Recommend City for a Residential Living

Author: M.P.Jananee

Date: 28th Jan 2020

Table of Contents

[INTRODUCTION 3](#_Toc31144998)

[Background 3](#_Toc31144999)

[Problem 3](#_Toc31145000)

[Interest 3](#_Toc31145001)

[DATA SOURCES 4](#_Toc31145002)

[METHODOLOGY 5](#_Toc31145003)

[RESULTS 8](#_Toc31145004)

[DISCUSSION 8](#_Toc31145005)

[CONCLUSION 8](#_Toc31145006)

[REFERENCES 9](#_Toc31145007)

[ACKNOWLEDGEMENT 9](#_Toc31145008)

# INTRODUCTION

## Background

Texas is the second largest state in the United States by area and population. There are several known cities in this large state.

The fourth most populous city in the USA is "Houston" in Texas. "Austin", the capital city of Texas is the 11th most populous city in the United States.

Houston has a wider land area (1651.1 km2) compared to Austin (809.9 km2).

Houston is a broad industrial base city and Austin is a major center for high tech, home to lakes and beautiful landscapes.

Amidst these two choices of city a person would like to know the recommended city for a good residential living in Texas.

Finding available residential places will be easy based on budget and plot sizes but the necessary basic amenities nearby is required for a comfortable living for a family. Hence this project might be helpful in identifying such places in this regard considering certain aspects as a sample reference.

Information about the cities based on certain categories within a specified area limit would help to make recommendations on the city to be chosen for living.

This project helps to identify the necessary data required for the recommendation, how to encode certain values, which model to be used and maps creation for better visualization of areas of interest.

## Problem

Choose a city for a residential living amidst the given list of cities, based on the venues that covers certain basic categories.

Data that comprises about the city's venues specific to categories would be required.

This project aims to select Austin or Houston as the city for a living based on specific categories of interest.

## Interest

This project can be used to explore and detect city of choice based on 3 basic important categories "Residential Buildings", Hospital", "Schools".

The cities would be clustered considering these areas of interest and folium maps would help to display the clusters clearly and depicts the density.

This would help any family who would like to settle in a city, choose place of residence considering certain factors.

# DATA SOURCES

For this project data is mainly fetched through Foursquare API requests only.

List of venues in a city within a radius of 1000m that belongs to category "Residential Buildings", "Hospital", and "School" are fetched.

The cities that are chosen to be compared within the state of Texas is "Austin", "Houston".

Its location (latitude, longitude) is known from the wiki "https://en.wikipedia.org/wiki/List\_of\_United\_States\_cities\_by\_population" for reference.

The URL "https://api.foursquare.com/v2/venues/search" will be used with appropriate query parameters for user authentication, location, category id.

The query parameter 'location' will be latitude and longitude of the cities being requested separately.

The query parameter 'categoryId' will be a list of categories that the user would like, to which the venue belongs.

Refer the link to know the category id of specific category "https://developer.foursquare.com/docs/resources/categories"

From the response of the venue search requests, latitude, longitude of the venue, name of the venue, category to which the venue belongs are retrieved.

These should be enough to know the presence of such venues within the specified radius.

But 'Distance' information is also retrieved from the response which states the distance of the venue from the center that would also help to group the venues nearby.

The resultant data frame will contain the name of the venue, location of the venue (latitude, longitude), distance from the specific city location and its category.

Encode the "Category" column using One Hot Encoder.

The categorized venues data frame for the cities 'Austin' and 'Houston' will be the main input for modelling, comparison and visualizing their venues.

This project considers only the input from foursquare site but the same can be merged with other datasets that might contain housing price, housing plans etc in order to deep drive into residential availability in such cities within the budget.

# METHODOLOGY

The objective to choose a suitable city for a residential living is tedious as there are many options and most of the cities are well-grown and renowned for different aspects. Often nowadays it takes long time to decide the place rather than shifting/moving to that place.

The current method chosen is to reduce such time for selection and to make quick decisions based on available data and choice of interest.

Initially when venues around a place were fetched, the list was huge and there were several categories (restaurants, ice cream shops, convention stores, bars, gyms, fitness centers and many more).

Currently have chosen 3 main categories, so filtering the remaining and choosing basic categories was a specific task. Venues that contain 'Residential Buildings - Condos' (Assisted/Private/House are not considered), 'Hospitals', 'Schools' only are chosen to compare the cities and choose one among them.

To predict whether a person would select a city or not for a living, might be predicted using 'Classification Model'. This would be purely a binary output.

Since our objective is to recommend a city based on certain factors, we might need to compare 2 cities, explore the different categories, check the density of its presence.

The city to be compared was chosen such that they are equally rated, popular in different domains, considerable difference in population density and land area.

There is no existing data on how many have selected between Austin and Houston and what would be the best prediction of it, basically there is no labelled data.

Based on above points, "Clustering" mechanism of modelling is chosen which will help us to identify a pattern of venues within a range of distance, and type of category

eg: <400m, > 400 - < 600, >600-<800, >800-<1000, >1000 - <1200

Hospitals within a range of 600-800 and residence in range of 800-1000.

'KMeans' Clustering with different number of clusters (NClusters = 3, NClusters=4) were tried and explored using folium maps.

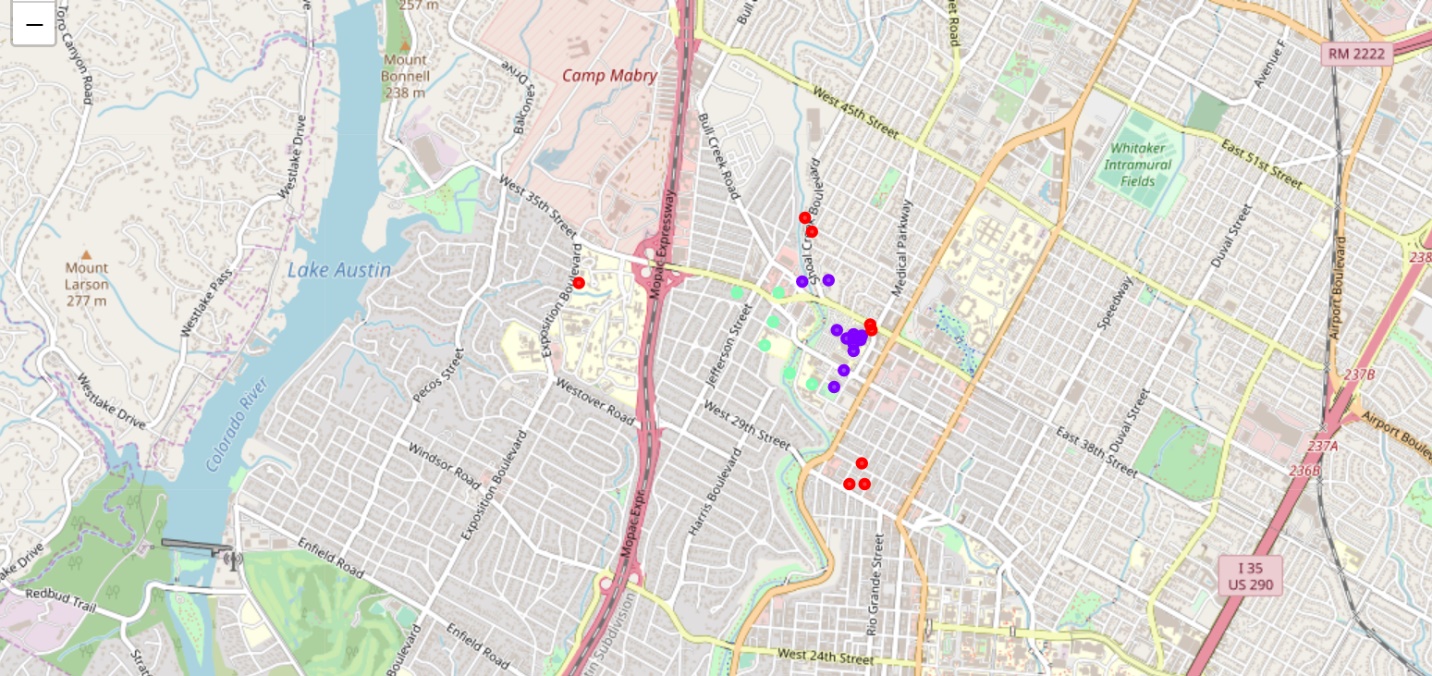
NClusters as 4 resulted in more expanded clusters which revealed that within a radius of 100m, there were many venues.

In case of 'Austin' city, cluster of distance of 600 - 900m resulted in another cluster of many venues concentrated in range of 800 - 900m.

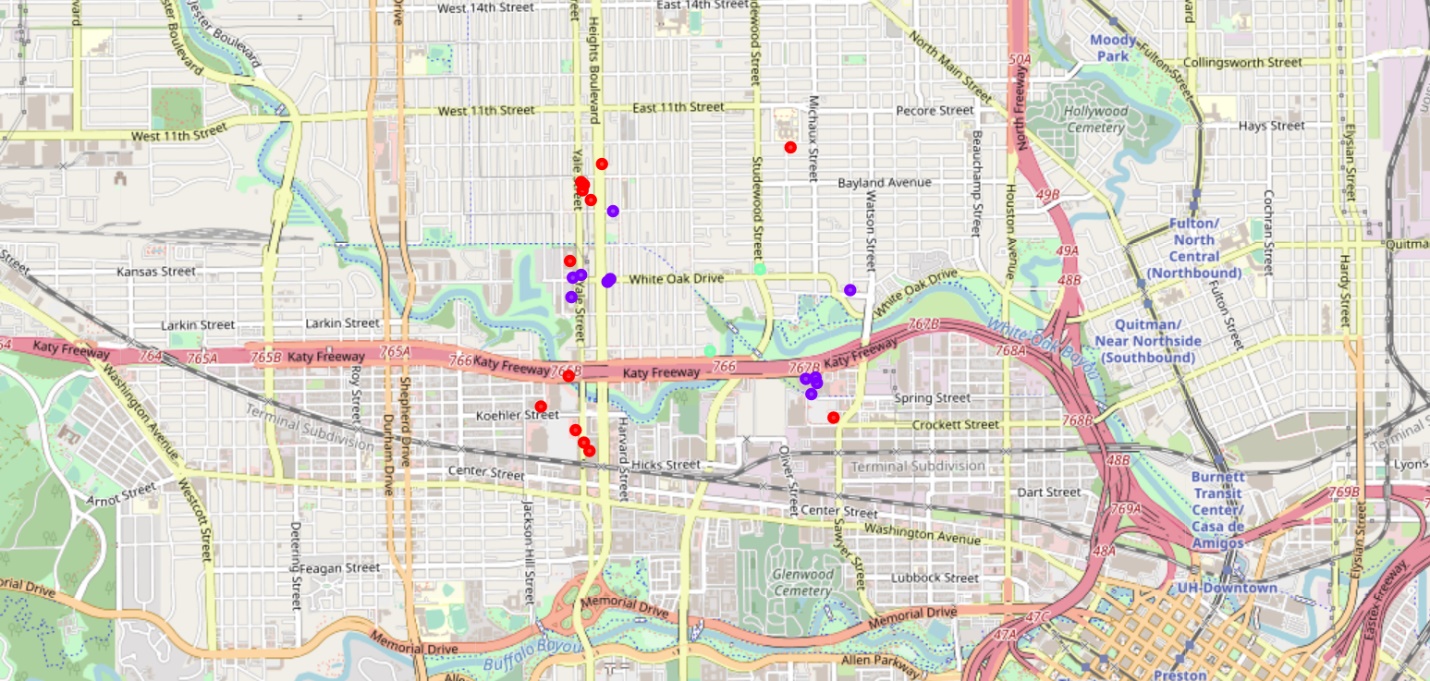
In case of 'Houston' city, cluster of distance of 900 - 1200m resulted in another cluster of many venues concentrated in range of 1000m and above.

Hence NCluster of 4 was chosen for the Clustering Modelling.

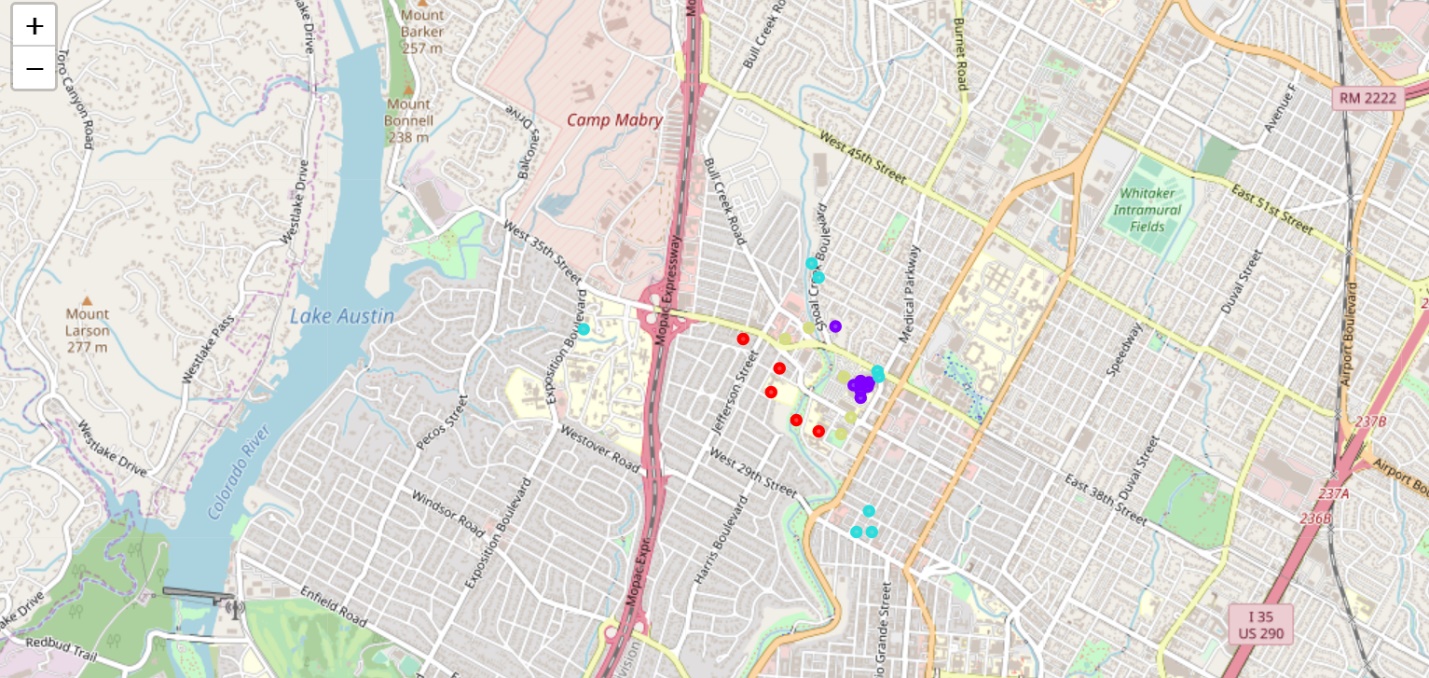
**MAP 1 – AUSTIN with N-Cluster as 3**



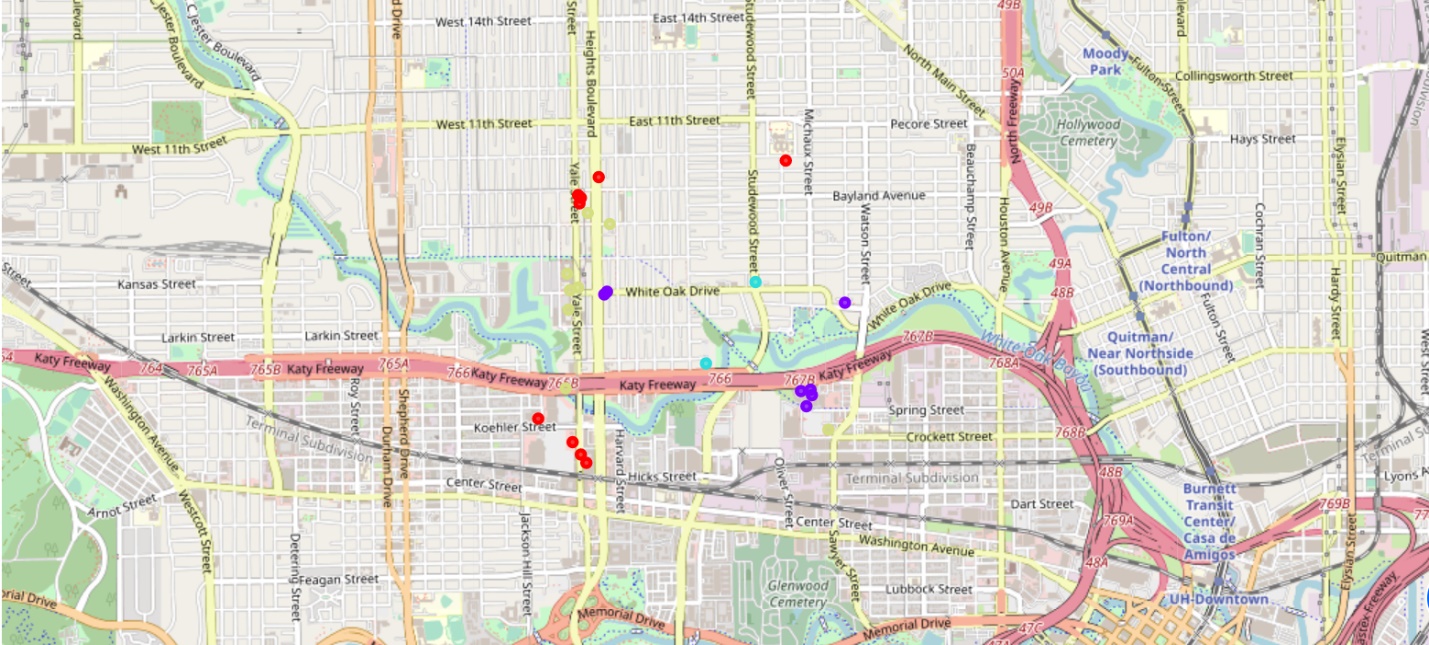
**MAP 2 – HOUSTON with N-Cluster as 3**



**MAP 3 – AUSTIN with N-Cluster as 4**



**MAP 4 – HOUSTON with N-Cluster as 4**



# RESULTS

Cities 'Austin' and 'Houston' are compared on Clusters modelled and its respective folium maps are displayed for better visualization. The clusters with its density are clearly shown in the maps.

Hospitals and Schools are much lesser in Houston and residences are more concentrated.

Although hospitals are more in Austin, hospital related areas and schools are widely spread.

Residences are sparsely located around 300m and more starting at distance 900m.

Based on above facts, within a specific radius of 1000m and specific categories, **"Austin"** can be suggested as chosen city for a residential living for a family.

# DISCUSSION

Although only few categories are considered, there can be multiple other categories to select venues nearby. For a residential living other factors apart from location can be considered along with this to find a much suitable city of choice.

Important aspect is also to consider comparable factors while recommending suggestions as it would be beneficial of performing modelling amidst such cases only.

Modelling helps us in predicting correct results for crucial tricky scenarios where mere manual observation will lead to errors.

This case study helps to avoid such manual errors and shows real facts about venues with not only numbers but a pictorial representation of it for easy selection and decision based on the results.

# CONCLUSION

In this case study, an example of comparing cities ‘Austin’ and ‘Houston’ using "Clustering" model is performed.

This project can be utilized for any such cities with different categories.

The results can be used by anyone who would like to settle in the city, for building contractors to know the existing residence density, for social workers or government to know whether the basic amenities such as school, hospital, convenient store are within the accessible range to the people. The result would also help any investor to know the whereabouts about nearby places to predict if he can invest in new shops/types of it.

The results showing folium maps with different clusters are self-detailed that quick decisions can be made based on such data.

# REFERENCES

<https://developer.foursquare.com/docs/api/venues/search>

<https://developer.foursquare.com/docs/api/venues/categories>

<https://developer.foursquare.com/docs/resources/categories>

<https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population>

# ACKNOWLEDGEMENT

Thanks to all the Teaching Staff of the entire courses to share such insights on Data Science in a readable, easier way and making us knowledgeable in this domain.

Thanks to Coursera for providing such great platform for learners.