IBM CAPSTONE PROJECT
THE BATTLE OF THE NEIGHBORHOODS (WEEK 1)

Background and Introduction

Problem Description

In today's society there are many ways to shop for everything. These ways include online shopping, in person brick and mortar, and there are services that businesses provide where they will go and shop for the specific items their customers want. Often times it is difficult for a personal shopper or an individual to buy things if they do not know what they specifically want. Even with today's recommender systems that can make recommendations based upon a profile, these systems have their weaknesses. Part of the problem is that there are so many choices and the overwhelming number of choices make it difficult for people to decide. According to this NYTimes article https://www.nytimes.com/2010/02/27/your-money/27shortcuts.html "Research also shows that an excess of choices often leads us to be less, not more, satisfied once we actually decide. There's often that nagging feeling we could have done better."

A city like New York is an enormous city and many times when US domestic visitors or foreigners are looking for a hotel in NYC the choices can be overwhelming. It is stated that there are anywhere from 20,000 to 30,000 licensed realtors in NYC. So not only is the problem of choosing a hotel extremely difficult but finding the right area to stay in adds to the complexity. Often times many tourists that visit NYC don't know anything about the city, but they can often, and without hesitation tell you what it is they like they do. They can very easiy share the types of experiences they like to have, the venues they visit the most, etc.

What I am proposing is to use Foursquare data and code a solution that pulls venue data from Fousquare and builds a dataframe that gives you the top ten venues by neighborhood. From there we can pull the foursquare data that has all of the hotels for that area. This would simplify the arduous task of picking a hotel from the 669 hotels that are currently available in NYC.

Data Description

The data driven solution will incorporate the use of a Jupyter Notebook as the developer environment that I will work from. While utilizing Python as the scripting language, I will leverage the many different Python libraries that handle JSON data, Dataframes, and mapping objects. The Dataframes will mostly contain location data that is extracted through the Foursquare API. Foursquare is a repository of location data that it obtains through interacting with its partners like Snap, Twitter, Google etc. The Foursquare database is rich with comprehensive location data and it will provide all of the data that we need for this evaluation.

Methodology

1. Collect the data
2. Inspect/Review/Understand the data
3. Prepare the data
4. Model the solution