The Battle of Neighborhoods: Baltimore

1. Introduction

Baltimore, Maryland is infamous for its significantly high crime rate. Violent crime spiked in 2015 after the death of Freddie Gray, which touched off riots and has resulted in the huge increasing in murders. Baltimore recorded 344 homicides in 2015. To put that in perspective, New York City, a city with a 2015 population of 8,491,079, recorded a total of 339 homicides in 2015. Baltimore had a 2015 population of 621,849; which means that in 2015 Baltimore had a homicide rate 14 times higher than New York City's.

However, Baltimore is also a tourism city with beautiful harbor and delicious seafood. The city hosted about 25 million visitors a year, which is about 8% of the population in U.S. In order to give a better guidance to the visitors, this project will focus on the analysis crime in Baltimore and will find out the safest area in Baltimore with view suggestions according to the Foursquare location data.

2. Data Acquisition and Data Cleaning

2.1 Data Acquisition

The data for this project are two sources. The first dataset is the crime in Baltimore for 2012-2017 from Kaggle. This data presents all the crimes recorded from 2012 to 2017 in Baltimore. This data set contains the following **columns**:

- **CrimeDate**: Indicate the crime data in form of MM/DD/YYYY
- **CrimeTime**: Indicate the crime time in form of HH/MM/SS
- **CrimeCode**: Indicate the crime code of the crime event
- **Location**: Indicate the street location of the crime event
- **Description**: Indicate the type of crime, eg: Robbery/Shooting...
- **Inside/Outside**: Indicate whether the crime event happened inside or outside
- **Weapon**: Indicate the weapon used for the crime event
- **Post**: Not clear
- **District**: Indicate the district of the crime happened
- **Neighborhood**: Indicate the neighborhood of the crime happened
- **Longitude**: Coordinates Information

- **Latitude**: Coordinates Information
- **Location 1**: In form of (Latitude, Longitude)
- **Premise**: Indicate the type of location where the crime event happened.
- **Total incidents**: Indicate the number of incidents under this record (usually equals to 1)

The second dataset is the Foursquare location data which contains the venue information grouped by neighborhood. This data set contains the following **columns**:

- **Neighborhood**: The name of neighborhood
- **Neighborhood Latitude**: Coordinates information
- **Neighborhood Longitude**: Coordinated information
- **Venue**: The name of venue
- **Venue Latitude**: Venue coordinates information
- **Venue Longitude**: Venue coordinates information
- **Venue Category**: The type of venue

2.2 Data Cleaning

The Baltimore crime data from 2012 to 2017 is imported as *df*

	CrimeDate	CrimeTime	CrimeCode	Location	Description	Inside/Outside	Weapon	Post	District	Neighborhood	Longitude	Latitud
0	09/02/2017	23:30:00	3JK	4200 AUDREY AVE	ROBBERY - RESIDENCE	I	KNIFE	913.0	SOUTHERN	Brooklyn	-76.60541	39.2295
1	09/02/2017	23:00:00	7A	800 NEWINGTON AVE	AUTO THEFT	O	NaN	133.0	CENTRAL	Reservoir Hill	-76.63217	39.3136
2	09/02/2017	22:53:00	9S	600 RADNOR AV	SHOOTING	Outside	FIREARM	524.0	NORTHERN	Winston-Govans	-76.60697	39.3476
3	09/02/2017	22:50:00	4C	1800 RAMSAY ST	AGG. ASSAULT	I	OTHER	934.0	SOUTHERN	Carrollton Ridge	-76.64526	39.2831
4	09/02/2017	22:31:00	4E	100 LIGHT ST	COMMON ASSAULT	O	HANDS	113.0	CENTRAL	Downtown West	-76.61365	39.2875
5	09/02/2017	22:00:00	5A	CHERRYCREST RD	BURGLARY	I	NaN	922.0	SOUTHERN	Cherry Hill	-76.62131	39.2486
6	09/02/2017	21:15:00	1F	3400 HARMONY CT	HOMICIDE	Outside	FIREARM	232.0	SOUTHEASTERN	Canton	-76.56827	39.2820
7	09/02/2017	21:35:00	3B	400 W LANVALE ST	ROBBERY - STREET	O	NaN	123.0	CENTRAL	Upton	-76.62789	39.3025
8	09/02/2017	21:00:00	4C	2300 LYNDHURST AVE	AGG. ASSAULT	O	OTHER	641.0	NORTHWESTERN	Windsor Hills	-76.68365	39.3137
9	09/02/2017	21:00:00	4E	1200 N ELLWOOD AVE	COMMON ASSAULT	I	HANDS	332.0	EASTERN	Berea	-76.57419	39.3055

Drop the columns that are not important for this project

	CrimeDate	CrimeTime	Location	Description	Weapon	District	Neighborhood	Longitude	Latitude
0	09/02/2017	23:30:00	4200 AUDREY AVE	ROBBERY - RESIDENCE	KNIFE	SOUTHERN	Brooklyn	-76.60541	39.22951
1	09/02/2017	23:00:00	800 NEWINGTON AVE	AUTO THEFT	NaN	CENTRAL	Reservoir Hill	-76.63217	39.31360
2	09/02/2017	22:53:00	600 RADNOR AV	SHOOTING	FIREARM	NORTHERN	Winston-Govans	-76.60697	39.34768
3	09/02/2017	22:50:00	1800 RAMSAY ST	AGG. ASSAULT	OTHER	SOUTHERN	Carrollton Ridge	-76.64526	39.28315
4	09/02/2017	22:31:00	100 LIGHT ST	COMMON ASSAULT	HANDS	CENTRAL	Downtown West	-76.61365	39.28756

One thing noticed in the weapon column is that for crime events with unidentified weapon, the description has both 'NaN' and 'Other'. 'NaN' means no weapon, and 'OTHER' means other weapon. See if we have any other descriptions of weapons.

	CrimeDate	CrimeTime	Location	Description	Weapon	District	Neighborhood	Longitude	Latitude
0	09/02/2017	23:30:00	4200 AUDREY AVE	ROBBERY - RESIDENCE	KNIFE	SOUTHERN	Brooklyn	-76.60541	39.22951
1	09/02/2017	23:00:00	800 NEWINGTON AVE	AUTO THEFT	NO WEAPON	CENTRAL	Reservoir Hill	-76.63217	39.31360
2	09/02/2017	22:53:00	600 RADNOR AV	SHOOTING	FIREARM	NORTHERN	Winston-Govans	-76.60697	39.34768
3	09/02/2017	22:50:00	1800 RAMSAY ST	AGG. ASSAULT	OTHER	SOUTHERN	Carrollton Ridge	-76.64526	39.28315
4	09/02/2017	22:31:00	100 LIGHT ST	COMMON ASSAULT	HANDS	CENTRAL	Downtown West	-76.61365	39.28756

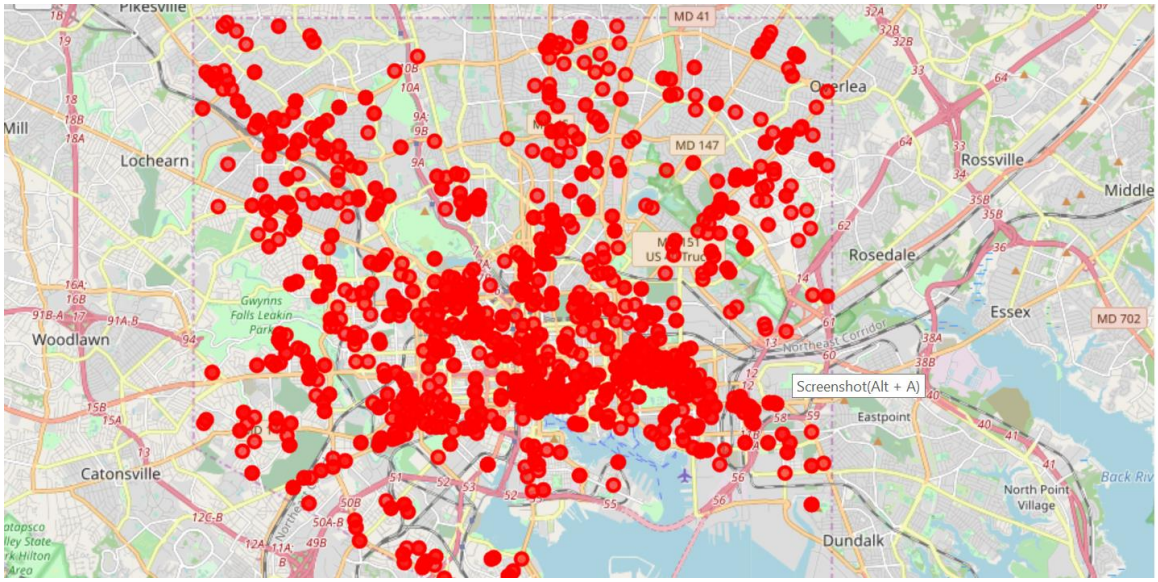
Now from the *df_bal* dataset, we have all the necessary items we need for analysis the crime events in Baltimore. Later, based on the visualization of the crime data, we will have an idea of the area with the lowest crime rate. Then we will have to use the **FOURSQUARE** data to analyze the attractions in that area and give suggestions to the tourists.

3. Methodology

3.1 Explore the crime data in Baltimore

3.1.1 Visualize the crime data in Baltimore

To have an intuitive feeling about the crime event in Baltimore, we will use the folium map the visualize the crime data. However, due to the large size of the dataset, only 1000 crime events in 2017 will be plotted into the map.



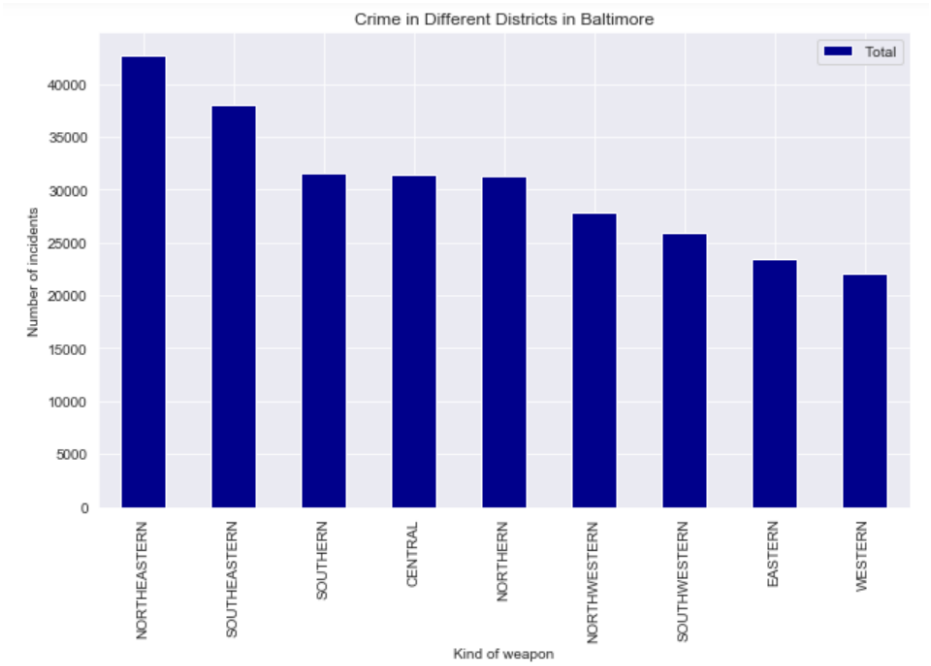
As we can see from the map, most of the crime events happened in downtown Baltimore which is near the harbor area. The areas in west, north and east are considered as more safe places since the density of crimes in relatively low.

3.1.2 Crime events in different districts

To have a statistical point of view, the total number of crimes in different district in Baltimore from 2012 to 2017 is shown as follows.

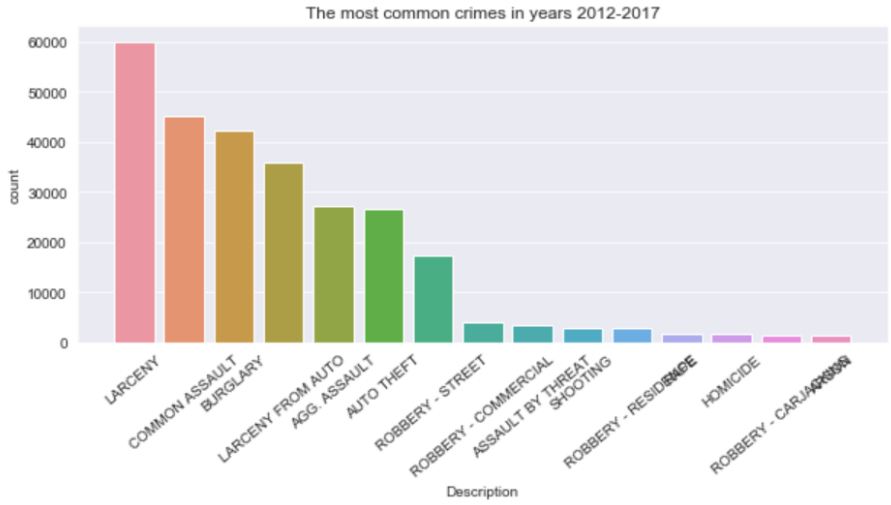
	District	Total
0	CENTRAL	31397
1	EASTERN	23465
2	NORTHEASTERN	42709
3	NORTHERN	31273
4	NORTHWESTERN	27895
5	SOUTHEASTERN	38023
6	SOUTHERN	31616
7	SOUTHWESTERN	25905
8	WESTERN	22039

According to the bar plot below, NORTHEASTERN has the highest crime events from 2012 to 2017, while WESTERN has the lowest crime events. Therefore, *WASTERN* will be considered as the safest area in Baltimore.



3.1.3 The most common crimes in Baltimore

Also, in order to have a general idea of the most common crimes in Baltimore, the bar plot groups by the different kinds of crime is shown in the following image. And larceny is the most frequent kind of crime in Baltimore.



3.2 Modeling

As a conclusion in 3.1.2, WESTERN is the safest area in Baltimore. Therefore, all the neighborhood in WESTERN will be listed in this section. We can find all the venues within a 500-meter radius of each neighborhood by connecting to the Foursquare API. This gives us a json file containing all the venues in each neighborhood which is converted to a pandas dataframe. This dataframe contains all the venues along with their coordinates and category in WESTERN Baltimore.

	CrimeDate	CrimeTime	Location	Description	Weapon	District	Neighborhood	Longitude	Latitude
22	09/02/2017	16:00:00	2000 N BENTALOU ST	AGG. ASSAULT	OTHER	WESTERN	Mondawmin	-76.65422	39.31091
37	09/02/2017	09:35:00	2500 PENNSYLVANIA AVE	LARCENY	NO WEAPON	WESTERN	Penn North	-76.64277	39.31033
40	09/02/2017	08:30:00	600 N CARROLLTON AVE	ROBBERY - STREET	FIREARM	WESTERN	Harlem Park	-76.63715	39.29567
70	09/01/2017	23:48:00	2000 W BALTIMORE ST	ROBBERY - STREET	NO WEAPON	WESTERN	Penrose/Fayette Street Ou	-76.64888	39.28810
78	09/01/2017	22:00:00	300 N FREMONT AVE	ARSON	NO WEAPON	WESTERN	Poppleton	-76.63103	39.29278

Also, the location information is obtained from google map and stored in the dataframe shown below.

	Neighborhood	Latitude	Longitude
0	Mondawmin	39.318	-76.659
1	Penn North	39.309	-76.644
2	Harlem Park	39.296	-76.642
3	Penrose/Fayette Street Ou	39.289	-76.654
4	Poppleton	39.295	-76.617
5	Bridgeview/Greenlawn	39.307	-76.655
6	Coppin Heights/Ash-Co-Eas	39.305	-76.637
7	Sandtown-Winchester	39.294	-76.653
8	Rosemont Homeowners/Tenan	39.313	-76.660
9	Panway/Braddish Avenue	39.293	-76.651

Now use the Foursquare API to obtain the venues of each neighborhood in WESTERN Baltimore and create a new dataframe called *bal__venues*

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Mondawmin	39.318	-76.659	Mamma Lucia	39.317113	-76.655321	Pizza Place
1	Mondawmin	39.318	-76.659	A&D Buffalo's	39.317732	-76.653998	Restaurant
2	Mondawmin	39.318	-76.659	Ross Dress for Less	39.316668	-76.658233	Clothing Store
3	Mondawmin	39.318	-76.659	Cinnabon	39.317104	-76.654277	Bakery
4	Mondawmin	39.318	-76.659	Family Dollar	39.317019	-76.658048	Discount Store

One hot encoding is done on the venues data. (One hot encoding is a process by which categorical variables are converted into a form that could be provided to ML algorithms to do a better job in prediction). The Venues data is then grouped by the Neighborhood and the mean of the venues are calculated, finally the 10 common venues are calculated for each of the neighborhoods.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Boyd-Booth	Grocery Store	Bar	Community Center	Market	Discount Store	Roller Rink	Sculpture Garden	Park	Flower Shop	Department Store
1	Bridgeview/Greenlawn	New American Restaurant	Convenience Store	Playground	Bus Stop	Breakfast Spot	Women's Store	College Bookstore	Community Center	Construction & Landscaping	Cruise
2	Coppin Heights/Ash-Co-Eas	Fried Chicken Joint	Wings Joint	Gym / Fitness Center	Market	Discount Store	Bakery	Roller Rink	Chinese Restaurant	Café	Deli / Bodega
3	Druid Heights	ATM	Café	Metro Station	Bookstore	Jazz Club	Basketball Court	Flower Shop	Business Service	Chinese Restaurant	Fried Chicken Joint
4	Easterwood	Chinese Restaurant	Pizza Place	Gym	Shopping Mall	History Museum	Discount Store	Market	Coffee Shop	Clothing Store	Plaza

Run *k*-means to cluster the neighborhood into 5 clusters. The reason to conduct a K- means clustering is to cluster neighborhoods with similar venues together so that people can shortlist the area of their interests based on the venues around each neighborhood.

4. Results

After running the K-means clustering we can access each cluster created to see which neighborhoods were assigned to each of the five clusters and the dataframe after cluster assigned is shown in below.

	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
0	Mondawmin	39.318	-76.659	3	Women's Store	Shopping Mall	Restaurant	Shoe Store	Clothing Store	Sandwich Place	Drugstore	Kids Store	Discount Store
1	Penn North	39.309	-76.644	0	Sandwich Place	Liquor Store	Health & Beauty Service	Pharmacy	Chinese Restaurant	Bookstore	Business Service	Deli / Bodega	Jazz Club
2	Harlem Park	39.296	-76.642	0	Playground	American Restaurant	Coffee Shop	Asian Restaurant	Pharmacy	Bus Stop	Bar	Women's Store	Discount Store
3	Penrose/Fayette Street Ou	39.289	-76.654	1	Intersection	Pizza Place	Pharmacy	Discount Store	Dive Bar	Convenience Store	Farmers Market	Restaurant	Train
4	Poppleton	39.295	-76.617	3	Restaurant	Non-Profit	Peruvian Restaurant	Lounge	Burger Joint	Italian Restaurant	Pizza Place	Deli / Bodega	Park

Now we can see the neighborhood which are assigned as the first cluster. In the first cluster, we can see that the most common venues are Café, Bar, and Bookstore.

	Latitude	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	39.309	Liquor Store	Health & Beauty Service	Pharmacy	Chinese Restaurant	Bookstore	Business Service	Deli / Bodega	Jazz Club	Department Store
2	39.296	American Restaurant	Coffee Shop	Asian Restaurant	Pharmacy	Bus Stop	Bar	Women's Store	Discount Store	Department Store
5	39.307	Convenience Store	Playground	Bus Stop	Breakfast Spot	Women's Store	College Bookstore	Community Center	Construction & Landscaping	Cruise
14	39.265	Food Truck	Shopping Mall	Buffet	Pier	Brewery	Intersection	Plaza	Donut Shop	Juice Bar
17	39.310	Bookstore	Park	Metro Station	Chinese Restaurant	Pharmacy	Café	Business Service	Plaza	Jazz Club
18	39.313	Café	Metro Station	Bookstore	Jazz Club	Basketball Court	Flower Shop	Business Service	Chinese Restaurant	Fried Chicken Joint
19	39.288	Intersection	Breakfast Spot	Discount Store	Spa	Boutique	Fried Chicken Joint	Pharmacy	Convenience Store	Cruise
24	39.288	Fried Chicken Joint	Fast Food Restaurant	Breakfast Spot	Restaurant	Dive Bar	Discount Store	Spa	Boutique	Convenience Store
25	39.289	Intersection	Museum	Boarding House	American Restaurant	Nightclub	Bus Stop	College Bookstore	Coffee Shop	Convenience Store

For the second cluster shown below, the most common venues are Restaurant and Bar.

	Latitude	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
3	39.289	Pizza Place	Pharmacy	Discount Store	Dive Bar	Convenience Store	Farmers Market	Restaurant	Train	Train Station
7	39.294	Fast Food Restaurant	Southern / Soul Food Restaurant	Dive Bar	Pizza Place	Farmers Market	Bar	Liquor Store	Fried Chicken Joint	Train
9	39.293	Fast Food Restaurant	Pharmacy	Dive Bar	Intersection	Farmers Market	Bar	Southern / Soul Food Restaurant	Liquor Store	Train
16	39.297	Pharmacy	Liquor Store	Furniture / Home Store	Fried Chicken Joint	Convenience Store	Bar	Women's Store	Discount Store	Department Store

The third cluster only has one neighborhood, which means it is a unique neighborhood and can not clustered with other neighborhoods.

	Latitude	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
11	39.308	Bus Stop	Fried Chicken Joint	Deli / Bodega	Women's Store	Drugstore	Community Center	Convenience Store	Cruise	Cycle Studio

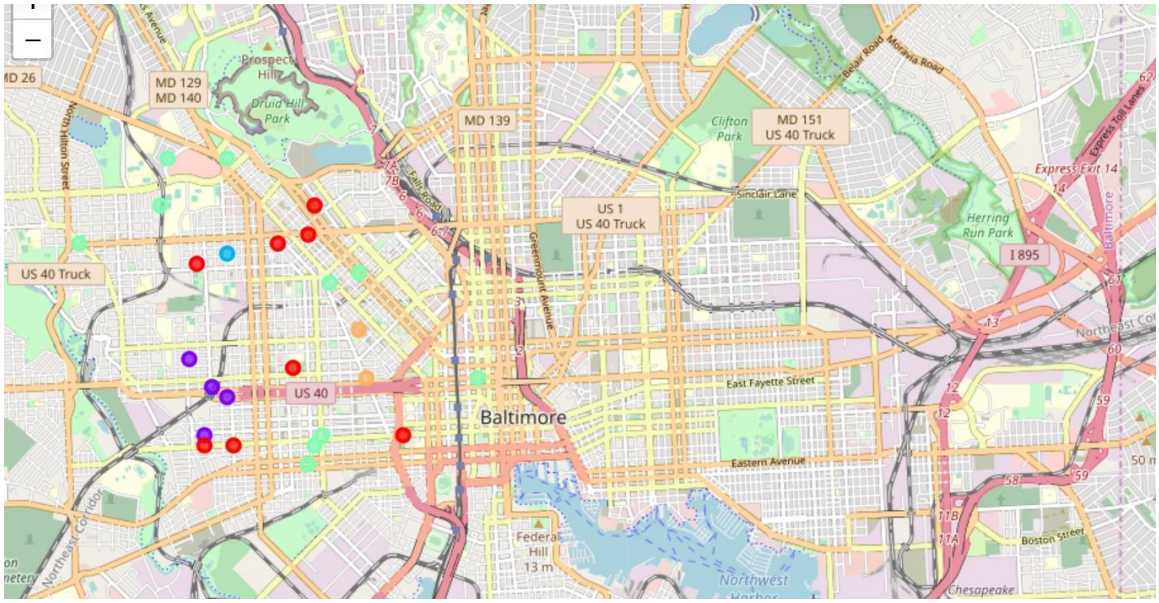
The fourth cluster shown below mainly contains all kinds of stores: shopping mall, kids store, drug store, mobile phone shop and so on.

	Latitude	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	39.318	Shopping Mall	Restaurant	Shoe Store	Clothing Store	Sandwich Place	Drugstore	Kids Store	Discount Store	Fast Food Restaurant
4	39.295	Non-Profit	Peruvian Restaurant	Lounge	Burger Joint	Italian Restaurant	Pizza Place	Deli / Bodega	Park	Sandwich Place
6	39.305	Wings Joint	Gym / Fitness Center	Market	Discount Store	Bakery	Roller Rink	Chinese Restaurant	Café	Deli / Bodega
8	39.313	Locksmith	Track Stadium	Fast Food Restaurant	Discount Store	Comic Shop	Community Center	Construction & Landscaping	Convenience Store	Cruise
10	39.318	Restaurant	Shoe Store	Mobile Phone Shop	Women's Store	Pizza Place	Bus Station	Bus Stop	Nail Salon	Convenience Store
12	39.288	Pizza Place	Gym	Shopping Mall	History Museum	Discount Store	Market	Coffee Shop	Clothing Store	Plaza
13	39.286	Gym	Shoe Store	Clothing Store	Chinese Restaurant	Market	Peruvian Restaurant	Liquor Store	Plaza	Rental Car Location
15	39.309	Drugstore	Gas Station	Grocery Store	Fried Chicken Joint	Bakery	Bank	Pizza Place	Fast Food Restaurant	Women's Store
21	39.289	Pizza Place	Peruvian Restaurant	Clothing Store	Chinese Restaurant	Market	Plaza	Discount Store	Shoe Store	Shopping Mall

The fifth cluster is also a small cluster with only two neighborhoods.

	Latitude	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
20	39.300	Bar	Community Center	Market	Discount Store	Roller Rink	Sculpture Garden	Park	Flower Shop	Department Store
22	39.295	Pharmacy	History Museum	Park	Donut Shop	Comic Shop	Community Center	Construction & Landscaping	Convenience Store	Cruise

The clustered neighborhoods are visualized by the folium library. According to the characterization of different clusters, neighborhoods with color green (Cluster 1), red (Cluster 4), and purple (Cluster 2) will be the most recommended neighborhood since most of the shops and restaurants will be in these areas.



5. Discussion

In this project, the crime data in Baltimore from 2012 to 2017 is analyzed and plotted. As mentioned in the introduction, Baltimore is very infamous of its crime rate and the visualized data proves that crime may happen in each corner of Baltimore. Most of the tourists will visit the Harbor area in Baltimore, however, the Harbor area has the highest crime density. Therefore, for tourist, Harbor area will not be recommended for the staying place in Baltimore. And Western is the safest area in Baltimore and finding a hotel in this area is the best choice for tourists according to the safety concern.

In addition, the attractions in Western Baltimore is also investigated. The neighborhoods in Western are clustered into five different groups and three of them are areas with plenty choices of restaurants and stores. For convince consideration, tourists are recommended to stay in those areas.

6. Conclusion

In conclusion, this project provides a better idea for tourists who want to visit Baltimore and may stay there for several time of the safest district in Baltimore and the most convenient areas in that district. However, due to the lack of ability of the author, only

safety and attraction are considered as factors. In the future study, some other factors such as transportation and hotel cost can also be considered when giving recommendations.