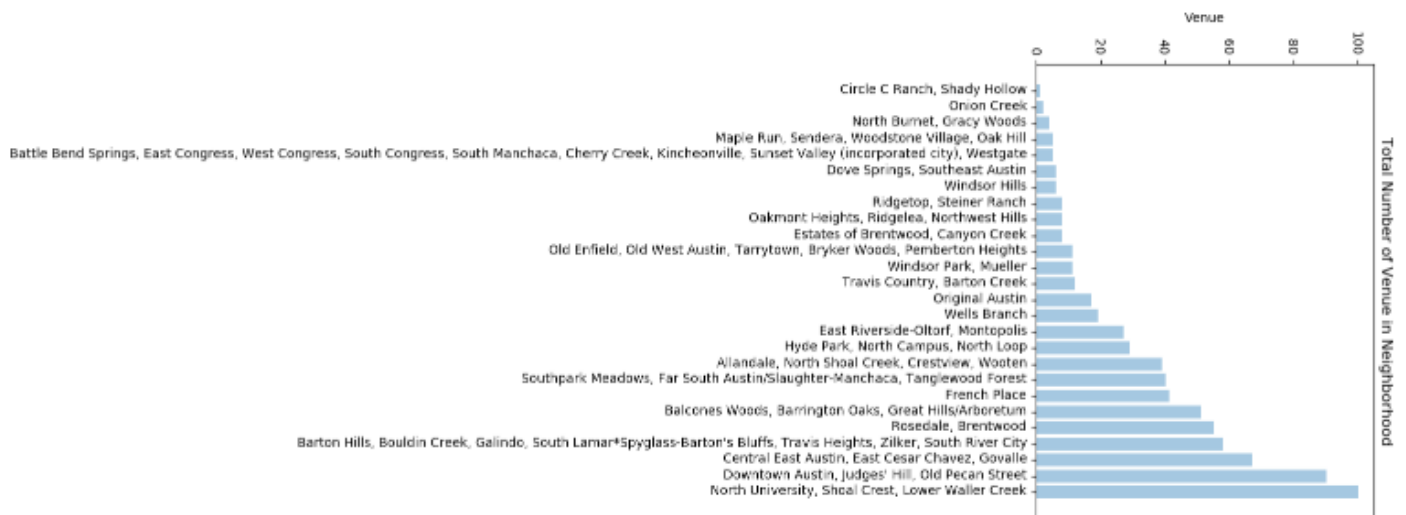
Out of all the venues polled by Foursquare, the noise needed to be cleared out for further analysis. A "Top 10" venue list was created for each neighborhood.

	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Allandale, North Shoal Creek, Crestview, Wooten	Pizza Place	Grocery Store	Food Truck	Spa	Ice Cream Shop	Bar	Bakery	Gym	Fast Food Restaurant	Jewelry Store
1	Balcones Woods, Barrington Oaks, Great Hills/A...	Indian Restaurant	Fast Food Restaurant	Gym / Fitness Center	Cosmetics Shop	Pizza Place	Chinese Restaurant	Mobile Phone Shop	Pet Store	Bakery	Rental Car Location
2	Barton Hills, Bouldin Creek, Galindo, South La...	Food Truck	Pizza Place	Mexican Restaurant	Park	Gym / Fitness Center	Bar	Taco Place	Coffee Shop	Thai Restaurant	Automotive Shop
3	Battle Bend Springs, East Congress, West Congr...	Food Truck	Convenience Store	Bakery	Health & Beauty Service	Field	Fast Food Restaurant	Farmers Market	Farm	Falafel Restaurant	Event Space
4	Central East Austin, East Cesar Chavez, Govalle	Coffee Shop	Mexican Restaurant	Cocktail Bar	Convenience Store	Brewery	Bar	Food Truck	Pub	Pizza Place	Fast Food Restaurant

Further analysis would need to be done in order to group our neighborhoods by common venues. Unsupervised K-Means clustering will be used in this instance. In order to group our venues, we will need to know what our optimal k value would be for clustering. The elbow method will be used to find this optimal k value.
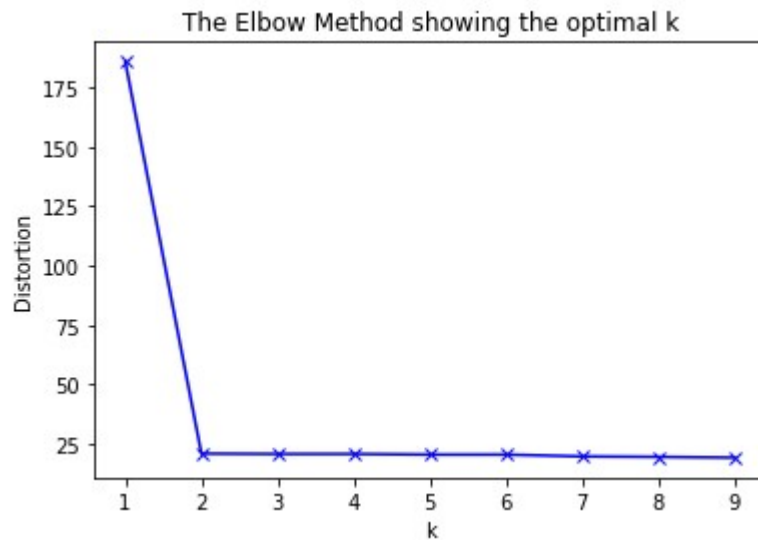

Our optimal k value would be 2 in this case. This is not an optimal case for k value clustering. All venues across Austin would be grouped into two bins. In spite of what the data is saying, the k value selected will be 5 for visual purposes.

The venues were grouped in a "Top 10" fashion, yielding this table:

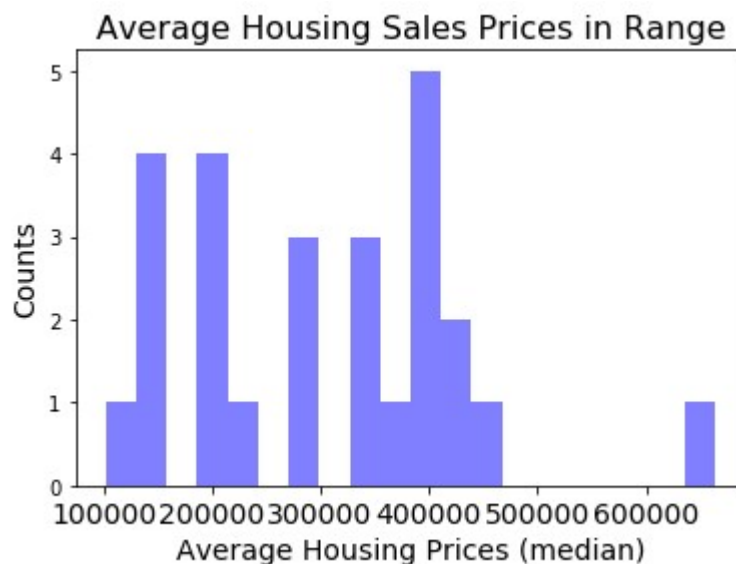|	ZipCode	Neighborhood	Lat	Long	HousingCost	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	78744	Dove Springs, Southeast Austin	30.1876	-97.7472	101990	2	Intersection	Mexican Restaurant	Home Service	Video Store	Dry Cleaner	Fishing Store	Field	Fast Food Restaurant	Farmers Market	Farm
1	78741	East Riverside-Oltorf, Montopolis	30.2315	-97.7223	133822	2	Bank	Pizza Place	Cosmetics Shop	Liquor Store	Video Store	Food Truck	Bus Station	Fast Food Restaurant	Mobile Phone Shop	Intersection
2	78758	North Burnet, Gracy Woods	30.3764	-97.7078	147457	2	Convenience Store	Grocery Store	Mexican Restaurant	Pizza Place	Fishing Store	Fast Food Restaurant	Farmers Market	Farm	Falafel Restaurant	Event Space
3	78753	Windsor Hills	30.3648	-97.6827	150718	1	Furniture / Home Store	Flea Market	Storage Facility	Rental Car Location	Electronics Store	Dry Cleaner	Field	Fast Food Restaurant	Farmers Market	Farm
4	78745	Battle Bend Springs, East Congress, West Congr...	30.2063	-97.7956	151641	2	Food Truck	Convenience Store	Bakery	Health & Beauty Service	Field	Fast Food Restaurant	Farmers Market	Farm	Falafel Restaurant	Event Space

The 1st Most Common Venue can also be displayed via a bar chart.

Number of Venues in Each Cluster
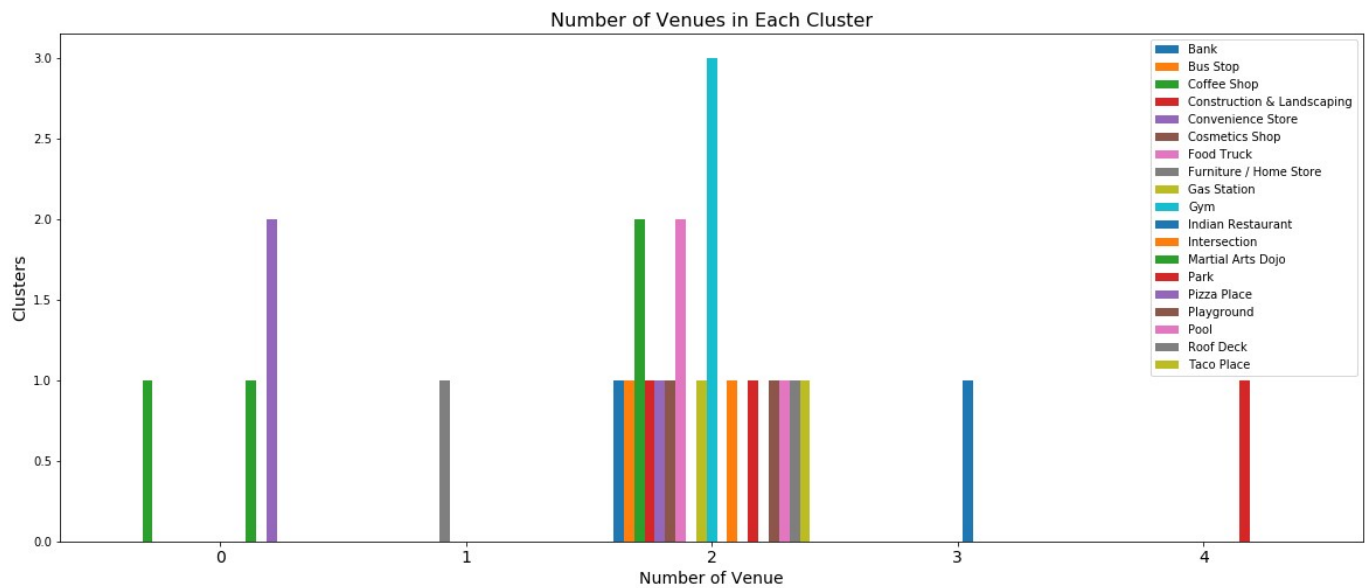
Clusters were then labeled according to their contents:

- Cluster 0: Pizza/Coffee
- Cluster 1: Home Stores
- Cluster 2: Total Austin
- Cluster 3: Indian Restaurants
- Cluster 4: Parks

Housing costs were also a point of concern in this study. Referring back to our first dataset, we can put the housing prices in Austin in a histogram for visualization:

Average Housing Sales Prices in Range

Housing costs were then binned in the following ranges:

- Less than 150,000 USD
- Ranging from 150,001-300,000 USD
- Ranging from 300,001-450,000 USD
- Ranging from 450,001-600,000 USD
- Greater than 600,000 USD

C. Results

The data is now able to be displayed in a centralized table for final viewing and visualization.

First, the clusters are added into the "Top 10" table:

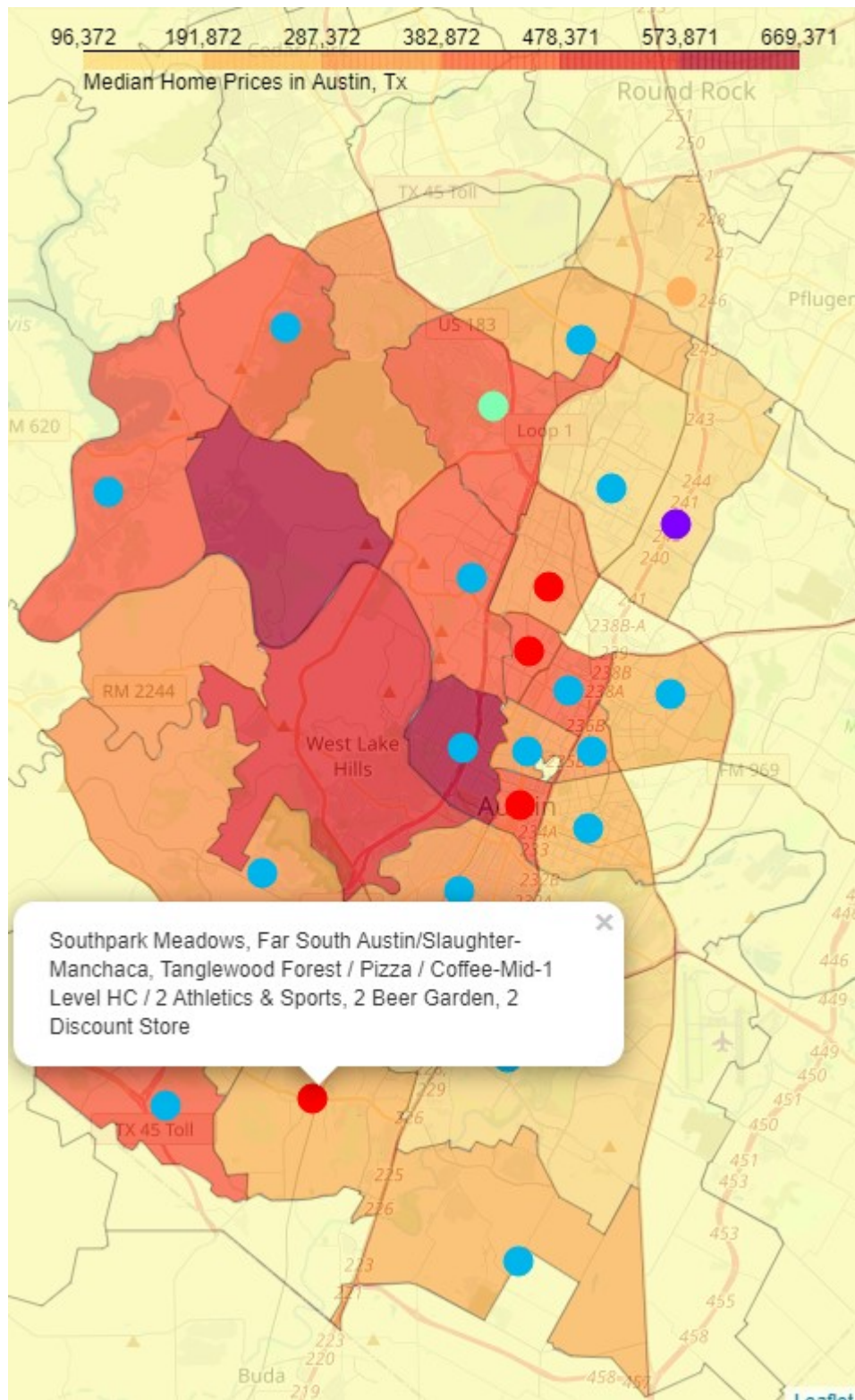| Lat	Long	HousingCost	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Borough	Join	Labels	Level_labels
30.1876	-97.7472	101990	2	Intersection	Mexican Restaurant	Home Service	Video Store	Dry Cleaner	Fishing Store	Field	Fast Food Restaurant	Farmers Market	Farm	Dove Springs, Southeast Austin	2 Intersection, 2 Mexican Restaurant, 1 Home S...	Total Austin	Low Level HC
30.2315	-97.7223	133822	2	Bank	Pizza Place	Cosmetics Shop	Liquor Store	Video Store	Food Truck	Bus Station	Fast Food Restaurant	Mobile Phone Shop	Intersection	East Riverside-Oltorf, Montopolis	2 Bank, 2 Cosmetics Shop, 2 Food Truck	Total Austin	Low Level HC
30.3764	-97.7078	147457	2	Convenience Store	Grocery Store	Mexican Restaurant	Pizza Place	Fishing Store	Fast Food Restaurant	Farmers Market	Farm	Falafel Restaurant	Event Space	North Burnet, Gracy Woods	1 Convenience Store, 1 Grocery Store, 1 Mexica...	Total Austin	Low Level HC
30.3648	-97.6827	150718	1	Furniture / Home Store	Flea Market	Storage Facility	Rental Car Location	Electronics Store	Dry Cleaner	Field	Fast Food Restaurant	Farmers Market	Farm	Windsor Hills	2 Furniture / Home Store, 1 Electronics Store,...	Home Stores	Mid-1 Level HC

Now the kmeans clustered and grouped zip codes are clearly defined on the generated map:
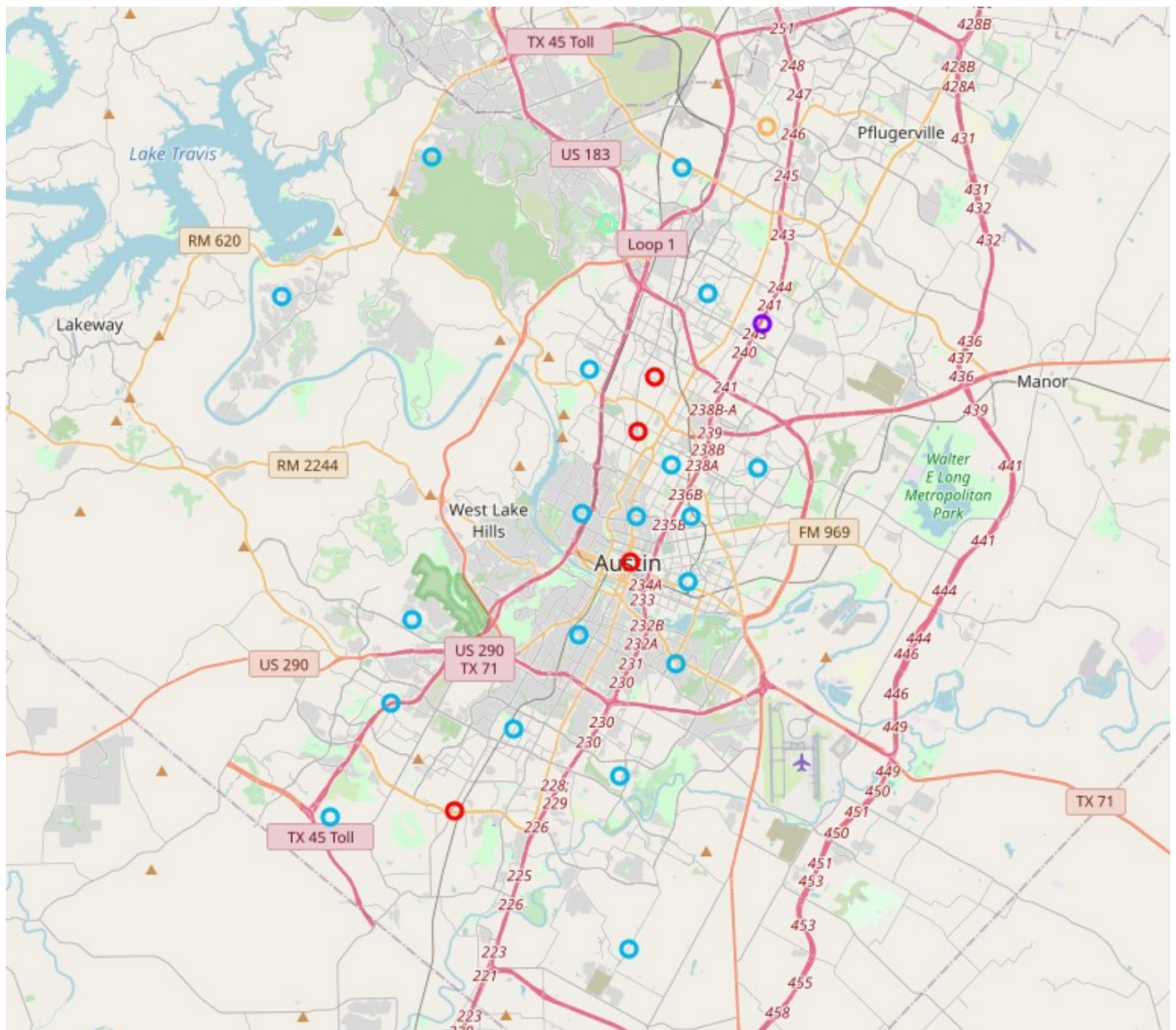
In the final map, we have a combined representation of all the data pulled and grouped for this project. A GEOJson file was downloaded from the Spacial Data Repository and used for this purpose. This map was then overlayed with the following information:

- Neighborhood name
- Cluster name
- Top 3 venues
- Color coded representation of median housing cost

Median Home Prices in Austin, Tx

Southpark Meadows, Far South Austin/Slaughter-Manchaca, Tanglewood Forest / Pizza / Coffee-Mid-1 Level HC / 2 Athletics & Sports, 2 Beer Garden, 2 Discount Store

D. Discussion

With the continuous influx of people relocating to Austin, Tx, this information could be valuable. Austin is a growing city with many different neighborhoods that sport many different venues and costs associated with each location.

The methods of grouping and clustering venues throughout the city still remains in question. Although using the elbow method to identify the optimal k value to be used for k means clustering is the tried and true way of partitioning observations, it would not have yielded much to visualize in this scenario. With k = 5, the clustering was forced to group together more venues into separate clusters. This method created a more visually attainable view of the data.

The median home values were also taken from 2018, as this was the most current data available. However, this would still provide the reader with an overall view of the median value distribution over the city. When newer data is available, the code can be updated.

This information would also be of knowledge that investors, small business owners, and budding entrepreneurs would like to see. By seeing the top businesses in a given area, along with the surrounding areas, one could easily call back the dataframe containing the raw data to see each separate business to forecast investment, construction, marketing, or advertising.

E. Conclusion

Planning a visit, move, or business venture always requires research. With a quick overview, one could get at a bird's eye view of the data of the Austin, Tx area. It could be seen what to expect when coming to this city. This gives a look-ahead to investors, property managers, or city planners as well.

F. References

1. "Austin, Texas." Wikipedia, Wikimedia Foundation, 4 May 2020, en.wikipedia.org/wiki/Austin,_Texas ().
2. Egan, John. "This Is How Much Austin's Population Could Surge through 2029, Study Says." CultureMap Austin, 15 Jan. 2020, austin.culturemap.com/news/city-life/01-15-20-austin-population-growth-through-2029-cushman-wakefield/ ().
3. "Foursquare Developer." Foursquare Developer, developer.foursquare.com/ ().
4. Gale Business: Demographics Now, www.gale.com/c/business-demographicsnow.
5. NYU Spatial Data Repository, geo.nyu.edu/ (geo.nyu.edu).
6. "Population Overview." Austin Chamber of Commerce, www.austinchamber.com/economic-development/austin-profile/population/overview ().