# The Battle of Neighborhoods

Cousera - IBM Applied Data Science

# Business Problem

**Urbanization is a Megatrend worldwide...**

**...more and more confusing cities**

**Where are good places for new businesses?**

# Criteria for a good location

**For businesses like restaurants, gyms or stores...**

**....the location crucial**

**Influence has:**

- **Crimes within a neighborhood**
- **Population density of a neighborhood**
- **Number of competitors within the neighborhood**

1. Crimes within a neighborhood
2. Number of competitors within the d

# Data Science as Solution

**Collect data about the neighborhoods of your city...**

**...evaluate the location with machine learning algorithms**

**Assume you are a founder of pizza place in Toronto**

# Collecting the Data

- **Dataset for geographic data on the neighborhoods**

- **Dataset for crime statistics on the neighborhoods**

- **Dataset about the neighborhood population**

- **Dataset about the number of competitors**
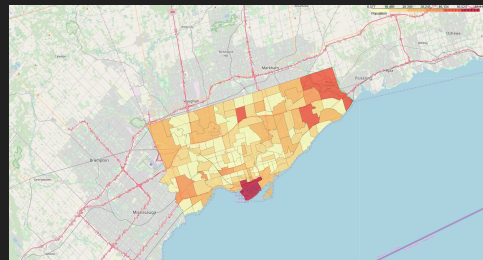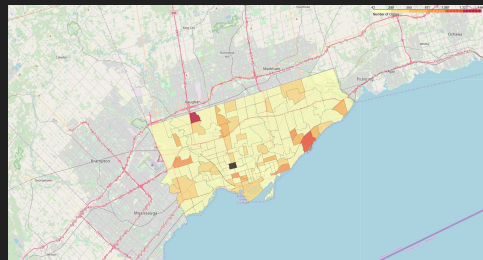




| Venue Category | Venue Id |
|---|---|
| Chinese Restaurant | 4bf58dd8d48988d145941735 |
| Sandwich Place | 4bf58dd8d48988d1c5941735 |
| Fast Food Restaurant | 4bf58dd8d48988d16e941735 |
| Pizza Place | 4bf58dd8d48988d1ca941735 |
| Fried Chicken Joint | 4d4ae6fc7a7b7dea34424761 |
| Wings Joint | 4bf58dd8d48988d14c941735 |

# Methodology

**Bring all data in one dataset together…**

**… find clusters in the data**

| AREA_SHORT_CODE | AREA_NAME | LONGITUDE | LATITUDE | Number of Crimes | Area m^2 | Population | Persons per m^2 | Sum Venue |
|---|---|---|---|---|---|---|---|---|
| 129 | Agincourt North (129) | -79.266712 | 43.805441 | 513.0 | 7264393.0 | 29113 | 0.004008 | 1 |
| 128 | Agincourt South-Malvern West (128) | -79.265612 | 43.788658 | 154.0 | 7875876.0 | 23757 | 0.003016 | 1 |
| 20 | Alderwood (20) | -79.541611 | 43.604937 | 223.0 | 4980675.0 | 12054 | 0.002420 | 2 |
| 95 | Annex (95) | -79.404001 | 43.671585 | 154.0 | 2791395.0 | 30526 | 0.010936 | 1 |
| 42 | Banbury-Don Mills (42) | -79.349718 | 43.737657 | 81.0 | 10045354.0 | 27695 | 0.002757 | 2 |

**K-means**

| Labels | LONGITUDE | LATITUDE | Number of Crimes | Area m^2 | Population | Persons per m^2 | Sum Venue |
|---|---|---|---|---|---|---|---|
| 0 | -79.356925 | 43.729088 | 810.142857 | 4.016187e+06 | 17517.000000 | 0.005073 | 0.857143 |
| 1 | -79.410995 | 43.718262 | 187.433735 | 5.914220e+06 | 20042.506024 | 0.004284 | 0.361446 |
| 2 | -79.381349 | 43.669528 | 148.000000 | 1.483088e+06 | 26825.250000 | 0.018447 | 9.500000 |
| 3 | -79.392426 | 43.686745 | 220.675676 | 2.365391e+06 | 17645.837838 | 0.008498 | 2.108108 |
| 4 | -79.435646 | 43.663522 | 303.500000 | 1.103498e+06 | 31304.000000 | 0.034273 | 1.500000 |

# Results

*Black Cluster: High crime,*
*          low population density, a few competitors*
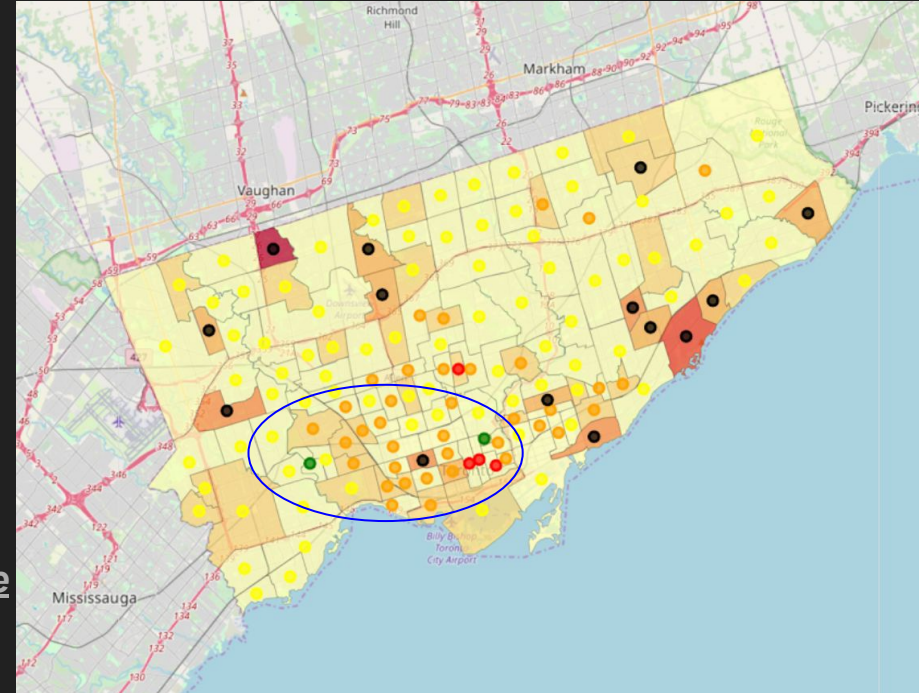*Yellow Cluster: Low crime,*
*          low population density, a few competitors*
*Red Cluster: Low crime,*
*          medium population density, a lot of competitors*
*Orange Cluster: Low crime,*
*          low population density, some competitors*
*Green Cluster: Low/medium crime,*
*          high population density, some competitors*

*The green cluster seem to be*
*          a optimal location for a new pizza place*

# Discussion and Conclusion

**K-means algorithm has identified two neighborhoods which seem to a very suitable location for a new pizza place**

**Data science can be a valuable contribution to find good location in big cities**

**For safe estimation of location more research is necessary - but the results of this first approach are promising**