## Selecting a Restaurant Location in Howard County, Maryland

## John Tamer

## December 16, 2020

## Introduction

Howard County is frequently cited for its affluence, quality of life, and excellent schools. Its estimated 2016 median household income of $120,194 raised it to the second-highest median household income of any U.S. county. Many of the most affluent communities in the area, such as Clarksville, Dayton, Glenelg, Glenwood, and West Friendship, are located along the Route 32 corridor in Howard County. The main population center of Columbia/Ellicott City was named second among Money magazine's 2010 survey of "America's Best Places to Live."

As the county continues to expand and the population continues to grow, there are certain, less-populated areas that are underserved by restaurants. Since many people must travel to other locations to dine out and get take-out food, traffic in some areas has greatly increased and the wait times at many restaurants continue to grow.

These factors combine to suggest that certain areas of Howard County constitute prime locations in which to open a new restaurant. However, that decision must factor in the competition as well as the number of potential customers for the new restaurant.

To that end, we propose an analytical approach that examines the main neighborhoods of the county, as designated by zip code, the numbers and types of venues in each, and the population. Based on a clustering of these factors, we will identify the best choice or choices for a new restaurant.

## Data Sources

Howard County maintains an online repository of data about the county (<https://data.howardcountymd.gov/)>. That cite provides the primary information about the neighborhoods and corresponding population numbers. In addition, Foursquare.com, an online service that provides location information, is used to identify the types and number of venues in each of the areas. Finally, the Folium Python library will be used to display maps of the area and to identify the location of clusters.

The data will be used to cluster neighborhoods according to the types of venues located in them as well as the population. The goal will be to find the cluster with the fewest restaurants and the highest population.

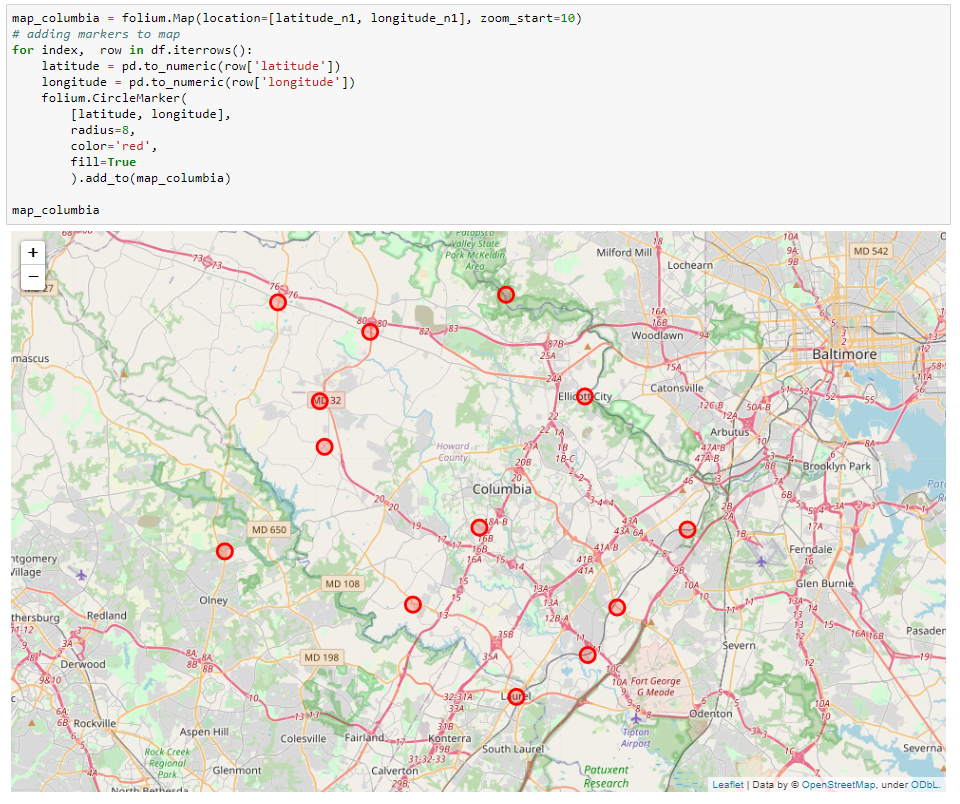
## Import Required Libraries

Important libraries include:

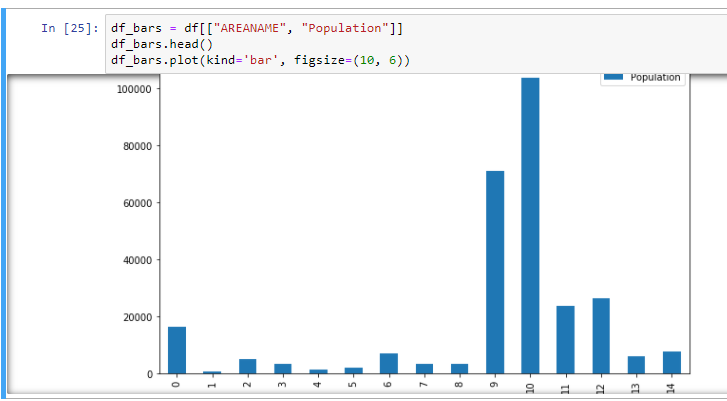
* Requests: for making API calls and receiving results
* Geocoder: for geocoding addresses
* Folium: for creating and displaying maps
* Sklearn: Scikit-Learn library for machine learning algorithms and utilities

## Methodology

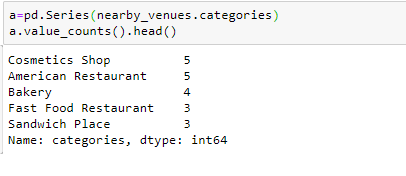
## The methodology employs a combination of geospatial analysis and machine learning. First, we select a geographic location and identify the neighborhoods contained therein. The map below shows the neighborhoods in Howard County, Maryland,



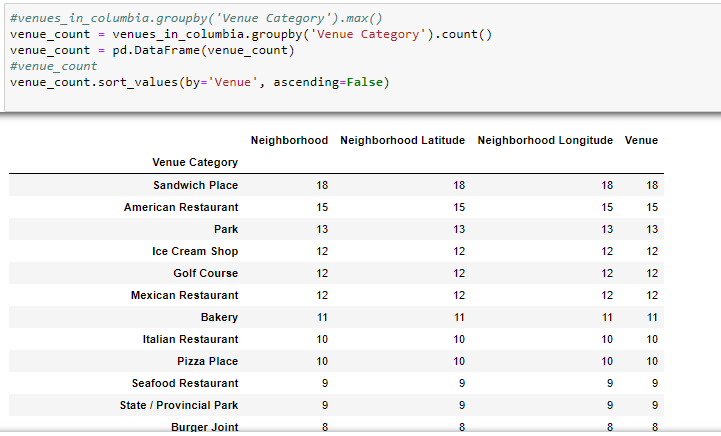
Next, we incorporate population information. One of the main determinants of where to location any business is the size of the population from which to draw customers. We can see from the bar chart below that two of the neighborhoods, Ellicott City and Columbia, are the most highly populated.



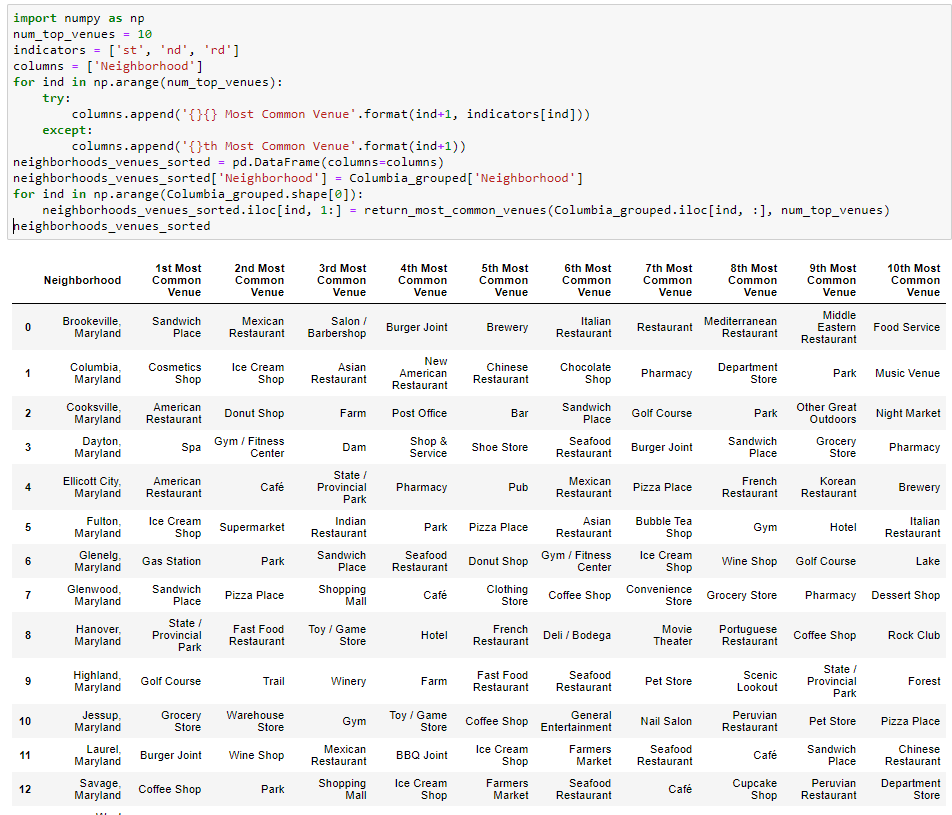
Next, we use the Foursquare API to find all venues within a radius around the center of our location. As a part of our exploratory data analysis, we determine the most frequently occurring categories.



As a preliminary answer to the business question, we can see that sandwich places, American restaurants, Mexican restaurants, Italian restaurants occur the most out of the restaurant categories.

These would not be good category choices for a new venue.

Next, we organize the categories by neighborhood and determine the top ten most frequently occurring categories in descending order.

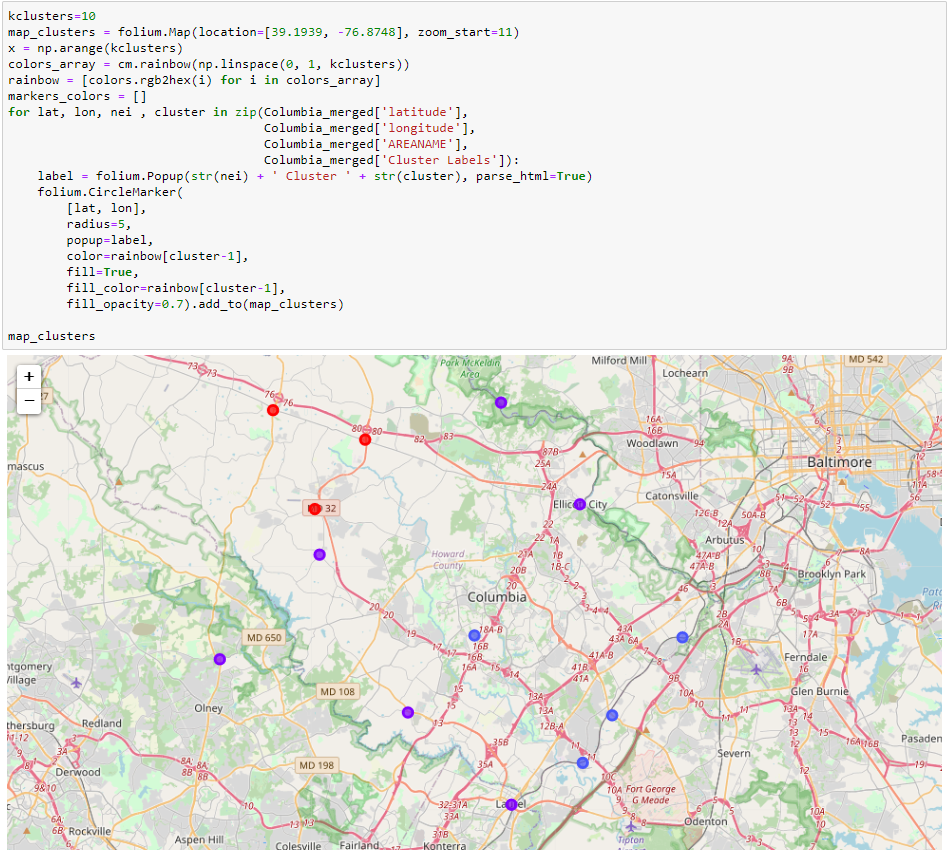


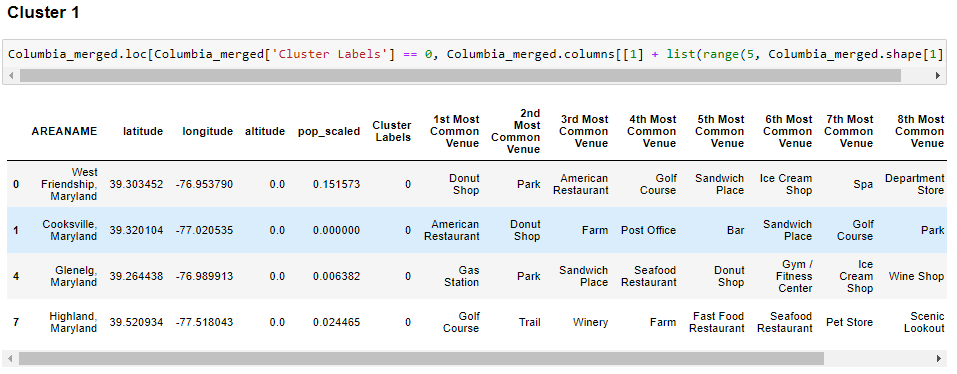
Next, we execute the K-Means clustering algorithm to group neighborhoods according to their venues and their population.

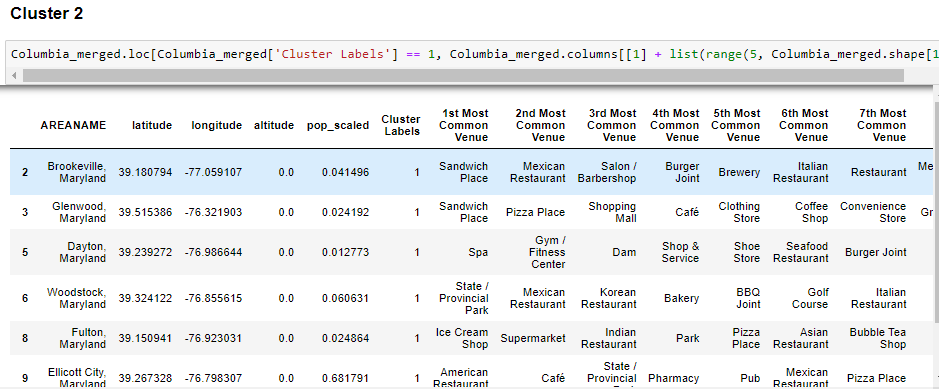
## Results

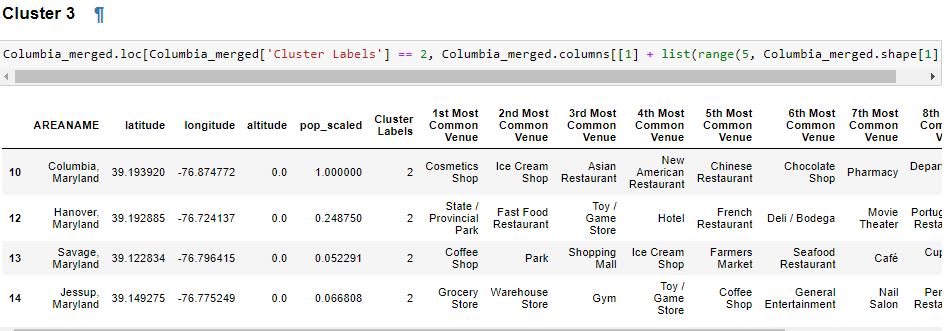
We can see from the map below, that there are three clusters indicated by colors of their location circles. We can also display the neighborhoods and their respective venues for each of the clusters.

Cluster 2 is not a good choice since most of the top occurring categories are restaurants. Cluster 1 is similarly not the best choice given the frequency of restaurants in the three or four top occurring categories. Cluster 3 represents a good choice. Other than a coffee shop and ice cream shop, there are no restaurants in the top occurring category and very few in the next several top occurring categories. We also see that Cluster 3 does contain the highest scaled population value as well.









**Discussion**

As a preliminary analysis, the above provide us with a good indication of possible neighborhoods within which to locate a new restaurant. Other factors that might eventually be considered and included in the analysis are the age distributions of the neighborhoods and perhaps the crime rates.

**Conclusion**

The purpose of this exercise is to determine a neighborhood in Howard County, Maryland in which to open a new restaurant. The select a candidate we located each of the neighborhoods on a map, used the Foursquare API to find nearby venues in each neighborhood. We also determined the most frequently occurring categories in each neighborhood. We used the k-Means clustering algorithm to group neighborhoods according to the types of venues and the scaled population values.

Three clusters were created by the algorithm. Cluster 2 is not a good choice since most of the top occurring categories are restaurants. Cluster 1 is similarly not the best choice given the frequency of restaurants in the three or four top occurring categories. Cluster 3 represents a good choice. Other than a coffee shop and ice cream shop, there are no restaurants in the top occurring category and very few in the next several top occurring categories. We also see that Cluster 3 does contain the highest scaled population value as well.