THE BATTLE OF NEIGHBORHOODS- PROJECT FINAL

NEW ITALIAN GOURMET RESTAURANT¶

1. INTRODUCTION
1.1 Background

Toronto is the provincial capital of Ontario and the most populous city in Canada, with a population of 2,954,024 as of July 2018. 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 encompasses a geographical area formerly administered by many separate municipalities. These municipalities have each developed a distinct history and identity over the years, and their names remain in common use among Torontonians. Former municipalities include East York, Etobicoke, Forest Hill, Mimico, North York, Parkdale, Scarborough, Swansea, Weston and York. Throughout the city there exist hundreds of small neighbourhoods and some larger neighbourhoods covering a few square kilometres.

1.2 Problem

The next project seeks to find the best area to open a new Gourmet Italian restaurant in the capital of Ontario Canada, Toronto. The goal of this exploratory study is to provide an tentative areas of greater sale and less competitiveness. Venice is a family restaurant of traditional Italian food that aims to differentiate itself from the rest of the restaurants in the municipality of Toronto through its gourmet gastronomic offer and made with the best local and organic products.

Although there is a wide variety of restaurants throughout the old Toronto, York and North York area, that range from Mexican food to Asian, our study focused on finding Italian competition and how to make this restaurant different from the others with a unique offer for the most demanding palates but also continuing with the idea of giving and having the best organic products in the area.

1.3 Interest

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. Based on the above, our study is limited to Toronto and Metropolitan area because the level of international business success is fruitful for investors

2. DATA DESCRIPTION
2.1 Data Sources

In this study the data will be collected in multiple ways, through the Foursquare Technology System and data collected from Toronto Open Data as well as data collection in segmentation and clustering.

The data from Foursquare will be used in the areas of great tourism as they are Old Toronto, York and North York, using the geolocation data in Foursquare we will find the Italian gourmet restaurants in the area as well as: tips, comments, distances, menu and rating and this data will be semi-structured, which is data that do not have a defined format, what they have are labels that facilitate separate one data from another.

2.2 Data cleaning

Before analysis the gathered data was prepared. The dataset was checked for missing data and outliers. For this the "outlier labeling rule" was used. All values outside the calculated range were considered outliers (Hoaglin & Iglewicz, 1987). The data was then analyzed using statistical software Watson IBM

In the following exploratory study of quantitative character it will help us to know the situation of the
 food sector of the municipality. As well as, analyze our competitors and the possible clientele that
 interests us for our business. The model has been designed based on a clutsering for segmentation
(K-means), because we are interested in doing an exploratory study with unspervised data.

As a database, I used GitHub repository in my study.
My master data which has the main components Borough, Average Price,
Latitude and Longitude informations of the city.

The strategy is based on mapping the above described data in section 2.0, in order to facilitate the choice of the best places for the new restaurant. The choice is made based on the demands imposed: proximity to other restaurants of the same category, other restaurants of the same concept, menus, prices, rating, tourist area and purchasing value venues to Toronto. The procesing of these DATA and its mapping will allow to answer the key questions to make a decision:

1. What is the area of Toronto with the best rental price that meets the established criteria?
2. What kind of Italian restaurants exist in the area?
3. What is the rating per restaurant of the same category?
4. What differentiates the other restaurants from our business plan?

3.2 Metodology execute – Mapping Data

This is the first section of the final project of Applied Data Science Capstone. First we will import the libraries that we will need.

I used python beautyfolSoup and CSV library to visualize geographic details of Toronto and its boroughs and I created a map of the capital with boroughs superimposed on top. I used latitude and longitude values to get the visual as below

Figura 2: Italian Restaurants

This method was easier than the Beautiful Soup but I put the Beautiful Soup method in the following blocks to meet the requirement of the assignment. The dataframe will consist of three columns: postalcode, borough, and neighborhood

3.2.1 Data transformed into pandas dataframe

Now that you have built a dataframe of the postal code of each neighborhood, we need to get the latitude and the longitude coordinates of each neighborhood.

3.2.4 Latitude and the longitude coordinates of each neighborhood.

```
In [30]: geo_data=pd.read_csv("https://cocl.us/Geospatial_data")
         geo_data
```

Out[30]:

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476
5	M1J	43.744734	-79.239476
6	M1K	43.727929	-79.262029
7	M1L	43.711112	-79.284577
8	M1M	43.716316	-79.239476
9	M1N	43.692657	-79.264848
10	M1P	43.757410	-79.273304
11	M1R	43.750072	-79.295849

# FIGURA 1: LATITUDE AND LONGITUDE

4. Results

I utilized the Foursquare API to explore the boroughs and segment them. I designed the limit as 100 venue and the radius 750 meter for each borough from their given latitude and longitude informations. Here is a head of the list Venues name, category, latitude and longitude informations from Forsquare API.

Most restaurant locations are distributed throughout the Old Toronto area, although we discovered that the most prestigious area is the one that is found in lawrence park, where there are only 3 Italian restaurant.
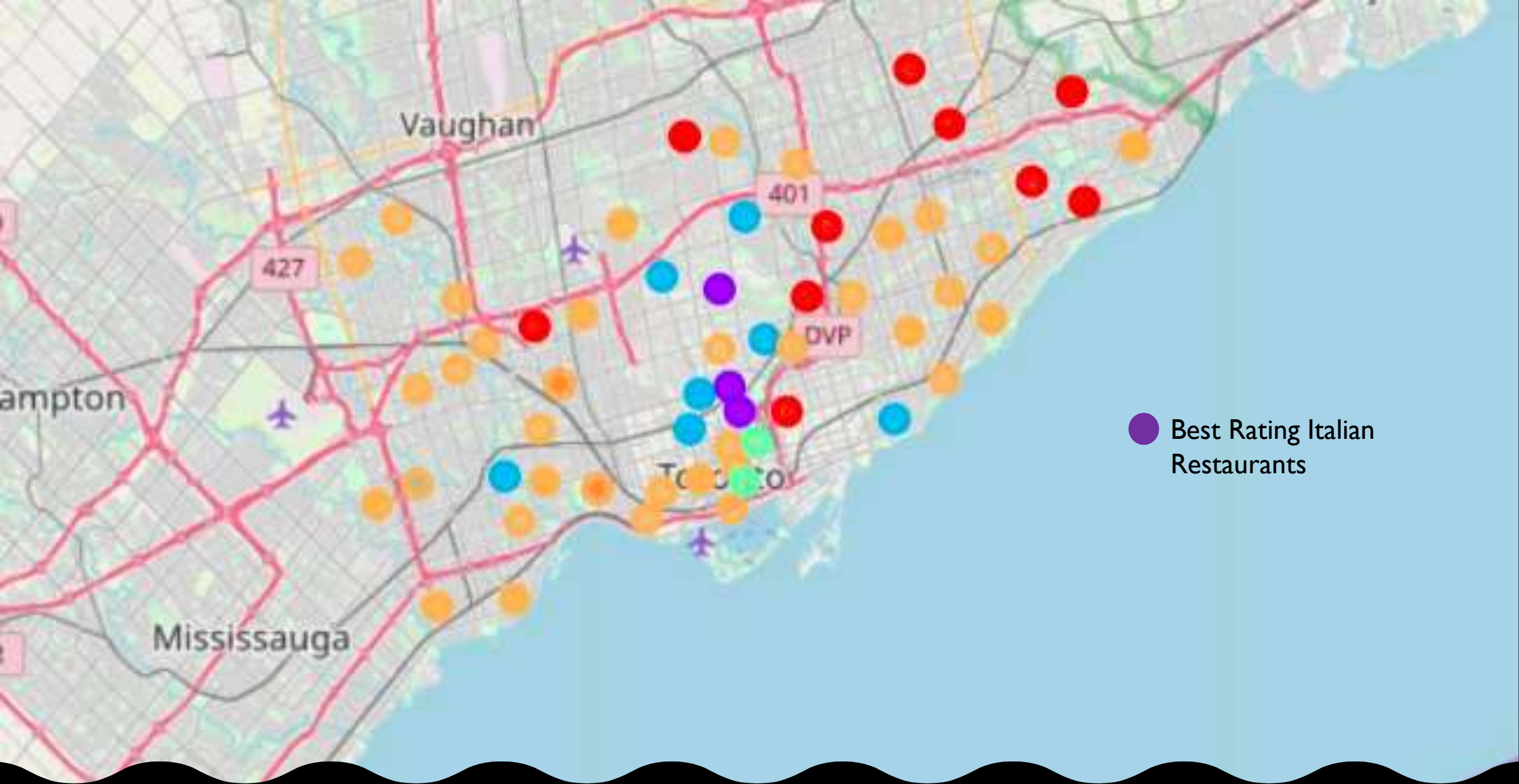
Best Rating Italian
Restaurants

Another one of the exclusive zones is Lawrence Hights where there are stores like chanel, diesel, Lacoste, Hugo Boss etc. Are only 1 Italian restaurant, that is not Gourmet and a bad rating by its consumers.

Best Rating Italian
Restaurants
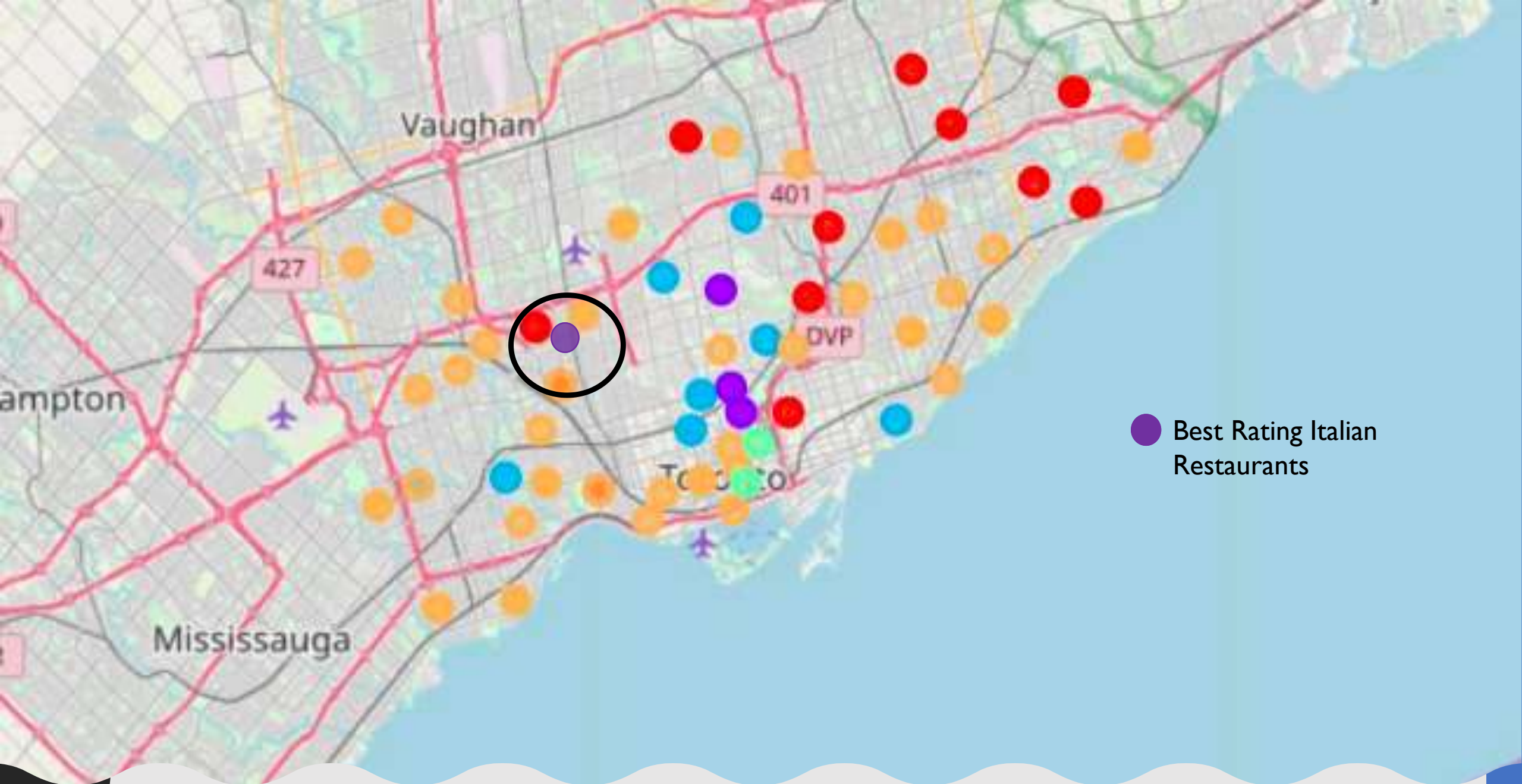
5. Conclusion and Recommendation

To finish this study we can say that the most successful areas are in Nort York, are Lawrence Park and North York, Lawrence Hights, because the category of the new restaurant is gourmet and in the area there is no other Italian of the same concept.

The competition does not offer any gourmet service or elaborated by local and organic products

6. Discussion¶

This analysis is performed pn limited data, this may be right or may be wrong. But if good amount of data is available there is scope to come up with better results.

This study was of exploratory character by which it leaves open some other ways to study as they are:

1. Departament zones
2. Type of client
3. Cultural areas
4. Quick accesses
5. Transit