Hotel Development in Philadelphia

Onur Yavuz

1.1 Introduction:

Philadelphia is one of the major cities of the United States situated in the East Coast. With its rich history, cultural attractions, and business opportunities it attracts millions of visitors annually. According to the Philadelphia Convention and Visitors' Bureau Website, (https://www.discoverphl.com/meet/choose-philadelphia/facts-and-figures/) , there are currently 39 hotel brands in the city center. For hotel investors, it is critical to understand consumer behavior and preferences of hotel customers for a successful venture. Demographic and socio-economic characteristics of travelers are the key determinants of hotel locations. Safety and proximity to leisure activities, including cultural and historical sites, dining and entertainment centers, are two important parameters when hotel customers decide where to stay during their travel and pricing. In this project, I analyze districts of Philadelphia in terms of venues and criminal data. The analysis would be of interest to hotel investors, investors in other branches of hospitality business and real estate.

1.2. Description of Data and Resources:

I analyze the city through Police districts. There are 22 districts whose ids are between 1 and 77. Below are the data-sets I used:

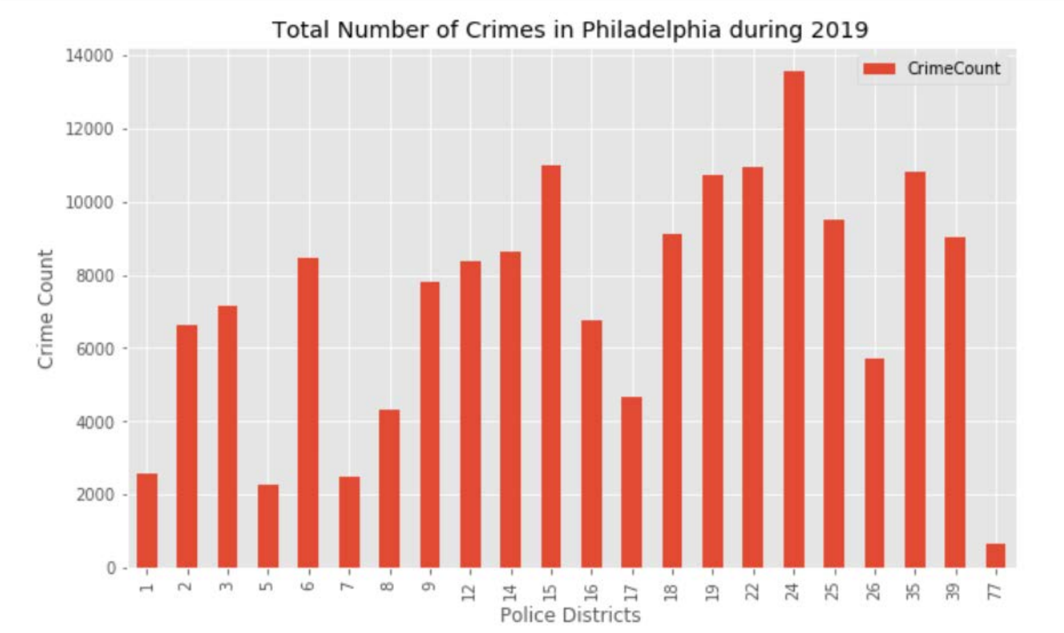
- Crime Data: Crime incidents data-set for 2019 from the website https://www.opendataphilly.org/ and the link for the data set is https://www.opendataphilly.org/dataset/crime-incidents
- Geo Data: Geojson file of police districts from the website https://www.opendataphilly.org/ and the link for the data set is http://data.phl.opendata.arcgis.com/datasets/62ec63afb8824a15953399b1fa819df2_0.geojson
- Venue Data: Common venues of Philadelphia from Foursquare API https://foursquare.com/

2.1. Methodology

Crime data contained information about the nature of the crime and specific addresses crimes took place. The latitude and longitude of the addresses were also provided. Here is a snapshot of the original data, where dc_district is the district id:

	objectid	dc_dist	psa	dispatch_date_time	dispatch_date	dispatch_time	hour_	dc_key	location_block	ucr_general	text_general_code	point_x	point_y	lat	lng
0	120	9	1	2019-11-12 09:59:00	2019-11-12	09:59:00	9.0	201909047628	100 BLOCK S BROAD ST	600	Thefts	-75.066628	40.049264	40.049264	-75.066628
1	128	9	2	2019-01-19 17:20:00	2019-01-19	17:20:00	17.0	201909002506	1400 BLOCK SPRING GARDEN ST	600	Thefts	-75.161446	39.962334	39.962334	-75.161446
2	53	77	A	2019-12-14 08:07:00	2019-12-14	08:07:00	8.0	201977007047	0 BLOCK PIA WAY	600	Thefts	-75.230706	39.883881	39.883881	-75.230706
3	54	77	A	2019-12-24 14:03:00	2019-12-24	14:03:00	14.0	201977007242	0 BLOCK PIA WAY	600	Thefts	-75.230706	39.883881	39.883881	-75.230706
4	55	77	A	2019-12-25 16:29:00	2019-12-25	16:29:00	16.0	201977007257	0 BLOCK PIA WAY	600	Thefts	-75.230706	39.883881	39.883881	-75.230706

As is seen from the table, there are multiple latitude/longitude information per district. To get estimates for the coordinates of the district, I found the means of the latitude/longitude per district. However, I used Google map to test which addresses these mean latitude/longitude values yielded and the resulting addresses were not meaningful. I have decided to use the crime data only for crime count and formed a database involving district number and crime count. Then I created a bar plot to view districts and the total crimes in 2019.



The geo data provided district id's, district addresses, and a list of longitude/latitude values for some addresses included in the district. Below is a snapshot.

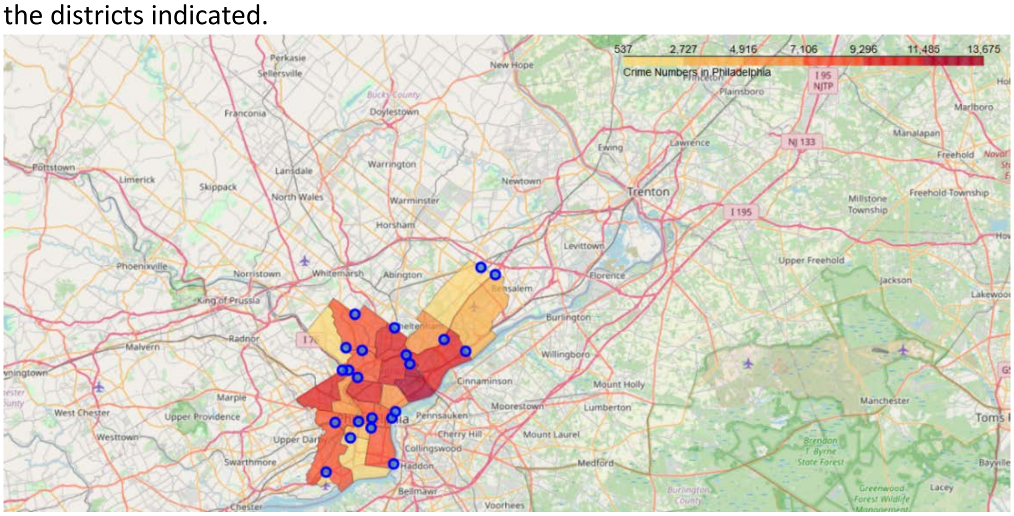

```
{'type': 'Feature',
 'properties': {'OBJECTID': 321,
  'AREA': None,
  'PERIMETER': 81903.64182498,
  'DISTRICT_': 1,
  'DISTRICT_ID': None,
  'DIST_NUM': 1,
  'SUM_AREA': None,
  'DIST_NUMC': '01',
  'LOCATION': '24th St. & Wolf St.',
  'PHONE': '686-3010',
  'DIV_CODE': 'SPD',
  'AREA_SQMI': 216350124.152872},
 'geometry': {'type': 'Polygon',
  'coordinates': [[[-75.1972400674602, 39.9294369069777],
    [-75.1969266694466, 39.9291830093794],
    [-75.1966551958808, 39.9293313020346],
    [-75.1962094965133, 39.9295635898883],
    [-75.1961188904874, 39.9296108123323],
```


As the addresses of the districts were not given in the standard form but rather in terms of intersections, geolocator did not return the longitude/latitude. Instead of using the addresses, I selected the first ordered pair in the coordinate set. I used Google Map to check some random entries and the addresses they yielded this time were within a few kilometers from the district address and I have decided to use the first ordered pair for each district.

I created a data frame using the geo data and merged it with the crime data. Here is a snapshot of the first five rows of the merged data frame (which I will refer as merged data from now on.)

	district	CrimeCount	Latitude	Longitude
0	1	2560	39.929437	-75.197240
1	2	6611	40.044541	-75.054444
2	3	7160	39.899322	-75.132053
3	5	2260	40.034678	-75.204300
4	6	8482	39.952940	-75.134366

Using the merged data, I created a choropleth map to illustrate the crime counts per district with the districts indicated.

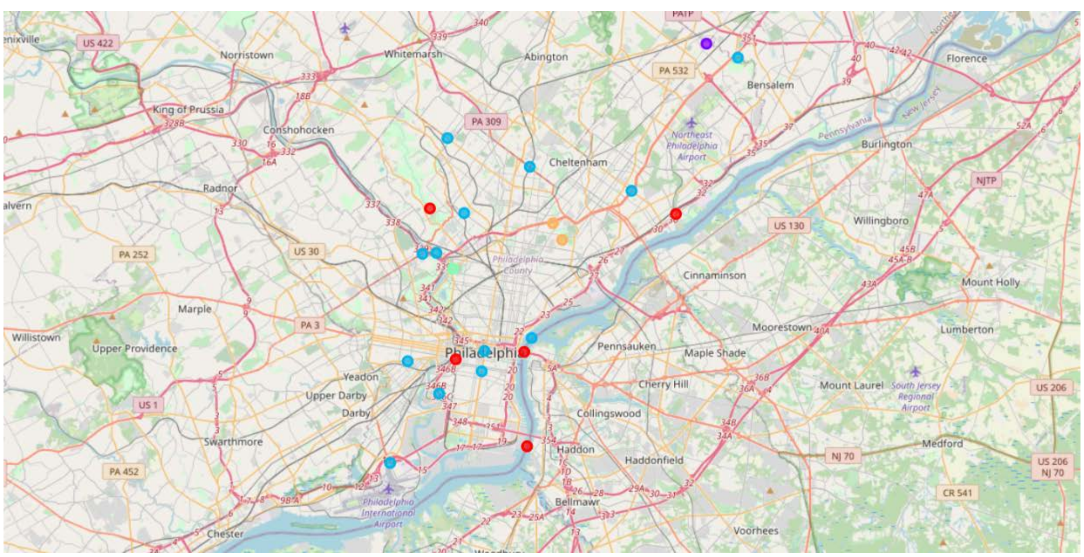


I obtained venue data from foursquare.com per district. Below are a snapshot of the data frame of the original venue data and a snapshot of the data frame which gives top ten venues per district.

	district	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	1	39.929437	-75.19724	Dew Drop Inn	39.927319	-75.194085	Bar
1	1	39.929437	-75.19724	GhettoHome	39.927620	-75.193275	Hookah Bar
2	1	39.929437	-75.19724	28th Street Supermarket	39.927152	-75.192743	Grocery Store
3	1	39.929437	-75.19724	Cow Chip Bingo	39.933115	-75.199841	Baseball Field
4	1	39.929437	-75.19724	Los Angeles	39.925805	-75.193989	Arcade

	district	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	1	Baseball Field	Arcade	Hookah Bar	Grocery Store	Bar	Ethiopian Restaurant	Flea Market	Fish Market	Film Studio	Fast Food Restaurant
1	2	Men's Store	Shopping Mall	Clothing Store	Mobile Phone Shop	Optical Shop	Shoe Store	Sandwich Place	Pharmacy	Deli / Bodega	Paper / Office Supplies Store
2	3	Park	Historic Site	Harbor / Marina	Donut Shop	Fish Market	Film Studio	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Ethiopian Restaurant
3	5	Park	Scenic Lookout	Beach	Bridge	Dive Bar	Trail	Eastern European Restaurant	Baseball Field	Fish Market	Film Studio
4	6	Japanese Restaurant	Bar	Park	Boat or Ferry	Trail	Other Great Outdoors	Beer Garden	Falafel Restaurant	Flea Market	Fish Market

There are 154 unique venue categories. I grouped the venue the data frame by "district" and formed a new data frame. Then I used **K-means algorithm** to cluster the districts with k=5 and formed the corresponding data frame. I marked the clusters on the Philadelphia map:



3.1. Results

I merged the data frame containing cluster labels with the data frame with the merged data (will refer as cluster data.)

	district	CrimeCount	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	10th Most Common Venue
0	1	2560	39.929437	-75.197240	2	Baseball Field	Arcade	Hookah Bar	Grocery Store	Bar	Ethiopian Restaurant	Flea Market	Fish Market	Film Studio	Fast Food Restaurant
1	2	6611	40.044541	-75.054444	2	Men's Store	Shopping Mall	Clothing Store	Mobile Phone Shop	Optical Shop	Shoe Store	Sandwich Place	Pharmacy	Deli / Bodega	Paper / Office Supplies Store
2	3	7160	39.899322	-75.132053	0	Park	Historic Site	Harbor / Marina	Donut Shop	Fish Market	Film Studio	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Ethiopian Restaurant
3	5	2260	40.034678	-75.204300	0	Park	Scenic Lookout	Beach	Bridge	Dive Bar	Trail	Eastern European Restaurant	Baseball Field	Fish Market	Film Studio
4	6	8482	39.952940	-75.134366	0	Japanese Restaurant	Bar	Park	Boat or Ferry	Trail	Other Great Outdoors	Beer Garden	Falafel Restaurant	Flea Market	Fish Market

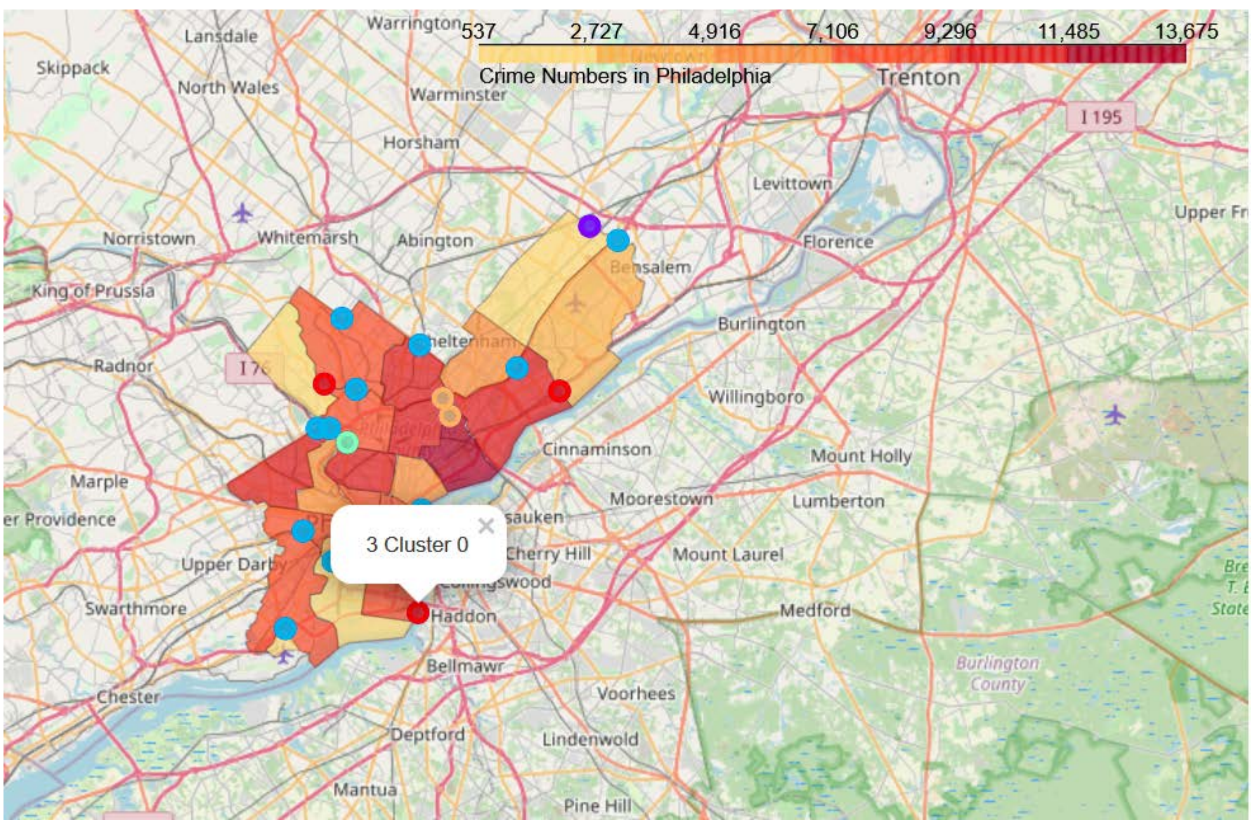
I examined each cluster separately and clusters 0 and 2 contained the majority of districts. Below are the corresponding data frames respectively:

district	Cluster Labels	CrimeCount	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
3	0	7160	Park	Historic Site	Harbor / Marina	Donut Shop	Fish Market	Film Studio	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Ethiopian Restaurant
5	0	2260	Park	Scenic Lookout	Beach	Bridge	Dive Bar	Trail	Eastern European Restaurant	Baseball Field	Fish Market	Film Studio
6	0	8482	Japanese Restaurant	Bar	Park	Boat or Ferry	Trail	Other Great Outdoors	Beer Garden	Falafel Restaurant	Flea Market	Fish Market
15	0	10981	Deli / Bodega	Music Venue	Park	Train Station	Train	Burger Joint	Yoga Studio	Ethiopian Restaurant	Film Studio	Fast Food Restaurant
18	0	9107	Food Truck	Park	Café	Trail	Playground	Train Station	Coffee Shop	Farmers Market	Liquor Store	Music Venue

district	Cluster Labels	CrimeCount	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
1	2	2560	Baseball Field	Arcade	Hookah Bar	Grocery Store	Bar	Ethiopian Restaurant	Flea Market	Fish Market	Film Studio	Fast Food Restaurant
2	2	6611	Men's Store	Shopping Mall	Clothing Store	Mobile Phone Shop	Optical Shop	Shoe Store	Sandwich Place	Pharmacy	Deli / Bodega	Paper / Office Supplies Store
8	2	4312	Convenience Store	Motel	Fish Market	Hotel	Rental Car Location	Italian Restaurant	Fast Food Restaurant	Gym Pool	Donut Shop	Border Crossing
9	2	7799	Hotel	Coffee Shop	Convenience Store	Sandwich Place	American Restaurant	Burger Joint	Public Art	Salad Place	Steakhouse	Donut Shop
12	2	8369	Diner	Deli / Bodega	Arts & Crafts Store	Indian Restaurant	Flea Market	Park	Breakfast Spot	Brewery	Mexican Restaurant	Middle Eastern Restaurant
14	2	8626	Bank	Shopping Plaza	Chinese Restaurant	Supermarket	Sandwich Place	Residential Building (Apartment / Condo)	Shopping Mall	Optical Shop	Gym	Home Service
16	2	6765	Pizza Place	Deli / Bodega	Trail	New American Restaurant	Food Truck	Fried Chicken Joint	Chinese Restaurant	Tennis Stadium	Donut Shop	Coffee Shop
17	2	4650	Pizza Place	Bakery	Mexican Restaurant	Café	Beer Garden	Gastropub	Italian Restaurant	Coffee Shop	Southern / Soul Food Restaurant	Sushi Restaurant
19	2	10745	Pizza Place	American Restaurant	Sandwich Place	Residential Building (Apartment / Condo)	Mexican Restaurant	Pharmacy	Steakhouse	Big Box Store	Sushi Restaurant	Mobile Phone Shop
26	2	5699	Pool	Beach	American Restaurant	Steakhouse	Bar	Beer Garden	Flea Market	Fish Market	Film Studio	Fast Food Restaurant
35	2	10797	Korean Restaurant	Diner	Pharmacy	Smoke Shop	Japanese Restaurant	Hardware Store	Bakery	Yoga Studio	Ethiopian Restaurant	Fish Market
39	2	9013	Pizza Place	Discount Store	Fried Chicken Joint	Deli / Bodega	Salon / Barbershop	Coffee Shop	Brewery	Shopping Mall	Film Studio	Breakfast Spot
77	2	666	Hotel	Hotel Bar	American Restaurant	Art Gallery	Boat or Ferry	Restaurant	Coffee Shop	Sushi Restaurant	Ethiopian Restaurant	Fish Market

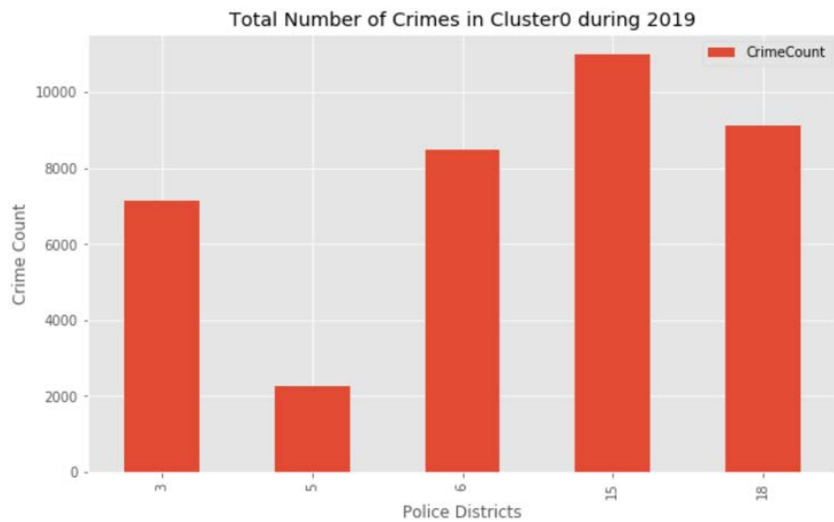
Top venues in Cluster 0 seem to be more recreation venues whereas those in Cluster 2 seem to be more dining and entertainment venues.

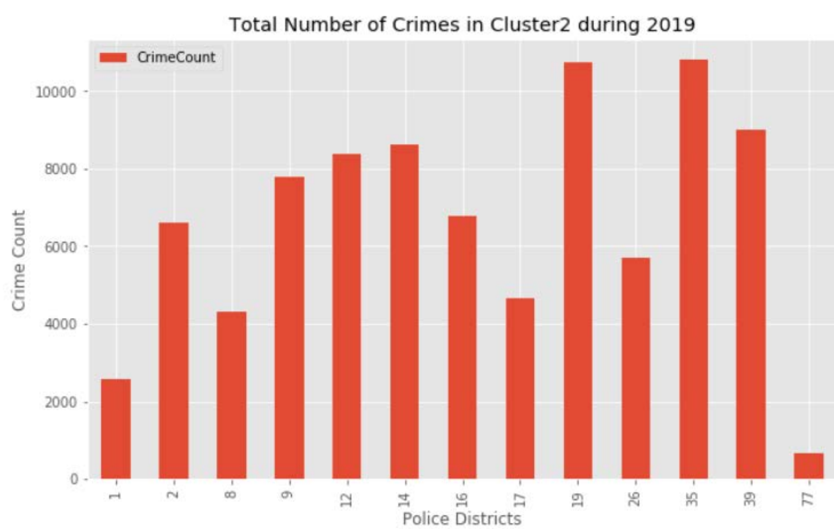
I added the choropleth map to the cluster map to obtain the following map so that one can compare the districts within the same cluster in terms of safety.



Blue dots in the above map represent cluster 2 and red ones represent cluster 0. It is also clear from the map that districts are mostly clustered at clusters 2 and 0.

One can also compare the safety of districts given cluster using bar charts. Below are the bar charts of crime data for clusters 0 and 2





4.1. Discussion

I selected longitude/latitude of districts from a list of longitudes/latitudes within the boundaries of the districts. A simple arithmetic mean of the longitudes/latitudes was not meaningful. A calculation of the geometric center of the given coordinates would make the analysis neater.

If I had data which connected police districts with neighborhoods and so was able to refer to regions with names would have made the analysis more readable.

K-means algorithm with different k values might give more accurate information.

Returned categories for venues are too specific. For example, Asian restaurant and Mexican restaurant are considered two different categories. For the sake of this project it might be better to combine them under the “restaurant” category.

I worked with raw crime numbers. If I had the population data for each district, comparing crime numbers to population in each district would give a better safety index.

5.1. Conclusion

Assuming all other factors are similar, an investor who would like to invest in a hotel development in Philadelphia can use the final map or bar charts given in the “Results “section to make a decision for location as they give safety information for districts with similar properties in terms of common venues. Similarly, a traveler who is searching for a hotel to stay, or someone who would like to buy or rent a house would benefit from the map and the charts.