

Predicting what are the best places to live in Houston when you start a new life

Andres Franco
December 2020

1. Introduction

Houston is the fourth largest city in the United States of America and one of the most diverse cities in the world.

When a person immigrates to the United States and lives in Houston for the first time, they must go through a process of adaptation and generally must begin to rebuild their life from scratch.

This process implies not having the privileges that other people have, such as having a car, having a credit history, among other things.

The data of the neighborhoods become more important at the moment that they provide information on the different amenities that do not present a significant cost of mobilization for their residents and where they consequently raise their quality of life.

1.2 Problem

As Houston is a city with many neighborhoods, the data collected will focus on the number of amenities found in the different super neighborhoods. This will allow the person to be moved a better overview of the amenities available in a larger area, to determine which of these super neighborhoods offers more amenities and therefore the cost of mobilization is lower.

The amenities that will be considered will be those that can increase the quality of life of people such as: parks, restaurants, gyms, schools, supermarkets, banks, gas stations, clothing stores.

1.3 Interest

The information should interest people with different profiles but will mainly be focused on people who are thinking of moving to Houston or people who want to start a new life in a different area of the city.

2. Data acquisition

2.1 Data sources and Data cleaning

The main source of data is the list of all the super neighborhoods in Houston with their respective latitudes and longitudes. These data were already extracted from different sources, mainly from the Houston city page where all the super neighborhoods are listed:

<https://www.houstontx.gov/superneighborhoods/>

the coordinates were extracted from the Google maps website, and some manually from Bing maps since some super neighborhoods were not in google maps as the government page lists them.

The information has already been collected in [this CSV file](#)

2.3 Feature selection

Together with the information of the 88 super neighborhoods and their coordinates, the Foursquare API will be used to build a data frame that allows to list the amenities for each super neighborhood.

These amenities will be filtered by amenities focused on improving the quality of life, their main categories being those that contain:

- Parks
- Restaurants
- Gyms
- Schools
- Supermarkets
- Banks
- Hospitals
- Telecom providers

***These categories may vary according to the Foursquare API, what is a requirement is that the selected categories are of first necessity and help to improve the quality of life of people**

Example of data frame during data cleaning

	Parks	Restaurants	Gyms	Schools	Supermarkets	Banks	Hospitals	Telecom Providers	Total
FOURTH WARD	1	16	4	2	3	3	1	2	32
SECOND WARD	1	10	5	2	3	3	1	2	27
DOWNTOWN	3	30	10	4	5	10	3	4	69
GREATER UPTOWN	1	20	4	4	5	6	2	3	45