

Applied Data Science: Capstone

Shopping Mall Location Choice in New York

Charles Xi Yan

27th March 2020

Contents

1	Introduction	1
1.1	Problem Description	1
1.2	Target of the Project	1
2	Data	2
2.1	Source of Data	2
2.2	Skills and Techniques	2

1 Introduction

As the biggest city in New York and one of the most famous tourism destination in the world, New York is always regarded as an ideal location for retail industry, especially for shopping malls. Shopping malls not only provide customers with an excellent platform for dining, shopping, film-seeing and other entertainment activities, but also offer retailers a good opportunity to expand their business and promote their products. As a fact, there are a lot of shopping malls and department stores in New York, say Macy's. However, as a complicated issue, opening a new shopping mall needs quite a serious and careful consideration because of the business running cost, such as the rent and the management cost. To some extent, a relatively ideal location makes the business on the halfway to the success.

1.1 Problem Description

The object of this project is to analyze the neighbourhoods in New York so that finally the best location for a shopping mall is determined. With data science methods and machine learning techniques learnt from IBM Data Science course, we aim to find a solution to such a question: if a real estate developer is planning to open a new shopping mall in New York, where should be recommended?

1.2 Target of the Project

This project offers useful advice especially to property developers or real estate investors for investing shopping malls in New York. This project is opportune as the shopping mall industry in New York, especially in Manhattan, has almost fully developed and some districts are suffering from oversupply of shopping malls. As Financial Times ever reported, New York has passed the peak era of shopping malls. However, the one with eye-catching features is still proved to have a great chance of success.

2 Data

We list the required dataset as follows:

1. List of neighbourhoods in New York.
2. Geographical coordinates of these neighbourhoods, i.e. the latitude and the longitude, in order to plot the map and collect the venue data.
3. Venue data, in order to perform clustering on the neighbourhoods.

2.1 Source of Data

The json data in the link (https://cocl.us/new_york_dataset) contains a list of neighbourhoods in New York with a total of 306 neighbourhoods in 5 boroughs. Wget command and relevant python packages are used to access the data. Geographical coordinates of neighbourhoods are obtained by using Geopy package.

Foursquare API is then used to collect the venue data of these neighbourhoods. As Foursquare API provides various categories of venue data, we directly focus on the shopping mall category for the project object.

2.2 Skills and Techniques

The project involves a big range of data science skills and machine learning techniques, from loading data by wget command, using Foursquare API, data cleaning, data wrangling, to K-mean clustering technique and Folium map visualization. All detail of applying these skill and techniques will be shown in the next section.