

Alcohol Outlet Rate and Crime Relationship

by Adil Warsi

1. Introduction

1.1 Background

Alcohol abuse can increase the likelihood that individuals will commit certain crimes such as assault or homicide either by reducing their inhibitions or judgment or by increasing their agitation and anger. There are three primary ways in which alcohol impacts crime. Being intoxicated in inappropriate circumstances can be a crime, such as drunk in public and driving under the influence (DUI). Alcohol abuse can increase the likelihood that individuals will commit certain crimes such as assault or homicide either by reducing their inhibitions or judgment or by increasing their agitation and anger. Finally, being the victim of certain types of crimes such as child or sexual abuse makes it considerably more likely that some individuals will develop alcohol abuse disorders later in life.

1.2 Problem

Alcohol can severely impact an individual's judgement, response time, actions, and aggression level. As a result, an intoxicated individual can put them and others in extreme danger of physical and emotional trauma. To protect everyone, all jurisdictions in the United States make it a crime to be intoxicated under certain circumstances. The most common include.

1.3 Interest

Exploring alcohol related crimes is important for understanding the full impact of alcohol on individuals and society at large. Criminality may not be the first symptom that people associate with problematic drinking, but it's a significant symptom, nonetheless.

So in this project we tried to find out the relationship between alcohol outlet which is the main source of alcoholic drinks and crime rate in the **San Francisco neighborhoods**.

1.4 Prediction Model

This project is done to find the most habitual place and neighborhood for work or leaving in San Francisco, by finding the direct relationship between alcohol outlet Crime rate per

neighborhood.

2 Data Preprocessing

2.1 Data Sources

A. Kaggle

I found the San Francisco crime data from Kaggle a .csv file which has various necessary features which require to analyze such as crime location, type of crime, etc. Other Features in data sets are: -

- I. **Dates** : The time on which crime was committed.
- II. **Category** : Type of crime committed.
- III. **Description** : Crime details.
- IV. **PdDistrict** : Comes under which district police department.
- V. **Address** : Location of crime.
- VI. **X and Y** : Coordinates of crime.

B. Folium

I utilized the Foursquare API to explore the neighborhood and find the Alcohol outlet crime relationship present in that neighborhood. I designed the limit of a radius 500 meter for each neighborhood from their given latitude and longitude informations. After cleaning the data, the important features are present in data sets are:-

- I. **Neighborhood** : Name of neighborhood.
- II. **Neighborhood Latitude and Longitude** : Coordinates of neighborhood.
- III. **Venue** : Name of Bar / Wine shop venue.
- IV. **Venue Latitude and Longitude** : Coordinates of Venue.

C. socrimedata.org

To find the frequency or rate of crime per neighborhood, I use the socrimedata.org which have detail crime report for each neighborhood. After processing the feature in datasets present are: -

- I. **Neighborhood** : Name of neighborhood.
- II. **Crime Rate** : No of crimes committed per neighborhood.

2.2 Data Cleaning

The neighborhood data is taken through Foursquare API, as the project is on alcohol crime relationship, so I have to first find out all the venues in San Francisco neighborhoods and then filter out only those venues which are in categories Bar, Liquor, Beer, Whiskey, Pub, and Wine.

There are several problems with the datasets. First, venues were identified by their names. However, there were different venues with the same names, which cause their data to mix with each other's. The neighborhood data is taken through Foursquare API, as the project is on alcohol crime relationship, so I have to first find out all the venues in San Francisco neighborhoods and then filter out only those venues which are in categories Bar, Liquor, Beer, Whiskey, Pub, and Wine.

Neighborhoods and co-ordinates they situated, I decided that it was not worth the large effort to make so, because such venues only accounted for ~1% of the data. Therefore, venues with duplicate names were removed.

Second, multiple entries existed for venues that serve alcohol drinks. This cause their location data to represent multiple samples with incomplete data. I wrote script to extract.

Total categories venues for these neighborhoods and discarded partial categories rows.

2.3 Feature Selection

After data cleaning, there were 260 samples and 7 features in the data. Upon examining the meaning of each feature, it was clear that there was some redundancy in the features.

Such total vs. the rate of crimes also existed between other features.

These two features contained very similar information with the difference being that the former feature increased with no of crimes, while the latter feature did not.

Such total vs. the rate of crimes also existed between other features. These features are problematic for two reasons. : (1) A time of crimes were duplicated in two features. (2) Date of crime were duplicated in multiple features.

In order to fix this, I decided to keep all features that were rates in nature, and drop them cumulative counterparts.

After discarding redundant features, I inspected the correlation of independent variables and found several pairs that were highly correlated (Pearson correlation coefficient > 0.9).

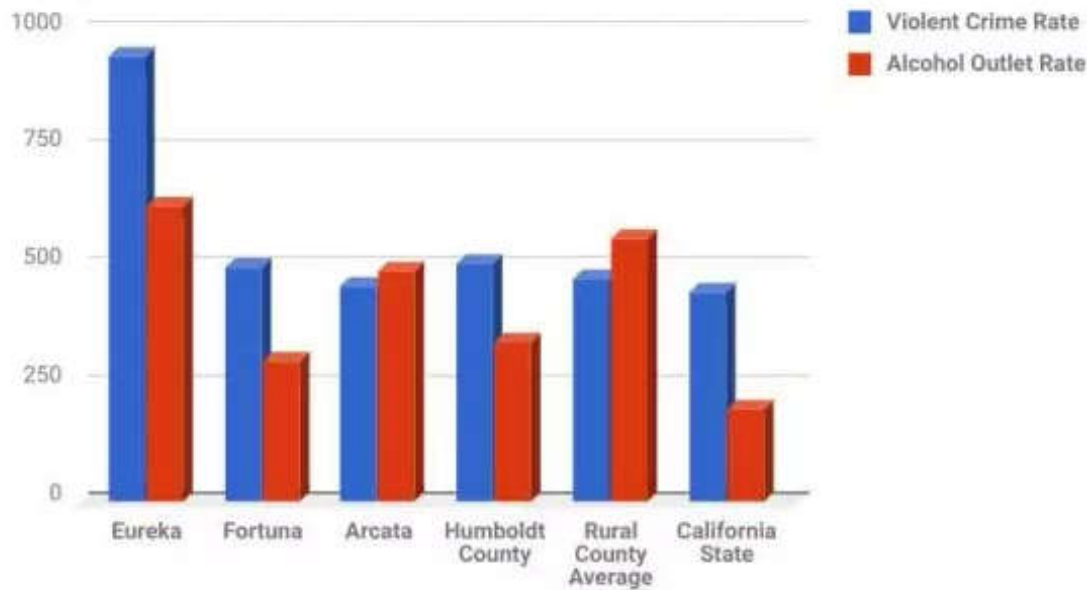
For example, crime attempted, crime made were highly correlated. This makes sense, after all, it is only the rate of crimes.

From these highly correlated features, only was kept, others were dropped from the dataset. After all, 7 features were selected.

3. Methodology

A number of individuals that serve time in jail have committed alcohol-related crimes. Offenses range from minor to serious and include property crime, public-order offenses, driving while intoxicated, assault and homicide. On average, roughly 40 percent of inmates who are incarcerated for violent offenses were under the influence of alcohol during the time of their crime. Many of these criminals had an estimated blood alcohol content (BAC) level of more than three times the legal limit at the time of their arrest.

Violent Crime & Alcohol Outlet Rate

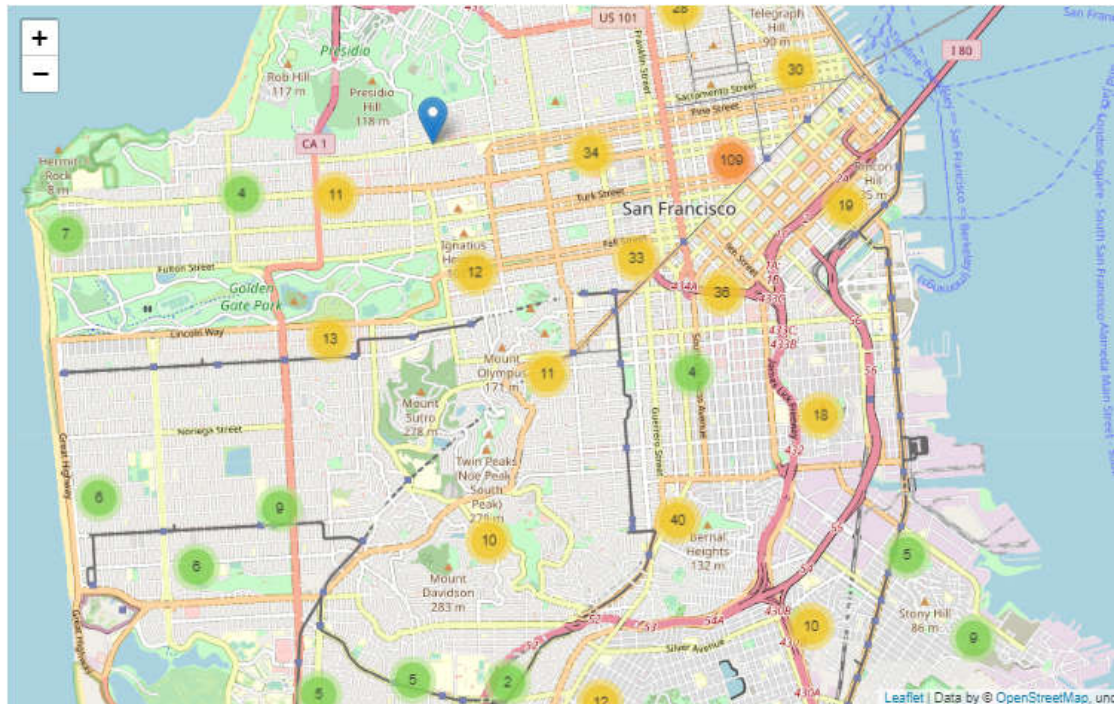


In Big Cities there are areas which are more pretend to crimes and also having the high alcohol outlets rate. So in these project we were finding the relationship between crimes and alcohol outlet rate in San Francisco city.

For finding the relationship we need the geographical and outlet venues San Francisco data, for which I used python Folium library to visualize geographical details and Foursquare API to get outlet venues location details respectively. So by using the outlet venues latitude and longitudes values, plot them on San Francisco map using folium.

Initially, I have to find out all the name of neighborhoods which were in San Francisco, for that I used web scrapping using python BS4 Library to get the data from "Wikipedia Neighborhood in San Francisco" webpage from which I get 94 neighborhood names. Then in Foursquare API using the neighborhood venues feature, then I listed out all the venues which are in San Francisco City. The dataset contains the features Name of Neighborhood, their latitude and longitude, name of venues and their latitude and longitude and venue category. Then from venue category I filtered out those venues which comes in category of Bar, Liquor, Beer, Whiskey, Pub, and Wine. Finally, I get the total 260 venues which are alcohol outlets.

By plotting the venues on San Francisco map I get the following result.



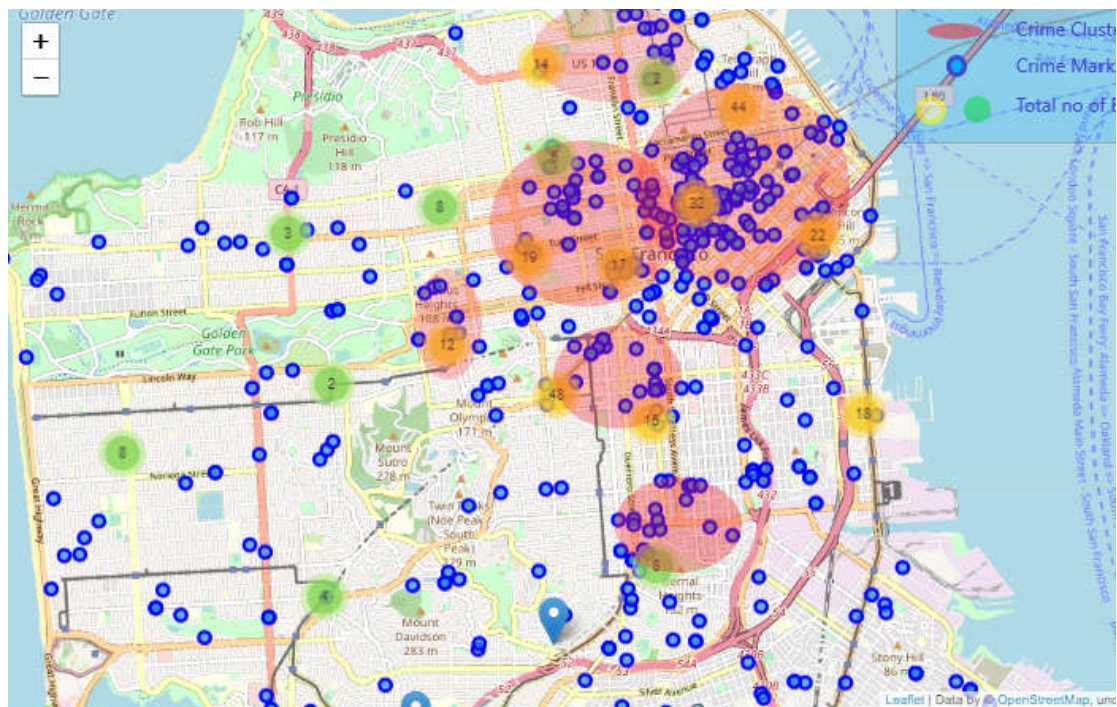
Which represents the no. of alcohol outlets per neighborhood.

I found San Francisco crime data on sfscrimedata.org, which has huge datasets of San Francisco crime having features of location and time range selection. They provide clean data so you don't have to spend time on cleaning. In preprocessing I have to find out the crime rate per neighborhoods so that I can merge it with alcohol outlet venues data. The Data head is

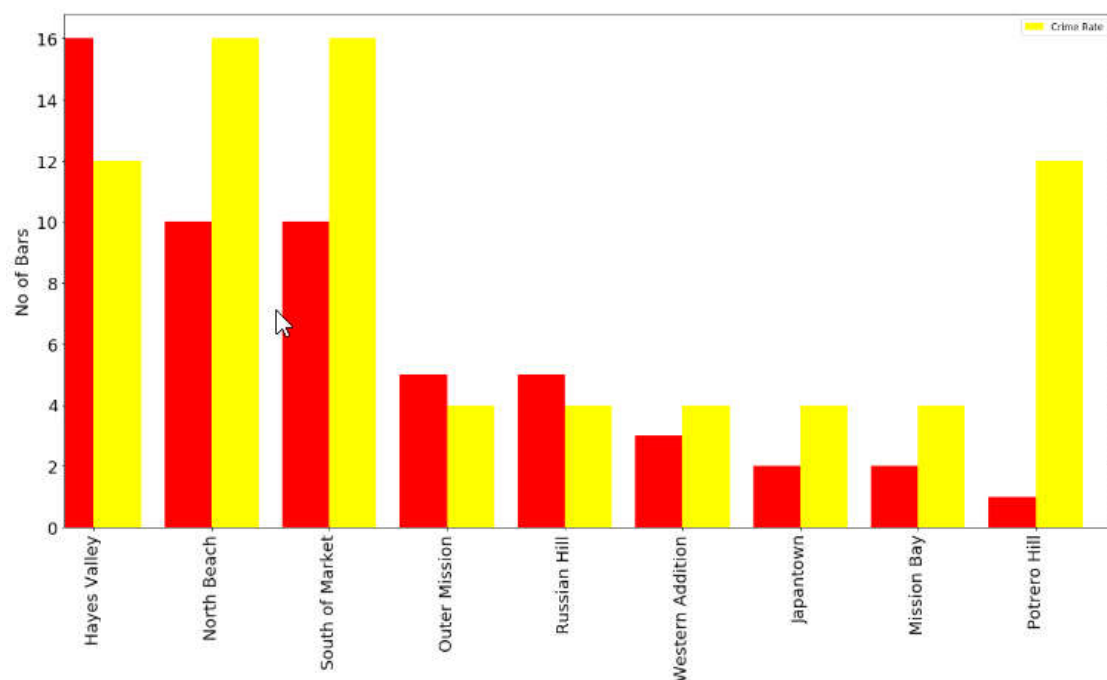
incident_date	incident_datetime	analysis_neighborhood	incident_category	incident_category.1	resolution	police_district	incident_descor
2019-07-07T00:00:00.000	2019-07-10T13:41:00.000	Bayview Hunters Point	Drug Violation	Drug Violation	Cite or Arrest Adult	Bayview	Firearm, Armed Possessive Controlled
2019-07-8T00:00:00.000	2019-07-08T20:44:00.000	Tenderloin	Drug Violation	Drug Violation	Cite or Arrest Adult	Tenderloin	Firearm, Armed Possessive Controlled
2019-07-4T00:00:00.000	2019-07-04T19:09:00.000	West of Twin Peaks	Liquor Laws	Liquor Laws	Open or Active	Taraval	Alcohol, Possessive Of By
2019-07-2T00:00:00.000	2019-07-02T16:35:00.000	Mission	Drug Violation	Drug Violation	Cite or Arrest Adult	Mission	Firearm, Armed Possessive Controlled
2019-07-2T00:00:00.000	2019-07-02T16:37:00.000	Mission	Drug Violation	Drug Violation	Cite or Arrest Adult	Mission	Firearm, Armed Possessive Controlled

4. Results

From the map it can clearly shows that the crime rate is high in those areas which has high no. of alcohol outlets.



In the above figure the red clusters represents those areas which has crime rate more than 6 per 500 meters. Those clusters also include high number of alcohol outlet rate.



Also from the above bar charts it clearly shows the that the Crime Rates and Alcohol outlets are highly correlated.

5. Conclusion

The total number of measurements and crime densities of the 94 neighborhoods in total can vary. As there is such a complexity, very different approaches can be tried in clustering and classification studies. Moreover, it is obvious that not every classification method can yield the same high quality results for this metropole.

I used the Kmeans algorithm as part of this clustering study. When I tested the Elbow method, I set the optimum k value to 5. For more detailed and accurate guidance, the data set can be expanded and the details of the neighborhood or street can also be drilled.

I also performed data analysis through this information by adding the coordinates of neighborhoods and crime rate as static data on GitHub. In future studies, these data can also be accessed dynamically from specific platforms or packages.

Moreover, it is obvious that not every classification method can yield the same high quality results for this metropole. In future studies, it can be used for to find out the more suitable place in the city.

6. Reference

- [1] San Francisco Neighborhoods - Wikipedia
- [2] Folium
- [3] Crime Data - Sfcrimedata.org
- [4] Google Data
- [5] www.alcoholrehabguide.org
- [6] www.addictioncenter.com