Capstone Project – Coffee Shop in Singapore

Description of Data

The data used to solve this problem is geolocation data collected from FourSqure.com. Foursquare has one of the largest database of 105+ million places and used by over 125,000 developers. Foursquare API will provide many categories of the venue data, we are particularly interested in the coffee shop category in order to help us to solve the business problem. Adequate explanation and discussion of the data is the following. Data is a single dataframe, containing at least a location of the coffee shop. Explanation of the location data is a standard tuple (lat, lng), where lat stands for latitude and lng for longitude. Some other metadata such as name, postal code, and etc., are also collected.

Example of the data:

Identifier	Name	Short Name	Address	Postal Code	Latitude	Longitude
5d170ae66d54f8002357e288	PPP Coffee	Coffee Shop	#02-19 Funan (107 North Bridge Road)	179095	1.291628	103.849741
52fddabd498e48893061a262	Starbucks Reserve Store (Starbucks Reserve)	Coffee Shop	#02-01/02/03 The Fullerton Waterboat House	049215	1.287049	103.853600
527ba4d1498e06c5aa18aaca	RONIN	Café	17 Hongkong St	059660	1.287708	103.847177
55880c7b498eef1e8810c90d	Hoshino Coffee	Café	#B2-55 Capitol Piazza (13 Stamford Road)	178905	1.292988	103.851441
5b348bd80d8a0f002cb3f0ed	The Stamford Brasserie	Café	2 Stamford Rd	178882	1.292836	103.853693

Data will be used as follow – by knowing the locations of the already existing coffee shops. It is possible to apply unsupervised learning technique like kernel density estimation (KDE) to determine the area of influence of the existing coffee shops.