IBM DATA SCIENCE CAPSTONE PROJECT REPORT
Part 1
An Analysis of Boston, MA Neighborhood Restaurant Data

Intro/Business Problem

The city of Boston, Massachusetts is a small city on the northeast coast of the United States. The city is well known for its many cultural and historical attractions that draw in approximately 20 million visitors per year from both the United States and abroad. Additionally the city is home to thousands of college students for much of the year.

The city itself is not very large area-wise and the tourist attractions tend to be contained in one small area of the city with more residential neighborhoods around that. The tourist areas do have some residential areas but these tend to be very upscale and expensive, such as: the North End and Back Bay. The residential areas outside of that zone tend to be more affordable and are where middle-class and college students tend to reside, such as: Allston and Roxbury.

I wish to do an analysis of these neighborhoods to see if there is a difference between the restaurant types that are most prevalent in the tourist areas versus the residential areas. This would assist someone who wishes to open a restaurant in these areas in determining which type of restaurant may be successful.

Data Description

As required by the assignment I will be using restaurant data gathering via a query to the Foursquare API. Foursquare is an American tech company founded in New York City in 2008. According to their Wikipedia page, "The company rose to prominence with the launch of its namesake local search-and-discovery mobile app, now known as Foursquare City Guide, which popularized the concept of real-time location-sharing and checking-in."

Also used will be zip code (postal code) data for the city scraped from the web page: https://en.wikipedia.org/wiki/Boston#Demographic_breakdown_by_ZIP_Code. From this data I will only use zip code and neighborhood data. I also took a screen capture from here: https://data.boston.gov/dataset/city-of-boston-boundary1/resource/7e6574df-a713-4617-b4f2-f4588fd211d7, but this data is not used in the analysis.