Searching for Optimal Locations for Cloud Kitchens in Mumbai

By Ameya Kulkarni

Email ID: ameya.kulkarni2014@gmail.com

Introduction

As a restaurant business owner, it's crucial to gauge how the pandemic has shifted consumer behaviour and their preferences towards food delivery versus dine-in.

The food delivery business got a shot in the arm during the government-imposed lockdowns. Apart from need fulfilment, the service also provided a sense of safety.

A report by Google and Boston Consulting Group estimated that the Indian food delivery market would reach $8 billion by 2022 [1].

The same report by Google and Boston Consulting Group also revealed another challenge for traditional restaurants: variety in cuisines (35 per cent) was one of the top reasons for recurrent use of online food ordering apps, other features like discounts and convenience comes later.

This means that a traditional restaurant is at a double disadvantage: managing increasing delivery costs and exploring multiple cuisines to stay popular in its area.

We can also safely assume that delivery executives shifting from the traditional model to this new 'gig' economy would seek some stability in monthly earnings. We can leverage this information to estimate the potential expense a food delivery business would typically incur for employing an delivery executive.

Recent developments like Zomato introducing a weekly 'minimum guarantee' payout corroborates our assumption. The new system is designed to ensure a minimum payment of INR 4,000-5,000 a week with a target of 60-65 hours a week [2].

Swiggy delivery executives in few metro cities in India had gone on strike in September'20 protesting against a 57% decrease in their per delivery payout and the removal of monthly incentives of INR 5,000 [2].

In Feb'21, Zomato increased its delivery executives' pay by 7-8% due to fuel prices touching INR 100 per litre. The company confirmed that a delivery executive travels about 100-120 km in a day and consumes 60-80 litres of fuel in a month, and the fuel price hike would result in monthly expenditures increasing by INR 600-800, which constitutes about 3% of their monthly income [3].

This places the average delivery executive's monthly income at around INR 25,000.

This can also be confirmed by viewing a job posting for such roles online.

We were lucky enough to find a listing of a restaurant business in Mumbai; the reason for the listing was rather griming: Covid-19.

It was a 3-year-old medium-sized restaurant, with about 30 employees. It registered sales of about INR 6-7 million and had an EBITA margin of 15%.

The restaurant was valued at INR 2.5 million, with a 5-year lease and additional physical assets of INR 1.5 million. It had an average order volume of 2,500 per month with an average order value of INR 400, resulting in average monthly revenue of INR 1 million.

If they had 7-8 delivery executives, their monthly expenses would be greater than INR 200,000, which is 20% of the revenue, and this doesn't include delivery asset costs [4].

The restaurant did mention that they relied on few food delivery start-ups for their delivery service.

Owning a fleet of delivery executives implies having the added expense of managing and maintaining manpower and delivery assets, without having any certainty of minimum revenue during daily operations and nowadays most restaurants rely on partnerships with food delivery start-ups/delivery firms to meet their delivery needs. Food delivery start-ups typically charge a commission rate of about 22-25% on order value for their service; the rate has only been increasing over the years [5].

This results in a shift from a fixed cost delivery model to a variable cost delivery model for the restaurants. The trade-off is moving delivery assets off the books, removing any and all delivery-related operating expenses, and making profit margins of the business more inelastic to decrease in order volume.

With the food delivery market slowly maturing, incentives and salaries for delivery executives will begin to match and exceed previous market levels.

It seems like the right time to explore a new type of model: the cloud kitchen model.

Rather than wasting space on a dining area, in a cloud kitchen model all space is utilised to optimise food production and delivery. These kitchen host delivery-only food brands.

This allows a single kitchen space to host multiple brands, therefore multiple types of cuisines, and route all deliveries through this single location. The model also works if some of the kitchen space is rented to others for setting up their brand. This may cannibalise your market, but you do also receive a cut in every order [6][7].

Another crucial aspect of the cloud kitchen model is that the business can completely focus on optimising delivery by selecting the right location for the business operations. The pandemic has resulted in vacancies in commercial properties, potentially making it the right time for exploring and setting up new locations based on this model [8].

For many businesses optimising transportation, a great cost-cutting exercise, and the benefits are easy to realise. With the changing times, it appears that the food delivery business is going to grow further.

Through this project, I hope to provide relevant data points and insights for selecting the best locations for setting up a cloud kitchen that maximises the serviceable population while minimising the service time and provide recommendations for the type of cuisines to serve and explore. The project will conclude the minimum number of cloud kitchen's locations that a business would need to set up to service the maximum population of the geography within an acceptable time limit and provide a list of cuisine recommendations for each location. We shall also provide some comments on whether such a cloud kitchen should rely on food delivery start-ups for their delivery service or setup their own.

Data

For this project, we have relied on the Foursquare location data (as instructed!) for details on existing restaurants, while for information about the city of Mumbai (location coordinates, populations, etc.), multiple websites have been scrapped.

The city of Mumbai is divided into administrative wards, and details about its population and areas are available on the web [9]. Unfortunately, location details about these wards are not available online. Pin code information about the ward offices was combined with latitude-longitude data of pin codes was used to identify the location of ward offices. It was assumed that these ward offices represent the central territory of each ward [10][11].

The city is also divided into suburbs with reliable location information, though inconveniently details about population and area are not [12].

We have used the suburb locations and the Foursquare location data to create a dataset of restaurants and other food establishments, their locations, and their category/type.

The category type serves as the base data for the cuisine recommendations. The Foursquare location data provided 70 unique categories, though the true number might be closer to 60 (once similar types are combined).

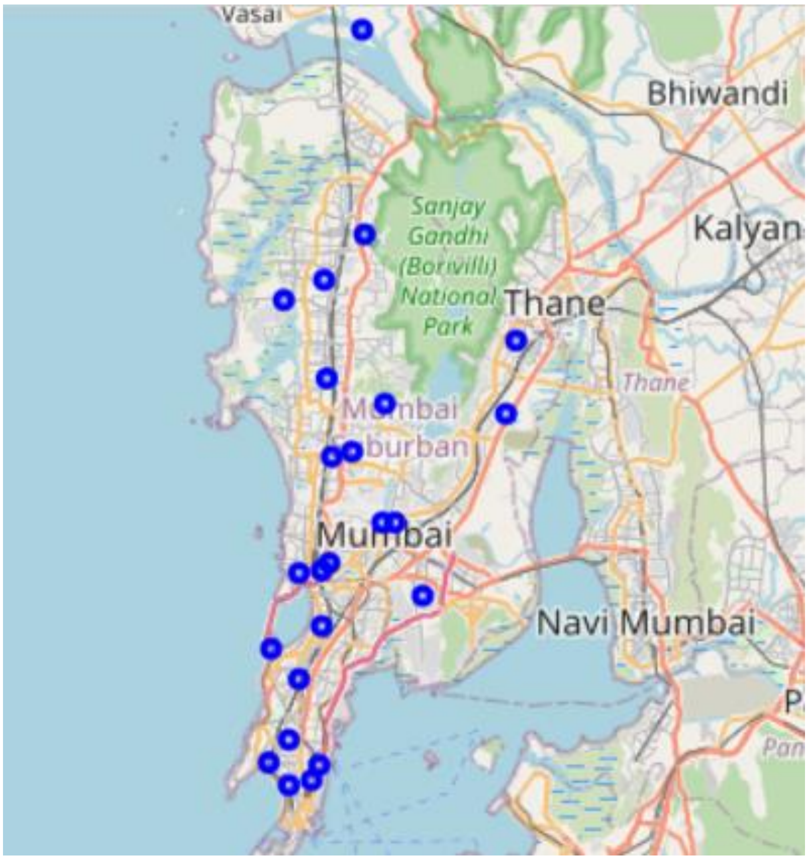
All other parameters required for this analysis has been derived from these datasets.

Overview of datasets

City-centre data:

- Total number of city centers = 24
- Total population of all city centers = 1,19,76,439
- Average city center population = 4,99,018
- Total area = 483 sq. km
- Average city center area = 20 sq. km
- Average inter-city center distance = 16 km
- Average number of venues within 15 km = 501
    - Dine-in: ~370
    - Café: ~50
    - Eatery: ~83
- Average delivery time from all venues = 1.02 hr
    - Average delivery time from venues within 15 km = 0.54 hr
    - Average delivery time by category (venues within 15 km)
        - Dine-in: ~ 0.54 hr
        - Café: ~0.55 hr
        - Eatery = ~0.55 hr

City-centre Locations (in blue)

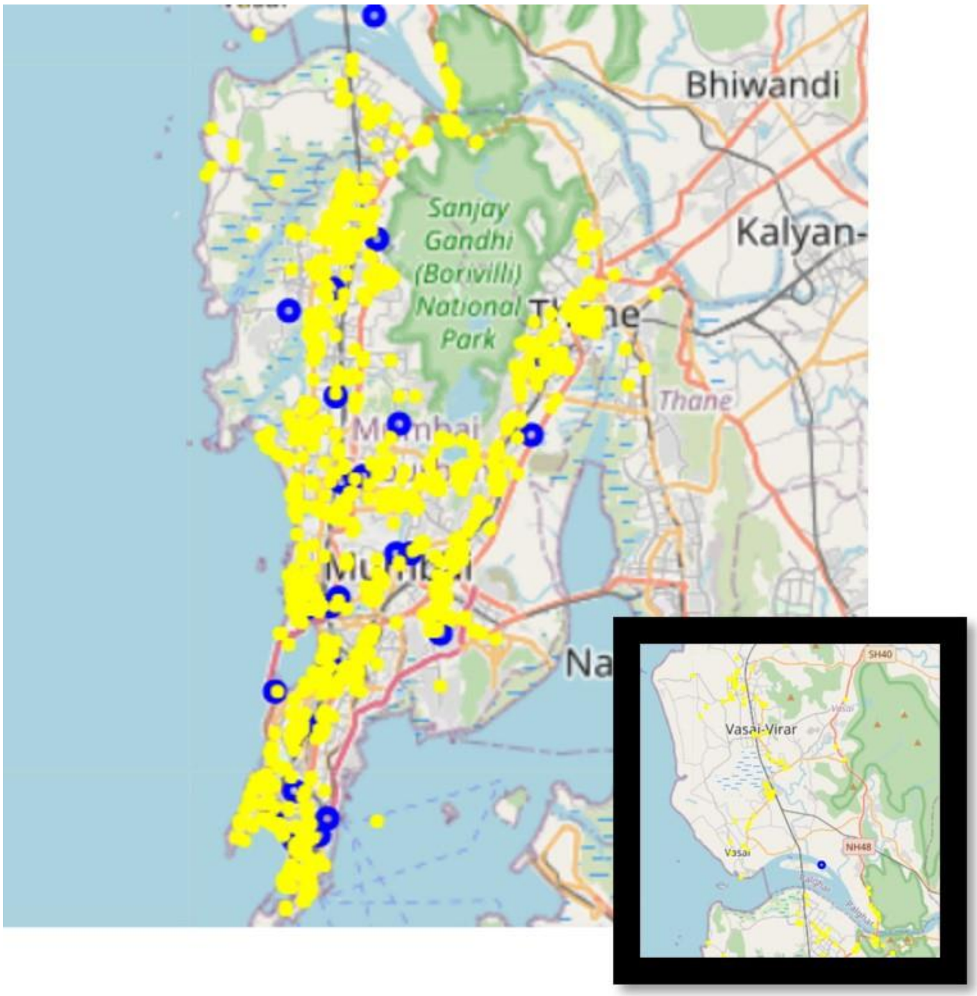


Venue location data:

- Total number of restaurants = 1,181
- Average number city centers within serviceable distance (15 km) = 9
- Average distance to city center within serviceable distance (15 km) = 7.2

Venue Type	Count	Average count of City centres within 15 km	Average distance to City centres within 15 km
Café	103	9	7.3
Dine in	844	8	7.2
Eatery	182	9	7.1

Venue Locations (in yellow)



Methodology

For a given set of city centres (population centres/neighbourhoods), an optimal location is at the right distance from each city-centre, such that the location can serve the maximum proportion of the total population of the set of city centres within a defined time limit. This is also referred to this as the serviceable population of the set of city centres.

We have assumed 45 mins or 0.75 hrs as the optimal time limit for delivery.

This raises two challenges: computing the time taken for servicing the population of a given set of city centres and a venue location and searching the area bounded by the set of city centres for the optimal location.

Computing the time taken to service a set of city centres

We have assumed a simple model of delivery. A delivery executive has to first travel to the *Location* of the restaurant/eatery. From there, the delivery executive must travel to a location near the *City-centre*. Based on a few assumptions, we can compute the population of the city-centre that the delivery executive can service within the predefined time limit.

The 45 min time limit implies that a delivery executive who receives an order, must reach the restaurant, wait for it to be prepared, and then collect the order and then travel to the residence of the customer to deliver it.

Assuming that the average speed of a delivery vehicle is 20 km/hr, that would result in a maximum distance of 15 km (D_{max}) that the delivery executive can travel without breaching the 45 min time limit.

To approximate the road distance an executive has to travel we have defined distance as the sum of its x-y or NS-WE components. This is based on the assumption that most roads are on an average parallel to the NS-WE axis.

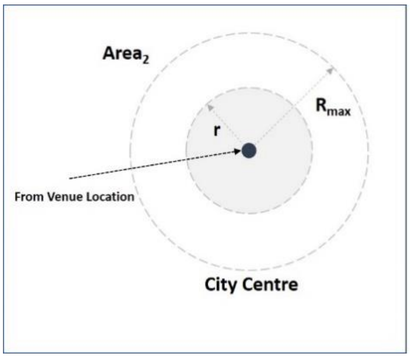
As the delivery executive is not generally waiting near any particular restaurant anticipating only those specific orders, we have assumed that the executive must travel an average initial distance of 2.5 km (D_0) to reach any restaurant.

This results in the maximum possible distance that the executive can travel to be,

= 15 km – 2.5 km = 12.5 km

We have assumed that the population of the city-centre is radially distributed around the location of the city-centre. Furthermore, we have defined the population density function such that all the population is concentrated within the area of the city-centre location.

We have been given sq. km values for each city-centre, and we assume a circular shape for the city-centre geography with radius R_{max}.



Following is the population density function used,

$P(r) = C_1 + C_2 \times r^2$

, r is the distance from the city-centre location.

To arrive at the potential serviceable population, we assume that all executives must travel to the city-centre location and then head out radially towards the customer's residence.

This additional constraint provides us with the simple delivery model mentioned earlier.

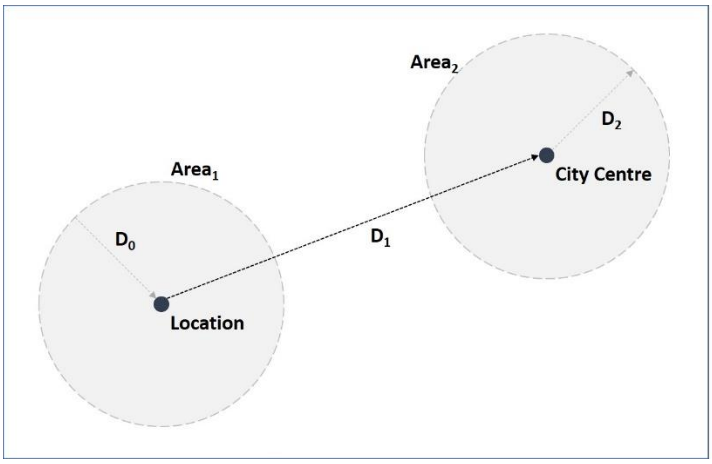
The maximum possible distance that the executive can travel, as computed above, is 12.5 km ($D_{max} – D_0$).

After travelling to the city-centre location (D_1, as shown in the diagram), the executive has a max range of D_2 that he/she can travel; the value is computed as,

$D_2 = D_{max} – D_1 – D_0$

If D_2 is greater than R_{max}, the value of D_2 is capped at R_{max}.

This provides us with a serviceable area, which combined with the population density function gives us the maximum population of the city-centre that the executive can service. We also refer to this as the serviceable population of the city-centre.



To compute delivery time to a particular city-centre from a venue location, we simply integrate over the city-centre area starting from the centre and going radially outward. This computation needs to be performed for all city centres within the set.

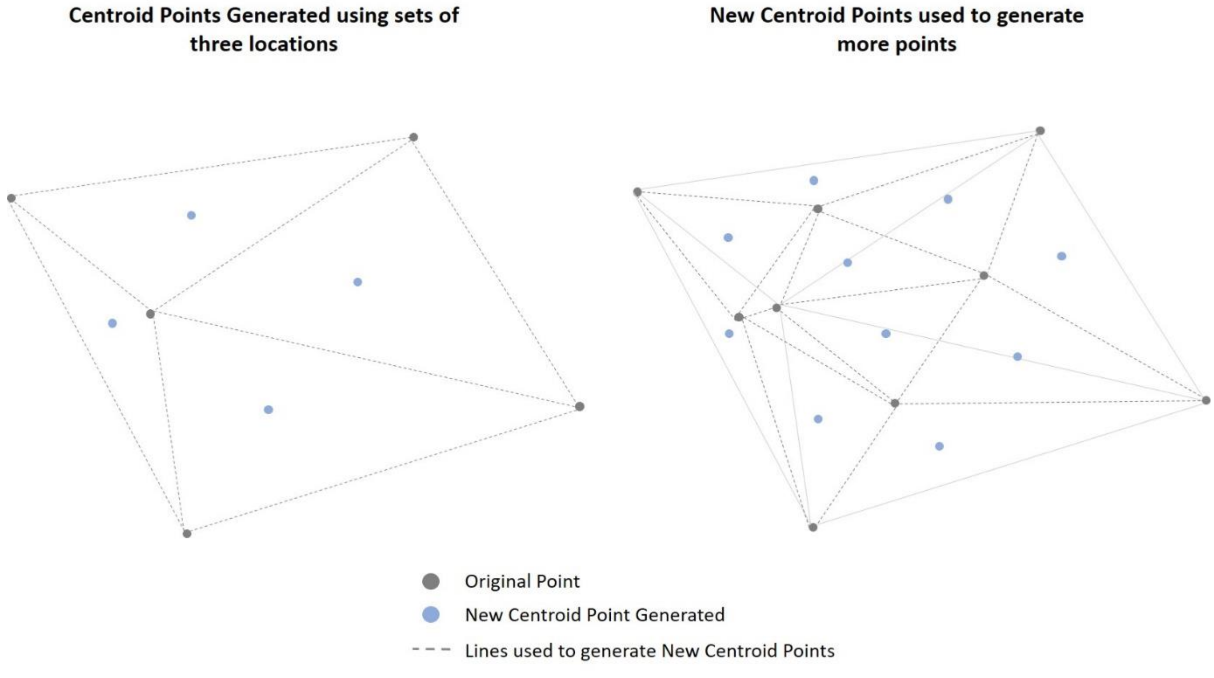
To arrive at the time taken to service a set of city centres from a venue location (restaurant), we simply take the weighted average using the population as a weight.

Search the area bounded by the set of city centres for the maximum

The next task would be to scour through multiple locations within the set of city centres to identify the optimal location that has the minimum service time for that given set of city centres.

Logically, such a point would exist within the area bounded by the city centres.

To search through this area, we compute centroids using the three city-centre locations with a 15 km limit on the length of any side of the triangle. Additionally, we also ensure that the area of the triangle is larger than 1 square kilometre and the minimum distance is 5 km. These centroids act as new points, and the centroid generation task is performed repeatedly with a decreasing limit on the length of a side, and an increased limit of minimum area. Then the above exercise is performed, and the total serviceable population for each point and given set of city centres is computed.



Note: Not all triangle generates are used to generate new centroid points, each triangle has to meet an upper limit of side length and a lower limit on area

The maximum value and its corresponding location represent the minimum service time for the given set of city centres and the optimal location.

Accuracy of the optimum depends on the depth of search within the area bounded by the set of city-centre locations a generating dense cluster is computationally expensive and time-consuming; thus, a faster and more efficient scoring mechanism is required.

Model for rating a location

To train the model, we have utilised the restaurant locations from the Foursquare location data. We computed the maximum potential serviceable population for each restaurant's set of city centres within a 15 km distance limit and generated a rating for each.

The rating of a location is defined as follows,

$$rating = \frac{(\text{minimum service time for a given set of city centres})}{(\text{service time of a location for given set of city centres})}$$

We create a multivariate polynomial regression model to help us compute the rating of a location and its set of city centres (defined above as city centres within 15 km).

The model takes certain location attributes as inputs and generate a rating for the location.

The inputs are the following three attributes,

1. Number of city centres within a serviceable distance
2. Average distance between the location and the city centres
3. Average potential serviceable population

The target variable is the rating of a location, as defined above.

We tested Linear Regression and Ridge Regression model, reduced model attributes.

We split the data into training and testing sets using sklearn.model_selection, then pre-processed the location attributes using sklearn.preprocessing.

In pre-processing, we used PolynominalFeatures to convert the data into a 6th order polynomial.

Searching for optimal locations

We generate a grid space over the entire area bounded by city-centre locations using the method mentioned above and then use the model to compute the ratings for each point in the grid space.

Locations with ratings close to 1 are considered optimal locations (for its particular set of city centres).

We then identify the minimum number of top-rated locations a business would need to set up cloud kitchens at, to service the maximum proportion of the population of the city of Mumbai.

We begin by sorting the city-centre list in ascending order based on how many top-rated locations are near it, then start selecting locations from those nearby top-rated locations.

Once a location is selected, we update the population numbers for each city-centre in its vicinity. We have considered that a cloud kitchen can service 70% of the population at any given city-centre.

If the population of a city-centre is completely satisfied, then all other top-rated locations that can service the city-centre are dropped from further search.

As the sorting logic skews the search towards city centres that are an outlier, we also try sorting in descending order, and few other combinations in between.

The set of best locations that have minimum unserved population is considered as the optimum.

Cuisine recommendations for best locations

Once we have obtained the list of best locations, we simply check the venue categories available at all city centres near our best locations.

We use the Foursquare location data to compute the average venue category at each city-centre, and then use the population of the city centres to compute the weighted average for the best location.

Results

The Linear Regression model with 3 inputs performed better than the other combinations.

Testing the model of 500 newly generated samples gave a mean squared error of 0.000700.

Model Performance

(A) Linear Regression, Polynomial of order 6, 3 inputs

(B) Linear Regression, Polynomial of order 6, 2 inputs

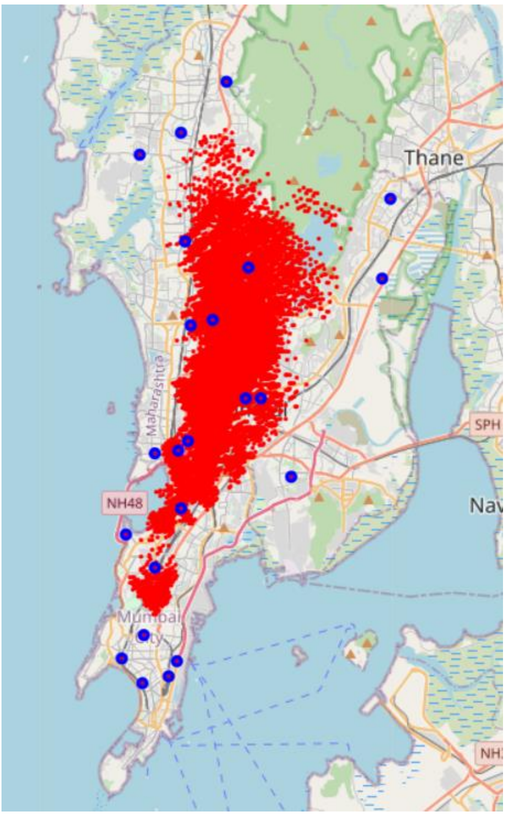
(C) Ridge Regression (alpha = 0.1), Polynomial of order 6, 3 inputs

(D) Ridge Regression (alpha = 0.1), Polynomial of order 6, 2 inputs

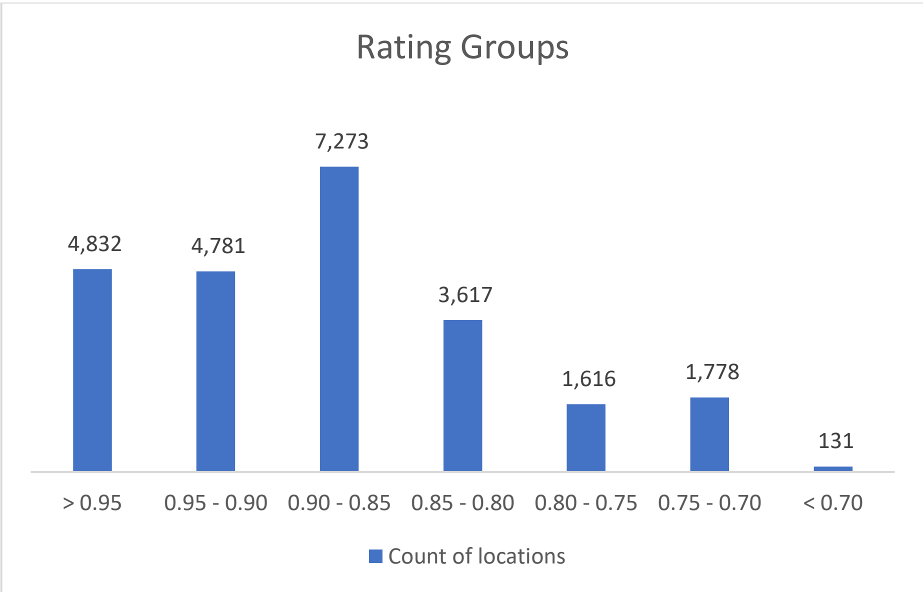
Model Evaluation	**(A)**	**(B)**	**(C)**	**(D)**
R^2	0.995311	0.960776	0.994489	0.960772
R^2	0.992297	0.958375	0.991564	0.958402
MSE	0.000456	0.003815	0.000536	0.003815
MSE	0.000636	0.003437	0.000696	0.003434

Parameter Evaluation (R^2)	(A)	(B)	(C)	(D)
Number of city centres within serviceable distance	0.913689	0.913689	0.913687	0.913687
Average distance between location and the city centres	0.317266	0.317266	0.317266	0.317266
Average potential serviceable population	0.194955		0.194954	

With 24 city centres as starting points, we generated a search space of 24,347 locations.



Using our model, we identified 4,832 points as top rated (rating greater than 0.95).



The model did rate 1.3% of the total locations with a rating greater than 1. We have ignored this from our analysis.

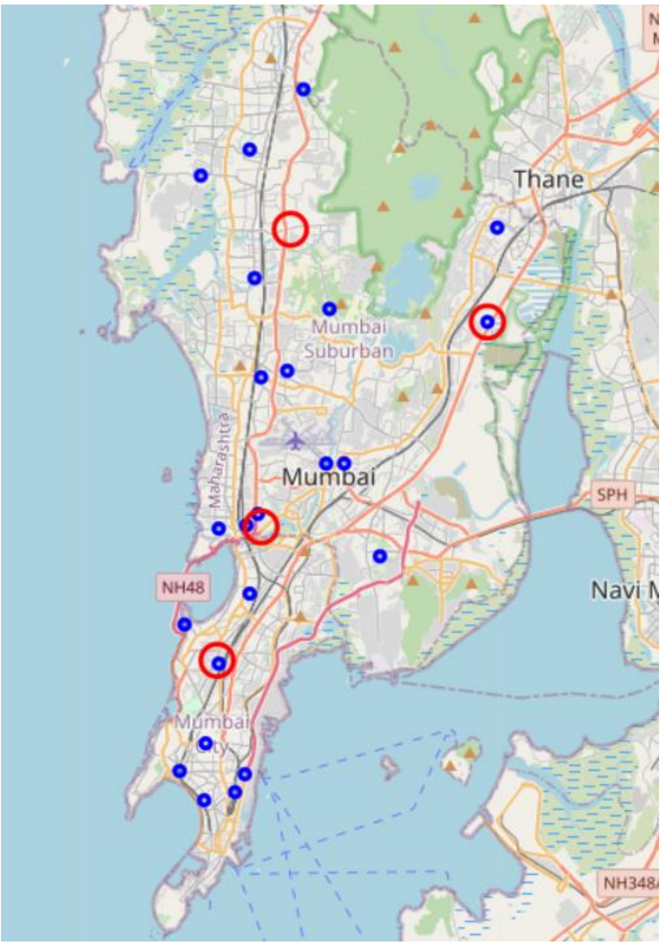
The analysis provided n locations as the minimum number of top-rated locations a business would need to set up cloud kitchens at, to service the maximum proportion of the population of the city of Mumbai.

City-centre Wise Details (considering best locations within 15 km)

- Average service time = 0.67 hr
- Average distance to best locations = 8 km
- Proportion of population unserved = 20.82%

Best Location Wise Details (considering city centres within 15 km)

- Average service time = 0.70 hr
- Average distance to city-centre = 9
- Average inter-best location distance = 18 km



For each location identified, we computed the top 5 venue categories that are frequently available at the nearby (within 15 km) city centres, bottom 5 venue categories that are infrequently available, and list of venue categories that are unable at any of the city centres.

Location	Top 5	Bottom 5	Absent
1	Indian Restaurant Café Fast Food Restaurant Chinese Restaurant Restaurant	German Restaurant Modern European Restaurant Fish & Chips Shop Brazilian Restaurant Multicuisine Indian Restaurant	Dhaba, English Restaurant, Cafeteria, Molecular Gastronomy Restaurant, Parsi Restaurant, Steakhouse, Chaat Place, Portuguese Restaurant
2	Indian Restaurant Café Fast Food Restaurant Chinese Restaurant Restaurant	Dhaba German Restaurant Multicuisine Indian Restaurant Gluten-free Restaurant English Restaurant	Cafeteria, Molecular Gastronomy Restaurant, Parsi Restaurant, Steakhouse, Chaat Place, Portuguese Restaurant
3	Indian Restaurant Café Fast Food Restaurant Chinese Restaurant Restaurant	English Restaurant Chaat Place Burrito Place Dumpling Restaurant Parsi Restaurant	Dhaba
4	Indian Restaurant Café Chinese Restaurant Fast Food Restaurant Bakery	North Indian Restaurant Creperie Afghan Restaurant Comfort Food Restaurant Mughlai Restaurant	Dhaba, English Restaurant, Burrito Place, Dumpling Restaurant

Currently the location selection has ensured a minimum of service time of 0.67 hr versus our initial benchmark of 0.75 hr

Discussion

The model allows us to select a location at random and quickly gauge its optimality for setting up a cloud kitchen model.

The time-consuming part of the project was generating the search space for the best locations. The algorithm was also utilised in computing the minimum service time for a given set of city centres. The centroid grid search approach is only limited by computational power and the number of seed points that we provide. By lowering the limit of the size of the triangle formed by points on the map, we can infinitely generate more points for search.

In our analysis, we used 24 city centres as seed points for generating the search space; purely since we had population and area size information about these locations.

To generate 24,347points, it took my system (i5, 8 GB memory) about 30 mins.

The distance-travel time computations and population density functions we have assumed leave much to be desired. If we could leverage Google Maps data, the inaccuracies in the distance-travel time computations can easily be resolved. For the population density function, if area boundary coordinates were made available, service time computation can be further refined.

The analysis also recommends which type of venue categories/cuisines the cloud kitchen can focus on. A business owner could focus on existing strengths or focus on the bottom 5 or the ones completely absent from the area. This also provides the owner with the opportunity to standardise all kitchens or allow variability among different kitchens based on local preferences.

The recommendation engine considers the presence of a type of venue category/cuisine near the city-centre as being analogous to the population showing a preference for it. Such an assumption is made due to the lack of restaurant-level order volume data.

Today’s food delivery start-ups are flush with this information, and thus can use predictive analytics to identify trends to make more accurate recommendations on present and upcoming consumer cuisine preferences.

As the cloud kitchen model can host multiple brands with different cuisines, it is up to the owner to try and explore which cuisine combinations work best for a given location.

The key aspect in creating a set of best locations was approaching it through the lens of population serviceability at a given city-centre: the goal was to ensure that we don’t deliberately overserve any city-centre. Hence while building the set of best locations, if any city-centre had all its population served, all locations that could serve the city-centre were dropped from consideration. Another key assumption was a cloud kitchen location could service about 70% of the population of the city-centre.

This number represents the size of the cloud kitchen: more space would imply more space for different cuisines and more capacity for producing orders.

Reducing this size would increase the number of locations as the algorithm would strive to maximise the proportion of the population it can service.

For the 4 best locations generated by our analysis, the proportion of the unserved population is 20.82%. Further investigation revealed that only 1 city-centre’s population has not been serviced at all, it is an outlier with an average inter-city-centre distance of 31 km; this implies that our location searching algorithm was able to build a set of locations that maximised serviceable population while ensuring minimum service time.

Conclusion

Our *Simple Delivery Model* can simulate possibilities of a delivery executive makes to deliver to locations near the city-centre, but it doesn't consider an executive making multiple deliveries during a single journey.

Food delivery start-ups are known to use the 'single journey multiple orders' tactic as it helps in reducing operating expenses (like fuel, maintenance) for the delivery executive while increasing revenue for the start-up from delivery charges.

As cloud kitchen hosts multiple brands with different cuisines, it can have a delivery executive make multiple deliveries in a single journey. This makes the demand for delivery executives less sensitive to fluctuations in order volume. The business can reduce their delivery costs without reducing an executive's compensation; it can simply reduce their delivery staff.

This also makes owning the delivery operation more competitive versus relying on food delivery start-ups.

As mentioned before, the project has focused on optimising the location selection aspect of the cloud kitchen model. The project structure and model can easily be tweaked to optimise location selection for other types of problems like selecting locations for setting up hospitals or schools.

Links

1. https://www.livemint.com/technology/tech-news/indian-online-food-delivery-market-to-hit-8-bn-by-2022-report-11580214173293.html
2. https://inc42.com/buzz/zomato-new-compensation-model-gig-economy/
3. https://thelogicalindian.com/uplifting/zomato-increases-delivery-executives-pay-to-make-up-for-sky-rocketing-fuel-prices-27066
4. https://www.smergers.com/business/food-delivery-business-for-sale-in-mumbai-india/t0dru/
5. https://www.thehindubusinessline.com/companies/amazon-foods-reduced-commission-rate-may-hurt-swiggy-zomatos-growth/article34036997.
6. https://www.thefoodcorridor.com/2019/12/05/everything-you-need-to-know-about-cloud-kitchens-aka-ghost-kitchens-in-2020/
7. https://limetray.com/blog/cloud-kitchen-business-model/
8. https://www.cushmanwakefield.com/en/india/insights/mumbai-marketbeat
9. http://www.demographia.com/db-mumbaidistr91.htm
10. https://indiamapia.com/Mumbai/Mumbai.html
11. https://memumbai.com/wards/
12. https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Mumbai