**Coursera Capstone**

## Capstone Project - The Battle of Neighborhoods

**Kripa Ranjit**

**Housing recommendation based on the crime rate in Austin area**

# Problem Statement

People usually want to live in a peaceful environment. Due to the crime rates in the area people tend to have negative living experience. People find it very hard to find a suitable place and neighborhood to accommodate them and their families. Due to crime rates and other multiple issues, housing has been struggling. It would be so much easier if people could look over the areas with less crimes when moving to a certain area.

With the annual crime rate dataset, people can derive useful information about the crime rates in different areas of Austin. This would help the people to make decision on where to move.

**Target audience: -**

Potential people who are planning to move to Austin from different city or state but are skeptical due to lack of knowledge and volatile crime rates.

**Stakeholders: -**

1. Sellers
2. Buyers
3. Real estate agents

# Data Section

Following sources of data are used while executing the Capstone Project: -

* **Data title: -**

Open Data published by data.austintexas.gov.

**Type of data: -**

Dataset in form of CSV file

**Duration: -**

January 2015 – December 2015

**Description of the dataset: -**

The data includes the crimes that occurred in Austin in 2015.

**Source: -**

https://data.austintexas.gov/Public-Safety/Annual-Crime-Dataset-2015/spbg-9v94/data

* **Data title: -**

Google Maps Geocoding API

**Type of data: -**

JSON

**Duration: -**

N/A

**Description of the data: -**

Location coordinates obtained by Gmaps API calls.

Location Information obtained from dataset is used to obtain the location coordinates from Google Maps.

**Source: -**

Google Cloud Platform/ Google Maps

* **Data title: -**

Foursquare location data

**Type of data: -**

JSON

**Duration: -**

N/A

**Description of the data: -**

Location coordinates obtained by Foursquare API calls.

To determine which area has the highest crimes.

**Source: -**

<https://foursquare.com/>

# Methodology

Annual Crime Dataset contains the crimes that occurred in the Austin area in the year 2015. This is an open dataset which is hosted on <https://data.austintexas.gov/Public-Safety/Annual-Crime-Dataset-2015/spbg-9v94/data>.

This project is focused on investigating the number of crimes that happened in Austin and recommend housing on that basis. In this manner the clients are able to move to the location which has a smaller number of crimes.

The automated script developed as a part of this project does the following: -

* 1. Parse the necessary data from the dataset which includes the data of crime.
  2. The data is cleansed, and any data of crime happened in the Austin area is presented. The data is further condensed by selecting the zip codes and street names.
  3. Unique “Street names” in the city of Austin where crimes occurred are filtered from the dataset.
  4. Location coordinates (latitude, longitude) of these street names are fetched by making API calls to Google Maps. A separate one-time Python script was developed to fetch this data and store it in a CSV file.
  5. The count function is used to get the number of crimes in each of the zip codes.
  6. Based upon the highest crime in any zip code, the map shows the plots for the location.
  7. The recommended locations are further fed into Foursquare API calls to determine various venues in proximity to them. All reported venues are then tabulated and presented to the user.
  8. Important facilities like Hospitals, Grocery stores, Elementary schools, High Schools are searched in vicinity of each location and then reported in a tabular form to the user.

# Results

With the data, I tried to extract Location which had the highest number of crimes along with the street name, the crime and its Latitude and Longitude. The street names are plotted on the map of Austin with the highest crime rate.

Important facilities are also presented to the user in a tabulated format to take care of his familial needs.

# Discussion

Based upon the findings in the results section, the user can take a conscious decision about choosing a street/ location based upon his/ her requirements.

The results section enlists places where a prospective client can buy a property based upon his needs and choices. With the highest rate, the client knows which area they should avoid.

Few possible cases are: -

* 1. A prospective client with elders in the family would be inclined to choose a location where hospitals and grocery stores are located in close proximity.
  2. A prospective client with kids in the family would choose a location where elementary and high schools are close-by. He would also like to choose a place with parks and other venues to accommodate his family are in the close vicinity.
  3. A bachelor would be inclined to choose a property which has pubs, bars, entertainment places close to the property.

# Conclusion

The decision of a buyer is influenced by the familial needs, personal biases and more importantly the safety of the family. So, based upon the findings summarized in the results and discussion sections, following conclusions can be made: -

* 1. While making recommendations to a prospective client, it is imperative to know which area they were looking for. This would help them know if the area they were targeting had any crimes in that year.
  2. Knowledge about the highest occurrence of crimes in any area can be very helpful for the client to decide which location is better for them.

# 