**Guiding Question**

This project was done to fulfill the requirements of the client, who approached “Group-2 Consulting” seeking recommendations for extending the boundaries of their bakery business. The client wished to know the current trends in the bakery market. The client was a highly mobile company, which meant that geographic analysis was relevant, including locations where bakeries are most common, or perform better or more poorly. Since every business must cater to customer requirements, the client asked us to further analyze the customers’ expectations and experiences with bakeries, including an analysis of featured attributes and a sentiment analysis of customers’ reviews about their experience. The client was very particular in knowing the performance of the reputed bakeries in the market, the items and the selling points of their competitors and the areas where their competitors performed poorly. The client was keen in understanding the check-in patterns of the customers, for instance, which day of the week attracted more customers? The client expected that the answer to all the above questions would guide them to know the pulse of the market and the customers so that they can deliver what is most attractive to customers, and therefore most profitable.

We at “Group-2 Consulting” planned to approach this case by performing a descriptive analysis of the datasets pertaining to bakery business using Python programming in ipython notebook. We were looking at data related to the transactions in the bakeries in the US, Canada, UK and Germany, which contained many variables including name, city and state of each bakery, time and date of customer check-ins, customer reviews, and customer survey responses about attributes featured in each bakery. It would have been ideal to also have access to information like each business’ cost and revenue data; customer age, sex, and other demographic information; and more information on the business, such as urban vs rural settings, franchise or independent, or size of the establishment. Such information would have helped us in devising a better marketing strategy and recommendations for our client. Unfortunately, we did not have access to any of this information. Nonetheless, using the data we acquired from the Yelp website, including data related to the Food & Restaurant industry, the business establishment details, the customer check-in details, the customer/user review and tips data and the photos that the users posted, we were convinced after initial descriptive analysis of the Yelp dataset that we can cater to the requirements of the client from the available data.

The Yelp dataset is not comprehensive by itself, the files were in JSON format and they need to be converted to csv to analyze them. The files were to be normalized for a better performance of the datasets. Despite all these shortcomings, we were able to arrive at required solutions.

After downloading the files in JSON format and converting and normalizing them for better performance given our limited hardware capacity (unfortunately this project was operating on a budget of exactly $0), we were able to conduct the analyses explained above. We based most of our findings and recommendations on business data, check-in data, and customer review and ratings.