



Candidate neighborhoods for opening new restaurants - Methodology and case study

IBM/Coursera's Applied Data Science Capstone course

Author: Paschoal Molinari

Date: 2021-03-02

Finding good neighborhoods is key for new restaurants

- **The business problem is finding neighborhoods in a specific city that are potentially good candidates for opening new restaurants.**
- **Usually finding a good neighborhood for a new business is a time and resource-consuming effort and has some blind spots of factual data about demand and competition.**
- **This problem affects many companies and entrepreneurs in the restaurant industry who need timely information for decision-making when choosing a new business location.**
- **Foursquare geospatial data can be used to uncover these spots.**

Data Acquisition

- The City of Toronto is the case study
- The Postal Code of Toronto's neighborhoods are from this link:
[https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada: M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)
- Latitude and Longitude Data come from geocoder.
- The geospatial data obtained from Foursquare (foursquare.com) about the different venues (not just restaurants) are engineered, tabulated and new statistical features are elaborated.

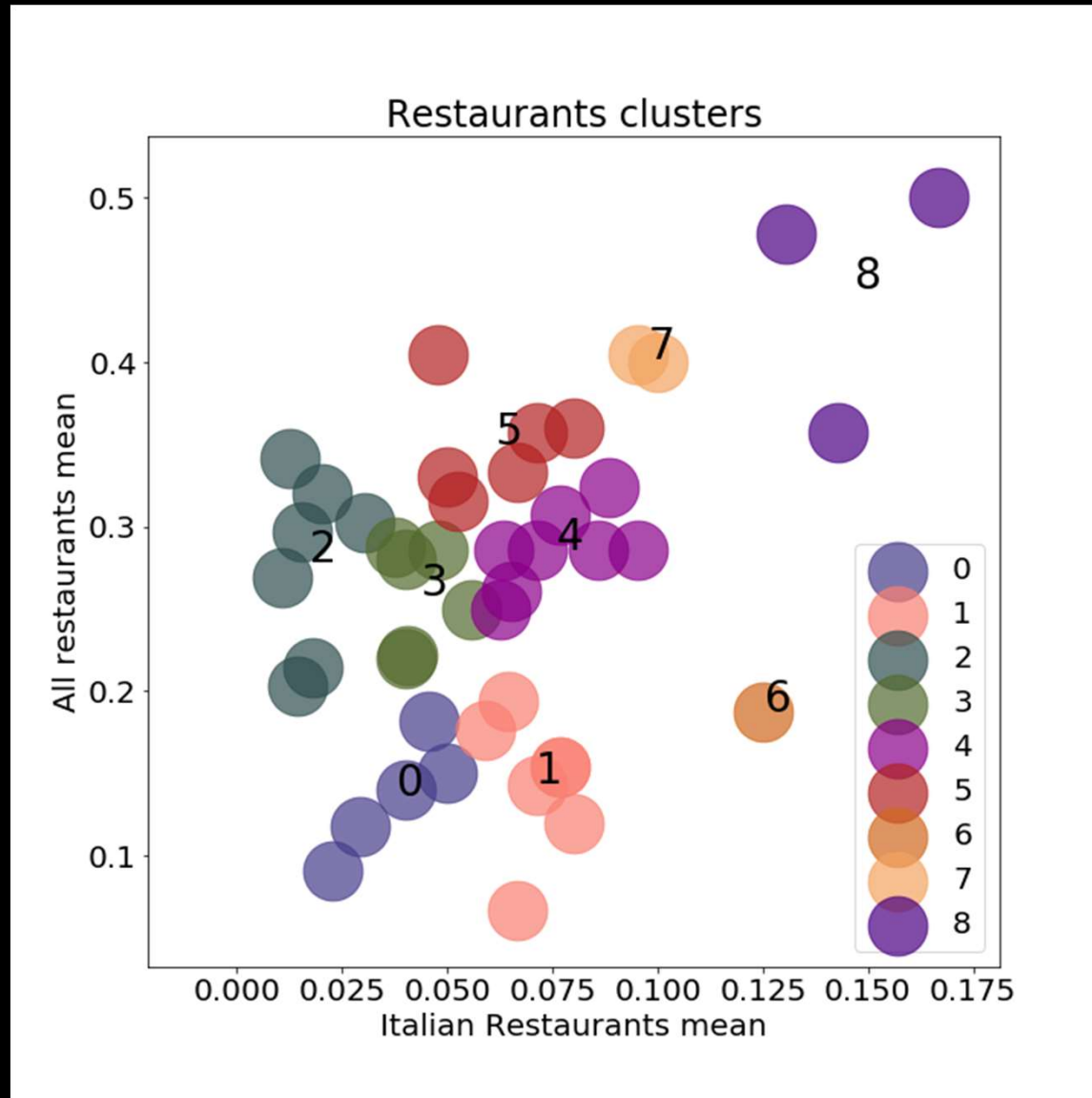
Methodology

Two new statistical features of high relevance are calculated from dozens of features:

- **Mean of restaurants by neighborhood (all categories of restaurants). The lower the mean the higher the need for restaurants in general.**
- **Mean of Italian restaurants by neighborhood. The lower the mean the higher the need for Italian restaurants.**

The Machine Learning K-Means is applied to cluster the neighborhoods using the features above. The final clusters are then sequenced by the size of the vector representing each cluster in ascending order.

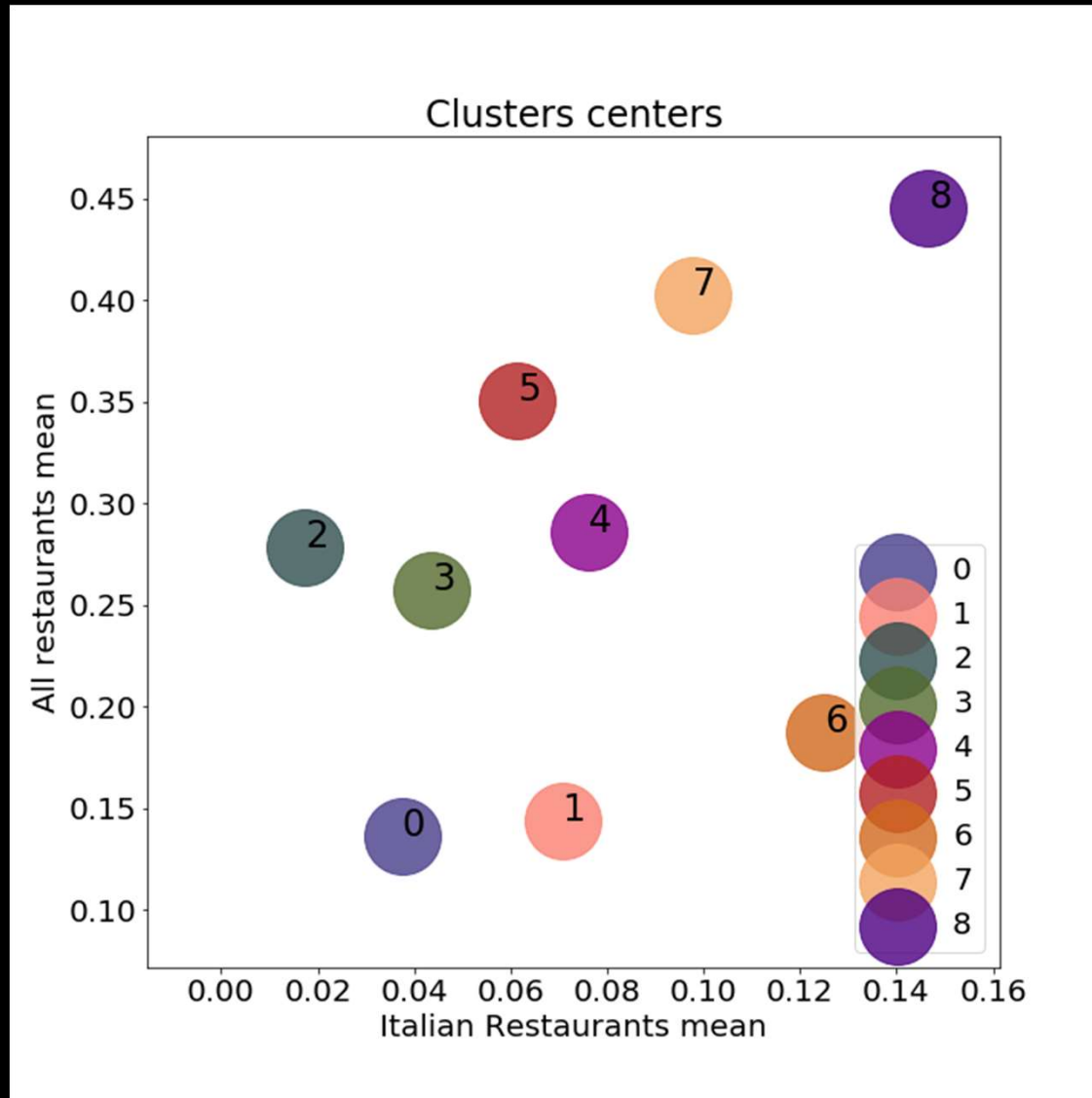
Visualization - Restaurants clusters



In the Toronto's case study nine clusters are generated by K-Means. Its sequenced ascending (0 to 8).

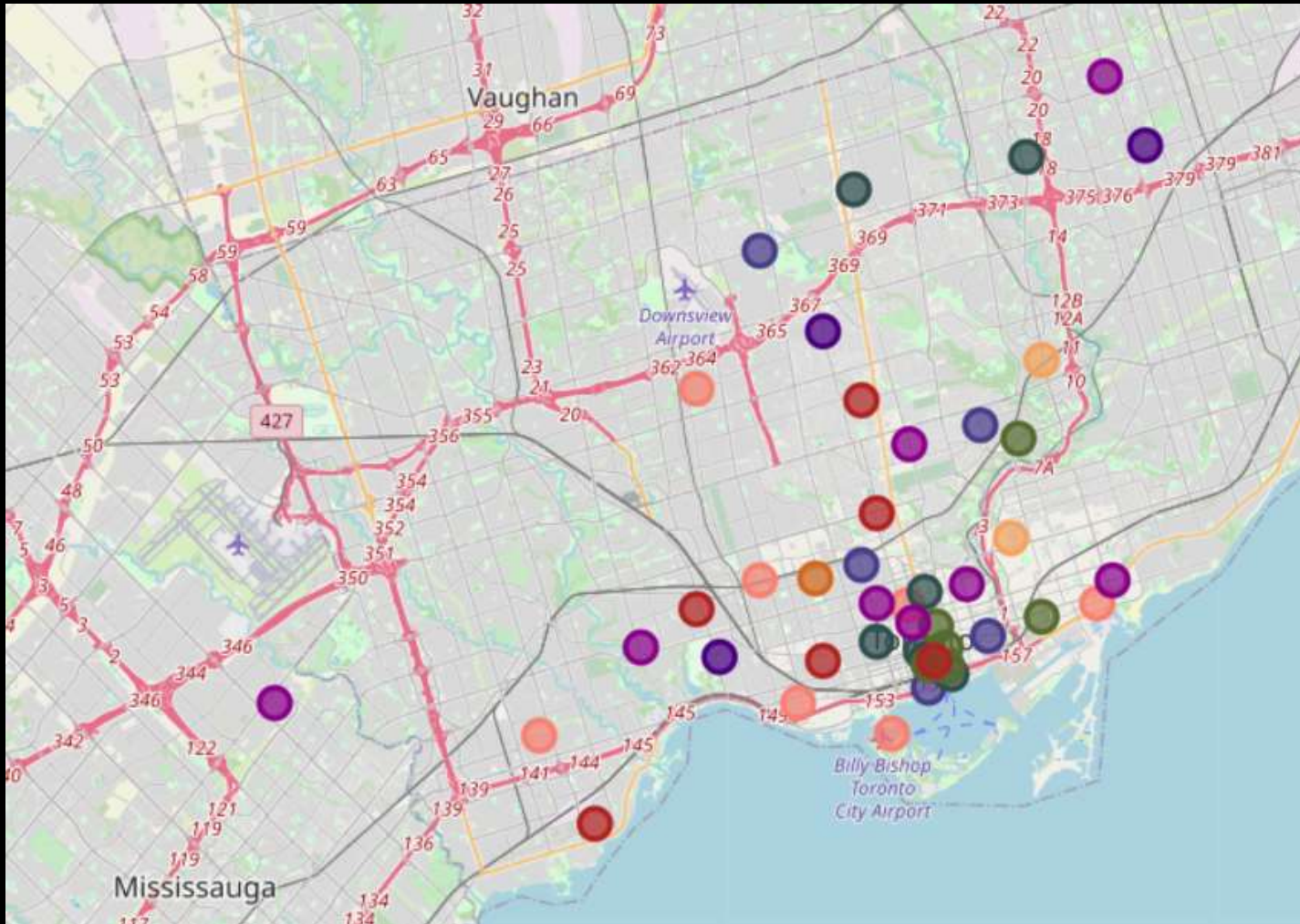
The clusters 0, 1, 2 and 3 are the most promising good candidates.

Visualization - Clusters centers



The clusters' centers give visual confirmation of the clusters' ascending sequence. The sequence gives them meaning.

Visualization - Geospatial Map of Toronto's clusters



The nine clusters are geographically sparsed.

Conclusion

Key aspects:

- The updated geospatial data from Foursquare is definitely of high quality and high value
- Machine Learning K-Means for clustering is very fast and practical to discover important information in statistical features
- The vectorization of the clusters' features' average is an effective way to sequence the clusters and give them meaning.

The objective of finding potentially good candidates neighborhoods to open a new restaurant was successfully achieved and is of high relevance to the business decision-making process.