Capstone Project Report

Segmenting and Clustering neighborhoods in Fredericton

July 12, 2019

1 Segmenting and Clustering Neighborhoods in Fredericton, NB

1.1 Applied Data Science Capstone Week 5 Peer-Graded Project Report

Introduction to the opportunity

Fredericton is the Capital City of the only Canadian fully-bilingual Province of New Brunswick and is beautifully located on the banks of the Saint John River. While one of the least populated provincial capital cities with a population base of less than 60 thousand residents, it offers a wide spectrum of venues and is a government, university and cultural hub.

As the city grows and develops, it becomes increasingly important to examine and understand it quantitatively. The City of Fredericton provides open data for everyone and encourages entrepreneurial use to develop services for the benefit of its citizens.

Developers, investors, policy makers and/or city planners have an interest in answering the following questions as the need for additional services and citizen protection:

1. What neighborhoods have the highest crime?
2. Is population density correlated to crime level?
3. Using Foursquare data, what venues are most common in different locations within the city?
4. Does the Knowledge Park really need a coffee shop?

Does the Open Data project have specific enough or thick enough data to empower decisions to be made or is it too aggregate to provide value in its current detail? Let's find out.

[73]:
```
from IPython.display import Image
from IPython.core.display import HTML
Image(url= "http://www.tourismfredericton.ca/sites/default/files/field/image/
↪fredericton.jpg")
```

[73]: <IPython.core.display.Image object>

1.2 Data

To understand and explore we will need the following City of Fredericton Open Data:

1. Open Data Site: http://data-fredericton.opendata.arcgis.com/

2. Fredericton Neighborhoods: http://data-fredericton.opendata.arcgis.com/datasets/neighbourhoods–quartiers

3. Fredericton Crime by Neighborhood: http://data-fredericton.opendata.arcgis.com/datasets/crimeby-neighbourhood-2017–crime-par-quartier-2017

4. Fredericton Census Tract Demographics: http://data-fredericton.opendata.arcgis.com/datasets/censustract-demographics–donn%C3%A9es-d%C3%A9mographiques-du-secteur-de-recensement

5. Fredericton locations of interest:https://github.com/JasonLUrquhart/Applied-DataScience-Capstone/blob/master/Fredericton%20Locations.xlsx

6. Foursquare Developers Access to venue data: https://foursquare.com/

Using this data will allow exploration and examination to answer the questions. The neighborhood data will enable us to properly group crime by neighborhood. The Census data will enable us to then compare the population density to examine if areas of highest crime are also most densely populated. Fredericton locations of interest will then allow us to cluster and quantitatively understand the venues most common to that location.

2 Methodology

All steps are referenced below in the Appendix: Analysis section.

The methodology will include: 1. Loading each data set 2. Examine the crime frequency by neighborhood 3. Study the crime types and then pivot analysis of crime type frequency by neighborhood 4. Understand correlation between crimes and population density 5. Perform k-means statistical analysis on venues by locations of interest based on findings from crimes and neighborhood 6. Determine which venues are most common statistically in the region of greatest crime count then in all other locations of interest. 6. Determine if an area, such as the Knowledge Park needs a coffee shop.

2.0.1 Loading the data

After loading the applicable libraries, the referenced geojson neighborhood data was loaded from the City of Fredericton Open Data site. This dataset uses block polygon shape coordinates which are better for visualization and comparison. The City also uses Ward data but the Neighborhood location data is more accurate and includes more details. The same type of dataset was then loaded for the population density from the Stats Canada Census tracts.

The third dataset, an excel file, “Crime by Neighborhood 2017” downloaded from the City of Fredericton Open Data site is found under the Public Safety domain. This dataset was then uploaded for the analysis. It’s interesting to note the details of this dataset are aggregated by neighborhood. It is not an exhaustive set by not including all crimes (violent offenses) nor specific location data of the crime but is referenced by neighborhood.

This means we can gain an understanding of the crime volume by type by area but not specific enough to understand the distribution properties. Valuable questions such as, “are these crimes occurring more often in a specific area and at a certain time by a specific demographic of people?” cannot be answered nor explored due to what is reasonably assumed to be personal and private information with associated legal risks.

There is value to the city to explore the detailed crime data using data science to predict frequency, location, timing and conditions to best allocated resources for the benefit of its citizens and its police force. However, human behavior is complex requiring thick profile data by individual and the conditions surrounding the event(s). To be sufficient for reliable future prediction it would need to demonstrate validity, currency, reliability and sufficiency.

2.0.2 Exploring the data

Exploring the count of crimes by neighborhood gives us the first glimpse into the distribution.

One note is the possibility neighborhoods names could change at different times. The crime dataset did not mention which specific neighborhood naming dataset it was using but we assumed the neighborhood data provided aligned with the neighborhoods used in the crime data. It may be beneficial for the City to note and timestamp neighborhood naming in the future or simply reference with neighborhood naming file it used for the crime dataset.

An example of data errors: There was an error found in the naming of the neighborhood “Platt”. The neighborhood data stated “Plat” while the crime data stated “Platt”. Given the crime dataset was simplest to manipulate it was modified to “Plat”. The true name of the neighborhood is “Platt”.

First Visualization of Crime Once the data was prepared; a choropleth map was created to view the crime count by neighborhood. As expected the region of greatest crime count was found in the downtown and Platt neighborhoods.

Examining the crime types enables us to learn the most frequent occurring crimes which we then plot as a bar chart to see most frequently type.

Theft from motor vehicles is most prevalent in the same area as the most frequent crimes. It’s interesting to note this area is mostly residential and most do not have garages. It would be interesting to further examine if surveillance is a deterrent for motor vehicle crimes in the downtown core compared to low surveillance in the Platt neighborhood.

Examining 2nd most common crime given it is specific: theft from vehicles After exploring the pivot table showing Crime_Type by Neighborhood, we drill into a specific type of crime, theft from vehicles and plot the choropleth map to see which area has the greatest frequency.

Again, the Platt neighborhood appears as the most frequent.

Is this due to population density?

Introducing the Census data to explore the correlation between crime frequency and population density. Visualizing the population density enables us to determine that the Platt neighborhood has lower correlation to crime frequency than I would have expected.

It would be interesting to further study the Census data and if this captures the population that is renting or more temporary/transient pollution, given the City is a University hub.

2.0.3 Look at specific locations to understand the connection to venues using Foursquare data

Loading the "Fredericton Locations" data enables us to perform a statistical analysis on the most common venues by location.

We might wonder if the prevalence of bars and clubs in the downtown region has something to do with the higher crime rate in the near Platt region.

Plotting the latitude and longitude coordinates of the locations of interest onto the crime choropleth map enables us to now study the most common venues by using the Foursquare data.

Analyzing each Location Grouping rows by location and the mean of the frequency of occurrence of each category we venue categories we study the top five most common venues.

Putting this data into a pandas dataframe we can then determine the most common venues by location and plot onto a map.

2.1 Results

The analysis enabled us to discover and describe visually and quantitatively: 1. Neighborhoods in Fredericton

2. Crime frequency by neighborhood
3. Crime type frequency and statistics. The mean crime count in the City of Fredericton is 22.
4. Crime type count by neighborhood.

   Theft from motor vehicles is most prevalent in the same area as the most frequent crimes. It's interesting to note this area is mostly residential and most do not have garages. It would be interesting to further examine if surveillance is a deterrent for motor vehicle crimes in the downtown core compared to low surveillance in the Platt neighborhood.

5. Motor Vehicle crimes less than $5000 analysis by neighborhood and resulting statistics. The most common crime is **Other Theft less than 5k** followed by **Motor Vehicle Theft less than 5k**. There is a mean of 6 motor vehicle thefts less than 5k by neighborhood in the City.
6. That population density and resulting visual correlation is not strongly correlated to crime frequency. Causation for crime is not able to be determined given lack of open data specificity by individual and environment.

7. Using k-means, we were able to determine the top 10 most common venues within a 1 km radius of the centroid of the highest crime neighborhood. **The most common venues in the highest crime neighborhood are coffee shops followed by Pubs and Bars**.

While, it is not valid, consistent, reliable or sufficient to assume a higher concentration of the combination of coffee shops, bars and clubs predicts the amount of crime occurrence in the City of Fredericton, this may be a part of the model needed to be able to in the future.

8. We were able to determine the top 10 most common venues by location of interest.

9. Statistically, we determined there are no coffee shops within the Knowledge Park clusters.

2.2 Discussion and Recommendations

The City of Fredericton Open Data enables us to gain an understanding of the crime volume by type by area but not specific enough to understand the distribution properties. Valuable questions such as, "are these crimes occurring more often in a specific area and at a certain time by a specific demographic of people?" cannot be answered nor explored due to what is reasonably assumed to be personal and private information with associated legal risks.

There is value to the city to explore the detailed crime data using data science to predict frequency, location, timing and conditions to best allocated resources for the benefit of its citizens and its police force. However, human behavior is complex requiring thick profile data by individual and the conditions surrounding the event(s). To be sufficient for reliable future prediction it would need to demonstrate validity, currency, reliability and sufficiency.

A note of caution is the possibility neighborhoods names could change. The crime dataset did not mention which specific neighborhood naming dataset it was using but we assumed the neighborhood data provided aligned with the neighborhoods used in the crime data. It may be beneficial for the City to note and timestamp neighborhood naming in the future or simply reference with neighborhood naming file it used for the crime dataset.

Errors exist in the current open data. An error was found in the naming of the neighborhood "Platt". The neighborhood data stated "Plat" while the crime data stated "Platt". Given the crime dataset was simplest to manipulate it was modified to "Plat". The true name of the neighborhood is "Platt".

Theft from motor vehicles is most prevalent in the same area as the most frequent crimes. It is interesting to note this area is mostly residential and most do not have garages. It would be interesting to further examine if surveillance is a deterrent for motor vehicle crimes in the downtown core compared to low surveillance in the Platt neighborhood.

It would be interesting to further study the Census data and if this captures the population that is renting or more temporary/transient pollution, given the City is a University hub.

Given the findings of the top 10 most frequent venues by locations of interest, the Knowledge Park does not have Coffee Shops in the top 10 most common venues as determined from the Foursquare dataset. Given this area has the greatest concentration of stores and shops as venues, it would be safe to assume a coffee shop would be beneficial to the business community and the citizens of Fredericton.

2.3 Conclusion

Using a combination of datasets from the City of Fredericton Open Data project and Foursquare venue data we were able to analyses, discover and describe neighborhoods, crime, population density and statistically describe quantitatively venues by locations of interest.

While overall, the City of Fredericton Open Data is interesting, it misses the details required for true valued quantitative analysis and predictive analytics which would be most valued by investors and developers to make appropriate investments and to minimize risk.

The Open Data project is a great start and empowers the need for a "Citizens Like Me" model to be developed where citizens of digital Fredericton are able to share their data as they wish for detailed analysis that enables the creation of valued services.

3 APPENDIX: Analysis

3.0.1 Load Libraries

[74]:

```
import numpy as np # library to handle data in a vectorized manner

import pandas as pd # library for data analysis
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)

import json # library to handle JSON files
```

```
!conda install -c conda-forge geopy --yes # uncomment this line if you haven't␣
↪completed the Foursquare API lab
from geopy.geocoders import Nominatim # convert an address into latitude and␣
↪longitude values

import requests # library to handle requests
from pandas.io.json import json_normalize # tranform JSON file into a pandas␣
↪dataframe

# Matplotlib and associated plotting modules
import matplotlib.cm as cm
import matplotlib.colors as colors

# import k-means from clustering stage
from sklearn.cluster import KMeans

# for webscraping import Beautiful Soup
from bs4 import BeautifulSoup

import xml

!conda install -c conda-forge folium=0.5.0 --yes
import folium # map rendering library

print('Libraries imported.')
```

Solving environment: done

All requested packages already installed.

Solving environment: done # All requested packages already

installed.

Libraries imported.

[75]:
```
r = requests.get('https://opendata.arcgis.com/datasets/
↪823d86e17a6d47808c6e4f1c2dd97928_0.geojson')
fredericton_geo = r.json()
```

[76]:
```
neighborhoods_data = fredericton_geo['features']
```

[77]:
```
neighborhoods_data[0]
```

[77]:
{'type': 'Feature',
'properties': {'FID': 1,

'OBJECTID': 1,

'Neighbourh': 'Fredericton South',

'Shape_Leng': 40412.2767429,

'Shape_Area': 32431889.0002},
'geometry': {'type': 'Polygon',

'coordinates': [[[-66.6193489311946, 45.8688925859664],

[-66.5986068312843, 45.8934317575498],

[-66.5998465063764, 45.8962889533894],

[-66.6005561754508, 45.8987959122414],

[-66.6007627879662, 45.9004150599189],

[-66.6005112596866, 45.9020341603803],

[-66.5993703992758, 45.9049409211054],

[-66.5983912356161, 45.9066536507875],

[-66.5950405196063, 45.9110977503182],

[-66.5924713378938, 45.9137165396725],

[-66.5975198697905, 45.9151915074375],

[-66.6016161874861, 45.9165914405789],

[-66.6063862416448, 45.9184662957134],

[-66.6102310310608, 45.9201848572716],

[-66.6193938469588, 45.9264149777787],

[-66.6194297795702, 45.9243466803461],

[-66.6206694546623, 45.9221345790227],

[-66.6241459348118, 45.9181100781124],

[-66.6249634017204, 45.9177976046497], [-66.6258796833102,
45.917910095299],

[-66.6292124330143, 45.9200348758374],

[-66.632733828928, 45.9225720071846],

[-66.6356353872957, 45.924409167803],

[-66.6362731911474, 45.9249840491044],

[-66.6381955858555, 45.9258900999313],

[-66.6400281490351, 45.9272147820915],

[-66.6469721261813, 45.9309512150791],

[-66.6492628301558, 45.9324257247173],

[-66.6501521622871, 45.9331254782868],

[-66.6504306400252, 45.9337564984884],

[-66.6505653873178, 45.9347436246005],

[-66.6503587748024, 45.9357182382069],

[-66.6520745569951, 45.9352246860213],

[-66.6532513500173, 45.9350872403269],

[-66.6541855979128, 45.9351122304785],

[-66.6557756159657, 45.9353808738969],

[-66.6597461695215, 45.9365616400027],

[-66.6692323789218, 45.9408659130747],

[-66.6702205257343, 45.9411720097543],

[-66.6705888350008, 45.9415718069541],

[-66.6717027459531, 45.9418654061867],

[-66.6805601346545, 45.9456570693391],

[-66.6808206460869, 45.945613344883],

[-66.690998558256, 45.9498794400526],

[-66.6932353633134, 45.9503791076107],

[-66.6956697977334, 45.9504478115476],

[-66.6955530167465, 45.9498607024316],

[-66.695014027576, 45.9498607024316],

[-66.6956248819692, 45.948261735435],

[-66.699766115429, 45.9452510552052],

[-66.6993978061625, 45.9450511702315],

[-66.6996762839006, 45.9448512845371],

[-66.6992271262585, 45.9446139193389],

[-66.7022364824603, 45.9407722096716],

[-66.7041049782513, 45.9393666396225],

[-66.7046080348104, 45.9387919073835],

[-66.7061441539463, 45.9390980155132],

[-66.7051919397451, 45.9388543785676],

[-66.7056949963042, 45.937405028971],

[-66.706611277894, 45.9362430230541],

[-66.7074107784969, 45.9356745059121],

[-66.7087133356588, 45.9350435075345],

[-66.7110938711618, 45.9342063302882],

[-66.7122526978783, 45.9309262230525],

[-66.7096026677901, 45.9293891917718],

[-66.6746402369322, 45.9061285859908],

[-66.6193489311946, 45.8688925859664]],

[[-66.6934150263703, 45.938648223393],

[-66.7001973067654, 45.9422339647247],

[-66.6939180829294, 45.9467626619838],

[-66.6912141539242, 45.9449262417569],

[-66.6899475293736, 45.9445014828376],

[-66.6890312477838, 45.9444702504357],

[-66.6889683657139, 45.9443827996167],

[-66.6899565125264, 45.9418404190785],

[-66.6934150263703, 45.938648223393]],

[[-66.6550120479742, 45.9291455121693], [-66.6557756159657, 45.9292704762017],

[-66.6599797314954, 45.9309387190672],

[-66.6629172224744, 45.9322757763752], [-66.6631867170597, 45.932475707408],

[-66.6631238349898, 45.9327880982037],

[-66.6619290756619, 45.9341813397283],

[-66.6616146653125, 45.9340751297235], [-66.6601863440107, 45.934818595486],

[-66.6591442982811, 45.9350997354041],

[-66.6586053091106, 45.9351059829416],

[-66.6564673187345, 45.9348748235837],

[-66.6542933957469, 45.9340501391045],

[-66.6529908385849, 45.9333129107794], [-66.652308118969, 45.9324569639043],

[-66.652191337982, 45.9319696305845],

[-66.6522721863576, 45.9313573339335],

[-66.6520476075366, 45.9305825815444],

[-66.6521284559121, 45.9301264722544],

[-66.6524428662616, 45.9296016295261],

[-66.6531166027247, 45.9293392062996],

[-66.6540508506202, 45.9291580085852],

[-66.6550120479742, 45.9291455121693]],

[[-66.6318085641854, 45.8878357293373], [-66.6328775593735, 45.8879357750148],

[-66.6341801165354, 45.8882108996987],

[-66.6351502970423, 45.8885422980769],

[-66.6362462416889, 45.8890987927924],

[-66.6370098096804, 45.8896365239624],

[-66.6381596532441, 45.8909183040123],

[-66.6385818614276, 45.8918186586532],

[-66.6387435581788, 45.8925689430378],

[-66.6385908445805, 45.8940757335582],

[-66.6327517952337, 45.900733882662],

[-66.62923039932, 45.9050971942525],

[-66.6276673307256, 45.9064848805016], [-66.626454605092, 45.9071974626627],

[-66.6253856099039, 45.9076662617274],

[-66.6230230407067, 45.9082913209882],

[-66.6205077579111, 45.9084913384651],

[-66.6180014582685, 45.9082413165064],

[-66.6181092561025, 45.9082100636823],

[-66.6170312777616, 45.9076037554142],

[-66.6161239793246, 45.9068661756028],

[-66.6150909167479, 45.9054972515047],
[-66.6147944727041, 45.9047533927481],
[-66.6146417591058, 45.9037907372083],
[-66.6146956580229, 45.9030155998367],
[-66.614974135761, 45.9020654166814],
[-66.617345688111, 45.8989772091164],
[-66.6203819937714, 45.8954199312614],
[-66.6263468072579, 45.8892363524244],
[-66.6281254715205, 45.8883672199348],
[-66.6291315846387, 45.8880795903605],
[-66.6304521081064, 45.8878732464875],
[-66.6318085641854, 45.8878357293373]]]}}

[78]:
```
g = requests.get('https://opendata.arcgis.com/datasets/
↪6179d35eacb144a5b5fdcc869f86dfb5_0.geojson')
demog_geo = g.json()
```

[79]:
```
demog_data = demog_geo['features']
demog_data[0]
```

[79]:
{'type': 'Feature',
'properties': {'FID': 1,
'OBJECTID': 501,
'DBUID': '1310024304',
'DAUID': '13100243',
'CDUID': '1310',
'CTUID': '3200002.00',
'CTNAME': '0002.00',
'DBuid_1': '1310024304',
'DBpop2011': 60,
'DBtdwell20': 25,
'DBurdwell2': 22,
'Shape_Leng': 0.00746165241824,
'Shape_Area': 2.81310751889e-06,
'CTIDLINK': 3200002,
'Shape__Area': 2.81310897700361e-06,

'Shape__Length': 0.00746165464503067},
'geometry': {'type': 'Polygon',
'coordinates': [[[-66.634784212921, 45.9519239912381],
[-66.6351046935752, 45.9507605156138],
[-66.6378263667982, 45.9510868696778], [-66.636944377136, 45.9521037018384],
[-66.634784212921, 45.9519239912381]]]}}

[ ]:

[80]:
```
import os
os.listdir('.')
```

[80]: ['Capstone Project Course.ipynb',
'Fredericton_Census_Tract_Demographics.csv',
'.DS_Store',
'Fredericton_Census_Tract_Demographics.xlsx',
'Crime_by_neighbourhood_2017.xlsx',
'Capstone Fredericton Crime and Police Station Location.ipynb',
'Boston_Neighborhoods (1).geojson',
'Fredericton Locations.xlsx',
'Week 3 Capstone - Segmenting and Clustering Neighborhoods in Toronto_Part 2.ipynb',
'Fredericton.jpg',
'Week 3 Capstone - Segmenting and Clustering Neighborhoods in Toronto_Part 2.pdf',
'Boston_Neighborhoods.geojson', '.ipynb_checkpoints',
'.git',
'Week 3 Capstone - Segmenting and Clustering Neighborhoods in Toronto.ipynb',
'Week 4 Capstone - Segmenting and Clustering Neighborhoods in Boston.ipynb',
'Week 3 Capstone - Segmenting and Clustering Neighborhoods in Toronto_Part 2.htm',
'Week 4 Capstone - Segmenting and Clustering Neighborhoods in Fredericton.ipynb',
'Week 4 Capstone - Segmenting and Clustering Neighborhoods in Fredericton -

Github submit.ipynb',

'Week 3 Capstone - Segmenting and Clustering Neighborhoods in Toronto_Part 2_files']

[81]:
```
opencrime = 'Crime_by_neighbourhood_2017.xlsx'
```

[82]:
```
workbook = pd.ExcelFile(opencrime)
print(workbook.sheet_names)
```

['Crime_by_neighbourhood_2017']

[83]:
```
crime_df = workbook.parse('Crime_by_neighbourhood_2017')
crime_df.head()
```

[83]:

	Neighborhood	From_Date	To_Date \
0	Fredericton South	2017-01-05T00:00:00.000Z	2017-01-26T00:00:00.000Z
1	Fredericton South	2017-03-04T00:00:00.000Z	2017-03-06T00:00:00.000Z
2	Fredericton South	2017-05-07T00:00:00.000Z	NaN
3	Fredericton South	2017-06-20T00:00:00.000Z	2017-06-21T00:00:00.000Z
4	Fredericton South	2017-07-09T00:00:00.000Z	2017-07-10T00:00:00.000Z

	Crime_Code	Crime_Type	Ward	City	FID
0	2120	B&E NON-RESIDNCE	7	Fredericton	1
1	2120	B&E NON-RESIDNCE	7	Fredericton	2
2	2120	B&E NON-RESIDNCE	12	Fredericton	3
3	2120	B&E NON-RESIDNCE	12	Fredericton	4
4	2120	B&E NON-RESIDNCE	7	Fredericton	5

[84]:
```
crime_df.drop(['From_Date', 'To_Date'], axis=1,inplace=True)
```

3.1 What is the crime count by neighborhood?

[128]:
```
crime_data = crime_df.groupby(['Neighbourhood']).size().to_frame(name='Count').
↪reset_index()
crime_data
```

[128]:

	Neighborhood	Count
0	Barkers Point	47
1	Brookside	54
2	Brookside Estates	9
3	Brookside Mini Home Park	5
4	College Hill	41
5	Colonial heights	9

6 Cotton Mill Creek 4

7 Diamond Street 1

8 Doak Road 1

9 Douglas 3

10 Downtown 127

11 Dun's Crossing 18

12 Forest Hill 12

13 Fredericton South 85

14 Fulton Heights 36

15 Garden Creek 13

16 Garden Place 4 17

26 Knowledge Park 18

27 Lian / Valcore 728 Lincoln 18

29 Lincoln Heights 18

30 Main Street 78 18

31 Marysville 3932 McKnight 433 McLeod Hill 334 Monteith / Talisman 1235 Montogomery / Prospect East 18

17 Gilridge Estates 3

18 Golf Club 7

19 Grasse Circle 1

20 Greenwood Minihome Park 2

21 Hanwell North8

22 Heron Springs 3

23 Highpoint Ridge 5

24 Kelly's Court Minihome Park 1

25 Knob Hill 4

36 Nashwaaksis 25

37 Nethervue Minihome Park 1

38 North Devon 113

39	Northbrook Heights	10
40	Plat	198
41	Poet's Hill	4
42	Prospect	81
43	Rail Side	3
44	Regiment Creek	1
45	Royal Road	7
46	Saint Mary's First Nation	25
47	Saint Thomas University	1
48	Sandyville	9
49	Serenity Lane	2
50	Shadowood Estates	5
51	Silverwood	12
52	Skyline Acrea	27
53	South Devon	68
54	Southwood Park	16
55	Springhill	1
56	Sunshine Gardens	10
57	The Hill	44
58	The Hugh John Flemming Forestry Center	3
59	University Of New Brunswick	15
60	Waterloo Row	9
61	Wesbett / Case	1
62	West Hills	5
63	Williams / Hawkins Area	17
64	Woodstock Road	41
65	Youngs Crossing	16

[153]:
```
crime_data .describe()
```

[153]: Count count 66.000000

mean 22.121212

```
std          34.879359
min           1.000000
25%           3.000000
50%           9.000000
75%          23.250000
max         198.000000
```

[86]:

```
crime_data .rename(index =str, columns={'Neighbourhood ':'Neighbourh ','Count':
 ↪'Crime_Count '}, inplace=True)
crime_data
```

[86]:

	Neighbourh	Crime_Count
0	Barkers Point	47
1	Brookside	54
2	Brookside Estates	9
3	Brookside Mini Home Park	5
4	College Hill	41
5	Colonial heights	9
6	Cotton Mill Creek	4
7	Diamond Street	1
8	Doak Road	1
9	Douglas	3
10	Downtown	127
11	Dun's Crossing	18
12	Forest Hill	12
13	Fredericton South	85
14	Fulton Heights	36
15	Garden Creek	13
16	Garden Place	4
17	Gilridge Estates	3
18	Golf Club	7
19	Grasse Circle	1
20	Greenwood Minihome Park	2

21	Hanwell North	8
22	Heron Springs	3
23	Highpoint Ridge	5
24	Kelly's Court Minihome Park	1
25	Knob Hill	4
26	Knowledge Park	1
27	Lian / Valcore	7
28	Lincoln	13
29	Lincoln Heights	14
30	Main Street	78
31	Marysville	39
32	McKnight	4
33	McLeod Hill	3
34	Monteith / Talisman	12
35	Montogomery / Prospect East	16
36	Nashwaaksis	25
37	Nethervue Minihome Park	1
38	North Devon	113
39	Northbrook Heights	10
40	Plat	198
41	Poet's Hill	4
42	Prospect	81
43	Rail Side	3
44	Regiment Creek	1
45	Royal Road	7
46	Saint Mary's First Nation	25

47	Saint Thomas University	1
48	Sandyville	9
49	Serenity Lane	2
50	Shadowood Estates	5
51	Silverwood	12
52	Skyline Acrea	27
53	South Devon	68
54	Southwood Park	16
55	Springhill	1
56	Sunshine Gardens	10
57	The Hill	44
58	The Hugh John Flemming Forestry Center	3
59	University Of New Brunswick	15
60	Waterloo Row	9
61	Wesbett / Case	1
62	West Hills	5
63	Williams / Hawkins Area	17
64	Woodstock Road	41
65	Youngs Crossing	16

[87]:
```
crime_data.rename({'Platt': 'Plat'},inplace=True)
crime_data.rename(index=str, columns={'Neighbourhood':'Neighbourh','Count':
↪'Crime_Count'}, inplace=True)
crime_data
```

[87]:

	Neighbourh	Crime_Count
0	Barkers Point	47
1	Brookside	54
2	Brookside Estates	9
3	Brookside Mini Home Park	5
4	College Hill	41
5	Colonial heights	9
6	Cotton Mill Creek	4

7	Diamond Street	1
8	Doak Road	1
9	Douglas	3
10	Downtown	127
11	Dun's Crossing	18
12	Forest Hill	12
13	Fredericton South	85
14	Fulton Heights	36
15	Garden Creek	13
16	Garden Place	4
17	Gilridge Estates	3
18	Golf Club	7
19	Grasse Circle	1
20	Greenwood Minihome Park	2
21	Hanwell North	8
22	Heron Springs	3
23	Highpoint Ridge	5
24	Kelly's Court Minihome Park	1
25	Knob Hill	4
26	Knowledge Park	1
27	Lian / Valcore	7
28	Lincoln	13
29	Lincoln Heights	14
30	Main Street	78
31	Marysville	39
32	McKnight	4
33	McLeod Hill	3
34	Monteith / Talisman	12
35	Montogomery / Prospect East	16
36	Nashwaaksis	25

37	Nethervue Minihome Park	1
38	North Devon	113
39	Northbrook Heights	10
40	Plat	198
41	Poet's Hill	4
42	Prospect	81
43	Rail Side	3
44	Regiment Creek	1
45	Royal Road	7
46	Saint Mary's First Nation	25
47	Saint Thomas University	1
48	Sandyville	9
49	Serenity Lane	2
50	Shadowood Estates	5
51	Silverwood	12
52	Skyline Acrea	27
53	South Devon	68
54	Southwood Park	16
55	Springhill	1
56	Sunshine Gardens	10
57	The Hill	44
58	The Hugh John Flemming Forestry Center	3
59	University Of New Brunswick	15
60	Waterloo Row	9
61	Wesbett / Case	1
62	West Hills	5
63	Williams / Hawkins Area	17
64	Woodstock Road	41

65 Youngs Crossing 16

[88]:
```
address = 'Fredericton, Canada '

geolocator = Nominatim()
location = geolocator .geocode(address)
latitude = location .latitude
longitude = location .longitude
print('The geograpical coordinate of Fredericton, New Brunswick is {}, {}.'.
 ↪format(latitude, longitude))
```

/anaconda3/lib/python3.7/site-packages/ipykernel_launcher.py:3:

DeprecationWarning: Using Nominatim with the default "geopy/1.18.1" `user_agent` is strongly discouraged, as it violates Nominatim's ToS https://operations.osmfoundation.org/policies/nominatim/ and may possibly cause 403 and 429 HTTP errors. Please specify a custom `user_agent` with

`Nominatim(user_agent="my-application")` or by overriding the default `user_agent`: `geopy.geocoders.options.default_user_agent = "my-application"`.

In geopy 2.0 this will become an exception.

This is separate from the ipykernel package so we can avoid doing imports until

The geograpical coordinate of Fredericton, New Brunswick is 45.966425,

-66.645813.

[89]:
```
world_geo = r'world_countries.json ' # geojson file

fredericton_1_map = folium .Map(location =[45.97 , -66.65 ], width=1000,␣
 ↪height=750,zoom_start =12)

fredericton_1_map
```

[89]: <folium.folium.Map at 0x1a 1f6b9278>

```
[90]: fredericton_geo = r.json()

threshold_scale = np.linspace(crime_data['Crime_Count'].
 ↪min(),crime_data['Crime_Count'].max(), 6,dtype=int)
threshold_scale = threshold_scale.tolist()
threshold_scale[-1] = threshold_scale[-1]+1

fredericton_1_map.choropleth(geo_data=fredericton_geo,␣
 ↪data=crime_data,columns=['Neighbourh', 'Crime_Count'],
    key_on='feature.properties.Neighbourh',␣
 ↪threshold_scale=threshold_scale,fill_color='YlOrRd', fill_opacity=0.7,
    line_opacity=0.1, legend_name='Fredericton Neighbourhoods')

fredericton_1_map
```

```
[90]: <folium.folium.Map at 0x1a1f6b9278>
```

3.2 Examine Crime Types

```
[131]: crimetype_data = crime_df.groupby(['Crime_Type']).size().to_frame(name='Count').
 ↪reset_index()
crimetype_data
```

[131]:

	Crime_Type	Count
0		4
1	ARSON	5
2	ARSON BY NEG	1
3	ARSON-DAM.PROP.	4
4	B&E NON-RESIDNCE	51
5	B&E OTHER	58
6	B&E RESIDENCE	151
7	B&E STEAL FIREAR	3
8	MISCHIEF OBS USE	1
9	MISCHIEF TO PROP	246
10	MISCHIEF-DATA	2
11	MOTOR VEH THEFT	40
12	THEFT BIKE<$5000	63