The Battle of the Neighborhoods (Week 2)

Applied Data Science Capstone by IBM/Coursera

Summary

Introduction: Business Problem

- ABC Lifestyle Food
- healthy (non-alcoholic) drinks and light food
- Customers are with different ethnicities, different ages and different education backgrounds
- Successful store in Manhattan
- Wants to expand in Toronto, Canada

Data

- The Neighborhood of current store in New York, "Lincoln Square"
- New York neighborhood list data.
- Toronto Wiki Page, for zip codes and their borough and neighborhood names.
- Folium data for visualizing neighborhoods.
- Foursquare data for common venues of given neighborhood.

Methodology

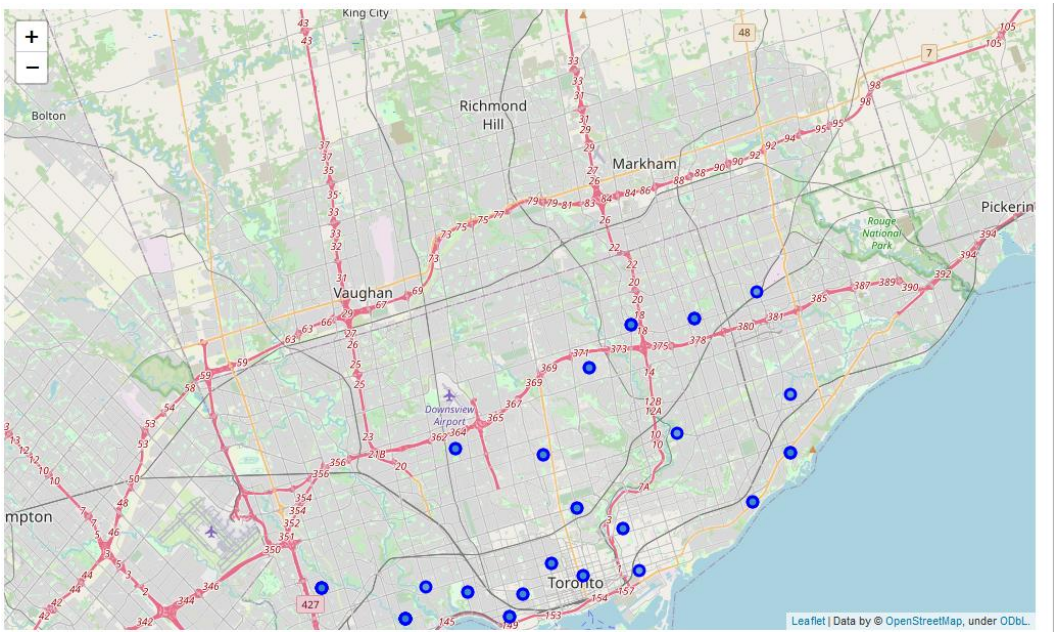
- Connect neighborhood data with Foursquare data
- Convert data to neighborhood – venue – category data
- Build cluster model using neighborhood – venue – category data on Manhattan
- Apply the model to Toronto neighborhood – venue – category data
- Find, in Toronto, the neighborhoods with the same cluster number as the neighborhood of “Lingcoln Square”

Analysis

1. download New York neighborhood
2. visualize NY neighborhoods
3. Extract Manhattan neighborhoods
4. Attach venues from Foursquare data
5. using one-hot, create neighborhood-venue-category data
6. download Toronto postal code-neighborhood data
7. Cleaning Toronto data
8. Create Toronto neighborhood-venue-category data
9. Subset both data with common categories
10. Aggregate data to one row each neighborhood
11. Build a K-means model on Manhattan data
12. Find cluster number for “Lincoln Square”
13. Apply model to Toronto data
14. Find neighborhoods with the same cluster number as “Lincoln Square” in Manhattan.

Results and Discussions

- The result is a list of neighborhoods in Toronto.
- Those neighborhoods are similar to “Lincoln Square” based on venue profile.
- The ABC Lifestyle Food can confidently use this list for their 2nd store.
- Additional features can be added, in real world, to make model better



Conclusion

- The consulting company successfully used machine learning technique
- Helped the ABC Lifestyle Foods find a list of suitable neighborhoods in Toronto
- This technique is useful for many companies thinking about their next expansion.
- This technique is different from traditional method,
- The machine learning method employed in this project is superior to the traditional method.