

# **MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR**

(A Govt. Aided UGC Autonomous & NAAC Accredited Institute Affiliated to RGPV, Bhopal)



**Project Report**

**on**

**Optimal Location Predictor**

**Submitted By:**

**Abhinav Chaturvedi**

**0901CS191003**

**Sanidhya Somwanshi**

**0901CS191107**

**Faculty Mentor:**

**Mr. Mir Shahnawaz Ahmad, Professor, Computer Science and Engineering**

**Dr. Anjula Mehto, Professor, Computer Science and Engineering**

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

**MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE**

**GWALIOR - 474005 (MP) est. 1957**

**MAY-JUNE 2022**

# **MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR**

(A Govt. Aided UGC Autonomous & NAAC Accredited Institute Affiliated to RGPV, Bhopal)



**Project Report**

**on**

**Optimal Location Predictor**

A project report submitted in partial fulfilment of the requirement for the degree of

**BACHELOR OF TECHNOLOGY**

**in**

**COMPUTER SCIENCE AND ENGINEERING**

Submitted by:

**Abhinav Chaturvedi**

**0901CS191003**

**Sanidhya Somwanshi**

**0901CS191107**

**Faculty Mentor:**

**Mr. Mir Shahnawaz Ahmad, Professor, Computer Science and Engineering**

**Dr. Anjula Mehto, Professor, Computer Science and Engineering**

Submitted to:

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE**

**GWALIOR - 474005 (MP) est. 1957**

**MAY-JUNE 2022**

# **MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR**

(A Govt. Aided UGC Autonomous & NAAC Accredited Institute Affiliated to RGPV, Bhopal)

## **CERTIFICATE**

This is certified that **Abhinav Chaturvedi** (0901CS191003) and **Sanidhya Somwanshi** (0901CS191107) has submitted the project report titled **Optimal Location Predictor** under the mentorship of **Mr. Mir Shahnawaz Ahmad** and **Dr. Anjula Mehto**, in partial fulfilment of the requirement for the award of degree of Bachelor of Technology in Computer Science and Engineering from Madhav Institute of Technology and Science, Gwalior.

**Mr. Mir Shahnawaz Ahamad**

Faculty Mentor

Professor

Computer Science and Engineering

**Dr. Manish Dixit**

Professor and Head

Computer Science and Engineering

**Dr. Anjula Mehto**

Faculty Mentor

Professor

Computer Science and Engineering

# **MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR**

(A Govt. Aided UGC Autonomous & NAAC Accredited Institute Affiliated to RGPV, Bhopal)

## **DECLARATION**

We hereby declare that the work being presented in this project report, for the partial fulfilment of requirement for the award of the degree of Bachelor of Technology in Computer Science and Engineering at Madhav Institute of Technology & Science, Gwalior is an authenticated and original record of my work under the mentorship of **Mr. Mir Shahnawaz Ahmad, Professor**, Computer Science and Engineering and **Dr. Anjula Mehto, Professor**, Computer Science and Engineering.

We declare that we have not submitted the matter embodied in this report for the award of any degree or diploma anywhere else.

Abhinav Chaturvedi  
0901CS191003  
3rd Year  
Computer Science and Engineering

Sanidhya Somwanshi  
0901CS191107  
3rd Year  
Computer Science and Engineering

# **MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR**

(A Govt. Aided UGC Autonomous & NAAC Accredited Institute Affiliated to RGPV, Bhopal)

## **ACKNOWLEDGEMENT**

The full semester project has proved to be pivotal to my career. I am thankful to my institute, **Madhav Institute of Technology and Science**, for allowing me to continue my disciplinary project as a curriculum requirement, under the provisions of the Flexible Curriculum Scheme (based on the AICTE Model Curriculum 2018), approved by the Academic Council of the institute. I extend my gratitude to the Director of the institute, **Dr. R. K. Pandit** and Dean Academics, **Dr. Manjaree Pandit** for this.

I would sincerely like to thank my department, **Department of Computer Science and Engineering**, for allowing me to explore this project. I humbly thank **Dr. Manish Dixit**, Professor and Head, Department of Computer Science and Engineering, for his continued support during the course of this engagement, which eased the process and formalities involved.

I am sincerely thankful to my faculty mentors. I am grateful to the guidance of **Mr. Mir Shahnawaz Ahmad, Professor**, Computer Science and Engineering and **Dr. Anjula Mehto, Professor**, Computer Science and Engineering for their continued support and guidance throughout the project. I am also very thankful to the faculty and staff of the department.

Abhinav Chaturvedi  
0901CS191003  
3rd Year  
Computer Science and Engineering

Sanidhya Somwanshi  
0901CS191107  
3rd Year  
Computer Science and Engineering

# Abstract

When opening a new restaurant, geographical placement is of prime importance in determining whether it will thrive. Although some methods have been developed to assess the attractiveness of candidate locations for a restaurant, the accuracy is limited as they mainly rely on traditional data sources, such as demographic studies or consumer surveys. In this project, we particularly take advantage of pre-existing restaurants to construct predictive features for assessing the attractiveness of candidate locations to expand a restaurant. Specifically, we investigate market competitiveness, and geographic characteristics of a location under consideration for a prospective restaurant. We devise the three sets of features and incorporate them into a classification model to predict the optimality of opening a restaurant at a candidate's desired location. Moreover, our experimental results suggest that market attractiveness and market competitiveness features are more predictive than geographic features. We are going to use Foursquare API for getting the information about pre-existing restaurants in that area.

**Keywords:** Geographical, Classification Model, Demographic, Market Competitiveness, Foursquare API

## सार:

एक नया रेस्तरां खोलते समय, भौगोलिक स्थिति यह निर्धारित करने में प्रमुख महत्व रखती है कि क्या यह पनपेगा। हालांकि रेस्तरां के लिए उम्मीदवार स्थानों के आकर्षण का आकलन करने के लिए कुछ तरीके विकसित किए गए हैं, सटीकता सीमित है क्योंकि वे मुख्य रूप से जनसांख्यिकीय अध्ययन या उपभोक्ता सर्वेक्षण जैसे पारंपरिक डेटा स्रोतों पर भरोसा करते हैं। इस परियोजना में, हम विशेष रूप से पूर्व-मौजूदा रेस्तरां का लाभ उठाते हैं ताकि एक रेस्तरां का विस्तार करने के लिए उम्मीदवार स्थानों के आकर्षण का आकलन करने के लिए भविष्य कहनेवाला सुविधाओं का निर्माण किया जा सके। विशेष रूप से, हम एक संभावित रेस्तरां के लिए विचाराधीन स्थान की बाजार प्रतिस्पर्धा और भौगोलिक विशेषताओं की जांच करते हैं। हम सुविधाओं के तीन सेट तैयार करते हैं और एक उम्मीदवार के वांछित स्थान पर एक रेस्तरां खोलने की इष्टतमता की भविष्यवाणी करने के लिए उन्हें एक वर्गीकरण मॉडल में शामिल करते हैं। इसके अलावा, हमारे प्रयोगात्मक परिणाम बताते हैं कि भौगोलिक विशेषताओं की तुलना में बाजार आकर्षण और बाजार प्रतिस्पर्धात्मक विशेषताएं अधिक अनुमानित हैं।

# TABLE OF CONTENTS

TITLE	PAGE NO.
<b>Abstract</b>	
सार	
<b>List of figures</b>	
<b>Chapter 1: Project Overview</b>	1
1.1 Introduction	1
1.2 Objective and Scope	1
1.3 Project Features	1
1.4 Feasibility	2
1.4.1 Operational Feasibility	2
1.4.2 Economic Feasibility	2
1.4.3 Legal Feasibility	2
1.5 System Requirements	3
<b>Chapter 2: Literature Review</b>	4
2.1 Quantitative Analysis	4
2.2 Competitor Analysis	4
2.3 Neighborhood Analysis	5
<b>Chapter 3: Preliminary Design</b>	6
3.1 Software Development Life Cycle Model	6
3.1.1 Rapid Application Development	6
3.2 Data Flow Diagram	6
3.3 Tools & Technologies	7
3.3.1 Python	7
3.3.2 Machine Learning	7
3.3.3 Jupyter Notebook	7
3.3.4 Visual Studio Code	7
3.3.5 Anaconda	7
3.3.6 Libraries Used	8
3.3.6.1 Numpy	8
3.3.6.2 Pandas	8
3.3.6.3 Seaborn	8



3.3.6.4 Matplotlib	8
3.3.6.5 Scikit-Learn	8
3.3.6.6 Pickle	8
3.4 What is API & how does it work?	9
3.4.1 Mapquest API	9
3.4.2 FourSquare API	9
<b>Chapter 4: Final Analysis And Design</b>	<b>10</b>
4.1 Calculation of grid & distance	10
4.2 Results	11
4.3 Result Analysis	12
4.4 Application	13
4.5 Problems Faced	13
4.6 Limitations	13
4.7 Conclusion	13
4.8 Future Scope	14
<b>References</b>	<b>15</b>

## LIST OF FIGURES

Figure Number	Figure caption	Page No.
3.2.1	Data Flow Diagram	6
3.4.1	How API works	9
4.1.1	Grid	10
4.2.1	Dataframe	11
4.2.2	Zones	11
4.3.1	Optimal Zones	12

# **Chapter 1: Project Overview**

## **1.1 Introduction**

Delhi, officially the National Capital Territory (NCT) of Delhi, is a city and a union territory of India containing New Delhi, the capital of India. Straddling the Yamuna river, primarily its western or right bank, Delhi shares borders with the state of Uttar Pradesh in the east and with the state of Haryana in the remaining directions. The NCT covers an area of 1,484 square kilometres. According to the 2011 census, Delhi's proper population was over 11 million, while the NCT's population was about 16.8 million. Delhi's urban agglomeration, which includes the satellite cities of Ghaziabad, Faridabad, Gurgaon and Noida in an area known as the National Capital Region (NCR), has an estimated population of over 28 million, making it the largest metropolitan area in India and the second-largest in the world. It is one of the largest cultural, economic, and educational centres of India. Delhi is home to the second highest number of billionaires and millionaires of any city in India. Delhi ranks fifth among the Indian states and union territories in the human development index. Delhi has the second highest GDP per capita in India. The diversity of the cuisine available is reflective of the social and economic diversity of Delhi. Indian, Chinese, Asian, Italian, Middle Eastern, Thai are some of the most popular in the city. We will analyse each of these restaurants to predict a suitable location for the respective restaurant.

## **1.2 Objective and Scope**

In this project we will try to find an optimal location for a restaurant. Specifically, this project will be targeted to stakeholders interested in opening an Indian restaurant in Delhi, India. Since there are lots of restaurants in Delhi we will try to detect locations that are not already crowded with restaurants. We are also particularly interested in areas with no Indian restaurants in the vicinity. We would also prefer locations as close to the city centre as possible, assuming that the first two conditions are met. We will use our data science powers to generate a few most promising neighbourhoods based on these criteria. Advantages of each area will then be clearly expressed so that the best possible final location can be chosen by stakeholders.

## **1.3 Project Features**

In particular, the concluded analysis will target stakeholders who want to set up one of the popular restaurants in Delhi, India. We will use various data science and analysis techniques to reach our goal of selecting optimal locations. The advantages of each area will then be clearly expressed so that the best possible final location can be chosen by stakeholders.

## **1.4 Feasibility**

### **1.4.1 Operational Feasibility**

In Operational Feasibility the degree of providing service to requirements is analyzed along with how easy the product will be to operate and maintain after deployment. Along with this other operational scopes are determining usability of product, Determining suggested solution by software development team is acceptable or not etc.

The project is feasible in terms of operations as it can be implemented anywhere with internet connectivity and system to process

### **1.4.2 Economic Feasibility**

In the Economic Feasibility study, the cost and benefit of the project are analyzed. This means under this feasibility study a detailed analysis is carried out of what will be the cost of the project for development which includes all required costs for final development like hardware and software resources required, design and development cost and operational cost and so on. After that, it is analyzed whether the project will be beneficial in terms of finance for the organization or not.

The project has an economical constraint as the API with more number of request and good internet connectivity require more budget.

### **1.4.3 Legal Feasibility**

In Legal Feasibility study project is analyzed from a legal point of view. This includes analyzing barriers of legal implementation of project, data protection acts or social media laws, project certificate, license, copyright etc. Overall it can be said that Legal Feasibility Study is a study to know if proposed project conform to legal and ethical requirements. The project is feasible legally.

The data and their respective data sources that are helpful for our is discussed below,

- Delhi city's major areas and neighborhoods are provided by the Wikipedia web page.
- Geographical coordinates of the areas are provided by Python's geocoder library.
- Venues in each locality of Delhi city are provided by Foursquare API. The coordinate information is used by the Foursquare API to fetch the venues within the specified radius and venue limits.

## **1.5 System Requirements**

Windows Based Requirements:

Computers running Microsoft Windows must meet the following minimum hardware and software requirements.

Microsoft Windows: 7/8/10/11

4 GB RAM minimum, 8 GB RAM recommended

1GB of available disk space minimum

1280 \* 800 minimum screen resolution

Software Requirement: Python 3.10.4

Hardware Requirement: Laptop/Computer

Internet Connectivity

## **Chapter 2: Literature Review**

### **2.1 Quantitative Analysis**

While in the past site selection may have been based on intuition, a wide spectrum of techniques is used for site selection and potential and current market study. This may include data analysis, sales forecasting, general area analysis of economic and demographic conditions, potential competition and growth, or simply checklists. One of the most effective and comprehensive techniques used by geographers for site evaluate is gravity modelling [4].

Quantitative research can be defined as any research that uses of numbers as the basis for generating inferences about the phenomenon under study. The statistical approaches to sampling, measurement, and data analysis, are a hallmark of quantitative research; statistics are genuinely relevant to quantitative approaches because it involves statistical modeling of the interrelationships between variables. Among the variety of methodologies and data analysis strategies that are employed in quantitative research, there is the relational or correlational research strategy [1]. Furthermore, correlational research is in charge of investigating the nature of the relationship between the variables (or factors) and getting and testing the theoretical model that might explain the resultant correlation.

When there are two quantitative variables (of interval on ration scale of measurement), it is possible to validate their relationship through mathematics and geometry statistical tests. If the mathematical qualities of a line are used to calculate the systematic change in the scores of a dependent variable (y) from an independent variable (x), its correlation is being calculated.

The procedure to overcome the best estimates of a variable y, taking into account its relationship with a variable x, is known as simple or bivariate linear correlation and regression analysis. This procedure consists of applying the formulas to a straight line to get y-intercept.

### **2.2 Competitor Analysis**

Competitor restaurants can help in better understanding the restaurant location demographics. However, competitor restaurant can pose threat to the restaurant based on similar concept. Complimentary restaurants are restaurants which have different restaurant concepts but similar price points. These complimentary restaurants can help in creating a market for the proposed restaurant [3].

## **2.3 Neighborhood Analysis**

The infrastructure of the neighborhood area is a key aspect in location analysis. The future growth of the proposed restaurant can be predicted depending on the presence of educational institutions, movie hall, and many others in that area [3].

In this project, we are using the Competitor Analysis Technique with the Quantitative (statistical & geometrical) approach.

## Chapter 3: Preliminary Design

### 3.1 Software Development Life Cycle Model

#### 3.1.1 Rapid Application Development

**Reason:** since the software size was not much large and there was a time-bound and the project was made in modules therefore in this project, I used Rapid Application Development. A software project can be implemented using this model if the project can be broken down into small modules wherein each module can be assigned independently to separate teams. These modules can finally be combined to form the final product.

### 3.2 Data Flow Diagram

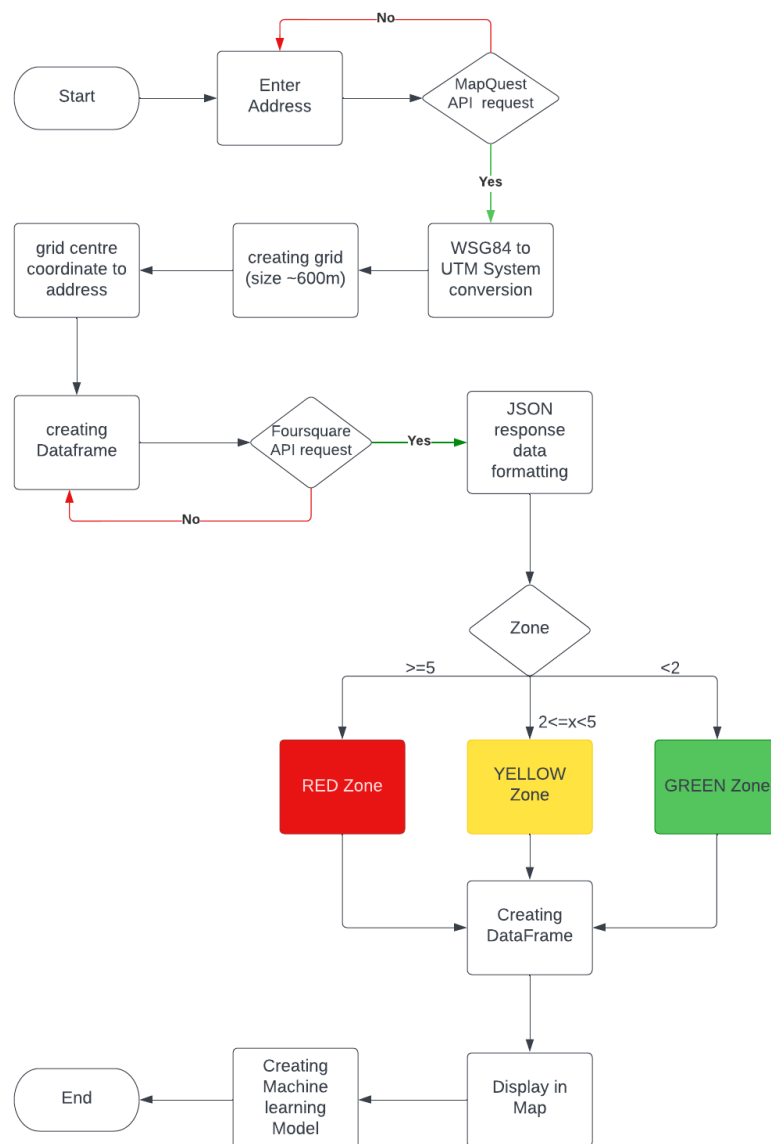


Fig 3.2.1 Data flow diagram



## **3.3 Tools & Technologies**

### **3.3.1 Python**

Python is a high-level, interpreted, interactive, and object-oriented scripting language. Python is designed to be highly readable. It uses English keywords frequently whereas other languages use punctuation, and it has fewer syntactical constructions than other languages.

### **3.3.2 Machine Learning**

Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention.

### **3.3.3 Jupyter Notebook**

Jupyter Notebook is a web-based interactive computational environment for creating notebook documents. A Jupyter Notebook document is a browser-based REPL containing an ordered list of input/output cells which can contain code, text, mathematics, plots and rich media. Underneath the interface, a notebook is a JSON document, following a versioned schema, usually ending with the ".ipynb" extension.

### **3.3.4 Visual Studio Code**

Visual Studio Code is a source-code editor made by Microsoft for Windows, Linux and macOS. Features include support for debugging, syntax highlighting, intelligent code completion, snippets, code refactoring, and embedded Git. Users can change the theme, keyboard shortcuts, preferences, and install extensions that add additional functionality.

### **3.3.5 Anaconda**

Anaconda is a distribution of the Python and R programming languages for scientific computing that aims to simplify package management and deployment. The distribution includes data-science packages suitable for Windows, Linux, and macOS. We used anaconda to get access to Anaconda Navigator and Anaconda Prompt.

### **3.3.6 Libraries Used**

#### **3.3.6.1 Numpy**

NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays. NumPy is open-source software and has many contributors.

#### **3.3.6.2 Pandas**

Pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. It is free software released under the three-clause BSD license. The name is derived from the term "panel data", an econometrics term for data sets that include observations over multiple time periods for the same individuals.

#### **3.3.6.3 Seaborn**

Seaborn is a data visualization library built on top of matplotlib and closely integrated with Pandas data structures in Python. Visualization is the central part of Seaborn which helps in exploration and understanding of data.

#### **3.3.6.4 Matplotlib**

Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK.

#### **3.3.6.5 Scikit-Learn**

Scikit-learn is a free software machine learning library for the Python programming language. It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, k-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.

#### **3.3.6.6 Pickle**

Pickle module is used for serializing and de-serializing a Python object structure. Any object in Python can be pickled so that it can be saved on disk. What pickle does is that it “serializes” the object first before writing it to file. Pickling is a way to convert a python object into a character stream. The idea is that this character stream contains all the information necessary to reconstruct the object in another python script.

### 3.4 What is API & how does it work?

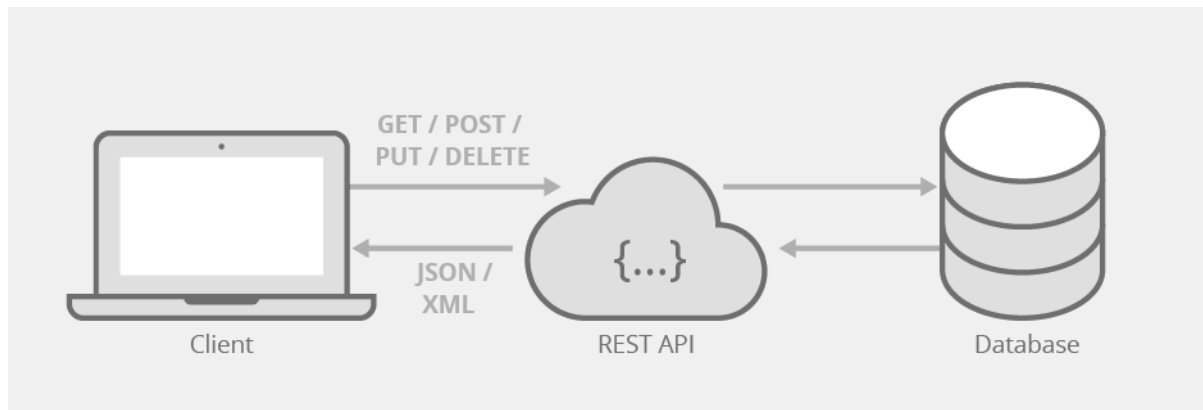


Fig 3.4.1 how API works

API is an acronym for Application Programming Interface that software uses to access data, server software or other applications and have been around for quite some time.

APIs communicate through a set of rules that define how computers, applications or machines can talk to each other. The API acts as a middleman between any two machines that want to connect with each other for a specified task.

#### 3.4.1 Mapquest API

The core technology of online mapping is a process called geocoding, in which the street address of a location is converted into specific geographic coordinates (longitude and latitude). Once a location is geocoded, it can be pinned to a precise location on an online map.

#### 3.4.2 Foursquare API

The Foursquare Places API provides location-based experiences with diverse information about venues, users, photos, and check-ins. The API supports real-time access to places, Snap-to-Place that assigns users to specific locations, and Geo-tag. Additionally, Foursquare allows developers to build audience segments for analysis and measurement. JSON is the preferred response format.

Foursquare allows users to input both a city and keywords related to what they're looking for there into its search bar. Users can then filter their search results — and read reviews on them — to choose a place to go. It will also start to recommend places to you based on your searches.

## Chapter 4: Final Analysis and Design

### 4.1 Calculation of grid & distance:

To accurately calculate distances we need to create our grid of locations in a Cartesian 2D coordinate system which allows us to calculate distances in meters (not in latitude/longitude degrees). Then we'll project those coordinates back to latitude/longitude degrees to be shown on the Folium map.

After converting the WGS84 spherical coordinate system (latitude/longitude degrees) to UTM Cartesian coordinate system (X/Y coordinates in meters)

Let's create a hexagonal grid of cells: we offset every other row, and adjust vertical row spacing so that every cell's center is equally distant from all neighbors.

After creating grid of cells, the distance between each cell will be calculated using Euclidean Distance:

$$d = \sqrt{[(x_2 - x_1)^2 + (y_2 - y_1)^2]}$$

- $x_{11}, y_{11}$  are the coordinates of one point.
- $(x_{22}, y_{22})$  are the coordinates of the other point.
- $d$  is the distance between  $(x_{11}, y_{11})$  and  $(x_{22}, y_{22})$ .

We have taken India Gate as centre for making grid. After calculation we got 364 neighborhood candidates.

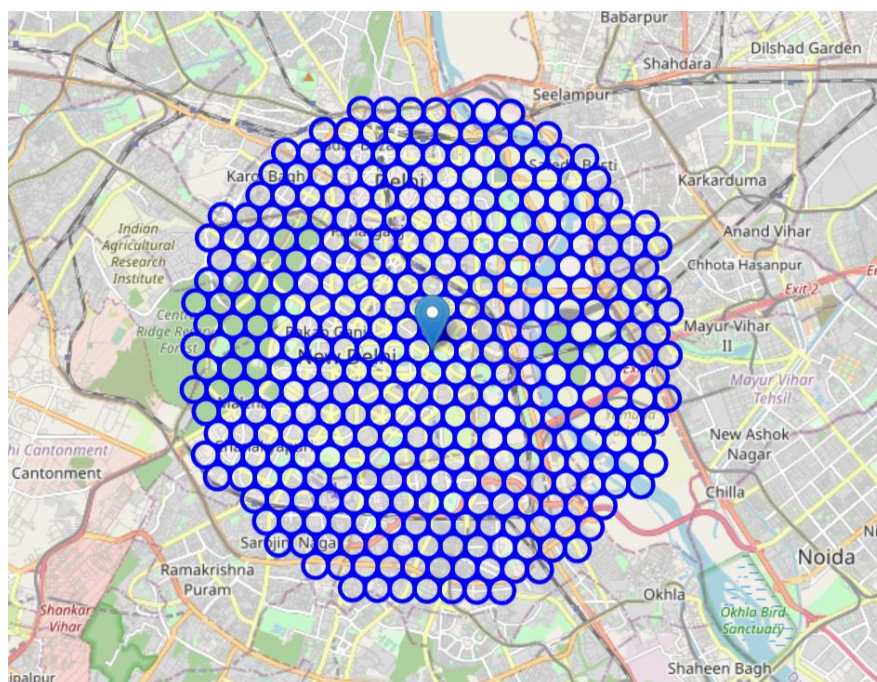


Fig 4.1.1 grid

## 4.2 Results

Our analysis shows that although there is a great number of restaurants in Delhi (~1600 in our initial area of interest which was 12x12km around Delhi centre), so we deduce the number of restaurant by applying a specific category (Indian Restaurant ) now, there are pockets of low restaurant density fairly close to the city centre. Highest concentration of restaurants was detected north and west from Delhi centre, so we focused our attention to areas south, south-east and east, corresponding to boroughs. We got 847 restaurants in 364 neighborhood coordinates.

	Neighborhood_Latitude	Neighborhood_Longitude	indian_Venue	indian_Venue_Latitude	indian_Venue_Longitude	Distance	Venue_Category
0	28.564017	77.210292	Himalaya Dhaba	28.561533	77.211215	290	Indian Restaurant
1	28.564017	77.210292	Rhythm Restro-Bar	28.560570	77.207794	454	Indian Restaurant
2	28.564017	77.210292	mann mess	28.560317	77.212707	474	Indian Restaurant
3	28.564017	77.210292	Sri Balaji South Indian Restaurant	28.562191	77.210866	210	South Indian Restaurant
4	28.563917	77.216422	Shri vinayaka restaurant	28.563822	77.219529	303	Indian Restaurant
...	...	...	...	...	...	...	...
842	28.662188	77.227694	The Chowk (Bar and Indian Kitchen)	28.662008	77.228237	56	Indian Restaurant
843	28.662188	77.227694	Comesum	28.660668	77.227810	169	Restaurant
844	28.662087	77.233829	Bittu Tikki Wala	28.661540	77.229294	447	Indian Restaurant
845	28.667126	77.212453	Aaoji Khaoji	28.665388	77.209558	342	Indian Restaurant
846	28.666824	77.230861	Hotel Samrat Jhansi	28.669731	77.233165	394	Indian Restaurant

847 rows × 7 columns

Fig 4.2.1 Dataframe

After directing our attention to this more narrow area of interest, we first created a dense grid of location candidates (spaced 300m apart); those locations were then filtered so that those with more than two restaurants in a diameter of 600m and those with an Indian restaurant closer were removed. Those location candidates were then clustered to create zones of interest that contain the greatest number of location candidates. Addresses of centers of those zones were also generated using reverse geocoding to be used as markers/starting points for more detailed local analysis based on other factors.

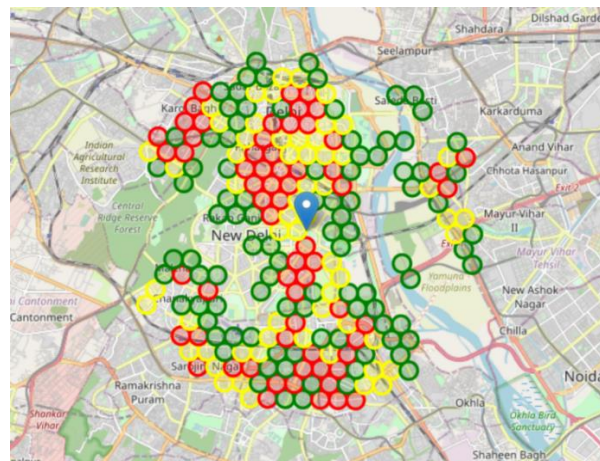


Fig 4.2.2 Zones



### 4.3 Result Analysis

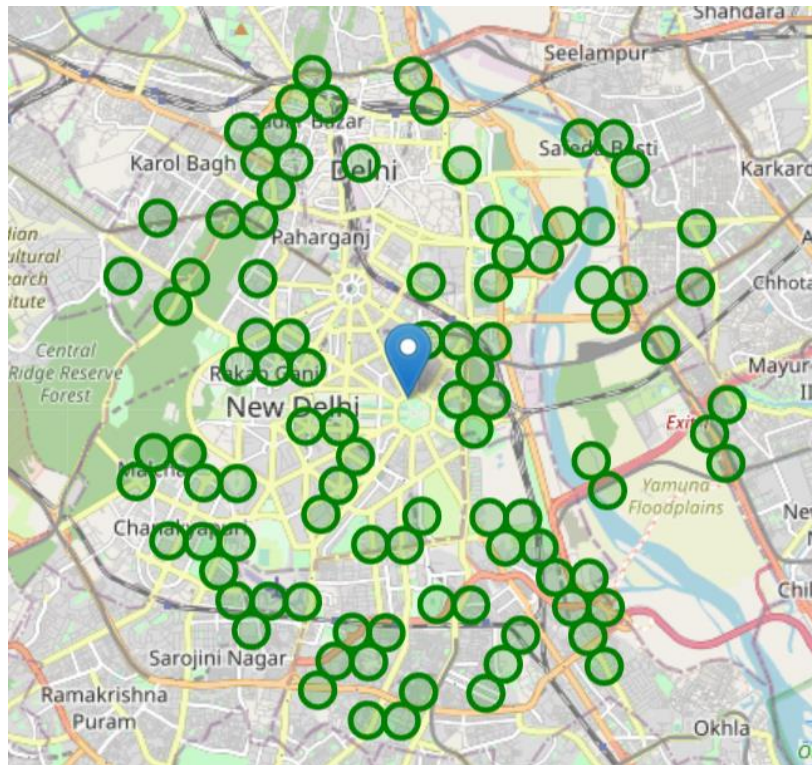


Fig 4.3.1 Optimal Zones

Result of all these green zones containing the largest number of potential new restaurant locations based on number of and distance to existing venues — both restaurants in general and Indian restaurants particularly. This, of course, does not imply that those zones are actually optimal locations for a new restaurant! The purpose of this analysis was to only provide info on areas close to Delhi center but not crowded with existing restaurants (particularly Indian) - it is entirely possible that there is a very good reason for a small number of restaurants in any of those areas, reasons which would make them suitable for a new restaurant regardless of lack of competition in the area. Recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually result in a location that has not only no nearby competition but also other factors taken into account and all other relevant conditions met.

On applying different machine learning algorithms to the result, where latitudes and longitudes will be input variables and zones will be the output variable.

The accuracy Score of the Random Forest Classifier is highest at about 0.5757. We can improve the accuracy score by increasing the dataset and applying overfitting algorithms by using ensemble learning, and cross-validation techniques.

## **4.4 Application**

This can be used by the stakeholders for getting information about the market place where they want to open a new restaurant. This will also help stakeholders by predicting the optimal place for opening of restaurant, and can also predict whether the location chosen by them comes in green zone, yellow zone or red zone.

## **4.5 Problems Faced**

While developing the project we faced various problems some of them are:

- Since Goggle map API is paid which leads to use an alternative of it that is MapQuest.
- Alternative of google map is not much accurate to that extinct.
- Connection Time out while API call is a big issue.

## **4.6 Limitations**

- Response of an API call for each neighborhood candidate is time taking.
- Number of API request per hour is quite low which affects the system efficiency.
- Low accuracy of Machine Learning Model due to small size of dataset.

## **4.7 Conclusion**

Purpose of this project was to identify Delhi areas close to the centre with a low number of restaurants (particularly Indian restaurants) in order to aid stakeholders in narrowing down the search for optimal location for a new Indian restaurant. By calculating restaurant density distribution from Foursquare data we have first identified general boroughs that justify further analysis, and then generated an extensive collection of locations that satisfy some basic requirements regarding existing nearby restaurants. Clustering of those locations was then performed in order to create major zones of interest (containing the greatest number of potential locations) and addresses of those zone centres were created to be used as starting points for final exploration by stakeholders. The final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighbourhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location

(proximity to park or water), levels of noise/proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.

## **4.8 Future Scope**

- We will try to enhance the dataset by including more cities (like Mumbai, Hyderabad, Pune, Bangalore, etc.)
- We will try to build a user interface.
- We will try to incorporate Restaurant Recommendation System based on Collaborative filtering.
- We will try to improve our data analysis and prediction by adding more features like life expectancy, and customer reviews on the demographic location.
- We will try to add a Customer review system for the pre-existing restaurants in that location which will help to build a community.



## References

1. Berumen Calderón, Mauro Felipe, Damayanti Estolano Cristerna, Angelica Selene Sterling Zozoaga, and Andreé Ricardo Berumen Calderón. "Model to assess the selection of the optimal location for a restaurant, a quantitative approach. Case study: Theme restaurants in Cancun, Mexico." *Journal of Foodservice Business Research* 24, no. 4 (2021): 457-501.
2. Hariharan, R., Arish Pitchai, and M. Dhilsath Fathima. "Prediction of Locations Using Unsupervised Learning Method to Open a Restaurant Branch." In *Conference on Multimedia, Interaction, Design and Innovation*, pp. 59-71. Springer, Cham, 2020.
3. Baksi, Bidisha Das, Varsha Rao, and C. Anitha. "A Survey on Local Market Analysis for a Successful Restaurant Yield." In *Emerging Technologies in Data Mining and Information Security*, pp. 249-257. Springer, Singapore, 2019.
4. Dock, Joel P., Wei Song, and Jia Lu. "Evaluation of dine-in restaurant location and competitiveness: Applications of gravity modeling in Jefferson County, Kentucky." *Applied Geography* 60 (2015): 204-209.