Battle of Neighborhood's

Recommendation System on the opportunity to open a Wine Bar in a Neighborhood of Toronto

By Massimo D'Alessandro

Presentation made under IBM's Professional Data Science Certification Coursera Capstone Project .

Business Problem

- A new investor has contacted us looking to open a Wine Bar in Toronto.
- The investor operates in the field of bar and restaurant and has many successful wine bars in the major cities of Italy, mainly in Milan.
- The investor wants to expand the wine bar activity in Canada starting from Toronto.
- The investor is looking for guidance to understand if Toronto could be a successful location, proving if there is a similarity with Milan, and which is the best neighborhood where to locate the Wine Bar.

DATA

REQUIREMENT

Data that will be required for solving the problem

- List of all Postal codes of Toronto and relative coordinates
- List of all Postal codes of Milan and relative coodinates
- List of all venues in each locality in the selected city

SOURCE

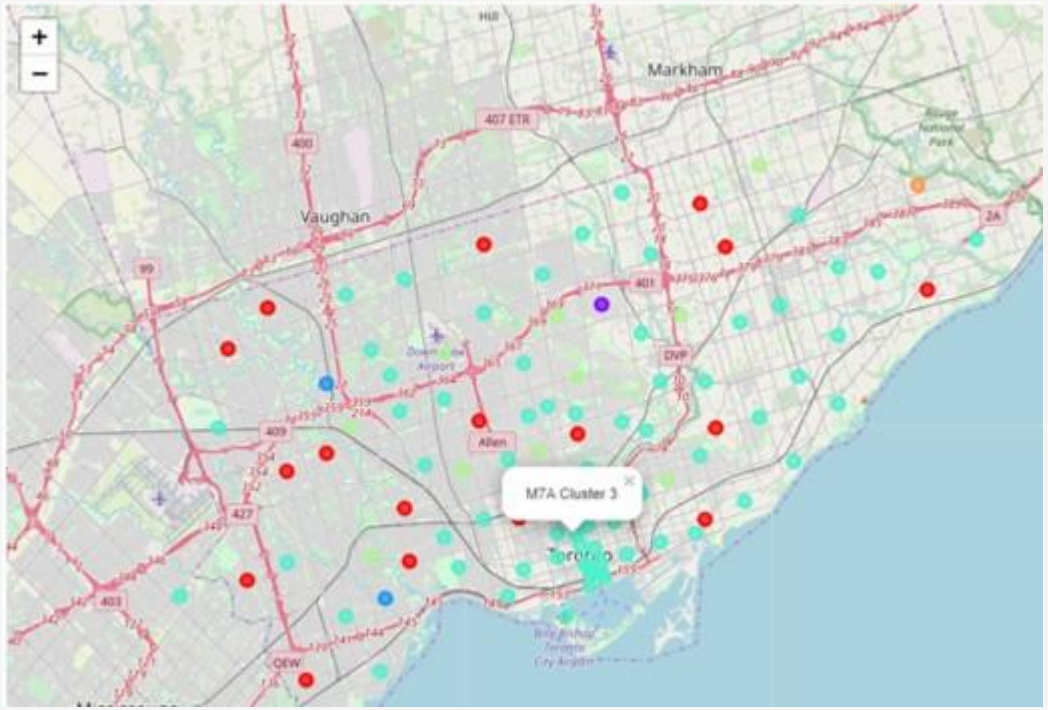
Wikipedia

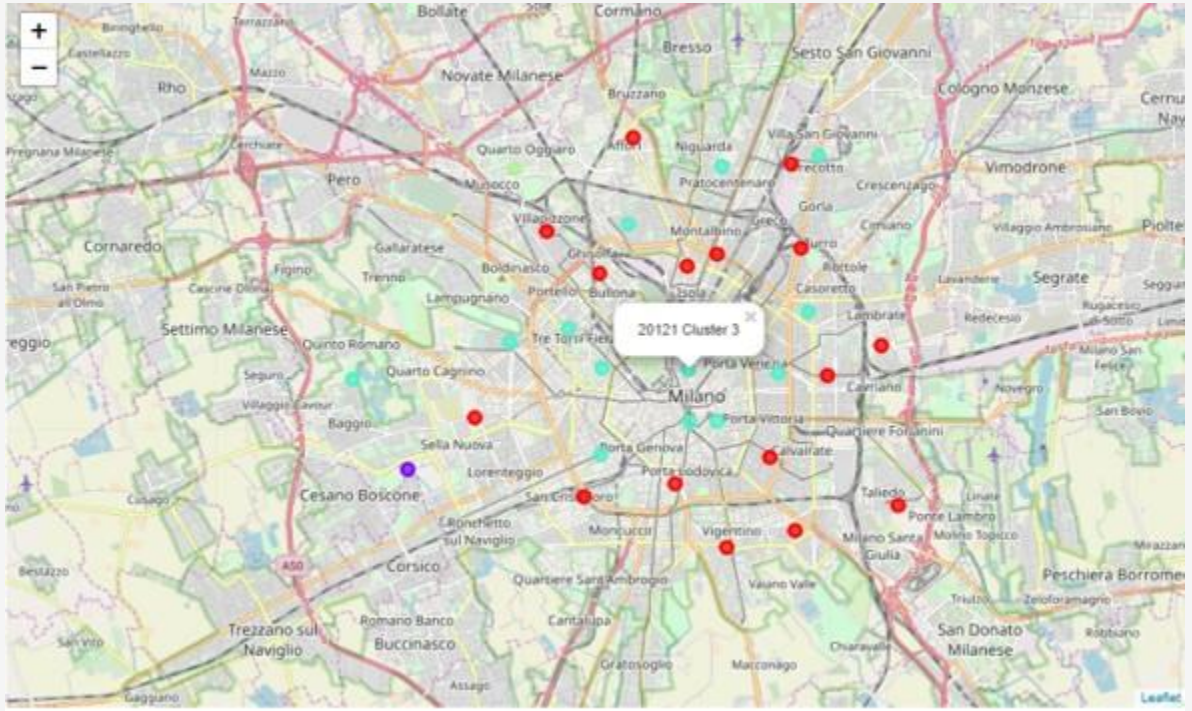
- **https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:**
- **https://worldpostalcode.com/italy/lombardia/milano**

Foursquare API

- For Venues in City and in Each locality in selected city.

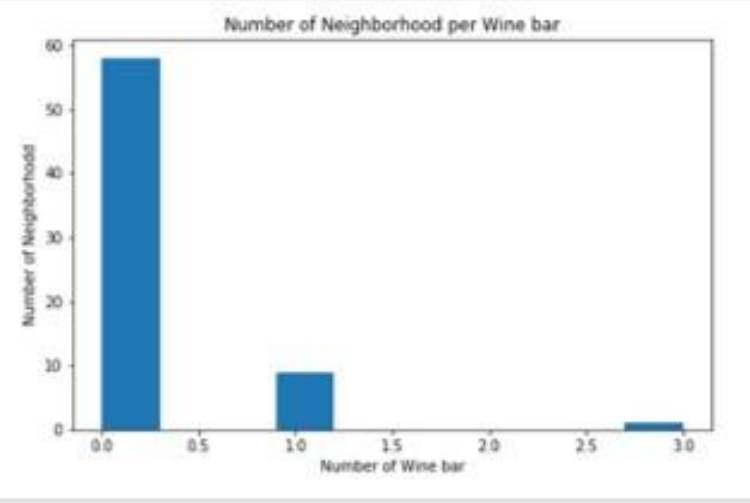
Using K-means clustering to prove a similarity between Toronto and Milan

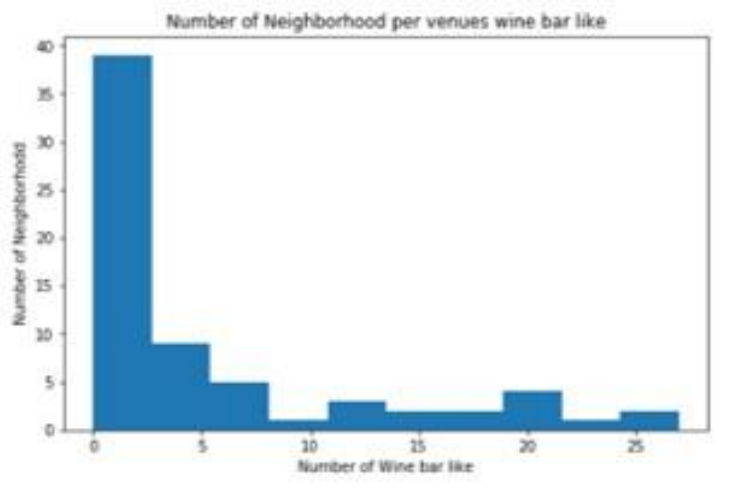




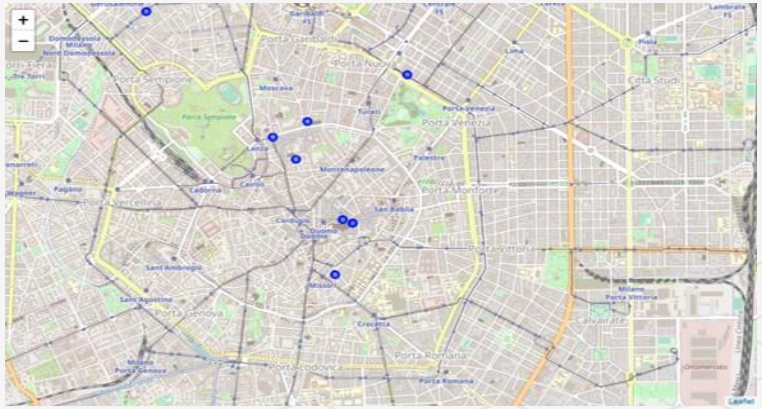
- In an area around the center, neighborhoods belong to the same cluster, proving that cities are similar
- This area will be the target to analyze, to understand if there are conditions to locate a wine bar in Toronto

Analysis of Wine bar and Wine bar like venues



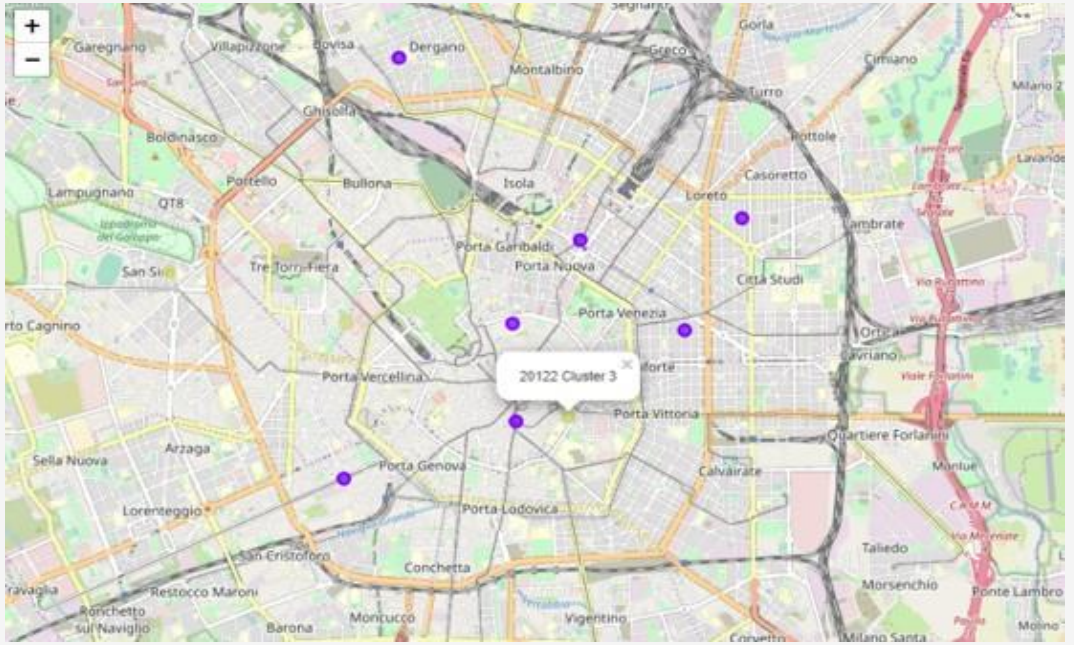


- Very few Neighborhoods have more than one wine bar
- in some area there is a high concentration of wine bar like venues; 25 venues in 3 area



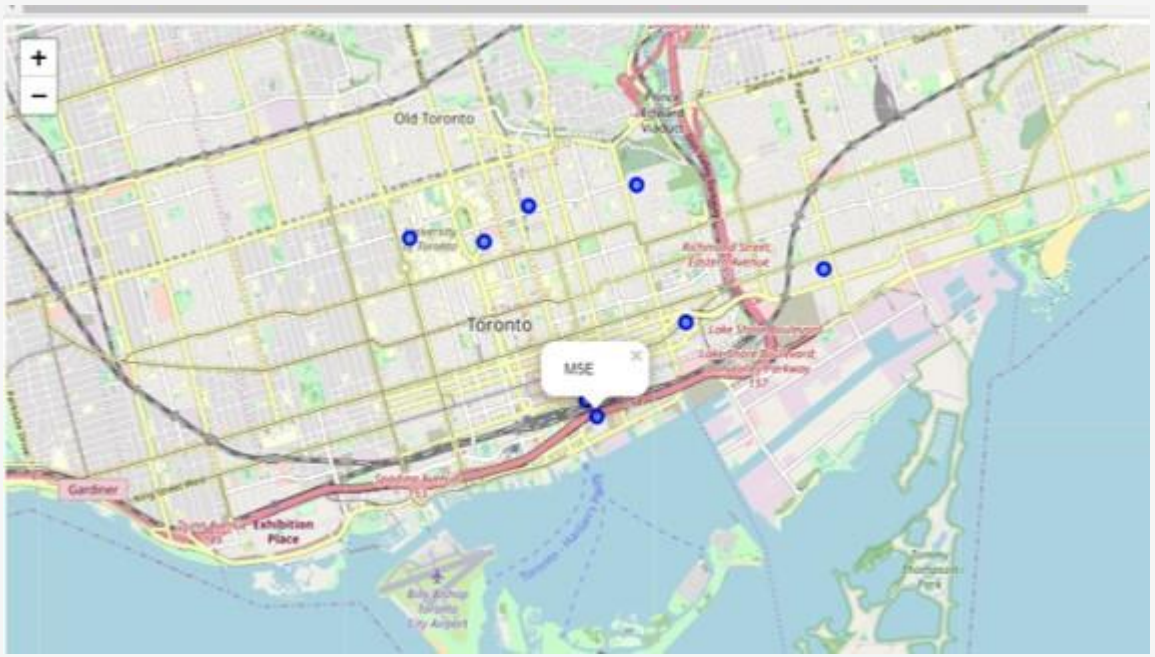
- All wine bar in Toronto are concentrated in the same area

Using K-means clustering to identify the most similar Toronto Neighborhood to the Milan neighborhood where the wine bar is located



- Cluster number 3 (yellow circle) represent similar neighborhoods in Toronto to the neighborhood in Milan where the wine bar is located
- The wine bar in Toronto could be locate in the cluster 3 neighborhood adding the followin condition:
  - The neighborhood of the cluster has no Wine Bar
  - The neighborhood of the cluster has more than 5 venues like Wine Bar among Wine Shop, Bar, Beer Bar, Café, Cocktail Bar, Coffee Shop
  - The neighborhood is not far from the center, less than 4 KMs

Conclusions

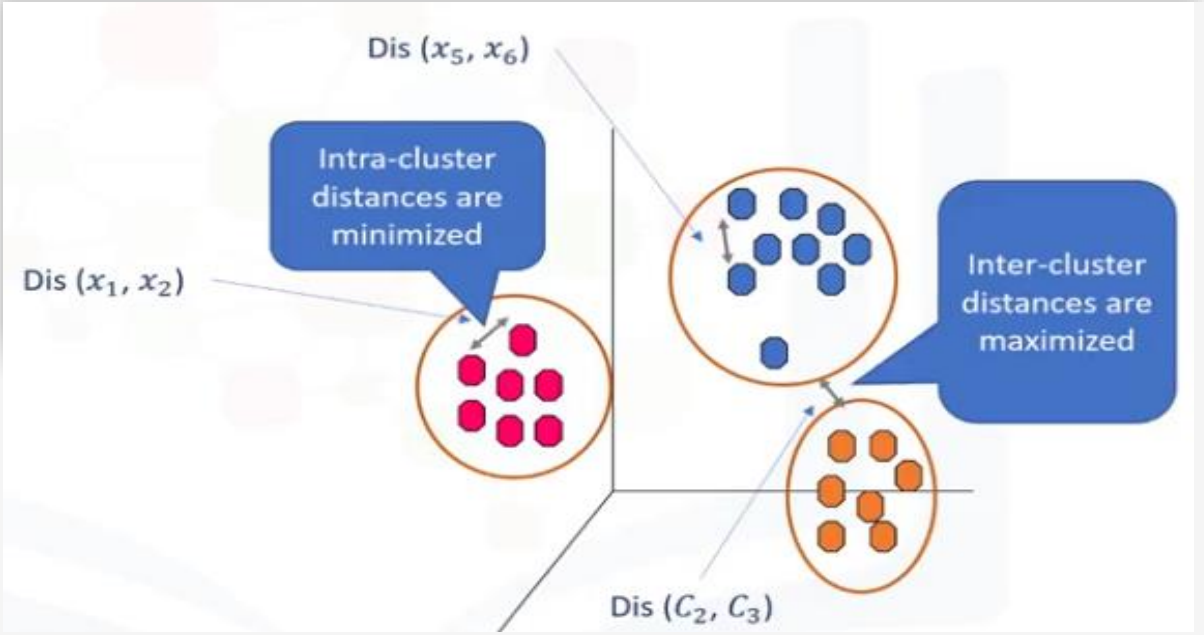


- The map represent the 9 neighborhoods where the wine bar has to be located, they are the most similar to the neighborhood of Milan where the successful wine bar is located.
- The same process can be applied to any other activity

K- Means Clustering (Unsupervised Machine Learning Algorithm)

Unsupervised Learning

Supervise, means to observe, and direct the execution of a task, project, or activity. We do not supervise the model, but we let the model work on its own to discover information that may not be visible to the human eye. It means, the unsupervised algorithm trains on the dataset, and draws conclusions on unlabeled data.



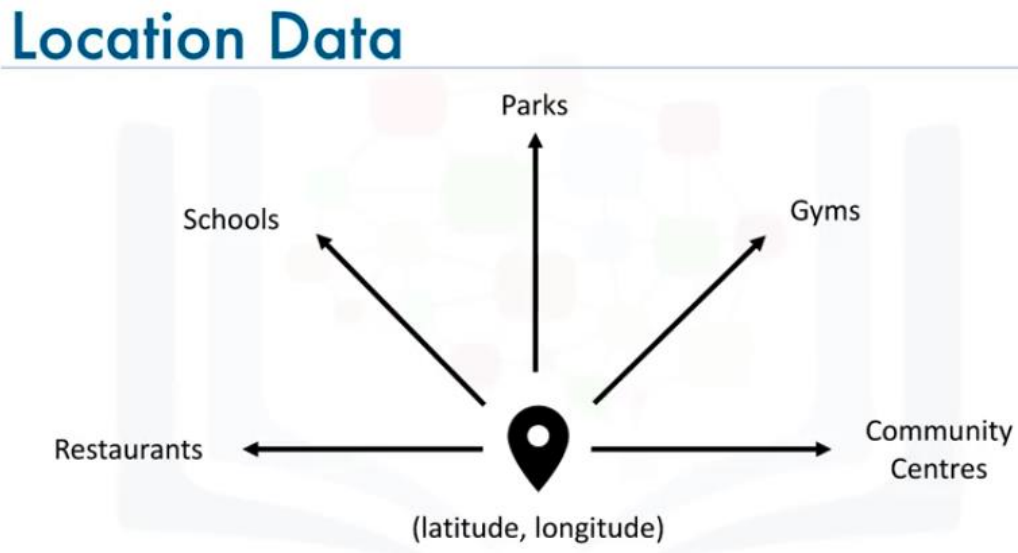
What is a Cluster?

1. A cluster is a group of data points or objects in a dataset that are similar to other objects in the group, and dissimilar to datapoints in other clusters.

K-Means Clustering

2. K-Means is a type of partitioning clustering, that is, it divides the data into K non-overlapping subsets or clusters without any cluster internal structure or labels. This means, it's an unsupervised algorithm.

Location Data Provider – Foursquare



1 Location Data

Location data is data describing places and venues, such as their geographical location, their category, working hours, full address, and so on, such that for a given location given in the form of its geographical coordinates (or latitude and longitude values) one is able to determine what types of venues exist within a defined radius from that location.

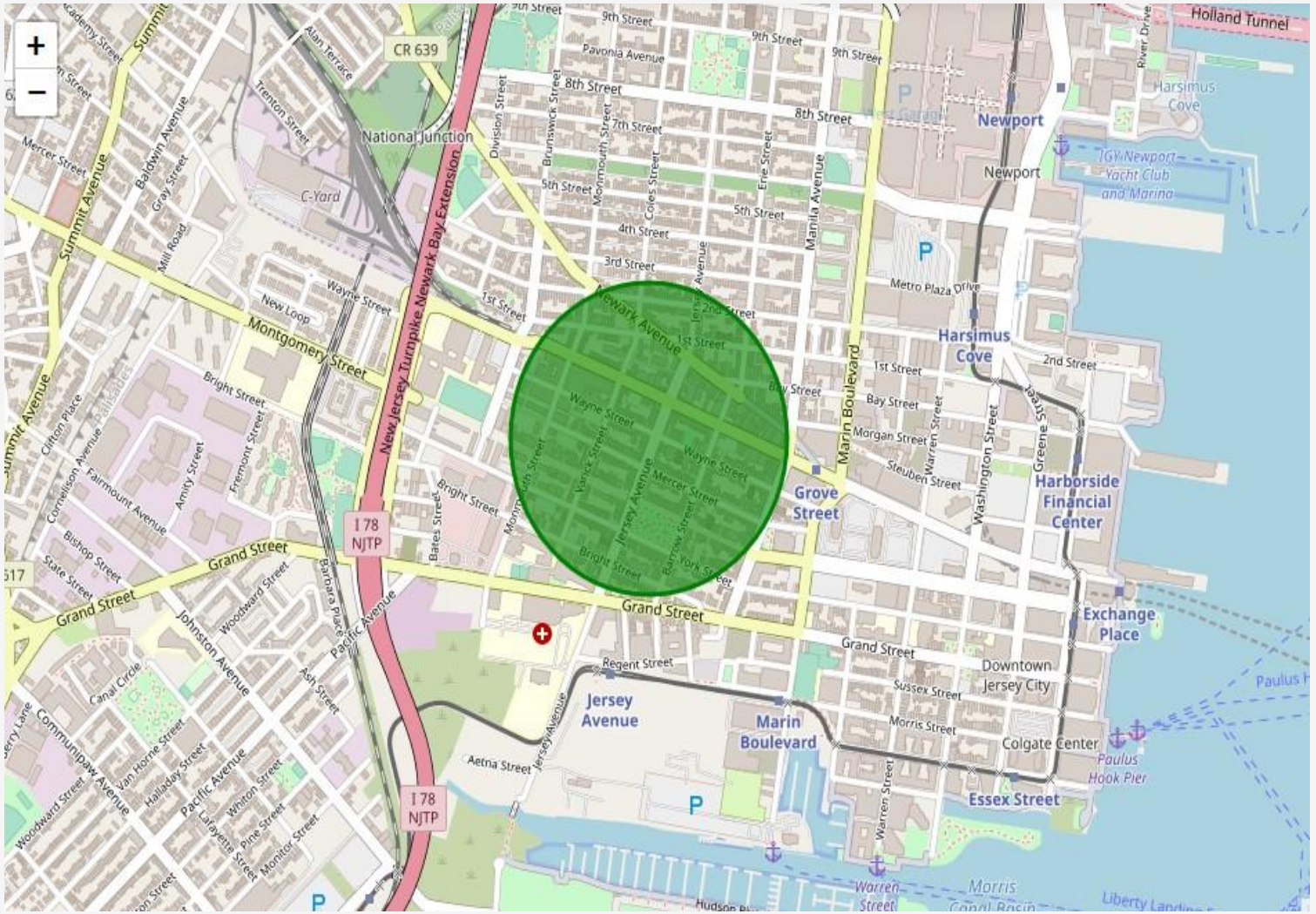
2 Foursquare

Foursquare is a technologic company that built a massive dataset of accurate location data.

Foursquare powers location data for Apple, Uber, Snapchat, Twitter and many more.

Their API and location data is currently being used by over 100,000 developers.

Results



Best City in USA for opening a Shopping Mall/Casino-

NEW JEARSEY

Best Locality in New Jersey –

Near Jersey Avenue

Discussions and Recommendations

1. In the Foursquare API, we have queried the Venues of a locality by specifying the LIMIT and Radius of our choice. We have chosen less LIMIT as the number of API calls that can be done using a free account in Four Square are less.
   - We can increase the limit for more accurate results.
   - We can increase the Radius for more venue results from each city

2. In the venue categories we are choosing only few out of 2000 that are available to give weights and identify the best cluster. Hence, assigning weights must be done relatively for each category and then considering more number of venue categories would actually yield a better output.