Capstone Project Assignment.

Safe Community to Run a Convenience Store

Grant Gloster

12 May 2019

1. Introduction:

Background

Running a convenience store in the world today can be a very dangerous job. But not every community is as risky as the. By compiling crime statistics from law enforcement agency, one can gauge these answers. Combining this data with the Foursquare data we can work out saturation levels for the same types of stores in the area and adjust the location to ensure you have less competition.

By knowing this information perspective shop owners can see where is a good place to establish a business and at what time of the day are safer to run their business.

Problem

Project aims to help investors to determine a safe and highly profitable community to invest into to build a convenience store using crime statistics and per capita income reports for communities of Chicago

Interest

Business investors who want to commit money into a profitable investment by building stores.

2. Data acquisition and Cleaning

Data Sources

Crime statistics are a common data resource that is made available to the public by law enforcement departments. Chicago crime stats are available here. This data is filtered on only 2018 incidents and the exported to a csv file that is uploaded to the “Data” folder in the project.

Chicago community boundary data is saved in a geojson file available here. This is used to map out the neighborhoods when displaying data on the map. The file is uploaded to the “Data” folder in the project.

Community names and ID’s are compiled and displayed on a webpage here. This is processed and saved in an csv file. The file is uploaded to the “Data” folder in the project.

Chicago Per Capita Income is available as a csv export here. The file is exported from the page then uploaded to the “Data” folder in the project.

Data Cleaning

Data is imported from the uploaded files. This is then processed to get a usable format to start the assessment process.

The crime incidents data needs to be cleaned up to remove unnecessary columns that won't be needed for data analysis. The features dropped are (**'Case Number', 'ID', 'FBI Code', 'X Coordinate', 'Y Coordinate', 'Year', 'Updated On', 'Location', 'Historical Wards 2003-2015', 'Zip Codes', 'Community Areas', 'Census Tracts', 'Wards', 'Boundaries - ZIP Codes', 'Police Districts', 'Police Beats'**).

Now the data is filtered on crime incidents where the "Location Description" is part of the following list (**'SMALL RETAIL STORE', 'DEPARTMENT STORE', 'CONVENIENCE STORE', 'GROCERY FOOD STORE'**). This will provide information on store related crimes in the different communities of Chicago.

Split the date into two separate features, date and time. Add an incident column at the end of the data that will be used to calculate community totals when processing the data.

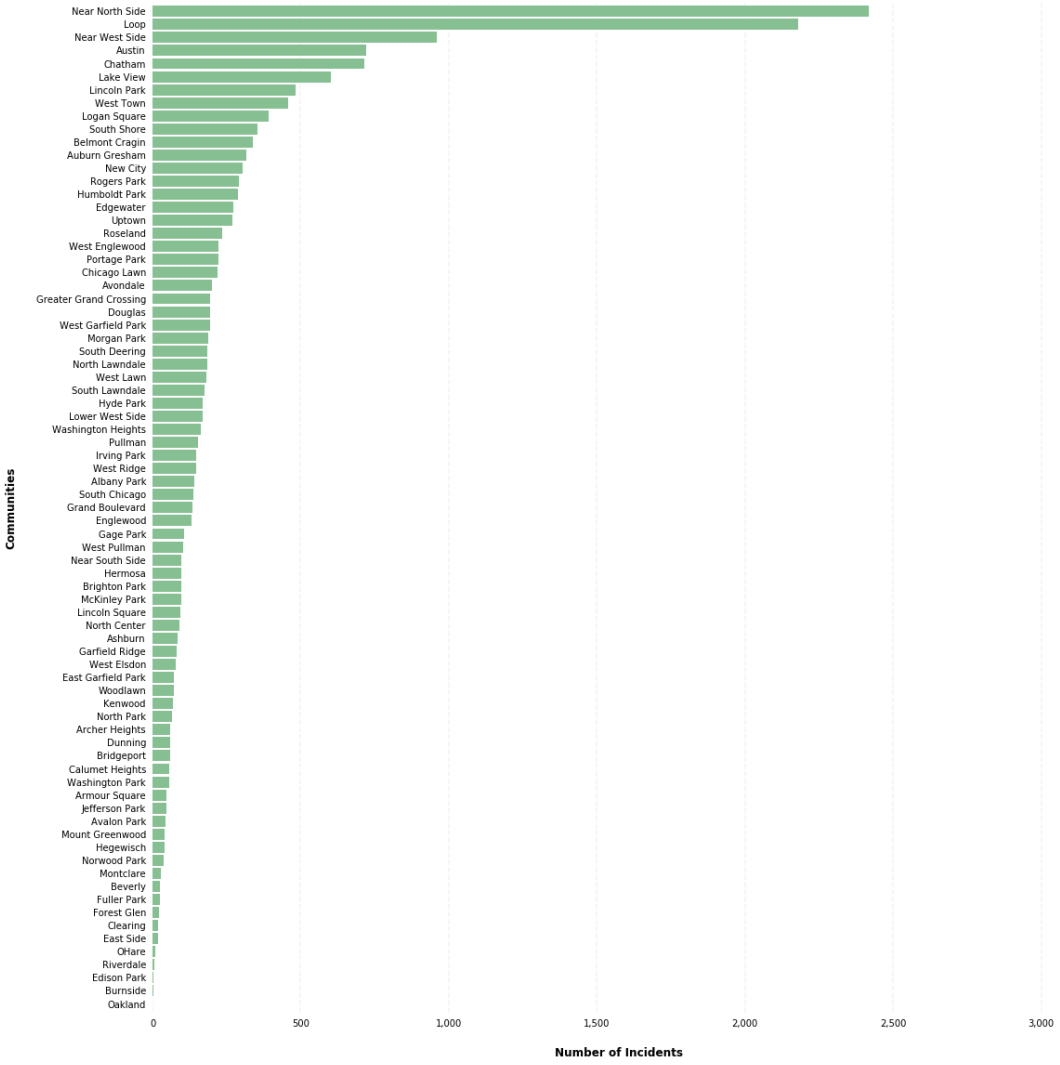
The income data is also cleaned up by removing all the unneeded columns (**'PERCENT AGED UNDER 18 OR OVER 64','PERCENT AGED 25+ WITHOUT HIGH SCHOOL DIPLOMA','PERCENT OF HOUSING CROWDED','COMMUNITY AREA NAME','PERCENT HOUSEHOLDS BELOW POVERTY','PERCENT AGED 16+ UNEMPLOYED'**). This is then merge with the community's data so that each community entry will have a per capita income and hard ship index.

3. Exploratory Data Analysis

Calculations

Taking the data available we group by Community Area and sum the Incidents on each row. This will give us a totals value for each community of the criminal incidents reported committed where location is a store. This will provide us information about the levels of crime in each community to help is make our first choice.

Taking the data we generate a bar graph which is ordered by total number of shop related crimes for the community's for 2018.

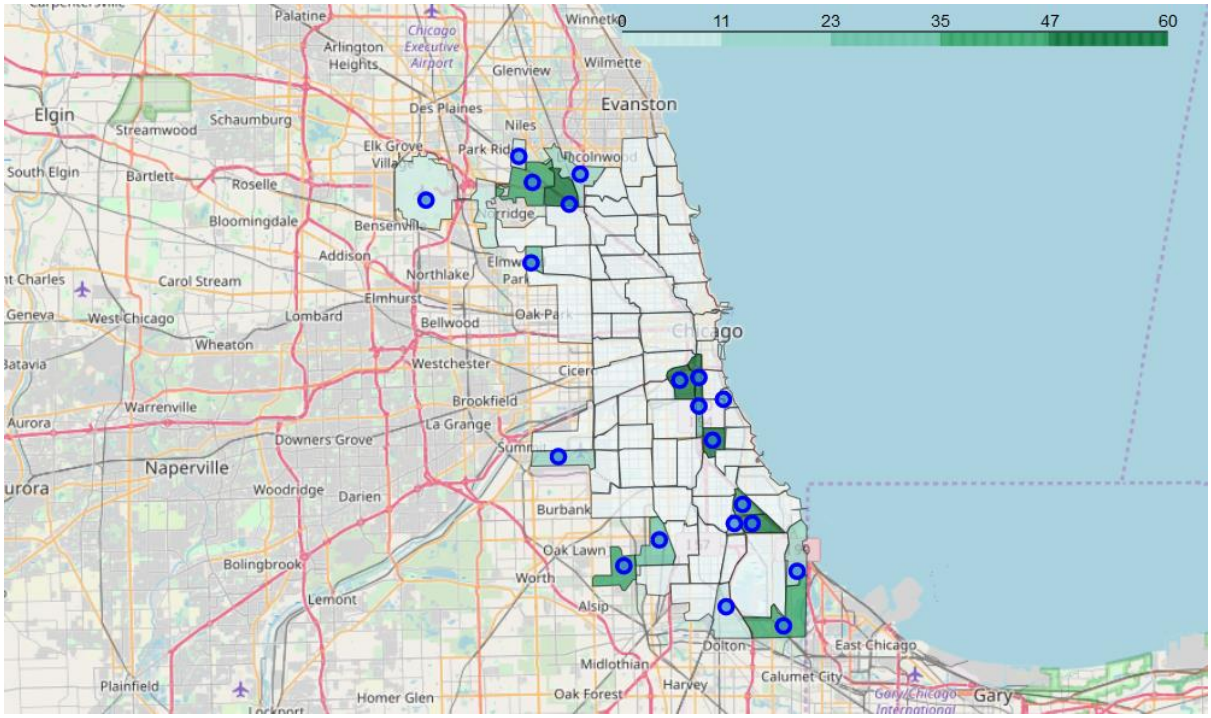


Using the data, we see there is a whole group of communities that have a very low annual incident counts. The bottom 20 communities are used for evaluation because the main purpose is to find the community that has the lowest incidents but also financially viable.

Community Area	Community Name	Incidents	PER CAPITA INCOME	HardshipIndex	Latitude	Longitude
36	Oakland	0.0	19252	78.0	41.823653	-87.60824
47	Burnside	3.0	12515	79.0	41.730035	-87.59671
9	Edison Park	4.0	40959	8.0	42.005733	-87.81400

Community Area	Community Name	Incidents	PER CAPITA INCOME	HardshipIndex	Latitude	Longitude
54	Riverdale	6.0	8201	98.0	41.667835	-87.60496
76	OHare	9.0	25828	24.0	41.973101	-87.90676
64	Clearing	19.0	25113	29.0	41.780588	-87.77338
52	East Side	19.0	17104	64.0	41.694618	-87.53338
12	Forest Glen	21.0	44164	11.0	41.991752	-87.75167
37	Fuller Park	24.0	10432	97.0	41.818089	-87.63255
72	Beverly	26.0	39523	12.0	41.718153	-87.67176
18	Montclare	27.0	22014	50.0	41.925309	-87.80089
10	Norwood Park	37.0	32875	21.0	41.985590	-87.80057
55	Hegewisch	40.0	22677	44.0	41.653646	-87.54698
74	Mount Greenwood	41.0	34381	16.0	41.698089	-87.70866
45	Avalon Park	45.0	24454	41.0	41.745035	-87.58865
11	Jefferson Park	47.0	27751	25.0	41.969738	-87.76311
34	Armour Square	48.0	16148	82.0	41.840033	-87.63310
40	Washington Park	55.0	13785	88.0	41.792534	-87.61810
48	Calumet Heights	57.0	28887	38.0	41.730035	-87.57921
60	Bridgeport	59.0	22694	43.0	41.837938	-87.65102

Using this list we plot out the community's based on their incident counts. We can review the data to if the communities are in the same geographical locations or separated.



Every marker indicates a different community center point. As shown they are strewn across the city so our assessment is not localized to only one part of the city.

Community Shops

Using FourSquare service provider, we retrieve all the shops that are listed for each community. This data is useful to provide a list of shops in the area to could be competitors for your business.

Using the data retrieved we calculate the frequency of each type of venue for the communities. We transpose the data and make each venue type as a different column value. Each row is a community and the value is the mean value of the venue frequency.

Community Name	ATM	Accessories Store	Adult Boutique	Antique Shop	Auto Dealership	Auto Garage	Auto Workshop	Automotive Shop	Bank	...	Storage Facility	Supermarket	Tanning Salon	Tattoo Parlor
Armour Square	0.000000	0.00	0.000000	0.000	0.000000	0.032258	0.000000	0.129032	0.000000	...	0.032258	0.000000	0.032258	0.000000
Avalon Park	0.028571	0.00	0.000000	0.000	0.028571	0.000000	0.000000	0.085714	0.000000	...	0.000000	0.000000	0.000000	0.028571
Beverly	0.000000	0.00	0.025641	0.000	0.000000	0.000000	0.000000	0.025641	0.025641	...	0.025641	0.000000	0.000000	0.000000
Bridgeport	0.000000	0.00	0.000000	0.000	0.000000	0.000000	0.000000	0.162791	0.069767	...	0.000000	0.000000	0.000000	0.000000
Burnside	0.000000	0.00	0.000000	0.000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000
Calumet Heights	0.000000	0.00	0.000000	0.000	0.000000	0.000000	0.076923	0.076923	0.000000	...	0.000000	0.000000	0.000000	0.000000
Clearing	0.045455	0.00	0.000000	0.000	0.000000	0.000000	0.000000	0.045455	0.045455	...	0.000000	0.000000	0.045455	0.000000
East Side	0.000000	0.00	0.000000	0.000	0.000000	0.000000	0.000000	0.166667	0.166667	...	0.000000	0.000000	0.000000	0.000000
Edison Park	0.000000	0.00	0.000000	0.000	0.000000	0.000000	0.000000	0.000000	0.031250	...	0.000000	0.000000	0.000000	0.000000
Forest Glen	0.000000	0.00	0.000000	0.000	0.000000	0.000000	0.000000	0.105263	0.000000	...	0.000000	0.000000	0.000000	0.000000
Fuller Park	0.000000	0.00	0.000000	0.000	0.000000	0.000000	0.000000	0.250000	0.000000	...	0.000000	0.000000	0.000000	0.000000
Hegewisch	0.000000	0.00	0.000000	0.000	0.000000	0.000000	0.000000	0.058824	0.235294	...	0.000000	0.000000	0.000000	0.000000
Jefferson Park	0.000000	0.00	0.000000	0.000	0.000000	0.000000	0.000000	0.076923	0.102564	...	0.000000	0.025641	0.000000	0.000000

Location, Location, Location

Because the data set has 91 columns and 20 rows it makes it difficult to assess the data visually. So, to make it more readable we need to calculate top 10 most common stores in the area. This will give

us an indication if there are a lot of convenience stores and what other types of stores are available that might bolster our business.

The data is ordered by hardship index from lowest to highest which will help show more affluent areas. Taking this and comparing it to the incidents counts we analyze the data to make a decision on the community to pick.

Community Area	Community Name	Incidents	PER CAPITA INCOME	HardshipIndex	Latitude	Longitude	1st Most Common Store	2nd Most Common Store	3rd Most Common Store	4th Most Common Store	5th Most Common Store	6th Most Common Store	7th Most Common Store	8th Most Common Store	9th Most Common Store	10th Most Common Store
9	Edison Park	4.0	40959	8.0	42.005733	-87.814004	Salon / Barbershop	Spa	Grocery Store	Photography Lab	Real Estate Office	Print Shop	Clothing Store	Cosmetics Shop	Nail Salon	Design Studio
12	Forest Glen	21.0	44164	11.0	41.991752	-87.751673	Financial or Legal Service	Miscellaneous Shop	Credit Union	Automotive Shop	Grocery Store	Currency Exchange	Salon / Barbershop	Nail Salon	Mobile Phone Shop	Boutique
72	Beverly	26.0	39523	12.0	41.718153	-87.671767	Salon / Barbershop	Boutique	Cosmetics Shop	Financial or Legal Service	Spa	Clothing Store	Miscellaneous Shop	Farmers Market	Women's Store	Storage Facility
74	Mount Greenwood	41.0	34381	16.0	41.698089	-87.708662	Cosmetics Shop	Salon / Barbershop	Women's Store	Electronics Store	Credit Union	Currency Exchange	Daycare	Deli / Bodega	Design Studio	Discount Store
10	Norwood Park	37.0	32875	21.0	41.985590	-87.800577	ATM	Automotive Shop	Furniture / Home Store	Flower Shop	Pet Store	Cheese Shop	Real Estate Office	Hardware Store	Farmers Market	Daycare
76	OHare	9.0	25828	24.0	41.973101	-87.906768	Newsstand	Electronics Store	Miscellaneous Shop	Chocolate Shop	Bookstore	Gift Shop	Credit Union	Convenience Store	Food & Drink Shop	Accessories Store
11	Jefferson Park	47.0	27751	25.0	41.969738	-87.763118	Salon / Barbershop	Bank	Automotive Shop	Financial or Legal Service	Pharmacy	Video Store	Laundry Service	Paper / Office Supplies Store	Other Repair Shop	Convenience Store
64	Clearing	19.0	25113	29.0	41.780588	-87.773388	Salon / Barbershop	Video Store	Nail Salon	ATM	Credit Union	Pharmacy	Office	Real Estate Office	Record Shop	Business Service
48	Calumet Heights	57.0	28887	38.0	41.730035	-87.579213	Salon / Barbershop	Furniture / Home Store	Hardware Store	Pharmacy	Gas Station	Laundry Service	Auto Workshop	Automotive Shop	Women's Store	Dry Cleaner
45	Avalon Park	45.0	24454	41.0	41.745035	-87.588658	Salon / Barbershop	Automotive Shop	Nail Salon	Boutique	Car Wash	Miscellaneous Shop	Health Food Store	Health & Beauty Service	Pet Service	Mobile Phone Shop
60	Bridgeport	59.0	22694	43.0	41.837938	-87.651028	Automotive Shop	Pharmacy	Bank	Grocery Store	Laundry Service	Convenience Store	Mobile Phone Shop	Massage Studio	Insurance Office	Salon / Barbershop
55	Hegewisch	40.0	22677	44.0	41.653646	-87.546988	Bank	Salon / Barbershop	Laundry Service	Grocery Store	Hardware Store	Pet Store	Discount Store	Market	Currency Exchange	Automotive Shop
18	Montclare	27.0	22014	50.0	41.925309	-87.800893	Automotive Shop	Salon / Barbershop	Furniture / Home Store	Grocery Store	Bank	Liquor Store	Optical Shop	Cosmetics Shop	Nail Salon	Clothing Store
52	East Side	19.0	17104	64.0	41.694618	-87.533387	Convenience Store	Video Store	Health & Beauty Service	Salon / Barbershop	Automotive Shop	Bank	Electronics Store	Credit Union	Currency Exchange	Daycare
36	Oakland	0.0	19252	78.0	41.823653	-87.608242	Salon / Barbershop	Gas Station	Health & Beauty Service	Discount Store	Grocery Store	Pharmacy	Furniture / Home Store	Liquor Store	Daycare	Optical Shop
47	Burnside	3.0	12515	79.0	41.730035	-87.596714	Salon / Barbershop	Convenience Store	Electronics Store	Cosmetics Shop	Credit Union	Currency Exchange	Daycare	Deli / Bodega	Design Studio	Discount Store
34	Armour Square	48.0	16148	82.0	41.840033	-87.633107	Automotive Shop	Gas Station	Spa	Salon / Barbershop	Grocery Store	Event Service	Massage Studio	Miscellaneous Shop	Nail Salon	Cosmetics Shop
40	Washington Park	55.0	13785	88.0	41.792534	-87.618105	Salon / Barbershop	Gas Station	Grocery Store	Liquor Store	Convenience Store	Record Shop	Recycling Facility	Mobile Phone Shop	Bookstore	Miscellaneous Shop
37	Fuller Park	24.0	10432	97.0	41.818089	-87.632551	Gas Station	Automotive Shop	Cosmetics Shop	Gourmet Shop	Liquor Store	Recycling Facility	Salon / Barbershop	Convenience Store	Discount Store	Credit Union
54	Riverdale	6.0	8201	98.0	41.667835	-87.604964	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Assumptions

Looking at the data it shows that Edison Park is at the top of the list. Low hardship index and incident value but the third most common store in the area is Grocery Store which means lots of possible competition. This also shows that there is a market for grocery stores in the area and the more unique the store the good possibility to be successful.

Forest Glen is a good option but it has a high incident rate for a community where the Grocery Store is the 5th most common store. So, lots of incidents for the few stores that are available. But it is the most viable because as you progress down the list the count of incidents goes down but the per capita income for the community goes down as well. This will affect the type of store you want to put up and the type of products you will sell to the public.

4. Conclusion

Using this type of format, you can review different cities and communities for a low incident area to invest into a store. The biggest problem will be cleaning up the data at the start and putting it in a usable format to be assessed. The data is mostly available for big cities but some areas do not supply all related data needed.

https://gist.github.com/3b65fa4698c75a3c3a11838385ef3fe4