Recommend City for a Residential Living

Author: M.P.Jananee

Date: 26th Jan 2020

Table of Contents

[INTRODUCTION 3](#_Toc30968438)

[Background 3](#_Toc30968439)

[Problem 3](#_Toc30968440)

[Interest 3](#_Toc30968441)

[Data Sources 4](#_Toc30968442)

# 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.