Coursera - IBM Data Science Professional

Applied Data Science - Capstone Project

Date: December 2020

Name: Ching Chieh (Marcos) Hsiao-Tseng

Analysis for the Lowest Crime Ridden Community Areas in Chicago in 2020

Introduction

The purpose of this capstone project will be to analyze the publicly available Chicago crime data for 2020 and present meaningful visualizations. The finding is intended for individuals, families, businesses, and others who, during these uncertain times, are considering relocating to a safer part of Chicago or looking to move into Chicago.

Data Description

Crimes Data: The city of Chicago has collected and makes its crime data easily accessible on its website (Chicago Data Portal, 2020). The dataset used in this analysis is Crimes - 2020. It is parsed only for crimes that led to an arrest. 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 in 2020. A detailed description of this dataset can be obtained from the Chicago Data Portal

Community Area Names: The city of Chicago is divided into 77 community areas for statistical and planning purposes (Wikipedia, 2020). The crimes data file lists community area numbers but not the names. The available table in Wikipedia is pulled and used to match the community area numbers with respective names.

GeoJSON Data: A geojson file (Kaggle, 2020) of the Chicago Community Areas was used to create the choropleth 2020 crime map.

Chicago Geolocation: An online geolocation search (Google, 2020) yielded the latitude and longitude values for Chicao.

Venues info: The Foursquare API (Foursquare, 2020) is the powerhouse used in the gathering of information on surrounding venues for the community areas.

Methodology

For this project a jupyter (IBM, 2020) notebook was used for the python coding along with github (GitHub, 2021) as the repository. Readily available libraries for python were deployed and used for the importing, handling, analyzing, clustering, and mapping of the data.

Action 1 started with the downloading of the crime data file. This was loaded onto a dataframe and reviewed. Clean-up was conducted by removing records that did not lead to an arrest and for records that had incomplete entries.

Action 2 involved grabbing the community area names from wikipedia and merging this to the dataset. As the Chicago crime data file only lists community area numbers, this step was necessary to match the numbers to the community area names for ease of identification.

Action 3 grouped the crime dataset by the community area names yielding the total number of crimes for each community area. This provided the data to construct the “Crimes That Led to Arrests - Chicago in 2020” visual bar chart.

Action 4 started with a google search for the geolocation of Chicago and a geojson file for the Chicago community areas. Assembling this together with the grouped crime dataset allowed for the creation of a choropleth map of the 'Crime Rate in Chicago in 2020'.

Action 5, firstoff, reduced the crime dataset by date to the latest 1000 records with the aim to produce a cluster map. The crime dataset was too large (over 30k+ records) to handle, so a decision was made to trim it down to a manageable size. That said, the date and range can easily be adjusted should the interested parties seek additional specifics. An interactive cluster map was generated with these 1000 records with zoomable capabilities that pinpoints and labels the type of crime for each incident.

Action 6 decoded the dataset to display in table format the top 1 thru 10 most common crime offenses that occured at each of the 77 community areas in Chicago in calendar year 2020.

Action 7 sorted the dataset resulting in a dataframe table listing the 10 community areas with the fewest crime incidents for the year.

Action 8 pulled Foursquare top venues info for these 10 safest community areas and displayed the findings in a table.

Results

Action 1: Crimes dataset

	Date	Block	Primary Type	Description	Location Description	Arrest	Community Area	Latitude	Longitude
0	03/18/2020 05:35:00 PM	003XX S CICERO AVE	NARCOTICS	MANUFACTURE / DELIVER - CRACK	SIDEWALK	1.0	25	41.876128	-87.745112
1	03/18/2020 11:52:00 PM	056XX S MARYLAND AVE	INTERFERENCE WITH PUBLIC OFFICER	RESIST / OBSTRUCT / DISARM OFFICER	STREET	1.0	41	41.792195	-87.604959
2	03/16/2020 12:05:00 PM	089XX S WALLACE ST	BATTERY	DOMESTIC BATTERY SIMPLE	RESIDENCE	1.0	71	41.731656	-87.638630
3	03/17/2020 09:30:00 PM	023XX S ST LOUIS AVE	BATTERY	AGG. DOMESTIC BATTERY - HANDS, FISTS, FEET, SE...	RESIDENCE	1.0	30	41.848873	-87.712480
4	02/24/2020 02:54:00 AM	052XX N SHERIDAN RD	THEFT	OVER $500	OTHER (SPECIFY)	1.0	77	41.976852	-87.655000
...	...	...	...	...	...	...	...	...	...
32603	12/26/2020 02:15:00 AM	009XX E 81ST ST	BATTERY	DOMESTIC BATTERY SIMPLE	RESIDENCE	1.0	44	41.747661	-87.601919
32604	12/26/2020 07:45:00 AM	0000X S MASON AVE	BATTERY	DOMESTIC BATTERY SIMPLE	SIDEWALK	1.0	25	41.879783	-87.773012
32605	12/26/2020 08:50:00 PM	004XX S PULASKI RD	CRIMINAL TRESPASS	TO VEHICLE	STREET	1.0	26	41.874744	-87.725459
32606	12/26/2020 02:18:00 PM	065XX S MARSHFIELD AVE	WEAPONS VIOLATION	UNLAWFUL POSSESSION - HANDGUN	STREET	1.0	67	41.775175	-87.665370
32607	12/26/2020 07:37:00 PM	003XX W 74TH ST	WEAPONS VIOLATION	UNLAWFUL POSSESSION - HANDGUN	STREET	1.0	69	41.759930	-87.632615

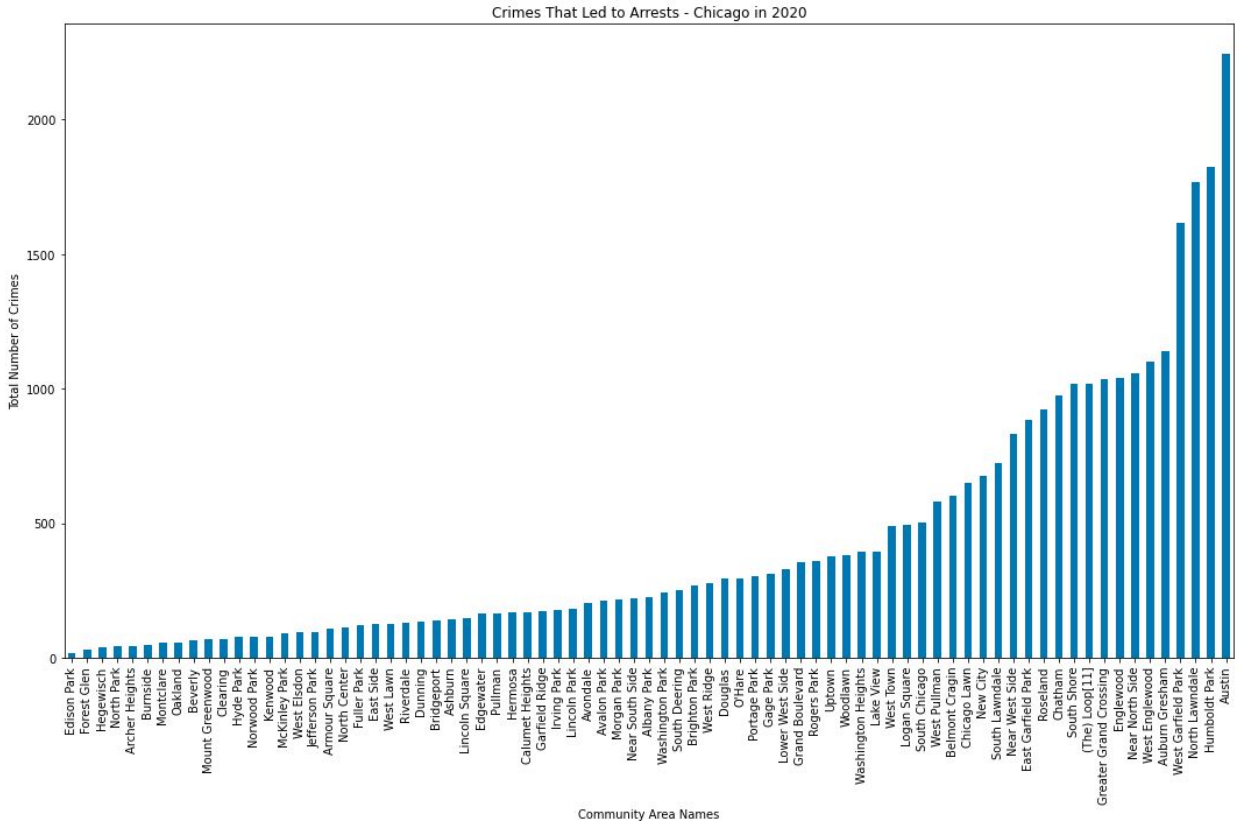
32608 rows × 9 columns

Action 2: Crimes dataset with Community Area Name defined

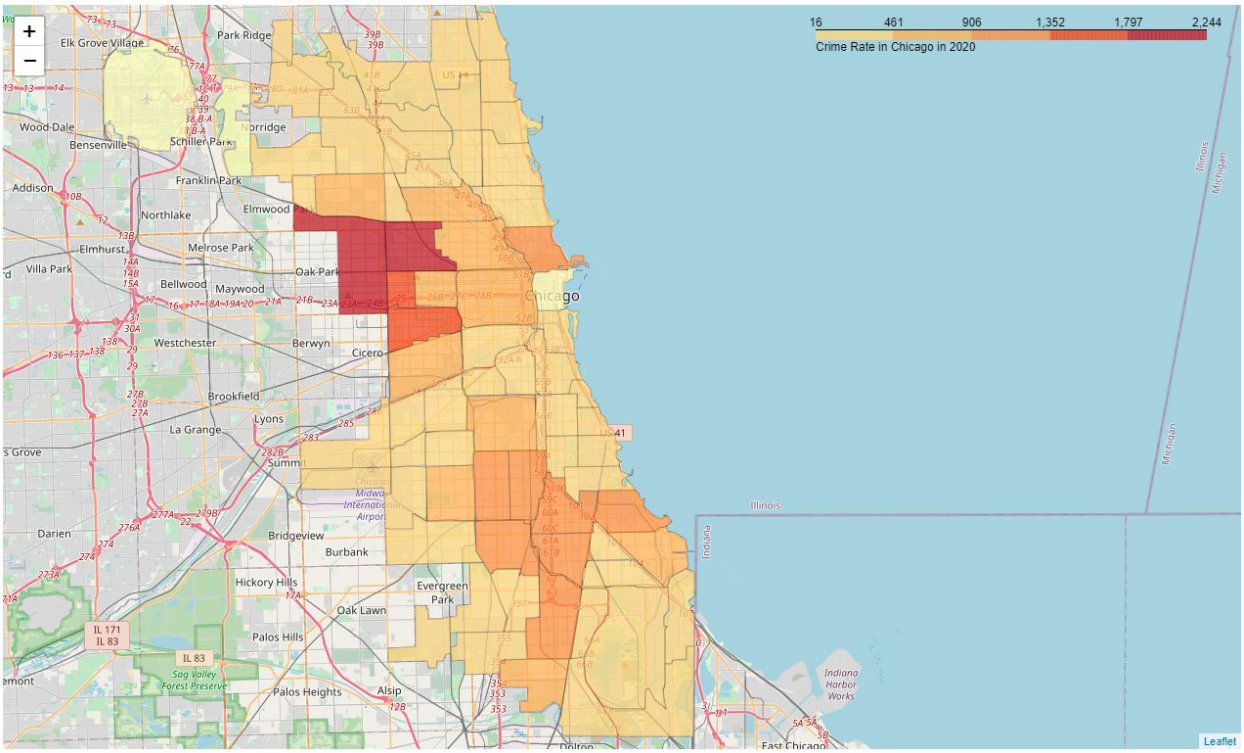
	Date	Block	Primary Type	Description	Location Description	Arrest	Community Area	Latitude	Longitude	Community Area Name
0	03/18/2020 05:35:00 PM	003XX S CICERO AVE	NARCOTICS	MANUFACTURE / DELIVER - CRACK	SIDEWALK	1.0	25	41.876128	-87.745112	Austin
1	03/18/2020 10:00:00 AM	051XX W LE MOYNE ST	BATTERY	DOMESTIC BATTERY SIMPLE	APARTMENT	1.0	25	41.907697	-87.755160	Austin
2	03/18/2020 05:06:00 PM	054XX W FERDINAND ST	CRIMINAL TRESPASS	TO RESIDENCE	RESIDENCE	1.0	25	41.888894	-87.761324	Austin
3	03/18/2020 11:37:00 AM	052XX W FULTON ST	NARCOTICS	POSSESS - HEROIN (WHITE)	SIDEWALK	1.0	25	41.885632	-87.755565	Austin
4	03/18/2020 08:15:00 AM	048XX W NORTH AVE	THEFT	RETAIL THEFT	CONVENIENCE STORE	1.0	25	41.909615	-87.747017	Austin
...	...	...	...	...	...	...	...	...	...	...
32603	05/01/2020 12:01:00 AM	032XX W BRYN MAWR AVE	PROSTITUTION	PANDERING	APARTMENT	1.0	13	41.983067	-87.709515	North Park
32604	12/02/2020 05:50:00 PM	040XX W PETERSON AVE	NARCOTICS	POSSESS - CANNABIS 30 GRAMS OR LESS	GAS STATION	1.0	13	41.990130	-87.728997	North Park
32605	12/09/2020 03:00:00 PM	057XX N TRIPP AVE	NARCOTICS	FOUND SUSPECT NARCOTICS	COMMERCIAL / BUSINESS OFFICE	1.0	13	41.985349	-87.734155	North Park
32606	12/01/2020 09:38:00 AM	063XX N MC CORMICK RD	BATTERY	AGGRAVATED PROTECTED EMPLOYEE - OTHER DANGEROU...	WAREHOUSE	1.0	13	41.996453	-87.712011	North Park
32607	12/06/2020 08:13:00 PM	040XX W FOSTER AVE	THEFT	RETAIL THEFT	CONVENIENCE STORE	1.0	13	41.975517	-87.729569	North Park

32608 rows × 10 columns

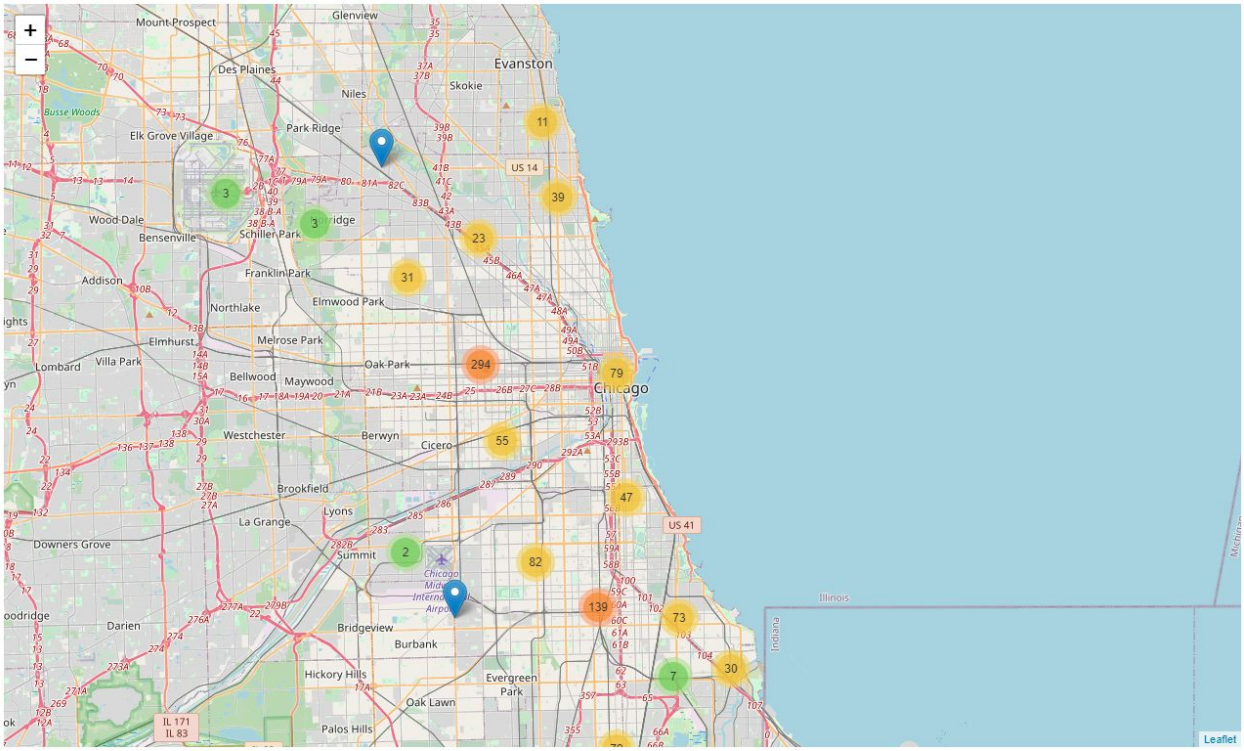
Action 3: Bar chart of the community areas with least to most crimes.



Action 4: Color coded choropleth map of the crime dataset by community area.



Action 5: Interactive cluster map pinpointing and identifying the newest 1000 indicidents. Try it out here (nbviewer, 2021). Scroll down to the map.



Action 6: List of the top 10 most common crimes for each community area.

	Community Area Name	1st Most Common Crime	2nd Most Common Crime	3rd Most Common Crime	4th Most Common Crime	5th Most Common Crime	6th Most Common Crime	7th Most Common Crime	8th Most Common Crime	9th Most Common Crime	10th Most Common Crime
0	(The) Loop[11]	THEFT	BATTERY	CRIMINAL TRESPASS	WEAPONS VIOLATION	NARCOTICS	ASSAULT	CRIMINAL DAMAGE	ROBBERY	PUBLIC PEACE VIOLATION	DECEPTIVE PRACTICE
1	Albany Park	NARCOTICS	BATTERY	THEFT	CRIMINAL DAMAGE	WEAPONS VIOLATION	ASSAULT	CRIMINAL TRESPASS	ROBBERY	OTHER OFFENSE	OFFENSE INVOLVING CHILDREN
2	Archer Heights	BATTERY	ASSAULT	THEFT	NARCOTICS	WEAPONS VIOLATION	OFFENSE INVOLVING CHILDREN	CRIMINAL TRESPASS	MOTOR VEHICLE THEFT	CRIMINAL DAMAGE	OTHER OFFENSE
3	Armour Square	WEAPONS VIOLATION	NARCOTICS	BATTERY	ASSAULT	THEFT	BURGLARY	CRIMINAL DAMAGE	DECEPTIVE PRACTICE	HOMICIDE	LIQUOR LAW VIOLATION
4	Ashburn	BATTERY	NARCOTICS	WEAPONS VIOLATION	ASSAULT	CRIMINAL TRESPASS	CRIMINAL DAMAGE	OTHER OFFENSE	MOTOR VEHICLE THEFT	THEFT	BURGLARY
...	...	...	...	...	...	...	...	...	...	...	...
72	West Lawn	THEFT	BATTERY	NARCOTICS	ASSAULT	OTHER OFFENSE	WEAPONS VIOLATION	CRIMINAL TRESPASS	CRIMINAL DAMAGE	MOTOR VEHICLE THEFT	DECEPTIVE PRACTICE
73	West Pullman	BATTERY	WEAPONS VIOLATION	NARCOTICS	OTHER OFFENSE	ASSAULT	CRIMINAL TRESPASS	THEFT	CRIMINAL DAMAGE	INTERFERENCE WITH PUBLIC OFFICER	HOMICIDE
74	West Ridge	BATTERY	NARCOTICS	ASSAULT	THEFT	WEAPONS VIOLATION	OTHER OFFENSE	CRIMINAL TRESPASS	CRIMINAL DAMAGE	MOTOR VEHICLE THEFT	BURGLARY
75	West Town	BATTERY	THEFT	ASSAULT	WEAPONS VIOLATION	NARCOTICS	CRIMINAL TRESPASS	CRIMINAL DAMAGE	OTHER OFFENSE	DECEPTIVE PRACTICE	MOTOR VEHICLE THEFT
76	Woodlawn	WEAPONS VIOLATION	BATTERY	NARCOTICS	ASSAULT	OTHER OFFENSE	THEFT	CRIMINAL DAMAGE	CRIMINAL TRESPASS	PUBLIC PEACE VIOLATION	INTERFERENCE WITH PUBLIC OFFICER

77 rows × 11 columns

Action 7: List of the 10 community areas with fewest crime incidents for Chicago for the 2020 calendar year.

	Community Area Name	Incidents
0	Edison Park, IL	16
1	Forest Glen, IL	28
2	Hegewisch, IL	40
3	North Park, IL	44
4	Archer Heights, IL	45
5	Burnside, IL	49
6	Montclare, IL	56
7	Oakland, IL	58
8	Beverly, IL	63
9	Mount Greenwood, IL	67

Action 8: List of most common venues as returned by Foursquare in December 2020 for these 10 safest Chicago community areas for calendar year 2020.

	Community Area Name	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	Archer Heights	Mexican Restaurant	Rental Car Location	Pizza Place	Donut Shop	Taco Place	Fast Food Restaurant	Sandwich Place	Bar	Discount Store	Seafood Restaurant
1	Beverly	Park	Sandwich Place	Pharmacy	Pizza Place	Grocery Store	Burger Joint	Italian Restaurant	Fast Food Restaurant	Pub	Seafood Restaurant
2	Burnside	Post Office	Convenience Store	Trail	Park	Tunnel	Construction & Landscaping	Other Great Outdoors	Yoga Studio	Food	Flower Shop
3	Edison Park	Italian Restaurant	Coffee Shop	Grocery Store	Pizza Place	American Restaurant	Pharmacy	Breakfast Spot	Salon / Barbershop	Deli / Bodega	Thai Restaurant
4	Forest Glen	Park	Coffee Shop	American Restaurant	Pizza Place	Bar	Grocery Store	Mexican Restaurant	Brewery	Middle Eastern Restaurant	Chinese Restaurant
5	Hegewisch	Mexican Restaurant	Fast Food Restaurant	Pizza Place	Bar	Sandwich Place	Seafood Restaurant	Discount Store	Grocery Store	Breakfast Spot	Café
6	Montclare	Mexican Restaurant	Fast Food Restaurant	Grocery Store	Thai Restaurant	Italian Restaurant	Coffee Shop	Pizza Place	Pharmacy	Park	Discount Store
7	Mount Greenwood	Pizza Place	Italian Restaurant	Fast Food Restaurant	Breakfast Spot	Grocery Store	Hot Dog Joint	Mexican Restaurant	Burger Joint	Coffee Shop	Discount Store
8	North Park	Pizza Place	Breakfast Spot	Italian Restaurant	Ice Cream Shop	Sandwich Place	Coffee Shop	Gym	Nature Preserve	Donut Shop	BBQ Joint
9	Oakland	State / Provincial Park	Antique Shop	Café	Pizza Place	Farm	Yoga Studio	Filipino Restaurant	Food Truck	Food & Drink Shop	Food

Discussion

Specific to addressing the business problem identified for this capstone project, it is advised for the client to consider the 10 community areas as identified in action 7 for relocation. Furthermore, action 8 provides the top venues for each area to be weighed against lifestyle and/or business needs.

Conversely, it is highly discouraged for a move into community areas with the highest numbers of crime incidents, see chart action 3.

The choropleth map per action 4 presents a quick "glance" with lighter colors indicating the safer community areas to steadily reddening worse areas. This is closely matched by the cluster map per action 5 denoting the last 1000 recorded crime incidents in Chicago for 2020.

Conclusion

Publically available data can be deconstructed, parsed, reassembled, analyzed, and visualized to provide insights and new understandings. The analysis and study to reveal the meanings hidden in the data is what data analysts/specialists/analysts/etc are trained to do.

It is the author's contemplation that the result derived from this capstone project has only scratched the surface with what can be done with this available Chicago Crimes - 2020 dataset: When do crimes most often occur? Where? What type of crimes are most prevalent during what time? Is there a pattern on which months have the highest number of crimes? Also, what types? If weather is added into the dataset, can a correlation be derived between how many/what type/time of crimes? And, what about these other variables if introduced into the dataset: socioeconomic, school performance, 2019 dataset and older… the resulting findings to these (musings) questions will commission stakeholders make better decisions!

Addendum

The majority of the code used in the python programming is derived from the teachings and instructions in the IBM Data Science Professional Certificate course (Coursera, 2020). The code is reused/reformatted/applied for the purpose of this final capstone project.

The copy of the Chicago Crimes - 2020 file used for the analysis was last retrieved on January 3, 2021. The resulting analysis and output is per the dataset constructed with this file. The Chicago Crimes - 2020 file is a live file that gets revised as there are new entries and/or updates.

References

Chicago Data Portal (2020). *Chicago Data Portal.* Retrieved December 26, 2020, from https://data.cityofchicago.org/browse?q=crimes&sortBy=relevance

Coursera (2020). *IBM Data Science Professional Certificate.* Retrieved January 2, 2021, from https://www.coursera.org/professional-certificates/ibm-data-science

Foursquare (2020). *Welcome to Foursquare Developers.* Retrieved December 26, 2020, from https://developer.foursquare.com/developer/

GitHub (2021). *Github.* Retrieved January 2, 2021, from https://github.com/

Google (2020). *Google.* Retrieved December 28, 2020, from https://www.google.com/

IBM (2020). *IBM Skills Network Labs.* Retrieved January 2, 2021, from https://labs.cognitiveclass.ai/

Kaggle (2020). *Chicago Community Areas GeoJSON.* Retrieved December 28, 2020, from https://www.kaggle.com/doyouevendata/chicago-community-areas-geojson

nbviewer (2021). *juptyer nbviewer.* Retrieved January 2, 2021, from https://nbviewer.jupyter.org/

Wikipedia (2020). *Community areas in Chicago.* Retrieved December 28, 2020, from https://en.wikipedia.org/wiki/Community_areas_in_Chicago