Finalize franchisee applicants for Artisan pizzeria.

(Hypothetical)

by
Swati Patil,
1 July 2021

Table of content:

1. Abstract
2. Introduction
2.1 Background
2.2 Problem
2.3 Interest
2.4 Literature review
3. Data Acquisition and cleaning
4. Methodology
5. Results findings
6. Conclusion
7. References

1. Abstract:

Artisan pizzeria has decided to allocate its franchisees to very few selected applicants in the Mumbai area. The company wants to serve all the city with very few (as few as possible) locations. Here I am clustering all existing pizza places with K-means and then fit all applicants in that cluster. I will find the closest applicants to the centroid of each zone/cluster. Here I provide optimum distance between each location so that city gets divided into equal zones without overlapping or too away from each other.

2. Introduction

2.1 Background: Artisan pizzeria (Hypothetical) is a well-established International pizzeria franchisor. It is a premium pizzeria chain than regular pizza joints available on every other street. Artisan Pizzeria offers handcrafted specialty pizzas by skilled pizzaiolos (A person who makes pizzas in a pizzeria) with locally sourced fresh ingredients. After gaining immense popularity and good response in many other countries, Artisan pizzeria now wants to enter the Indian market, starting with Mumbai. Mumbai is the financial capital of India, it is very densely populated (44,500 people per square kilometers)and famous for its heavy traffics and busy streets. Artisan pizzeria received a considerable number of franchisee applications in Mumbai from various locations.

2.2 Problem: Even after a background check and initial screening based on different primary factors, the company has many applications remaining. Being a premium specialty chain and maintaining their quality, Artisan pizzeria has franchising policies preventing them from allowing too many pizzerias nearby. The company wants to target customers from all existing regular pizza joints who wish for a premium experience by providing them one Artisan location as an option in the vicinity. Artisan does not want to avoid competition; they want to fetch customer from the current market and create their own customer base. Location is a crucial factor when we want to allow a limited number of locations yet want to cover all the existing market.

2.3 Interest: The company wants to know

1. The precise number of franchisees to allow for the city that will cover all Mumbai.

2. Divide city into areas of existing pizza places and their centroids.

3. Determine following distances.

a. What is the minimum safe distance between two franchisees to avoid overcrowding and internal competition?

b. What is the maximum distance between franchise and existing pizza places so that consumers should not hesitate to reach to Artisan?

2.4 Literature review: According to Statist research study, "What is the maximum length of time you would drive to a place to eat?" was conducted in November 2016 by interviewing 719 people aged18 and older. Most of the people are in a category of 16 to 30 minutes. According to Brightlocal's findings by author Ross Marchant, consumers are willing to drive 17 minutes to a local restaurant / café. From OALLEY.NET (IT provide an interactive map to create an area based on travel times), the average traffic speed in Mumbai is 15kmph, and it's as low as 10kmph in busy hours.

3. Data Acquisition and cleaning:

3.1 Data Source: Geo locations of all the existing pizza places is collected using foursquare API. As this is a hypothetical situation, there are no actual franchisee applicants. I collected all existing café's and their locations just like pizza places and treated them as applicants. Let's assume applicants provide this data in online application forms.

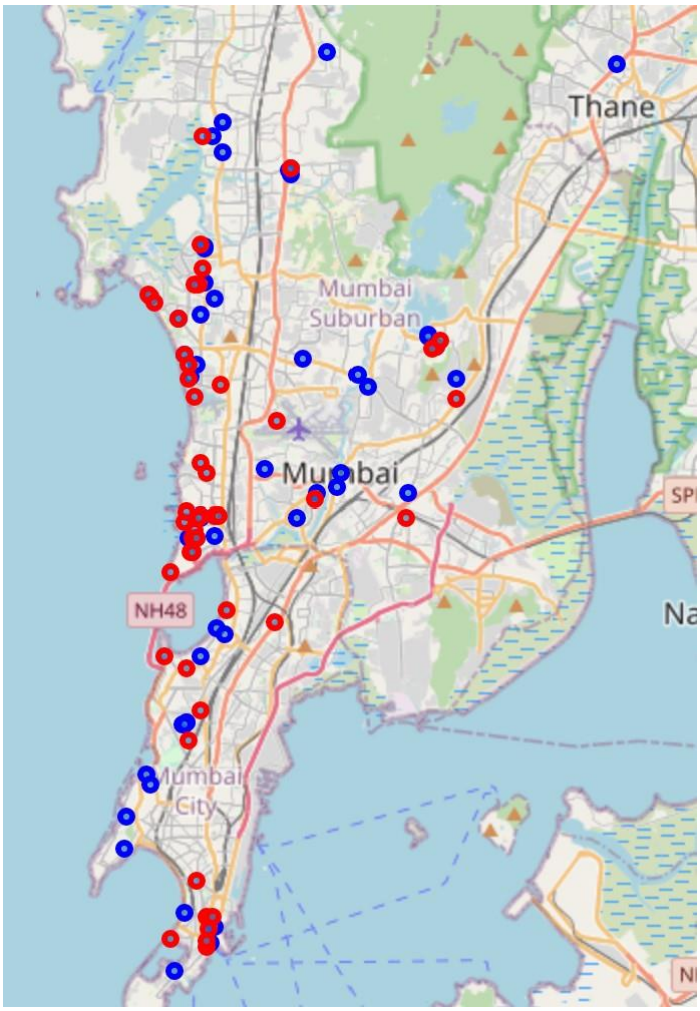
3.2 Feature selection: Form four square API, I gathered info about each pizza place like their Name, Addresses, Locations(Coordinates), Category, Rating, etc. For this situation, I need only location details, and I keep Names for reference purpos.

3.3 Exploratory Data Analysis: As mentioned in the literature review section, consumers don't mind traveling around 17 to 30 minutes to their preferred restaurant/café. 15km/h is the average speed of Mumbai, So let's try to add the following constraints for clustering.

a. The minimum distance between two franchisees (a) must be more than 10 km(40 min).

b. The maximum distance between franchise and other pizza places in that area (b) must be 5km (20 min).

Considering these two conditions, the divided clusters will be non-overlapping zones of radius 5km each.



- Applicants
- Existing pizza Places

The above Image help visualizing the distribution of applicants Vs. and existing pizza places.

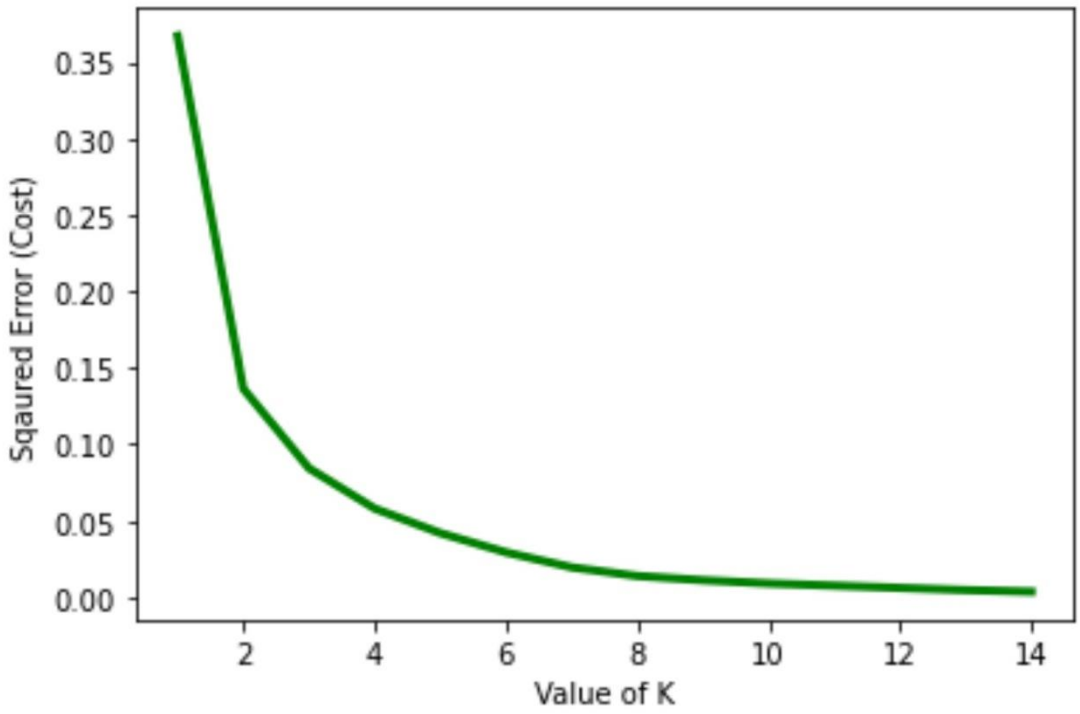
The distribution of pizza places and applicants is very similar and uniform to each other.

4. Methodology:

I approached the clustering problem by implementing the **k-means algorithm**. k-means is a distance-based method that iteratively updates the location of k cluster centroids until convergence.

The main user-defined parameters of the k-means algorithm are the **distance function** (often Euclidean distance) and the **number of clusters k**. This parameter needs to be set according to the application or problem domain. (There is no magic formula to set k.) In a nutshell, k-means groups the data by minimizing the sum of squared distances between the data points and their respective closest centroid.

To find out the optimum value of k, I used the elbow method. I plotted squared error for k 1 to 9 on the graph below.

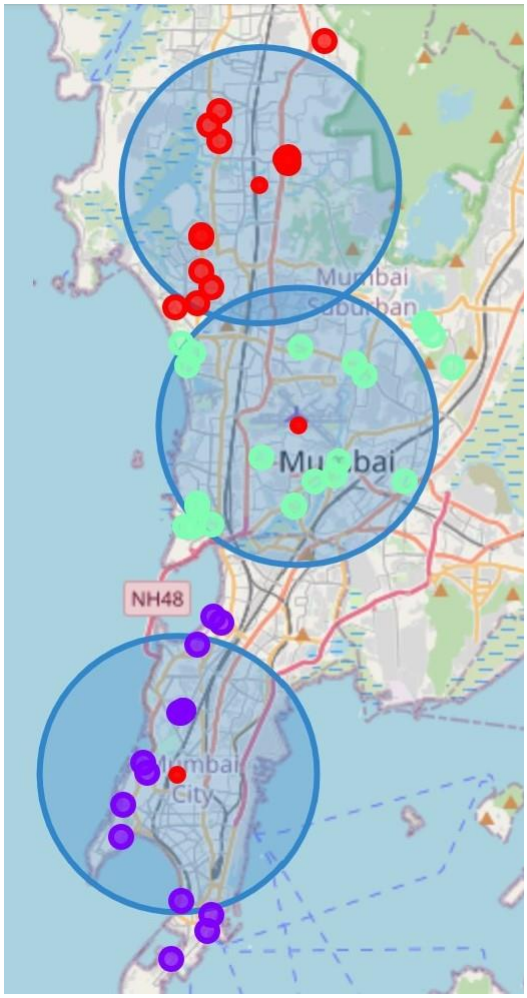


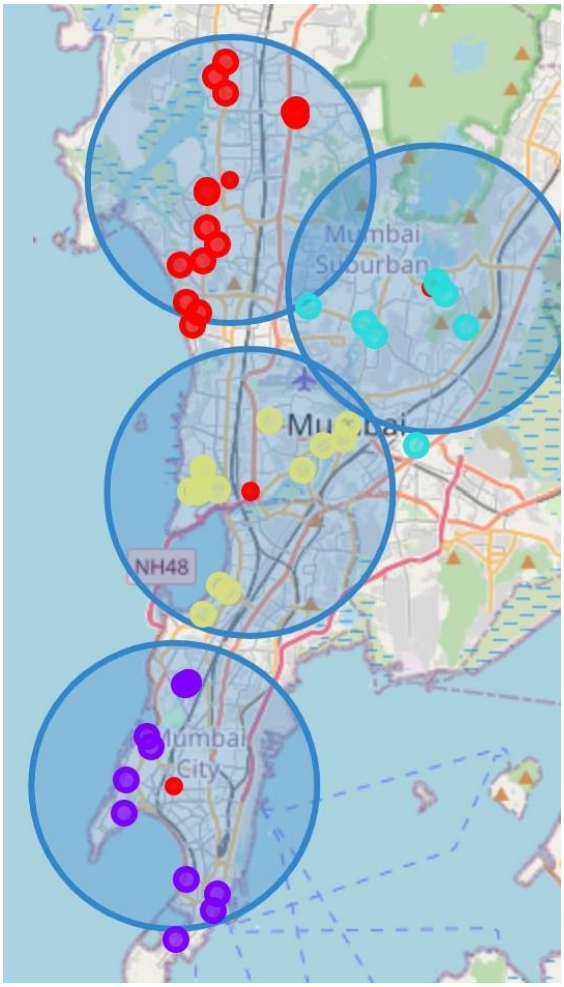
In the above plot, I observed the elbow of the graph is at k = 3. But after visualizing K= 4 gave the best results with minimal outliers, better coverage of the area, and least overlapping. I also labeled applicants to corresponding clusters using predict method.

To find out the ranking, I found out the distance of each applicant from the respective centroid and ranked them in ascending order.

5. Results findings:

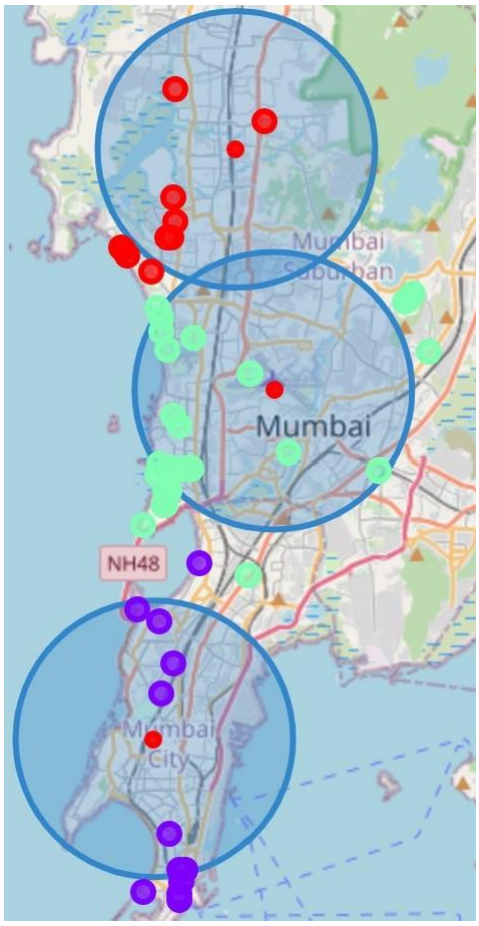
Following are the visualization details for k=3 and k=4.

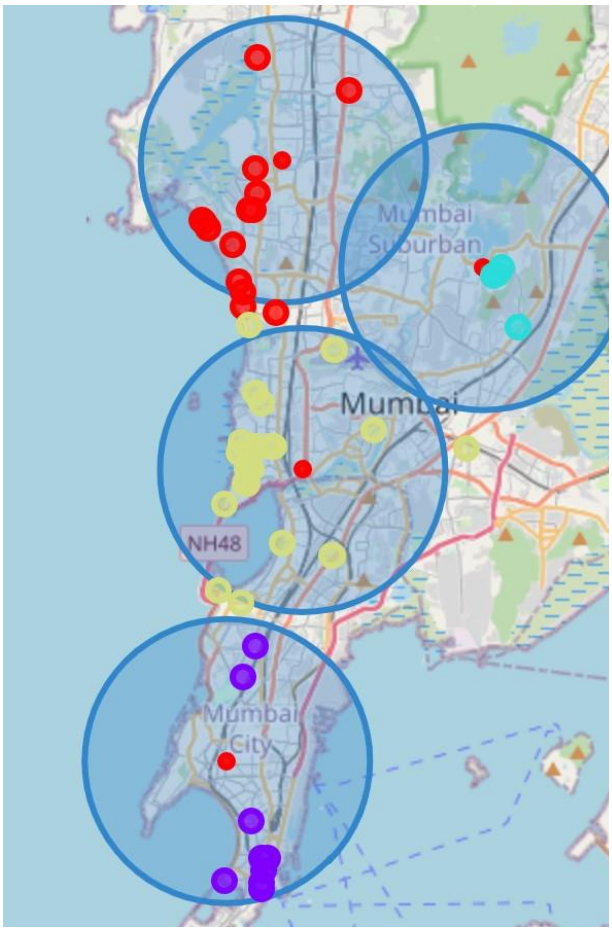




K= 3 **K=4**

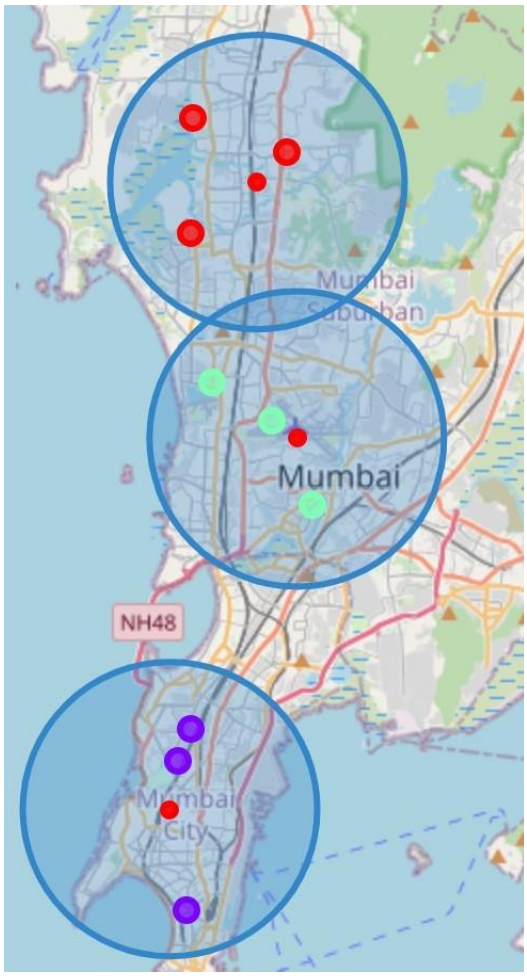
All pizza places clustered with respective centroids.

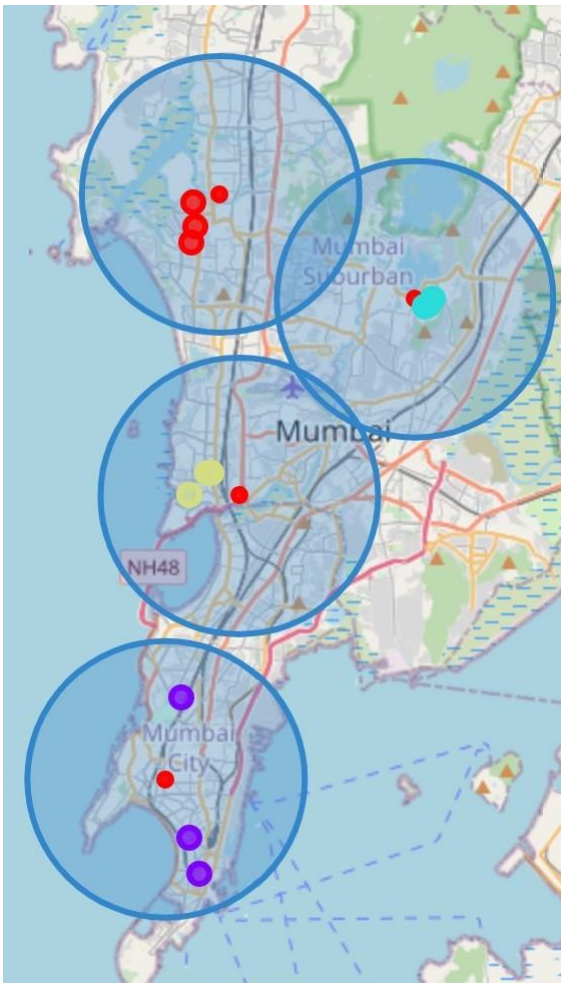




K= 3 **K=4**

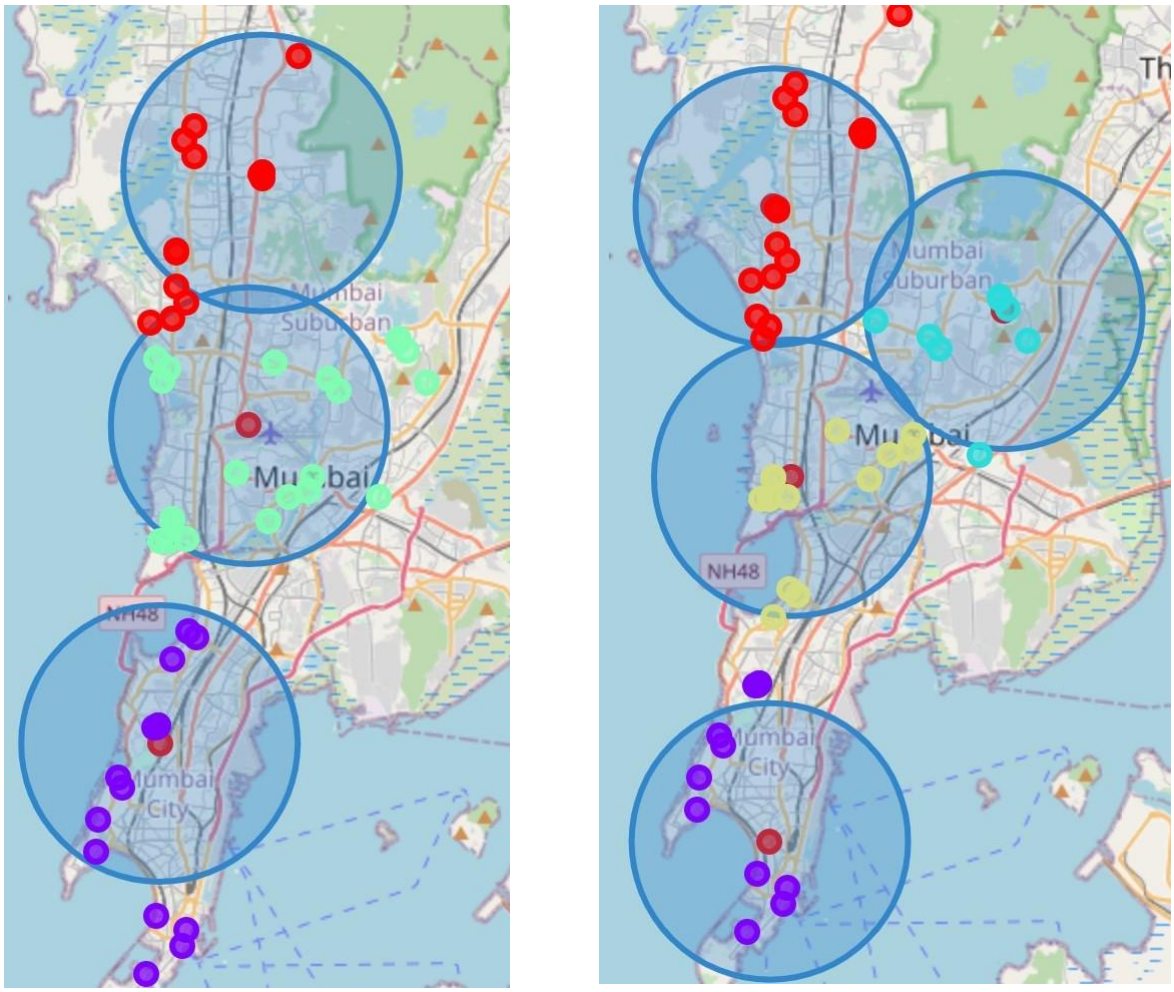
All applicants are divided into respective clusters.





K= 3 **K=4**

Top 3 Applicants around the centroids.



K= 3 **K=4**

Pizza places After changing center to the nearest applicant to the centroid.

After observing all the above Images, though the elbow method gave k=3 as the optimal value for K . In the real world, K = 4 gave the best results To evenly distribute and cover all areas and all competitor pizza places. Hence I selected K=4 to finalize the applicants.

6. Conclusion:

Dividing the city into four zones/clusters using the k-means algorithm, I achieved the following desired results.

1. All pizza places are within 20min (5km) reach from at least one of the selected franchise locations.
2. Less overlapping evenly distributed zones.
3. Successfully achieved to keep the minimum distance between two franchises, i.e., about 10km

7. References:

1. https://foursquare.com/
2. https://www.oalley.net/
3. https://www.brightlocal.com/research/local-business-travel-times/
4. https://www.statista.com/statistics/659362/time-consumers-would-use-to-drive-to-a-place-to-eat-us/

.