

Introduction/Business Problem

Colleges and Universities employ a large number of young professionals in both administrative and research positions. As careers grow and change, individuals often change schools to align their research interests and continue to grow professionally. Even though a new school and/or program may be a good fit professionally, it's often hard to determine where to live that is close enough to the school but will also align with personal interests and community expectations.

A young professional working at Columbia University has just accepted a new position at Harvard University. She loves her current community in the Manhattan neighborhood of Morningside Heights, but is now looking to move from New York City, NY to Cambridge, MA. Before she moves, she wants to learn more about where she will be living and find a location to live that is (1) similar to her current Manhattan neighborhood of Morningside Heights and (2) has as many Coffee Shops as possible.

To explore her new home and choose a location, she will:

1. Understand the locations of the Cambridge neighborhoods by viewing them on a map.
2. Determine which Cambridge neighborhoods are the most similar to her Manhattan neighborhood of Morningside Heights
3. Identify the number of Coffee Shops within a 500 meter radius in each Cambridge neighborhood.

Data

1. List of New York City neighborhoods with neighborhood name, latitude, and longitude. (This file is available at https://geo.nyu.edu/catalog/nyu_2451_34572)
2. List of Cambridge neighborhoods with neighborhood name, latitude, and longitude. (This file is available at <https://geo.nyu.edu/catalog/harvard-cambridge14cddneighborhoods>)
3. Latitude and Longitude of Cambridge, MA (obtained via geopy library)
4. List of venues with categories (Foursquare API data for New York City, NY and Cambridge, MA)

The data will be used to explore and choose a location as follows:

1. The latitude and longitude of Cambridge, MA (#3) from geopy will be used to generate a Folium map. The latitude and longitudes of Cambridge neighborhoods will be pulled from the list of Cambridge neighborhoods (#2) and points will be added to the existing Folium map.
2. The Borough, Neighborhood, Latitude and Longitude will be extracted from the json files of the list of New York City neighborhoods (#1) and placed into a pandas dataframe. A new dataframe that only contains the information for the neighborhood of Manhattan will be created.
3. The list of Cambridge neighborhoods (#2) and placed into a single pandas dataframe.

4. The Manhattan and Cambridge dataframes will be merged.
5. A list of venues information will be obtained for all neighborhoods in the merged dataframe by using a GET request through the Foursquare API and passing the latitudes, longitudes, and a radius of 500 meters.
6. The category type for each venue will be extracted and a dataframe containing neighborhood, neighborhood latitude, neighborhood longitude, venue, and venue category will be created.
7. The dataframe will be grouped by neighborhood and the frequency of occurrence of each category will be determined.
8. The grouped frequency dataframe will be used to cluster the neighborhoods into 5 clusters using k-mean clustering.
9. The clusters will be searched for the one that contains the Manhattan neighborhood of 'Morningside Heights'.
10. The top 5 most common venues and a count of those venues for each Cambridge neighborhood in the cluster containing Morningside Heights will be printed.
 - a. Note: If none of the similar neighborhoods have Coffee Shops identified as a top 5 most common venue category then this number should be increased until the number of coffee shops can be determined.
11. The neighborhood with the greatest number of coffee shops will be obtained. If more than one neighborhood has this maximum number of shops, other assumptions can be explored.
12. A recommendation will be given based on (1) similarity as determined by clustering, (2) the highest number of coffee shops, (3) other assumptions as needed.