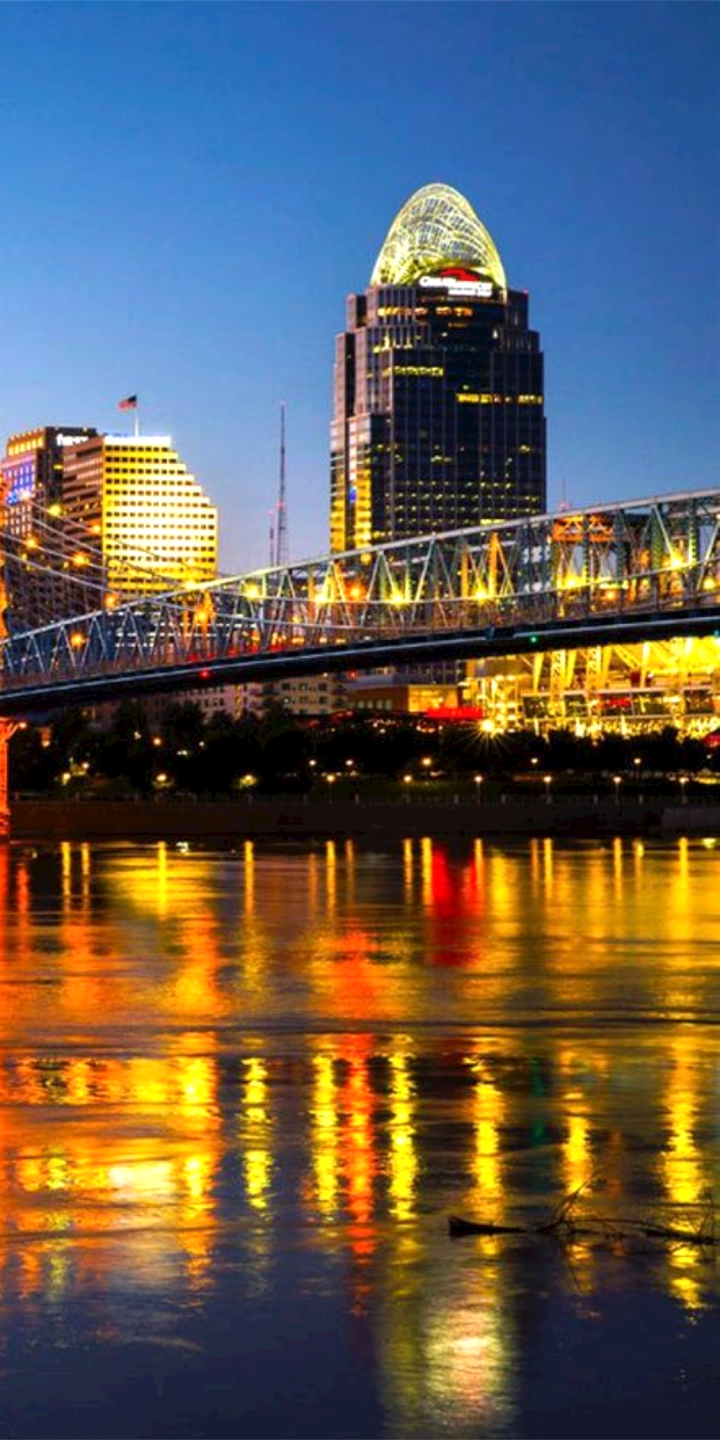


AUTHOR: SANJANA ARE

The Taste of Cincinnati

An Exercise in K-Means Clustering

Background and Problem

Known as the “Queen City of the West,” Cincinnati, Ohio is home to a diverse array of global cuisines. From the renowned American Tom & Chee to the homely French Graeter’s, one is hardpressed to find a more culturally rich metropolis in the Midwest. In fact, thanks to a recent revitalization of the restaurant scene, USA Today has consistently named Cincinnati as ”one of six small cities with big food scenes“ since 2012. Furthermore, Cincinnati’s food scene ranks among the nation’s most affordable, with the average price of a three-course meal for 2 at a mid-range restaurant at only $40, second to only San Antonio, Texas at $35. With new restaurants taking hold every year, the city’s food scene shows no signs of slowing down.

For the tourist, or even city native, Cincinnati’s food scene can often be quite overwhelming. With a myriad of restaurants from different cuisines to choose from, it is hard to objectively determine which restaurants are the most worth visiting. In this paper, I hope to distinguish the upper echelon of restaurants using a machine learning model.

This classification is of interest to many different stakeholders, namely tourists and city natives who are seeking to explore Cincinnati’s bustling food scene. Furthermore, such a classification is also of interest to food critics whose livelihoods depend upon providing scathing critiques of renowned restaurants. Finally, this classification is of interest to restaurants themselves, who gain significant exposure by being in the upper echelon.

The Data

The raw data used in this analysis includes 64 restaurants within 1000 meters of the heart of Cincinnati (geographical coordinates 39.1014537°, -84.5124602°). Each restaurant's name, latitude, longitude, unique ID, and venue type is recorded upon the initial read of the Foursquare API. Upon further querying, the number of Foursquare users who have liked a restaurant is also recorded. The number of likes and the venue type are then utilized to categorize each restaurant into a different cluster based on a k-means clustering algorithm. The latitude and longitude of each restaurant is utilized to map each cluster using the Folium package.

Data was collected using the Foresquare API, a free tool that allows developers to access location-based experiences with diverse information about venues, users, photos, and check-ins.

Methodology

1

Imports

Imported geopy, sklearn, folium, and json.

2

Foresquare API Setup

Instantiated Foursquare API credentials (client server, client ID, version ID).

3

Initial Data Collection

Used Foursquare API to collect data about 100 venues within 1000 meters of the heart of Cincinnati (39.1031 N, 84.5120 S).

4

Advanced Data Collection

Used Foursquare API to collect data about Foursquare users who liked venues. Filtered venues into broad food categories.

5

K-Means Clustering

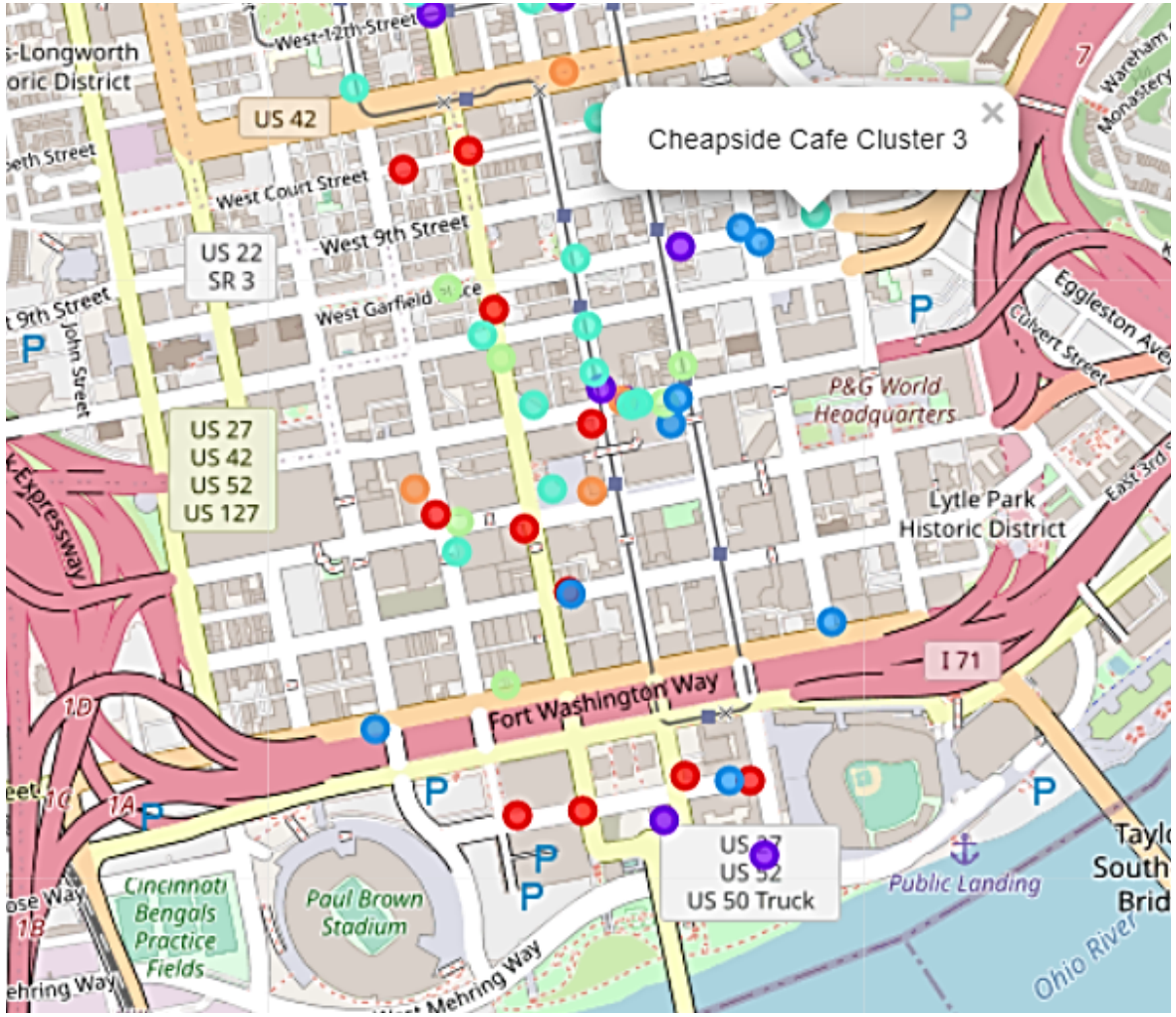
One-hot encoded data to prepare for clustering. Plotted MSE vs. k-value to determine 6 as optimal k-value.

6

Data Visualization

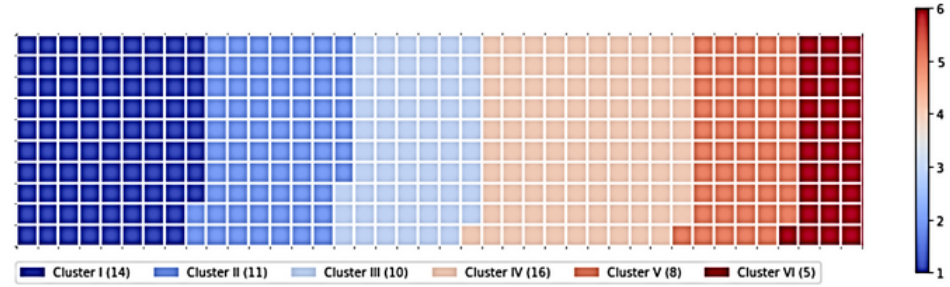
Created Folium map to visualize clusters. Created waffle map to visualize distribution of venues in each cluster.

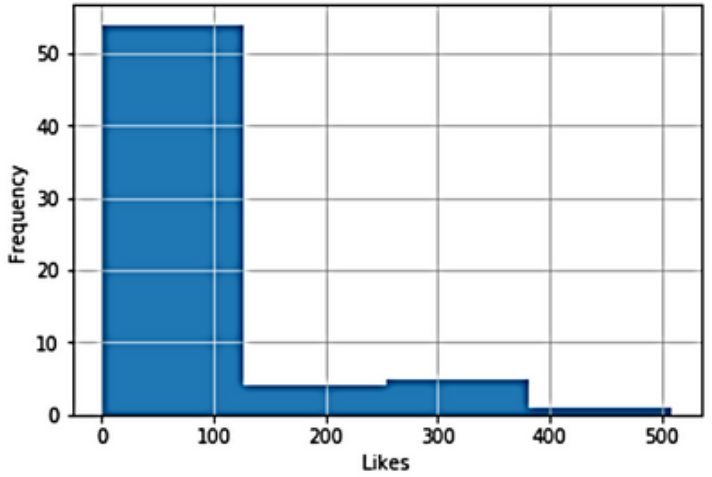
Data Visualization




```
Total number of tiles is 400
Cluster I: 88
Cluster II: 69
Cluster III: 62
Cluster IV: 100
Cluster V: 50
Cluster VI: 31

<Figure size 432x288 with 0 Axes>
```





Results

Cluster 1

“Great” restaurants from various ethnicities (European, Hispanic, and American)

Cluster 5

“Great” alternative restaurants (ice cream shops, coffee shops, and breakfast spots)

Clusters 0, 2, 3, and 4

Restaurants that are “above average,” “below average,” or “poor”

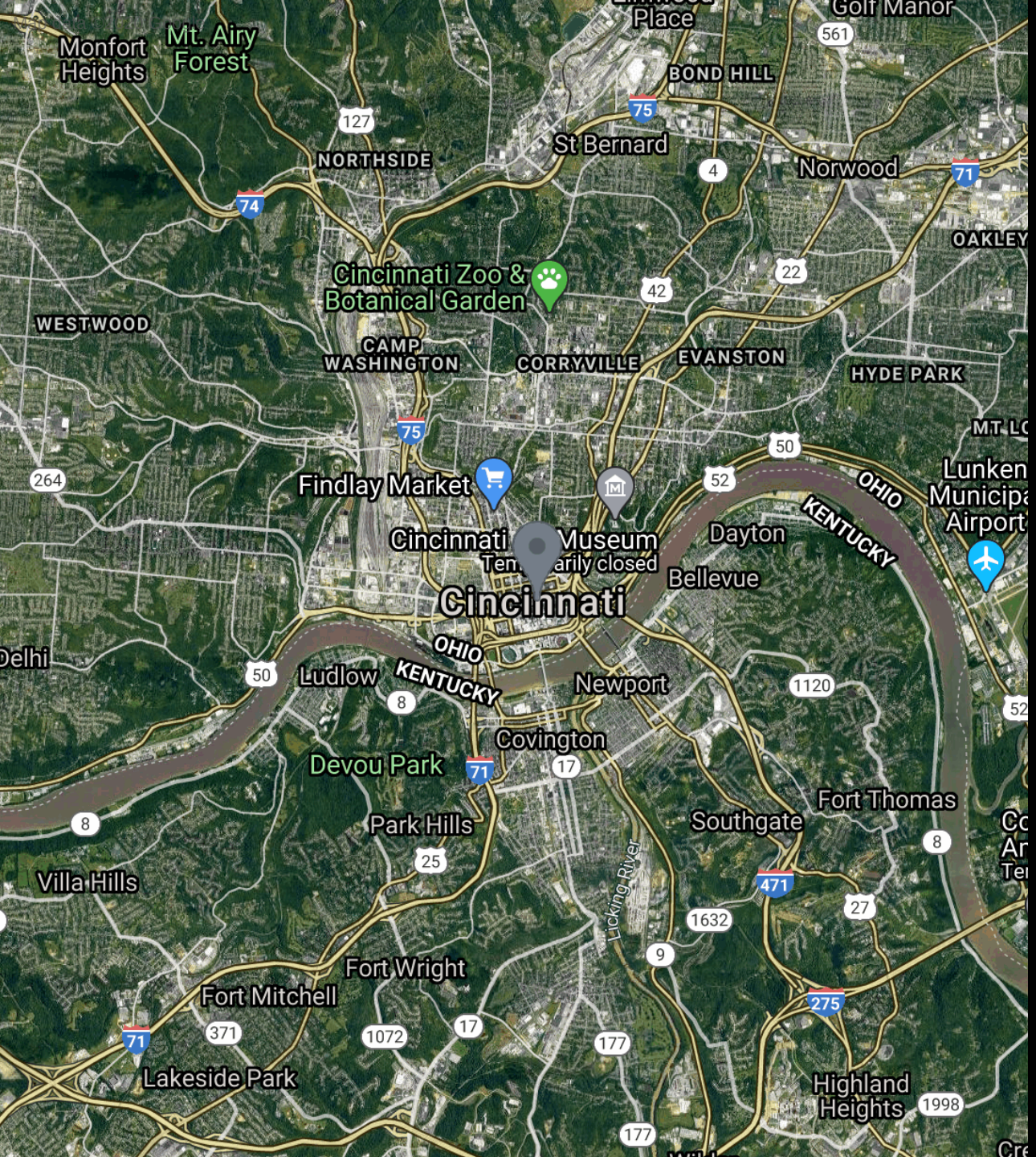
Results Cont.

Cluster 2 and Cluster 4 consist of all the “poor” restaurants, Cluster 0 consists of all the ”below average” restaurants, and Cluster 3 consists of all the “above average” restaurants. Of notable interest to tourists, locals, food critics, and restaurants are Cluster 1 and Cluster 5, both of which consist of ”great” restaurants. Cluster 1 features food from various ethnicities (European, Hispanic, and American) while Cluster 5 features alternative restaurants such as ice cream, coffee shops, and breakfast spots. Depending on which type of food a stakeholder is craving, they may choose between Cluster 1 and Cluster 5.

Discussion

Surprisingly, restaurants are evenly distributed over all 6 clusters despite the overwhelming right skew in the Foursquare likes data shown in Figure 3. While the upper echelon of “great” restaurants seems to be limited, there are roughly equal proportions of ”poor,” “average,” and ”above average” restaurants in Cincinnati. Specifically, there are very few “great” alternative restaurants in Cincinnati, a point of interest to potential entrants. Thus, our recommendations for the relevant stakeholders are as following. **Locals/tourists/restaurants should explore restaurants in Cluster 1 and Cluster 5. Potential entrants should seek to establish a restaurant such that it falls in Cluster 5.**

Of course, the analysis presented here is far from complete. Only 2 factors are analyzed: the number of Foursquare users who liked a specific restaurant and the broad categorization of each restaurant. Several other factors are also of interest: the general price-point of each restaurant, each restaurant’s proximity to other tourist attractions, and each restaurant’s Michelin rating, to name just a few. While this data was not immediately available for this analysis, with the use of web scraping techniques, it can be obtained for future analyses.



Conclusion

As diverse as the food scene in Cincinnati is, the upper echelon of restaurants is limited. This is to the benefit to tourists, locals, and food critics who may find it difficult to determine which restaurants are the most worth exploring. Specifically, for all relevant parties, the restaurants in Cluster 1 and Cluster 5 should be of special interest. Furthermore, as the food scene in Cincinnati continues to expand, restaurants must seek to position themselves in the upper echelon of restaurants, the segment that is the least saturated.