

Finding optimal location for a market in HYDERABAD

G.S.ROHITH

Introduction

Background

Opening a physical store is not without risks. One the most obvious risks to evaluate before opening a new store is receiving enough customers to make the business profitable. Therefore it is essential to pick the right location to make sure it is convenient for the customers and there is enough demand for a new store in the area.

Business problem

In this project, we tackle the problem of a low-cost supermarket chain trying to decide in which area of Madrid (Spain) they should open their new store in order to maximize the revenue. It is important to note that the city of Madrid consists of 21 districts and 131 neighborhoods with great differences between them.

The goal is to identify the optimal neighborhood for opening a store taking different factors into consideration such as the types of neighborhood (a residential area would be ideal), the amount of people living in those areas (the higher the population the higher the food demand

Using data science, geospatial analysis and machine learning techniques, this project aims to provide a solution for this problem and recommending the best neighborhood for opening the low- cost supermarket.

Data

The following sections describe the data that is needed for answering this business question.

Wikipedia

The first data that we need is the list of neighborhoods in Hyderabad. The Wikipedia page “List of neighborhoods of Hyderabad” shows a table with the name of each neighborhood for each of the district. In our project, we will work directly with the neighborhoods and ignore the districts since this way we can perform a more granular analysis of the areas.

Geospatial data

Since the plan is to target residential areas, we need to analyze the type of food venues present in each neighborhood. With the Foursquare API we can explore the different food venues, considering that a big density of bars and restaurants over very few supermarkets will most likely refer to a business or recreational area where people don't usually buy at supermarkets. In the other hand, a large proportion of supermarkets over the rest of food venues might indicate it is a residential area where people normally make their food shopping.

Before we can make use of the Foursquare API we need to convert the neighborhood names into a pair of latitude and longitude coordinates. We can query the Foursquare API using the HTTP GET method on the explore endpoint indicating the geographical coordinates, venue categories and radius.

The following figure shows the location of all the neighborhoods within the city of Madrid, it is important to remark that any areas and towns outside the city have not been considered for this project.

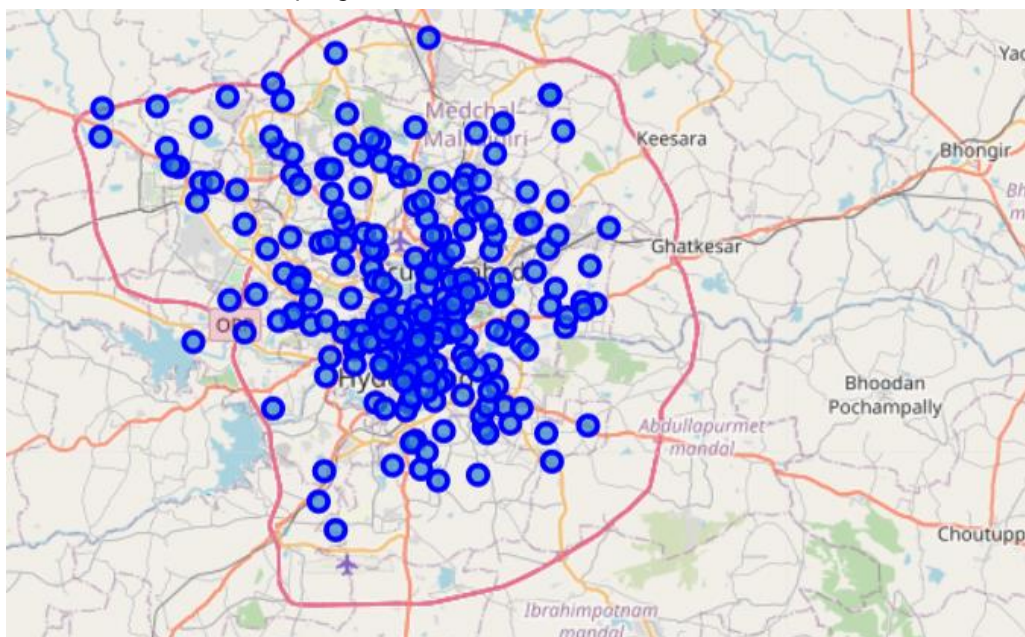


Fig 1: Hyderabad neighborhoods

Census data

Finally we will need data from the census of Madrid. We can obtained this data from that are accessible from the geo iq website. Particularly we are interested in the population

Below we can see the type of information that is required for this analysis.

Neighbourhood	Population	Latitude	Longitude
AMBERPET	181589	17.38582	78.51836
SHAIKPET HYDERABAD	123675	17.40373	78.39310
SAROORNAGAR	122440	17.35442	78.53921
PRAKASH NAGAR	99543	17.36069	78.41634
BEGUMPET	92340	17.44729	78.45396

Table 1: Geographical and census data

Methodology

Neighborhood segmentation

Since initially we don't know how many different types of neighborhood we can find in Hyderabad, we are using the elbow method to obtain the optimal number (k) of clusters.

Although the figure below shows 2 as the optimal number of clusters, we are using 5 (the second best k) since this way we can break down more the number of neighborhoods that we are going to analyze against the census data.

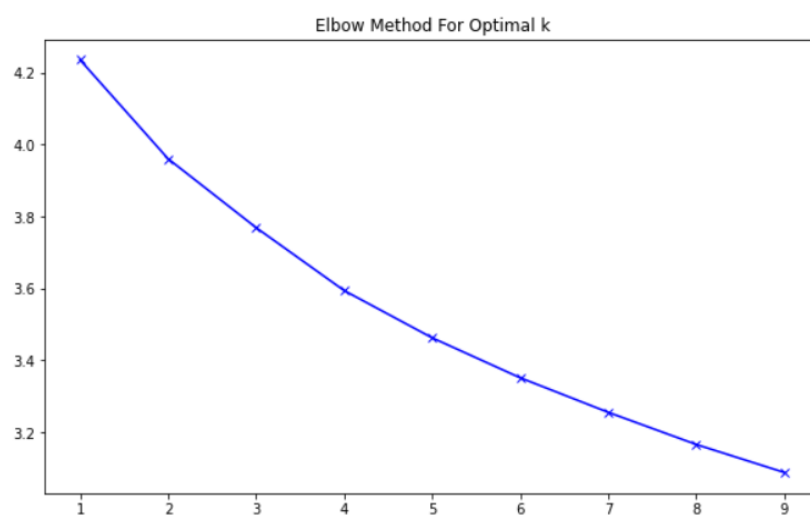


Fig 2: Elbow method for the optimal number of clusters

With K-means algorithm we can group the 243 neighborhoods into 5 clusters depending on the most popular venues in those areas. Below we can visualize what are the most common venues and the cluster (from 0 to 4) that has been assigned to each area.

	Neighbourhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	AMEERPET	17.43535	78.44861	4	Bakery	Women's Store	Electronics Store	Miscellaneous Shop	Clothing Store	Building	Indian Restaurant	Temple	Mobile Phone Shop	Residential Building (Apartment / Condo)
1	BEGUMPET	17.44729	78.45396	3	Residential Building (Apartment / Condo)	Office	Indian Restaurant	Temple	Café	Restaurant	Coworking Space	Dentist's Office	Building	Housing Development
2	SR NAGAR	17.49808	78.46963	4	Bank	Pharmacy	Department Store	Breakfast Spot	Automotive Shop	Salon / Barbershop	Housing Development	Government Building	Clothing Store	Motorcycle Shop
3	PRAKASH NAGAR	17.36069	78.41634	3	Residential Building (Apartment / Condo)	Bank	Indian Restaurant	Fast Food Restaurant	Electronics Store	Market	Bakery	Restaurant	Grocery Store	Chinese Restaurant
4	PUNJAGUTTA	17.42616	78.45210	0	Indian Restaurant	Multiplex	Jewelry Store	Fast Food Restaurant	Bank	Women's Store	Electronics Store	Kids Store	Miscellaneous Shop	Professional & Other Places

Table 2: Neighborhood clusters and most popular venues

	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	BEGUMPET	Residential Building (Apartment / Condo)	Office	Indian Restaurant	Temple	Café	Building	Restaurant	Dentist's Office	Coworking Space	Housing Development
3	PRAKASH NAGAR	Residential Building (Apartment / Condo)	Bank	Grocery Store	Restaurant	Indian Restaurant	Bakery	Electronics Store	Market	Fast Food Restaurant	Café
5	BALKAMPET	Residential Building (Apartment / Condo)	Indian Restaurant	Temple	Office	Café	Bakery	Automotive Shop	Bank	Coworking Space	Hospital
7	BHARAT NAGAR	Residential Building (Apartment / Condo)	Bank	Indian Restaurant	Housing Development	Bar	Mobile Phone Shop	Temple	Salon / Barbershop	Bus Station	Event Space
8	ERRAGADDA	Bank	Residential Building (Apartment / Condo)	Doctor's Office	Automotive Shop	Office	Bakery	Bus Station	Housing Development	Bar	Hospital
9	BORABANDA	Residential Building (Apartment / Condo)	Bank	Bakery	Automotive Shop	Office	Department Store	Convenience Store	Clothing Store	Temple	Tech Startup
10	MOTI NAGAR	Residential Building (Apartment / Condo)	Bank	Office	Department Store	Convenience Store	Asian Restaurant	Housing Development	Indian Restaurant	Automotive Shop	Bakery

Table above shows the cluster 3

In order to name the different clusters it is necessary to explore the most representative venues in the neighborhoods of each cluster. Finally we come with the following representation:

Cluster	Type of area	Common venues
0	Recreation	Restaurants
1	Industrial	Factory
2	Education	Colleges
3	Residential	Houses
4	Business	Offices

Table 3: Types of neighborhood

Table 3: Types of neighborhood

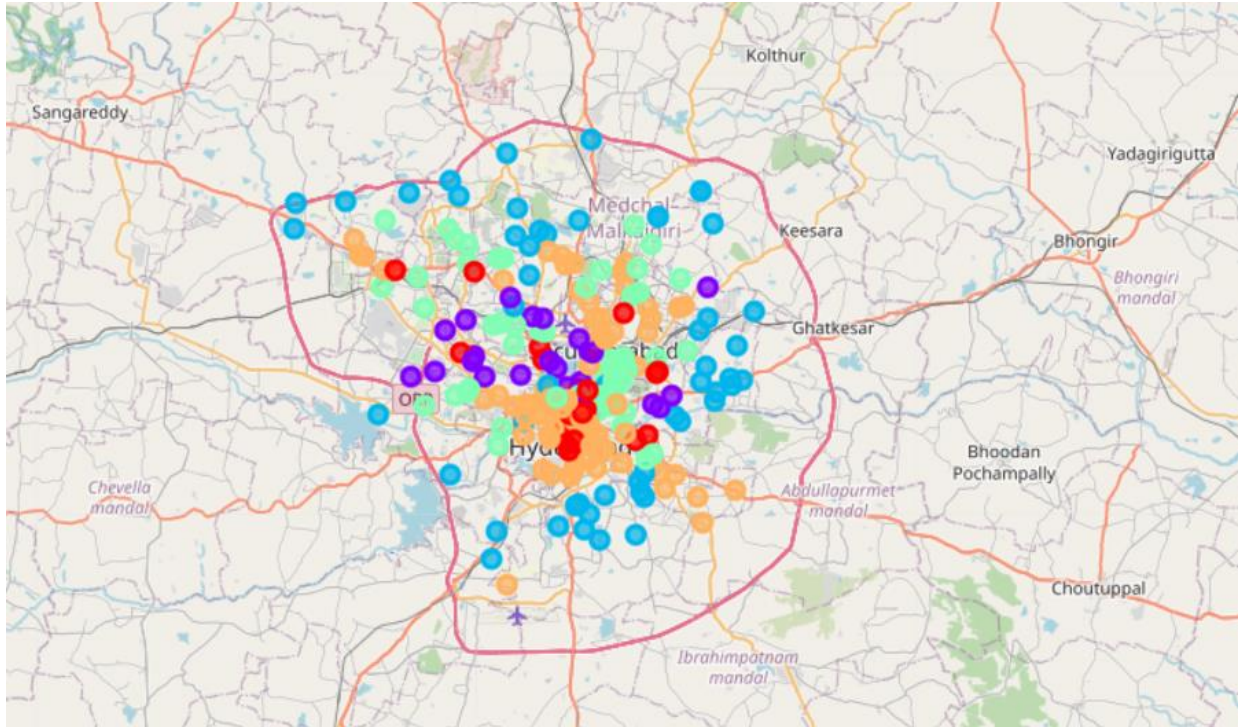


Fig 3: Map of Madrid with different clusters of neighborhoods

Therefore we will analyze the cluster 3 (light blue in the map above) in detail since this is the one that refers to residential areas. This cluster contains 56 neighborhoods.

Census analysis

Merging the geographical data of the residential neighborhoods together with the census data we can know the count of population that exists in each neighborhood. This will help us identify the neighborhoods where the population is high so that demand for supermarket may increase

Neighbourhood	Population	Latitude	Longitude
AMBERPET	181589	17.38582	78.51836
SHAIKPET HYDERABAD	123675	17.40373	78.39310
SAROORNAGAR	122440	17.35442	78.53921
PRAKASH NAGAR	99543	17.36069	78.41634
BEGUMPET	92340	17.44729	78.45396

Table 4: Geographical, census and market data for residential neighborhood.

Results and discussion

The following table represents the residential top 10 neighborhoods in cluster 3 considering the population

	Neighbourhood	Population	Latitude	Longitude
11	AMBERPET	181589	17.38582	78.51836
27	SHAIKPET HYDERABAD	123675	17.40373	78.39310
47	SAROORNAGAR	122440	17.35442	78.53921
1	PRAKASH NAGAR	99543	17.36069	78.41634
0	BEGUMPET	92340	17.44729	78.45396
32	NIZAMPET	84398	17.51831	78.38189
10	VIDYANAGAR	81023	17.40458	78.51031
9	BAGH LINGAMPALLY	79623	17.39931	78.49964
29	HAFEEZPET	79148	17.48990	78.35220
44	A. S. RAO NAGAR	78194	17.41120	78.50824

Fig 4: The top 10 neighborhoods in the cluster 3 w.r.t population

Above we can clearly see that Amberpet neighborhood having the highest population around 181k and we also know that the cluster 3 is the residential area so it is profitable to open a new store in the above areas.

Conclusion

The neighborhoods of Hyderabad were analyzed with the purpose of finding the ideal location for a supermarket. We applied machine learning techniques such as k-means clustering to find different clusters so that we could focus in only one type of neighborhood (residential). Further data such as population and market venues have been used to reduce the number of potential areas.

This project could be improved by only taking certain venue categories into consideration when performing the clustering segmentation. We could for example identify the key types of venue that define a residential area such as schools, pharmacies, small markets and corner shops, and the types of venue that discard a residential area such as night clubs, theatres and so on. Another improvement could be achieved by only handling certain groups of ages and social classes that would normally shop in a supermarket.

Although this project focuses particularly in a supermarket, it could easily be amended for any type of business and city, as long as the corresponding data are available to be included in the analysis.

References

Wikipedia article: https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Hyderabad

Foursquare API: <https://developer.foursquare.com>

Geo iq (for population): <https://geoiq.io/data>