Coursera - Applied Data Science Capstone - Submission 1

Introduction/Business Problem

I want to move to a neighborhood in Boston and open a bar business that specializes in microbrewery beers. I have not decided on a neighborhood and want to compare and contrast different neighborhoods of Boston in terms of how many bars exist already so as to assess competition and increase the likelihood of business success. Later when the bar is established, I want to offer a food menu and need to explore the type and quantity of existing food establishment in these neighborhoods. I want to figure which neighborhoods are similar to each other and what is the bar scene in those already as input to my decision to locate a bar in one of these neighborhoods.

How data will be used to address the problem.

As I am not already familiar with Boston, I have to gather a list of the neighborhoods and their respective location data in terms of latitude and longitude. This data will be used to feed into the FourSquare API in order to find features regarding the respective neighborhoods including the bar and restaurant data. I want to profile the different neighborhoods for their existing bar types and counts. Using a K-means clustering algorithm on this data will group similar neighborhoods together. Ideally, I would be able to use other factors such as commercial rental rate data and median income in the different neighborhoods to give more information regarding factors in any decision to locate a new bar.