**A description of the problem and a discussion of the background.**

My company has been contacted by XYZ Hospitality group who is looking to expand their business and open a new restaurant in New York City. XYZ Hospitality Group specializes in Italian and French Restaurants and they report they have greater success when opening a restaurant when there is less competition from similar cuisines within the surrounding area. The client also reports that they have had greater success with a property that has catering opportunities nearby including event spaces, museums or large offices where lunch contracts can be created.

Opening a restaurant is a very risky and difficult task where a majority of newly opened restaurants will fail within their first year due to lack of planning. This is why proper planning and analysis is vital for an increased probability of success. The old adage “Location, Location, Location” could not be truer when it comes to opening a restaurant especially in the city of New York. The use of data analysis can ultimately result in a more successful business both by selecting a location with less competition and more catering opportunities based off of surrounding venues.

**A description of the data and how it will be used to solve the problem.**

For my analysis I will be using date obtained from NYU that provides centroids of the neighborhoods throughout New York City entitled “New York City Neighborhood Names”. This dataset will be used to identify and segment neighborhoods within New York City to help identify an area of promise to open a potential restaurant in one of the five boroughs of New York City. I will then utilize foursquare data in order to identify different venues in the different neighborhoods of New York City to determine if there is significant competition in the area or if there are catering/partnering opportunities within the area that would be beneficial for the client (XYZ Hospitality group). I will cluster my data into 5 different clusters that will further demonstrate potential competition within the surrounding area that would be detrimental for the client to invest in a certain cluster.

Here is an example of the data set obtained:

