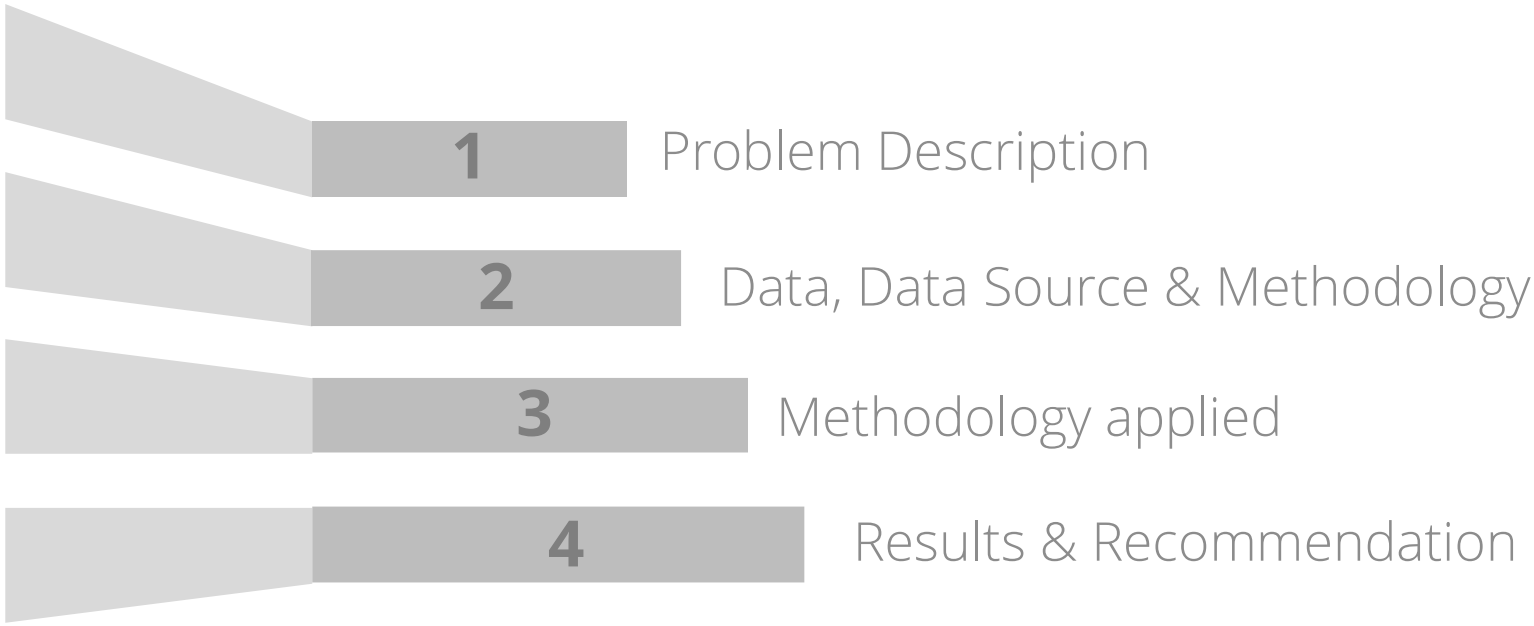
Capstone Project - The Battle of Neighbourhoods

Where to open a restaurant?

Dan Wyss

| Agenda |


1
Problem Description
2
Data, Data Source & Methodology
3
Methodology applied
4
Results & Recommendation

Problem Description

Business Case	The purpose of this project is to provide an analysis to the fictitious question in which neighbourhood of the boroughs “Bronx” or “Queens” is it worth to open a restaurant. To answer this question data will be analysed to detect venues which are in competition with the intention to open a new restaurant so that hard competition can be avoided and the idea of opening a restaurant will have a promising starting point.
Approach	The following approach will be implemented: 1. Identify the neigborhoods of the Bronx and of Queens 2. Get the respective latitude and longitude of the neighborhoods 3. Visualize them on a map to get an geographical overview 4. Get the venues within a neighborhood and the respective venue category 5. Identify unique categories of the venues 6. Identify the top 3 venue categories of each neighborhood 7. Run k-means to cluster the neighborhood into 3 clusters 8. Visualize the outcome of the clustering on the map 9. Examine clusters

Data, Data Source & Methodology

Data of New York	To identify the neighborhoods, including latitude and longitude thereof, of the “Bronx” and of “Queens which are boroughs of New York a dedicated json file will be downloaded from: https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json And converted into a panda data frame structure for further analysis.
Venue data	To obtain the data of the venues with a neighborhood and the respective category the Foursquare API will be used . The data consists o the following parameters: 1. Name of the venue 2. Category of the venue 3. Latitude of the venue 4. Longitude of the venue
Methodology	After having merged the data received from the two source mentioned above, the data mining will be continued to 1. Classify the categories of the venues 2. Define the occurrence of each category for the venues and identify the top 3 categories of each neighborhood 3. Run k-means to cluster the neighborhood into 3 clusters 4. Visualize the outcome of the clustering on the map & examine the clusters

Results

Map with Clusters

Most Common Venues

Cluster 0

1st : Beach, Art Gallery, Nightlife
2nd : Surf Spot, Park, Bar
3rd: Bar, Spa, Donut Shop

Cluster 1

1st : various Restaurants, different Stores
2nd : Fast Food and different stores and shops
3rd: Restaurants, Fast Food, Stores

Cluster 2

1st : Pizza Places, restaurants, parks
2nd : Pizza places, donut shops, stores
3rd: fast food, bars, park

Recommendation

Due to the fact that neither restaurants, nor various fast food shops are heavily distributed within cluster zero it is strongly recommended that a restaurant should be opened within the following neighborhoods of cluster 0:

- Broad Channel
- Breezy Point
- Averne
- Rockaway Beach
- Neponsit
- Belle Harbor
- Rockaway Park
- Roxbury
- Hammels