LEVERAGING MACHINE LEARNING AND SOCIOECONOMIC INDICATORS TO RECOMMEND A BEST LOCALITY FOR RESTAURANT IN INDORE

A PROJECT REPORT

in partial fulfillment for the award of the degree of

MASTER OF TECHNOLOGY

in

DATA SCIENCE

Submitted by

TARUNA SAHU

DS7A-2117

EXTERNAL SUPERVISOR	**INTERNAL SUPERVISOR**
NAME:	**NAME:**
Designation:	Designation:

SCHOOL OF DATA SCIENCE AND FORECASTING

(UNIVERSITY TEACHING DEPARTMENT)

DEVI AHILYA VISHWAVIDYALAYA

Indore (M.P.)

April, 2023

SCHOOL OF DATA SCIENCE AND FORECASTING

DEVI AHILYA VISHWAVIDYALAYA INDORE (M.P.)

STATEMENT OF ORIGINALITY

In accordance with the requirements for the Degree of Master of Technology in DATA SCIENCE, in SCHOOL OF DATA SCIENCE AND FORECASTING, I present this report entitled LEVERAGING MACHINE LEARNING AND SOCIOECONOMIC INDICATORS TO RECOMMEND A BEST LOCALITY FOR RESTAURANT IN INDORE. This report is completed under the Supervision of:

NAME OF EXTERNAL SUPERVISOR

Designation & Affiliation:

NAME OF INTERNAL SUPERVISOR

Designation & Affiliation:

I declare that the work presented in the report is my own work except as acknowledged in the text and footnotes, and that to my knowledge this material has not been submitted either in whole or in part, for any other degree at this University or at any other such Institution.

(……)

Name and Signature of the Student

Date:

SCHOOL OF DATA SCIENCE AND FORECASTING

DEVI AHILYA VISHWAVIDYALAYA INDORE (M.P.)

RECOMMENDATION

This dissertation entitled LEVERAGING MACHINE LEARNING AND SOCIOECONOMIC INDICATORS TO RECOMMEND A BEST LOCALITY FOR RESTAURANT IN INDORE submitted by TARUNA SAHU towards the partial fulfillment of Degree of Master of Technology in Data Science of Devi Ahilya Vishwavidyalaya, Indore is a satisfactory account of his project work and is recommended for the award of degree.

External Supervisor

Internal Supervisor **Head of Department**

SCHOOL OF DATA SCIENCE AND FORECASTING

DEVI AHILYA VISHWAVIDYALAYA

INDORE (M.P.)

CERTIFICATE

This is to certify that the dissertation entitled LEVERAGING MACHINE LEARNING AND SOCIOECONOMIC INDICATORS TO RECOMMEND A BEST LOCALITY RESTAURANT IN INDORE submitted by TARUNA SAHU is approved for the award of Master of Technology in Data Science.

INTERNAL EXAMINER **EXTERNAL EXAMINER**

DATE:

ABSTRACT

The craze for food is expanding each day. People prefer to celebrate any occasion of their life in a restaurant, which results in growing demand for a good restaurant with the best amenities. Every year 60% of restaurants fail in their first year.The Aim of this Project is to use a data-driven approach to provide the best location for a restaurant in the city. This study uses various factors which are included for the recommendation of the location such as competition, area, etc. This study uses a combination of proprietary datasets and data sets available publicly to develop a comprehensive analysis framework. The analysis involves exploring the relationships between different variables and identifying the most significant factors that influence the success of an Indian restaurant.

The result can help businessmans, entrepreneurs and restaurants to make an informed decision about opening a restaurant in the locality. This analytic approach can help the restaurant business to gain more profitability and success by providing valuable information

ACKNOWLEDGEMENT

I would like to express my deepest gratitude to Head of the Department Prof. (Dr.) V.B. Gupta, School of Data Science and Forecasting DAVV Indore for giving me a chance to work on this project.

I would also like to thank Dr. Vandit Hedau for his continuous encouragement for betterment of the project.

It is their help and support, due to which we became able to complete the project and technical report. Without their support this work would not have been possible.

Name of the Student with Signature

Taruna Sahu

TABLE OF CONTENTS

Chapter Name			Page no.
	Abstract		**5**
	Acknowledgment		**6**
	List of Symbol		**8**
	List of Figure		**9**
1.	**Introduction**		**10**
	1.1.	Problem Statement	10
	1.2.	Objective	10
	1.3.	Approach	10
	1.4.	Organization Report	10
2.	**Literature review**		**11**
	2.1.	Python	11
	2.2.	Pandas	11
	2.3.	Geopy Geocoder	11
	2.4.	Foursquare API	12
	2.5.	Folium	12
	2.6.	Machine Learning	12
	2.6.1.	Unsupervised Machine Learning	12
	2.6.2.	K- means Clustering	13
3.	**Methodology**		**14**
	3.1.	Data Collection	14
	3.2.	Data Preprocessing	18
	3.3.	Data Loading	19
	3.4.	Model	20
	3.5.	Analysis	22
4.	**Result**		**28**
5.	**Conclusion**		**29**
6.	**Future Scope**		**30**
7.	**References**		**31**

LIST OF SYMBOLS, ABBREVIATIONS AND NOMENCLATURE

1. ML: Machine Learning
2. csv: Comma Separated Values
3. AI: Artificial intelligence
4. UML: Unsupervised Machine Learning
5. km : Kilo Meter

LIST OF FIGURES

1. K- Means Clustering
2. Localities of Indore
3. Selling Price per Square Feet
4. Population and Area of the Localities
5. Indore Localities and their Coordinates
6. Indore Map
7. Venue and Venue Category
8. One Hot Encoding based on Venue Category
9. One Hot Encoding based on Restaurant Types
10. Clustering based on Venue Category
11. Clustering based on Restaurant Types
12. Clustering based on Population Area
13. Types of Restaurant in Indore

1. INTRODUCTION

1.1 The Problem Statement:

The main reason for failure of a restaurant is starting the restaurant without prior analyzing the factors that can affect the restaurant success. According to the National Restaurant Association 60% of restaurants shut down within the first year of operations, and up to 80% of restaurants close their operations in the first five years. The major reason is poor location and high rentals i.e. either you are located far from the target audience or the location you prefer drains out your pocket.

1.2 Objective:

The objective of this project is to suggest an optimal locality for opening a restaurant to evaluate affecting factors like demographics, competition, traffic and accessibility.

1.3 Approach:

This study uses Unsupervised Machine Learning techniques to create a group of famous localities of indore based on population, area, and venue.

1.4 Organization of report:

- chapter 2: theoretical studies about the technology used.
- chapter 3: data collection, preprocessing and modeling
- chapter 4: Result and Analysis
- chapter 5: Conclusion
- Chapter 6: Reference

2. Literature Review

2.1 Python:

Python is an interpreted programming language . It was first introduced in 1991 by Guido van Rossum. Python has been used for various applications such as AI, Web development, Data Analysis, Scientific Computing. Python is a very easy language, syntax of this language easy to read and learn. Python can be used on various platforms like Windows, Mac, Linux etc. Python focuses on code readability. It can perform automatic memory management.

2.2 Pandas:

Pandas is a Python Library developed by Wes Mckinney in 2008. Pandas used to help handle and analyze large and complex datasets. Pandas has functions to clean and explore datasets. Pandas can delete rows, columns from datasets.

2.3 Geopy Geocoder:

Geopy is a Python library for providing access to geocoding services like Bingmap, google maps etc. It helps to change physical addresses into geographic coordinates i.e. latitude and longitude. Geopy also provides calculation of distance between two points, converting geographic coordinates into address, and geotagging

2.4 Foursquare API:

Foursquare API is used to get location based data and services. This API also gives information about the venues, users and tips. It gives real time updates. This API allows developers to create location based applications,

recommendations and travel and tourism apps . This API also Includes authentication mechanisms to secure user data.

2.5 Folium:

Folium is a python library used to visualize maps. Folium has a wide range of functions to create maps with labels, zoom controls. Folium also allows developers to mark, pop-ups , and custom icons to add in maps. It supports a wide range of data formats like Geojson , CSV, Pandas dataframe .

2.6 Machine Learning:

Machine Learning is a subset of AI that enables computers to self learn without being explicitly programmed. Machine learning learns from past data and identifies patterns, relationships and makes predictions.

There are three types in ML: Supervised learning, Unsupervised learning and Reinforcement learning.

2.6.1 Unsupervised ML:

Unsupervised learning is used to analyze and group the unlabeled datasets. It is used to find hidden relationships and patterns between the unlabeled datasets. Clustering, Anomaly detection , and Dimension reduction are the types of unsupervised ML.

Clustering: Clustering algorithm is used to group data points based on their characteristics or attributes.

Anomaly detection: this algorithm used to identify data points which are significantly different from the rest of datasets.

Dimensionality reduction: This algorithm is used to reduce the number of variables in a dataset.

2.6.2 K-Means Clustering:

K-means Clustering Algorithm is an UML, which groups the unlabeled datasets into different clusters. K is used to define parameters that represent clusters to be created. K means algorithm works by dividing the dataset into k clusters , algorithm assigning each data point into clusters based on distance such as Euclidean distance. This algorithm is easy to implement and compute. This is commonly used in customer segmentation, image segmentation, anomaly detection.

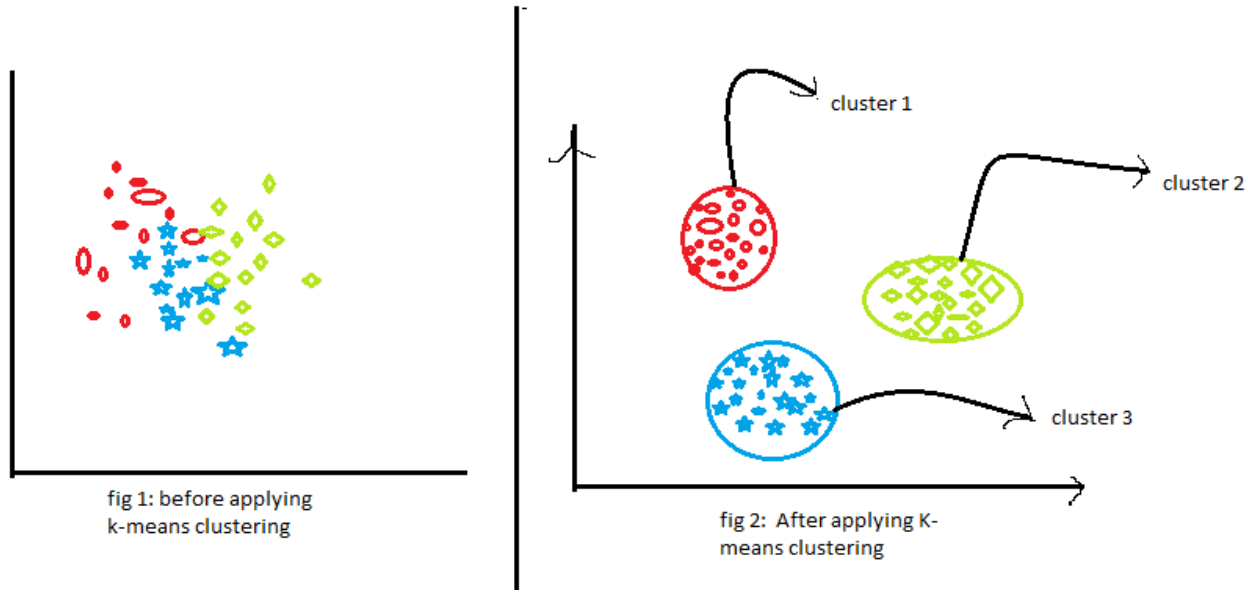


Fig 1: K-Means Clustering

3. METHODOLOGY

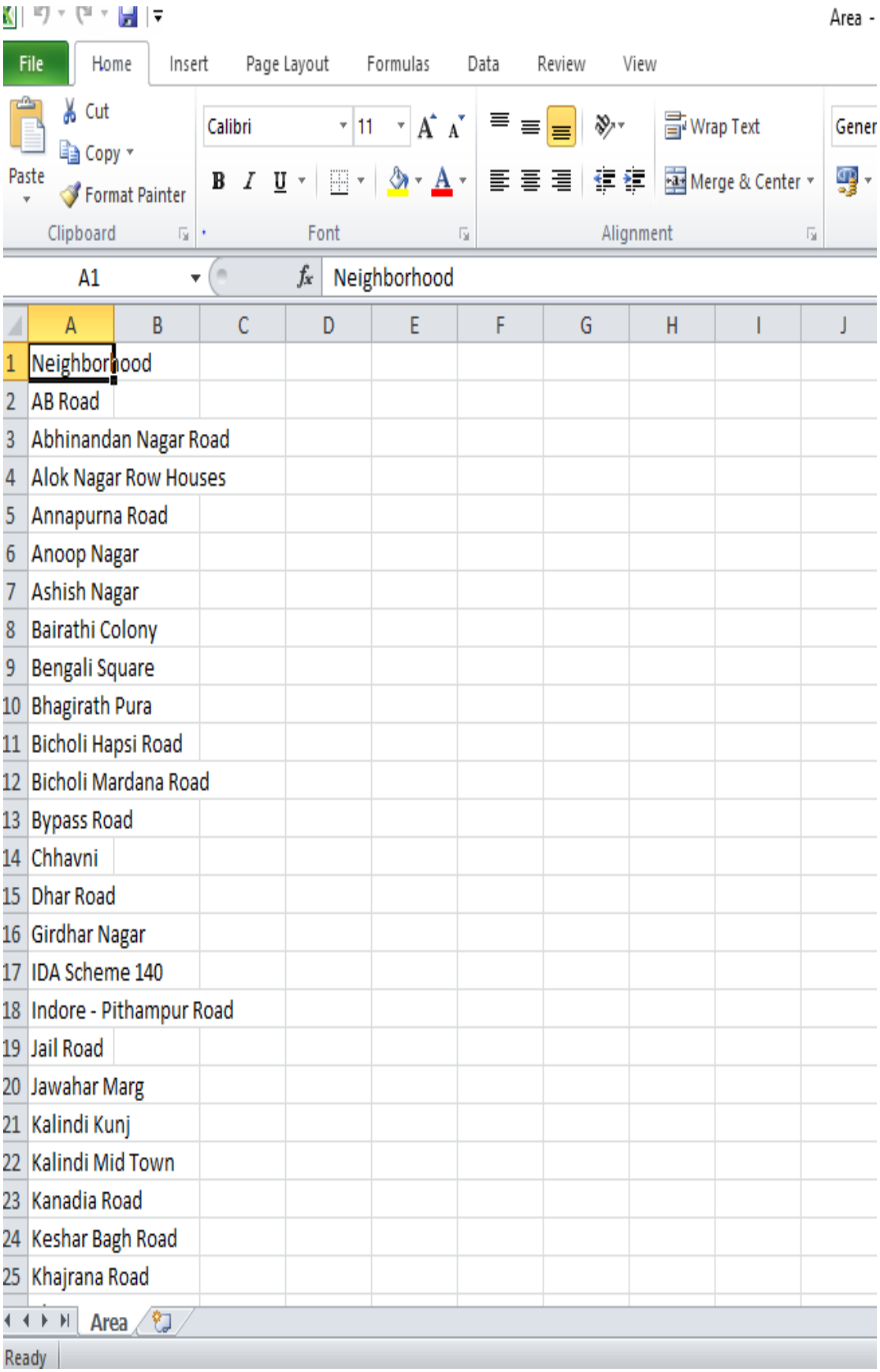
This Section includes all the methods and techniques that are used in this project. The various steps to get an optimal result are as follows:

- Data Collection
- Data Preprocessing
- Data Loading
- Model
- Result

The Project is based on Unsupervised Machine Learning algorithm i.e. K means algorithm. In this Project i used k mean algorithm to cluster famous localities of Indore based on their Population, Venues and Restaurant category.

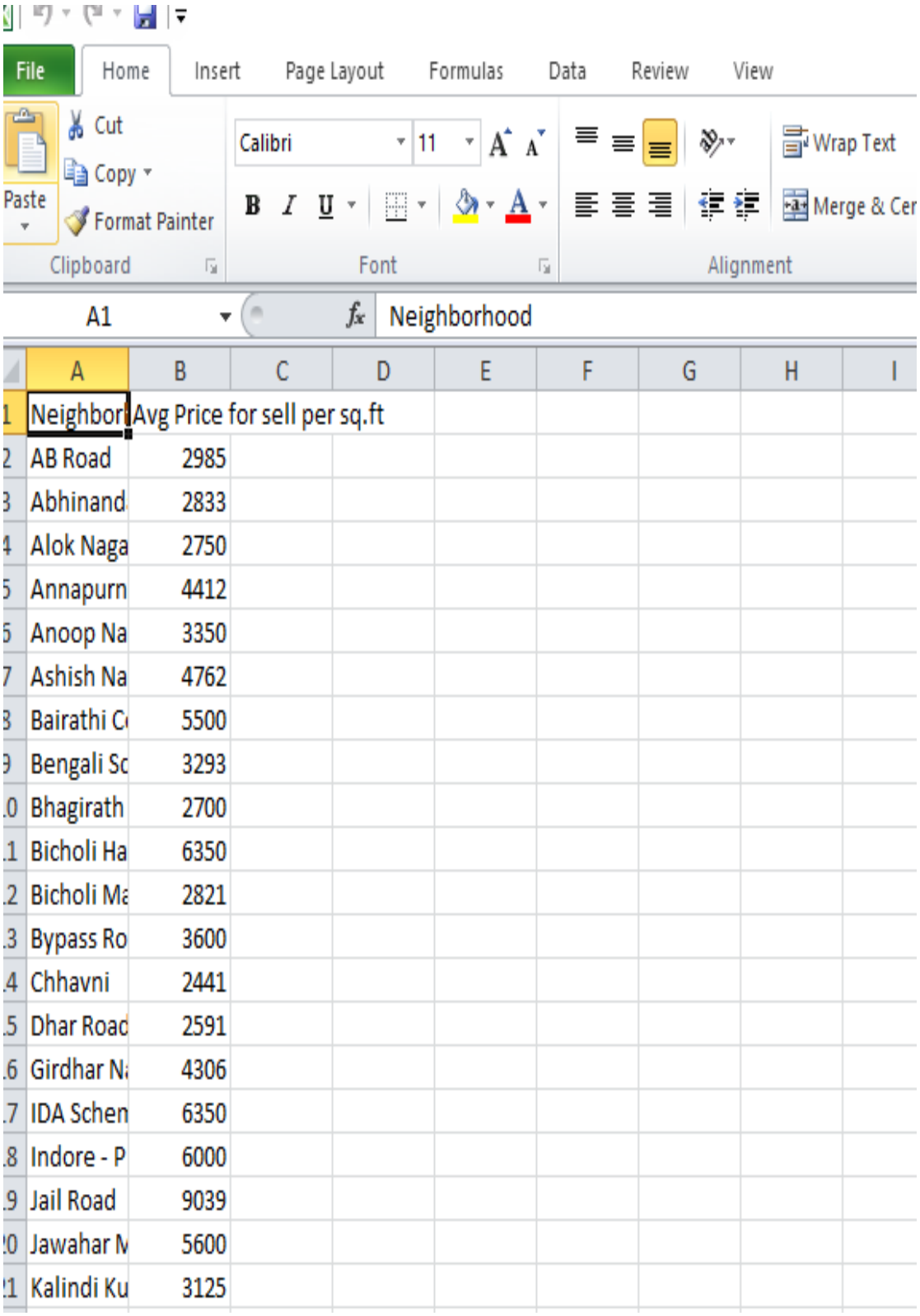
3.1 Data Collection:

The first step of the Project is to collect the data. Here I used web scraping to collect the locality of the indore. The name of the localities is stored in the file name as “Area.csv” . The area wise Population , selling price per square feet and area in km also collect from the web scraping and store in filename “Area_pop.csv” and “Area_range.csv” in csv format.



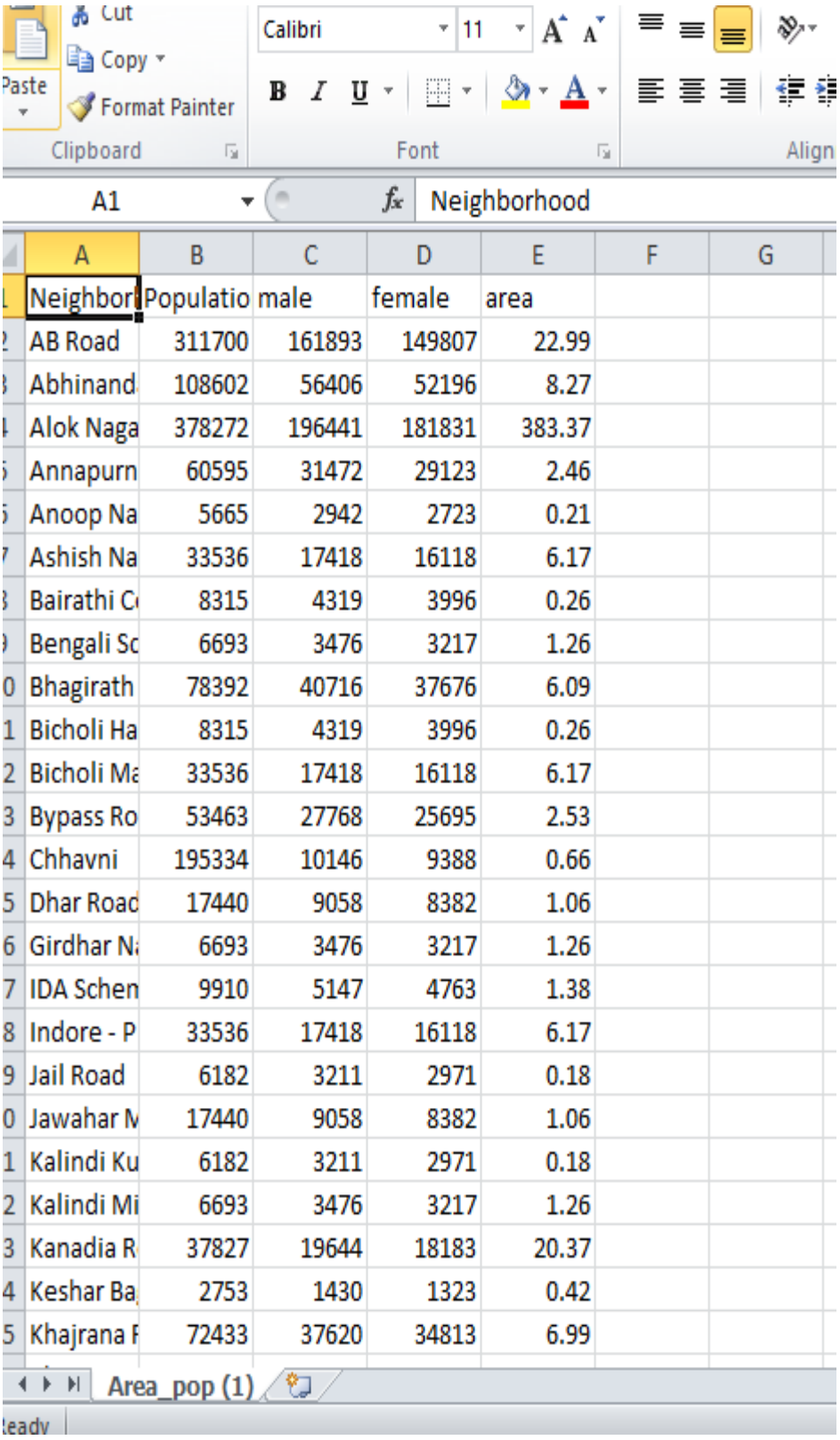
	A
1	Neighborhood
2	AB Road
3	Abhinandan Nagar Road
4	Alok Nagar Row Houses
5	Annapurna Road
6	Anoop Nagar
7	Ashish Nagar
8	Bairathi Colony
9	Bengali Square
10	Bhagirath Pura
11	Bicholi Hapsi Road
12	Bicholi Mardana Road
13	Bypass Road
14	Chhavni
15	Dhar Road
16	Girdhar Nagar
17	IDA Scheme 140
18	Indore - Pithampur Road
19	Jail Road
20	Jawahar Marg
21	Kalindi Kunj
22	Kalindi Mid Town
23	Kanadia Road
24	Keshar Bagh Road
25	Khajrana Road

Fig2: Localities of Indore



	A	B	C	D	E	F	G	H	I
1	Neighborh	Avg Price for sell per sq.ft							
2	AB Road	2985							
3	Abhinand	2833							
4	Alok Naga	2750							
5	Annapurn	4412							
6	Anoop Na	3350							
7	Ashish Na	4762							
8	Bairathi C	5500							
9	Bengali Sc	3293							
10	Bhagirath	2700							
11	Bicholi Ha	6350							
12	Bicholi Ma	2821							
13	Bypass Ro	3600							
14	Chhavni	2441							
15	Dhar Road	2591							
16	Girdhar Na	4306							
17	IDA Schen	6350							
18	Indore - P	6000							
19	Jail Road	9039							
20	Jawahar M	5600							
21	Kalindi Ku	3125							

Fig 3: Selling Price per square feet



	A	B	C	D	E
1	Neighborh	Populatio	male	female	area
2	AB Road	311700	161893	149807	22.99
3	Abhinand	108602	56406	52196	8.27
4	Alok Naga	378272	196441	181831	383.37
5	Annapurn	60595	31472	29123	2.46
6	Anoop Na	5665	2942	2723	0.21
7	Ashish Na	33536	17418	16118	6.17
8	Bairathi C	8315	4319	3996	0.26
9	Bengali Sc	6693	3476	3217	1.26
10	Bhagirath	78392	40716	37676	6.09
11	Bicholi Ha	8315	4319	3996	0.26
12	Bicholi Ma	33536	17418	16118	6.17
13	Bypass Ro	53463	27768	25695	2.53
14	Chhavni	195334	10146	9388	0.66
15	Dhar Road	17440	9058	8382	1.06
16	Girdhar Na	6693	3476	3217	1.26
17	IDA Schen	9910	5147	4763	1.38
18	Indore - P	33536	17418	16118	6.17
19	Jail Road	6182	3211	2971	0.18
20	Jawahar M	17440	9058	8382	1.06
21	Kalindi Ku	6182	3211	2971	0.18
22	Kalindi Mi	6693	3476	3217	1.26
23	Kanadia R	37827	19644	18183	20.37
24	Keshar Ba	2753	1430	1323	0.42
25	Khajrana F	72433	37620	34813	6.99

Fig 4: Population and Area of the localities

3.2 Data Preprocessing:

This section includes data preprocessing , so for the preparation first I read the file “Area.csv” into a dataframe then I used geopy.geocoder python library to get Latitude and Longitude of the localities. The latitude and Longitude add to the data frame.

```
Code  + Text
df.head()
]
```

	Neighborhood	latitude	longitude
0	AB Road	22.766544	75.900694
1	Abhinandan Nagar Road	22.761933	75.864985
2	Alok Nagar Row Houses	22.735187	75.888129
3	Annapurna Road	22.694320	75.838780
4	Anoop Nagar	22.731192	75.893259

Fig 5: Indore locality and their coordinates

After getting the latitude and longitude i created a map to show all these localities, in that folium library is used.

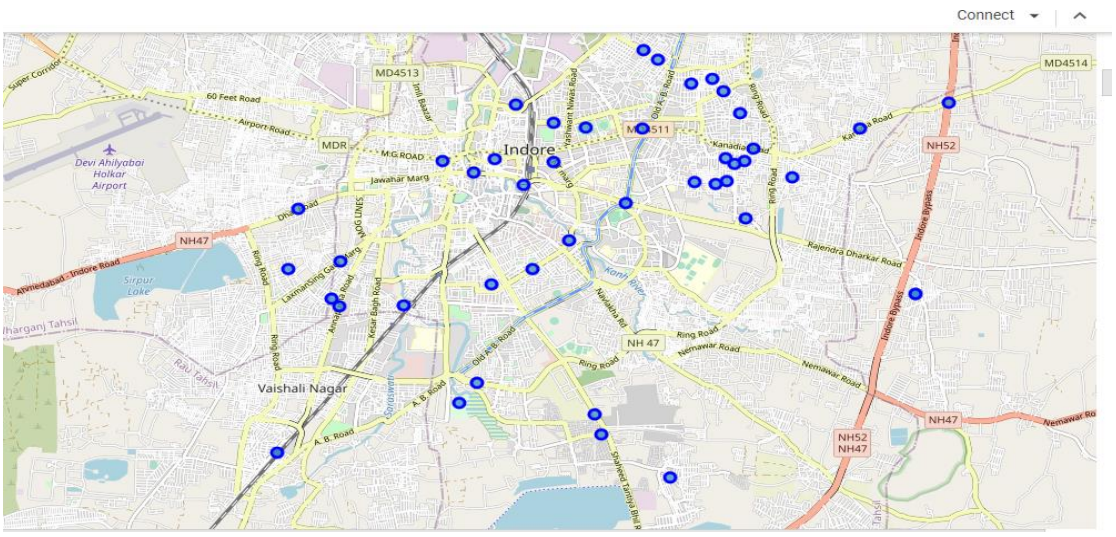


Fig 6: Indore Map

The Foursquare API is used to get nearby venues of the given locality.

so with the foursquare api i have created another dataframe called venue.

“venue” dataframe includes columns neighborhood, venue name, venue category and venue latitude and longitude.

```
# DataFrame by nearby Venues
print(venues.shape)
venues.head()
```

```
(177, 7)
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	AB Road	22.766544	75.900694	Lemosys Infotech	22.764291	75.901827	IT Services
1	AB Road	22.766544	75.900694	kbc	22.762606	75.898903	Bakery
2	Alok Nagar Row Houses	22.735187	75.888129	Shoppers Stop Indore	22.736239	75.891010	Shopping Mall
3	Alok Nagar Row Houses	22.735187	75.888129	Zodiac Retail Shop	22.733242	75.888941	Men's Store
4	Alok Nagar Row Houses	22.735187	75.888129	Hotel Shreemaya Residency	22.735864	75.890749	Hotel

Fig 7: Venues and venue category

3.3 Data Loading:

For loading the data into the model, Venues are grouped by neighborhood. after grouping the data one hot encoding done based on the venue category.

```
onehot.head()
```

	Neighborhood	ATM	Accessories Store	Airport	Art Gallery	Asian Restaurant	BBQ Joint	Bakery	Bar	Bus Station	...	Snack Place	South Indian Restaurant	Sporting Goods Shop	Stadium	Street Art	Tea Room	Train Station	Wing Join
0	AB Road	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	
1	AB Road	0	0	0	0	0	0	1	0	0	...	0	0	0	0	0	0	0	
2	Alok Nagar Row Houses	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	
3	Alok Nagar Row Houses	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	
4	Alok Nagar Row Houses	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	

5 rows × 59 columns

Fig 8: One hot Encoding based on Venue category

To check the competition I also analyze the types of restaurant available at the locality. and do one hot encoding based on the types of restaurant.

```
indore_onehot.head()
```

	Neighborhood	Asian Restaurant	Dumpling Restaurant	Fast Food Restaurant	Indian Restaurant	Restaurant	South Indian Restaurant
1	Alok Nagar Row Houses	1	0	0	0	0	0
2	Anoop Nagar	0	0	1	0	0	0
3	Ashish Nagar	0	0	0	0	1	0
4	Bengali Square	0	0	0	1	0	0
5	Bengali Square	0	0	0	0	1	0

Fig 9: One hot Encoding based on Types of Restaurant

As you can see there are six types of restaurant in indore.

3.4 Model:

The data is now ready to put in the model. For Analyzing the optimal location K mean clustering model is used. Three K mean models are created for the recommendation of the best location.

3.4.1 Clustering based on venue category:

The first K mean model fitted with the data where the one hot encoding was done on the venue category. Here the value of k is 5 i.e. 5 groups are made depending on venue category. The clusters are shown on the map with different colors.

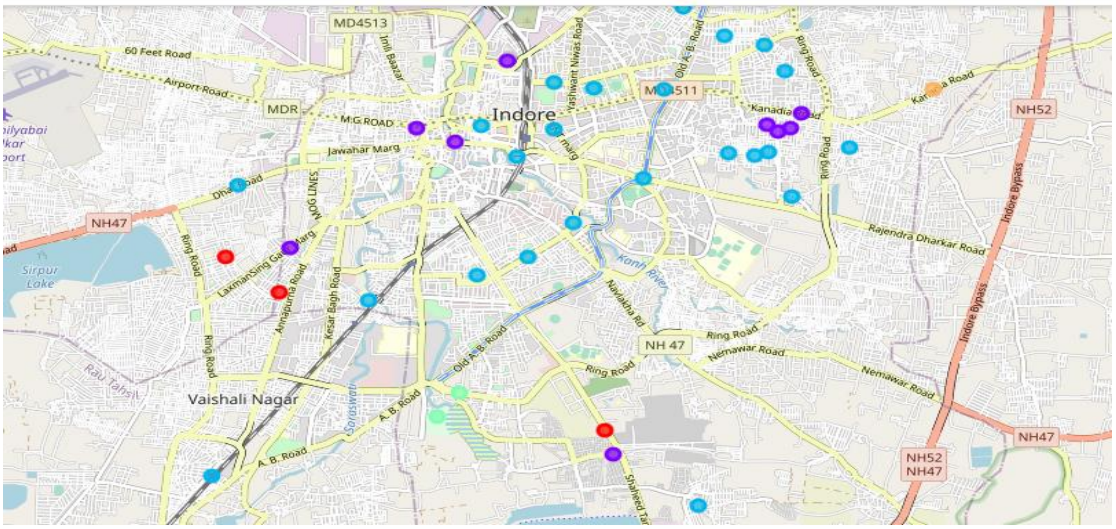


Fig 10: Clustering based on venue category

3.4.2 Clustering based on Restaurant Types :

The first K mean model fitted with the data where the one hot encoding was done on the restaurant types. Here the value of k is 8 i.e. 8 groups are made depending on the restaurant category. The clusters are shown on the map with different colors.

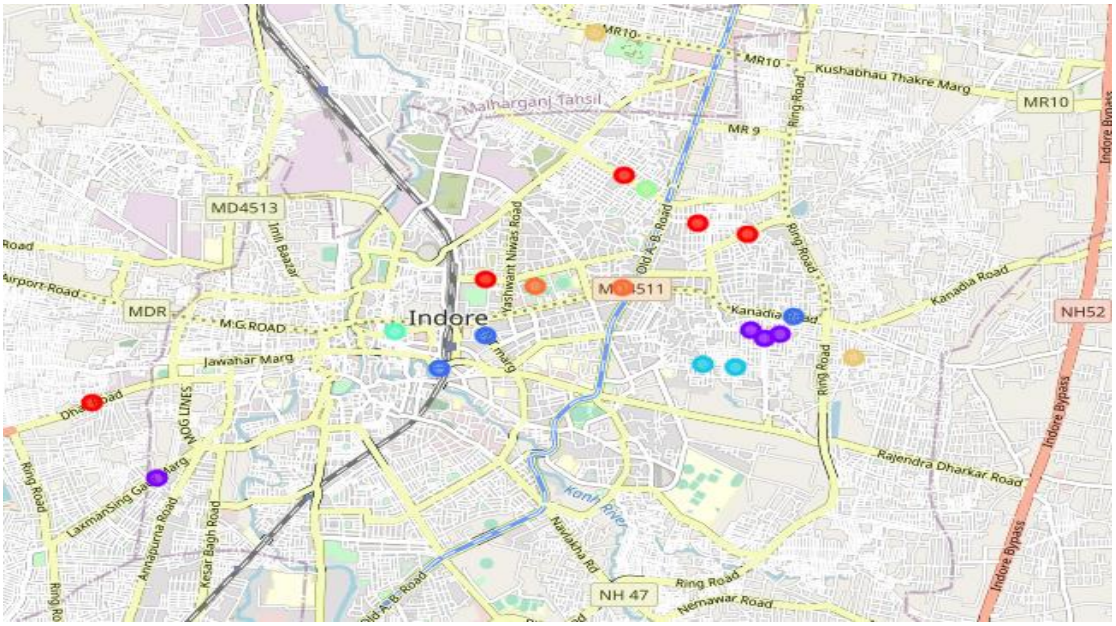


Fig11: Clustering based on Restaurant Type

3.4.3 Clustering based on Population and Area:

The first K mean model fitted with the data where the population and area merged on the neighborhood. Here the value of k is 5 i.e. 5 groups are made depending on the population. The clusters are shown on the map with different colors.

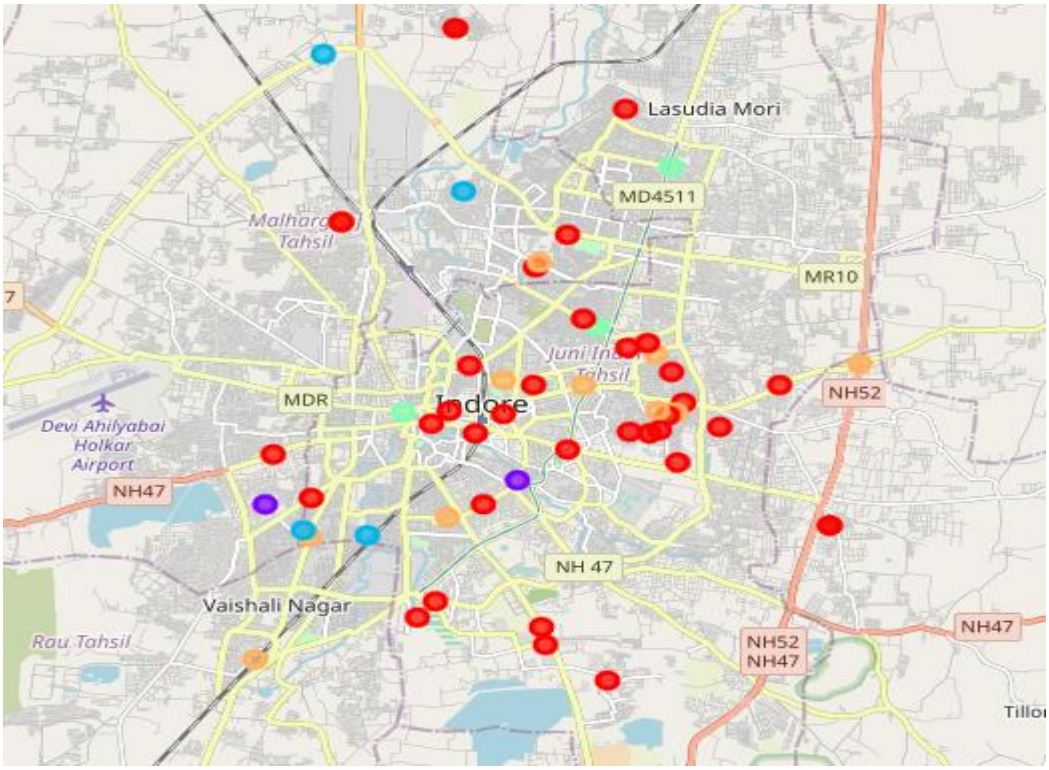


Fig 12: Clustering based on Population Area

3.5 Analysis :

In this project the three clustering models are created based on different criteria. so as the result we get three different clusters.

3.5.1 Clustering based on venue category:

There are a total of five clusters created. After analyzing each cluster we get :

- cluster 1 has the most common venue is an ATM.

```
[ ] #Cluster 1
    merged.loc[merged['Cluster Labels'] == 0,merged.columns[[0] + list(range(4, merged.shape[1]))]]
```

	Neighborhood	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
31	Shiv Shakti Nagar Road	ATM	Men's Store	Food & Drink Shop	Plaza	Lake	Market	Mattress Store	Motel	Movie Theater	Multiplex
34	South Tukoganj	ATM	Café	Pub	Lake	Market	Mattress Store	Men's Store	Motel	Movie Theater	Multiplex
37	Vijay Nagar	ATM	Plaza	Juice Bar	Lake	Market	Mattress Store	Men's Store	Motel	Movie Theater	Multiplex

```
[ ] merged.loc[merged['Cluster Labels'] == 0,merged.columns[[0] + list(range(4, merged.shape[1]))]]['1st Most Common Venue'].value_counts().head()

    ATM    3
    Name: 1st Most Common Venue, dtype: int64
```

- cluster 2 has the most common venue snack place and hotel.

35	Road	Snack Place	Coffee Shop	Restaurant	ATM	Plaza	Lake	Market	Mattress Store	Men's Store
38	Vishnu Puri Colony	Art Gallery	Snack Place	Indian Restaurant	ATM	Plaza	Lake	Market	Mattress Store	Men's Store

```
] merged.loc[merged['Cluster Labels'] == 1,merged.columns[[0] + list(range(4, merged.shape[1]))]]['1st Most Common Venue'].value_counts().head()

    Snack Place          3
    Hotel                2
    Coffee Shop          1
    Stadium              1
    Indian Restaurant    1
    Name: 1st Most Common Venue, dtype: int64
```

- cluster 3 has the most common venue hotel and indian restaurant.

```
[ ] merged.loc[merged['Cluster Labels'] == 2,merged.columns[[0] + list(range(4, merged.shape[1]))]]['1st Most Common Venue'].value_counts().head()

    Hotel               3
    Indian Restaurant   2
    Restaurant          2
    Bus Station         2
    Ice Cream Shop      2
    Name: 1st Most Common Venue, dtype: int64
```

- cluster 4 has the most common venue park and bar.

```
[ ] #Cluster 4
    merged.loc[merged['Cluster Labels'] == 3,merged.columns[[0] + list(range(4, merged.shape[1]))]]
```

	Neighborhood	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
25	Regional Park Road	Park	Bar	ATM	Pub	Lake	Market	Mattress Store	Men's Store	Motel	Movie Theater
26	Ring Road	Bar	ATM	Pub	Lake	Market	Mattress Store	Men's Store	Motel	Movie Theater	Multiplex

```
[ ] merged.loc[merged['Cluster Labels'] == 3,merged.columns[[0] + list(range(4, merged.shape[1]))]]['1st Most Common Venue'].value_counts().head()

    Park    1
    Bar     1
    Name: 1st Most Common Venue, dtype: int64
```

- cluster 5 has the most common venue as an ATM.

```
[ ] #Cluster 5
    merged.loc[merged['Cluster Labels'] == 4,merged.columns[[0] + list(range(4, merged.shape[1]))]]
```

	Neighborhood	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
10	Kanadia Road	Paper / Office Supplies Store	ATM	Plaza	Juice Bar	Lake	Market	Mattress Store	Men's Store	Motel	Movie Theater

After analyzing the cluster 1 is suitable for opening a restaurant because of the lack of restaurants.

3.5.2 Clustering based on Restaurant Types:

There are a total of eight clusters created. After analyzing each cluster we get :

- cluster 1 has the most common venue Fast food restaurant.

```
indore_merged.loc[indore_merged['Cluster Labels'] == 0,
                  indore_merged.columns[[0] + list(range(4, indore_merged.shape[1]))]]
```

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
4	Anoop Nagar	Fast Food Restaurant	Asian Restaurant	Dumpling Restaurant	Indian Restaurant	Restaurant
13	Dhar Road	Fast Food Restaurant	Asian Restaurant	Dumpling Restaurant	Indian Restaurant	Restaurant
23	Khajrana Road	Fast Food Restaurant	Asian Restaurant	Dumpling Restaurant	Indian Restaurant	Restaurant
36	Nai Duniya	Fast Food Restaurant	Asian Restaurant	Dumpling Restaurant	Indian Restaurant	Restaurant
46	Race Course Road	Fast Food Restaurant	Asian Restaurant	Dumpling Restaurant	Indian Restaurant	Restaurant

- cluster 2 has the most common venue Indian restaurant.

```
indore_merged.loc[indore_merged['Cluster Labels'] == 1,
                  indore_merged.columns[[0] + list(range(4, indore_merged.shape[1]))]]
```

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
47	Rajendra Nagar	Indian Restaurant	Asian Restaurant	Dumpling Restaurant	Fast Food Restaurant	Restaurant
49	Rani Sati Gate	Indian Restaurant	Asian Restaurant	Dumpling Restaurant	Fast Food Restaurant	Restaurant
53	Sainath Colony	Indian Restaurant	Asian Restaurant	Dumpling Restaurant	Fast Food Restaurant	Restaurant
61	Sneh Nagar	Indian Restaurant	Asian Restaurant	Dumpling Restaurant	Fast Food Restaurant	Restaurant
66	Tilak Nagar Main Road	Indian Restaurant	Asian Restaurant	Dumpling Restaurant	Fast Food Restaurant	Restaurant
71	Vishnu Puri Colony	Indian Restaurant	Asian Restaurant	Dumpling Restaurant	Fast Food Restaurant	Restaurant

- cluster 3 has the most common venue Indian restaurant.

```
indore_merged.loc[indore_merged['Cluster Labels'] == 2,
                  indore_merged.columns[[0] + list(range(4, indore_merged.shape[1]))]]
```

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
7	Bengali Square	Indian Restaurant	Restaurant	Asian Restaurant	Dumpling Restaurant	Fast Food Restaurant
52	RNT Marg	Indian Restaurant	Restaurant	Asian Restaurant	Dumpling Restaurant	Fast Food Restaurant
60	Siyaganj	Indian Restaurant	Restaurant	Asian Restaurant	Dumpling Restaurant	Fast Food Restaurant

- cluster 4 has the most common venue restaurant.

```
indore_merged.loc[indore_merged['Cluster Labels'] == 3,
                  indore_merged.columns[[0] + list(range(4, indore_merged.shape[1]))]]
```

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
4	Anoop Nagar	Fast Food Restaurant	Asian Restaurant	Falafel Restaurant	Indian Restaurant	Restaurant
13	Dhar Road	Fast Food Restaurant	Asian Restaurant	Falafel Restaurant	Indian Restaurant	Restaurant
23	Khajrana Road	Fast Food Restaurant	Asian Restaurant	Falafel Restaurant	Indian Restaurant	Restaurant
26	Khatiwala Tank	Fast Food Restaurant	Asian Restaurant	Falafel Restaurant	Indian Restaurant	Restaurant
46	Race Course	Fast Food Restaurant	Asian Restaurant	Falafel Restaurant	Indian Restaurant	Restaurant

- cluster 5 has the most common South Indian restaurant.

```
indore_merged.loc[indore_merged['Cluster Labels'] == 4,
                  indore_merged.columns[[0] + list(range(4, indore_merged.shape[1]))]]
```

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
35	MTH Compound	South Indian Restaurant	Asian Restaurant	Dumpling Restaurant	Fast Food Restaurant	Indian Restaurant

- cluster 6 has the most common venue as an Asian restaurant.

```
indore_merged.loc[indore_merged['Cluster Labels'] == 5,
                  indore_merged.columns[[0] + list(range(4, indore_merged.shape[1]))]]
```

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
2	Alok Nagar Row Houses	Asian Restaurant	Dumpling Restaurant	Fast Food Restaurant	Indian Restaurant	Restaurant

- cluster 7 has the most common venue restaurant.

```
indore_merged.loc[indore_merged['Cluster Labels'] == 6,
                  indore_merged.columns[[0] + list(range(4, indore_merged.shape[1]))]]
```

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
5	Ashish Nagar	Restaurant	Asian Restaurant	Dumpling Restaurant	Fast Food Restaurant	Indian Restaurant
57	Scheme No.54	Restaurant	Asian Restaurant	Dumpling Restaurant	Fast Food Restaurant	Indian Restaurant

- cluster 8 has the most common venue for Indian restaurant.

```
indore_merged.loc[indore_merged['Cluster Labels'] == 7,
                  indore_merged.columns[[0] + list(range(4, indore_merged.shape[1]))]]
```

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
38	New Palasia	Indian Restaurant	Fast Food Restaurant	Asian Restaurant	Dumpling Restaurant	Restaurant
73	Yeshwant Niwas Road	Indian Restaurant	Fast Food Restaurant	Asian Restaurant	Dumpling Restaurant	Restaurant

After analyzing cluster 1 it is suitable to open an Indian restaurant because of Fast food restaurants as common venues.

3.5.3 Clustering based on Population and Area :

There are a total of eight clusters created. After analyzing each cluster we get :

- cluster 1 has the most common venue hotel.

	Neighborhood	Population	area	Normalized_population	Normalized_area	Avg Price for sell per sq.ft	Normalized_range	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Mc Comm Ver
2	Anoop Nagar	5665	0.21	0.014976	0.000548	3350	0.370616	Hotel	Market	Café	Fast Food Restaurant	Pool Hall	Juice E
3	Ashish Nagar	33536	6.17	0.088656	0.016094	4762	0.526828	Restaurant	Lake	Bakery	Bus Station	ATM	Juice E
4	Bengali Square	6693	1.26	0.017694	0.003287	3293	0.364310	Coffee Shop	Indian Restaurant	Restaurant	Juice Bar	Lake	Mar
5	Bicholi Hapsi Road	8315	0.26	0.021982	0.000678	6350	0.702511	Hotel	Pool Hall	Indian Restaurant	Juice Bar	Lake	Mar
6	Bicholi Mardana Road	33536	6.17	0.088656	0.016094	2821	0.312092	Hotel	Pool Hall	Indian Restaurant	Juice Bar	Lake	Mar

- cluster 2 has the most common venue Bus Station.

	Neighborhood	Population	area	Normalized_population	Normalized_area	Price for sell per sq.ft	Normalized_range	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	M Com Ve
7	Chhavni	195334	0.66	0.516385	0.001722	2441	0.270052	Bus Station	Zoo	Indian Restaurant	Wings Joint	Train Station	Juice Bar	Lake	Ma
32	Shiv Shakti Nagar Road	250456	13.72	0.662106	0.035788	4565	0.505034	Men's Store	Garden	Food & Drink Shop	ATM	Ice Cream Shop	Juice Bar	Lake	Ma

- cluster 3 has the most common venue Bus Station.

	Neighborhood	Population	area	Normalized_population	Normalized_area	Price for sell per sq.ft	Normalized_range	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	M Com Ve
14	Lokmanya Nagar	107816	6.35	0.285022	0.016564	6018	0.665782	Indian Restaurant	Train Station	Bakery	Tea Room	Farmers Market	Plaza	Juice Bar	L

- cluster 4 has the most common venue Juice Bar.

	Neighborhood	Population	area	Normalized_population	Normalized_area	Price for sell per sq.ft	Normalized_range	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7t Mos Commo Venu
0	AB Road	311700	22.99	0.82401	0.059968	2985	0.330236	Indian Restaurant	ATM	Pool Hall	Juice Bar	Lake	Market	Men' Stor
1	Alok Nagar Row Houses	378272	383.37	1.00000	1.000000	2750	0.304237	Hotel	Asian Restaurant	Café	Shopping Mall	Pool Hall	Juice Bar	Lak
24	Rajwada	311700	22.99	0.82401	0.059968	9039	1.000000	Historic Site	Juice Bar	Bakery	Snack Place	Flea Market	Pool Hall	Lak

- cluster 5 has the most common venue Indian and Fast Food Restaurant.

	Neighborhood	Population	area	Normalized_population	Normalized_area	Price for sell per sq.ft	Normalized_range	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	Co V
12	Khajrana Road	72433	6.99	0.191484	0.018233	2752	0.304458	Market	BBQ Joint	Garden	Fast Food Restaurant	ATM	Pool Hall	
13	Khatiwala Tank	89991	4.12	0.237900	0.010747	3834	0.424162	Fast Food Restaurant	Wings Joint	Tea Room	ATM	Plaza	Juice Bar	
19	New Palasia	61258	2.11	0.161942	0.005504	5669	0.627171	Café	Indian Restaurant	Fast Food Restaurant	Hot Dog Joint	Plaza	Snack Place	C
22	Race Course Road	78392	6.09	0.207237	0.015885	5200	0.575285	Ice Cream Shop	Department Store	Smoke Shop	Fast Food Restaurant	Pool Hall	Juice Bar	
23	Rajendra Nagar	53186	6.31	0.140603	0.016459	3000	0.331895	Indian Restaurant	Diner	Train Station	Bus Station	Juice Bar	Lake	M
29	Sainath Colony	77085	3.58	0.203782	0.009338	3360	0.371723	Hotel	Snack Place	Coffee Shop	Indian Restaurant	Train Station	Plaza	
	Tilak Nagar								Snack	Coffee	Indian	Train		

After analyzing the cluster 2 is suitable for opening a restaurant because of the lack of restaurants.

4. RESULTS

After Analyzing the each cluster the model recommends Vijay Nager, South Tukoganj, Khajrana Road, Anoop Nager, Dhar Road, Chhavni as the best locations for opening a restaurant .

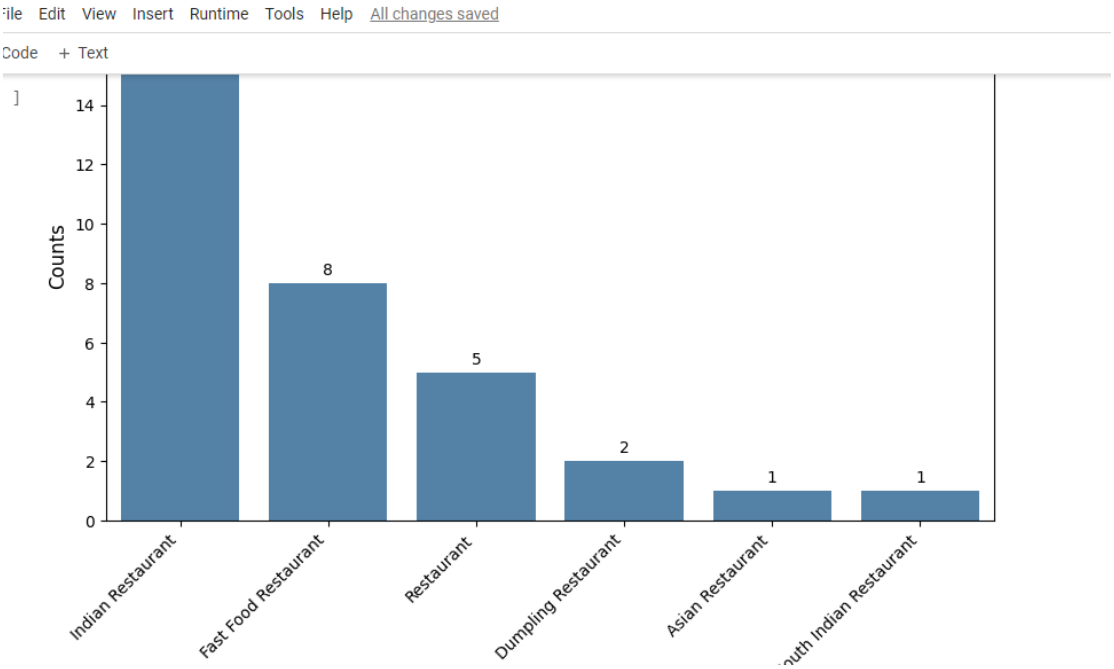


Fig 13: Types of Restaurant in Indore

The Bar Chart shows that Indore has the most Indian Restaurant followed by Fast Food Restaurant .

5. Conclusion

The project involved collecting data on various socioeconomic indicators such as population density, income levels, and restaurant density in different localities of Indore. This data was then used to train a machine learning model, specifically a K-means clustering algorithm, to group localities based on their similarities and identify potential locations for the Indian restaurant.

The results of the project showed that the machine learning model was able to accurately identify the best locality for the Indian restaurant based on the socioeconomic indicators. The recommended locality had a high population density, high income levels, and relatively low restaurant density, indicating a high demand for Indian cuisine in that area. Overall, the project demonstrates the power of machine learning in identifying optimal locations for businesses such as restaurants, and highlights the importance of considering socioeconomic indicators when making such recommendations.

This approach can be extended to other business types and in other cities, providing valuable insights to entrepreneurs and investors looking to start or expand their businesses.

6. Future Scope:

Expanding the scope to other cities: While the project focused on recommending the best locality for an Indian restaurant in Indore, the same approach could be applied to other cities in India or around the world. This would involve collecting data on the relevant socioeconomic indicators in those cities and training the machine learning model to make recommendations based on that data.

Incorporating additional data sources: The project relied primarily on socioeconomic indicators to recommend the best locality for the Indian restaurant. However, additional data sources such as online reviews, social media sentiment analysis, and foot traffic could be incorporated to provide a more comprehensive analysis of the potential location.

Developing a web application: The project could be further developed into a web application that allows users to input their preferences and receive recommendations for the best locality for their business. This would involve building a user-friendly interface that incorporates the machine learning model and the relevant data sources.

Expanding the business types: While the project focused on identifying the best locality for an Indian restaurant, the same approach could be applied to other business types such as cafes, bars, or retail stores. This would involve modifying the machine learning model and the relevant data sources to suit the specific business type.

7. References

1. https://en.wikipedia.org/wiki/Indore#:~:text=Indore%20
2. https://www.makaan.com/price-trends/property-rates-for-buy-in-indore
3. https://geoiq.io/places/Indore/EnJPosJucr
4. https://www.mapsofindia.com/indore/localities/
5. https://docs.python.org/3/
6. https://medium.com/@aboutiana/a-brief-guide-to-using-foursquare-api-with-a-hands-on-example-on-python-6fc4d5451203
7. https://geopy.readthedocs.io/en/stable/