OPENING A NEW RESTAURANT IN LOS ANGELES CITY

Aug-17 2019

Outlines

- ✓ Business Problem
- ✓ Introduction
- ✓ Data and Sources
- ✓ Methodology
- ✓ Results
- ✓ Discussion
- ✓ Conclusion

Business Problem

- As increasing numbers of consumers want to dine out, the number of restaurants has skyrocketed in Loss Angeles.
- In this regard, this report is written to propose neighborhood(s) suitable to open a new restaurant based on available data of Los Angeles.
- Unsupervised machine learning will be used to achieve the target. In this report, we will present the background of the work and also the source of data which will be used.
- This report is useful for investors in Los Angeles that are interested to open a restaurant in one of neighborhoods of Los Angles. Los Angeles has many neighborhoods and each region offers a variety of dining options. So, it is useful to use available data to inspect all areas before.

Introduction

- Many people in Los Angeles eating out at least once a week and the restaurant industry continues to thrive.
- **The Importance of Location**: A restaurant's location influences many aspects of your operation, including the menu! In addition to being one of the most significant determinants of your financial viability, decisions regarding location are not easy to undo.

Important Factors

To choose a location for the new restaurant, the following aspects are important:

- **The demographics**: Ensure the target market of your restaurant matches the demographics of the area. Factors like the income in a certain radius definitely matter.
- **Crime Rates**: High crime rates can make potential customers uncomfortable, and if they feel they'll be mugged walking to their cars, it will drive away business.
- **Availability of services, shopping center, etc.**: One of the important factors is the availability of, world-class shopping, recreation center, good urban parks, near a body of water, Good traffic, etc.
- **Available restaurant and food services**: When there are many restaurants available in a region of Los Angeles It says that the region was good for open a restaurant. But, can we extend it to the future? So, having a record of available services is useful.

Data sources

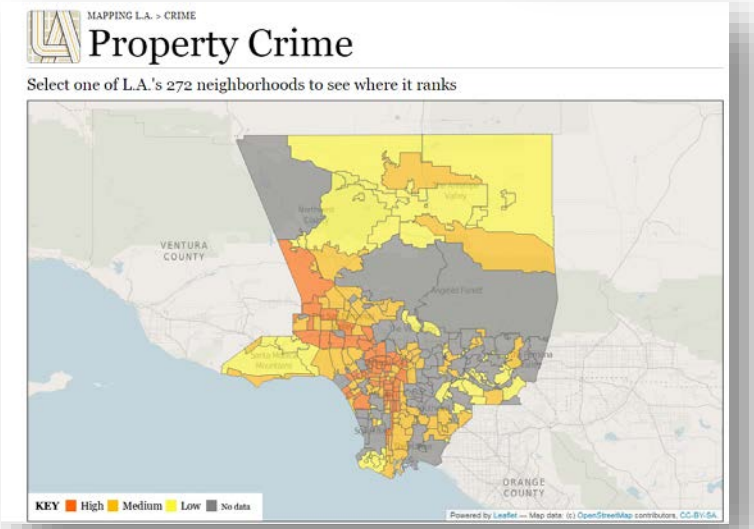
The data that we need to complete this project:

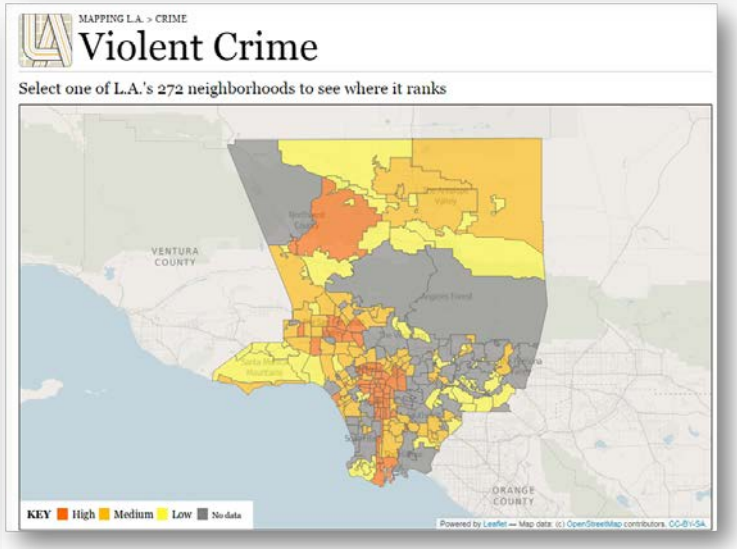
- Name and Coordinates of Los Angeles neighborhoods,
- Income rate of residents in each neighborhood,
- The crime rate in each area of Loss Angeles,
- No. and type of amenities like shopping centers, recreation centers, etc.
- Available restaurant and service foods in each region.

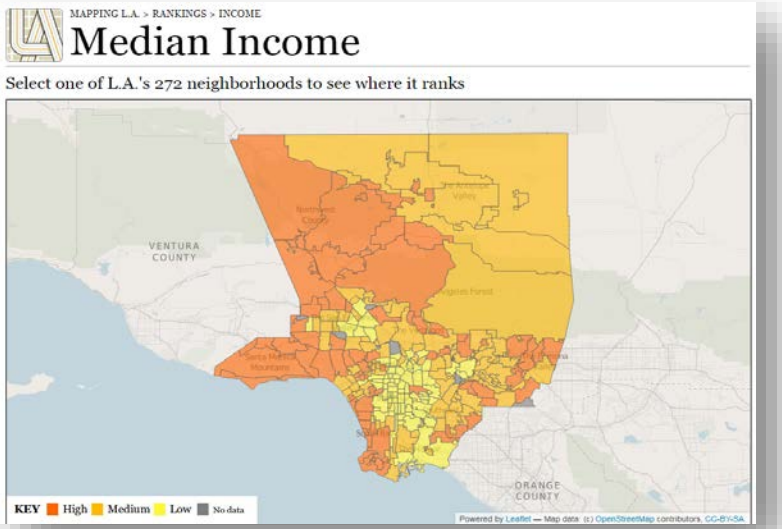
Extracted data from websites

set	slug	the_geom	kind	external_i	name	display_na	sqmi	type	name_1	slug_1	latitude	longitude	location
L.A. County N	acton	MULTIPOL	L.A. Count	acton	Acton	Acton L.A.	39.33911	unincorporated-area			-118.17	34.49736	POINT(34.497355239240846 -118.16981019229348)
L.A. County N	adams-nor	MULTIPOL	L.A. Count	adams-nor	Adams-No	Adams-No	0.80535	segment-of-a-city			-118.3	34.03146	POINT(34.031461499124156 -118.30020800000011)
L.A. County N	agoura-hill	MULTIPOL	L.A. Count	agoura-hill	Agoura Hill	Agoura Hill	8.14676	standalone-city			-118.76	34.14674	POINT(34.146736499122795 -118.75988450000015)
L.A. County N	agua-dulce	MULTIPOL	L.A. Count	agua-dulce	Agua Dulc	Agua Dulc	31.46263	unincorporated-area			-118.317	34.50493	POINT(34.504926999796837 -118.3171036690717)
L.A. County N	alhambra	MULTIPOL	L.A. Count	alhambra	Alhambra	Alhambra I	7.623814	standalone-city			-118.137	34.08554	POINT(34.085538999123571 -118.13651200000021)
L.A. County N	alondra-pa	MULTIPOL	L.A. Count	alondra-pa	Alondra Pa	Alondra Pa	1.139894	unincorporated-area			-118.335	33.88962	POINT(33.889617004889644 -118.33515598608159)
L.A. County N	artesia	MULTIPOL	L.A. Count	artesia	Artesia	Artesia L.A.	1.632204	standalone-city			-118.08	33.8669	POINT(33.866895999126271 -118.08010100000017)
L.A. County N	altadena	MULTIPOL	L.A. Count	altadena	Altadena	Altade	8.710338	unincorporat			-118.136	34.19387	POINT(34.193870502232173 -118.13623898201556)
L.A. County N	angeles-cr	MULTIPOL	L.A. Count	angeles-cr	Angeles C	Ang	7					1394	POINT(34.313937005895312 -117.9223952817848)
L.A. County N	arcadia	MULTIPOL	L.A. Count	arcadia	Arcadia	Arca						3323	POINT(34.133229999123017 -118.03041899311202)
L.A. County N	arleta	MULTIPOL	L.A. Count	arleta	Arleta	Arleta L.A.	3.096179	segment-of-a-city			-118.431	34.2431	POINT(34.243099999121583 -118.4307575)
L.A. County N	arlington-h	MULTIPOL	L.A. Count	arlington-h	Arlington H	Arlington H	1.031415	segment-of-a-city			-118.323	34.04491	POINT(34.04491049912405 -118.3234085)
L.A. County N	athens	MULTIPOL	L.A. Count	athens	Athens	Athens L.A.	1.332753	unincorporated-area			-118.305	33.92369	POINT(33.923692505935 2 -118.30465647554277)
L.A. County N	atwater-vill	MULTIPOL	L.A. Count	atwater-vill	Atwater Vil	Atwater Vil	1.776894	segment-of-a-city			-118.262	34.13107	POINT(34.131066356759177 -118.26237347966236)
L.A. County N	avalon	MULTIPOL	L.A. Count	avalon	Avalon	Avalon L.A.	2.744697	standalone-city			-118.327	33.33695	POINT(33.336954499133086 -118.32733223477572)
L.A. County N	avocado-h	MULTIPOL	L.A. Count	avocado-h	Avocado H	Avocado H	2.948459	unincorporated-area			-118.001	34.04088	POINT(34.040881003821966 -118.0012614768012)
L.A. County N	azusa	MULTIPOL	L.A. Count	azusa	Azusa	Azusa L.A.	9.871436	standalone-city			-117.912	34.13747	POINT(34.13746999912302 -117.91246849999999)
L.A. County N	vermont-sl	MULTIPOL	L.A. Count	vermont-sl	Vermont-S	Vermont-S	1.442453	segment-of-a-city			-118.29	33.98391	POINT(33.983913999124709 -118.29035750000011)
L.A. County N	baldwin-hil	MULTIPOL	L.A. Count	baldwin-hil	Baldwin Hi	Baldwin Hi	2.883705	segment-of-a-city			-118.358	34.01197	POINT(34.01197027055953 -118.35774600000005)
L.A. County N	baldwin-pa	MULTIPOL	L.A. Count	baldwin-pa	Baldwin Pa	Baldwin Pa	6.778016	standalone-city			-117.975	34.08111	POINT(34.081109499123691 -117.97519050000002)
L.A. County N	bel-air	MULTIPOL	L.A. Count	bel-air	Bel-Air	Bel-Air L.A.	6.373321	segment-of-a-city			-118.458	34.10206	POINT(34.102056999123342 -118.45841550000007)
L.A. County N	bellflower	MULTIPOL	L.A. Count	bellflower	Bellflower	Bellflower	6.195312	standalone-city			-118.129	33.88801	POINT(33.888013499126606 -118.12903150000017)
L.A. County N	bell-garder	MULTIPOL	L.A. Count	bell-garder	Bell Garde	Bell Garde	2.480285	standalone-city			-118.15	33.96569	POINT(33.965685999125014 -118.14993600000002)

Coordinates



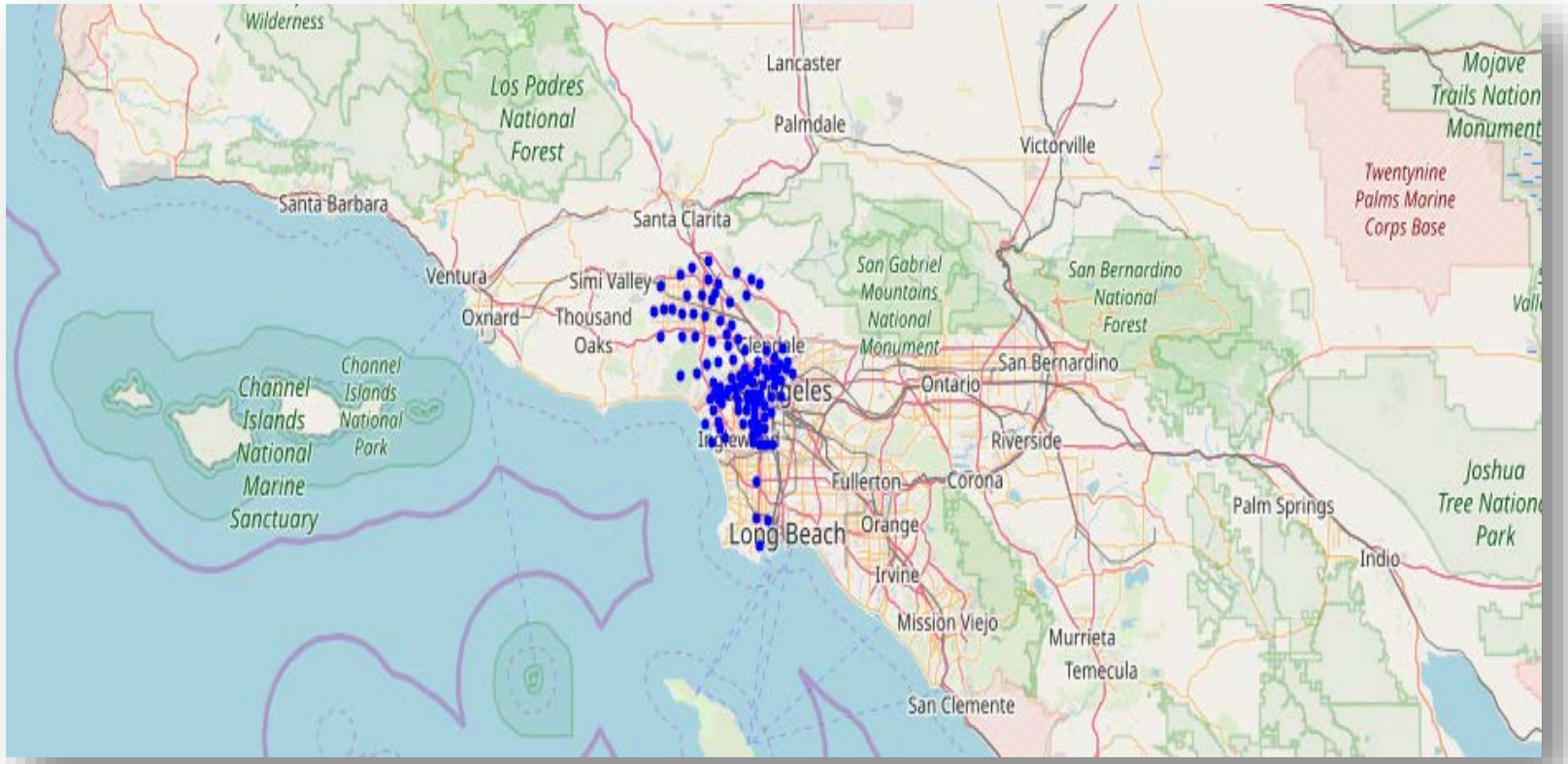




Methodology

- ✓ Data Extraction/Preparation
  - Neighborhood Name, Location and Area Size
  - Crime data
  - Income rate
- ✓ Exploratory Data Analysis
- ✓ Exploration of Los Angeles Neighborhoods
  - Selection of Target Area
  - Exploration of Venues in All Neighborhood
  - Categorize Venues
  - Analyze Each Neighborhood (Ranking)
- ✓ Cluster Neighborhoods
  - K-Mean Method
  - Agglomerative method

Prepared data frame



Exploratory data Analysis

- The size of data set Containing Neighborhood, Area size and Coordinates of Los Angeles city is (110, 7). The dataframe has 110 unique neighborhoods.
- Pacific Palisades with the area of 22.83 is the largest, and Larchmont with the area of 0.48 is the smallest neighborhoods in Los Angeles city.
- Fairfax with the property crime rate of 345.8 has the highest, and Century City with the property crime of 33.1 has the lowest rate in Los Angeles city.
- Chesterfield Square with the violent crime rate of 191.2 has the highest, and Bel-Air with the violent crime of 0.0 has the lowest rate in Los Angeles city.
- Bel-Air with the median income of 207938.0 $ has the highest income, and Downtown with the median income of 15003.0 $ has the lowest income in Los Angeles city.

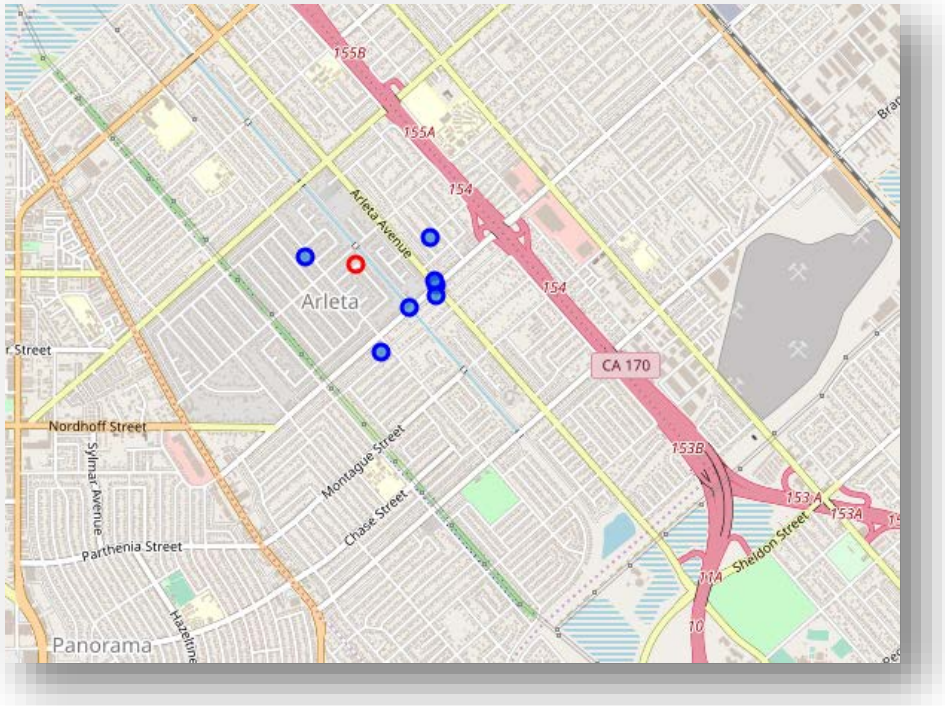
Exploration of Neighborhoods

Top 2 food-related venues for each neighborhood

	Neighborhood	1st Most Common Food Venue	2nd Most Common Food Venue
0	Arleta	Thai Restaurant	Bakery
1	Canoga Park	Mexican Restaurant	Ice Cream Shop
2	Chatsworth	Wings Joint	Vietnamese Restaurant
3	Encino	Japanese Restaurant	Deli / Bodega
4	Lake Balboa	Mexican Restaurant	Burger Joint

Top 2 other venues for each neighborhood

	Neighborhood	1st Most Common Other Venue	2nd Most Common Other Venue
0	Arleta	Video Store	Historic Site
1	Canoga Park	Sports Bar	Furniture / Home Store
2	Chatsworth	Park	Wine Bar
3	Encino	ATM	Supplement Shop
4	Lake Balboa	Convenience Store	Pharmacy



Map of Venues (blue) around neighborhood Arleta (red)

Cluster Neighborhoods

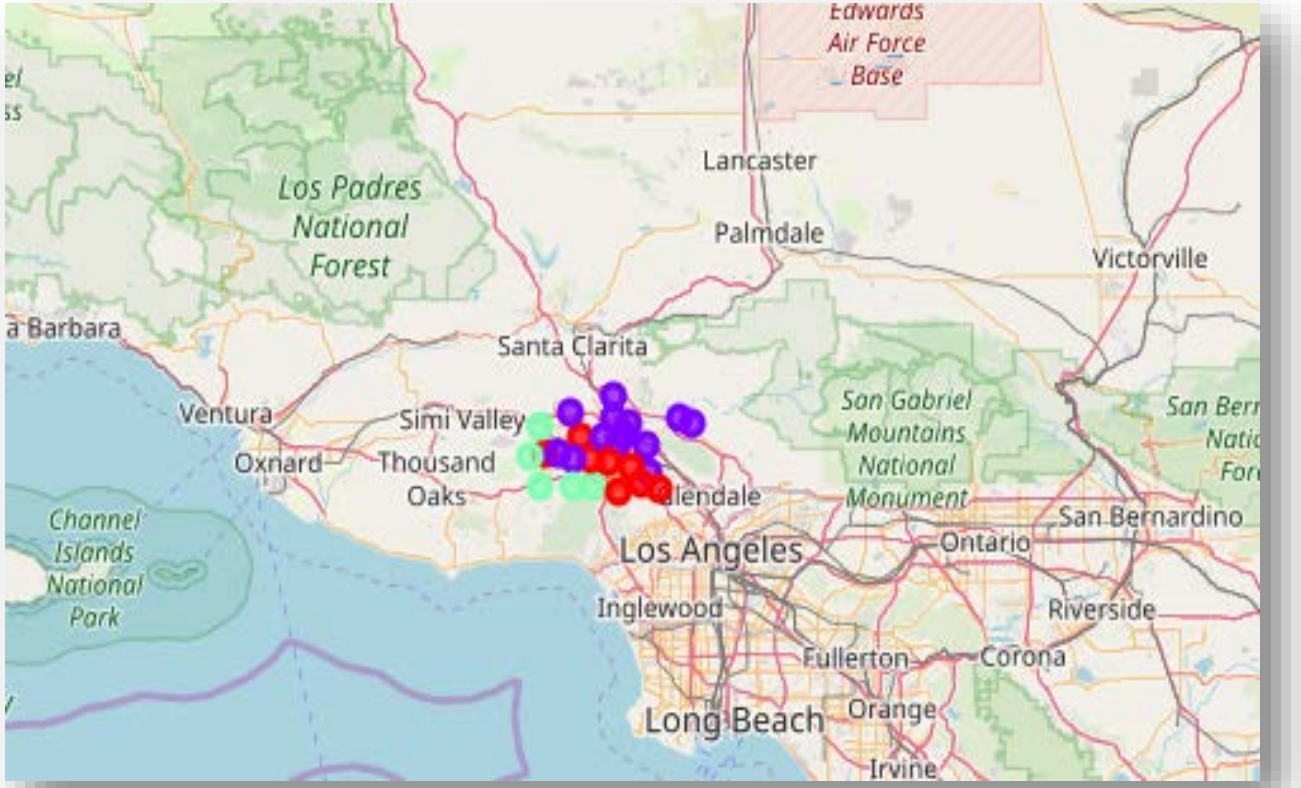
- K-Means is a type of partitioning clustering which divides the data into K non-overlapping sphere-like clusters without any labels. It is relatively efficient on medium to large-sized data sets. This method tries to minimize intra-cluster distances and maximize the inter-cluster distances.

All data used in the clustering process (first 5 rows)

	Neighborhood	Square_Mile	Violent Crime Per Capita	Property Crime Per Capita	Median Income	sum_food_places	sum_other_places
0	Arleta	3.096179	12.0	66.8	65649.0	3.0	4
1	Canoga Park	4.348518	22.8	157.2	51601.0	5.0	5
2	Chatsworth	15.243597	18.6	111.3	84456.0	0.0	2
3	Northridge	9.467487	15.6	141.6	67906.0	11.0	2
4	Encino	9.499707	9.0	106.8	78529.0	13.0	9

Results

The below figure depict three clusters using K-mean method.



Discussion

The below table shows the comparison of different features for all clusters.

Attribute	cluster 1			cluster 2			cluster 3		
	MIN	AVG	MAX	MIN	AVG	MAX	MIN	AVG	MAX
Area Size	8.52	11.36	15.25	1.2	5.4	9.46	3	6.18	12.45
Violent Crime Rate	8.4	14.44	20.1	8.2	16.81	28.1	3.9	15.8	23.8
Property Crime Rate	87.9	115.8	148.5	95.1	127.7	157.2	50.4	77.5	122.9
Median Income	73000	86000	103000	41000	58800	73000	42000	62500	121000
No. of Food Realted Place	0	3.4	13	4	9.4	17	0	3.5	9
No. of Other Place	1	3.8	9	2	9.4	20	1	3.1	9

Discussion

- Cluster 1 (cyan)

This cluster contains 5 neighborhoods. The average area size of this cluster is around 11 which is the highest among all clusters. the violent and property crime rates are 14.5 and 115.8, respectively which introduce this cluster as the highest crime rate after cluster 2. The income range for this cluster is 73000 to 103000. The ratio of the number of food-related venues to other kind of venues is 0.9.

- Cluster 2 (red)

This cluster contains 8 neighborhoods. The average area size of this cluster is around 5.4 which is the lowest among all clusters. the violent and property crime rates are 16.8 and 127.7, respectively which is the highest crime rate after cluster 2. The income range for this cluster is 41000 to 73000. This is the lowest income rate among all clusters. The ratio of the number of food-related venues to other kind of venues is 1. It means that on average the number of venues related to food is equal to the number of other venues like the gym, etc.

- Cluster 3 (purple)

This cluster contains 13 neighborhoods. The average area size of this cluster is around 6.2. the violent and property crime rates are 15.8 and 77.5, respectively which is the lowest crime rate among all clusters. The income range for this cluster is 42000 to 121000. The ratio of the number of food-related venues to other kind of venues is 0.88.

Conclusion

- Based on the above cluster descriptions, **cluster 3** is the best areas to open a restaurant. This cluster has minimum criminal rates, minimum food-related venues to other type ratio and wide range of incomes.

	Square_Mile	Violent Crime Per Capita	Property Crime Per Capita	Median Income	Cluster Labels	1st Most Common Food Venue	2nd Most Common Food Venue	sum_food_places	1st Most Common Other Venue	2nd Most Common Other Venue	sum_other_places
0	3.096179	12.0	66.8	65649.0	2	Thai Restaurant	Bakery	3.0	Video Store	Historic Site	4.0
6	3.187498	15.9	102.9	75675.0	2	Wings Joint	Vietnamese Restaurant	0.0	Church	Park	2.0
7	5.308118	23.2	79.7	52456.0	2	Pizza Place	Fast Food Restaurant	2.0	Grocery Store	River	2.0
8	5.869115	21.2	122.9	42791.0	2	Latin American Restaurant	Indian Restaurant	9.0	Wine Bar	Mobile Phone Shop	9.0
9	7.137554	18.0	65.5	49066.0	2	Breakfast Spot	Food Truck	2.0	Home Service	Wine Bar	1.0
10	3.647996	23.8	101.2	44468.0	2	Mexican Restaurant	Wings Joint	1.0	Skating Rink	Park	3.0
11	5.586170	3.9	50.4	121428.0	2	American Restaurant	Vietnamese Restaurant	1.0	Park	Electronics Store	3.0
12	5.865491	18.5	69.7	54771.0	2	Vietnamese Restaurant	Fast Food Restaurant	9.0	Convenience Store	Supermarket	2.0
14	12.456388	15.3	60.2	65783.0	2	Pizza Place	Mexican Restaurant	7.0	Business Service	Garden Center	1.0
15	4.004117	7.9	60.2	68720.0	2	Food Court	Wings Joint	1.0	Concert Hall	Park	2.0
16	9.423924	19.8	92.8	51290.0	2	Taco Place	Food Truck	4.0	Convenience Store	Electronics Store	4.0
19	10.015818	8.5	58.6	58001.0	2	Wings Joint	Vietnamese Restaurant	0.0	Lake	Trail	2.0
24	4.777241	17.5	76.4	62535.0	2	Ice Cream Shop	Fried Chicken Joint	6.0	Convenience Store	Home Service	5.0

- Within the selected cluster and based on ranked food venues or even other venues we can offer the type of food-related venues in different neighborhood. It should be noted that in this table the two top venues are shown. One may consider the less common type of venues. Looking at both, the most common and less common food-related is recommended.