2020

Identifying under developed areas for future investment in Jaipur

Prakhar Garg

MBA, Finance

6/20/2020

Table of Contents

EXECUTIVE SUMMARY

As the metro cities started getting congested, there is growing trend of shifting of companies to Tier-1 cities. This opens up a opportunity to real-estate developers and property investors to identify areas in these cities, which are under developed, early before the general public for earning good returns.

With this objective, this study is conducted on, one of the upcoming metro cities, Jaipur to identify the less developed area using data analytics and machine learning algorithms like clustering. Three clusters of 'Most Developed', 'Moderately Developed' and 'Under Developed' areas are formed using K-means clustering.

This project founds 12 such areas in Jaipur which are underdeveloped and could be a potential profitable investment for property developers.

INTRODUCTION

With the rising population of metro cities and increasing commuting times, there is a growing trend of companies shifting their base to Tier 1 cities and in turn resulting in their development. Thus, identifying the less developed areas in these cities could result in handsome profits for real estate developers.

Jaipur is one of those new-aged cities attracting India Inc. towards it. It is projected to record real GDP growth of 8.7% during 2015-2020 and set to emerge as the new A-class metro with a combined household income of ₹80,000 crores by 2021. Therefore, finding the neighborhoods, which are underdeveloped, early could provide good returns to the property investors.

Business Problem

The objective of this study is to find the under developed areas in upcoming metro city Jaipur for investment purpose. Using data science methodology and machine learning techniques like clustering, this study aims to provide solution to answer the following business question: **What are the neighborhoods that are currently under-developed and could be a potential investment opportunity in Jaipur City?**

Target Audience

This study will help real-estate developers and property investors to identify potential investment opportunity in the form of underdeveloped areas in upcoming metro city Jaipur.

DATA

Data Required

Following data is needed to solve the business problem:

- List of Neighborhoods in Jaipur. This defines the scope of this study which is confined to the city of Jaipur, capital city of Indian state of Rajasthan.
- Latitude and Longitude Coordinates of those neighborhoods. This is required in order to plot the maps and get the venue data using FourSquare API.
- Categories and Venues data using Foursquare API. These data will be used to cluster the neighborhoods.

Data Sources

Following are the sources of our data:

- Maps of India website (https://www.mapsofindia.com/pincode/india/rajasthan/jaipur)
- Geocoder package will be used to get the latitude and longitude coordinates for each neighborhood
- Foursquare Developer Access to Venue and Categories Data (https://developer.foursquare.com)

METHODOLOGY

The methodology will include:

1. Data retrieval, exploration and wrangling
2. Performing K-means clustering algorithm to segment neighborhood
3. Visualizing neighborhood clusters using boxplot
4. Labelling the resulting clusters

Firstly, we will need to get the list of all neighborhoods in the city of Jaipur. Fortunately, the list is available on the Maps of India website (https://www.mapsofindia.com/pincode/india/rajasthan/jaipur). Web scrapping will be done through Python requests and BeautifulSoup package to extract the area list. We will get the following table with 562 rows:

	Location	Pin Code
0	A C jobner	303329
1	A.G.office	302005
2	Ashoknagar	302001
3	Achalpura	303908
4	Achrol	303002

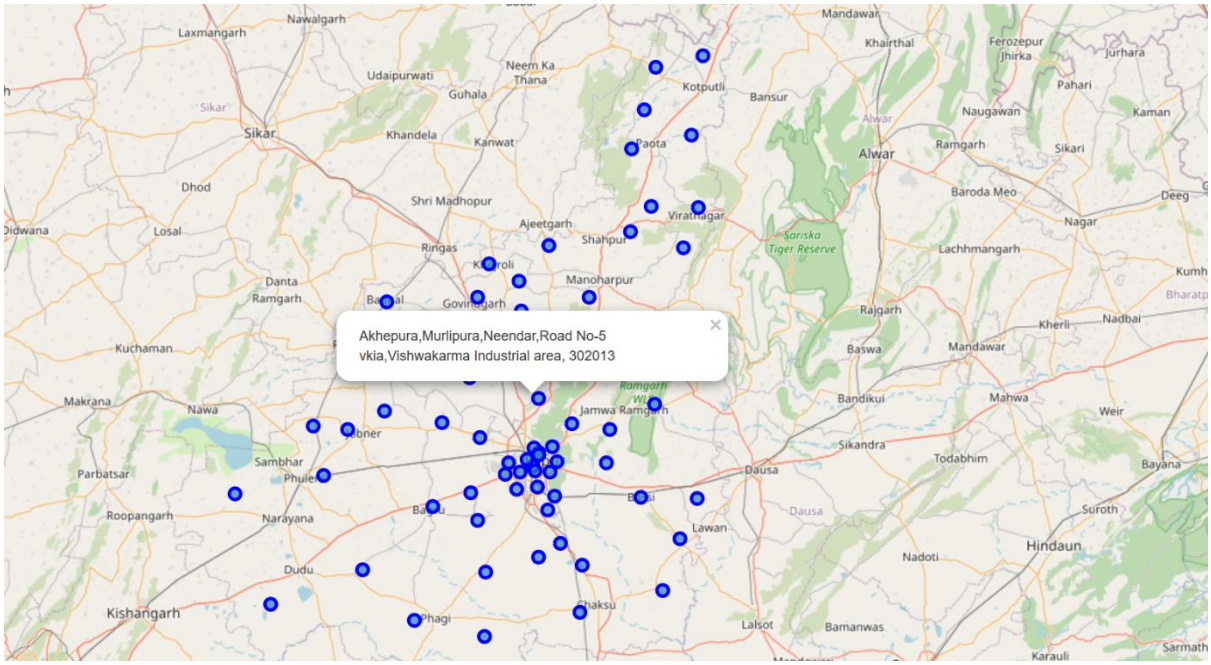
This list is an elaborative list and contains the neighborhoods under same Pin Code as different row. However, we require the list with unique pincodes with their associated neighborhoods clubbed by delimiter. Following list is generated after doing the above changes with 77 rows:

	Pin Code	Location
0	302001	Ashoknagar,Bhura Tiba,C.P.m.g,Chandpole Bazar,...
1	302002	Amer Road,Govind Nagar,Kishanpole Bazar,Shivaj...
2	302003	Haldiyon Ka rasta,Jaipur City,Krishi Upaj mand...
3	302004	Arjun Lal sethi nagar,Janta Colony,Jawahar Nag...
4	302005	A.G.office,High Court,Legislative Assembly,N.C...

Although we have retrieved the Neighborhood data, it is necessary to fetch the geographical coordinates of each Neighborhood to use FourSquare API. To do so, we will use the Geocoder package which will allow us to convert address into geographical coordinates in the form of latitude and latitude. Once retrieving this data, our dataframe looks like the following:

	Pin Code	Location	Latitude	Longitude
0	302001	Ashoknagar,Bhura Tiba,C.P.m.g,Chandpole Bazar,...	26.930038	75.805092
1	302002	Amer Road,Govind Nagar,Kishanpole Bazar,Shivaj...	26.942591	75.838600
2	302003	Haldiyon Ka rasta,Jaipur City,Krishi Upaj mand...	26.910635	75.849662
3	302004	Arjun Lal sethi nagar,Janta Colony,Jawahar Nag...	26.887956	75.831860
4	302005	A.G.office,High Court,Legislative Assembly,N.C...	26.899511	75.791935

After generating the above dataframe, we will use the folium maps to check the sanity of geographical coordinates data returned by Geocoder package plotted in the city of Jaipur. The following map is generated:



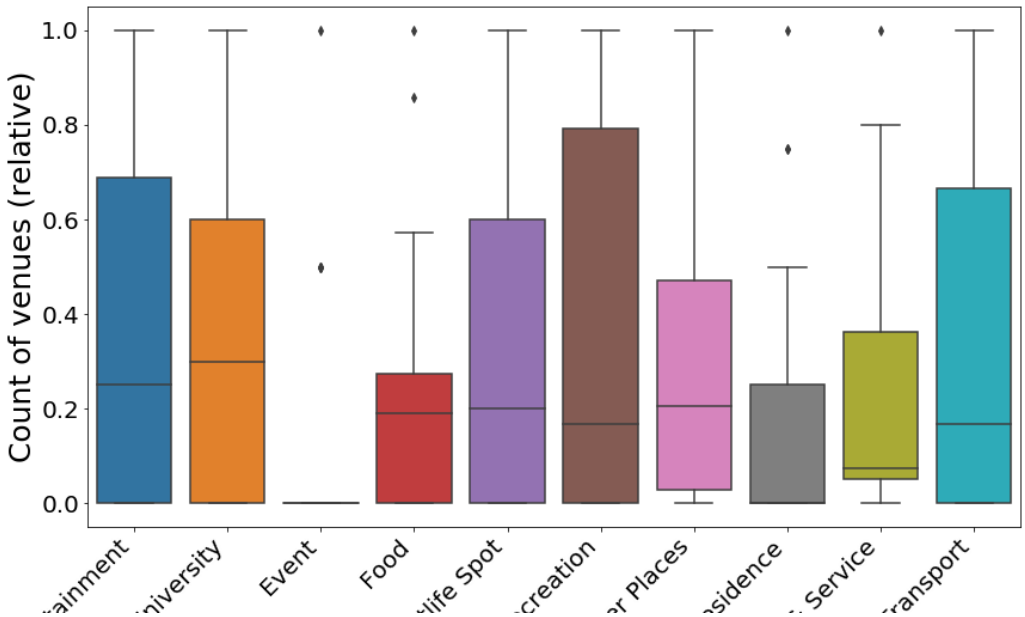
Analyzing this map, we realize that this data covers whole Jaipur District. But we want to do analysis for only the main Jaipur City. After careful examination of the graph and based on our prior knowledge of the city's geography, we have identified that the Pin Codes starting with '303***', and thereafter, are outside the main city and therefore our analysis. Thus we will drop the outside city areas from our study and following dataframe with 26 rows will be generated:

	Pin Code	Location	Latitude	Longitude
0	302001	Ashoknagar,Bhura Tiba,C.P.m.g,Chandpole Bazar,...	26.930038	75.805092
1	302002	Amer Road,Govind Nagar,Kishanpole Bazar,Shivaj...	26.942591	75.838600
2	302003	Haldiyon Ka rasta,Jaipur City,Krishi Upaj mand...	26.910635	75.849662
3	302004	Arjun Lal sethi nagar,Janta Colony,Jawahar Nag...	26.887956	75.831860
4	302005	A.G.office,High Court,Legislative Assembly,N.C...	26.899511	75.791935
5	302006	Ajmer Road,Industrial Estate,Jaipur R.s.,Khati...	26.914695	75.777270
6	302011	Airport Sanganer	26.925730	75.806590
7	302012	Bindayaka,Dhankya,Jhotwara,Jhotwara Industrial...	26.960270	75.665821
8	302013	Akhepura,Murlipura,Neendar,Road No-5 vkia,Vish...	27.044835	75.805821
9	302015	Bajaj Nagar,Bapu Nagar,Barkat Nagar,Gandhi Nag...	26.890574	75.797725
10	302016	Bani Park,Collectorate,Dwarikapuri,Shastri Nag...	26.939149	75.793650
11	302017	Malviya Industrial area,Malviya Nagar,Mrec Jai...	26.837535	75.843069
12	302018	Amer Clark hotel,Durgapura	26.855660	75.803751
13	302019	Shyam Nagar	26.887180	75.761123

Next, the Foursquare API is used to find the top 10 categories, a representation of development in each neighborhood. For each category under each neighborhood, the number of venues are found which is the basis of our study to define development in the following neighborhood. The resulting dataframe looks like following:

	Pin Code	Location	Latitude	Longitude	Arts & Entertainment	College & University	Event	Food	Nightlife Spot	Outdoors & Recreation	Professional & Other Places
0	302001	Ashoknagar,Bhura Tiba,C.P.m.g,Chandpole Bazar,...	26.930038	75.805092	2	1	0	6	3	5	4
1	302002	Amer Road,Govind Nagar,Kishanpole Bazar,Shivaj...	26.942591	75.838600	1	0	0	11	0	1	8
2	302003	Haldiyon Ka rasta,Jaipur City,Krishi Upaj mand...	26.910635	75.849662	1	0	0	0	0	1	3
3	302004	Arjun Lal sethi nagar,Janta Colony,Jawahar Nag...	26.887956	75.831860	2	5	2	6	5	6	6
4	302005	A.G.office,High Court,Legislative Assembly,N.C...	26.899511	75.791935	4	1	0	21	4	3	9

Before doing further exploration, the venue data is normalized and the resulting dataframe is visualized using the following BoxPlot:



Since the event data has not much venues, we have dropped this from our analysis.

Lastly, we perform clustering on the venues dataframe by using K-means clustering. K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the

nearest cluster, while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms and is suitable for solving our business problem by forming clusters of neighborhood. The neighborhood are divided into 3 clusters based on the number of venues in each category in each neighborhood. The results allow us to identify the areas which are developed and underdeveloped. Based on the clusters generated based on venues, we are able to identify the neighborhood holding investment opportunity for property developers.

RESULTS

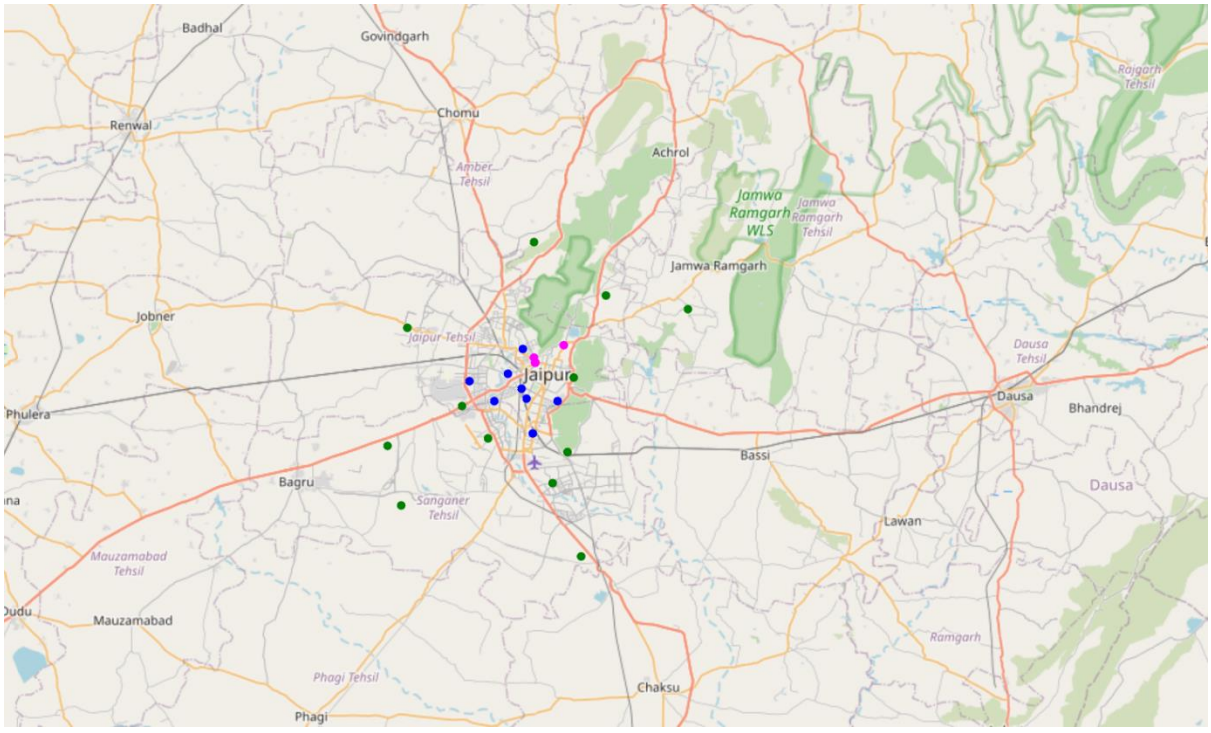
The result from the K-mean clustering show that we can categorize the neighborhoods into 3 clusters based on the number of venues in Top 10 categories in each neighborhood:

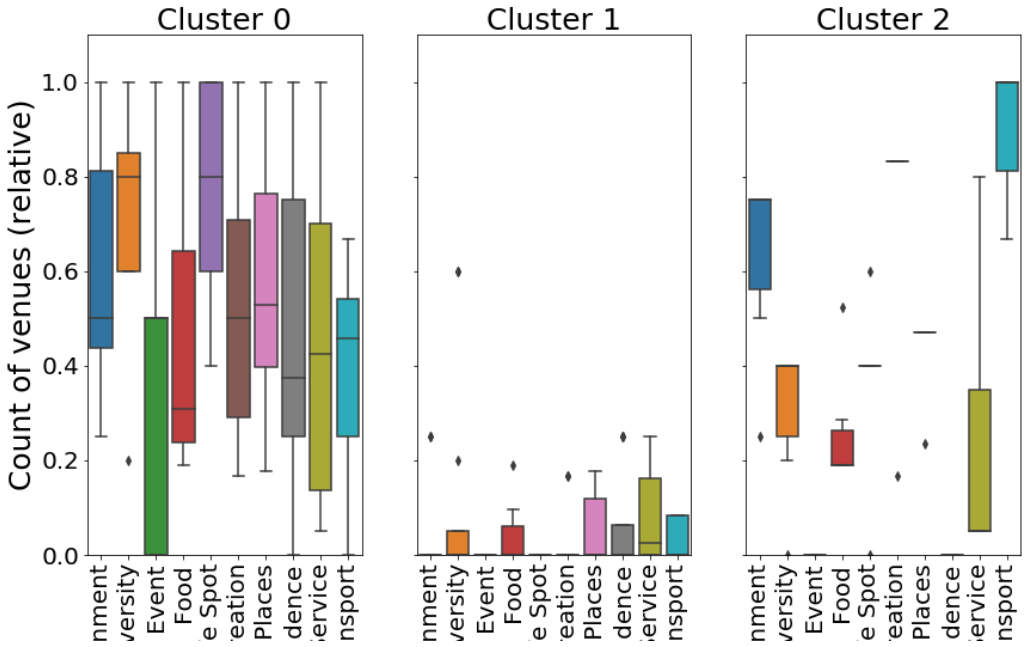
Cluster 0: 8 ‘Most Developed’ Neighborhoods

Cluster 1: 12 ‘Under Developed’ Neighborhoods

Cluster 2: 6 ‘Moderately Developed’ Neighborhoods

The results of the clustering algorithms are visualized using the map and box plot below. The map is showing cluster 0 in blue colour, cluster 1 in green colour, and cluster 2 in magenta colour.





DISCUSSION

After observing the results from clustering algorithm, it is clear that cluster 1 is the area which is least developed followed by cluster 2, and lastly cluster 0 which is considerably most developed. Therefore, this project recommends property developers and investors to buy and develop property in Cluster 1 Neighborhoods. Moreover, Cluster 1 could also be the potential investment destination. But, they should avoid Neighborhoods in Cluster 1.

Limitations and Suggestions for future Research

In this project, we have considered only one factor i.e. number of venues in Top 10 categories in each categories. However, other factors could also influence the identification of under developed area like population concentration. Future research could find such data and utilize it in forming clusters using K-means.

CONCLUSION

This study encompasses identify the business problem, specifying the data required, extraction and preparation of data, running machine learning clustering algorithm and lastly providing recommendation to the relevant stakeholders such as property developers and investors to identify the underdeveloped Neighborhoods. The Neighborhoods in Cluster 1 are the least developed and thus the most preferred for investment. The findings of this study will help developers in locating profitable areas for investment.

REFERENCES

Pin Code of Jaipur in Rajasthan. Maps of India. Retrieved from

https://www.mapsofindia.com/pincode/india/rajasthan/jaipur

Foursquare Developers Documentation. Foursquare. Retrieved from

https://developer.foursquare.com/docs

Business Standard (2019, December 12) Surat, Jaipur to emerge as new metros. Business Standard. Retrieved from

https://www.business-standard.com/article/economy-policy/surat-jaipur-to-emerge-as-new-metros-by-2018-ey-117030901171_1.html