Neighborhood Clustering - Dallas, Texas, USA

If someone is looking to open a restaurant, where would you recommend that they open it?

2. Data

In order to do the research in this report, data will be gathered through the following sources:

1. For this analysis I will be using a New York City dataset that has data on all five boroughs of New York City as well as longitude and latitude of each neighborhood. The dataset allows me to explore the neighborhoods in Manhattan, NYC. Below is the data sample:

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Manhattan	Chinatown	40.715618	-73.994279
2	Manhattan	Washington Heights	40.851903	-73.936900
3	Manhattan	Inwood	40.867684	-73.921210
4	Manhattan	Hamilton Heights	40.823604	-73.949688

2. In order to identify all the coffeeshops in Manhattan neighborhoods I will be using Foursquare location dataset. Below is the data sample:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
38	Hudson Yards	40.756658	-74.000111	Oslo Coffee Roasters	40.760512	-73.998985	Coffee Shop
58	Hudson Yards	40.756658	-74.000111	Blue Bottle Coffee	40.753846	-74.002250	Coffee Shop
61	Hudson Yards	40.756658	-74.000111	Romeo and Juliet Coffee	40.760726	-73.997724	Coffee Shop

3. The Foursquare API will be used to collect information regarding neighborhoods with a smallest number of existing coffeeshops with the help of a new column ‘frequency’. Below is the data sample:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	frequency
38	Hudson Yards	40.756658	-74.000111	Oslo Coffee Roasters	40.760512	-73.998985	Coffee Shop	3
58	Hudson Yards	40.756658	-74.000111	Blue Bottle Coffee	40.753846	-74.002250	Coffee Shop	3
61	Hudson Yards	40.756658	-74.000111	Romeo and Juliet Coffee	40.760726	-73.997724	Coffee Shop	3