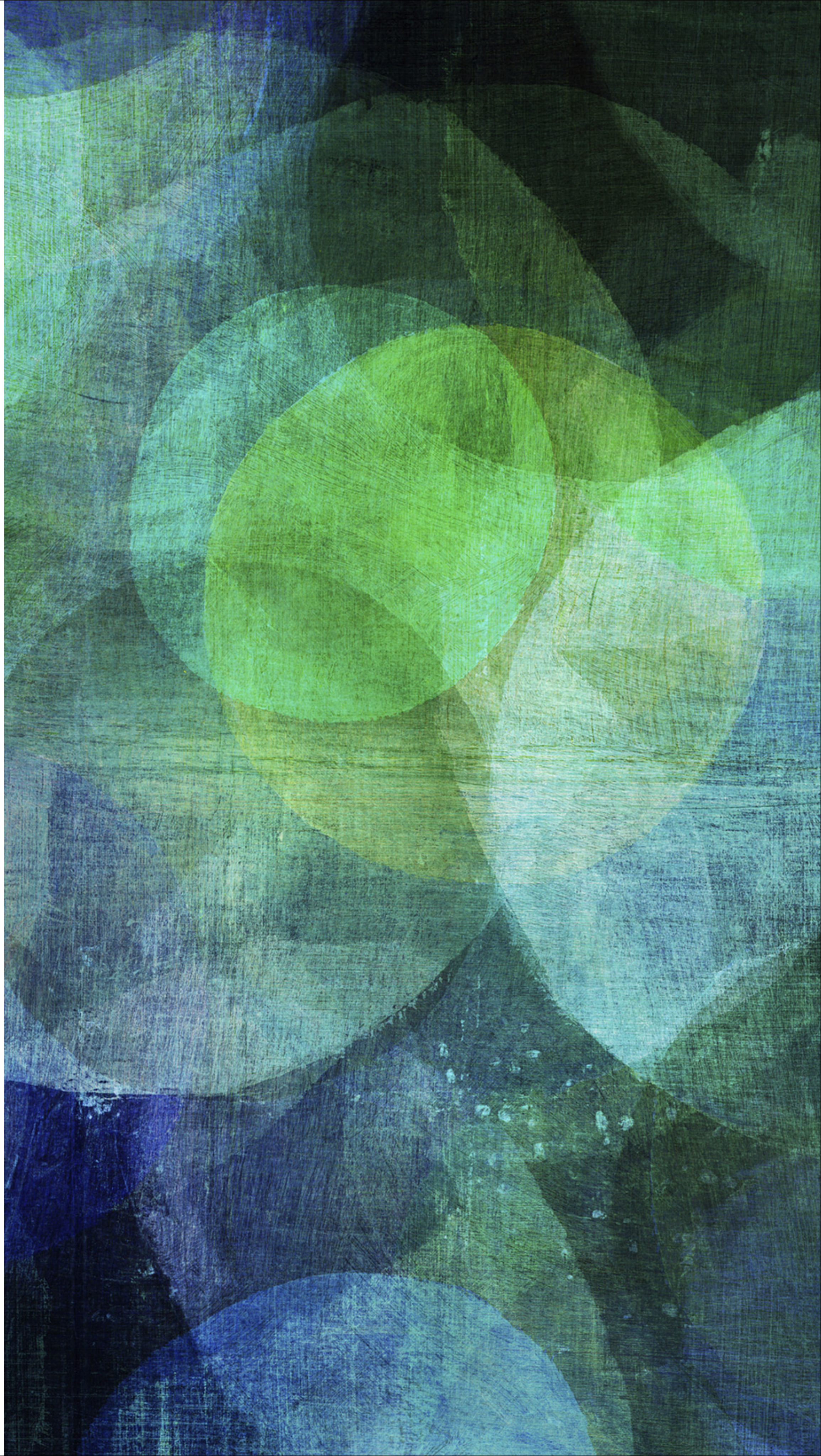
FINDING THE BEST PLACE TO OPEN A RESTAURANT IN TORONTO

INTRODUCTION



INTRODUCTION

- Objective : Finding the best neighbourhoods in Toronto to open a new restaurant of a specific type
  - Chinese restaurant, Italian restaurant, English pub ...

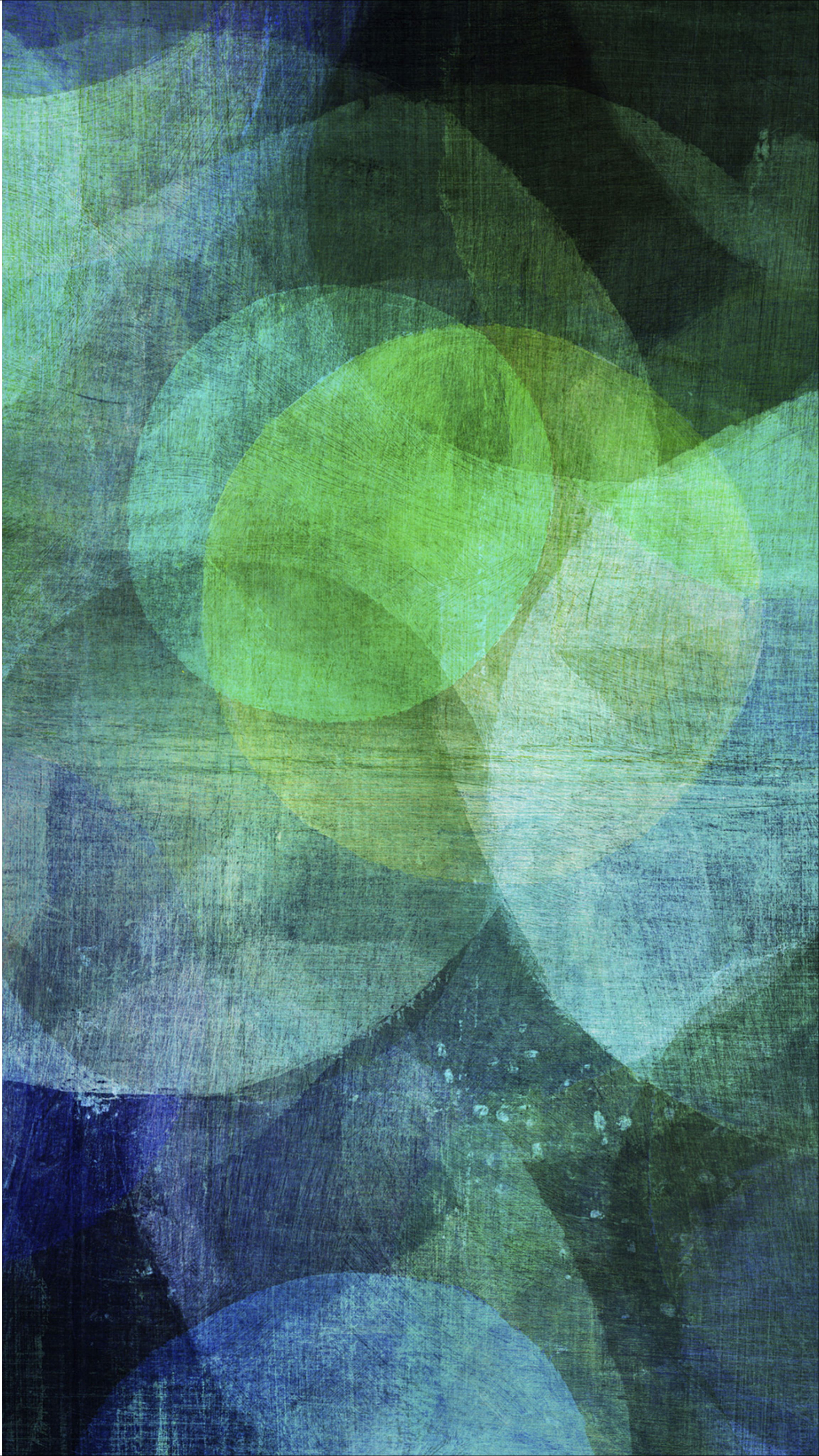
- The best neighbourhoods are the ones where :
  - The habitants would likely eat in this kind of restaurant
  - The competition is limited (e.g. there is a a reasonable number of existing restaurants of the same type in the neighbourhood).

INTRODUCTION - METHODOLOGY

1. Use of demographic data by neighbourhoods
    - It allows to gauge people's taste in each neighbourhood
    - For example, we assume that areas with a majority of chinese people would be good for chinese restaurants

2. Use of FoursquareAPI data to analyse the competition by neighbourhoods
    - We obtain the list of venues of each neighbourhood, and we can count the number of restaurants of each type.
    - The neighbourhoods with the fewest number of existing restaurants of the same type as the one we want to open will be considered to have the least competition

DATA



DEMOGRAPHIC DATA

- Neighbourhood information data

Neighbourhood

Category	Topic	Data Source	Characteristic	City of Toronto	Agincourt North	Agincourt South-Malvern West	Alderwood	Annex	Banbury-Don Mills	Bathurst Manor	Bay Street Corridor	Ba Vi
Neighbourhood Information	Neighbourhood Information	City of Toronto	Neighbourhood Number	n/a	129	128	20	95	42	34	76	52

Neighbourhood number (or CDN)

- Ethnic origin data

Neighbourhood

Category	Topic	Data Source	Characteristic	City of Toronto	Agincourt North	Agincourt South-Malvern West	Alderwood	Annex	Banbury-Don Mills	Bat Ma
Ethnic origin	Ethnic origin population	Census Profile 98-316-X2016001	North American Aboriginal origins	"35,630"	40	105	305	475	230	90
Ethnic origin	Ethnic origin population	Census Profile 98-316-X2016001	First Nations (North American Indian)	"27,610"	25	90	200	345	175	75
Ethnic origin	Ethnic origin population	Census Profile 98-316-X2016001	Inuit	515	0	0	15	20	10	0
Ethnic origin	Ethnic origin population	Census Profile 98-316-X2016001	Métis	"8,465"	10	25	100	115	60	0

Ethnic origin

Number of people by neighbourhood and ethnic origin

LIST OF VENUES BY NEIGHBOURHOOD USING THE FOURSQUAREAPI

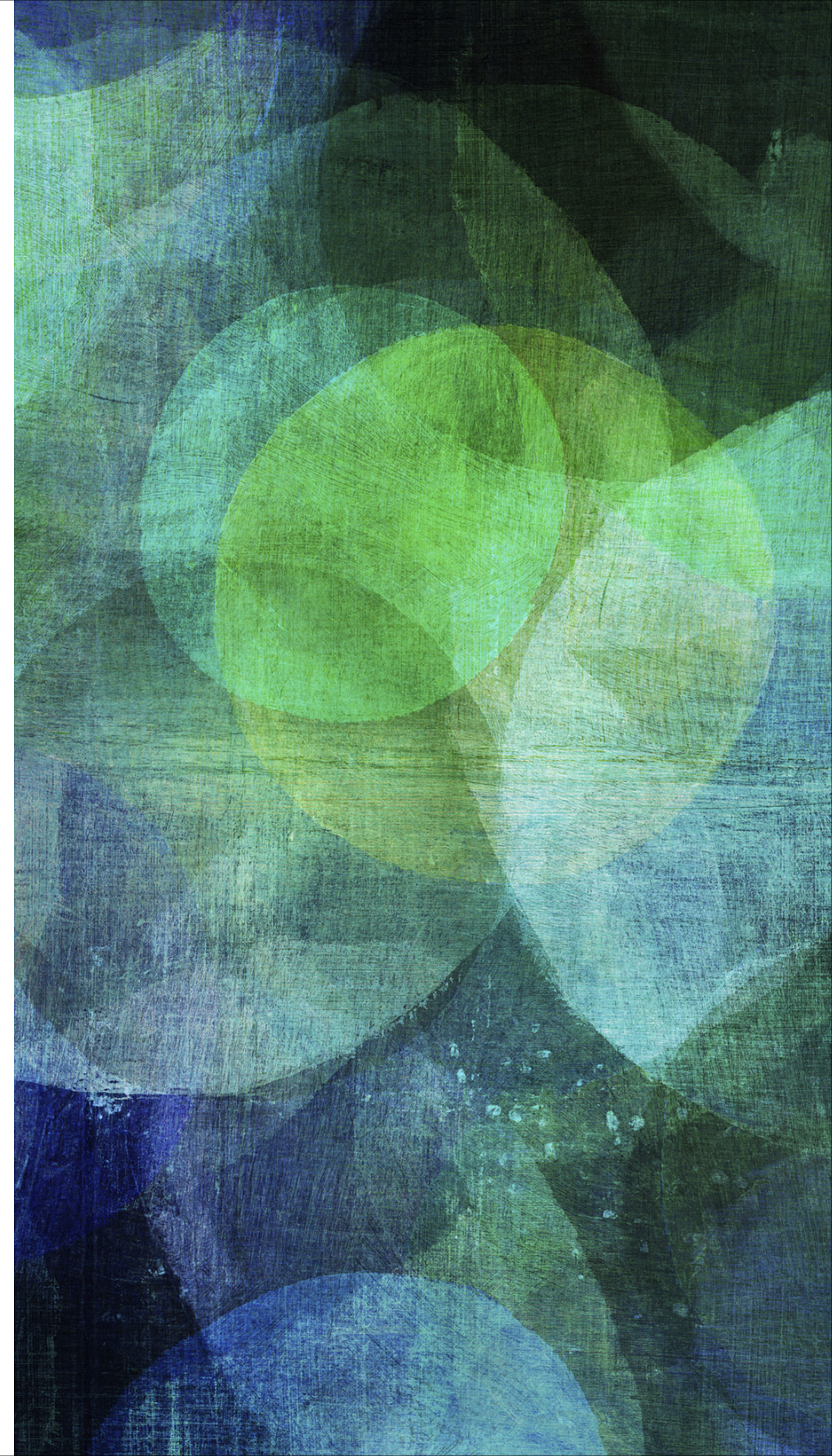

```
"response": {
  "headerLocation": "Lawrence Park South",
  "headerFullLocation": "Lawrence Park South, Toronto",
  "headerLocationGranularity": "neighborhood",
  "totalResults": 5,
  "suggestedBounds": {
  "ne": {
  "lat": 43.7182
  "lng": -79.410
  },
  "sw": {
  "lat": 43.7056
  "lng": -79.427
  }
  },
```

```
"venue": {
  "id": "5bbc5b610d173f002c4148f4",
  "name": "Wooden Woodworking Canada Inc.",
  "location": {
    "address": "22 Melham Ct, #1",
    "lat": 43.71708562332913,
    "lng": -79.41977977752686,
    "labeledLatLngs": [
      {
        "label": "display",
        "lat": 43.71708562332913,
        "lng": -79.41977977752686
      }
    ],
    "distance": 575,
    "postalCode": "M1B 2T7",
    "cc": "CA",
    "city": "Scarborough",
    "state": "ON",
    "country": "Canada",
    "formattedAddress": [
      "22 Melham Ct, #1",
      "Scarborough ON M1B 2T7",
      "Canada"
    ]
  },
  "categories": [
    {
      "id": "545419b1498ea6ccd0202f58",
      "name": "Service à domicile",
      "pluralName": "Réparations et services à domicile",
      "shortName": "Services à domicile",
      "icon": {
      "prefix": "https://ss3.4sqi.net/img/categories_v2/shops/hardware_",
      "suffix": ".png"
    },
    "primary": true
    }
  ],
```


- In this example :

1. The place comes from the Lawrence Park South : headerLocation tag
2. The name of the place is Wooden Woodworking Canada Inc
3. The category of the place is Service à domicile (english translation : Home service) : Group -> Items -> Venue -> Categories -> Name tag

METHODOLOGY

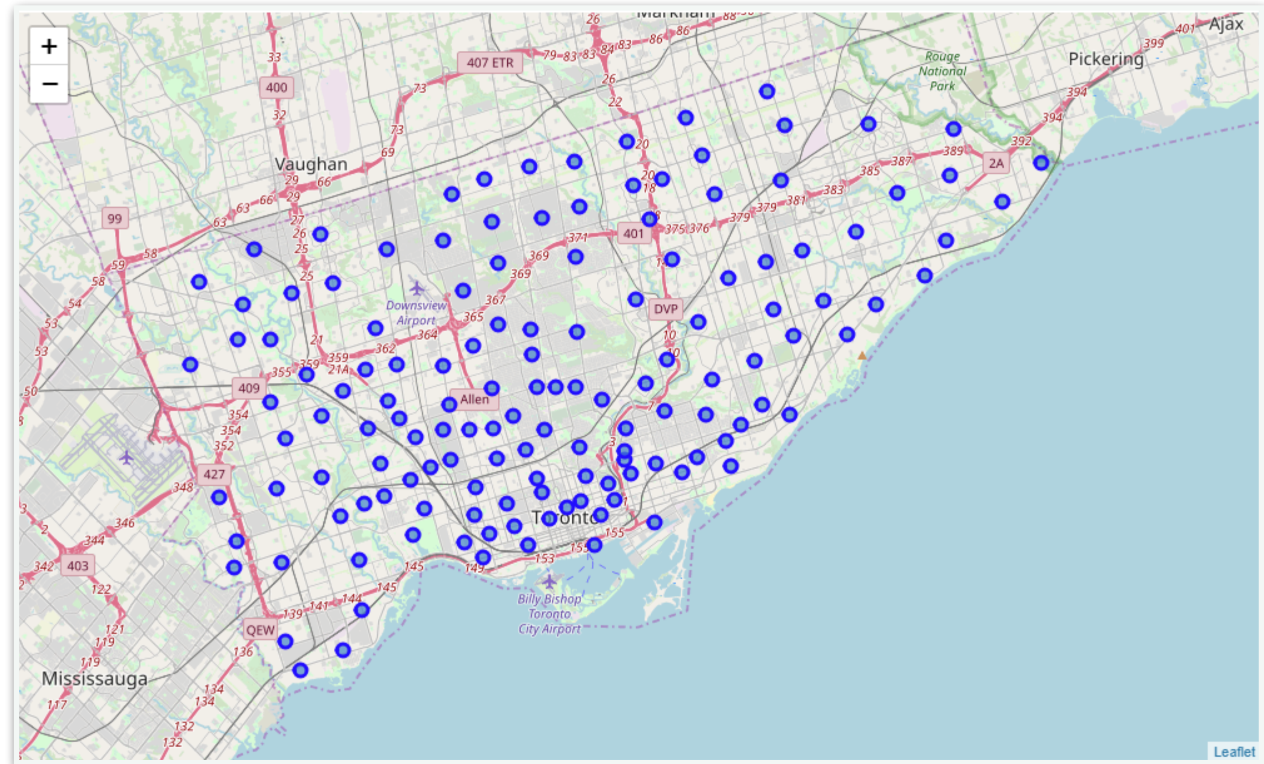


DATA ANALYSIS 1/2

- Get each neighbourhood's coordinates
  - Use the Geocoder package

CDN	City_Area	Latitude	Longitude
129	Agincourt North	43.80930	-79.26707
128	Agincourt South-Malvern West	43.78735	-79.26941
20	Alderwood	43.60496	-79.54116
95	Annex	43.66936	-79.40280
42	Banbury-Don Mills	43.74041	-79.34852

- Draw the neighbourhoods on the map of Toronto
  - Use the Folium package
  - Good way to visualise the position of each neighbourhood
  - It confirms that our dataset covers the whole city



DATA ANALYSIS 2/2

- Find out the top most common ethnic origins by neighbourhood
  - In order to prepare our data for clustering by demographic data

Agincourt North

Origin	Count
Chinese	16950.0
Sri Lankan	2230.0
East Indian	2090.0
Filipino	1465.0
Canadian	1295.0

Alderwood

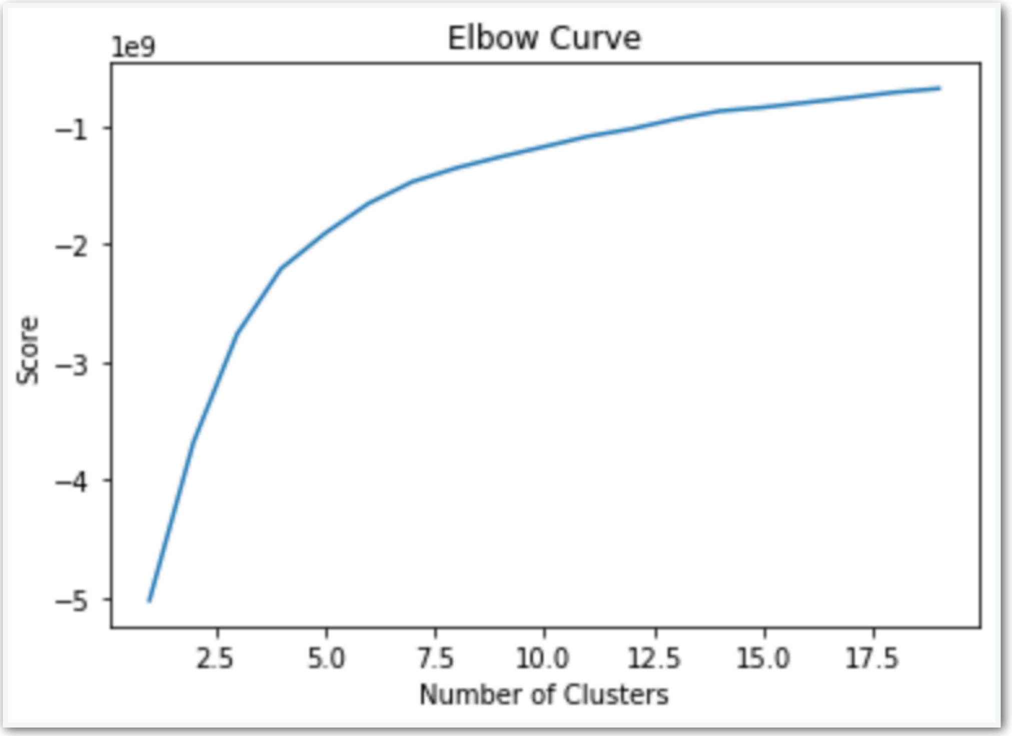
Origin	Count
English	2320.0
Canadian	2245.0
Irish	1900.0
Scottish	1720.0
Italian	1275.0

MACHINE LEARNING - ALGORITHM USED

- K-Means algorithm
  - One on the most used algorithm for unsupervised learning and clustering
  - Typically used for scenarios like understanding the population demomgraphics, market segmentation, social media trends, anomaly detection, etc… where the clusters are unknown to begin with
  - One of the simplest clustering algorithm to implement and to run
  - It is less time consuming than other, more complex algorithms

MACHINE LEARNING - CLUSTERING

- Find the optimal number of clusters : Elbow method
  - Find the most appropriate number of clusters in a dataset, by running several K-means algorithm and comparing the sum of squared distances of samples to the nearest cluster centre
  - The elbow is visible for K = 5, this is the value we use



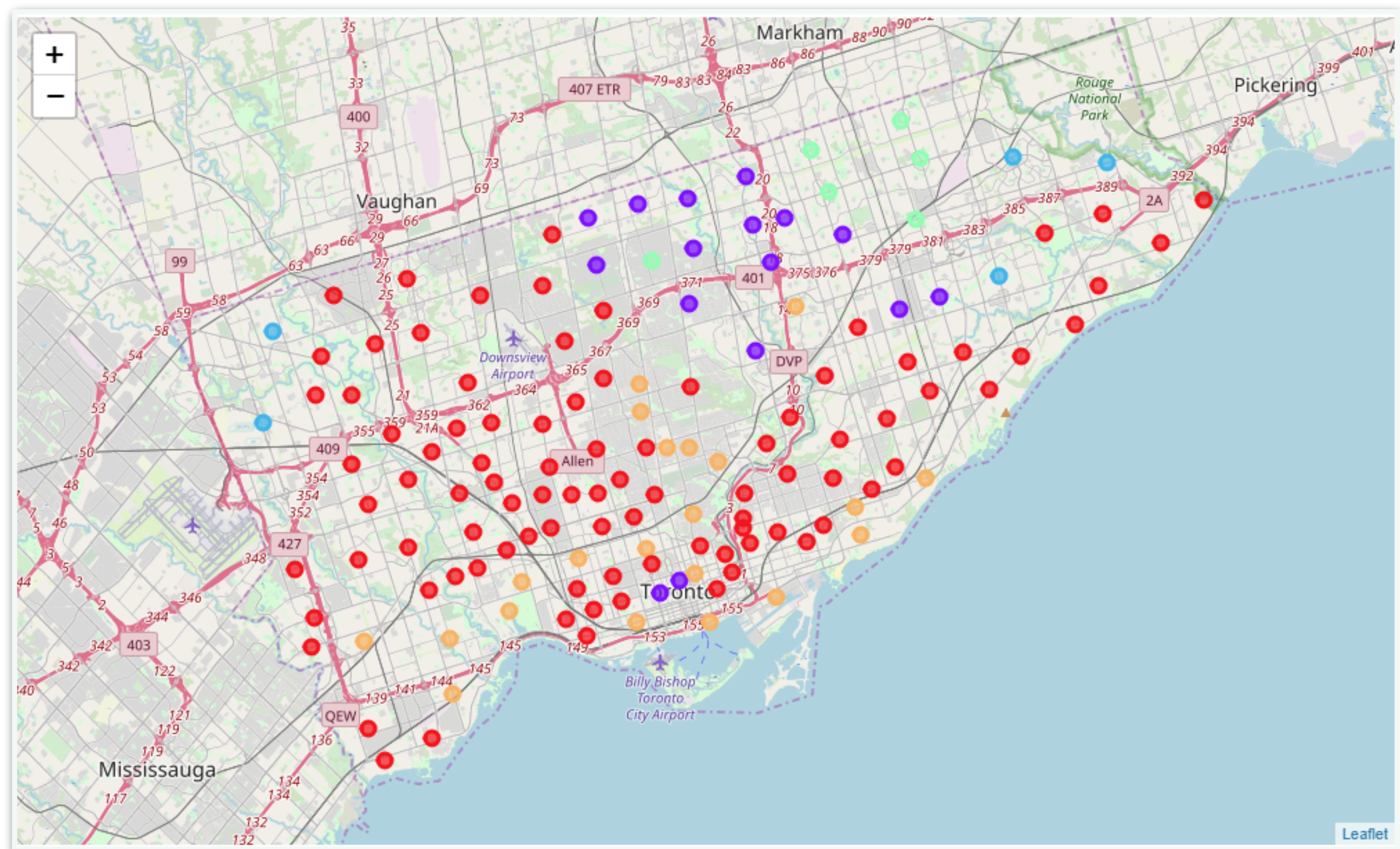
MACHINE LEARNING – CLUSTERING

- Add the clusters labels to the neighbourhoods dataframe
  - Example with two neighbourhoods

Cluster

CDN	City_Area	Latitude	Longitude	Cluster Labels	1st Most Common Origin	2nd Most Common Origin	3rd Most Common Origin	4th Most Common Origin	5th Most Common Origin	6th Most Common Origin	7th Most Common Origin	8th Most Common Origin	9th Most Common Origin	10th Most Common Origin
129	Agincourt North	43.80930	-79.26707	3	Chinese	Sri Lankan	East Indian	Filipino	Canadian	English	Tamil	Jamaican	Scottish	Irish
20	Alderwood	43.60496	-79.54116	0	English	Canadian	Irish	Scottish	Italian	Polish	German	French	Ukrainian	Portuguese

- Draw the clusters on the map
  - One colour by cluster



MACHINE LEARNING – CLUSTERS OBTAINED

➤ We obtain the following clusters (top 3 neighbourhoods)

Cluster 0 : European and Canadian people

CDN	City_Area	Latitude	Longitude	Cluster Labels	1st Most Common Origin	2nd Most Common Origin	3rd Most Common Origin	4th Most Common Origin	5th Co O
20	Alderwood	43.604960	-79.541160	0	English	Canadian	Irish	Scottish	Ita
57	Broadview North	43.689370	-79.354290	0	English	Irish	Scottish	Greek	Ca n
11	Eringate-Centennial-West Deane	43.661910	-79.577380	0	Canadian	English	Italian	Irish	Sc

Cluster 1 : Asian people

CDN	City_Area	Latitude	Longitude	Cluster Labels	1st Most Common Origin	2nd Most Common Origin	3rd Most Common Origin	4th Most Common Origin	5th Cor Oı
127	Bendale	43.75963	-79.25739	1	Chinese	East Indian	Filipino	Canadian	En
47	Don Valley Village	43.78552	-79.35017	1	Chinese	Filipino	East Indian	Iranian	En
126	Dorset Park	43.75533	-79.27746	1	Filipino	East Indian	Chinese	Canadian	Sri Lar

Cluster 2 : Indian people

CDN	City_Area	Latitude	Longitude	Cluster Labels	1st Most Common Origin	2nd Most Common Origin	3rd Most Common Origin	4th Most Common Origin	5th Cor Oı
132	Malvern	43.80977	-79.22084	2	East Indian	Sri Lankan	Filipino	Chinese	Jar
2	Mount Olive-Silverstone-Jamestown	43.74721	-79.58826	2	East Indian	Iraqi	Jamaican	Canadian	So
131	Rouge	43.80766	-79.17405	2	East Indian	Sri Lankan	Canadian	Filipino	Jar

Cluster 3 : Chinese people

CDN	City_Area	Latitude	Longitude	Cluster Labels	1st Most Common Origin	2nd Most Common Origin	3rd Most Common Origin	4th Most Common Origin	5th Cor Oı
129	Agincourt North	43.80930	-79.26707	3	Chinese	Sri Lankan	East Indian	Filipino	Ca n
128	Agincourt South-Malvern West	43.78735	-79.26941	3	Chinese	East Indian	Filipino	Sri Lankan	Ca n
117	L'Amoreaux	43.79726	-79.31220	3	Chinese	East Indian	Canadian	Sri Lankan	Fili

Cluster 4 : Irish, Scottish and English people

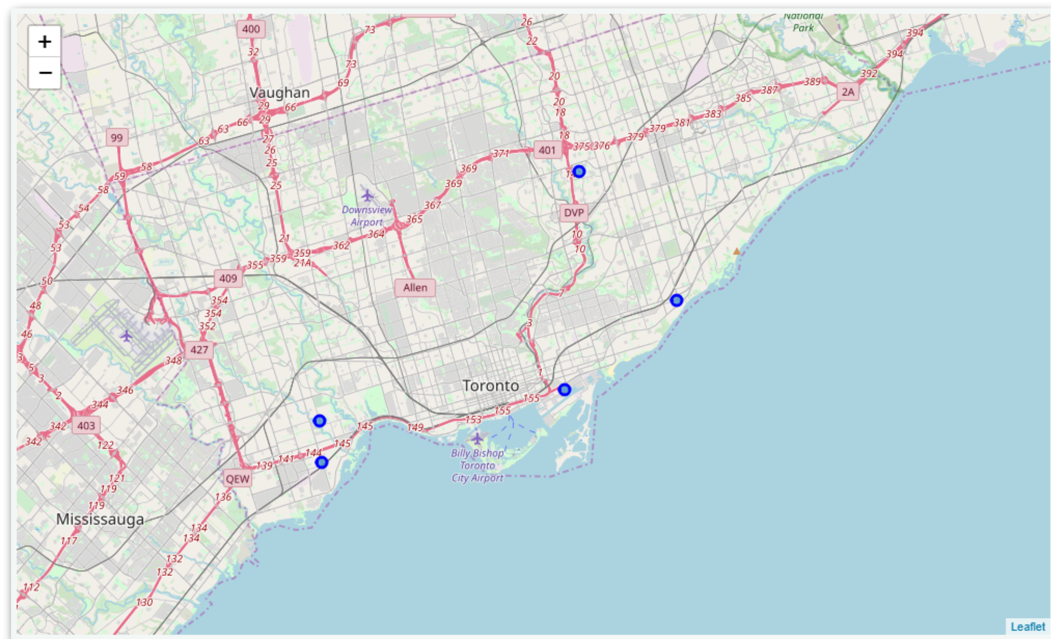
CDN	City_Area	Latitude	Longitude	Cluster Labels	1st Most Common Origin	2nd Most Common Origin	3rd Most Common Origin	4th Most Common Origin	5tl Co C
95	Annex	43.66936	-79.40280	4	English	Irish	Scottish	Canadian	G
122	Birchcliffe-Cliffside	43.69472	-79.26460	4	English	Irish	Canadian	Scottish	Fr
75	Church-Yonge Corridor	43.66024	-79.37868	4	English	Irish	Scottish	Chinese	Ca n

FOURSQUAREAPI – COMPETITION ANALYSIS

- Example : Analysis of competition for a **new Irish/ English pub**
  - Isolate the cluster that contains neighbourhoods with a lot of Irish/English people (Cluster 4)
  - For each neighbourhood, retrieve the list of existing pubs using FoursquareAPI
  - Count the number of pubs for each neighbourhood. We assume that the neighbourhoods with the fewer number of existing pubs, within this cluster, are the best spots (**more potential clients + less competition**)
  - Draw the top 5 neighbourhoods on the map

CDN	Area Latitude	Area Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
95	43.66936	-79.40280	The Madison Avenue Pub	43.667947	-79.403486	Pub
95	43.66936	-79.40280	Duke of York	43.669186	-79.397527	Pub
75	43.66024	-79.37868	Churchmouse & Firkin	43.664632	-79.380406	Pub
62	43.68415	-79.29911	Grover Pub and Grub	43.679181	-79.297215	Pub
62	43.68415	-79.29911	Mullins Irish Pub	43.680348	-79.289370	Pub

Count	CDN	City_Area	Latitude	Longitude	Cluster Labels
4	16	Stonegate-Queensway	43.63718	-79.50058	4
15	122	Birchcliffe-Cliffside	43.69472	-79.26460	4
21	45	Parkwoods-Donalda	43.75613	-79.32880	4
29	70	South Riverdale	43.65221	-79.33820	4
30	17	Mimico (includes Humber Bay Shores)	43.61729	-79.49885	4



CONCLUSION

- In this project, we managed to cluster the city of Toronto using **demographic data by neighbourhoods**. This helps us identify which neighbourhoods are the most adequate for opening a new restaurant of a specific type.

- Then, we managed to identify the neighbourhoods with **the less competition** within these adequate neighbourhoods, in order to optimise the performance of this new business.

- **Food service contractors** can use similar data analysis in order to find **the best spots to open a new restaurant** of a specific type.