**CANNABIS IN DENVER AND CRIMES REPORTED RELATED TO MARIJUANA**

*Suman Saurabh*

1. **Introduction**

**1.1 Background**

The Colorado amendment 64, which was passed by voters on November 6, 2012, led to legalization in January 2014[1]. Many reports suggests that use of marijuana has decreased in teenagers from 25% to 21%, which is an opposite effect of legalizing “the weed”[2]. There has been also a decline in the report of marijuana-abuse related hospitalization from the 2005-2014. The state has also been able to make a good revenue, with $266 million in taxes alone in the year of 2018. Most of the sales (98%) has been reported to be from recreational marijuana. There has not been conclusive evidence of the extremely bad effects of marijuana on mental health of a used except inhaling smoke can always be dangerous and detrimental to lungs.

One the other side of the story the City and County of Denver reports data on its website on marijuana related crime[3]. The website mentions that the crime related are in “clear connection or relation to marijuana”. Now these crimes reports can really hurt setting up new business or consumers who use are recreational in nature.

* 1. **Problem Statement**

The problem statement addressed in this report is, “How the crimes related to marijuana can affect the new businesses location where to setup? and which areas and stores can be put under more surveillance by authorities to better control it? Or where to go for a safer marijuana procurement?

To answer these questions, I decided to investigate the data and see how these reported crime neighborhoods are related to the marijuana sales locations in the city because there cannot be an illegal marijuana. The approach to the study is that data provided by the city of Denver will be used to identify the neighborhoods with maximum marijuana related crime will be identified. The types of crime that are mostly reported related to marijuana will be identified. Next, marijuana stores and businesses will be identified using Foursquare database and mapped. Then on the same map neighborhood with most reports will be plotted. Finally, we will run K-means clustering to identify the centroid of the clusters of the stores which are safer for new business and users who procure marijuana. Basic statistical analysis will be done to identify the spread of crimes reported over time., the types and the neighborhood which sees most reports and machine learning K-means will be used to identify the clusters centers.

1. **Data**

**2.1 Brief account on data source**

The data to be used in the project is crimes related to marijuana reported by City of Denver and Foursquare location data base. A brief description is included below:

* Denver Metropolitan Crimes Related to Marijuana (reported on website of City of Denver)[1].
  + Uses: Basic statistical analysis will be done to identify the spread of crimes reported over time, the types and the neighborhood which sees most reports
* Foursquare database to identify the marijuana stores in the Denver metropolitan area.
  + Uses: To run K-means clustering on latitude and longitude data, to identify the centroid of the clusters of the stores which are safer for new business and users who procure marijuana[2].
  1. **Data Description**

The data on crimes related to marijuana reported by city of Denver is a .csv file that can be accessed from the website “Open Data Catalogue” hosted by city of Denver for free public access.

The data is a relational structured data in form of an excel sheet with 14 columns and 1203 entries in rows. The columns contain data on the following features:

* Incident\_Id (Primary Key)- Unique for each reported incident.
* First\_Occurance\_Date- Date of first occurrence of a crime reported in the neighborhood or the address related to marijuana.
* Last\_Occurance\_Date – Date of first occurrence of a crime reported in the neighborhood or the address related to marijuana (both first and last occurrence date are same for most reports, except for majority of last dates are missing).
* Report date for each incident.
* Address of each incident
* Geo\_X and Geo\_Y (geographic location system for the reported incident, not latitude or longitude data for the addresses of incidents).
* District
* Precinct
* Offense code, type\_Id and category\_Id
* MJ Relation (Industry or Non-Industry)- related to marijuana industry or not.
* Neighborhood\_Id

**2.3. Data Cleaning**

No data cleaning was required because the data was available in a very structured form, no missing data, except for Last\_Occurance\_Date, which will not be used in this project. Hence, the next step was to select the features related to the project.

**2.4 Feature selection**

Features that can help in temporal, spatial and nature of the crime reported were selected as the features of our interest. Hence, the selected features are:

1. Incident\_Id (Unique for each incident)
2. Report date – to track the numbers over time.
3. Incident address – to lookup latitude and longitude data
4. Offense type/category – to identify what kind of offense are the most likely reported related to marijuana.
5. Neighborhood\_Id – to identify safe neighborhood for business, tourists and users.
6. **Exploratory Data Analysis**

**Scatter plot:**

A scatter plot of crimes reported date against the neighborhood on y-axis and report date on x-axis revealed that the number of crimes reported because of marijuana is decreasing over the years, but for some neighborhoods it was high in the beginning and is still relatively high. Some of these neighborhoods identified with the exploratory scatter plot are:

* Elyria-Swansea
* Overland
* Athmar-park
* Northeast-park-hill
* Montebello

The scatter plot is presented as Figure 1 in Appendix I.

**Sorting and Pivot:**

1. Identifying the most common type of reported crime: To identify the crime category that has been reported over time, we pivoted the dataframe for the crime category and sorted the values by numbers in descending order. The crime that were reported in highest numbers were burglary- *business by force, theft, criminal mischief, robbery and shoplifting* (Table1, Appendix II). However, the number of burglaries to business by force outweighs all the other categories by far. Hence, setting up business in the neighborhoods with high crime reports is going to hurt a new business setup.
2. Identifying the neighborhoods with most numbers of crime reported: Similar exercise to that described above was done for the neighborhoods to identify the top 10 neighborhoods with most numbers of crimes reported presented in Table (Appendix II).
3. Use of describe method to get primary statistical figures:

* Most numbers of incidents, (6) were reported on 16th May 2019, which is recent.
* Most of the incidents reported has been tagged to be industry related (966 out of 1203)
* Most common crime reported was burglary to businesses (674 out of 1203, which is more than half of it).
* The neighborhood where most crimes were reported was Elyria-Swansea (107 incidents).

**Latitude and Longitude lookup:**

The addresses in the dataframe were used for lookup of latitude and longitude using the geopy module. However, not all the addresses were available for latitude and longitude conversion, hence google maps were used for the lookup.

**Foursquare database for JSON file of Denver metropolitan and lookup of marijuana businesses in the area:**

The foursquare database was used for retrieving Denver metropolitan JSON file. Next, Marijuana related businesses were looked up in the vicinity of Denver downtown area with a radius of 5km. 29 businesses were reported from the lookup exercise. The description of the data is below:

* The shape of the data retrieved was a 30 by 18 relational database table.
* The columns were: ['id', 'name', 'categories', 'referralId', 'hasPerk', 'location.address',

'location.lat', 'location.lng', 'location.labeledLatLngs', 'location.distance', 'location.postalCode', 'location.cc', 'location.city', 'location.state', 'location.country', 'location.formattedAddress', 'location.crossStreet', 'venuePage.id'].

* The data was cleaned to get: Name, category, latitude, longitude, distance, city, state, country, cross-street, Id.
* Describe method was used to learn that mean distance of all the stores from downtown is about 3.7 km, minimum is 225m and a maximum of 11 km within a radius of 50km from downtown.
* Next the name, category, latitude and longitude features were selected and plotted on folium to get Figure 3.

1. **Predictive modeling using K-Means clustering:**

The data was cleaned again a little by removing government buildings related to marijuana like Colorado department of revenue Medical Marijuana.

K-means clustering algorithm was run on the latitude and longitude data of the marijuana businesses in the Denver metropolitan areas to identify various clusters of the businesses and to identify which cluster is closer to the neighborhoods where there is higher chances for business burglary.

The initial iteration was run with 5 cluster centers, the result is shown in Figure 4 (Appendix I).

Next, K-means metrics of distortion and inertia was adapted to determine, what should be the number of clusters for this dataset and it was determined that 2 clusters give the best result using the elbow method (Figure 5 and Figure 6). The map with two clusters shows the result of two clusters (Figure 7).

1. **Discussion**

The distortion kept going down for increasing number of K. The cluster of stores near the downtown are dense and much safer than the stores in Elyria Swansea and towards south Broadway and University of Denver campus. It is difficult to tell anything about the safety of people who go to buy medical or recreational marijuana in Montebello neighborhood because of lack of data. The stores in the Edgewater and Glendale neighborhood are quite in good numbers and certainly safer and less crime related to marijuana reported there.

Hence, for new stores and customers of medical or recreational marijuana, new businesses opportunities lies safer in Denver downtown (for obvious reasons), however, Edgewater and Glendale neighborhood seems safer too and probably cheaper to rent a business there than Denver downtown (can't be established using the data presented in this study, but a general decrease in property value going away from downtown is expected).

1. **Recommendations based on the study**

* **For New Business**: Since the most common reported crime is burglary to businesses using force, it is an important issue to setup the location in a safer neighborhood. Safer choice is downtown area, the Edgewater neighborhood and the Glendale neighborhood area. However, the areas around down can be expensive to setup business but better doing, that’s problem for later. Hence, just for economics sake Glendale and Edgewater neighborhoods are safer and probably less expensive than downtown area.
* **For Users:** Easy to choose downtown area if that is closer, however, the safer neighborhood are gain the Glendale and Edgewater neighborhood.
* **For Authorities:** Neighborhood and clusters of businesses near the cluster 1 near Elyria Swansea should have more surveillance to avoidincreased count of similar crimes.

**REFERENCES**

[1] Cannabis in Colorado 2020. https://en.wikipedia.org/wiki/Cannabis\_in\_Colorado.

[2] Ingraham C. Economic policy n.d. https://www.washingtonpost.com/news/wonk/wp/2016/06/21/colorado-survey-shows-what-marijuana-legalization-will-do-to-your-kids/ (accessed August 4, 2020).

[3] Open Data Catalog 2020. https://www.denvergov.org/opendata/dataset/city-and-county-of-denver-crime-marijuana (accessed August 4, 2020).

**APPENDIX I**

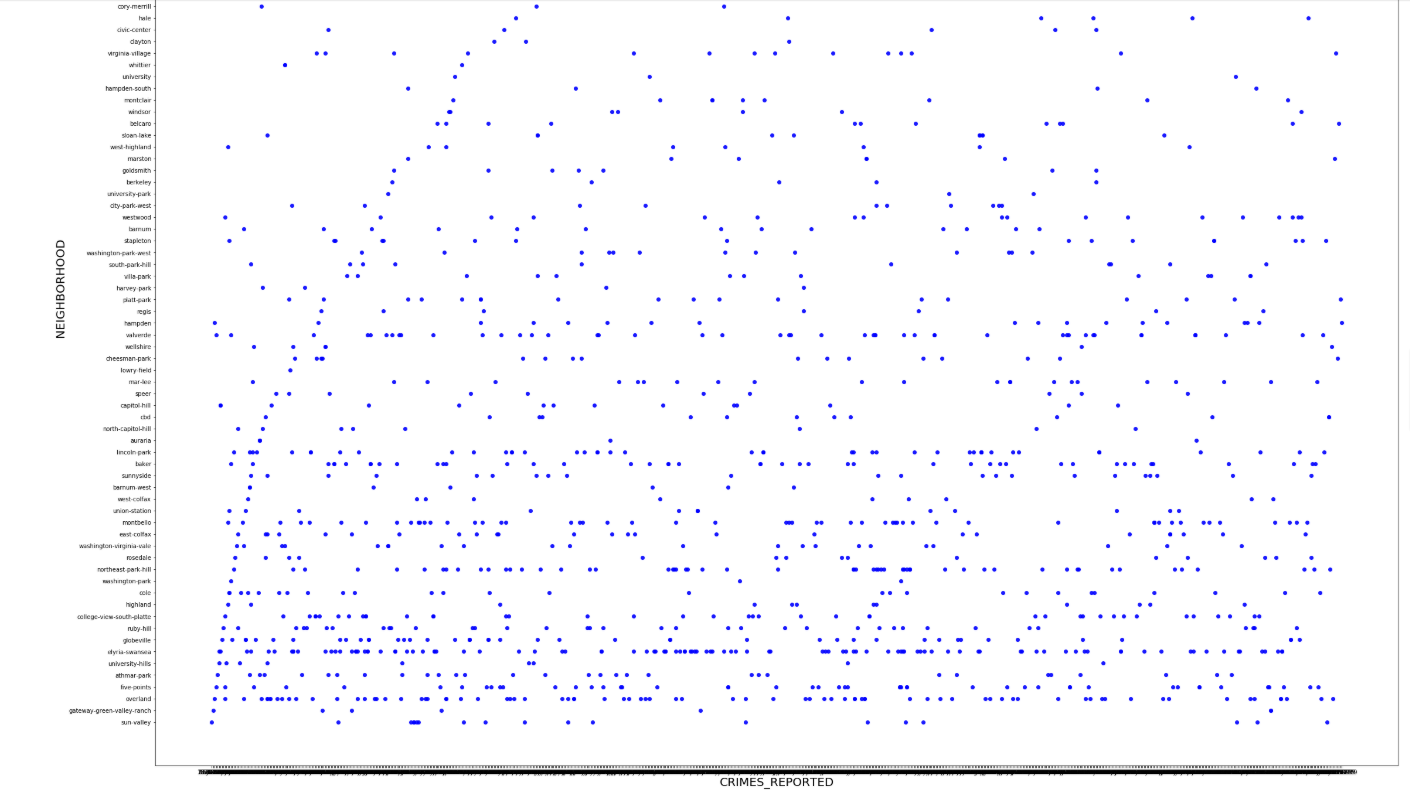


Figure 1. Neighborhood vs Report date of crimes related to marijuana. (Not visible, but neighborhood like Elyria Swansea and overland are the lower in the y-axis, they seem to be most affected neighborhood throughout the reported history. However, overall, the numbers seems to decrease every year for all neighborhoods).

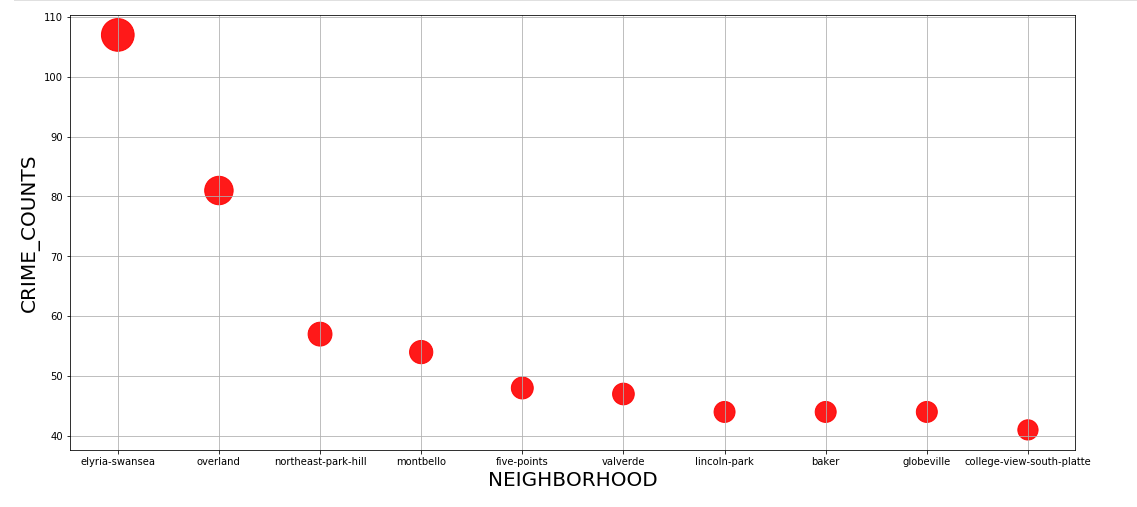


Figure 2. Crimes reported in the various neighborhoods

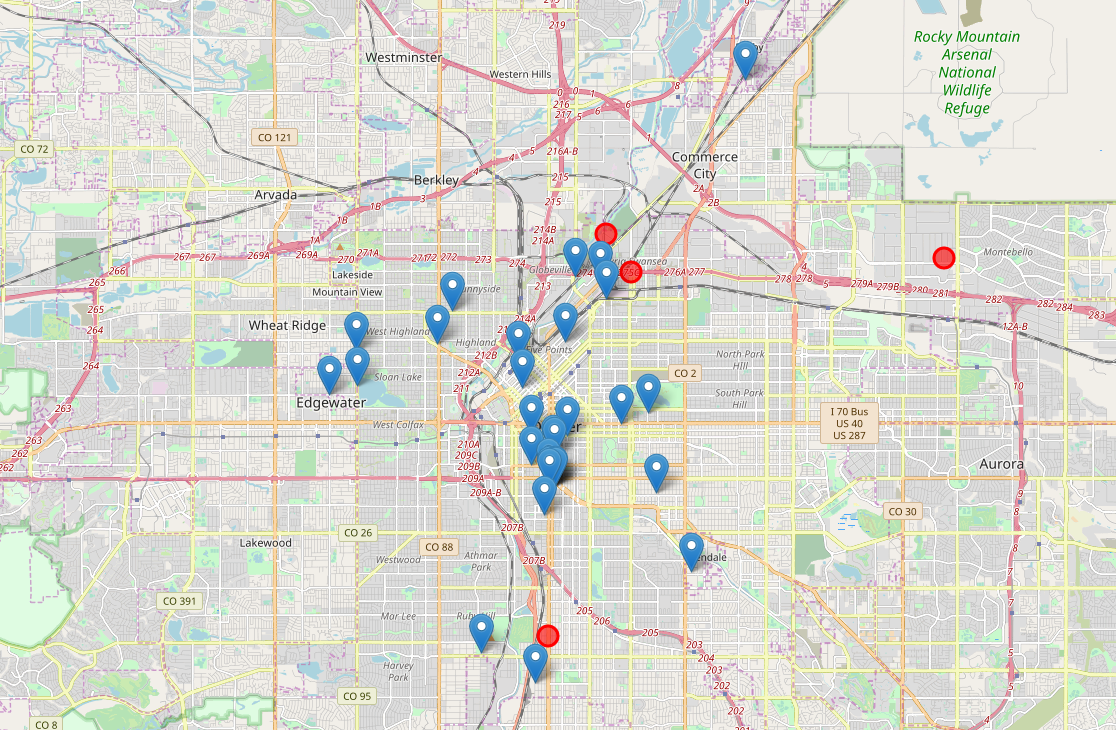


Figure 3. Folium used for plotting the marijuana businesses in the blue markers and the neighborhoods with highest reported crimes in red circles (pop ups available in the notebook version).

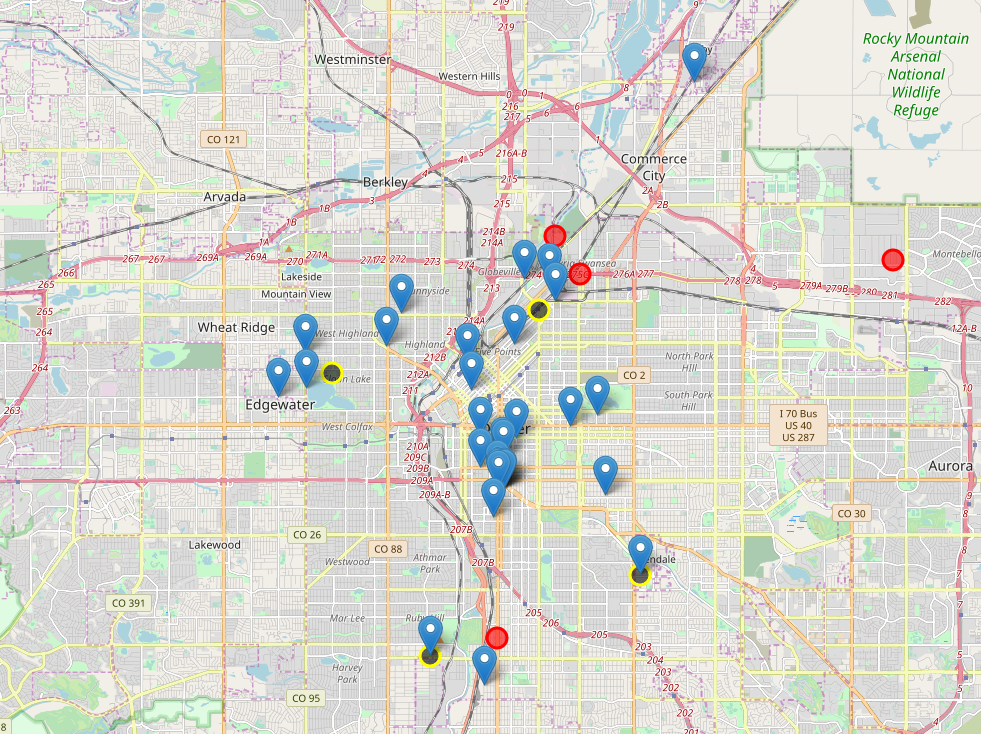


Figure 4. Folium used for plotting the marijuana businesses in the blue markers and the neighborhoods with highest reported crimes in red circles (pop ups available in the notebook version) and black with yellow outline circles for the cluster centers identified by K-means.

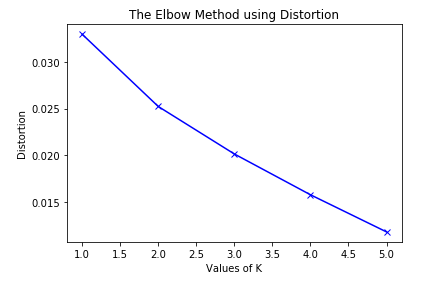


Figure 5. The elbow method used for K-means plotting distortions vs values of K

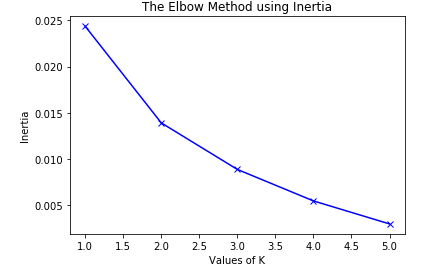


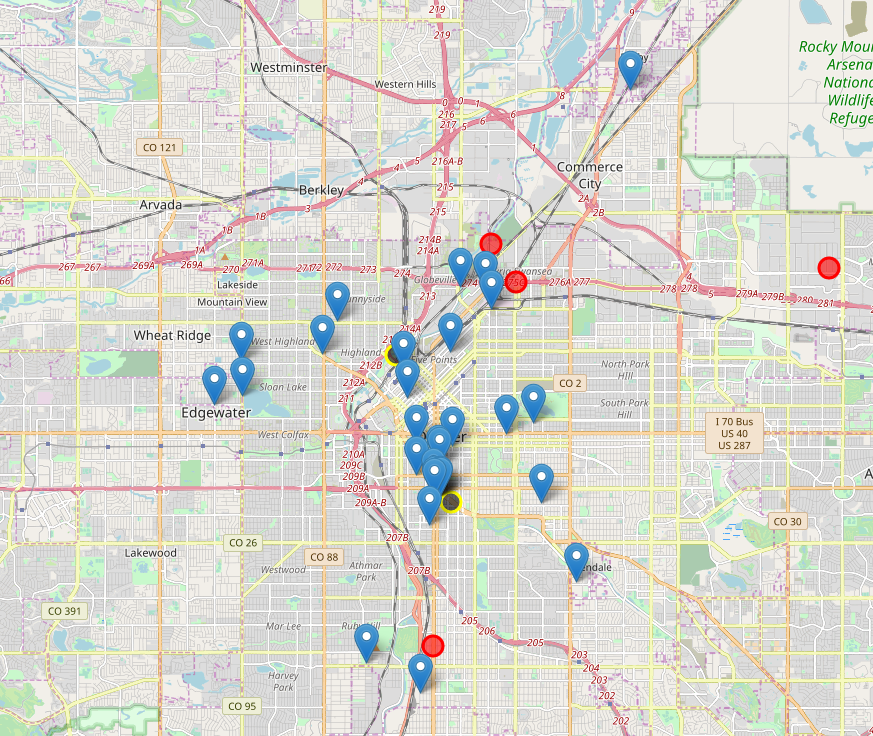
Figure 6. The elbow method used for K-means plotting Inertia vs values of K

Figure 7. Two cluster result for the marijuana stores and neighborhood.

**APPENDIX II**

Table 1

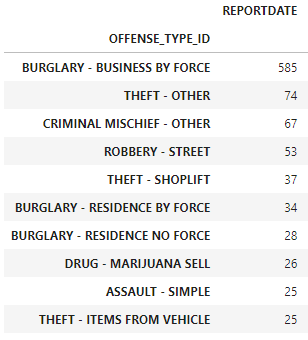


Table 2

