

Capstone Project-Battle of The Neighborhoods

Marc-Henri Bleu-Laine

July 2019

Background

Opening an African restaurant in the city of Atlanta is a tough problem for business owners

- Which location is best?
 - Where is the competition?
 - Is the location safe?
 - Is it expensive to rent?
- Taking into consideration all these parameters and more adds complexity
 - Solving this problem can lead to more people opening their restaurants
 - Spreading of the overall African Culture
 - Potential large market due to small competition (limited number of African restaurants)



Problem Statement

Determining the optimal location to open an African restaurant in Atlanta, GA

For this project the optimal location is defined as:

- Top places of the neighborhood are restaurants
- No direct competition with other African restaurants
- Low crime rate
- Cheap rent



Data Description & Methodology

→ dataflow



- Zillow Provides rent of different neighborhoods for year 2019



- APD COBRA Dataset Provides 2019 crime counts and types in the city of Atlanta for a given neighborhood

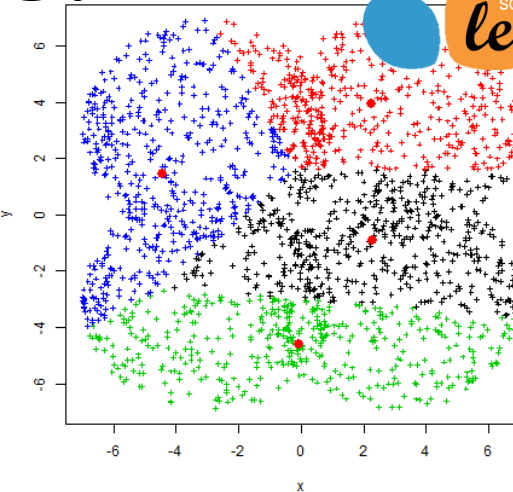


- Foursquare Provides venues in all neighborhoods (using combined neighborhoods from the two other datasets)
- Other features are names, coordinates, and categories

Final Dataset

Per neighborhood: most common venues, crime rates, average rent, number of African restaurants

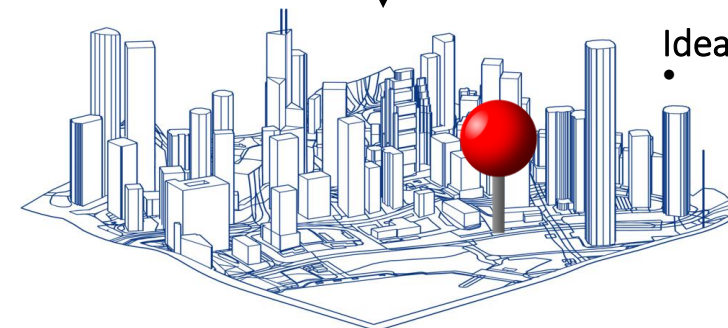
K Means Clustering



- **K-Means Algorithm** Creates clusters of neighborhoods

TOPSIS

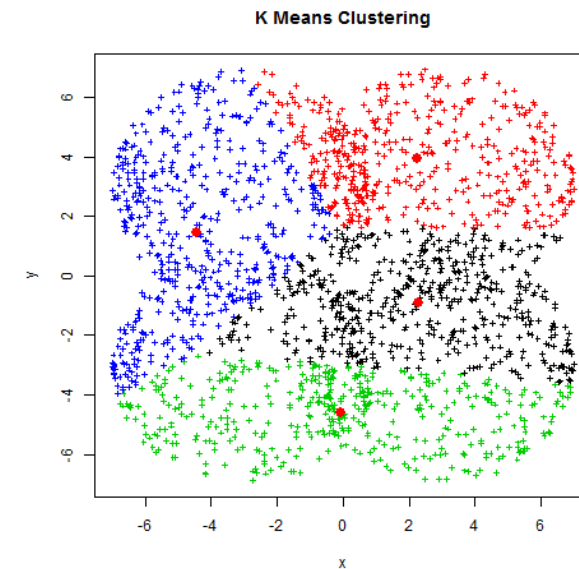
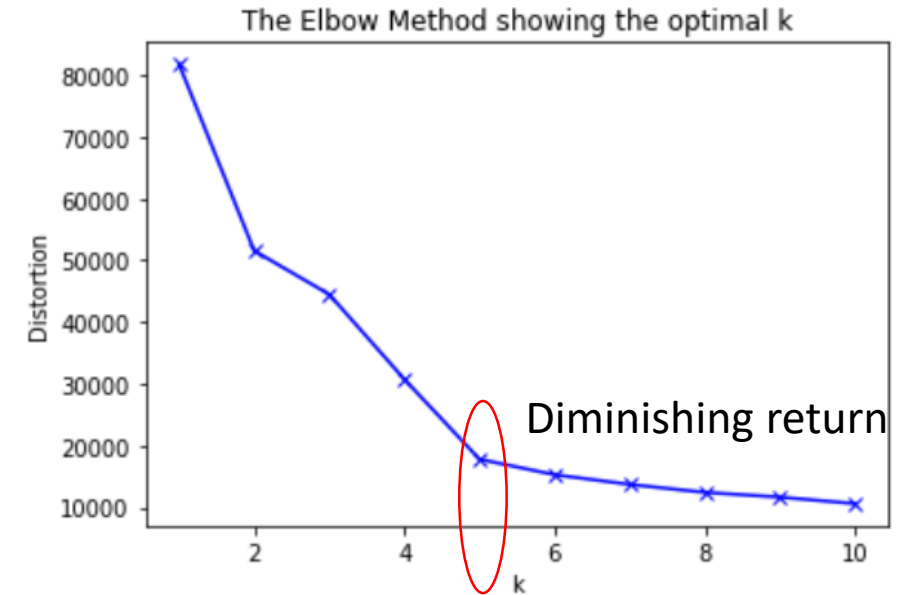
- **TOPSIS** Allows selection of a cluster of neighborhoods



- **Ideal cluster determined** Further Analysis can be done to select one final neighborhood

Neighborhood Clustering

- Before using K-Means to cluster the neighborhoods in the final dataset, the elbow method was used
 - The method consist in finding the percentage of variance contained in the clusters.
 - As more clusters are added, more variance is explained by them.
 - After some point there is diminishing return in adding more clusters (variance is only slightly more explained with additional clusters), the number of cluster at that moment is then the optimal one
- The ideal number of clusters found was 5
- K-Means from the sklearn library is used to cluster
 - It initially randomly establish a number of centers k for the clusters, then find the distance from all the points in the dataset to these centers.
 - The points that are closer to a center are considered to be in the same cluster
 - New centroids can be found by averaging the points in a given center, and new distance from the data points to the centroid can be calculated
 - The process is repeated until convergence is achieved (centroids no longer move)



Neighborhood Clustering-Cluster Descriptions

Cluster 0

- The second largest cluster created
- Contains most of the African restaurants
- Highest crime rate
- Highest average rent
- Diverse neighborhoods in terms of venues but seems to contain the most food places in its most common ones

Cluster 2

- It has the second lowest crime rate
- Quite a good amount of restaurants in its most common venues
- Has two African restaurants
- Second highest rent

Cluster 4

- Third largest rent
- Okay amount of restaurants
- One African restaurant
- Third largest crime rate

Cluster 1

- Smallest cluster
- Lowest rent price
- No African restaurants
- Lowest crime rate
- Does not look appealing in terms of restauration (only the 7th and 10th most common venues are ones)

Cluster 3

- It has no African restaurants
- It is the second largest cluster and has the second largest crime rate
- Average rent is on the lower sideAn okay amount of restaurants (lots of fast-food)
- Second largest crime rate

Decision Making

A variant of the final dataset was used

- After exploring the clusters, the restaurant popularity feature was created
- Ranked from 3 to 1 with 3 being a cluster with a high overall restaurant count and 1 being a cluster with a low count
- Other features were number of African restaurants, crime rate, and average rent

Technique for Order Preference by Similarity to Ideal Solution (TOPSIS)

- Used for multi-attribute decision making
- Internal scale needs to be chosen for qualitative data (restaurant popularity)
- Stakeholder specifies direction of the feature (minimize average rent for example) and the relative importance by associating it with a weight

Cluster	Weights:	0.4	0.2	0.3	0.1
	Direction:	Minimize	Minimize	Maximize	Minimize
		Rent	Crime	Restaurant_popularity	African_restaurant
0		1870.39375	130.400	3	3
1		10666.40000	11.000	1	0
2		5470.90000	11.660	2	2
3		1188.98000	77.775	3	0
4		2931.90000	36.720	2	1

Result & Conclusion

- The goal of this work was to help potential business owner to open an African restaurant in the city of Atlanta, GA
- Cluster 3 was found to contain the best neighborhoods based on the data and the preferences that were used. The final choice of the neighborhood is left to be done by the stakeholder
- The methodology included fusing multiple datasets and using the newly created one to cluster the neighborhoods using machine learning. The methodology also included a decision making step to select which cluster contains the best neighborhood
- For future work it would be useful to:
 - Include the ranking of the venues to see which one you are competing against
 - Acquire missing data for some neighborhoods so that they can be included in the cluster
 - Add the different types of crime in the dataset and weight them differently
 - Cluster within a cluster to find even more insights about the neighborhoods
 - More rigorous quantification of restaurant popularity could be done

Data Description & Methodology

