Clustering Neighborhoods in New York City and Segmentatio

Introduction to the Problem

once i was walking in the streets of NYC and i came to know that if someone have to find out that hov each other based on some parameters. how would he find it out? so to tackel this problem i came to k i choosed NYC is that the data i needed for the analysis of my project is publically available. i dint had

Solution

here I will onvert addresses into their equivalent latitude and longitude values. Also, I will use the Four New York City. You will use the explore function to get the most common venue categories in each ne group the neighborhoods into clusters. I will use the k-means clustering algorithm to complete this ta visualize the neighborhoods in New York City and their emerging clusters

Way to the Solution

1. Download and Explore Dataset
2. Explore Neighborhoods in New York City
3. Analyze Each Neighborhood
4. Cluster Neighborhoods
5. Examine Clusters

Data Discription

This Neighborhood Dataset has a total of 5 boroughs and 306 neighborhoods. In order to segment th essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each bor coordinates of each neighborhood. Luckily, this dataset exists for free on the web. Feel free to try to fi the link to the dataset: https://geo.nyu.edu/catalog/nyu_2451_34572 For convenience, I downloaded

Use geopy library to get the latitude and longitude values of New York (

In order to define an instance of the geocoder, we need to define a user agent. We will name our agen
capstone project.

Creating Map using folium

Folium is a great visualization library. Feel free to zoom into the above map, and click on each circle m
neighborhood and its respective borough.

However, for illustration purposes, let's simplify the above map and segment and cluster only the neig
original dataframe and create a new dataframe of the Manhattan data.

USING FOURSQURE API

Next, I am going to start utilizing the Foursquare API to explore the neighborhoods and segment them

2. Explore Neighborhoods in Manhattan

Let's create a function to repeat the same process to all the neighborhoods in Manhattan

we found out that how many unique categories can be curated from all the returned venue

3. Analyze Each Neighborhood

----Battery Park City---

SN. venue ---> freq

0 Park ---> 0.07

1 Coffee Shop ---> 0.07

2 Hotel ---> 0.05

3 Gym ---> 0.04

4 Shopping Mall ---> 0.03

4. Cluster Neighborhoods

I used k-means clustering tocluster the Neighborhoods in the new york city.

i run the k-means clustering algorithm with k=5 clusters.

you can play arround it do it by changing the value of k

5. Examine Clusters

Now, you can examine each cluster and determine the discriminating venue categories that distinguis
categories, you can then assign a name to each cluster. that you can see in the code.

RESULT AND CONCLUSION

as you can see in the code that there are 5 clusters. each clusters are having neighborhoods and corr
are 1 to 10 most common places

the clusters are made based on the similarity of teh most common place in the each cluster.