LOCATION OF MANHATTAN'S NEXT NEIGHBORHOOD BREAKFAST SPOT

IBM Applied Data Science Capstone
Capstone Project

Vineet Raichur
May, 2020

Business Problem

- Identify the location of Manhattan's next neighborhood breakfast spot
- Manhattan is one of the most expensive real estate market in the U.S.
- Choosing the right location will be critical to ensure profitability of the proposed breakfast spot
- To help narrow down the search, this analysis will aim to identify a few suitable neighborhoods
- The assumption here is that the ideal neighborhood will be one with
  - *No or very few breakfast spots*
  - *More residential buildings than offices*
  - *High number other food and beverage businesses (e.g., restaurants, coffee shops and bakeries)*
- Further research will help locate the buildings/spaces that meet the budget constraints

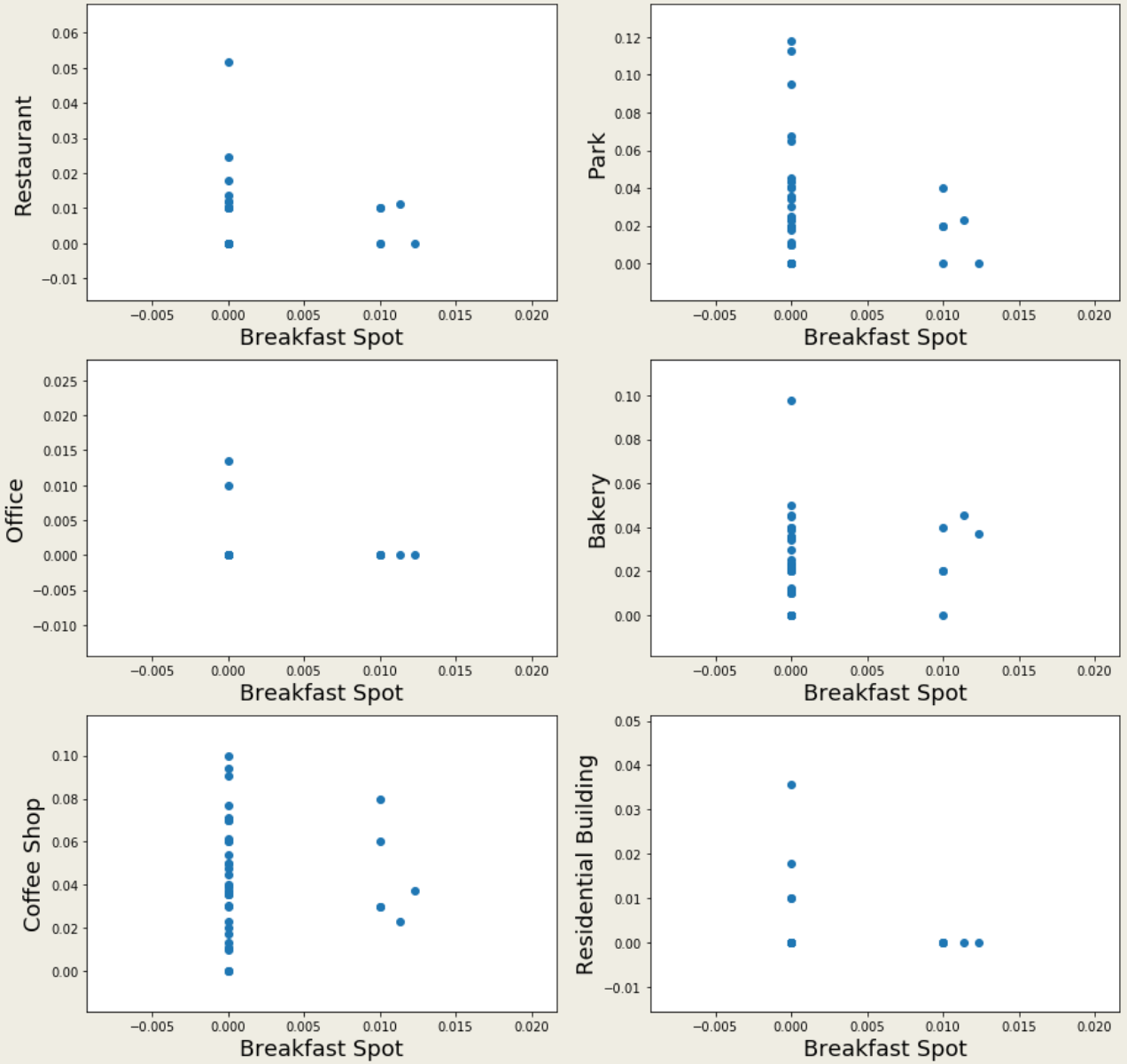
Data Used

- List of neighborhoods in Manhattan
  - *Using dataset available from NYU Spatial Data Repository – https://geo.nyu.edu/catalog/nyu_2451_34572*
  - *Data contains list of neighborhoods and their latitude and longitude coordinates*
- Venue data
  - *Used Foursquare API to gather venues data*
  - *Gathered list of top 100 venues within a 500 meter radius of the neighborhood*
  - *Foursquare data contains venue name, category and geographical coordinates*
  - *List of venues was converted into quantitative frequency of occurrence of venues in each category in each neighborhood*
- Venue categories used
  - *Breakfast spot, Restaurants, Park, Office, Bakery, Coffee Shop and Residential Buildings*

Exploratory Data Analysis

- Two clusters appear in the scatter plots - one with and one without breakfast spots
- Locations where no breakfast spots exist, there is a wide range in the frequency of occurrence of other venues

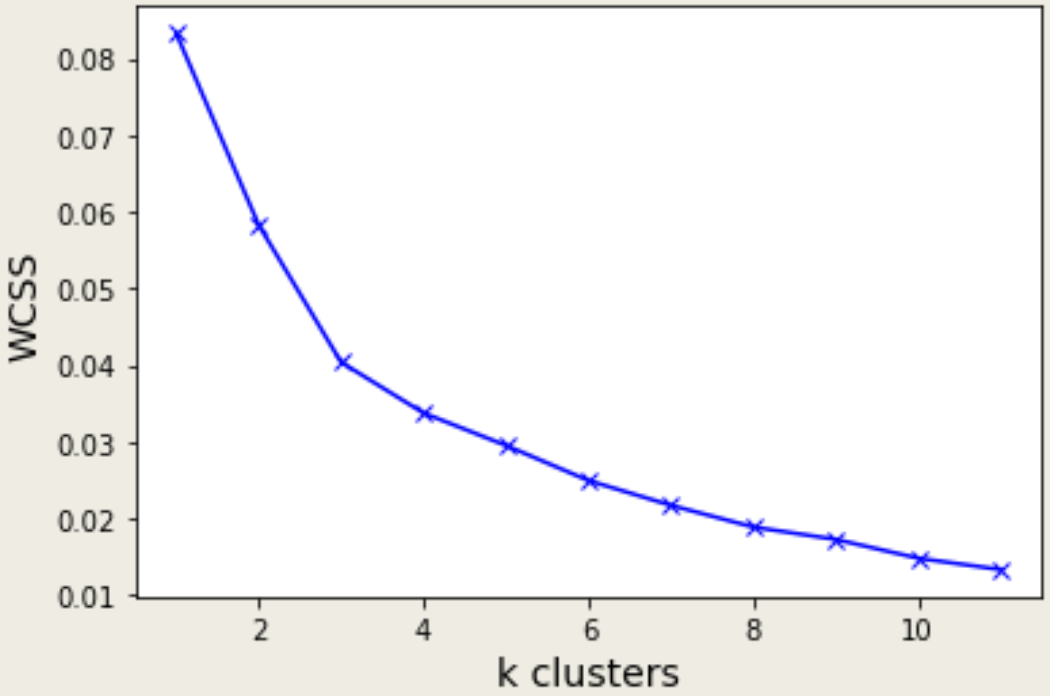
Figure: Scatter plots of frequency of occurrence of breakfast spots vs all other types of venues



Analysis

- We performed clustering on the data using k-means clustering approach
- We used Within Cluster Sum of Squares (WCSS) measures to determine optimal number of clusters
- Decided to run the k-means clustering with 6 clusters

Figure: WCSS values at each clustering trial with k number of clusters



Results

- Cluster 3 meets the requirement with
  - *No breakfast spots and no offices*
  - *High occurrence of residential buildings and parks*
  - *Occurrence of restaurants, bakeries and coffee shops*

Table: Average frequency of occurrence of each venue type in each cluster

Cluster Labels	Breakfast Spot	Restaurant	Park	Office	Bakery	Coffee Shop	Residential Building
0	0	0	0.1040	0	0	0.0709	0
1	0.0017	0.0035	0.0140	0.0008	0.0186	0.0735	0.0008
2	0	0.0062	0.0393	0	0.0141	0.0429	0.0079
3	0.0036	0.0146	0.0278	0	0.0476	0.0154	0
4	0	0.0068	0.0926	0.0068	0	0.0068	0
5	0.0022	0.0020	0.0041	0	0.0282	0.0306	0

Recommendations

- All 8 neighborhoods in 3rd cluster have no breakfast spots and no offices, but only 3 of them have residential buildings
- Three neighborhoods – Hudson Yards, Roosevelt Island and Turtle Bay – are good candidates for the new breakfast spot

Table: Average frequency of occurrence of each venue type in each neighborhood in Cluster 3

Cluster Labels	Neighborhood	Breakfast Spot	Restaurant	Park	Office	Bakery	Coffee Shop	Residential Building
2	Civic Center	0	0	0.0430	0	0.0215	0.0538	0
2	Flatiron	0	0.01	0.03	0	0.01	0.02	0
2	Gramercy	0	0.0119	0.0238	0	0.0119	0.0476	0
2	Hudson Yards	0	0.0179	0.0357	0	0	0.0357	0.0179
2	Roosevelt Island	0	0	0.0357	0	0	0.0357	0.0357
2	Sutton Place	0	0	0.0408	0	0.0204	0.0612	0
2	Tribeca	0	0	0.0649	0	0.0390	0.0390	0
2	Turtle Bay	0	0.01	0.04	0	0.01	0.05	0.01

Conclusion

- Our aim in this project was to help narrow down the search for the location of a new neighborhood breakfast spot in Manhattan
- We gathered location data on the neighborhoods in Manhattan and venue data for these neighborhoods from Foursquare
- We used the k-means clustering approach to cluster the neighborhoods based on the frequency of occurrence of certain types of venues
- Based on this analysis we found that Hudson Yards, Roosevelt Island and Turtle Bay would be ideal neighborhoods for the breakfast spot
- This analysis can be further refined for instance, by adding population densities and average commercial space rents in each neighborhood