Areas of high criminality in Mexico City

Alberto Ruiz Gayosso

December 19, 2019

Introduction

Mexico city is one of the most populous cities in the world. With a population of about 8 918 653 people, the capital of Mexico represents a very important cultural and financial centre in the world. Despite this condition, one of the main problems in Mexico city is the high rates of criminality that affects virtually every person living in it. Being the most important financial centre of the country, it is fundamental to identify areas with high rates of criminality and a huge influx of people. On the one hand, criminal activities such as homicides, robbery or kidnapping are a major concern for people moving to these areas because of the need to get to their workplaces, on the other hand, activities like extortion might force business owners to close or entrepreneurs to avoid these areas leading to a decrease in economic activity.

In this report we analyze neighborhoods with high rates of criminality and their most common venues using the Foursquare API. With the information obtained, a clustering algorithm is applied to find the type of crimes that are associated to a particular category of venues. Temporal patterns are also studied in order to determine the hours and days with higher rates of criminality.

The results obtained are intended to help people working in these places to avoid hours with high criminal activity, also, investors willing to open new businesses might benefit from this information by locating the best areas for a particular type of venue.

Data

Data description

The dataset used in this study was obtained from the web page of Mexico City data [1] and consists of records of crimes under investigation in Mexico city from years 1973 to 2019, in this report we only focus in the last four years. Since not all crimes are reported to the police, the dataset is only a sample of the total number of crimes, therefore it is assumed that the sample is a representative one. This dataset consists of 865 108 rows and 18 columns.

Information about crime category, date and time of occurrence, borough and neighborhood in which crimes occur and location coordinates (latitude and longitude) are provided in the dataset. Further filtering of data was made to restrict the crime categories considered, specifically to those related to the criminal activities mentioned in the previous section. The Foursquare API [3] was used to collect data about the ten most common venues in each neighborhood.

Finally, geojson data for boroughs in Mexico City was downloaded from the same web page [1] and used to create maps in Folium library to gain visual insight.

Feature selection

As previously mentioned, the crime dataset is filtered since many columns are redundant or irrelevant for the purposes of this study. Only columns related to date and place of ocurrence are kept. Figure 1 shows the first rows of the dataset.

	Year	Month	Datetime	Crime	Crime_category	Borough	Neighborhood	Longitude	Latitude
0	2016	Septiembre	2016-09-16 12:48:00	DAÑO EN PROPIEDAD AJENA CULPOSA POR TRÁNSITO V...	DELITO DE BAJO IMPACTO	IZTAPALAPA	MINERVA	-99.110789	19.350066
1	2016	Julio	2016-07-08 10:00:00	FALSIFICACION DE TITULOS AL PORTADOR Y DOCUMEN...	DELITO DE BAJO IMPACTO	COYOACAN	COPILCO EL BAJO	-99.181971	19.340739
2	2016	Septiembre	2016-09-08 00:00:00	DDH OTRAS MATERIAS	HECHO NO DELICTIVO	MIGUEL HIDALGO	LOMAS DE CHAPULTEPEC I SECCIÓN	-99.205634	19.429673
3	2016	Diciembre	2016-12-07 00:00:00	DDH OTRAS MATERIAS	HECHO NO DELICTIVO	CUAUHTEMOC	DOCTORES	-99.153406	19.424215
4	2016	Julio	2016-07-01 12:00:00	FALSIFICACION DE DOCUMENTOS	DELITO DE BAJO IMPACTO	TLALNEPANTLA DE BAZ	NaN	NaN	NaN

Figure 1: First rows of the dataset showing the relevant features for the study.

From figure 1 can be observed that there are rows with null values in latitude and longitude columns, this rows were remove since the locations of each neighborhood is required for the posterior analysis of the data. Also, the column datetime was converted to numpy datetime datatype for easier handling. The columns **Crime** and **Crime_category** are similar in content, however the first one is very specific showing around 130 types of crimes while the second only contains 15 crime categories. Therefore, the **Crime_category** column will be used for the clustering algorithm.

Methodology

Temporal patterns of crime activity

To better understand data we need to make an exploratory analysis so we can decide which areas of the city have the most impact on the security of the city. For this purpose we begin our study by looking for temporal patterns of criminal activity. In Figure 2, the number of crimes per day from January 2016 to November 2019 is shown. It is evident from this figure that there was an increase in the number of crimes per day since the beginning of 2016 up to finals of 2018. In January 2019 a sudden decrease is observed which might be due to the change in government that generated a lot of expectation among the population. Additionally, in Figure 2 it is shown the number of crimes for each hour and day of the week, we can observe that the pattern is the same for every day with only a difference in the height of the peak for the weekend. This peak is always occurring at 12:00 PM and it is a very pronounced one.

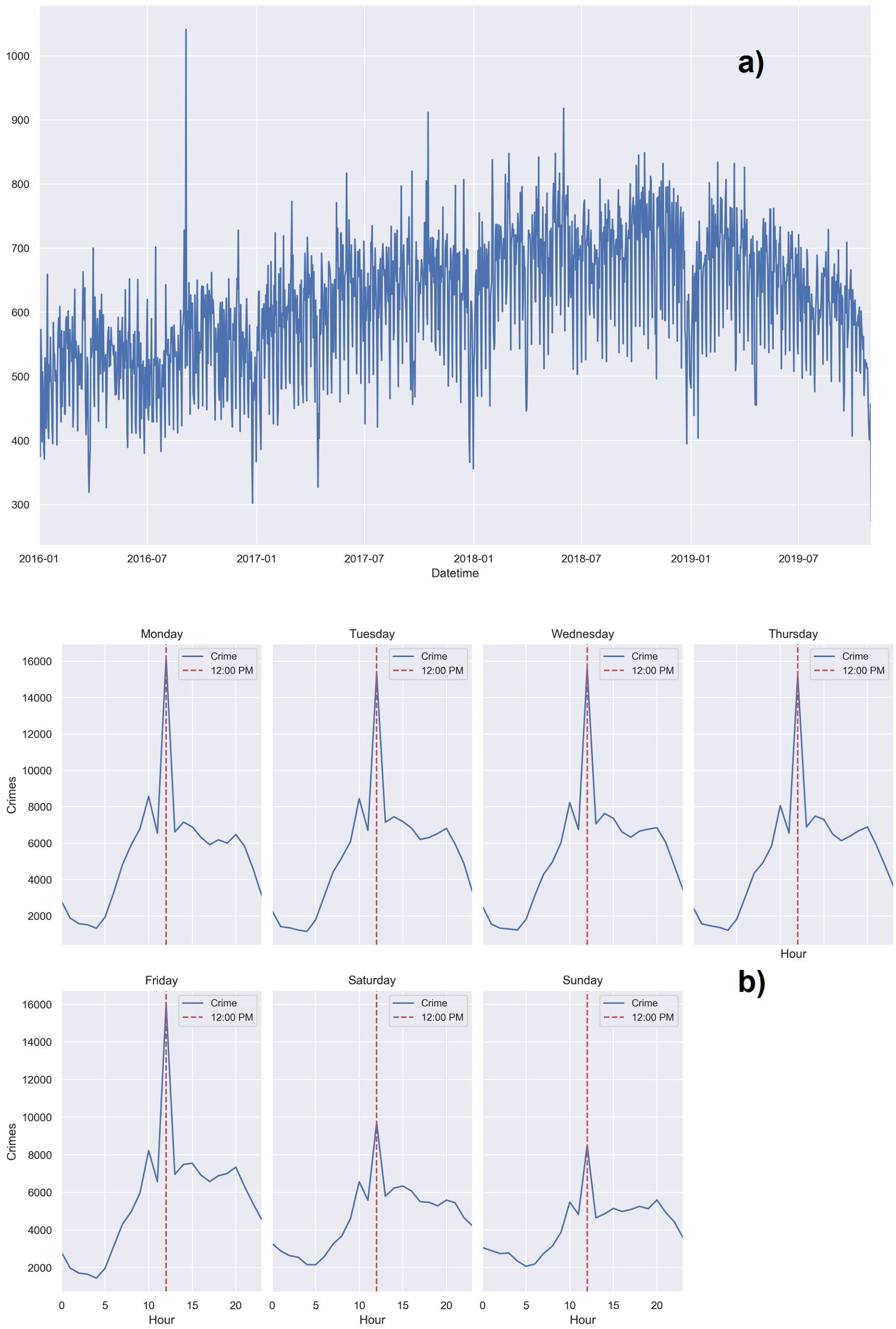

Figure 2: Temporal patterns of criminal activity. In a) crimes per day from January 2016 to November 2019 are shown, an increase of the number of crimes per day is observed from January 2016 to November 2019. The number of crimes at very hour for each day of the week are shown in b). Here, a pronounced peak is observed at 12:00 AM exceeding the 15000 crimes per hour, this peak is lower on weekends which is about 8000 crimes per hour.

Areas with highest rates of criminality

The second part of the study consists in determining the areas with high rates of criminal activity, here, the analysis is in the most part visual. In Table 1, the number of crimes in each borough of Mexico city are presented, these areas can be visualized in Figure 3 where it is evident that areas of higher criminal activity correspond to the central region of the city. This region is also central to the economic activity which means that there is a huge influx of people into it and contains most of the Metro system stations that serve more than one billion passengers each year. For this reasons, the boroughs in this central region are of particular importance for this study. In Figure 4, the neighborhoods with more than 200 crimes over the four years are shown. About 75% of the total number neighborhoods have less than or equal to 252 crimes which means that only a small fraction of them might be considered as dangerous. In fact, of the 525 neighborhoods with more than 200 crimes only 80 of them concentrate the 50% of the total number of crimes over the four years. As expected, these neighborhoods are located in boroughs of the central region of the city. This region comprises Cuauhtémoc, Iztapalapa, Gustavo A Madero, Benito Juárez, Coyoacán, Iztacalco, and Miguel Hidalgo boroughs.

Table 1: **Number of crimes in each borough of Mexico city from 2016 to 2019.** The boroughs with higher number of crimes correspond to the central region of Mexico city.

Borough	Crimes
Cuauhtémoc	136604
Iztapalapa	130555
Gustavo A Madero	86436
Benito Juárez	74506
Coyoacán	58033
Alvaro Obregón	57469
Miguel Hidalgo	57280
Venustiano Carranza	50168
Tlalpan	49681
Azcapotzalco	41881
Iztacalco	36854
Xochimilco	26374
Tláhuac	19907
La Magdalena Contreras	13043
Cuajimalpa de Morelos	12198
Milpa Alta	5641

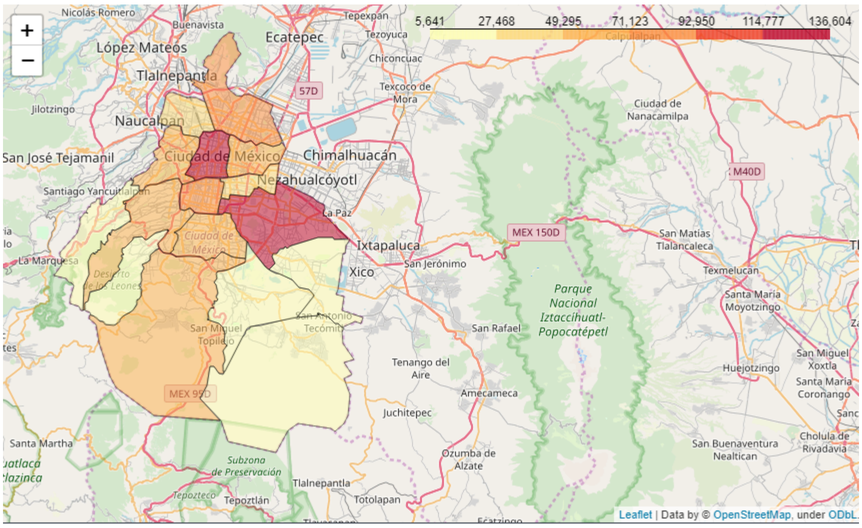

Figure 3: Here, the number of crimes in each borough is shown codified by the color bar. It is evident that boroughs in the central region of the city are the most dangerous.

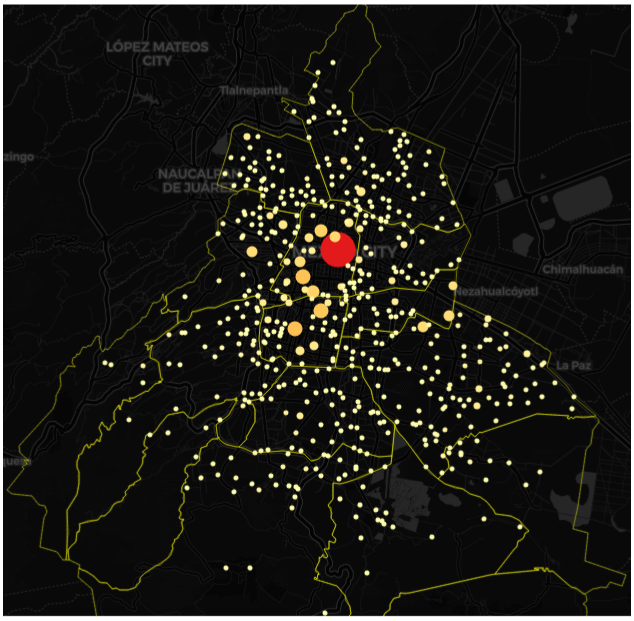

Figure 4: Neighborhoods with more than 200 crimes over the four years. The radius and color of the circles represent the total number of crimes in each neighborhood. Again, the larger circles are concentrated around the central region of the city.

Clustering of neighborhoods

For the third part of the study we identify the 100 most common venues in each neighborhood within a radius of 500 meters using Foursquare API. Table 2 shows the 5 most common venues in all the city, Mexican restaurants and Taco places are the most common venues. This is not a surprise since tacos are the most famous dish in Mexico. However it is important to note that Mexican restaurants and taco places are not exclusive of the central region of the city but they are located all around the city, therefore this type of venues might not be very important in determining the clusters of neighborhoods. Moreover, not all neighborhoods have 100 venues within a radius of 100 meters so we are only considering those with more than 10 different venues.

Table 2: **Most common venues in neighborhoods with more than 200 venue categories.**

Venue	Frequency
Mexican Restaurant	1720
Taco Place	1320
Coffee Shop	636
Restaurant	500
Bakery	455

In figure 5 the results of the K-means clustering algorithm are shown. The 15 crime categories and the corresponding venue categories for each neighborhood were used to train the algorithm, location coordinates were not considered.

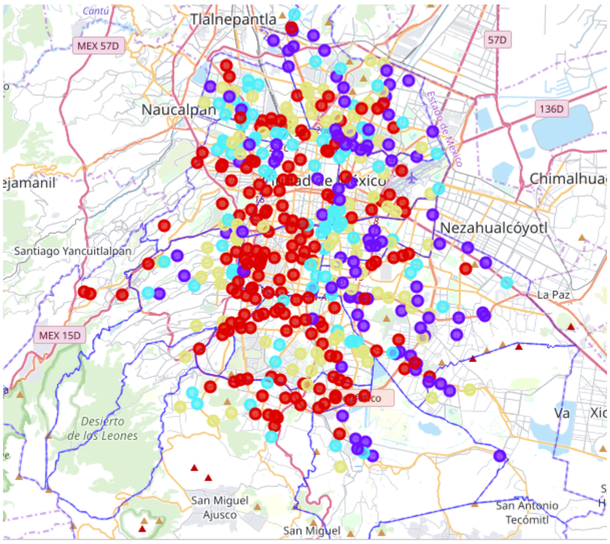

Figure 5: Neighborhoods grouped into 4 clusters according to crime and venue categories. Circles with the same color represent members of the same cluster.

To better understand the different the groups, it is necessary to look for the most common crime and venue categories which determine each one of the clusters. Once determined the defining categories, a name can be assigned to each cluster to represent its characteristics. In table 3, the most common crimes and venues categories for each cluster are displayed.

Results

In the methodology section, we discussed the temporal patterns of crime activity (Figure 2), both the number of crimes per day from January 2016 to November 2019 and the number of crimes per hour for each day of the week. In the first case it is evident an increase in the number of crimes from 2016 until the end of 2018, at this point, a sharp decrease is observed in a brief period time which might be related to the change of goverment. After this event, the number of crimes rises again to recover its original trend. In the second case, it is interesting to note that each day of the week follows the same pattern being the only difference between them in the height of the peak in the number of crimes per hour. This peak reveals that criminals prefer to commit crimes at midday when people are concentrated in working areas. This is also related to the fact that the most dangerous bouroughs and neighborhoods are located in the central region of the city where economic activity is larger, this can be seen in figures 3 and 4.

In respect of the data presented in Table 3, it is observed that there is no clear relationship between the most common crimes and the most common venues in a particular cluster. It is clear that the characteristics of the clusters are very similar and that the separation of neighborhoods into different groups is based on small differences in the number of occurrences of a particular crime or venue category in each neighborhood.

Discussion

In this study we analyze criminal activity and its relationship with the most common venues in Mexico city. For this purpose, data of crimes under investigation that contains information about crime category, date of occurrence and localization of the crime scene is used. We explore temporal patterns to gain insight into the increment of criminality from January 2016 to

November 2019 and the hours of highest criminal activity. As discussed in the previous section, the results displayed in Figure 2 reveal the evolution of criminal activity during this period of time and which hours and days are preferred by criminals. Future studies might be focused on finding the relationship between different crime categories and the hours in which they are most frequent.

As a second result, the most dangerous boroughs and neighborhoods were found. These results are displayed in figures 3 and 4 where we can observe how the majority of crimes are commited in the central region of the city. This is consistent with that fact that this region concentrates most of the working areas and present a large influx of people from other parts of the city.

For the final part of the study we apply a clustering algorithm to the set of the most dangerous neighborhoods to find a relationship between the most common crime and venue categories. The latter obtained using the Foursquare API. The results of the application are shown in Figure 5 and the most common categories are summarized in Table 3. From the data observed in this table we can conclude that there is no significant relationship between the most common venues in a neighborhood and the most common crimes commited in it.

References

Mexico City Crime Data: https://datos.cdmx.gob.mx/explore/dataset/victimas-en-carpetas-de-investigacion-pgj/export/

Table 3: Most common crimes and venues of each cluster. Each column represent neighborhoods with the same cluster label. The numbers following the crime or venue category stand for

Cluster Label	0	1	2	3
1st Most Common Crime	Low-level crimes 165	Low-level crimes 92	Low-level crimes 83	Low-level crimes 93
2nd Most Common Crime	Car Theft 87 Robbery on the Street 76	Car Theft 48 Robbery on the Street 38	Car Theft 42 Robbery on the Street 39	Car Theft 56 Robbery on the Street 36
3rd Most Common Crime	Robbery on the Street 84 Car Theft 65	Robbery on the Street 54 Car Theft 30	Robbery on the Street 41 Car Theft 35	Robbery on the Street 56 Car Theft 33
4th Most Common Crime	Business Robbery 113	Business Robbery 52 Delivery Man Robbery 18	Business Robbery 56 Delivery Man Robbery 10	Business Robbery 60 Delivery Man Robbery 9
5th Most Common Crime	Delivery Man Robbery 40 Account Holder Robbery 29 Business Robbery 27 Robbery in Metro System 22	Delivery Man Robbery 42 Business Robbery 25	Delivery Man Robbery 28 Business Robbery 12 Homicide 11	Delivery Man Robbery 28 Homicide 18 Business Robbery 14
6th Most Common Crime	Delivery Man Robbery 41 Account Holder Robbery 28 House Robbery 24 Homicide 21	Homicide 29 Robbery on Public Transport 22 Delivery Man Robbery 18	Homicide 21 Robbery on Public Transport 19	Delivery Man Robbery 24 Homicide 22 House Robbery 14
7th Most Common Crime	Robbery on Public Transport 25 Homicide 25 Rape 21	Homicide 23 Robbery on Public Transport 21 Rape 17	Homicide 14 House Robbery 13 Robbery on Public Transport 11 Account Holder Robbery 11	Homicide 14 Robbery on Public Transport 14 Delivery Man Robbery 14
8th Most Common Crime	Account Holder Robbery 21 Robbery on Public Transport 21 Rape 21	Rape 20 House Robbery 18 Account Holder Robbery 16 Homicide 13	Account Holder Robbery 13 Robbery on Public Transport 13 Rape 13 Homicide 13	Rape 16 House Robbery 16 Robbery on Public Transport 15 Account Holder Robbery 12
9th Most Common Crime	Rape 35	Rape 21 Account Holder Robbery 17 Robbery on Public Transport 15 House Robbery 15	Robbery on Public Transport 17 Rape 12 House Robbery 11 Homicide 10	Rape 18 Robbery on board a Taxi 12 House Robbery 12
10th Most Common Crime	Rape 21 House Robbery 19	Account Holder Robbery 17 Robbery on board a Taxi 13 Rape 11	House Robbery 13 Robbery on board a Taxi 11	Rape 14 Robbery on board a Taxi 12 Account Holder Robbery 11

1st Most Common Venue	Mexican Restaurant 49 Coffee Shop 22	Mexican Restaurant 33 Taco Place 21	Mexican Restaurant 81	Taco Place 74
2nd Most Common Venue	Mexican Restaurant 25 Taco Place 22	Mexican Restaurant 17 Taco Place 17 Convenience Store 9	Taco Place 33	Mexican Restaurant 43 Convenience Store 12 Taco Place 12
3rd Most Common Venue	Taco Place 24 Mexican Restaurant 16 Coffee Shop 16	Bakery 7 Mexican Restaurant 6 Pizza Place 6 Taco Place 6 Convenience Store 5	Taco Place 12 Convenience Store 8	Convenience Store 10 Mexican Restaurant 8 Seafood Restaurant 7 Restaurant 6 Pizza Place 6 Taco Place 6 Bakery 5
4th Most Common Venue	Coffee Shop 14 Mexican Restaurant 13 Bakery 12 Restaurant 11 Taco Place 10	Mexican Restaurant 11 Bakery 6 Gym 6 Restaurant 5 Gym / Fitness Center 4	Taco Place 7 Convenience Store 6 Restaurant 5 Gym / Fitness Center 4	Mexican Restaurant 9 Convenience Store 8 Pizza Place 6 Seafood Restaurant 5 Coffee Shop 5
5th Most Common Venue	Bakery 12 Coffee Shop 11	Pharmacy 6 Seafood Restaurant 5 Convenience Store 5 Restaurant 4 Coffee Shop 4 Bar 4	Coffee Shop 7 Taco Place 6 Bar 6 Bakery 4 Convenience Store 4	Convenience Store 5 Bakery 5 Food Truck 4 Park 4 Restaurant 3
6th Most Common Venue	Restaurant 15 Taco Place 13	Restaurant 6 Gym 5 Taco Place 4 Coffee Shop 4	Bakery 5 Seafood Restaurant 4 Athletics & Sports 4 Market 4	Market 6 Restaurant 5 Burger Joint 5 Mexican Restaurant 5 Seafood Restaurant 4 Pizza Place 4
7th Most Common Venue	Pizza Place 7 Ice Cream Shop 7 Restaurant 6 Bakery 6 Taco Place 6 Sushi Restaurant 5 Coffee Shop 5 Italian Restaurant 5 Bar 5 Gym / Fitness Center 5	Mexican Restaurant 4 Clothing Store 4 Park 4 Coffee Shop 3 BBQ Joint 3 Diner 3 Food Truck 3 Caf 3 Convenience Store 3	Market 6 Caf 5 Bakery 4 Gym 4 Bar 4 Restaurant 4	Food Truck 5 Restaurant 4 Ice Cream Shop 4 Pharmacy 4 Seafood Restaurant 4 Coffee Shop 3

8th Most Common Venue	Restaurant 10 Coffee Shop 10 Seafood Restaurant 9	Taco Place 6 Restaurant 4 Caf 4 Mexican Restaurant 3 Gym 3	Restaurant 6 Park 6 Taco Place 6 Snack Place 5 Bakery 4	Mexican Restaurant 6 Bakery 5 Paper / Office Supplies Store 4 Park 3 Food Truck 3
9th Most Common Venue	Bakery 9 Restaurant 7 Caf 6 Seafood Restaurant 6 Pizza Place 5 Clothing Store 5 Japanese Restaurant 4 Breakfast Spot 4 Gym / Fitness Center 4	Taco Place 6 Bakery 5 Seafood Restaurant 5 Clothing Store 3 Food Truck 3 Gym / Fitness Center 3 Gym 2 Coffee Shop 2	Burger Joint 4 Food Truck 4 Bakery 3 Park 3 Coffee Shop 3 Athletics & Sports 3	Bakery 6 Seafood Restaurant 5 Park 4 Restaurant 4 Grocery Store 3
10th Most Common Venue	Taco Place 8 Pizza Place 7 Sushi Restaurant 6 Bakery 6 Restaurant 6 Seafood Restaurant 6 Gym 5	Coffee Shop 4 Bakery 4 Food 3 BBQ Joint 3 Caf 3 Hotel 3 Pharmacy 3 Electronics Store 3	Bakery 5 Park 4 Caf 4 Pizza Place 3	Ice Cream Shop 5 Bakery 3 Burger Joint 3 BBQ Joint 3 Seafood Restaurant 3 Restaurant 3 Fast Food Restaurant 2 Bar 2 Brazilian Restaurant 2