Data

To address the question, we will use the following data:

- List of neighborhoods in Brooklyn, NY. This is the area of focus of the specific retailer – identifying the optimum location in Brooklyn to open the new grocery store/restaurant is the project scope.
- Latitude and longitude coordinates of these neighborhoods. This is required to plot the maps and get the venue data.
- Venue data, more specifically data related to grocery stores and restaurants. These data will be used to perform clustering analysis on the neighborhoods mentioned above.

Sources of data/manipulation techniques

- The dataset exists for free on the web: https://geo.nyu.edu/catalog/nyu_2451_34572. It contains a total of 5 boroughs and 306 neighborhoods. We will be based on the lab instructions to retrieve the data of New York and then we will focus on Brooklyn.
- We will use geopy library to get the latitude and longitude of Brooklyn.
- We will create a map(folium) to visualize Brooklyn and its neighborhoods.
- The we will get the geographical coordinates of the neighborhoods in the area.
- Foursquare API keys will be used to get the venue data for those neighborhoods along with many categories of the venue data. We are rather interested in grocery stores and restaurants around to solve the business problem.

What follows

Data section is followed by Methodology section where we will discuss in detail the steps taken in this project, from data analysis to machine learning techniques (K-means clustering) that were applied to finally identify the suggested location.