Finding the Perfect Neighborhood in San Antonio, Texas

DAVID RISIUS

FEBRUARY 17, 2020

Problem Statement

- Given a list of preferred criteria about a neighborhood, we would like to find an initial set of neighborhoods to begin searching for a new home in the San Antonio area. Our initial set of criteria is as follows:

  - Median Home Price: I am looking for a single-family house within the $200K–$350K range. There are multiple neighborhoods both above and below this range so these will be eventually filtered out. We would also like to find those neighborhoods where the median home price is increasing over time in case we would like to sell the home in the future

  - Good Schools: Since I have school-aged children, good schools in the neighborhood are very important.

  - Active lifestyle: Proximity to parks or other outdoor recreation is important. The ability to walk or bike versus drive to these areas is also important.

  - Diversity of Activities: I would like the neighborhood to have a wide range of venues available nearby. For instance, I wouldn't want all the top venues in the neighborhood to be gas stations or BBQ joints. A wide range of venues such as dining, shopping, and recreation would be important.

Data

► Geographic Data from San Antonio AreaConnect

► Texas School Performance Data from TxSmartSchools.org (Smart Score)

► Median Home Price Data Over Time from Zillow Research

► Joined and Cleaned Data Sets

	Zipcode	City	State	AreaCode	County	Latitude	Longitude	2012-12	2019-12	price_bins	price_labels	Smart Score
0	78201	San Antonio	TX	210	Bexar	29.472	-98.537	88509.0	156320.0	(100000, 200000]	1	4.0
1	78202	San Antonio	TX	210	Bexar	29.422	-98.466	60016.0	129942.0	(100000, 200000]	1	3.0
2	78203	San Antonio	TX	210	Bexar	29.415	-98.462	71213.0	150560.0	(100000, 200000]	1	1.5
3	78204	San Antonio	TX	210	Bexar	29.397	-98.500	77524.0	137329.0	(100000, 200000]	1	3.0
4	78205	San Antonio	TX	210	Bexar	29.424	-98.487	184158.0	259457.0	(200000, 300000]	2	2.0

► Combine this data with FourSquare API Venue Data

# Methodology

- Using the FourSquare data by San Antonio postal code, I use k-means clustering with k=7 to group each neighborhood according to the most popular venues.

- I use one-hot encoding on the San Antonio FourSquare data to determine the mean frequency of occurrence of different venue types for each postal code. This provides multiple clusters.

- I analyze the clusters to determine which have the desirable characteristics such as diversity of venues and proximity to parks and entertainment.

- Next, I filter out the zip codes based on median home prices and smart school scores. Combining the clusters with the school and home price data provides me with a short-list of neighborhoods to focus for my future home search.
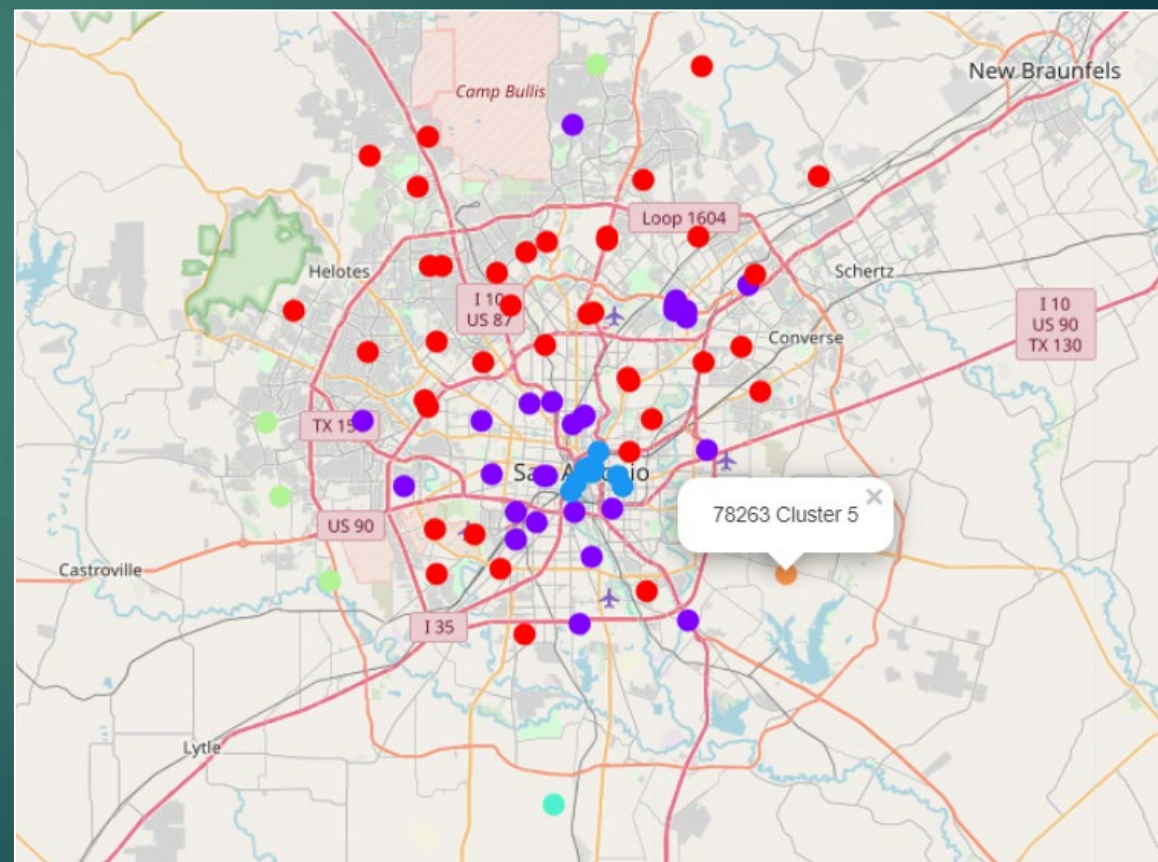
Analysis (1 of 2)

San Antonio Neighborhoods are Generally Clustered Geographically

▶ Initial mapping of neighborhoods
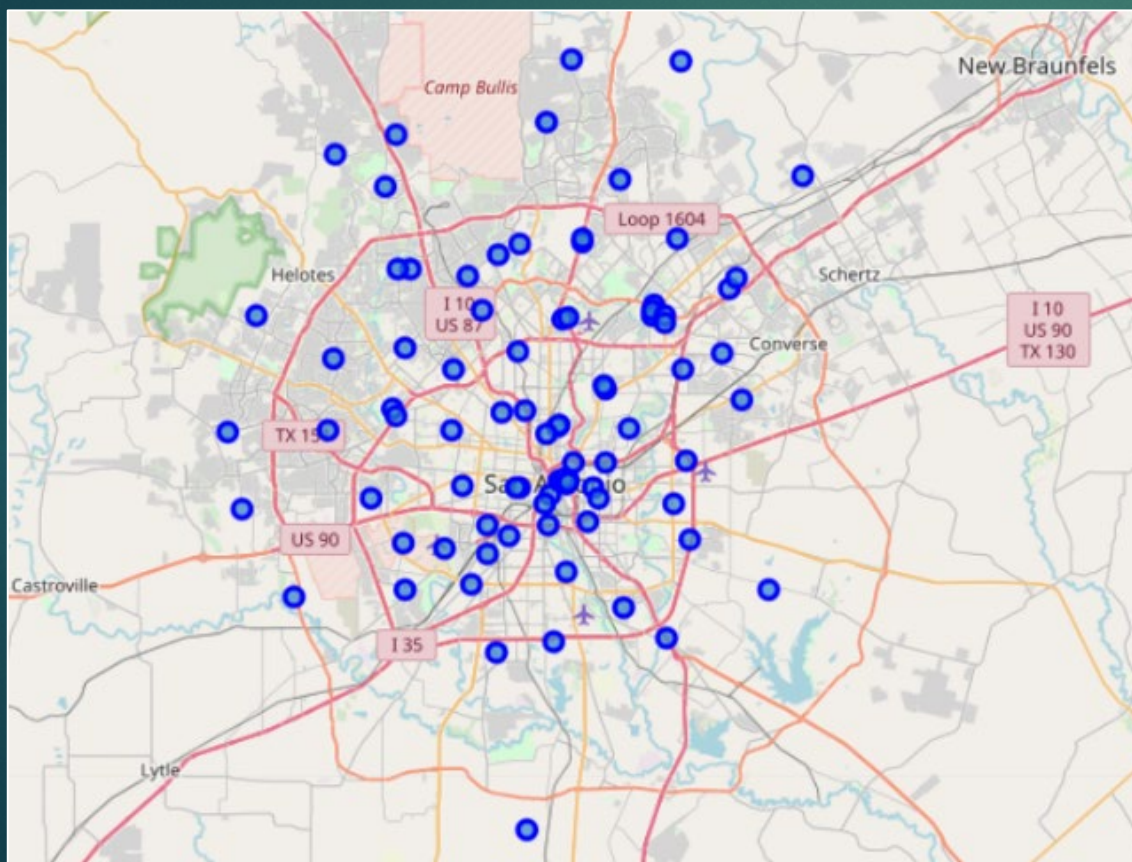
▶ After K-means Clustering

Analysis (2 of 3)

Generally, I would likely prefer cluster 0, 4, and 5 with a diversity of restaurants and shopping. Cluster 1 has too much fast food and Cluster 2 is mainly for tourists.

Cluster	Color	Characteristics
0	Red	Restaurants of all types, coffee shops, ice cream shops
1	Purple	Fast food and Mexican restaurants, discount stores, bars
2	Blue	Hotels, high-end restaurants, plazas, museums (this is the tourist district)
3	Light Blue	This only contains 1 postal code in south. Massage, studio, zoo, fish market, flea market
4	Light Green	On far west and far north sides. Parks, pools, golf, seafood, pharmacies
5	Orange	Gym, zoo, seafood, flea market, restaurants (this could probably be combined with cluster 3
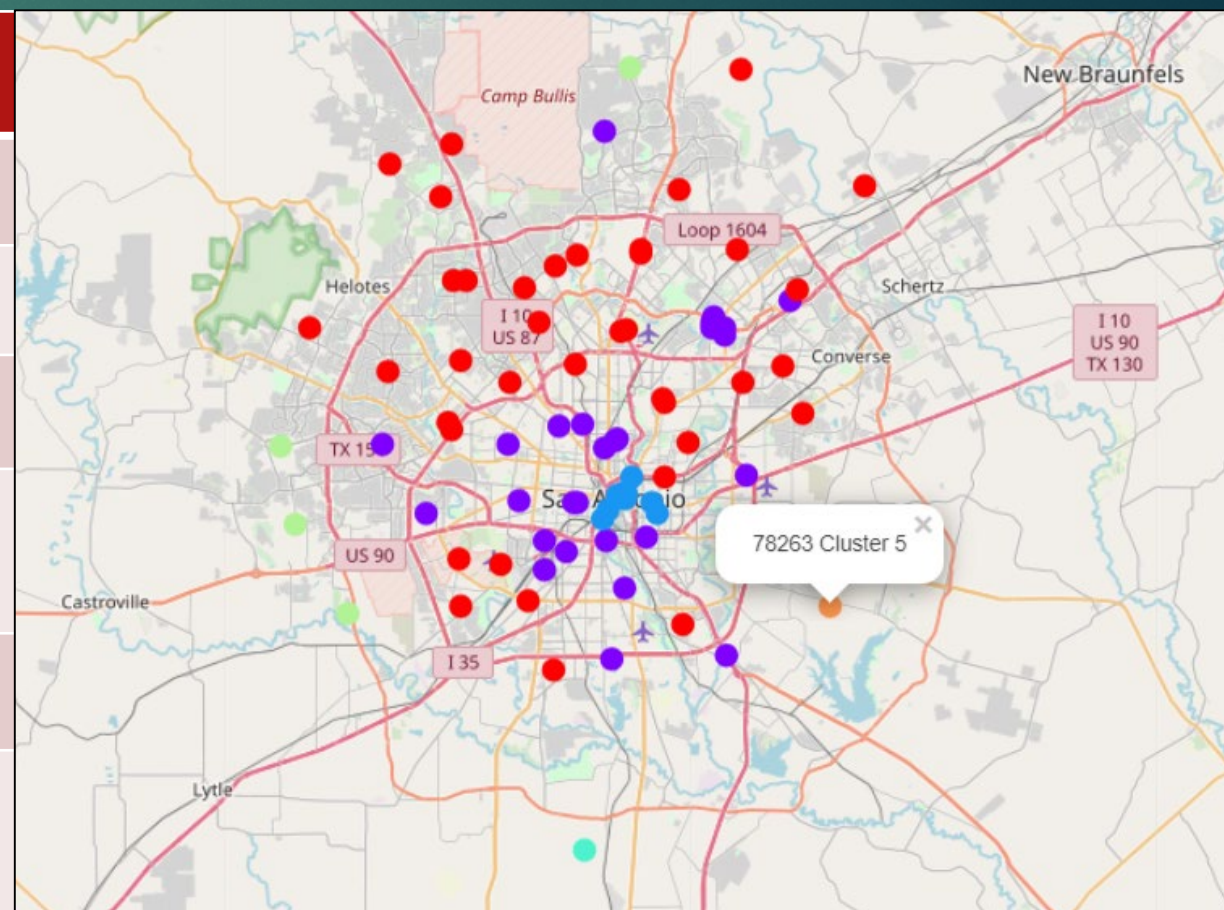
Analysis (3 of 3)

After filtering, most of the remaining neighborhoods are in the North or Far West sides. This will form the basis of our final list of six neighborhoods to start the home search

► Filter by desired median home price ($200K-$350K)

► Analyze the remaining list for top venues, home prices, and school information
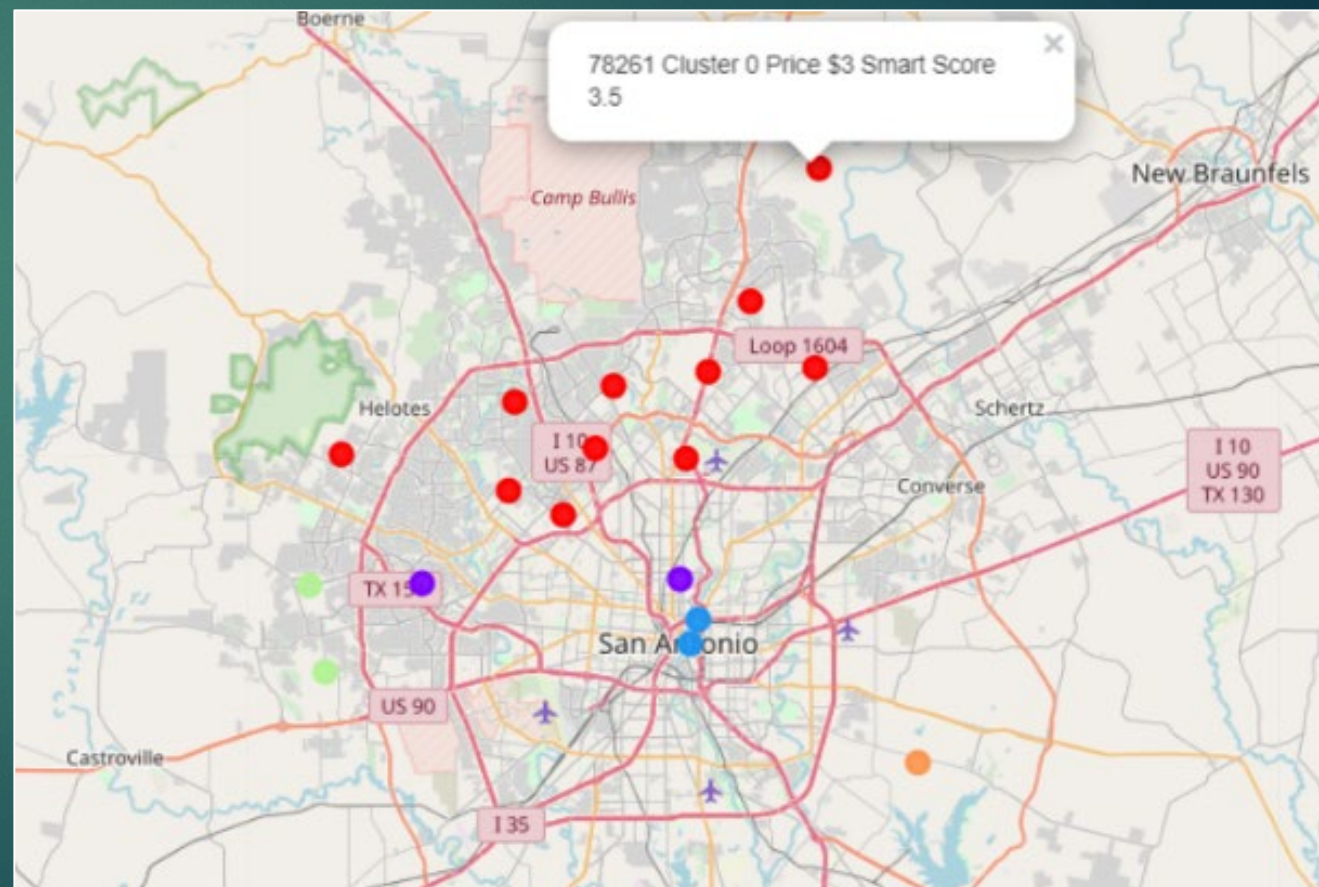
Conclusions (1 of 2)

The final step to get to the six neighborhoods was subjective, analyzing the maps on the neighborhoods to find areas with the best school scores. The map and table below show the final list of neighborhoods

Postal Code	Median Home	School	Venues
78212	$190K	8.5	Bars, Parks, Restaurants
78231	$327K	NA	Gym, Coffee shops
78249	$224K	12.0	Restaurants, shopping
78251	$197K	10.5	Restaurants
78253	$263K	10.0	Park, theater
78254	$226K	5.5	Personal care, shopping
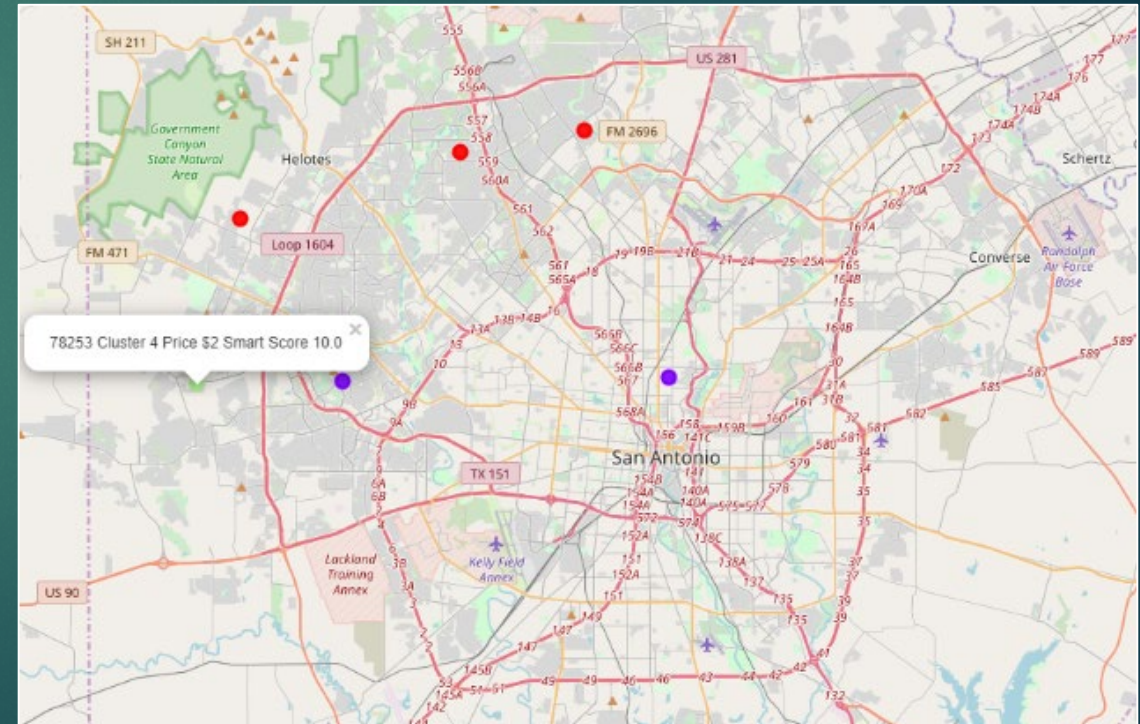
Conclusions (2 of 2)

► In this project, I used multiple data sources to determine initial search areas for finding a new house. Using the FourSquare data along with other research data on schools and home prices can assist us to find the right areas to live based on personal and family preference. Some other important aspects for this research that we did not consider in the research were traffic and commute times, which are pretty significant in San Antonio. Most of the neighborhoods I chose were on the edge of the city where the commute to downtown can run upwards of an hour.

► One interesting conclusion that I did not expect was that clustering the FourSquare venue data can tell us quite a bit about housing prices in an area. When I filtered the median housing prices, most of the purple (cluster 1) clusters filtered out. For future research, one might attempt to model median housing prices in a particular neighborhood based on the FourSquare, school scores and other data sources. They could help realtors and home buyers/sellers better determine housing prices to set.

► The complete notebook with all code and data files can be find on my GitHub page.