

An aerial view of the Manhattan skyline at sunset. The sky is filled with soft, orange and pink clouds, and the sun is low on the horizon, casting a warm glow over the city. The dense cluster of skyscrapers and buildings is visible, with the Empire State Building and other iconic structures standing out. The water of the Hudson River is visible in the distance.

Capstone Project - From Singapore to Manhattan, New York City (NYC)

GOH JIE DA

Scenario and Background

A few weeks ago, a close friend of mine has been offered an attractive relocation package to New York City. However, he is still in a dilemma if he would like to take up the offer, citing various reasons and concerns such as:

- Missing his family
- Forgoing the convenience to his current work place
- Ability to enjoy similar level of comfort in the rental apartment with his given budget
- Uncertainty of having good Chinese cuisines within close proximity to his rental apartment

Based on the above reasons, he asked if I could use a data driven approach to address his concerns and allow him to make a more informed decision.



The Problem



The Data Driven Approach

The Problem

How to find an apartment in Manhattan, NYC, that meets the following conditions:

1. Furnished rental apartment with a private room
2. Monthly rental of US\$5000/month or less
3. Located within 800m to a subway metro station in Manhattan
4. Wide variety of chinese cuisines within close proximity of 1 km radius to his rental apartment

Data and Libraries

Leverage on the following libraries and datasets:

1. Geopy library to obtain the latitude and longitude values of New York City (NYC)
2. Folium to visualise the Geospatial Data
3. Foursquare API to attain the amenities within Manhattan's neighborhood in NYC
4. New York University's Spatial Data Repository for 2014 New York City Neighborhood Names Dataset
5. A python Craigslist wrapper to retrieve listings results from Craigslist site.
6. NYC Transit Subway Entrance And Exit Data from MTA Headquarters, New York City Transit

Data Driven Approach

- Perform spatial analysis and visualisation, tapping on Folium and data obtained by Geopy library, by exploring the neighborhoods in Manhattan in greater details.
- Utilise Foursquare API to obtain the amenities within each Manhattan's neighborhood proximity.
- Implement one hot encoding and subsequently group rows by neighborhood with the mean of the frequency of occurrence in each category (of the amenities)
- Apply K-means algorithm to cluster the neighborhoods and further examine each cluster.
- Identify neighborhoods with Chinese Restaurant being the Top 2 Most Common Venue
- Employ the python Craigslist wrapper to retrieve listings results, that meet the criterias set out in the problem statement, from Craigslist site based on the above neighborhoods identified.
- Compute the distance for each listing results to its nearest subway station using haversine formula
- Lastly, analyse the available listings that are within 800m from its nearest subway station and has a wide variety of chinese cuisines within the proximity of 1 km radius based on the neighborhoods identified to make a more informed decision.

An aerial photograph of the New York City skyline, featuring numerous skyscrapers and buildings. The Empire State Building is prominent in the center. The text "Thank You" is overlaid in a large, bold, dark red font. The background is slightly faded to make the text stand out.

Thank You