

## **1. Data acquisition and cleaning**

The data to use in this project, is data based on United States Cities by crime rate based on crime report statistics from 2017 for the topmost populous cities in America. The data is found on the following link: [https://en.wikipedia.org/wiki/List\\_of\\_United\\_States\\_cities\\_by\\_crime\\_rate](https://en.wikipedia.org/wiki/List_of_United_States_cities_by_crime_rate). In the United States, some crimes have been identified in different states and they are present in different types. There are 3 types of crimes Violent crimes, property crimes and Arson crimes. The data provided is based on Crime rates per 100,000 people per year and the population numbers are based on U.S. Census estimates for the year end. The file contains data for 36 states, their population, and different crimes in different cities. The safe place must be a place with less rate number of crimes. Based of the information in the file, we will explore the states cities and the crimes recorded which is most important in choosing the safe place to do the business.

Foursquare API will help to collect data to locate different places and get coordinates of states, cities, neighborhoods and get knowledge on the venues in different cities. With the data related to the venues, we can now the existence of competitors of the business and also to know other options of business that can be done in a city. The Nevada state is the state of interest and details information must be analysed in order to decide the right place for restaurant business. Nevada state has 121 cities and the information about the population is available on the following link: [https://www.nevada-demographics.com/cities\\_by\\_population](https://www.nevada-demographics.com/cities_by_population). The population data helps as to know where there are a large population so that we can get clients. The data for neighborhoods that will be used is at the link: <http://www.city-data.com/nbmaps/neigh-Henderson-Nevada.html>. This data helps to get information on the neighborhoods and their population.