Maurizio Milazzo
08-Dec-2019

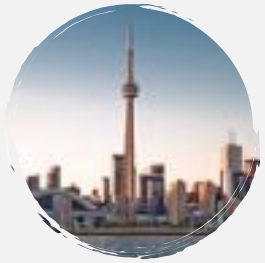
THE BATTLE OF NEIGHBORHOODS

PRESENTATION

CAPSTONE PROJECT

Indroduction

In this section, I show where to find the best opportunities to find an ideal location for opening a restaurant.



Toronto

Review of best Place in Toronto

Business Problem

Choice data from the OpenData catalog on Toronto City

Crime

We look on the Major Crime Indicators (MCI) Open Data

Division

We create a map with the 19 Division in Toronto

Hot Spot

We look for the best place with Hot Spot exists around Toronto

Success Criteria

We found the best place wich all the mentioned criteria

Toronto - Facs

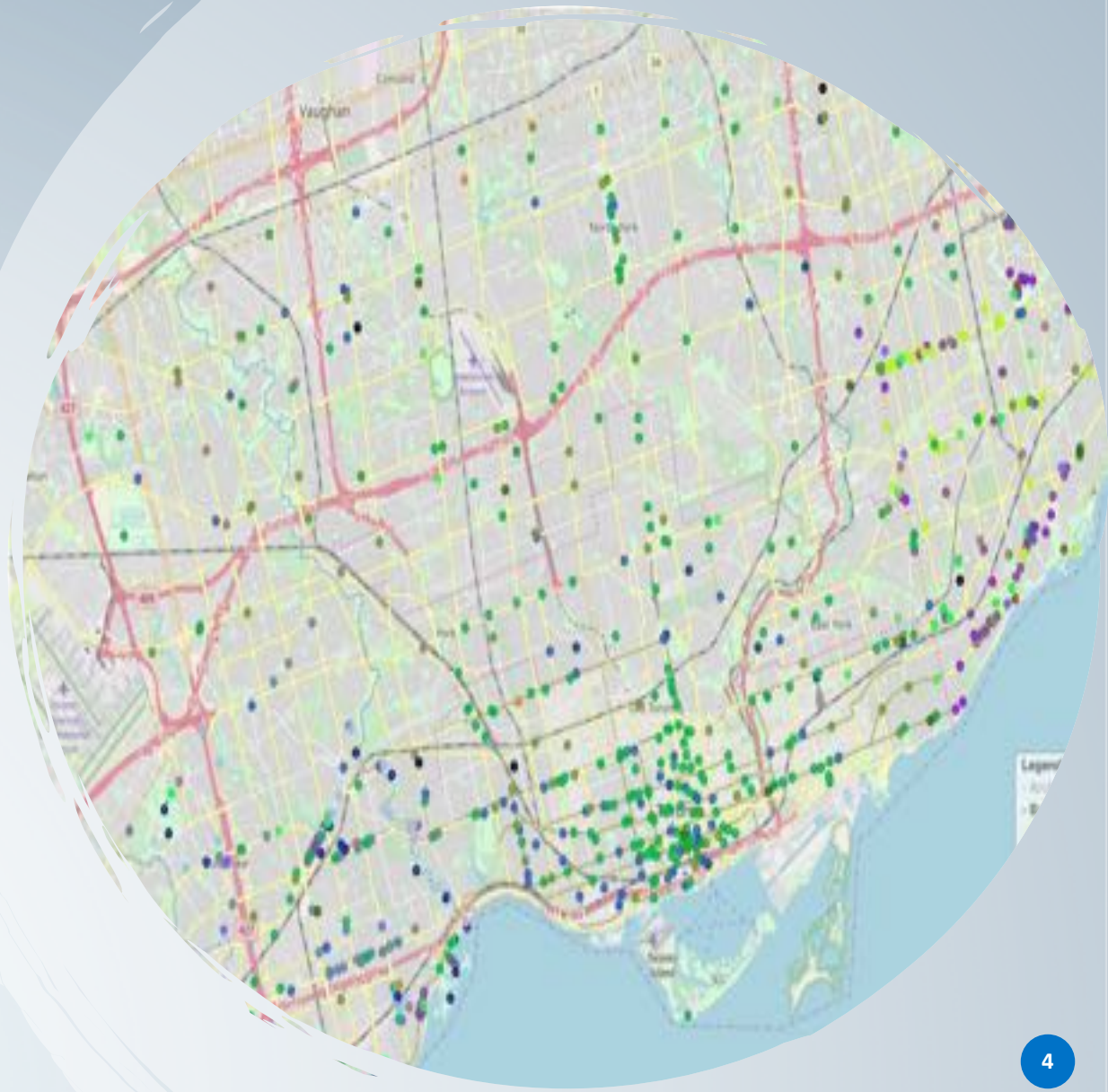
Toronto is the provincial capital of Ontario and the most populous city in Canada and the fourth largest populous city in North America, with a population of 2,731,571 as of 2016. Current to 2016, the Toronto census metropolitan area , of which the majority is within the Greater Toronto Area, held a population of 5,928,040, making it Canada's most populous CMA. The city is the anchor of the Golden Horseshoe, an urban agglomeration of 9,245,438 people (as of 2016) surrounding the western end of Lake Ontario. Toronto is an international centre of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world.

People have travelled through and inhabited the Toronto area, located on a broad sloping plateau interspersed with rivers, deep ravines, and urban forest, for more than 10,000 years. After the broadly disputed Toronto Purchase, when the Mississauga surrendered the area to the British Crown, the British established the town of York in 1793 and later designated it as the capital of Upper Canada. During the War of 1812, the town was the site of the Battle of York and suffered heavy damage by American troops. York was renamed and incorporated in 1834 as the city of Toronto. It was designated as the capital of the province of Ontario in 1867 during Canadian Confederation.

Decision Factors on Restaurant Location

- Toronto Population
- Demographic data of Toronto City
- What hot spots are there in Toronto like entertainment zones, parks etc.?
- What restaurants are there at this location?
- What restaurants names is there?
- Segmentation of the police into 19 divisions

and so on.....



Descrption of the used Data

Police Boundaries

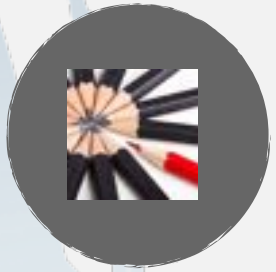
Data shows which geographical areas in the City are associated to each Toronto Police Division.

Dine Safe

Information to the public concerning the Toronto Public Health Food Premises Inspection and Disclosure system.

Cultural Hotspot

Cultural Hotspot spans a vast geographic area and many local neighbourhoods with unique features.



Distance

Calculate the distance between crime and location and hot spot

One-hot

Create artificial crime data. We will be assigned to the restaurant and hot spot.

Police Boundery

The police borders consist of 19 divisions.
We will divide all restaurant and hot spot in these divisions. Then we will visualize which places are highly medium or slightly endangered by adding the criminal data to them.

	AREA_LONG_CODE	LATITUDE	LONGITUDE	Shape__Area
9	0	43.633336	-79.368454	1.079671e+07
10	11	43.654151	-79.465607	3.570932e+07
14	12	43.695261	-79.493046	4.742078e+07
16	13	43.692443	-79.439954	3.559172e+07
0	14	43.649990	-79.417189	2.710943e+07
6	22	43.636142	-79.535267	1.300893e+08
13	23	43.714309	-79.577914	1.103716e+08
2	31	43.748497	-79.519215	8.177649e+07
3	32	43.756900	-79.432549	1.170648e+08
4	33	43.764411	-79.352777	1.012368e+08
17	41	43.730915	-79.276205	8.530402e+07
5	42	43.810977	-79.242229	1.625190e+08
1	43	43.766467	-79.199932	1.137953e+08
11	51	43.655470	-79.364360	1.645457e+07
12	52	43.653342	-79.388915	9.579489e+06
15	53	43.701340	-79.385936	5.988492e+07
18	54	43.700513	-79.325294	3.874682e+07
8	55	43.665352	-79.323255	3.933033e+07
7	58	43.622809	-79.381290	8.005564e+06

```
df_policeBoundery.groupby('AREA_LONG_CODE')['AREA_LONG_CODE','LATITUDE','LONGITUDE','Shape__Area']
.head(20).sort_values('AREA_LONG_CODE')
```

Restaurants Types

Here the data is grouped into the individual species. In Toronto has more than 7000 declared restaurants. Next, follow about 2700 Take Out's. After that, there are 2300 shopping possibilities, us. In total there are about 17478 possibilities to eat.
For each of these data records, the name of the Lokas and its position is available. The data is available under the City of Toronto's Open Data Portal.
https://open.toronto.ca/dataset/dinesafe/

TYPE	ID	NAME	ADDRESS	LATITUDE	LONGITUDE
Restaurant	7097	7097	7097	7097	7097
Food Take Out	2747	2747	2747	2747	2747
Food Store (Convenience/Variety)	2375	2375	2375	2375	2375
Child Care - Catered	615	615	615	615	615
Food Court Vendor	501	501	501	501	501
Supermarket	461	461	461	461	461
Bakery	401	401	401	401	401
Child Care - Food Preparation	364	364	364	364	364
Food Processing Plant	243	243	243	243	243
Student Nutrition Site	217	217	217	217	217
Food Depot	216	216	216	216	216
Cafeteria - Private Access	184	184	184	184	184
Cocktail Bar / Beverage Room	181	181	181	181	181
Banquet Facility	163	163	163	163	163
Food Caterer	160	160	160	160	160
Butcher Shop	154	154	154	154	154
Community Kitchen (Meal Program)	136	136	136	136	136
Cafeteria - Public Access	118	118	118	118	118
Private Club	92	92	92	92	92
Hot Dog Cart	77	77	77	77	77
Retirement Homes(Licensed)	75	75	75	75	75
Boarding / Lodging Home - Kitchen	72	72	72	72	72
Secondary School Food Services	69	69	69	69	69
Food Bank	68	68	68	68	68
Institutional Food Services	67	67	67	67	6

```
out_df.groupby('TYPE').count().sort_values(by=['ID'], ascending=False)
```

Colouring per type

For each type of restaurant a color is calculated to improve the visualization on the map.

1)
```
out_df_grpType['colorGrp'] = out_df_grpType
.apply(lambda x : list(np.random.choice(range(256), size=3)), axis=1)
```

2)
```
result = pd.merge(out_df, out_df_grpType, on='TYPE')
```

	TYPE	colorGrp
14	Bake Shop	[68, 16, 116]
7	Bakery	[62, 103, 10]
8	Banquet Facility	[246, 196, 130]
34	Bed & Breakfast	[222, 138, 23]
6	Boarding / Lodging Home - Kitchen	[57, 79, 128]

	ID	NAME	TYPE	ADDRESS	LATITUDE	LONGITUDE	colorGrp
0	9008018	'K' STORE	Food Store (Convenience/Variety)	99 CARLTON ST	43.66205	-79.37747	[120, 155, 57]
1	10356847	1000 VARIETY	Food Store (Convenience/Variety)	1000 PAPE AVE	43.68829	-79.34863	[120, 155, 57]
2	10383994	18 SMOKE SHOP & VARIETY	Food Store (Convenience/Variety)	222 SPADINA AVE	43.65141	-79.39758	[120, 155, 57]
3	10301731	3 JIT VARIETY	Food Store (Convenience/Variety)	1452 BLOOR ST W	43.65718	-79.44885	[120, 155, 57]
4	10330988	338 VEGETABLE & FRUIT	Food Store (Convenience/Variety)	631 GERRARD ST E	43.66561	-79.35074	[120, 155, 57]

Cultural Hot Spot

- Performance
- Exhibition / Visual Arts
- Screen Based
- Library
- Multipurpose
- Heritage

https://open.toronto.ca/dataset/cultural-spaces/

chs_df = gpd.read_file("./data/CULTURAL_HOTSPOT_WGS84.shp")

chs_df.head()

	TYPE
Public Art	101
Hot Eats	78
Business	77
Creative	42
Community	41
Heritage	36
Library	35
History	27
Nature	21
Architecture	9
Park	5

	PNT_OF_INT	DESCRPTION	SOCIAL_MED	WEBSITE	CATEGORY	LOCATION	X	Y	LONGITUDE	LATITUDE
0	21 Points in Equilibrium (Sculpture)	This sculpture is by James Southerland. It is ...	None	None	Public Art	150 Borough Dr	324341.227	4847914.922	-79.257067	43.772936
1	Crucified Again (Sculpture)	Crucified Again shows the body of a tortured m...	None	None	Public Art	450 Scarborough Golf Club Road	327641.001	4845799.116	-79.216170	43.753806
2	A Tall Couple (Sculpture)	Louis Archambault (1915-2003) created A Tall C...	None	None	Public Art	-79.187369, 43.783078	329947.860	4849060.368	-79.187369	43.783078
3	Warden Underpass Mural (Mural)	The Warden Underpass Mural provides a visual h...	None	None	Public Art	-79.273286, 43.693461	323059.616	4839081.887	-79.273286	43.693461
4	Sustenance (Mural)	Sustenance, also known as the "Western Gateway...	None	None	Public Art	-79.255266, 43.700826	324509.747	4839904.258	-79.255266	43.700826

Crime Data

The Toronto Police Service is dedicated to delivering police services, in partnership with our communities, to keep Toronto the best and safest place to be.
Our Principles Actively Accountable and Trusted Transparent and Engaged Inclusive and Collaborative Affordable and Sustainable Resources

- Strategic Plan
- The Way Forward
- TPS Scorecard
- Information Strategy
- TPSB Monthly Statistical Briefings

mci_df = gpd.read_file("./data/MCI_2014_to_2018.shp")

Data Type		Open Data		Data Analytics		Maps		
Category	Data	Download	Currency	Year-to-Date	Historical	Interactive (Year-to-Date)	Interactive (Historical)	Static
Crime	Assault	↓↓	2014-2018					N/A
Crime	Auto Theft	↓↓	2014-2018					N/A
Crime	Break and Enter	↓↓	2014-2018					N/A
Crime	Robbery	↓↓	2014-2018					N/A
Crime	Sexual Assault	N/A	N/A		N/A		N/A	N/A
Crime	Theft Over	↓↓	2014-2018					N/A
Crime	Homicide	↓↓	2004-2018					
Crime	Shootings	N/A	N/A				N/A	
Crime	Stabbings	N/A	N/A	N/A	N/A	N/A	N/A	
Crime	Bike Thefts	↓↓	2014-2018	N/A	N/A	N/A	N/A	N/A
Crime	All Major Crime Indicators	↓↓	2014-2018					N/A
Crime	Arrests	N/A	N/A	N/A	N/A	N/A	N/A	N/A

ANALYTIC APPROACH

ANALYTIC Approach

Toronto City has a total of 19 division and 17000 dining options with around 470 hot spots.

For the years 2014 - 2018 167525 cases are registered.

- Part 1 - Clustering of restaurants

- Part 2 - Clustering from Hot Spot

- Only the criminal data from the year 2018 are consulted.

All data will be filtered by OpenData and used for this project.

All data is processed by the same ETL process.

• Load the data and explore data (CSV, Shape, Txt)

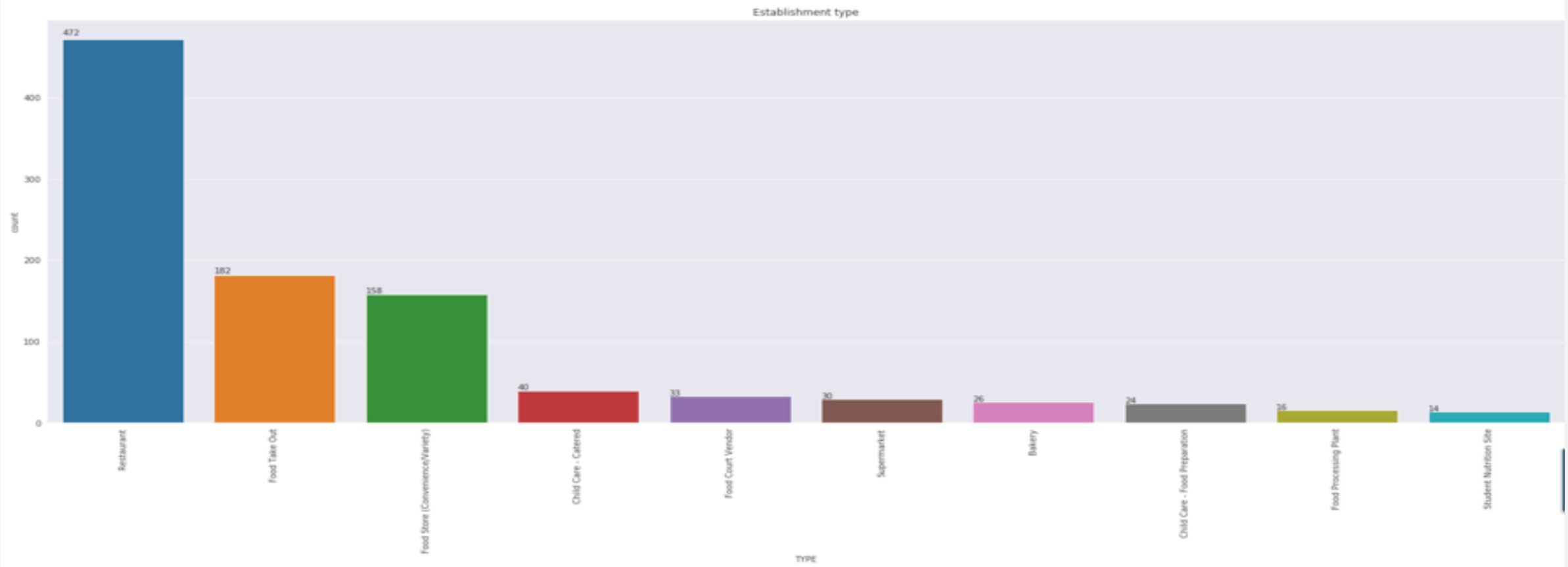
• Transform the data of nested python dictionaries into a pandas dataframe.

• Dataframe contains the geographical coordinates.

• Data will used from OpenData from City of Toronto.

• Folium libraries used to create a map of Toronto city .

Methodology 1
Establishment type (most Top 1000)

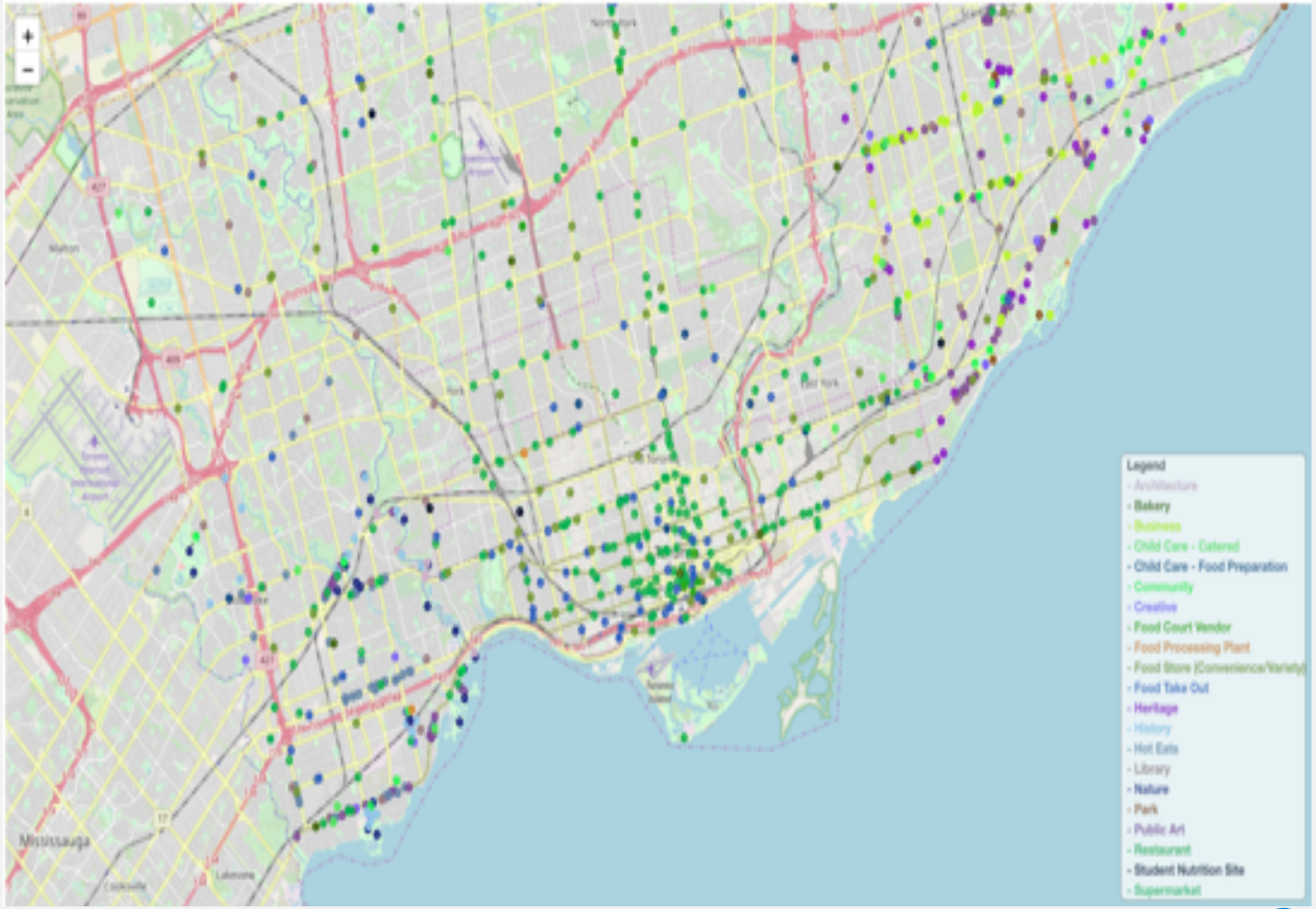


Catering:
Restaurant and other possibilities to be able to eat oneself
- We are using the data of restaurant (Top 1000) of all types
- There are a total of 472 restaurants which is by far the highest number in Toronto.
- And the lowest are the student nutrition site

Methodology 1
Visualize establishment in Toronto

```
# map of Toronto, CA using latitude and longitude values

map_markets
= folium.Map(location=[latitude, longitude], zoom_start=12)

# add markers to map
for lat, lng, name, types, colo in zip(slHLocToronto['LATITUDE'],
slHLocToronto['LONGITUDE'], slHLocToronto['NAME'’],
slHLocToronto['TYPE'], slHLocToronto['colorGrp']):
        label = '{}, {}, {};{}'.format(name, types, lat, lng)
        colo = colo.replace('[','').replace(']','').split(',')
        label = folium.Popup(label, parse_html=True)
        folium.CircleMarker(
            [lat, lng],
            radius=3,
            popup=label,
            color=f'#{int(colo[0]):02x}',
            fill=True,
            fill_color=f'#{int(colo[0]):02x}',
            fill_opacity=0.7,
            parse_html = False).add_to(map_markets)

map_markets
```

The geograpical coordinate of Toronto, CA are 43.653963, -79.387207.



Methodology 1
Most used Word for location names

0 'K',STORE
1 1000,VARIETY
2 18,SMOKE,SHOP,&,VARIETY
3 3,JIT,VARIETY
4 338,VEGETABLE,&,FRUIT
Name: NAME, dtype: object

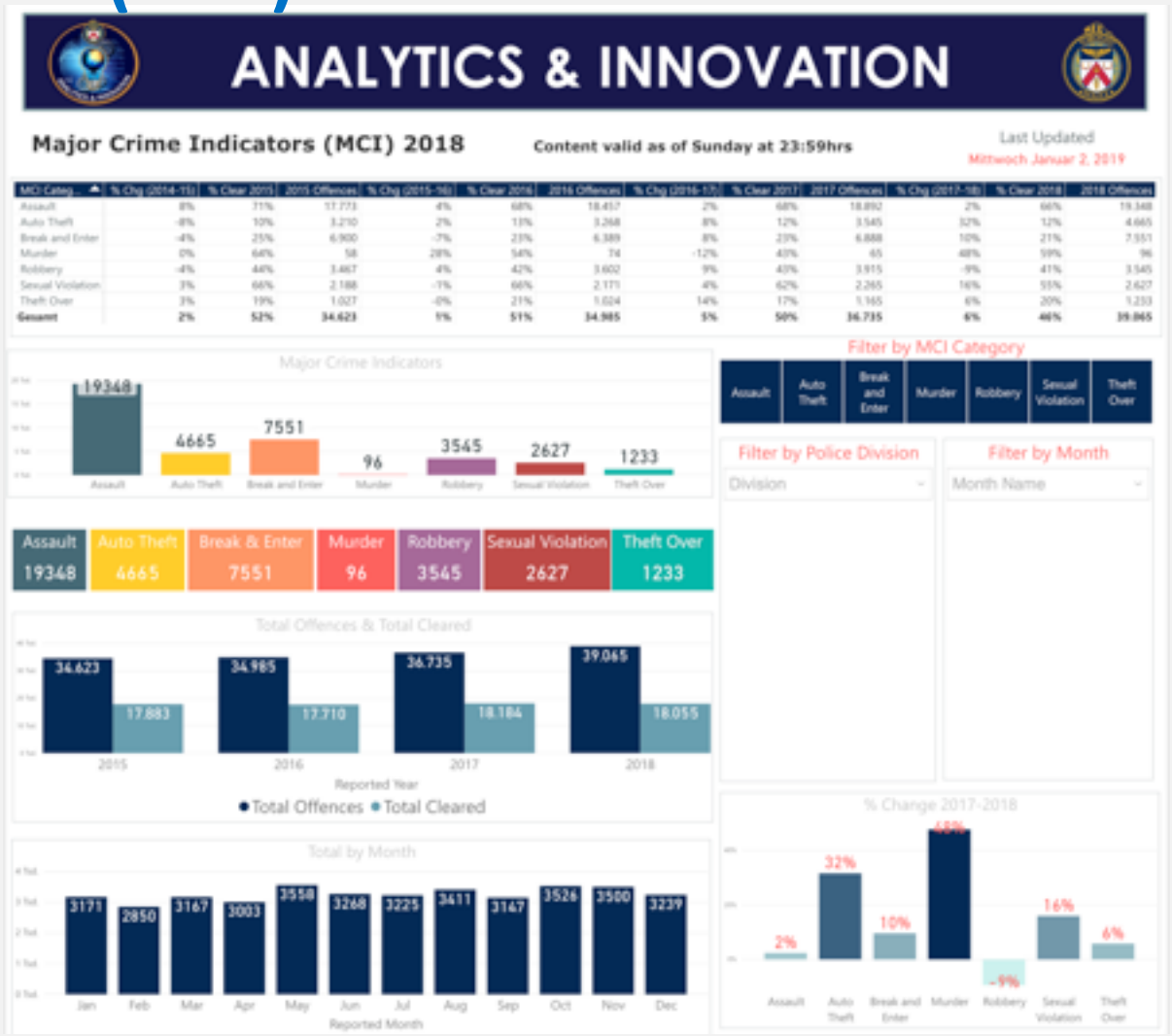
```
# Start with one review:
text = " ".join(review for review in dfTorontoNAME)

wordcloud = WordCloud(stopwords=stopwords,
max_font_size=60, max_words=200,
background_color="white").generate(text)
plt.figure(figsize=(50, 20))
plt.imshow(wordcloud, interpolation="bilinear")
plt.axis("off")
plt.show()
```

Methodology 2
The Toronto Major Crime Indicators (MCI) Data

Major Crime Indicators (MCI) Dashboard.

The information contained in these dashboards refers to 2018 Year-to-Date as well as previous full Year End content.

This data are been aggregated. We load the realdata from the opendata Catalog. https://opendata.arcgis.com/datasets/98f7dde610b54b9081dfca80be453ac9_0.zip?outSR=%7B%22wkid%22%3A102100%2C%22latestWkid%22%3A3857%7D

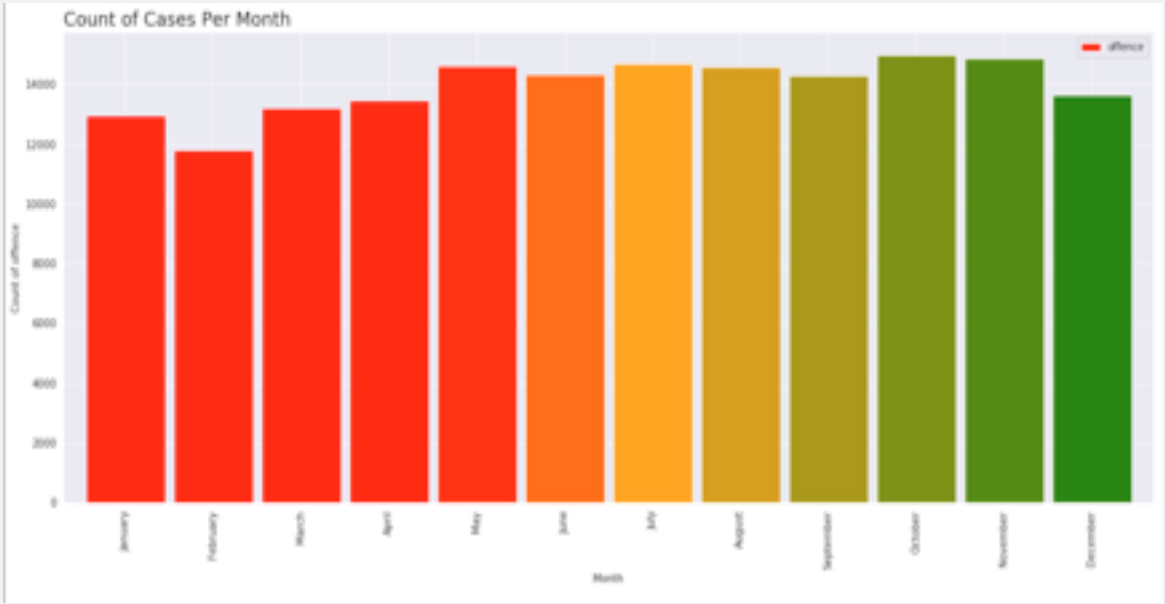


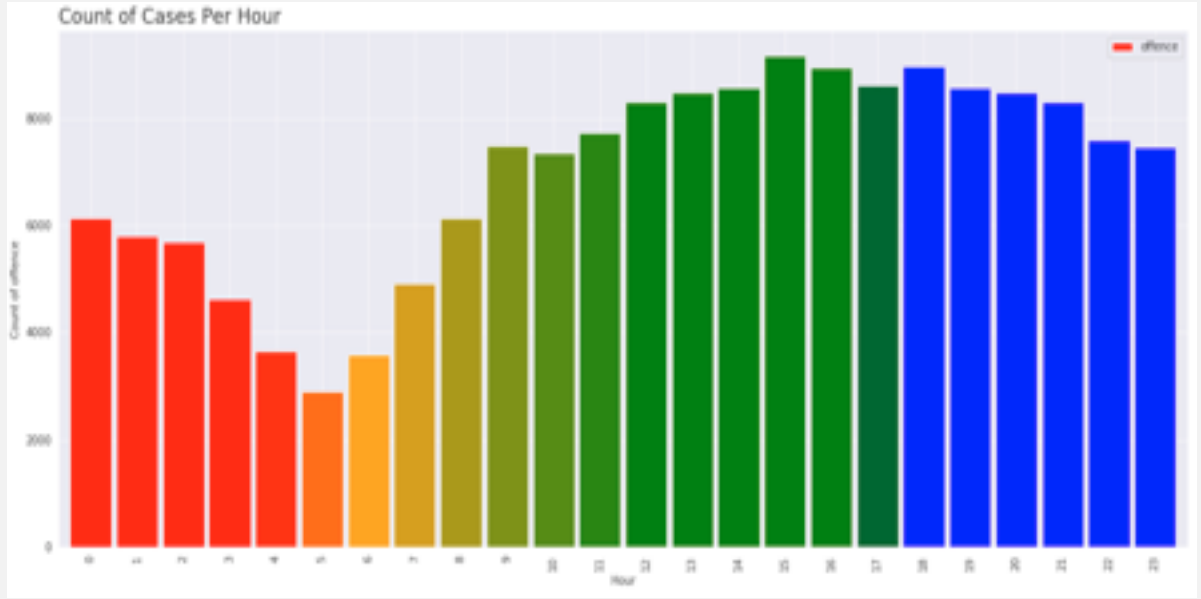
http://data.torontopolice.on.ca/pages/major-crime-indicators

Methodology 2
The Toronto Major Crime Indicators (MCI) Data

	reportedda	premisetyp	offence	reportedye	reportedmo	reported_1	reported_3	reportedho	MCI	Division	Neighbourh	Lat	Long	geometry
0	2014-06-20 13:20:00+00:00	Apartment	Theft Over	2014	June	20	Friday	13	Theft Over	D52	Bay Street Corridor (76)	43.659229	-79.385193	POINT (-79.38519 43.65923)
1	2014-07-02 02:58:00+00:00	Outside	Pointing A Firearm	2014	July	2	Wednesday	2	Assault	D32	Newtonbrook West (36)	43.777592	-79.425400	POINT (-79.42540 43.77759)
2	2014-07-02 02:58:00+00:00	Outside	Robbery With Weapon	2014	July	2	Wednesday	2	Robbery	D32	Newtonbrook West (36)	43.777592	-79.425400	POINT (-79.42540 43.77759)
3	2014-07-02 05:40:00+00:00	House	B&E	2014	July	2	Wednesday	5	Break and Enter	D42	Malvern (132)	43.801727	-79.210373	POINT (-79.21037 43.80173)
4	2014-07-02 20:57:00+00:00	Commercial	Assault	2014	July	2	Wednesday	20	Assault	D42	Milliken (130)	43.835884	-79.254334	POINT (-79.25433 43.83588)

```
dfMCI = pd.read_csv('./data/Toronto_MCI.csv', parse_dates=['reportedda'])
```






```
fig = dfMCI.groupby([dfMCI['reportedda'].dt.month], sort=True).count().plot(y = 'offence', kind='bar', figsize=(20,8), width=0.9, color=rvb(x/N))
```

Methodology 2
The Toronto Major Crime Indicators (MCI) Data

```
# Instantiate a feature group for the incidents in the dataframe
offence = folium.map.FeatureGroup()

for lat, lng, prem, in zip(df_top_crimes.Lat, ...)
    offence.add_child(
        folium.CircleMarker(
            [lat, lng], .....)

# add incidents to map
crime_map.add_child(offence)


# display world map
crime_map
```

The geograpical coordinate of Toronto, CA are 43.653963, -79.387207.

Methodology 2
The Toronto Major Crime Indicators (MCI) Data

```
mc = MarkerCluster()
crimedina_cluster
= folium.Map(location=[latitude, longitude], zoom_start=12)

for row in df_top_crimes.itertuples():
    mc.add_child(folium.Marker(location=[row.Lat, row.Long],
popup=row.offence))

# add Cluster
crimedina_cluster.add_child(mc)

for row in slDinaToronto.itertuples():
    mc.add_child(
      folium.Marker(
          location=[row.LATITUDE, row.LONGITUDE],
          popup=row.TYPE))

# add Cluster
crimedina_cluster.add_child(mc)

crimedina_cluster
```

The geograpical coordinate of Toronto, CA are 43.653963, -79.387207.

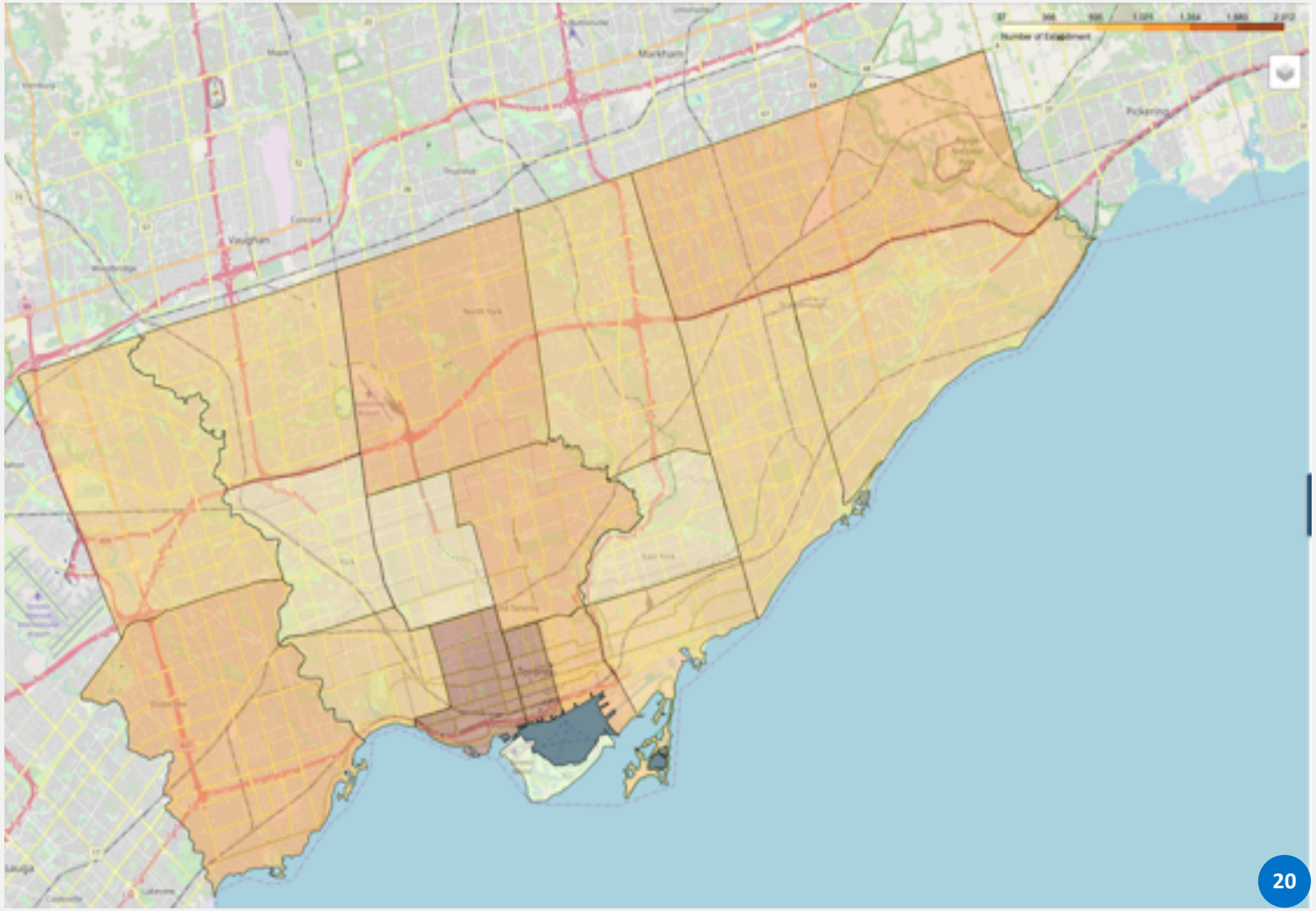
Methodology 4
Create with police boundery a map of Toronto Restaurants and Culture Hot Spot

```
# create a map
m = folium.Map(location=[43.653963, -79.387207])

## add chloropleth layer
folium.Choropleth(
    geo_data='./data/Police Boundaries Data.geojson',
    name='Number of Estabiliment',
    data=z,
    columns=['AREA_LONG_CODE','TYPE'],
    key_on='feature.properties.AREA_LONG_CODE',
    fill_color='YlOrBr',
    fill_opacity=0.7,
    line_opacity=0.1,
    legend_name='Number of Estabiliment'
).add_to(m)
```

Data

```
dfToronto = pd.read_csv('./data/Toronto_dinasafe.csv')
x = dfToronto[['AREA_LONG_CODE','TYPE']]
     .groupby(['AREA_LONG_CODE'])
z = x.count().reset_index()
```

The geograpical coordinate of Toronto, CA are 43.653963, -79.387207.



coursera

RESULT

Result

Generating crime data to calculate a prediction

For the clustering to find a good places for the new restaurant we create an artificial crime data record.

```
df_onhotCrime_dmy = dfMCI[['Division','LATITUDE', 'LONGITUDE']]
```

The crimes will be split between an act 0 and no act 1.
The newly created artificial record will be assigned to the restaurant and hot spot.

```
df_onhotCrime_dmy['art_crime'] = np.random.randint(0, 2,
df_onhotCrime_dmy.shape[0])
```

The next step is to assign a random date.

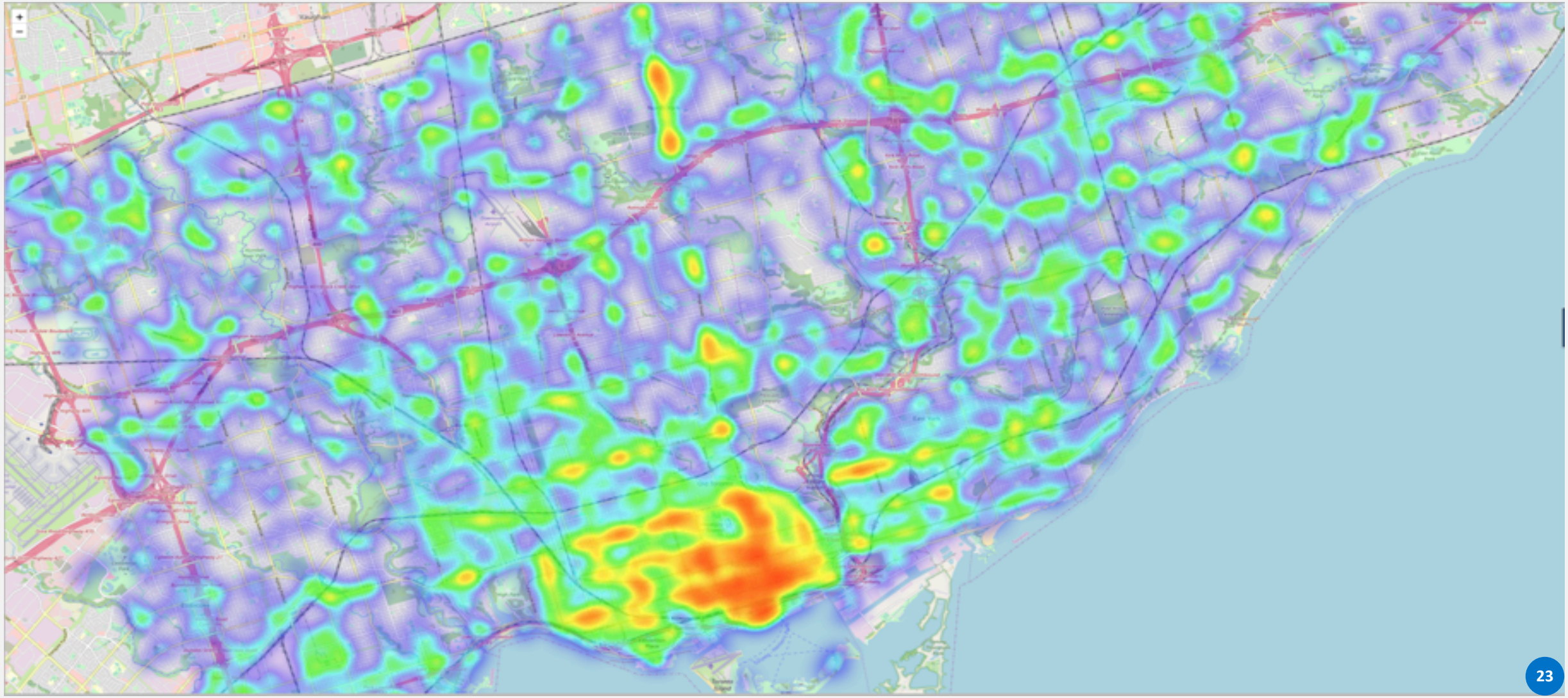
At the end a prediction should be made whether a crime will be committed in the desired place or not.

At the end we found 3000 bad place and 500 good places.

We look now for one place for every 19 Division.

Result
Headmap of places that are not good to open a restaurant

The geograpical coordinate of Toronto, CA are 43.653963, -79.387207.



Result
The new place for open a restaurant

```
# Get the Information for the first new places
new_place_name = new_Dina_places.Name[0]
new_place_latitude = new_Dina_places.Point[0][0]
new_place_longitude = new_Dina_places.Point[0][1]

# Create the Folium Map
heatmap = folium.Map(
    location=[new_place_latitude, new_place_longitude])

# List comprehension to make out list of lists of Crimes
heat_data = [[row['LATITUDE'], ....)]

# Plot the crimes on the map
HeatMap(heat_data,
                ...
      overlay=True).add_to(heatmap)

# Add the Venue to the Map
folium.Marker(
   location=[new_place_latitude, new_place_longitude],  ...
).add_to(heatmap)

heatmap
```

1000 Pape Ave, Toronto, East York, Ontario, Kanada, 43.6882889, -79.3485799

DISCUSION

- Opportunities to explore real estate prices in the various districts.
- In District 54 - East York it has the smallest density of Restaurants and kitchens.
- The risk can certainly be taken with great menus.
- It also shows that the crime rate is not too high.
- Unfortunately, it also has not many hot spots that can attract visitors.
- Therefore a local cuisine is more recommended than a foreign one.

CONCLUSION

- Analysis based on large amounts of data
- Execute process with more information districts with prices.
- Concentration of the restaurant business
- Consider further possibilities of different districts that have a good number of restaurants but no hot spot places.
- Look out for the neighbourhood or the restaurant type nearby which makes it better than us.