



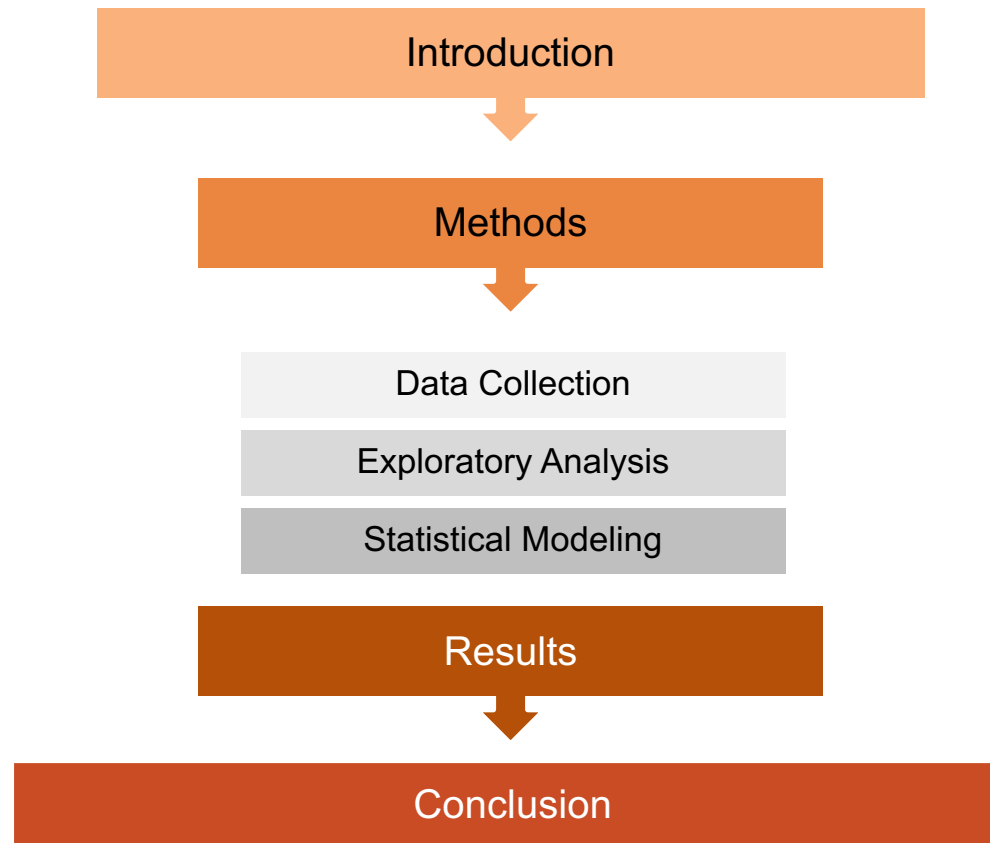
# Capstone Project

## The Battle of Neighborhoods

Finding Ideal Neighborhoods in Toronto

Haoyuan Gao

# Table of Contents



# INTRODUCTION

## BUSINESS PROBLEM

- Providing ideal neighborhood recommendation and lifestyle guide
- Match users' preferences to the physical and social elements of the neighborhoods

## TARGET GROUP

- New residents to the city
- Residents who are looking for new residential places
- People who are interested in exploratory

# METHODS

## Data Collection

- Toronto Open Data
- FourSquare API

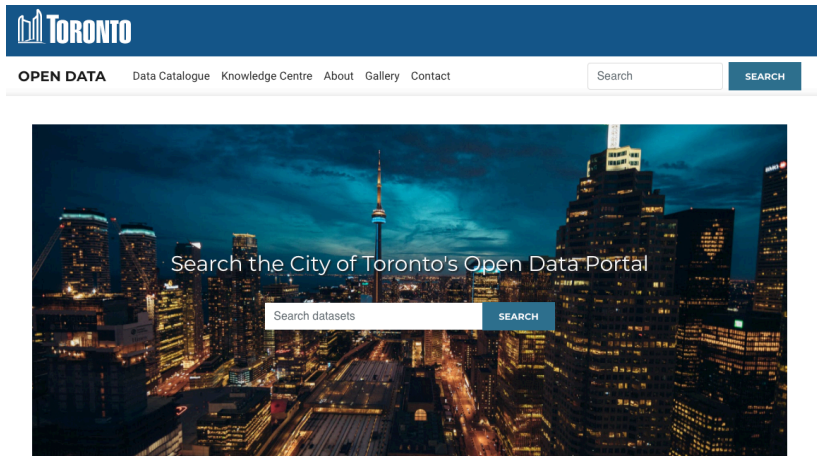
## Exploratory Analysis

- Folium

## Statistic Modeling

- K-Means Clustering
- Correlation Matrix

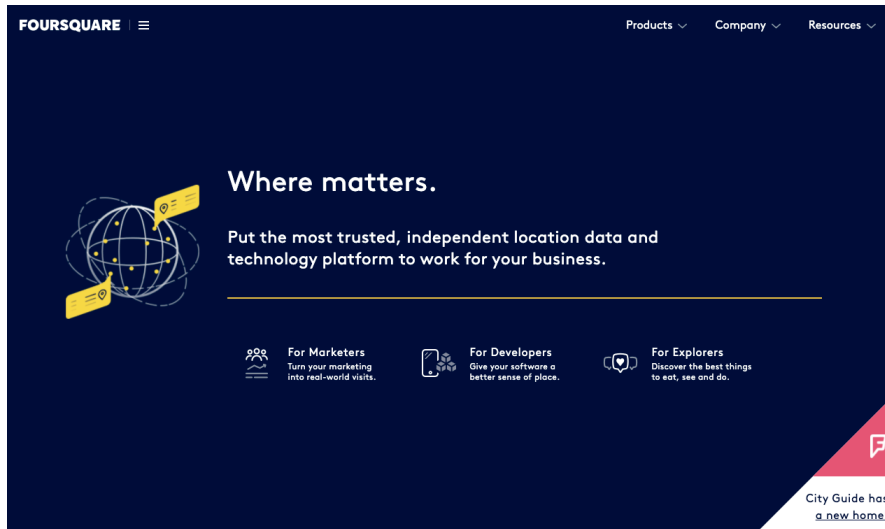
# Data Collection



## Toronto Open Data

- Economics
- Demographics
- Crime
- The boundaries of Toronto neighborhoods & geolocation
  - Latitude and longitude

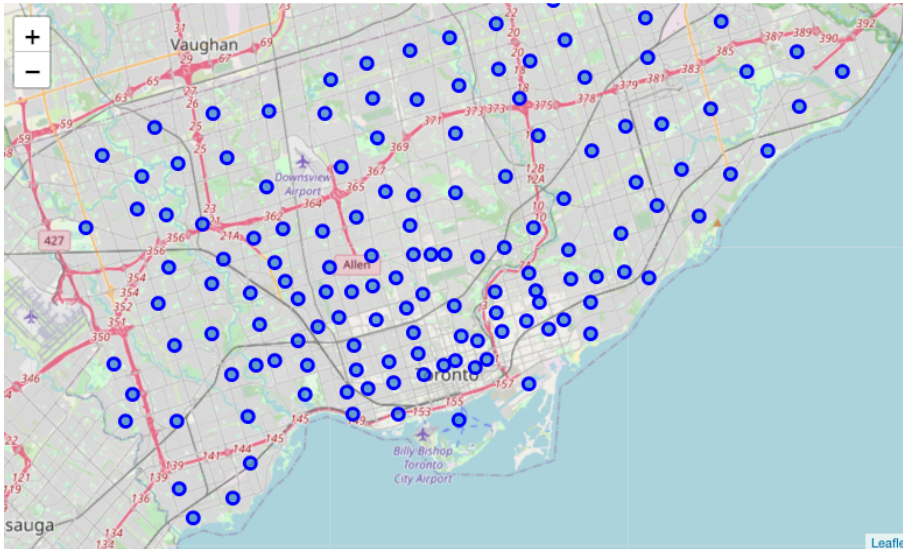
# Data Collection



## FourSquare API

- Venue information
  - Venue location
  - Geographical coordinates
  - Venue category

# Exploratory Analysis



## Folium

- Folium is used for visualizing geospatial data in this project. It takes the location parameters, latitude and longitude, and generates an interactive map around the given coordinates. All cluster visualization is generated using Folium and its Marker class.

# Statistical Modeling

## K-Means Clustering

- Grouping observations within a dataset
  - Identify similarities
- Optimal Number of clusters
  - The Elbow Method

## Correlation Matrix

- Show the correlation coefficients between two variables
- Summarize data for further analysis



# Results

## K-Means Clustering

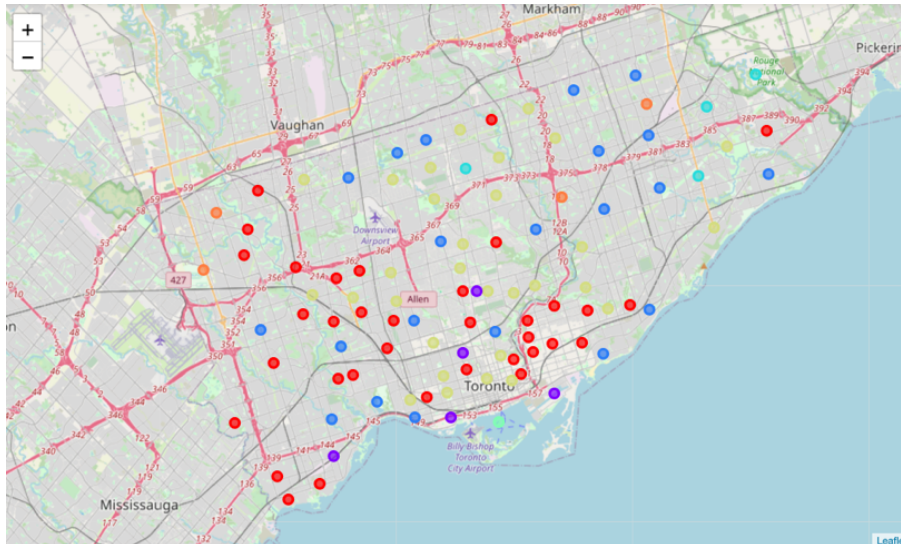
- The percentage of different age groups
- The type of venue components existed
- Suggests similar neighborhoods by components

## Correlation Matrix

- Indicate the coefficients between crime and economic factors

# K-Means Clustering

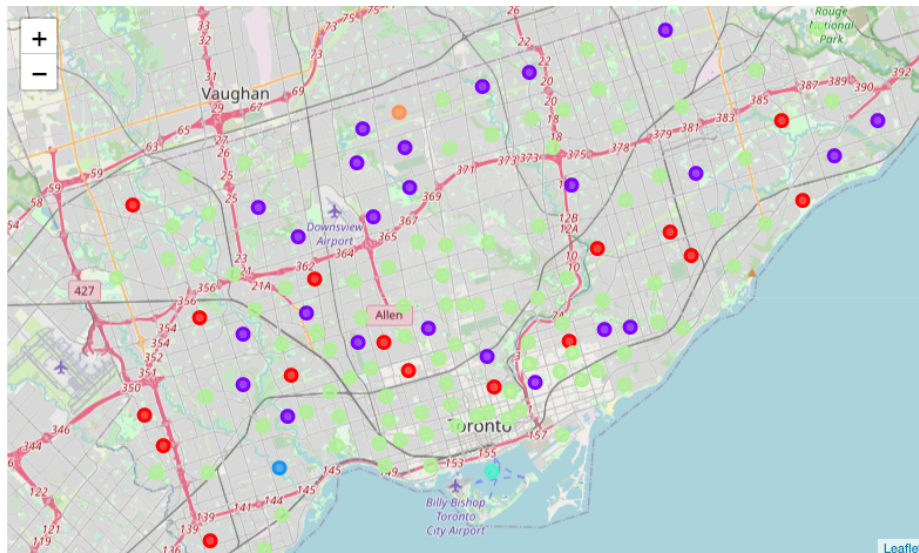
The K-Means clustering analysis could help to identify the components of similar neighborhoods.



## Age-group

- Identify the percentage of different age groups that live in the neighborhoods
- Suggest similar community population components with map location provided.

# K-Means Clustering

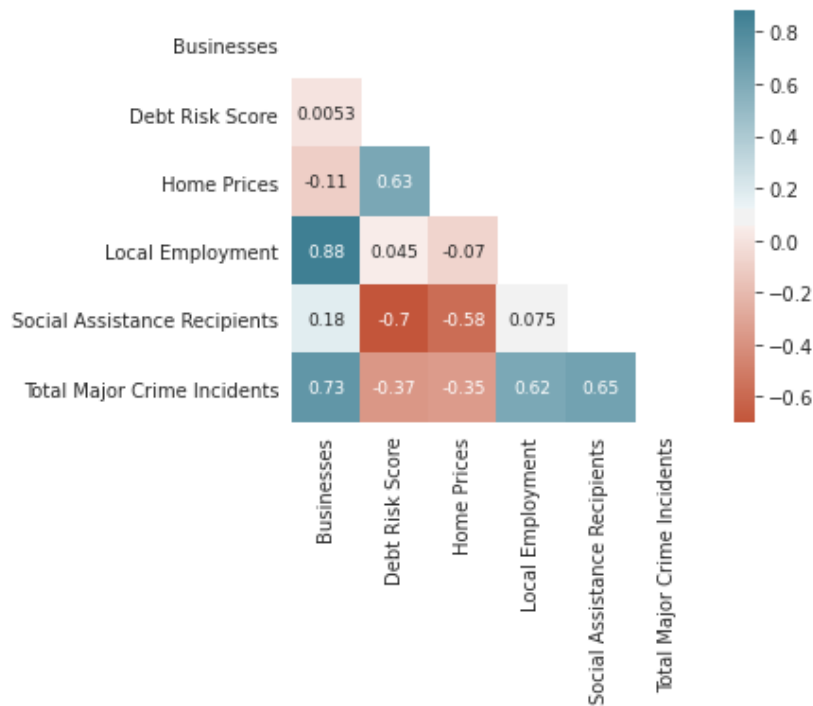


## Venues

- Identify the top 5 most common venues within a neighborhood

# Correlation Matrix

A correlation matrix is created by using the economics and crime data to view the correlations between different economic factors and safety.



## Crime and Economics

- Total number of registered **business** has **strong positive** linear relationships on the number of **total major crime incidents** and **local employment**.
- The number of **total major crime incidents** has **moderate positive** relationships on the number of **local employment** and **social assistance recipients**.

# CONCLUSION

---

## Project Summary

Providing residential recommendations among 140 neighborhoods in Toronto by applying statistical modeling on the demographics, economics, geolocation, crime and the nearby venues data.

---

---

## Further Analysis

Created user profiles by pre-define lifestyles

---

# RESOURCES

- Social Development, Finance & Administration (2014, December 31). Wellbeing Toronto - Economics, Open Data Dataset. Retrieved November 11, 2020, from: <https://open.toronto.ca/dataset/wellbeing-toronto-economics/>
- Social Development, Finance & Administration (2017, February 28). Wellbeing Toronto - Demographics, Open Data Dataset. Retrieved November 11, 2020, from: <https://open.toronto.ca/dataset/wellbeing-toronto-economics/>
- Social Development, Finance & Administration (2020, November 2). Neighborhoods, Open Data Dataset. Retrieved November 11, 2020, from: <https://open.toronto.ca/dataset/neighbourhoods/>
- Social Development, Finance & Administration (2017, February 28). Wellbeing Toronto - Safety, Open Data Dataset. Retrieved November 11, 2020, from: <https://open.toronto.ca/dataset/wellbeing-toronto-safety/>