Coursera Capstone
IBM Applied Data Science Capstone

Finding the Optimal Location for a New Museum in New York City

By Abbie Maemoto

July 2020

Introduction

New York City is a melting pot of culture, art, and history. This New York reputation is in large part due to the thousands of historic museums located all across New York. These museums are hubs of tourism, attracting people from all walks of life. In these museums, tourists lose track of time, immersing themselves in the worlds of art, fashion, performance, and science. The world-wide attraction to these museums is an important factor in New York's economy. A popular museum can help sustain nearby hotel and restaurant industries. As most museums are non-profits, the popularity of a museum is largely dictated by the number of members a museum can obtain. Therefore, the location of the museum is of utmost importance to its success. Property developers, in the interest of maintaining museum popularity, must be careful in deciding the location of the museum and thus, in this study, we will look at the optimal locations for the opening of New York's newest museum.

Business Problem

The objective of this study is to identify the best locations in the city of New York City to open a new museum. Through implementing data science extraction and machine learning analysis methods, we aim to answer the business question: If a property developer is looking to build a new museum in New York City, USA, where is the optimal location to open the museum?

Target Audience

The target audience of this study are investors and property developers looking to plan the construction and opening of a brand-new museum in New York City. Finding the optimal location for the museum will help obtain valuable investors for the project. As New York is already home to the second-highest number of museums (approx. 2500) in the United States, finding the best possible location will be critical for being able to compete with the existing museums.

Data

To tackle this project, we will need the following information:

a) A list of the neighborhoods in New York City, NY. This will help confine the focus of our study specifically to the region of New York City which is the largest city in New York that is home to over eight million people.

b) The coordinates (latitude and longitude) of the neighborhoods found in part a

c) Surrounding venue data that can contribute to a further understanding of whether the neighborhood could foster the development of a new museum. This will be used for clustering later in the project.

The list of neighborhoods was found at https://en.wikipedia.org/wiki/Neighborhoods_in_New_York_City which consists of 259 neighborhoods. We will utilize data mining techniques to extract the data from this website, supplemented by Python and beautifulsoup packages. Then, using the geocoder package, we will obtain

the coordinates of the various neighborhoods. Next, we will use FourSquare to analyze the venue data for all of the neighborhoods. This project will consist of various data analysis tools, including web scraping, implementing API through FourSquare, data wrangling, K-means clustering, and visualization with the Folium package.