PEST CONTROL BUSINESS OPPROTUNITIES IN NYC

Data Mining report

MAY 1, 2021

SBEAT
AKL NZ

Problem Statement

A young entrepreneur in New York city is wanting to open a mobile pest control business with a strong focus on food industry. The business concept is chosen to have employees travelling to the customers by escooters or ebikes allowing for extra savings on fleet and travel time as well as other costs associated. However, a basic requirement for storage and cleaning facilities are stated as follows:

- The location for the base needs to be chosen as closely to most of the job locations as possible

- Location detailed down to a neighbourhood level

Essentiality of this report is in the tight competition between major pest control firms in NYC which are monopolizing the areas.

That is why this report will play a vital role to help our customer to try and use the competitive advantage of a zero-fleet operational approach.

The task will involve some exploratory analysis of Food Venues in NY as well as the potential in the services required.

The results will be used for other pest control businesses with a different set of requirements but also specialised in Food Businesses. This report is targeted to help the Restaurant owners to identify potential areas suitable to open a bar or cafe.

DATA

Dataset 1

The data we are going to use is a Foursquare API data retrieval consisting of venues and categories of venues in the whole New York city.

Contains the following fields:

- name
- categories
- lat
- lng

Sample:

	name	categories	lat	lng
0	Lollipops Gelato	Dessert Shop	40.894123	-73.845892
1	Rite Aid	Pharmacy	40.896649	-73.844846
2	Carvel Ice Cream	Ice Cream Shop	40.890487	-73.848568
3	Walgreens	Pharmacy	40.896528	-73.844700
4	Dunkin'	Donut Shop	40.890459	-73.849089
5	Cooler Runnings Jamaican Restaurant	Caribbean Restaurant	40.898083	-73.850259
6	Subway	Sandwich Place	40.890468	-73.849152
7	Central Deli	Deli / Bodega	40.896728	-73.844387
8	Koss Quick Wash	Laundromat	40.891281	-73.849904

We are going build the hot encoded table and count the values for each Neighbourhood therefore we are not much concerned with the geo data

Dataset 2

The second dataset is a NYC Rat Sightings dataset freely available on https://www.kaggle.com/new-york-city/nyc-rat-sightings

New York City rodent complaints can be made online, or by dialing 3-1-1, and the New York City guide Preventing Rats on Your Property discusses how the New York City Health Department

inspects private and public properties for rats. Property owners that fail inspections receive a Commissioner's Order and have five days to correct the problem. If after five days the property fails a second inspection, the owner receives a Notice of Violation and can be fined. The property owner is billed for any clean-up or extermination carried out by the Health Department.

Data is from 2010-Sept 16th, 2017 and includes date, location (lat/lon), type of structure, borough, and community board.

Contains the fields :

Unique Key Created Date Closed Date Agency Agency Name Complaint Type Descriptor Location Type Incident Zip Incident Address Street Name Cross Street 1 Cross Street 2 Intersection Street 1 Intersection Street 2 Address Type City Landmark Facility Type Status Due Date Resolution Action Updated DateCommunity Board Borough X Coordinate (State Plane) Y Coordinate (State Plane) Park Facility Name Park Borough School Name School Number School Region School Code School Phone Number School Address School City School State School Zip School Not Found School or Citywide Complaint Vehicle Type Taxi Company Borough Taxi Pick Up Location Bridge Highway Name Bridge Highway Direction Road Ramp Bridge Highway Segment Garage Lot Name Ferry Direction Ferry Terminal Name Latitude Longitude Location

---

We only use the fields of Borough and count the grouped values

Dataset 3

New York Borough data in JSON format

Contains :

- Borough
- Neighborhood
- lat
- lng

Dataset 4

Obtained via a web scraping technique from the website (https://www.nycbynatives.com/nyc_info/new_york_city_zip_codes.php)

NYC zip codes for each Borough. Contains:

- Borough

- ZIP

Cleaned sample:

	zip	borough
0	**10001**	**Manhattan**
1	**10002**	**Manhattan**
2	**10003**	**Manhattan**
3	**10004**	**Manhattan**
4	**10005**	**Manhattan**

We will use this table to join it on zip with the rat sightings table.

Methodology

As the methodology we chose to combine the data from the New York city Neighbourhoods geo information and utilize a Foursquare API to get the list of venues in each neighbourhood.

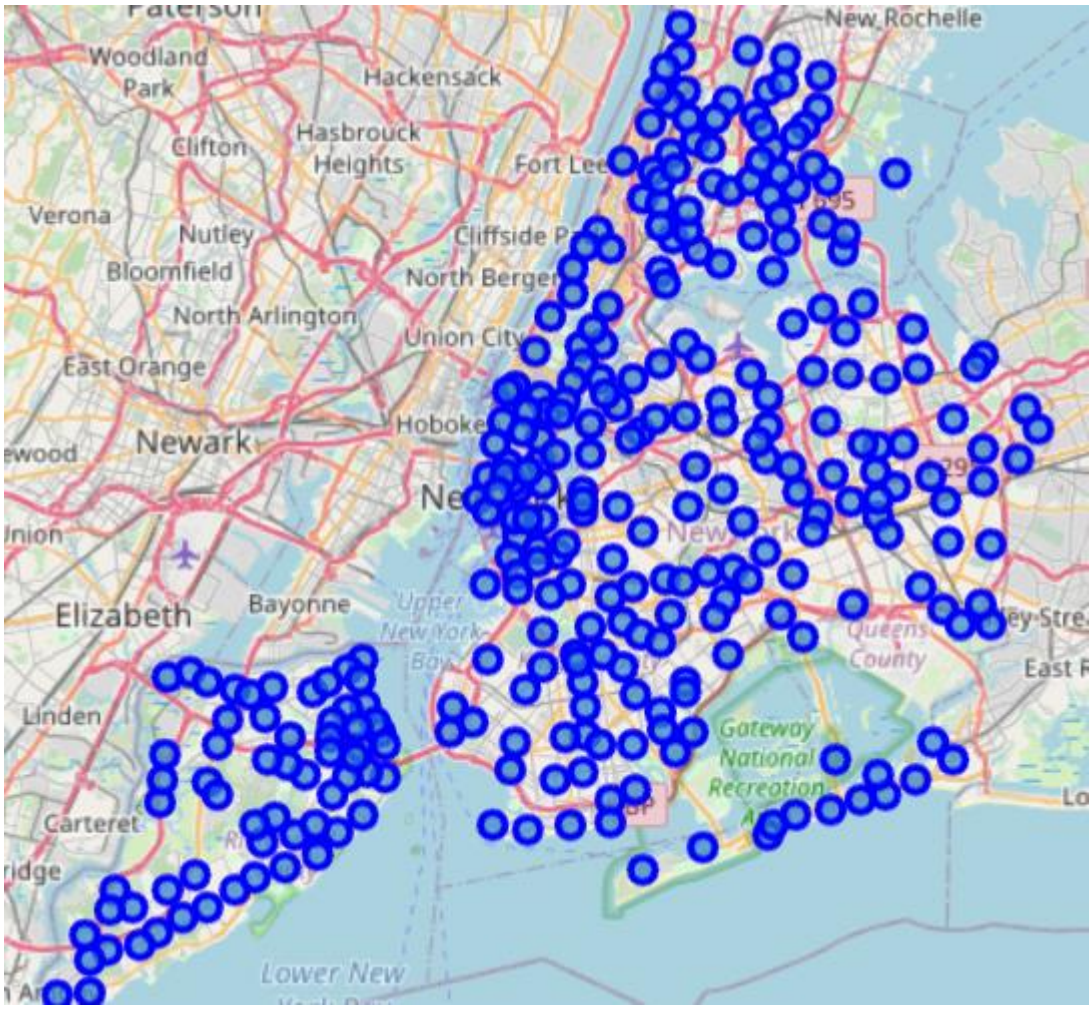
Consecutively we performed some analysis on the resulting data table to select only food related venues from the whole list followed by grouping it and combining into total Food Venues for Borough and Neighbourhood.

We had to utilize the NYC zip codes to join the two tables on the zip code to get the resulting table of rat sightings and food venues.

As the main KPI for the business problem we chose the ratio of Rats sightings for each venue and borough.

A sorting technique was performed to sort the results so that we have a descending table.

On the picture below we see the neighborhoods we will be exploring for the business opportunities



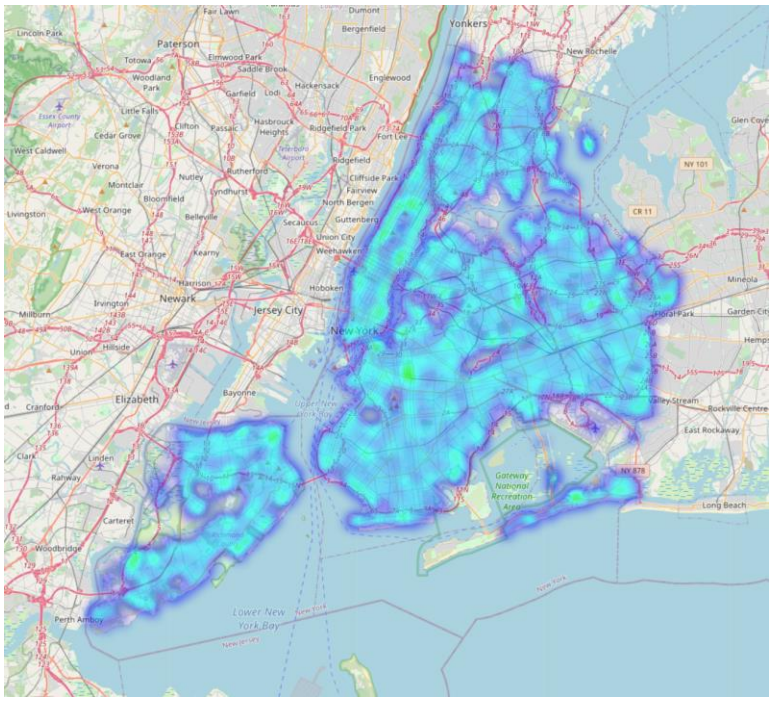
Discussion

During the data mining we have discovered that Queens, Manhattan and Bronx have the most food venues with most pests sighted.

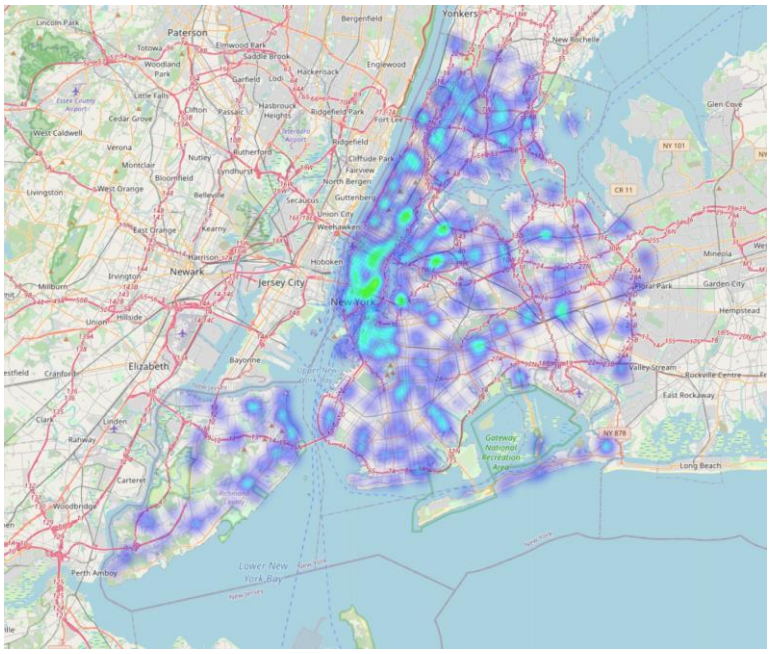
However, after performing some data analysis purely on the rats data, we found that Brooklyn is the most infested Borough. However, our business problem was to prioritize the food venues which made the Neighbourhood of Astoria that is part of Queens our top suggestion.

Manhattan is also very much affected by the pests which made Murray Hill into top 3 neighbourhood option.

Below is the heatmap of the rat sightings in NYC with a clear indication of downtown areas being affected the most.



In comparison below is the heatmap of the food venues which almost coincide with the pest infested areas:



After data wrangling and joining datasets by zip codes and Borough names we can make some conclusions and apply some KPIs being derived as rat per venue ratio multiplied by 1000.

We will be using that KPI to grade each neighborhood for attractiveness to our business venture.

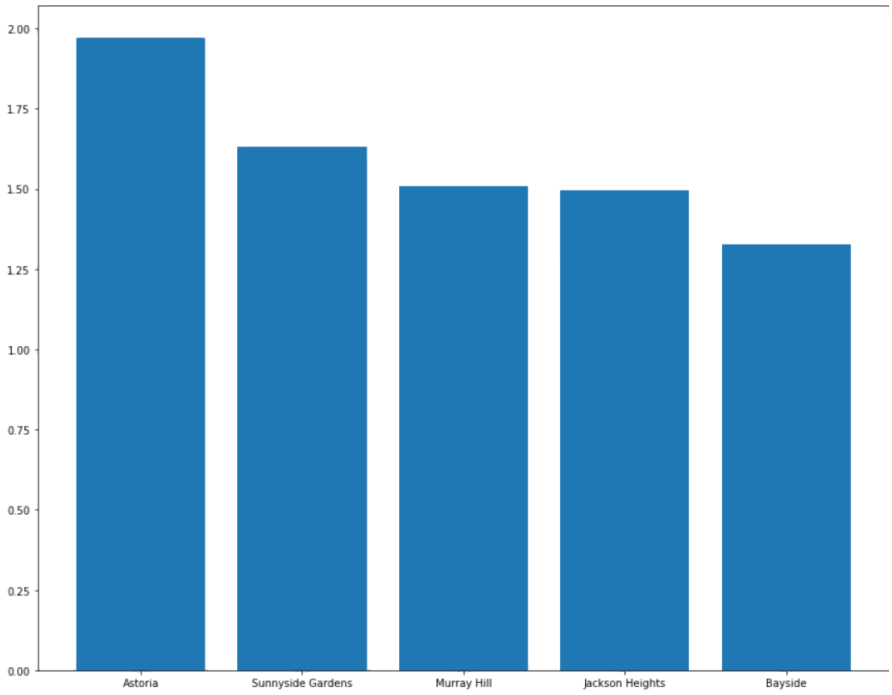
	Borough	RatsNum
0	Bronx	41332.0
1	Brooklyn	69136.0
2	Manhattan	52352.0
3	Queens	29412.0
4	Staten	9792.0

While Brooklyn being the most pest infested we will investigate the KPI for each Borough

During the data mining we have discovered that Queens, Manhattan and Bronx have the most food venues with most pests sigthed.

However after performing some data analysis purely on the rats data we found that Brooklyn is the most infested Borough. However our business problem was to prioritize the food venues which made the Neighborhood of Astoria that is part of Queens our top suggestion.

Manhattan is also very much affected by the pests which made Murray Hill into top 3 neighborhood option.



Results

As we can see in the final results the most pest infested, and food business dense areas are Astoria and Sunnyside Gardens