# Expatriation – choosing the best neighborhood to move in to Houston with ML. A practical case.

## Table of Contents

# 1 Introduction

## 1.1 Background

Everyday, people move out of their home country to another one. The United Nations estimated the number of people living outside their home country to be around 232 millions[1].

There is several motivations of why one could decides or be forced to expatriate to a foreign country. To list a few, this could be for:

- Business purpose; an employee is sent to another country by its company
- To flew conflict region or persecution
- For economical purpose; be able to find a job or simply for better salary conditions
- etc.

Whatever the reason is, expatriates will encounter a common challenge: leaving their home, there region they know well to another home, another region yet to be known.

## 1.2 Problem

Choosing the right spot isn't easy. Each particular situation could narrow down the choice, to a continent, a country or a town. But even when the destination town has been defined, the expatriate still face a large panel of choices. E.g., the choice could lays anywhere from one hour drive radius from the future work location. As diverse as the world can be, different choices can lead to completely different experiences.

Choosing a neighborhood that correspond to one aspiration is not an easy task. The success in choosing generally come with a solid knowledge of what a location can offer. This comes after several trials, errors and is helped by friends discussions, feedback and recommendations.
Expatriates will typically have limited knowledge about the destination prior to the expatriation as they don't live there yet. Also, as for friend recommendations, they will have few acquaintance to counsel prior or short after their arrival.

Still, it is possible to get help and recommendations through numerous local associations, books or simply over internet. This said, it is hard to find tailor made recommendations as one aspirations are by nature individuals.

It worth noting that expatriation candidate have a limited time to dedicate in finding the right spot to move in as there are many challenges encountered during the expatriation process. In fact, the choice of the neighborhood to move in is probably one of the last thing an expatriate candidate will focus in. Indeed, what would be the point to put so much effort finding the good neighborhood while the Visa application is still ongoing and the outcome uncertain? Is it really relevant while the location of the job

is not yet known? Expatriate with family must take care also that their child have are accepted by a proper school.

Even though the location is tackled at the end of the process, it deserve much attention as the expatriate will live there for a significant period of their life. In most case, renting means to commit for at least about 1 year. Buying, for even longer.

## 1.3 Interest

In the author opinion, being able to find the right neighborhood is a major factor of success of an expatriation. If there is other obvious factor of success as Visa/legal, work, etc. A bad geographical situation will brings problem that will come in addition to the other problems an expatriate have to face.

The destination location is important, for the expatriate that seek to enjoy is life as anyone else and this start with where he lives. What quality of life can be expected from a bad neighborhood?

The cost for companies expatriating employees are not negligible. The cost for employee should not be forgotten has well. We can list for instances:

- Visa/work permit and legal fee
- Moving cost
- Employee compensation
- Spouse quitting his job
- etc.

A bad experience will obviously leads to an early renunciation/stop of the expatriation experience or poorer employee morale. Without question, this cause a negative impact on the expected return of investment for both the company, the employee and his family.

## 1.4 Scope of this work

In this work, we develop an algorithm that aims to help choosing the best location for an expatriate based

- on his current location or a location he knows and would appreciate to live in
- a target town where he aims to settle

The purpose is to help the subject to choose the best neighborhood possible of a town that is unknown to him by clustering the destination town neighborhood according to their markers and find which ones are the closest from a neighborhood the expatriate know well and likes to live in.

In this work, we will focus on one practical case. The case chosen here is the case of an expatriate leaving in his hometown Liege, a town of 200,000 inhabitants of Belgium to move in to Houston, a town of 2,300,000 inhabitants of U.S.A. [2][3]

Having a current locations that he likes for its venues and surroundings, the goal set here is to determine, using machine learning algorithms, which neighborhood of the destination town he should choose.

Best neighborhood will be chosen under the following criteria:

1. Destination town neighborhood alike to home neighborhood in term of nearby venues
2. Safest places in therm of criminality for a given housing budget.

The first part of this work will be done by splitting the destination town in same size square. For each subdivision, nearby venues data available from the API Foursquare will be listed, namely: Name, category and location. A clustering algorithm will be applied in order to group the subdivision that are alike. Then, the same model will be applied to the origin home neighborhood to determine which sub-region of the destination town are most alike.

For the second part, criminal record from HPD will be compiled and a criminality score will be attributed to each sub-region according of the offense type (more point awarded for a homicide than for a simple felony). In parallel, house pricing per $ft^2$ from Redfin will be compiled and attributed to each destination town sub-division. Then, a multi-objective analysis will be conducted to sort out the safest neighborhood by house price range.

Other features than the one cited above will not be taken into account in the analysis. For instances, following feature are excluded of the analysis:

- housing quality,
- distance from work locations
- weather difference (e.g. flood area)
- road conditions / public transportation
- Population density

At the end, a map generate using Folium library will be displayed with the best candidate per price ranges so the expatriate can choose it’s new neighborhood according to his financial abilities.

# 2 Data

## 2.1 Origin of data

In this work, following data have been used.

1. Data on venues are coming from Foursquare API (https://developer.foursquare.com)
2. Coordinate of specific address have been obtained with the functionality ‘Search Nearby’ from Google Maps (https://www.google.be/maps/)
3. Criminal record data for Houston in 2019 have been taken from the NIBRS (https://www.houstontx.gov/police/cs/Monthly_Crime_Data_by_Street_and_Police_Beat.htm)
4. Housing pricing per zipcode have been provided by **Redfin**, a national real estate brokerage. (https://www.redfin.com/blog/data-center/)

## 2.2 About data used in this work

### 2.2.1 venues information from Foursquare API

Foursquare has built a massive dataset of accurate location data all across the world. In this work, we are using the Foursquare API in order to get information about venues in a given range of a location.

Data of Foursquare API has been used in order to assess what kind of venues where available across the destination and origin town. Then, we have process the venues information in order to cluster the different subdivision of the destination town in several categories and compared them to the origin town.

Bellow are some extract of the Foursquare API used in this work.

| | Cell_index | Cell_Latitude | Cell_Longitude | Venue | Venue_Category | Venue_Latitude | Venue_Longitude |
|---|---|---|---|---|---|---|---|
| 0 | 0 | 29.601804 | -95.561416 | Walgreens | Pharmacy | 29.600119 | -95.563614 |
| 1 | 0 | 29.601804 | -95.561416 | Starbucks | Coffee Shop | 29.599570 | -95.563985 |
| 2 | 0 | 29.601804 | -95.561416 | Tornado Burger | Burger Joint | 29.611505 | -95.564204 |
| 3 | 0 | 29.601804 | -95.561416 | El Vaquero | Mexican Restaurant | 29.589083 | -95.564442 |
| 4 | 0 | 29.601804 | -95.561416 | CVS pharmacy | Pharmacy | 29.600460 | -95.565012 |

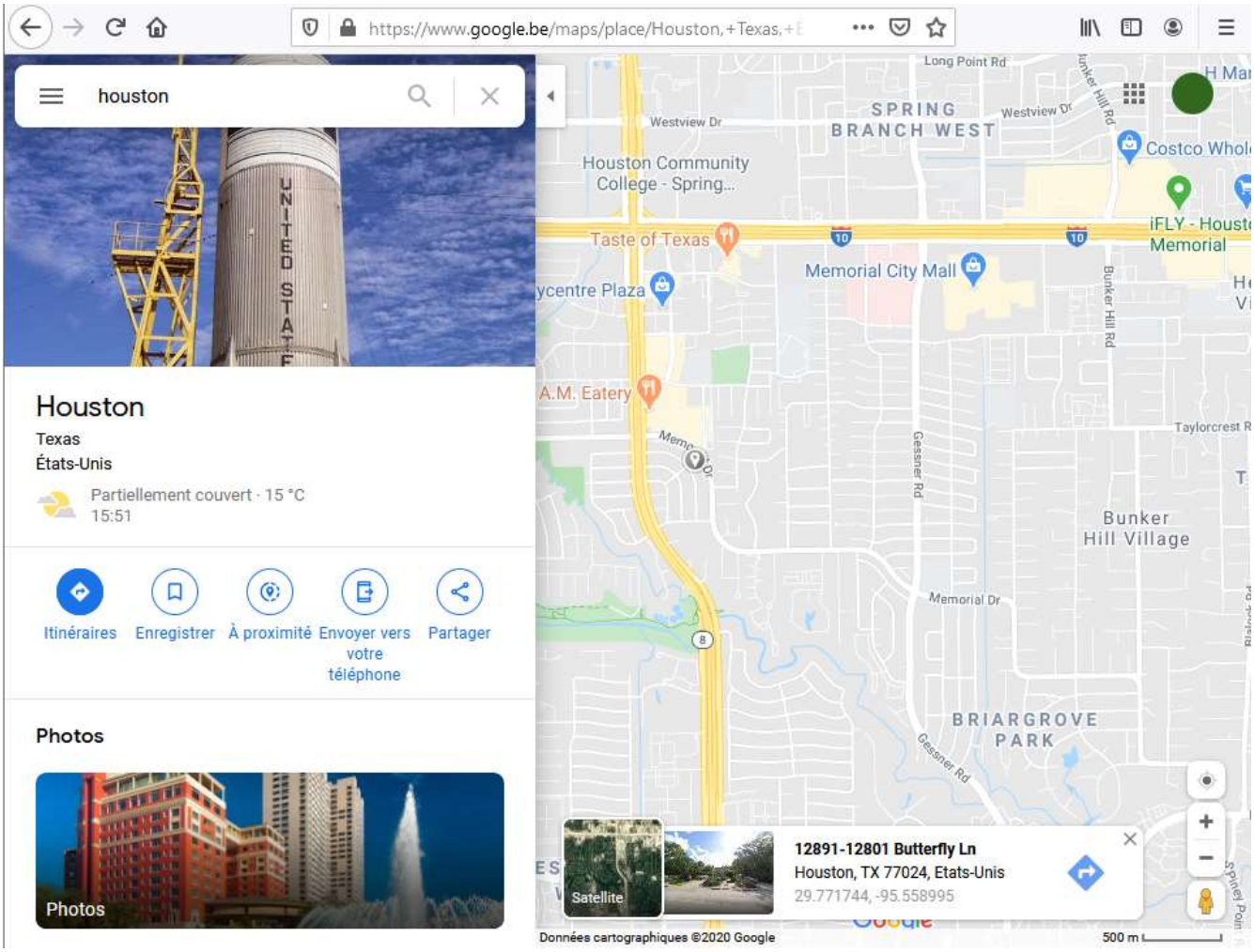
### 2.2.2 Coordinate from ‘Search Nearby’ from Google Maps

Maps is a service provided by Google allowing user to navigate on a world map and get information on what venues there is nearby.

The services has been using in this works for

- preliminary location explorations;
- and to get coordinates of the center of the destination town and of the selected origin address

The figure below show a snapshot of Google Maps centered on Houston:




## 2.2.3 Criminal records

The dataset used contains a breakdown of Group "A" Offenses for which HPD wrote police reports. The data is broken down by police districts and beats, and displayed by street name and block range.

Those data have been used first assigning each entries to a subdivision of the destination town, then a criminality score have been determined by summing each entries. A weight have been given according to the offense type:

1. no physical damage (theft, fraud) = 1 point
2. simple assault = 2 points
3. aggravated assault = 3 points

4. Rape = 5 point
5. Homicide = 10 point

An extract of the orignal data is shown here under:

| | A | B | C | D | E | F | G | H | I | J | K | L | M |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Incident | Occurrence Date | Occurrence Hour | NIBRS Class | NIBRSDescription | Offense Count | Beat | Premise | Block Range | StreetName | Street Type | Suffix | ZIP Code |
| 2 | 5619 | 01/01/2019 | 0 | 290 | Destruction, damage, vandalism | 1 | 9C30 | Residence, Home (Includes Apartment) | 9622 | SAN CARLOS | | | 77013 |
| 3 | 17319 | 01/01/2019 | 0 | 35A | Drug, narcotic violations | 1 | 7C10 | Highway, Road, Street, Alley | 0 | EAST | FWY | | 77020 |
| 4 | 18119 | 01/01/2019 | 0 | 290 | Destruction, damage, vandalism | 1 | 16E40 | Residence, Home (Includes Apartment) | 16718 | LONE QUAIL | CT | | 77489 |
| 5 | 19019 | 01/01/2019 | 0 | 520 | Weapon law violations | 1 | | Residence, Home (Includes Apartment) | 1909 | MELBOURNE | | | 77026-000 |
| 6 | 20519 | 01/01/2019 | 0 | 13A | Aggravated Assault | 1 | 15E30 | Residence, Home (Includes Apartment) | 4034 | OSBY | DR | | 77025 |
| 7 | 20519 | 01/01/2019 | 0 | 23H | All other larceny | 1 | 15E30 | Residence, Home (Includes Apartment) | 4034 | OSBY | DR | | 77025 |
| 8 | 20519 | 01/01/2019 | 0 | 290 | Destruction, damage, vandalism | 1 | 15E30 | Residence, Home (Includes Apartment) | 4034 | OSBY | DR | | 77025 |
| 9 | 20519 | 01/01/2019 | 0 | 35A | Drug, narcotic violations | 1 | 15E30 | Residence, Home (Includes Apartment) | 4034 | OSBY | DR | | 77025 |
| 10 | 31119 | 01/01/2019 | 0 | 13B | Simple assault | 1 | 19G20 | Residence, Home (Includes Apartment) | 4700 | KIRKWOOD | RD | S | 77072 |
| 11 | 31119 | 01/01/2019 | 0 | 290 | Destruction, damage, vandalism | 1 | 19G20 | Residence, Home (Includes Apartment) | 4700 | KIRKWOOD | RD | S | 77072 |
| 12 | 31119 | 01/01/2019 | 0 | 13A | Aggravated Assault | 2 | 19G20 | Residence, Home (Includes Apartment) | 4700 | KIRKWOOD | RD | S | 77072 |
| 13 | 34419 | 01/01/2019 | 0 | 13B | Simple assault | 1 | 5F10 | Residence, Home (Includes Apartment) | 7429 | LONG POINT | RD | | 77055 |
| 14 | 34819 | 01/01/2019 | 0 | 290 | Destruction, damage, vandalism | 1 | 15E30 | Residence, Home (Includes Apartment) | 4065 | SILVERWOOD | DR | | 77025 |
| 15 | 34819 | 01/01/2019 | 0 | 13B | Simple assault | 2 | 15E30 | Residence, Home (Includes Apartment) | 4065 | SILVERWOOD | DR | | 77025 |
| 16 | 39919 | 01/01/2019 | 0 | 13B | Simple assault | 1 | 20G10 | Residence, Home (Includes Apartment) | 3500 | WOODCHASE | DR | | 77042 |
| 17 | 40219 | 01/01/2019 | 0 | 13A | Aggravated Assault | 1 | 13D20 | Residence, Home (Includes Apartment) | 6502 | ROXBURY | RD | | 77087 |
| 18 | 40519 | 01/01/2019 | 0 | 13B | Simple assault | 1 | 19G20 | Residence, Home (Includes Apartment) | 6819 | COOK | RD | | 77072 |
| 19 | 41419 | 01/01/2019 | 0 | 290 | Destruction, damage, vandalism | 1 | 7C20 | Residence, Home (Includes Apartment) | 4521 | KASHMERE | ST | | 77026 |
| 20 | 43019 | 01/01/2019 | 0 | 35A | Drug, narcotic violations | 1 | 10H40 | Highway, Road, Street, Alley | 1160 | GRAY | ST | | 77002 |
| 21 | 43019 | 01/01/2019 | 0 | 35B | Drug equipment violations | 1 | 10H40 | Highway, Road, Street, Alley | 1160 | GRAY | ST | | 77002 |
| 22 | 43719 | 01/01/2019 | 0 | 13B | Simple assault | 2 | 11H40 | Residence, Home (Includes Apartment) | 1114 | CHOATE | CIR | | 77017 |
| 23 | 44419 | 01/01/2019 | 0 | 13B | Simple assault | 2 | 10H40 | Bar, Nightclub | 2600 | TRAVIS | ST | | 77006 |
| 24 | 45919 | 01/01/2019 | 0 | 120 | Robbery | 1 | 1A30 | Highway, Road, Street, Alley | 2200 | SOUTHWEST | FWY | | 77098 |

## 2.2.4 Housing pricing from Redfin

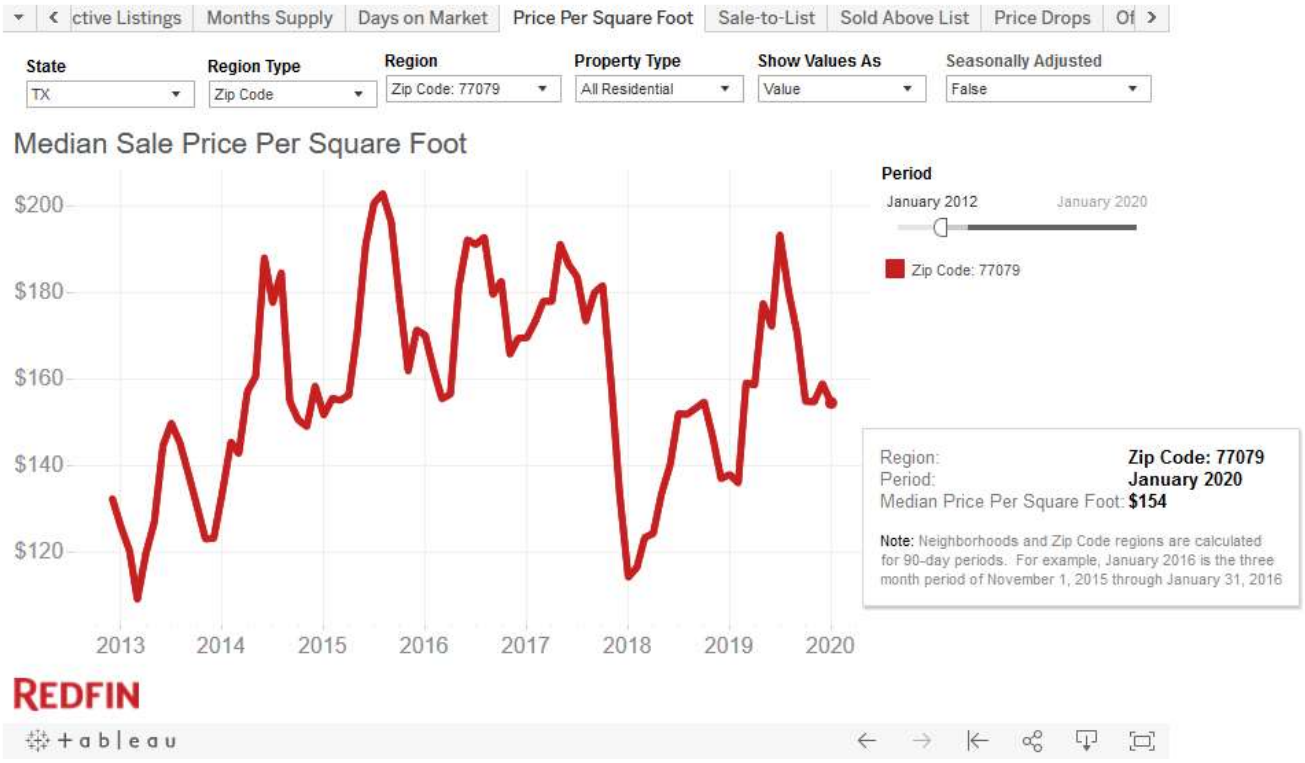
Housing pricing in the destination Town have been extracted from Redfin database, a US national real estate brokerage.

Price per sq foot for renting and buying have been used for each zipcode have been used. Prices have been assigned for each destination town subdivision.

For the subdivision of interest, the prices have been processed with the criminality score calculated using HPD dataset in order to determine the best subdivision: for each category of subdivision determined using Foursquare, what was the safer subdivision for a price range.

Here under is an example of information provided by Redfin

# Home Prices, Sales & Inventory



# 3 References


1. DAILY MAIL, *"More people than ever living outside their home country". 12 September 2013.* [viewed 5 March 2020].
   Available from: http://www.dailymail.co.uk/news/article-2418902/More-people-living-outside-home-country-Number-migrants-worldwide-hits-232-million.html
2. WIKIPEDIA, *"Liege"* 1 March 2020.
   [viewed 5 March 2020.].
   Available from: https://en.wikipedia.org/wiki/Liège
3. WIKIPEDIA, *"Houston"* 1 March 2020.
   [viewed 5 March 2020.].
   Available from: https://en.wikipedia.org/wiki/Houston
4. https://developer.foursquare.com
5. https://www.houstontx.gov/police/cs/Monthly_Crime_Data_by_Street_and_Police_Beat.htm
6. https://www.google.be/maps/
7. https://www.redfin.com/blog/data-center/