Police Stations VS Donut Shops

Igor Khatybov

May 2020

Table of Contents

Introduction

Background

Do police officers and detectives like donuts in real life as much as in movies? As a person who grew up in another country and now living in the United States, I am still trying to figure out the way this country lives.

And I have been curious about this question a lot. I always wanted to do some research on this subject, and I could not miss this opportunity.

Problem

This project aims to show the relationship between the locations of police stations and food joints in order to determine the popularity of donut shops near the stations. Also, the stations will be clustered to determine a similarity between them.

Three cities take part in this project: Chicago, New York City, and Los Angeles.

Interest

Even though the reason for this subject was pure curiosity, this project can be used as the first step in marketing research, for example, to find out what is the most popular brand of a donut shop near police stations.

Data

Data sources

Web-pages with addresses of police stations:

- Chicago: https://www.chicago.gov/city/en/depts/cpd/dataset/police_stations.html
- New York City: https://www1.nyc.gov/site/nypd/bureaus/patrol/precincts-landing.page
- Los Angeles: http://www.lapdonline.org/our_communities/content_basic_view/6279

FourSquare API (https://foursquare.com/city-guide) was used to get information about food joints for each police station.

OpenStreetMap Nominatim API (https://nominatim.org/release-docs/develop/api/Overview/) was used to get coordinates for the cities and each police station.

Feature selection

- Police station
    - Name
    - Address
    - Latitude
    - Longitude
- Food Joint
    - Category
    - Distance to Police station

Data cleaning

Addresses of police stations:

- Chicago: The dataset already had all the needed information with no missing values.
- New York City: The data was scraped from the web-page in tabular format. However, I removed the borough names from the table since I didn't need this information.
- Los Angeles: The data didn't have any convenient structure to scrap; therefore I copied this data and convert it to CSV format manually.

Coordinates of police stations:

- New York City: several addresses of police stations were not recognized by the API and were fixed manually.

Food Joints:

There was one problem with this API: I could not get a full list of food joints, because the API has a limit of 100 results for each location and gives those locations based on its own criteria of popularity. To get the most complete list of food joints for each police station, I made two requests for each station:

- For all food joints
- Search request for a "donut" keyword

There were two missed values in the whole dataset, and they were fixed manually.

Data Analysis

The goal of this project is to determine if there is a relationship between police station locations and donut shop locations, i.e. is the donut shop category one of the most popular categories near police stations.

In the first step, I collected and cleaned the required data for locations of police stations and popular food joints around each station.

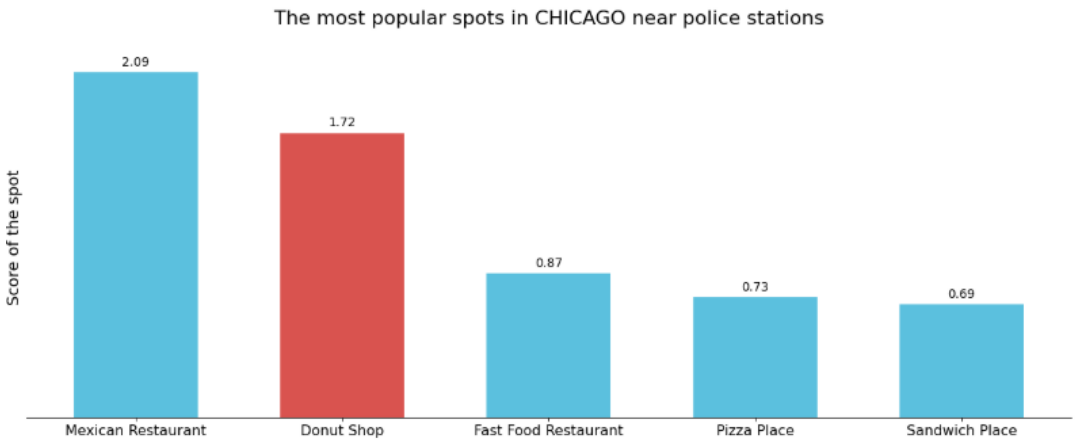
In the second step, I calculated a score for each category of food joints for each police station.

$$Score = \frac{Number\ of\ food\ joints\ from\ the\ same\ category}{Median\ of\ the\ distance\ for\ the\ category\ to\ police\ station}$$

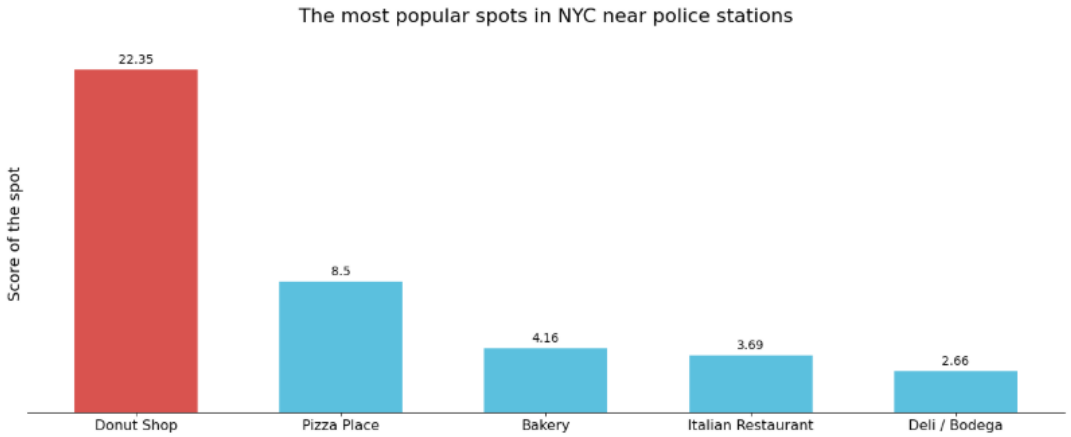
Using the sum of this score for each category, I sorted the data to find the most popular categories near police stations for the city. This answered the main question of this project.

In the third step, I did cluster for each city and tried to find relationships between different police stations using K-Means clustering.

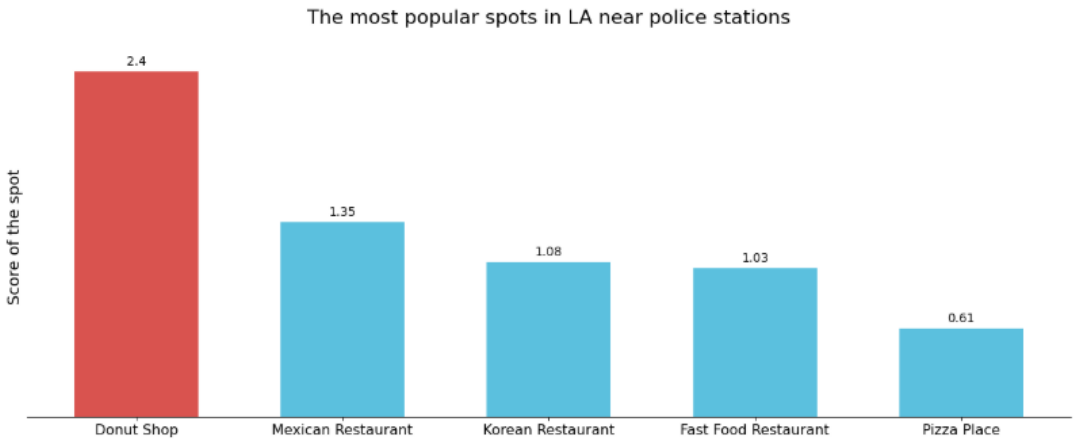
The most popular food joints near police stations for the city



Donut Shop on the second place.



Donut Shop has a much higher score than the other categories.



Donut Shop has a much higher score than the other categories as well. It is interesting, that Korean restaurants have a higher score than fast-food joints.

Clustering

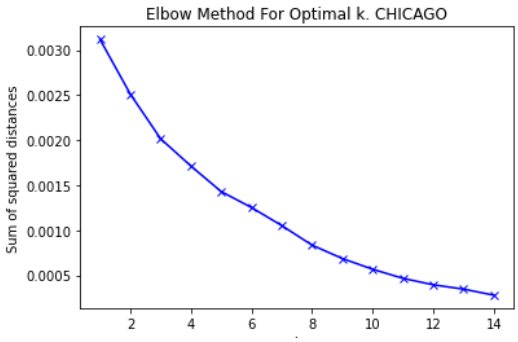
I separated police stations to different groups to see if there any relationships between them.

I used K-Means for clustering.

In order to do that, I determined steps:

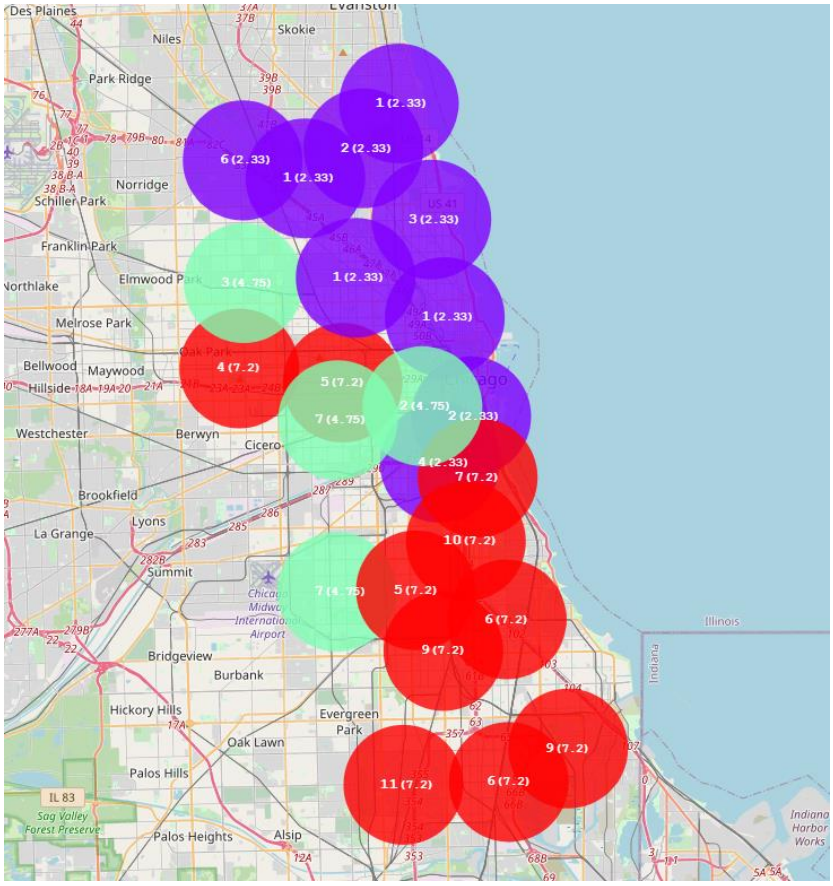
- Find best K for clustering using the elbow method
- Cluster all stations by the top categories for each PS
- Find position for Donut Shops in top categories for each cluster

Chicago



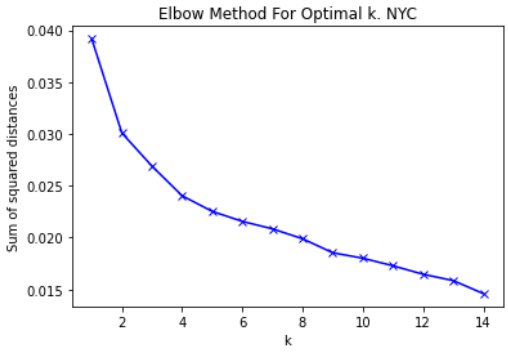
Best K is 3

The numbers represent the position of Donut shop in top categories for each police station and the average position of the Donut Shop category for each cluster respectively



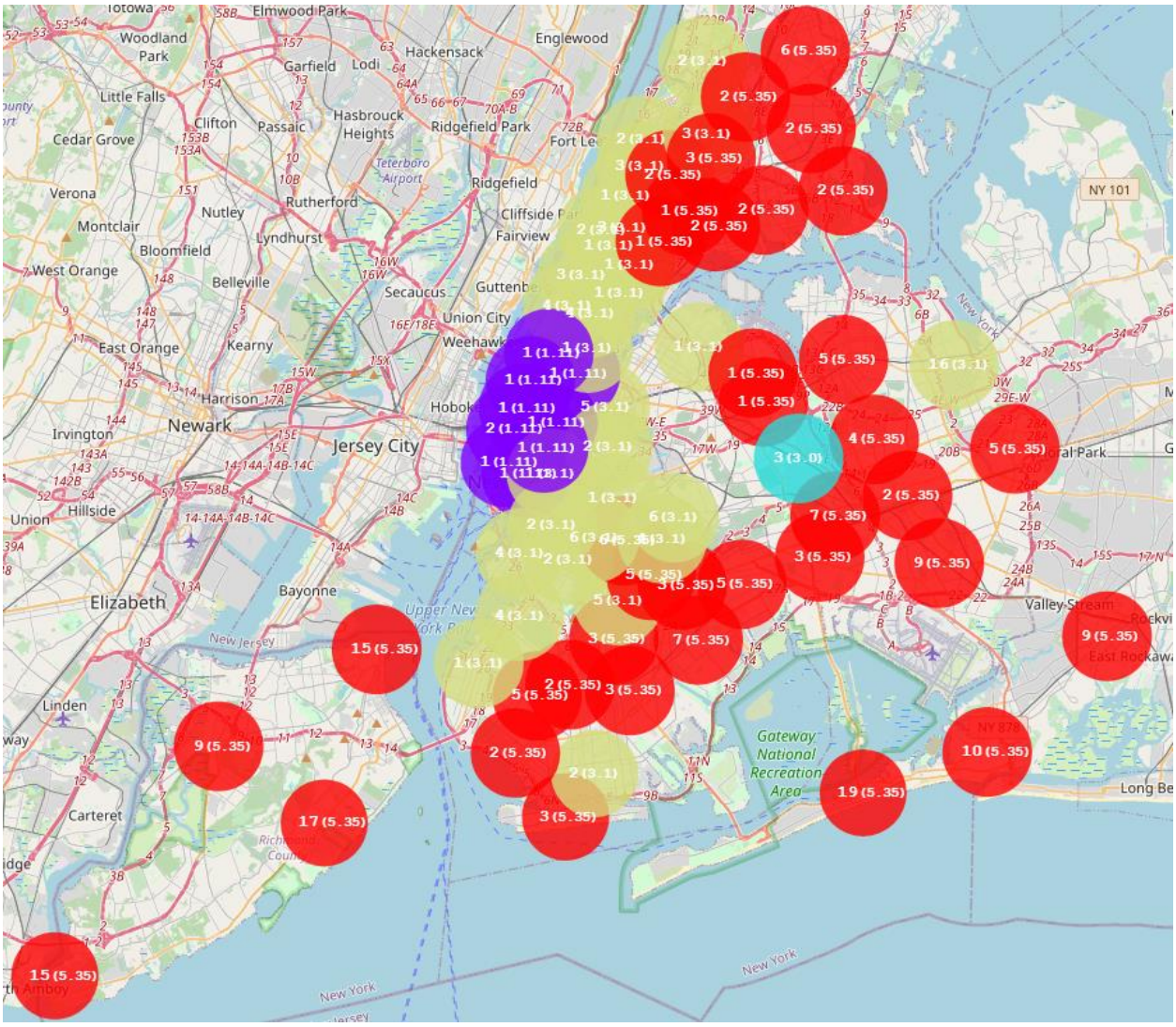
Donuts shops are much more popular in the north part of the city than in the south part. The third cluster is a link between two big clusters.

New York City

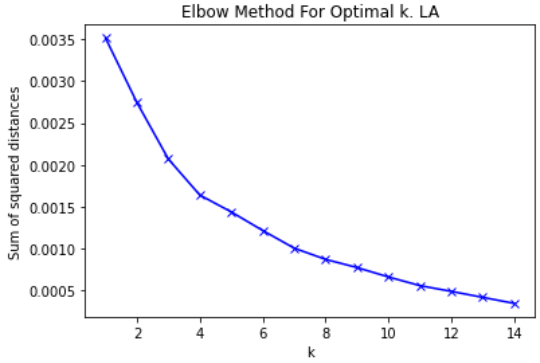


Best K is 4

The numbers represent the position of Donut shop in top categories for each police station and the average position of the Donut Shop category for each cluster respectively

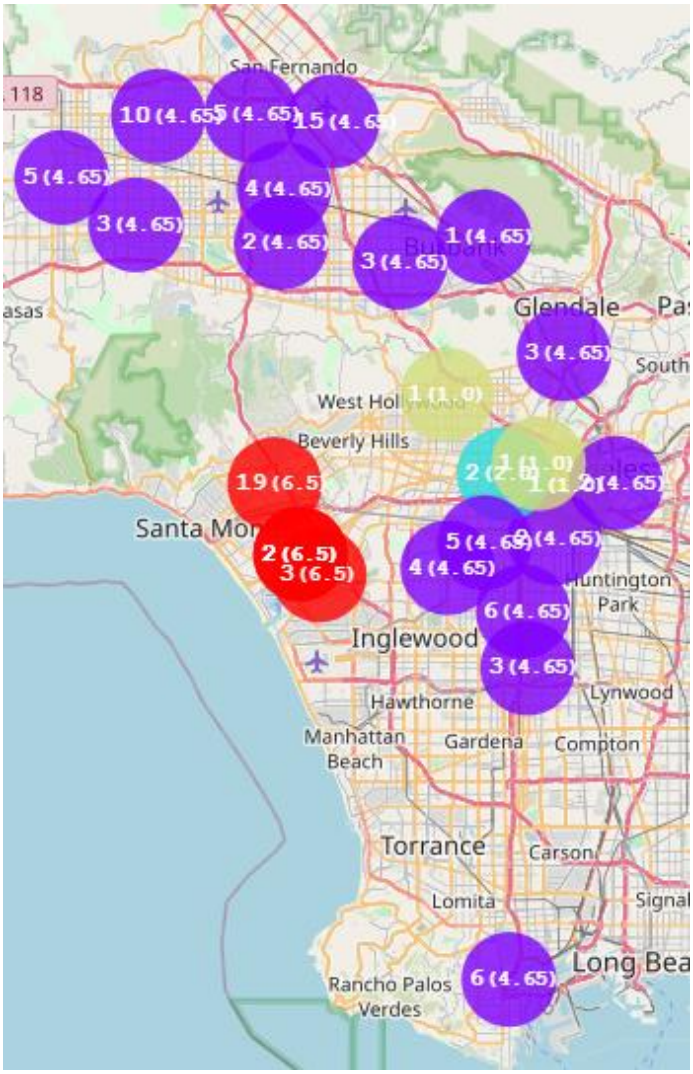


Los Angeles



Best K is 4

The numbers represent the position of Donut shop in top categories for each police station and the average position of the Donut Shop category for each cluster respectively



Results

The analysis shows that donuts shops are popular near police stations.

The clustering, however, shows that more extensive analysis should be made, especially in New York City and Los Angeles since clusters, in general, are not very different and have outliers.

Discussion and Conclusion

The purpose of this project was to find out if there is a relationship between police station locations and the popularity of donut shops near stations.

The relationship was defined, but analysis showed that there are steps in this project that could be improved:

- List of the food joints
    - It makes sense to increase the radius around each police station for more objective evaluation.
    - Get the full list of food joints within the radius regardless of popularity defined by the API.
- Make a more extensive analysis of the outliers in clustering.

This concludes the results of our project, but there are further steps that can be performed, for example:

- To find the most popular brands of donut shops near police stations.
- To find the most popular brands of donut shops within the cities and compare the result with donut shops near the stations.