

Identifying Cafe Density for Coffee Bean Distribution

Applied Data Science Capstone Project Report

Locating Where Coffee is Sold

- Coffee is product with universal appeal.
- Businesses in the coffee industry might reasonable expect a steady demand in future years.
- With limitless options, finding the ideal location for a coffee bean supply operation is an essential study.
- Finding places where coffee is sold and shops that might affect coffee sales is inherently interesting to potential coffee bean suppliers and other potential stakeholders:
 - Café owners
 - Real estate investors
 - Residents
 - Tourists

Data Extraction & Features

- **IBM Developer Skills Network:**

- Borough
- Neighborhood
- Latitude
- Longitude

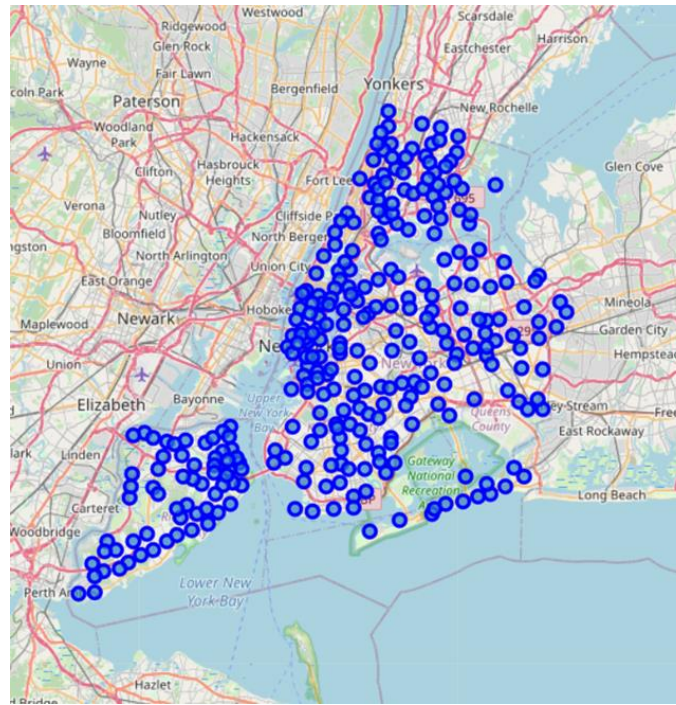
- **Foursquare:**

- Neighborhood
- Neighborhood latitude
- Neighborhood longitude
- Venue
- Venue Latitude
- Venue Longitude
- Venue Category

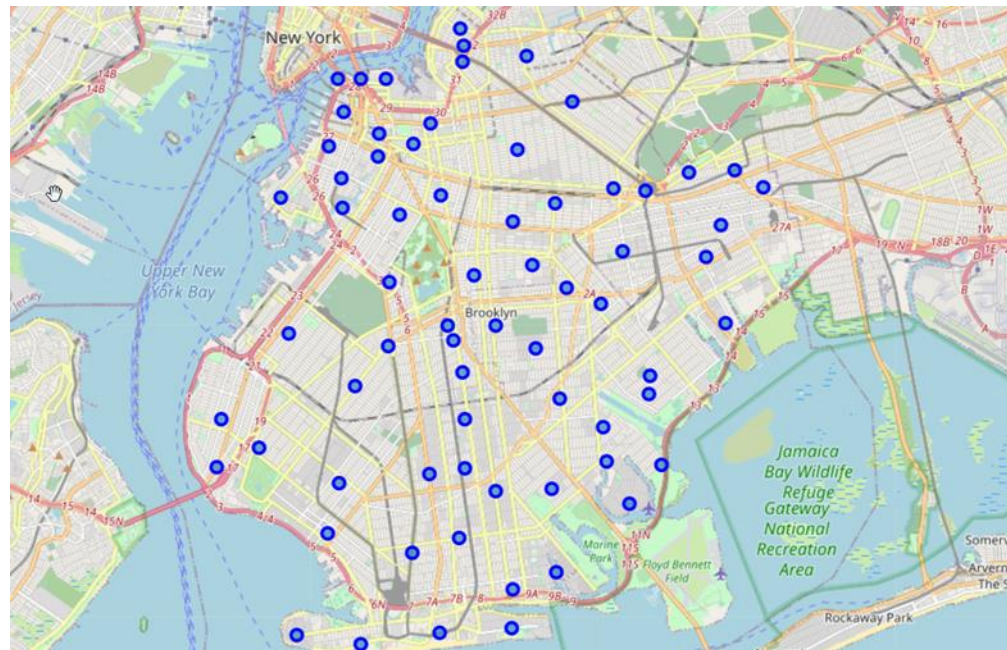
	name	categories	lat	lng
0	Pilo Arts Day Spa and Salon	Spa	40.624748	-74.030591
1	Bagel Boy	Bagel Shop	40.627896	-74.029335
2	Georgian Dream Cafe and Bakery	Caucasian Restaurant	40.625586	-74.030196
3	Leo's Casa Calamari	Pizza Place	40.624200	-74.030931
4	Cocoa Grinder	Juice Bar	40.623967	-74.030863

- **5 boroughs, 306 neighborhoods, 291 venue categories**

New York Geographical Data Visualization



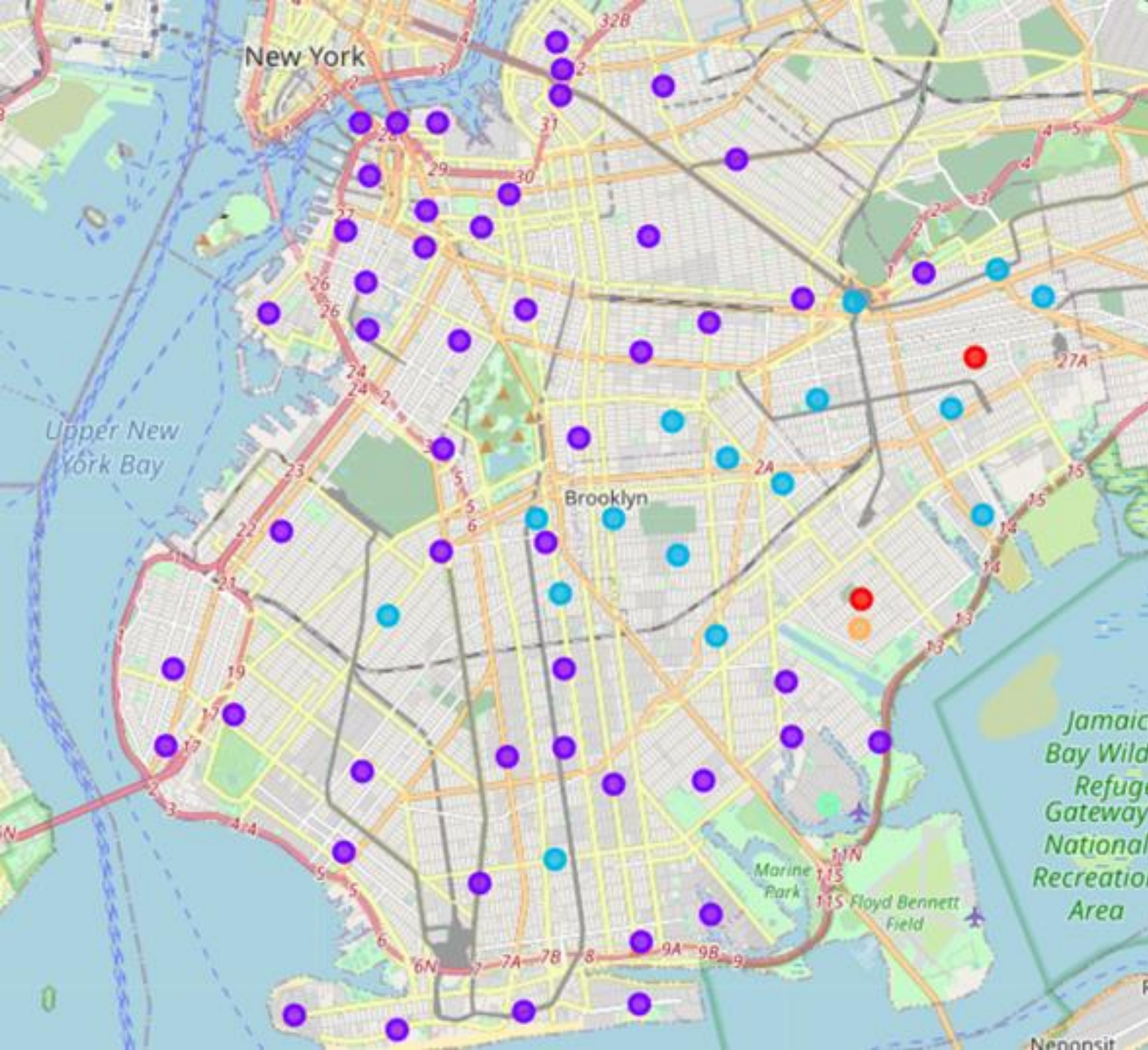
- *Folium* used to plot the geographic coordinates of NYC's neighborhoods (left)
- *Visualization of the dataframe exclusively for Brooklyn (below)*



	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Bath Beach	Pizza Place	Chinese Restaurant	Pharmacy	Peruvian Restaurant	Bubble Tea Shop	Donut Shop	Fast Food Restaurant	Gas Station	Italian Restaurant	Women's Store
1	Bay Ridge	Spa	Italian Restaurant	Greek Restaurant	Pizza Place	American Restaurant	Bar	Bagel Shop	Hookah Bar	Sushi Restaurant	Sandwich Place
2	Bedford Stuyvesant	Coffee Shop	Deli / Bodega	Pizza Place	Café	Bar	Tiki Bar	New American Restaurant	Playground	Boutique	Fried Chicken Joint
3	Bensonhurst	Chinese Restaurant	Italian Restaurant	Sushi Restaurant	Ice Cream Shop	Donut Shop	Shabu-Shabu Restaurant	Cha Chaan Teng	Liquor Store	Noodle House	Smoke Shop
4	Bergen Beach	Park	Harbor / Marina	Athletics & Sports	Playground	Baseball Field	Fast Food Restaurant	Event Space	Factory	Falafel Restaurant	Farm

Brooklyn Dataframe Sample of Most Common Venue Types by Neighborhood

- Head of the new dataframe created solely for the venues within a 500 meter radius of each of Brooklyn's neighborhoods
- One hot encoding used to evaluate the mean frequency of occurrence of venue categories



Clustering Neighborhoods

- *k-means* was ran to cluster the neighborhoods of Brooklyn into five different clusters.
- The results revealed fairly distinct neighborhood characteristics, some containing parks and marinas, with others containing spas and shabu-shabu restaurants.
- Cluster 1 was identified as an opportunistic group of neighborhoods for the possible opening of a coffee bean supply business.
- Additional details of the clusters are located in the project report.

Results & Conclusion

- **Findings:**

- Three of the five clusters contained neighborhoods with niche geography and possible demographics
- Cluster 2 consisted of more ethnic-themed, dinner type restaurants and not very many places to drink or purchases coffee
- Cluster 1 was comprised of the most venues catering to customers desiring a more upscale leisure style of retail space including cafes, spas, boutiques, and bars.

- **Conclusion:**

- Cluster 1, located primarily on Brooklyn's north, west, and south sides, would be a reasonable place to start a search for the placement of a coffee bean supply business for proprietors looking to operate in the vicinity of places that sell coffee or stores that might affect coffee sales.
- Data contained in the study mainly considered quantitative data, with the absence of other possible input including residential sentiment and observational trends affecting communities.
- Report is open for additional models and analysis for future expansion of insight.