Analyzing Chicago attractions along with crime data

Yasmeen Abu-Kheil

October 1, 2019

1. Introduction:

Tourists find it difficult to select the optimal locations for their Hotels. They would like to select locations that are nearby attraction areas, safe to stay and walk in as well as near to famous restaurants and transportations. In this project, we will analyze and cluster Chicago city based on tourism attractions, crime rates, public transportations and famous restaurants. Then, we will provide recommendations on the top best hotel locations that optimizes tourists' needs and requirements.

2. Data acquisition and cleaning

2.1 Data sources

In this section, the data used to solve the above problem are described as follow:

- Chicago Crime Data to determine the crime rate at different Chicago areas
- Top attractions and restaurants acquired from FourSqaure website
- Within the top attractions, find the restaurants, transportations and hotels and their geographical locations

2.1.1 Crime Data

This dataset reflects reported incidents of crime (with the exception of murders where data exists for each victim) that occurred in the City of Chicago from 2001 to present, minus the most recent seven days

(Source: https://ibm.box.com/shared/static/svflyugsr9zbqy5bmowgswqemfpm1x7f.csv)

2.1.2 FourSquare Data

We will use the FourSquare website to request the top 50 sites in Chicago

2.1.3 Restaurants, transportations and hotels

After getting the top 50 cites in Chicago, we will find category restaurants, transportations and hotels

2.2 Data content and cleaning

The content of the crime data is described below:

```
['ID',
 'CASE_NUMBER',
 'DATE',
```

```
'BLOCK',
'IUCR',
'PRIMARY_TYPE',
'DESCRIPTION',
'LOCATION_DESCRIPTION',
'ARREST',
'DOMESTIC',
'BEAT',
'DISTRICT',
'WARD',
'COMMUNITY_AREA_NUMBER',
'FBICODE',
'X_COORDINATE',
'Y_COORDINATE',
'YEAR',
'UPDATEDON',
'LATITUDE',
'LONGITUDE',
'LOCATION']
```

The following information were extracted from the crime data for the years from 2009:

Crime_Columns_Interest = ['DATE', 'BLOCK','PRIMARY_TYPE', 'DISTRICT','WARD', 'COMMUNITY_AREA_NUMBER', 'YEAR','LATITUDE', 'LONGITUDE'];

	DATE	BLOCK	PRIMARY_TYPE	DISTRICT	WARD	COMMUNITY_AREA_NUMBER	YEAR	LATITUDE	LONGITUDE
0	04/04/2011 05:45:00 AM	043XX S WABASH AVE	THEFT	2	3.0	38.0	2011	41.815933	-87.624642
1	12/30/2010 04:30:00 PM	083XX S KINGSTON AVE	THEFT	4	7.0	46.0	2010	41.743665	-87.562463
2	02/02/2016 07:30:00 PM	033XX W 66TH ST	THEFT	8	15.0	66.0	2016	41.773455	-87.706480
3	09/29/2010 07:59:00 AM	006XX W CHICAGO AVE	THEFT	12	27.0	24.0	2010	41.896447	-87.644939
4	11/30/2016 01:15:00 AM	050XX N KEDZIE AVE	THEFT	17	33.0	14.0	2016	41.972845	-87.708600

The foursquare website were used to obtain the top 50 attractions, restaurants, transportations and hotels within 500 radius of Chicago city. Samples of obtained data are shown below:

	index	name	categories	lat	lng
0	0	The Blackstone Hotel	Hotel	41.873343	-87.624765
1	1	Hilton Chicago	Hotel	41.872591	-87.624726
2	3	Hilton Chicago Continental Ballroom	Hotel	41.871778	-87.625333
3	4	Hotel Blake	Hotel	41.875164	-87.629485
4	6	The Congress Plaza Hotel	Hotel	41.875168	-87.624715
5	8	Travelodge	Hotel	41.874626	-87.625645
6	9	H.Chicago	Hotel	41.876235	-87.623808
7	10	CARA Compound @ Congress Hotel	Hotel	41.875478	-87.624290

	index	name	categories	lat	lng
0	0	Metra - Van Buren Street	Train Station	41.876947	-87.623031
1	1	CTA Bus Stop 4725	Bus Station	41.875855	-87.623837
2	2	CTA Bus Stop 17255	Bus Station	41.875784	-87.626256
3	3	NICTD's South Shore Line at Van Buren and Jackson	Train Station	41.876605	-87.622964
4	4	CTA Bus Stop 73	Bus Station	41.876320	-87.624346
5	5	CTA Bus Stop 1586	Bus Station	41.874750	-87.624400
6	6	Southshore Train	Train Station	41.877017	-87.623100
7	7	CTA Bus Stop 74	Bus Station	41.874716	-87.623769
8	8	CTA Tower 12	Train Station	41.877195	-87.626261
9	9	CTA Bus Stop 1584	Bus Station	41.873364	-87.623928

	name	categories	lat	lng
0	Auditorium Theatre	Theater	41.876058	-87.625303
1	SummerDance	Dance Studio	41.873804	-87.623335
2	Auditorium Building	Concert Hall	41.875858	-87.624565
3	Buddy Guy's Legends	Music Venue	41.873034	-87.626135
4	Chicago Architecture Foundation	Museum	41.878556	-87.624550
5	Museum of Contemporary Photography	Art Museum	41.874382	-87.624524
6	Symphony Center (Chicago Symphony Orchestra)	Concert Hall	41.879275	-87.624680
7	The Art Institute of Chicago	Art Museum	41.879609	-87.623572
8	Thorne Miniature Rooms	Museum	41.879532	-87.623680

	name	categories	lat	lng
0	Cafecito	Cuban Restaurant	41.875724	-87.626386
1	Dunkin'	Donut Shop	41.876768	-87.624575
2	Fontano's Pizza and Subs	Pizza Place	41.877157	-87.624348
3	Himalayan Restaurant	Indian Restaurant	41.874129	-87.626064
4	Mercat a la Planxa	Tapas Restaurant	41.873212	-87.624481
5	Harold's Chicken Shack	Fried Chicken Joint	41.873954	-87.626081
6	Blaze Pizza	Pizza Place	41.875233	-87.627468
7	Epic Burger	Burger Joint	41.875002	-87.627525
8	Just Salad	Salad Place	41.878226	-87.626779

3. Methodology

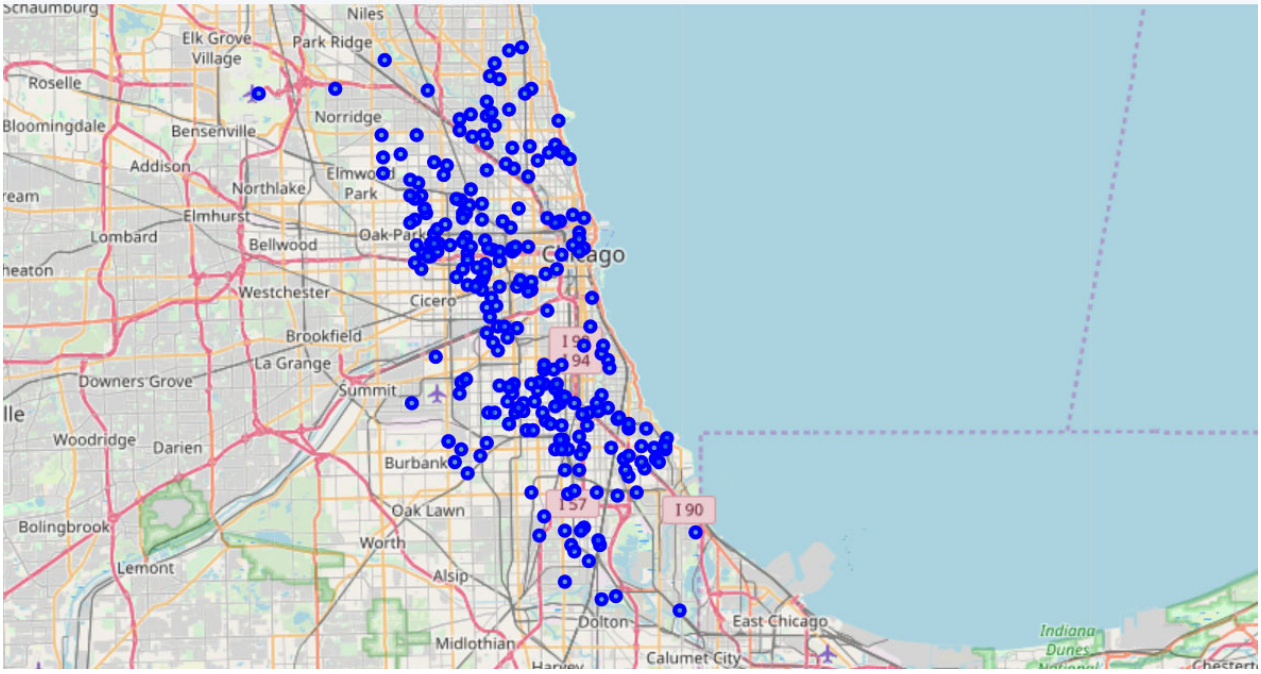
3.1 Data Visualizations

- Crime Data

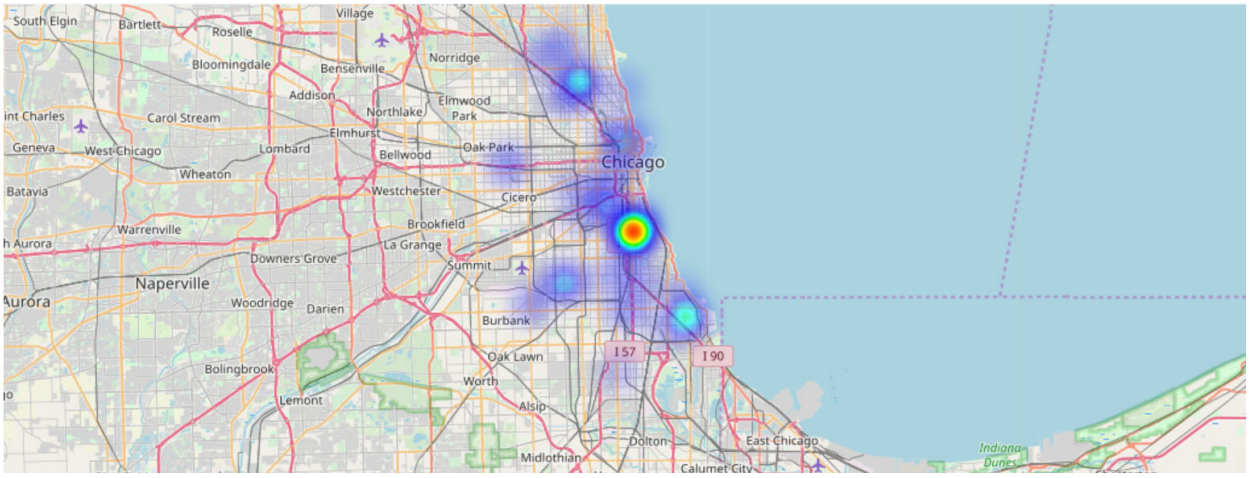
There are 29 unique crime category from 2009 and these category were ordered from the most common occurring category to the least occurring ones. Samples of the categories are shown below:

	PRIMARY_TYPE	Count
27	THEFT	61
2	BATTERY	38
5	CRIMINAL DAMAGE	29
13	NARCOTICS	21
3	BURGLARY	18
20	OTHER OFFENSE	18
1	ASSAULT	15
24	ROBBERY	9
12	MOTOR VEHICLE THEFT	8

The location map that presents the crimes locations from 2009 till now in shown below:

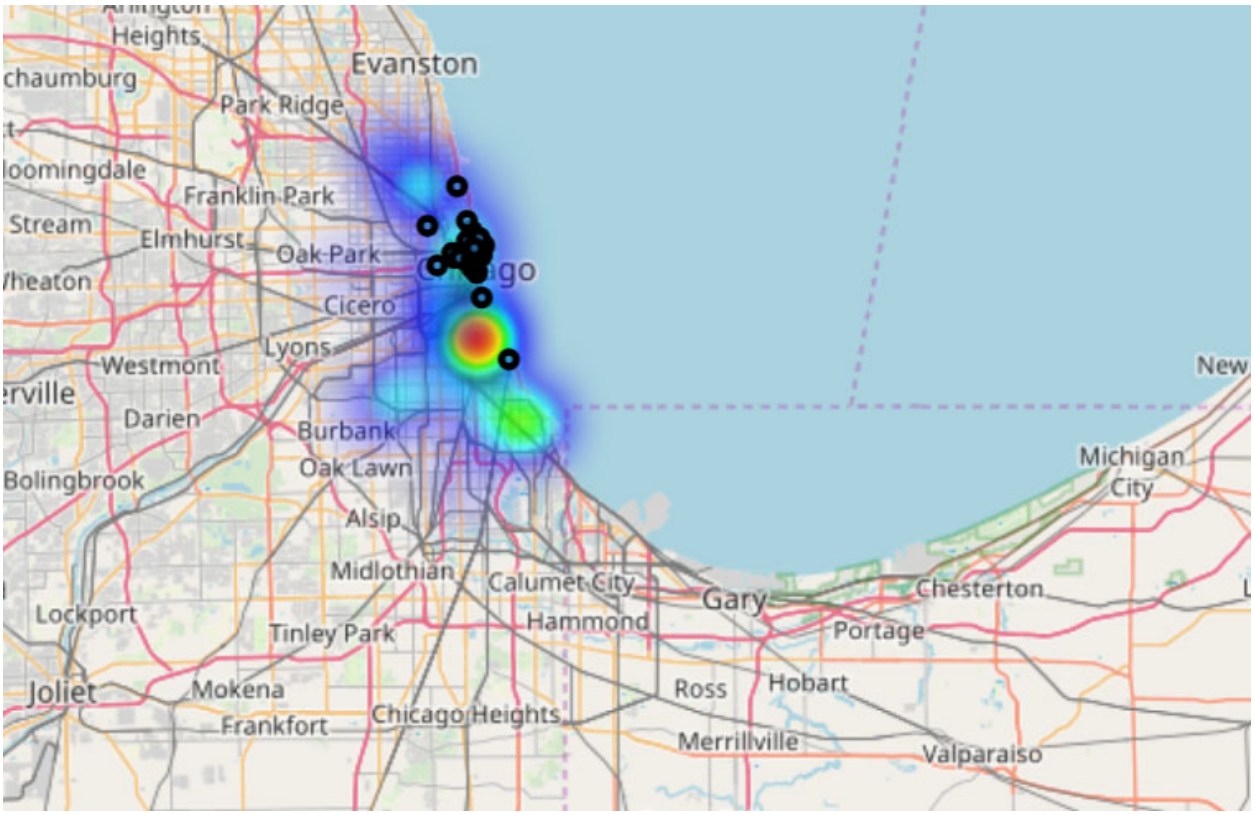


The heat map for the top three crime categories were generated:

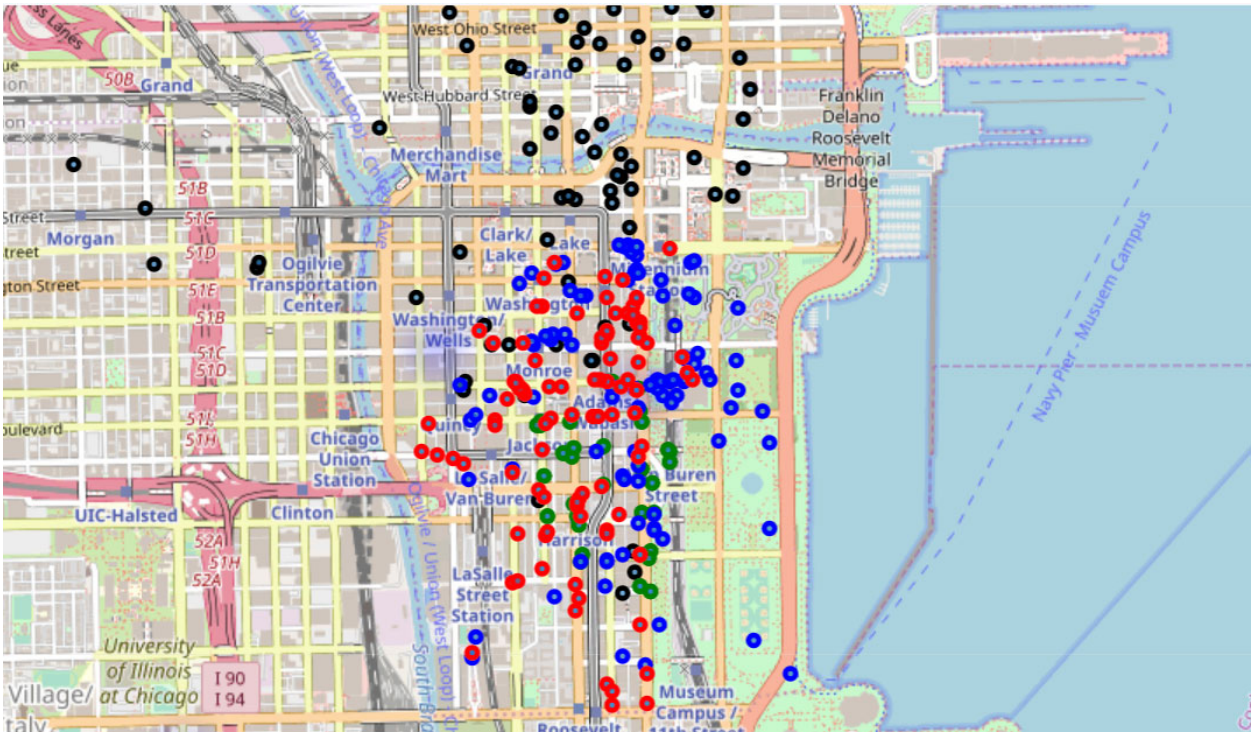


- Location Data

On top of the above heat map, the extracted hotel locations were added and combined with the crime occurrence data:

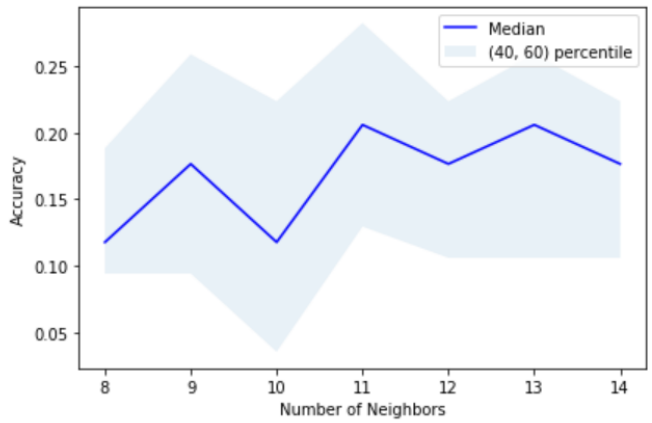


On top of the above heat map, the extracted attractions, food and transportation locations were added and combined with the crime occurrence data:

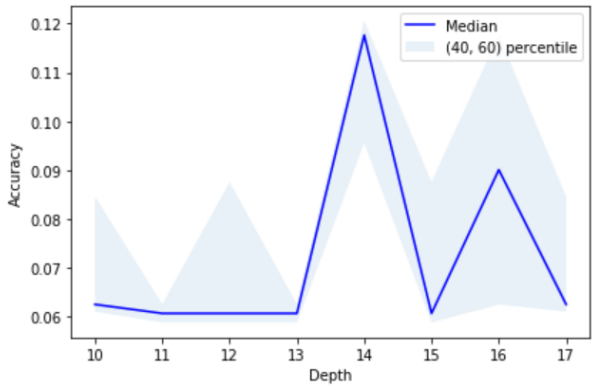


3.2 Data Analysis

- **K Nearest Neighbor (KNN)**



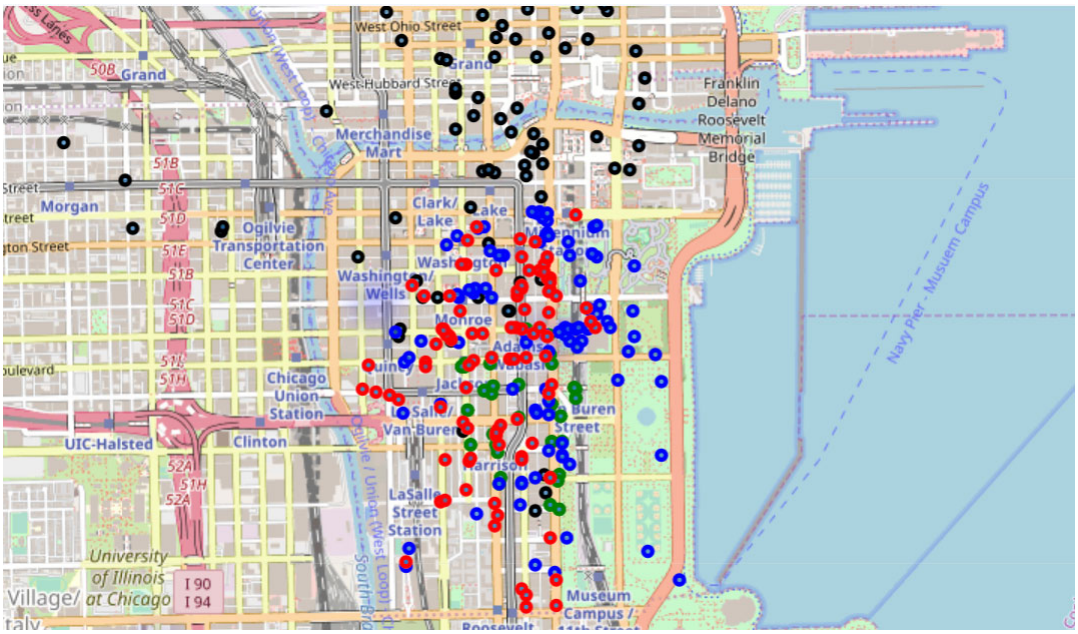
- **Decision Tree**

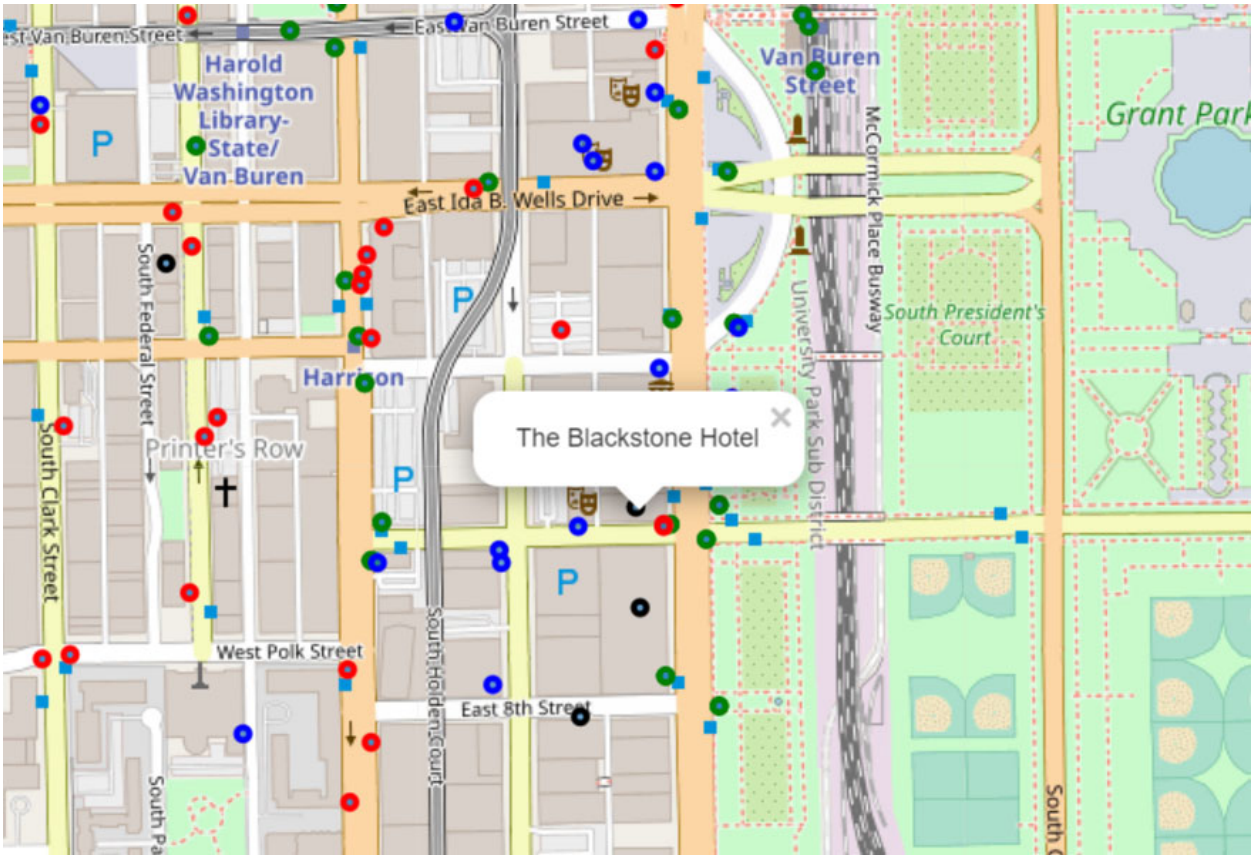


Algorithm	Jaccard	LogLoss
KNN	0.401198	1.240934e+01
Tree	1.000000	9.992007e-16

4. Results & Discussion

Based on the below figure, tourism attractions, crime rates, public transportations and famous restaurants were visualized in the map. The best hotel that optimizes tourists' needs and requirements is the black stone hotel.





5. Conclusions

The use of data science can help tourists to find and select the optimal locations for their Hotels. They would like to select locations that are nearby attraction areas, safe to stay and walk in as well as near to famous restaurants and transportations. In this project, we analyzed and clustered Chicago city based on tourism attractions, crime rates, public transportations and famous restaurants. We also provided recommendations on the top best hotel locations that optimizes tourists' needs and requirements.