

Introduction/Business Problem

For my project I decided to explore whether there is a correlation between neighborhood types derived using ML algorithm and different crime types also broken down by some clusters using K-means. To me this looks to very interesting thing to explore: maybe there are some hidden dependencies that we can discover. For example: liquor shops may draw more DUI crimes nearby. So, let's see and dig into the problem and figure out what do we need to start this project.

Data Sources:

1. Philadelphia Police Department crimes data set. It includes both violent and nonviolent offences. It has location, general crime type, datetime and latitude and longitude of that address <https://www.opendataphilly.org/dataset/crime-incidents>
2. Foursquare API – similar to the course material, I will use it information about neighborhoods
3. Neighborhood shape data in GeoJson format. This will be used to display neighborhoods borders on the map and also latitude and longitude of crime will be compared vs shapes from this source to map each crime into appropriate neighborhood. https://github.com/azavea/geo-data/tree/master/Neighborhoods_Philadelphia

Methodology:

I used the datasets from above to generate vector datasets ready to be used for clustering with Scikit Learn library.

I used K-means clustering method for analyzing for patterns and grouping. To come up with the optimal clusters amount I used Elbow approach for both crime rates and neighborhood analysis. The inertia amount was starting to show less degree of declining with cluster amounts of 10 for both subsets.

Results:

The final dataset consists of each neighborhood clustered by venues and crimes. Also, each line includes top ten venues and top ten crimes

Discussions/Conclusions:

Clearly, there are some dependencies among different neighborhoods and crime types.

For example, neighborhoods clusters under 0,3 and 6 cluster only have crimes clustered 0,7 and 8 respectively.

Some neighborhoods (cluster 2,4,5) showed predominant crime type (crime clusters 5,6), while roughly half of neighborhood clusters had no clear dominant clustering in crime.

Let's try to determine if there are any similarities in clusters and whether we can spot any trends. Again, we will focus only on Neighborhood clusters that only have 1 or 2 predominant crime types (0,3,6) in it, ignoring neighborhoods with all sort of crimes.

Crime cluster 0 is a cluster with predominant all other offences and thefts from vehicles. Then Thefts, Vandalism/Fraud are trailing behind. While all of this is not really great, still it is much more "safer" than Cluster crime 7 for example, where "All other assaults" are taking 1 and the second most frequently happening place.

Let's see if we draw any conclusions: Neighborhood cluster 0 which draws so many "light" crimes is a cluster where many service establishments are located: Home Service/Vintage Store/Barbershop/Zoo Exhibit – all nonfood places with some exceptions. Accordingly, these spots with mass gatherings, though not as social as eating/drinking draw many "light crimes". On the contrary, Neighborhood cluster 6 is more of "eatery" style borough with Thai/Fast Food/Filipino restaurants. It looks like that these (not necessary these specific types of restaurants, but more like general places where people can get food and drinks) – tend to attract more violent type of crimes.

Of course, it is not a final conclusion and many other steps can be recommended to improve the accuracy (specifically on the crimes side, as there are too many "Other" types of crimes.) Still, I think as a working hypothesis that there is a working correlation between amount of eateries/bars and number of "medium and more serious than medium" crimes.