

# PR\_Assign\_W5

December 4, 2020

## 1 Final Project Report: Analysis of COVID cases in Ontario, CA

Owned by Asghar Sadeghi, PhD, Dec. 2020

### 1.1 Week 5 - Part A: Statement of the Problem

In this study, we plan to analyze the COVID-19 cases in Ontario province. Currently, we observe the second wave of pandemic and the government plans to impose additional restrictions on different counties. Given the venues in the neighborhood of a medical center and the number of confirmed cases, we try to cluster similar cities/counties in terms of COVID behaviour and consult the government to make decisions accordingly. In the first section we analyze the data, then we try to map it, later the neighborhood venues will be extracted using FourSquare, and as the final step the data will be clustered to 5 different categories to restrict the interactions. The first audience of this study would be the Government of Ontario and the Mayors and City councils, but the citizens also could be the second audience.

#### 1.1.1 Import libraries

Libraries imported. - Confirmed

### 1.2 Week 5 - Part B: Data Section

The data is live COVID data that is being posted on the "<https://data.ontario.ca>". It includes all reported cases since the start of pandemic. The columns for each case are the exact episode date, age grouped by decade, gender, outcome (recovered, active, and death), Reporting PHU ID, name, postal code, latitude and longitude. Furthermore, using the Foursquare website, the venues near each public health unit (PHU) is extracted to find a relation between venues and the number of confirmed cases. Data URL: <https://data.ontario.ca/dataset/f4112442-bdc8-45d2-be3c-12efae72fb27/resource/455fd63b-603d-4608-8216-7d8647f43350/download/conposcovidloc.csv>

### 1.3 Week 5 - Part C: Basic analysis of the received data

In the first step, we drop the cases that include 'NaN' values in their cells. Data shows handful of cases at the start of pandemic have less information, but as times go on, the information package is more complete.

After cleaning, number of confirmed cases to be analyzed in the rest of study are: 121733

A summary of data is shown in Table 1. Dataframe includes more columns, but for the sake of brevity, we just show some of the main columns.

Table 1: Daily COVID-19 Cases, Ontario Province

```
[5]: Accurate_Episode_Date Age_Group Client_Gender \
0      2020-11-03      60s      MALE
1      2020-11-23      30s      FEMALE
2      2020-11-26      40s      FEMALE
3      2020-11-19      70s      FEMALE
4      2020-11-20      30s      MALE

      Reporting_PHU
0      Peel Public Health
1  York Region Public Health Services
2      Durham Region Health Department
3      Peel Public Health
4      Peel Public Health
```

Now, data is grouped on a weekly basis for a better representation on plot. The onjective here is to see the confirmed cases per week and observe the trend of confirmed cases.

Table 2: Confirmed cases on a weekly basis

```
[6]: Row_ID
Accurate_Episode_Date
2019-12-23/2019-12-29      2
2019-12-30/2020-01-05      2
2020-01-06/2020-01-12      1
2020-01-13/2020-01-19      1
2020-01-20/2020-01-26      3
```

Confirmed cases are divided based on the gender to see the significance of each gender exposure to virus.

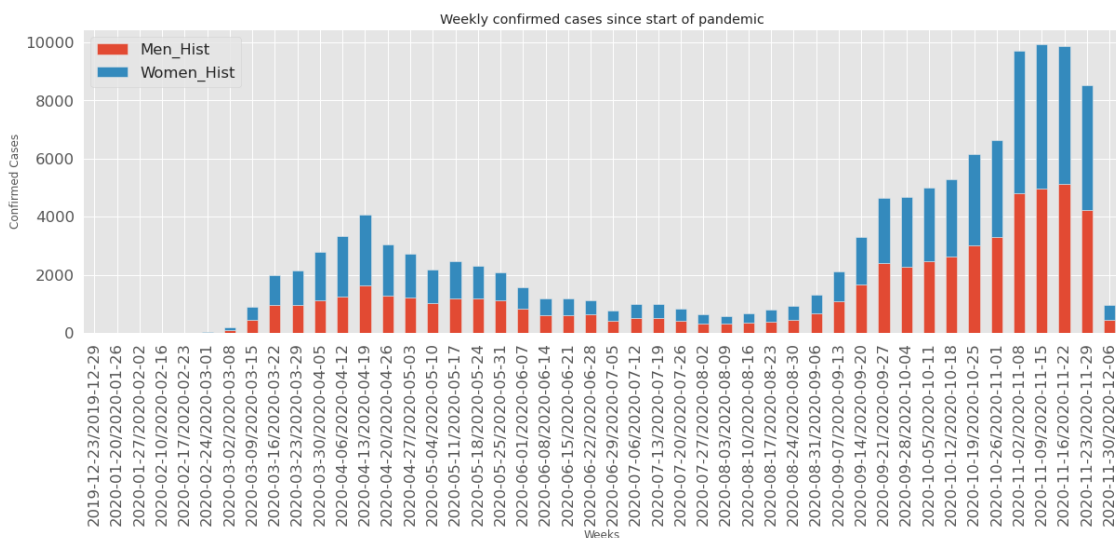


Figure 1: Weekly visualization of confirmed cases based on gender

Figure 1 shows that we are exactly in the middle of second wave. The pandemic trend graphically seems to be identical for each gender. In the next plot, we like to see how the trend of fatality has changed in the second wave compared to the first one.

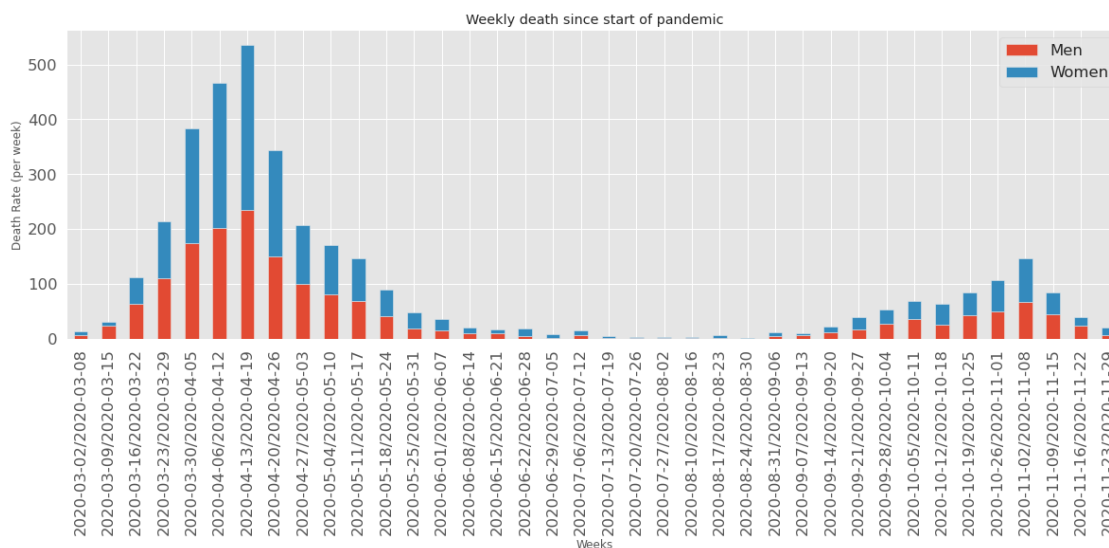


Figure 2: Fatality rate on a weekly basis since the start of pandemic

COVID-19 virus was an unprecedented pandemic. Figure 2 shows how the health knowledge of both health personnel and people has improved over time. The fatality rate of the first wave was too high while in the second wave the fatality rate dropped by 80% while affection rate increased by 150%, from 4,000 cases per day to approximately 10,000 cases per day. Cumulative fatality shows that since the start of pandemic we had more than 3500 cases died.

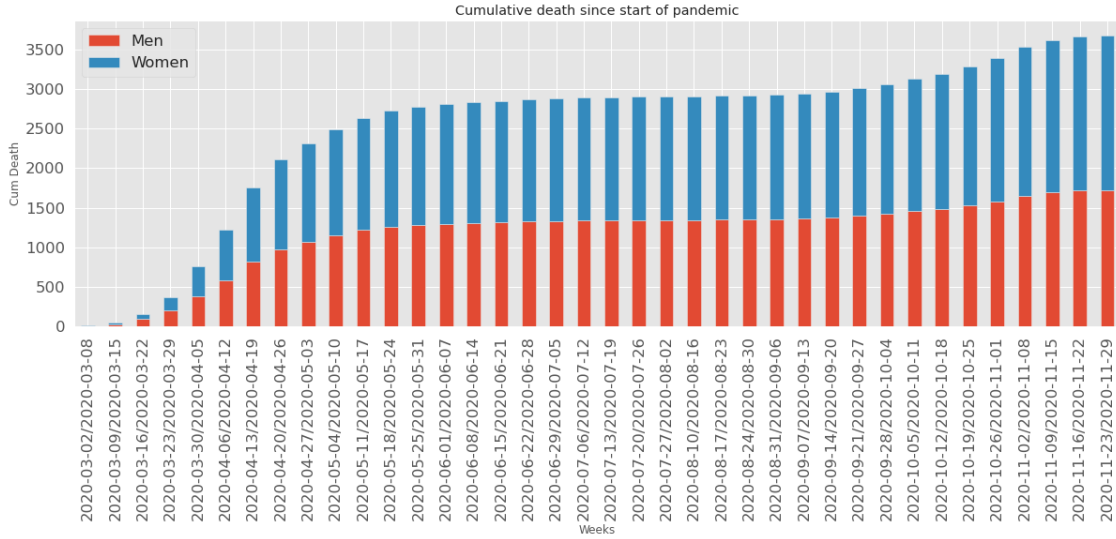


Figure 3: Cumulative fatality in Ontario Province

Another important measure is the age group of dead people since the start of pandemic. This information is highly demanded to classify and focus on the most vulnerable groups in the society. Figure 4 presents the distribution of fatality by the age group. It shows that seniors must be carefully looked after because the most cases are in the range of 60s and older. Another observation is that as age goes higher, the ratio of female to male fatality increases.

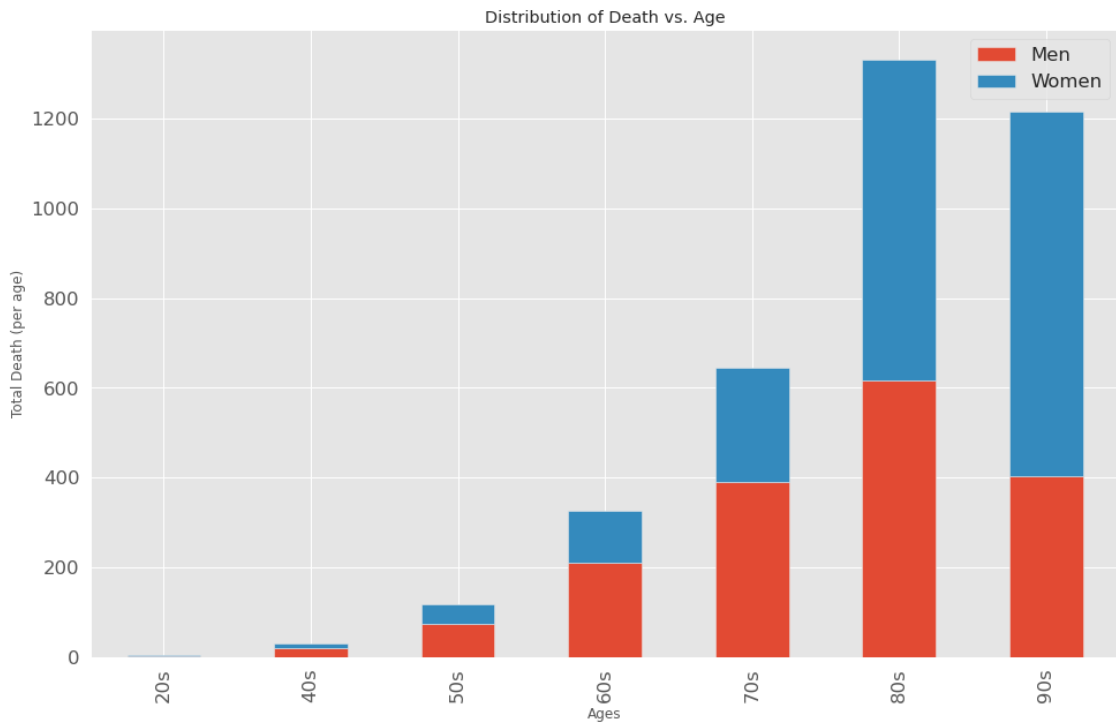


Figure 4: Fatality versus the age group

Figure 4 shows the fatality rate decreases for the age range of 90s compared to 80s while this plot does not account for the number of confirmed cases for each

[11]: Text(0.5, 1.0, 'Fataillity Percentage vs. Age')

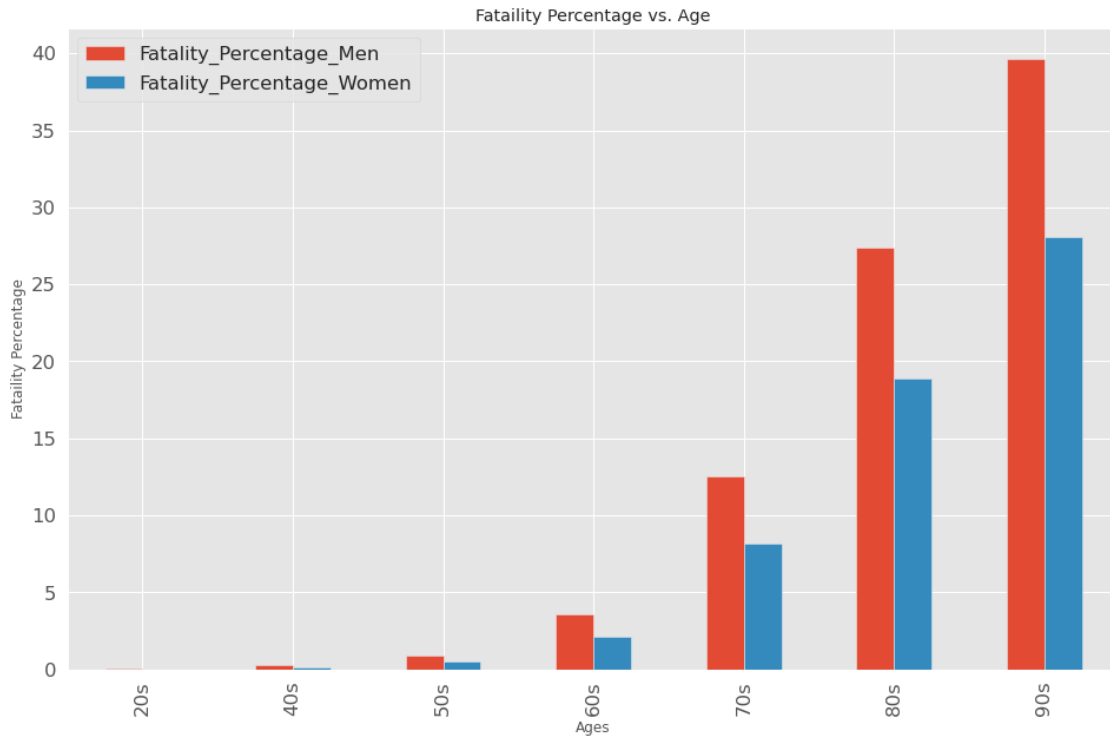


Figure 5: Fatality percentage vs. the age group

Figure 5 shows that the fatality percentage monotonically increases vs age group and the percentage is lower for women.

#### 1.4 Week 5 - Part D: Map representation of confirmed cases

Number of cases are summed up for each PHU.

Table 3: Number of total confirmed cases by each PHU

```
[12]:
```

	Reporting_PHU	Row_ID
0	Algoma Public Health Unit	61
1	Brant County Health Unit	613
2	Chatham-Kent Health Unit	515
3	Durham Region Health Department	4692
4	Eastern Ontario Health Unit	872

The location of each PHU and the total percentage are added to each PHU row in Table 4 and are shown in Figure 6.

Table 4: Location and address of each PHU

[13]:

	Reporting_PHU	Row_ID	Reporting_PHU_Postal_Code	\
0	Algoma Public Health Unit	61	P6B 0A9	
1	Brant County Health Unit	613	N3R 1G7	
2	Chatham-Kent Health Unit	515	N7M 5L8	
3	Durham Region Health Department	4692	L1N 0B2	
4	Eastern Ontario Health Unit	872	K6J 5T1	

	Reporting_PHU_Latitude	Reporting_PHU_Longitude	Percentage_Cases
0	46.532373	-84.314836	0.050110
1	43.151811	-80.274374	0.503561
2	42.403861	-82.208561	0.423057
3	43.898605	-78.940341	3.854337
4	45.029152	-74.736298	0.716322

The geograpical coordinate of Sudbury is 46.49272, -80.991211.

[28]: <folium.folium.Map at 0x7efbecf75e10>

Figure 6: PHUs on map. The blue circles represents the percentage of the confirmed cases (cannot be shown)

## 1.5 Week 5 - Part E: Finding nearby venues

### 1.5.1 Define Foursquare Credentials and Version

The credentials of the app created in FourSquare is given here:

```
CLIENT_ID: 5CI5C01LEUXWI12VXC00100UHPD5RHARRWPV02DQULDLVNM
CLIENT_SECRET:0NCZVFFNU3WGX0VVX4UXFBSSKSUKVEYJWG1ZB22L03SQ04MK
```

### 1.5.2 Nearby Venues

Nearby venues are extraced for the following PHUs:

```
Algoma Public Health Unit
Brant County Health Unit
Chatham-Kent Health Unit
Durham Region Health Department
Eastern Ontario Health Unit
Grey Bruce Health Unit
Haldimand-Norfolk Health Unit
Haliburton, Kawartha, Pine Ridge District Health Unit
Halton Region Health Department
Hamilton Public Health Services
Hastings and Prince Edward Counties Health Unit
Huron Perth District Health Unit
```

Kingston, Frontenac and Lennox & Addington Public Health  
 Lambton Public Health  
 Leeds, Grenville and Lanark District Health Unit  
 Middlesex-London Health Unit  
 Niagara Region Public Health Department  
 North Bay Parry Sound District Health Unit  
 Northwestern Health Unit  
 Ottawa Public Health  
 Peel Public Health  
 Peterborough Public Health  
 Porcupine Health Unit  
 Region of Waterloo, Public Health  
 Renfrew County and District Health Unit  
 Simcoe Muskoka District Health Unit  
 Southwestern Public Health  
 Sudbury & District Health Unit  
 Thunder Bay District Health Unit  
 Timiskaming Health Unit  
 Toronto Public Health  
 Wellington-Dufferin-Guelph Public Health  
 Windsor-Essex County Health Unit  
 York Region Public Health Services

where, they are 2355 different venues

Venues are inserted into a dataframe where the top five cases are:

Table 5: List of Venues

[20]:

	Reporting_PHU	Neighborhood	Latitude	Neighborhood	Longitude	\
0	Algoma Public Health Unit		46.532373		-84.314836	
1	Algoma Public Health Unit		46.532373		-84.314836	
2	Algoma Public Health Unit		46.532373		-84.314836	
3	Algoma Public Health Unit		46.532373		-84.314836	
4	Algoma Public Health Unit		46.532373		-84.314836	

	Venue	Venue	Latitude	Venue	Longitude	Venue	Category
0	Shogun Sushi		46.530801		-84.319091	Sushi Restaurant	
1	Burger Don		46.522580		-84.319638	Burger Joint	
2	Fratellis		46.542842		-84.318774	Italian Restaurant	
3	YMCA		46.521494		-84.316275	Gym / Fitness Center	
4	North 82		46.527674		-84.319183	Steakhouse	

## 1.6 Week 5 - Part F: Neighborhood Venues vs Confirmed Cases in Ontario

In this section, number of confirmed cases is related to the number of venues in the surrounding counties. Figure 7 shows that there is a positive relation between confirmed cases and the number of venues in close proximity. The correlativity increases when we plot confirmed cases per venue.

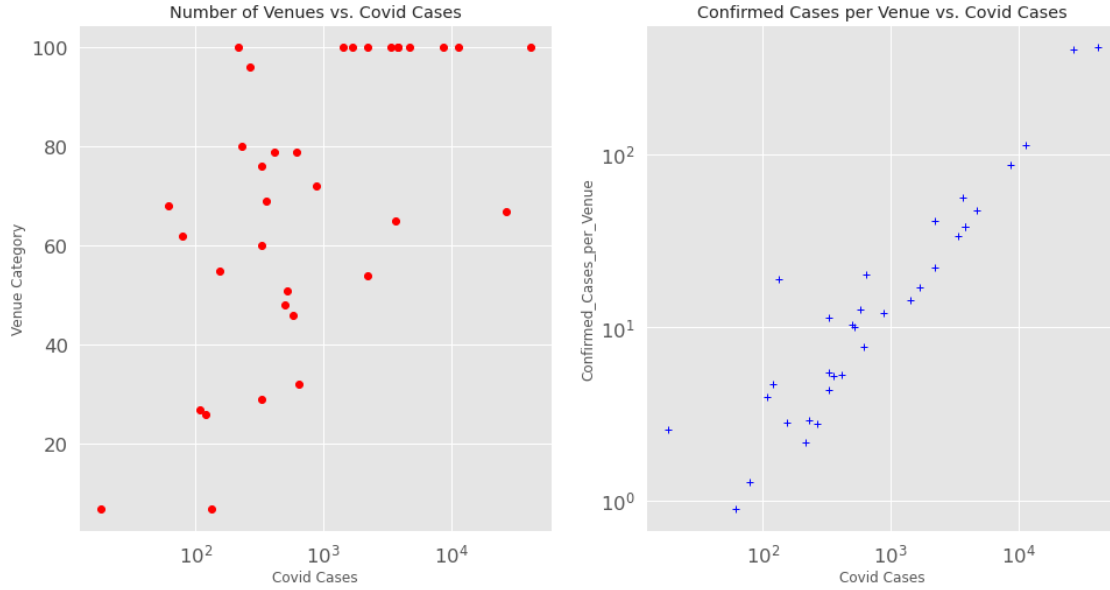


Figure 7: a) Number of venues vs. number of confirmed cases, and b) Confirmed cases per venue vs Confirmed cases in the close neighborhoud

### 1.7 Week 5 - Part G: Clustering the neighborhoods based on confirmed cases

The regions are clustered in 5 different zones to impose additional restrictions. "KMeans" function is used to cluster the venues based on the confirmed cases per venue.

[25]: <folium.folium.Map at 0x7efbeb9d4a58>

Figure 8: Ontario map including clusters (cannot be shown)

[26]: <seaborn.axisgrid.FacetGrid at 0x7efbeb9ad828>



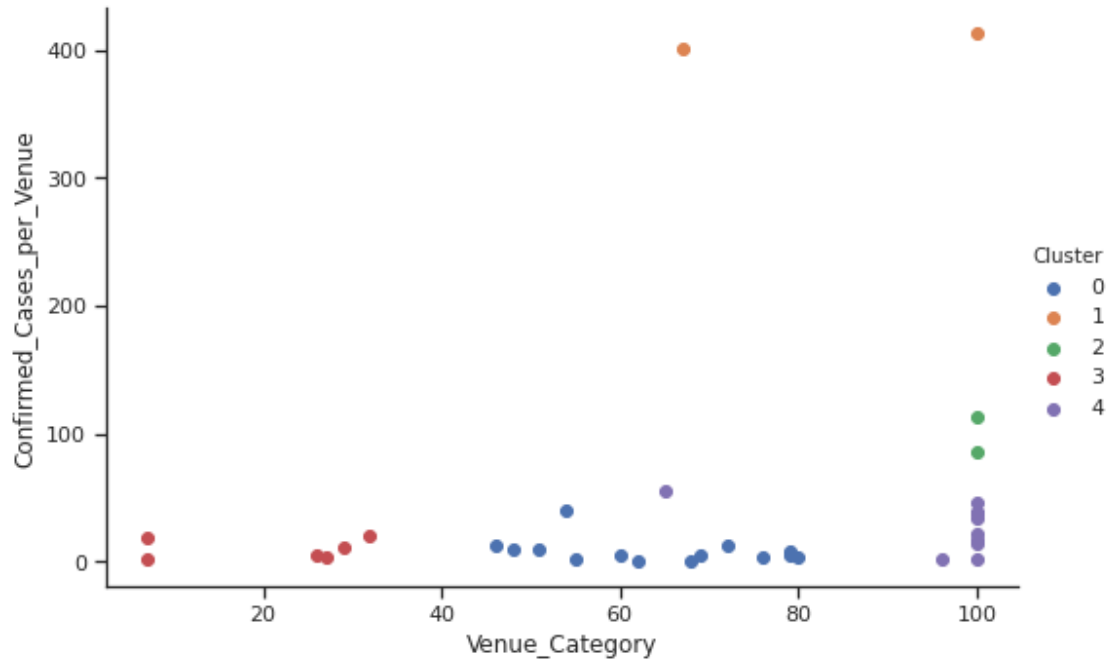


Figure 9: Confirmed cases per venue showing different clusters

## 1.8 Week 5 - Part H: Discussion and Conclusions

Results show that we can manage cities for the next phase of restrictions based on findings of this analysis. Cluster map reveals that Toronto Downtown and Brampton region are associated with the highest level of risk which implies imposing most restrictions. Northern Toronto (newmarket region) and Ontario could be dealt similarly as the second level of risks. Detroit, London, Hamilton and some other same-size cities are in the third category. Smaller cities such as Sudbury are in the next level of decision making. And finally, the lowest risk zones can be defined as Timmins and Simcoe.