

A Survey on Local Market Analysis for a Successful Restaurant Yield



Bidisha Das Baksi, Varsha Rao and C. Anitha

Abstract Establishment of a new restaurant requires a paramount investment in it. Thus, a thorough analysis of different factors is important to the determination of the probable rate of success of the restaurant. Among the different factors, location plays a vital role in the determination of the success of the restaurant unit. The demographics of the location, existing cluster of restaurants established in the location, and the growth rate of the location need to be studied prior to selection of the optimal location. However, it is a cumbersome job to find the correct location, by analyzing each of these different aspects of the location. This paper studies the different aspects of a location which makes it a determining factor in the prediction of success of a restaurant. The proposed work aims at the creation of a web application that determines the locations suitable for the establishment of a new restaurant by using techniques of machine learning and data mining.

Keywords Restaurant market · Clustering algorithm
Location-based predictive analysis · Web application

1 Introduction

A new venture requires a thorough analysis of market conditions. Restaurant business is one of the most profitable businesses of today's time, and at the same time it involves immense risk. New food joints require large capital investment and operational costs. In [1], the author states that when the new outlet fails to break even, it closes in a

B. Das Baksi · V. Rao · C. Anitha (✉)
Department of Computer Science and Engineering, The National Institute
of Engineering Mysuru, Mysuru 570008, Karnataka, India
e-mail: anithac.cse@nie.ac.in

B. Das Baksi
e-mail: bidisha.bksh@gmail.com

V. Rao
e-mail: varshar186@gmail.com

short span of time, thereby incurring huge losses. Location of the new venture plays an important role in determining its success. Also, as in [2], it is demonstrated that the local socio-demographic condition has significant impact on the profitability of a restaurant. Emerging trends in the food industry play an important role while designing the business plan. Many other factors contribute to the type of menu and the price that can be charged.

It is imperative for restaurateurs to understand the likes and dislikes of the customers [3]. Since location is an important factor in determining success, a thorough analysis of the location needs to be performed before the investment is made. This analysis constitutes investigation of various parameters such as the demographic conditions, local preferences, tourist population, competition in the area, economy of the area, and so on [4, 5]. When an investor wants to open a new restaurant in a city, the investor is unsure of the area in the city that he/she should invest in. Opportunities for new restaurants in various regions can be established based on people's preferences.

To solve the above-stated problem, the proposed system intends to explore factors that determine the success rate of a restaurant by creating an application that accepts various attributes of the new venture as an input and provides the investor with a list of the areas that are suitable for his/her venture and also the approximate annual revenue that can be generated from each of these locations. We aim to design an automated system to determine locations suitable for the establishment of a restaurant business by applying concepts of K-means clustering. The primary objective of the proposed system is to help investors make an informed and optimal decision about selecting a location for their restaurant business before opening a new outlet in the city. This analysis helps investors take a calculated risk.

2 Related Work

This section summarizes the contributions by various researchers' for setting up and running a successful restaurant. There are many factors that determine the success of a restaurant. However, location is deemed to be the most significant factor among others. It is important that the restaurant is in the vicinity of that community, which has sufficient count of the ideal customers that the restaurant targets. Here, we focus on only those contributions that consider location as a major factor for analysis.

2.1 *Analysis Based on Demographics and Psychographics*

In [4], the author states that restaurant style and cuisine type helps in defining the demographics and psychographics of the customer, respectively.

For the identification of the ideal customer, the target market is selected and the ideal customer profile is defined. This involves the determination of three aspects

of the ideal customer: the demographics, the psychographics, and the behavior. An approximate consumer profile would include demographics (age, gender, education, income, family, life, and geographic location), buying habits (role, frequency, online/mobile/offline, and social proof), psychographics (attitudes, values, lifestyle, personality, and interests/hobbies) and persona types (competitive, spontaneous, methodical, and humanistic).

The next step involves the determination of the area where the ideal customer crowd is in majority. For example, in America, a list of zip codes can be created, using Quick Fact Census. It will give information about the demographics identical to each of these zip codes.

Threshold value for population is decided on the basis of capital investment and also the size of business. Once the population is estimated, the percentage of crowd is calculated that can serve as the ideal customer for the restaurant. Based on this percentage of crowd, the list of zip codes can be ranked and ideal location can be confirmed.

Customer Activity Analysis

Customer activity plays an important factor, and it can be of two types: Foot traffic and vehicle traffic. The effect of traffic can be an additional factor to be considered in making decision regarding the best place for restaurant concept.

SAP digital consumer insight tool is a software which provides foot traffic report by leveraging location-enabled mobile phones. The data can be even broken down into age, gender, home zip code, etc. Thus, this tool can be used to estimate the foot traffic.

Vehicle traffic comprises vehicular traffic in the area near to the restaurant locality. Fortunately, all cities, countries, and states periodically conduct a traffic count.

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.

Neighborhood

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.

2.2 Effect of Socio-demographic Factors on Location Pattern

In [2], a number of socio-demographic factors are considered to understand their influence on restaurant location. According to the authors, factors such as cost of establishment, traffic flow, ingress, parking, visibility demographics, competition,

and municipal regulations affect the selection of an ideal location for establishing a restaurant. Also, the assessment of current demand characteristics and how they are likely to evolve over a period of 10 or 20 years are also crucial for selecting the location. In [2], a geographic region is divided into subregions on the basis of zip codes. The probability of a zip code serving as an ideal region for the development of the business is determined. The restaurant location patterns across the US were expressed using the negative binomial (NB) regression model for each zip code. The likelihood of observing a count of restaurant number y_i in zip code i is considered as [2]

$$f(y_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \quad (1)$$

The (conditional) expectation of y_i , λ_i , is specified as a log-linear function of a set of explanatory variables x_i , as [2]

$$\ln E(x_i) = \ln(\lambda_i) = x_i \beta \quad (2)$$

where β is the vector of coefficients of x_i .

In NB model, a gamma distributed term is incorporated in (2), such that

$$\ln E(x_i) = \ln(\lambda_i^*) = x_i \beta + v_i \quad (3)$$

The size of each zip code area varies significantly across the country; therefore, an exposure variable $\ln(AREA_i)$ is included in x_i with coefficient constraint as 1.

Therefore, if both sides of (3) are subtracted by the term $\ln(AREA_i)$, the left-hand side becomes $\ln E\left(\frac{y_i}{AREA_i/x_i}\right)$, which represents the restaurant density.

Thus, restaurant growth from t_1 to t_2 in a particular zip code i can be predicted as

$$\widehat{growth}_i = \frac{y_i(t_1)}{y_i(t_2)} = \exp(\Delta x_i \hat{\beta}) \quad (4)$$

Several socio-demographic factors were considered, and the parameters considered for analysis were based on these factors. The parameters include the population count, median age, median income, educational attainment population, and renter occupation. Their analysis indicated that the higher the parameter values, the larger the supply of restaurants. On the other hand, the other factors which did not contribute were household size, male percent, and owner occupation. In [2], authors conclude that a prime location is directly linked customer in terms of intention, satisfaction, loyalty, and retention. According to [2], smaller cities amid rural areas appear more attractive than more highly urbanized areas to certain restaurant types. Some restaurant brands were directly affected by the housing tenure and owner occupancy percentage. Ethnic neighborhood characteristics also mattered for some brands.

2.3 Other Contributions

Many online un-authored information [5] and articles [4, 6] can be found to facilitate restaurateurs in considering important factors such as industry trends, location market area, competition, customer psychological, and potential revenue projection by providing questionnaires and survey forms. It can be difficult for a restaurateur to understand these factors and it is also time-consuming. According to [3], people rate a restaurant not only based on food but also on dine-scape factors such as facility aesthetics, lighting, ambience, layout, table setting, and servicing staff. These factors priority vary depending on location.

With the abundance of data available on the above factors, analysis on it can be done using various mathematical models and data analysis methods like multiple regression, neural networks, Bayesian network model, random forest, SVM, and many more can be used in predicting potential revenue of restaurant depending on various factors [1, 7]. Not only revenue prediction but also customer preference can be determined. Analysis provides more insights to the restaurateurs and can be used as reference for decision-making.

3 Proposed Work

The shortcomings of the referred papers were observed to be the following. In [4], the authors have categorized the income into three levels: Low, mid, and high. But the author has failed to elaborate on the range in which these three levels are decided. Further, the author has classified high income into high and high disposable income. This type of categorization leads to ambiguity and may not be suitable for all geographic locations around the globe.

In [2], the authors conclude the location of a restaurant and its type is heavily influenced by population density, ethnicity, and urbanization of an area. It highlights relationships between demand characteristics and total restaurant supply. Some market factors are universal in how they influence supply, such as higher population density, median household income, median age, and ethnicity. The factors such as ownership styles and local government policies did not find any weightage in the analysis. In [2], the model predicts business future growth of US-based restaurants. This model is not suitable for Indian-based restaurants. Also, prediction of business growth is not easy as the Indian market is volatile. A lot of other factors were not considered in the analysis such as the local government regulations and competition between different restaurant types.

Considering the Indian scenario, the GDP has grown rapidly in the last few years; accordingly, the lifestyle of people has changed. This transformation is visible with more and more people eating out frequently. Restaurants are cropping up in every corner of the city to meet this increased demand. In order for restaurants to survive in this highly competitive business, location plays a vital role. The other factors include

Table 1 Proposed dataset I

Area/Ward no.	Gender (M/F)	Annual income	Age (range)	Preferences (Chinese/Indian/Continental)	Amount usually spent on eating outside	Amount likely to be spent per meal	No. of members in the household
---------------	--------------	---------------	-------------	--	--	------------------------------------	---------------------------------

capital investment, operational costs, theme, and so on. Location of a new venture plays an important role in determining its success.

The government of India is also coming up with great initiatives such as the development of smart cities. We have proposed a system which provides the investor, a number of suitable locations for investment. The system determines the locations that possess the threshold characteristics required for a successful restaurant. Initially, data related to food preferences, type of cuisine, economy, type of population, ambience preferences, etc. are collected from the customers which are stored in the database and used for the market analysis. The proposed system takes as its input, the investment amount, targeted age group, type of cuisine, ambience, and related information from the investor and provides the suitable locations along with the approximate success rate at each location by using a suitable clustering algorithm on the collected data. The investor can choose among the given locations and take appropriate decision based on the analysis.

3.1 System Design and Datasets

This section describes the proposed system design and the datasets considered. The proposed system is a web application that any investor can use before localizing the restaurant location. Since the data considered need to be dynamically updated to reflect the changes in every part of the city, most of its content directly depends on the data available through various government and authorized websites. The rest of the data need to be collected through surveys. The data thus collected are proposed to be grouped on the basis of ward numbers as per the city corporation's information.

Dataset Representing User Demographics: Dataset I

This set consists of data related to user age, cuisine preferences, annual income, and expenditure habits. A survey form is created and distributed among the citizens of the restaurant to collect information from them (Table 1). The data so collected are cleaned and wrangled to form the desired dataset. This dataset helps to learn about the current industry trend and the market demands (Table 2).

Dataset Representing Areas of Importance: Dataset II

This set consists of areas or structures of economic interest. This can be collected through Google Maps API. This dataset helps to determine the relative importance of different regions (Table 3).

Table 2 Dataset I after data wrangling

Subregion group no.
Male (%)
Female (%)
Mean age range
Mean income
Mean no. of members in the household
Chinese cuisine (%)
Continental cuisine (%)
Indian cuisine (%)
Approximate population
Mean amount likely to be spent on a meal
Mean amount usually spent on food per month

Table 3 Proposed dataset II

Area/Ward no.	Region of interest (Mall/Park/Highway/Flyover/Theatre, Metro)
---------------	---

Table 4 Dataset II after data wrangling

Subregion group no.	Malls (%)	Airport (Yes/No)	Parks (%)	No. of flyovers	Movie halls (%)	Relative importance score
---------------------	-----------	------------------	-----------	-----------------	-----------------	---------------------------

Table 5 Proposed dataset III

Restaurant location (Ward No.)	Specialty cuisine (Indian/Chinese/Continental)	Investment	Average cost for two	No. of customers visit per day	Popularity rating	Zomato rating
--------------------------------	--	------------	----------------------	--------------------------------	-------------------	---------------

Dataset II is wrangled (Table 4). The areas are grouped into subregion and relative importance scores are assigned to each subregion. Based on the type of area considered, the weightage for each of the regions of interest is dynamically decided and considered. For example, the weightage for airport may differ from a Metropolitan city and a town city. The weightage given to each region of interest is stored in a database and updated in small intervals of time, depending on the current scenario of the city/town. For example, the percentage of malls in a particular region multiplied to its assigned weightage adds up to the relative importance score for that region.

Dataset Consisting of Existing Restaurant Data: Dataset III

This set consists of attributes of an existing restaurant business. This dataset can be formed by performing a survey and collecting information from manager of each restaurant. This dataset is used to train the model for predicting annual revenue of a restaurant (Table 5).

Table 6 User input for proposed work

Restaurant type (Indian/Chinese/Continental)
Annual Investment
Average cost for two
Crowd target per day
Points of interest (Mall/Movie Hall/Park/School/ etc.)

User Input

The investor gives input about the restaurant through the website. A survey form is put up on the website which is filled by the user and submitted to the server. The survey form is as shown in Table 6.

Design

The data collected in dataset I (Table 1) is subjected to k-means algorithm where the attributes chosen for the k-means algorithm are male percent, female percent, age range, cuisine preference, mean income, mean no. of members in the household, cuisine percentage, approximate population, mean amount to be spent on a meal, and the mean amount spent on dining outside per month. The optimal number of clusters that is to be formed for a region should be selected after visualization of data. The clusters thus formed have unique characteristics that differentiate them from one another and the members of the same cluster resemble one another in some way or the other. The idea of growth of restaurants on the basis of location factors, as studied in [2], is taken into consideration and each of the clusters is approximated to be suitable for certain types of restaurants. For example, as shown in [2], a higher male percentage in most areas leads to scarcity of all types of restaurants in that area except that of vegetarian restaurants which shows a higher rate of establishment around those areas, with higher male population. The input accepted from the user through the survey form is analyzed and converted into another data point, which is mapped against the clusters formed through k-means clustering algorithm. The appropriate cluster is selected and the subregions in that cluster are displayed as the preferable location for the setting up of the restaurant. The region of importance score is considered, and the subregions are ranked on the basis of that. The dataset III (Table 5) is used to train a machine learning model which is used to predict the annual revenue that can be generated by the restaurant to be established in the regions that are chosen to be suitable for its establishment. The result of this prediction can be used to re-rank the subregions in the chosen cluster.

4 Conclusion

This paper studies the different factors of a location which makes it an important aspect that is to be considered, prior to investing in a new venture at an unknown place. Guidelines for the optimal selection of location for a new venture are existing.

Some researches prove the correlation between the location factors and the growth of restaurant in that location. Based on these studies, our proposed system aims at creating a web application which recommends the appropriate location for the establishment of a new restaurant, thereby reducing the task of the investors in studying each individual factor. However, the proposed system is yet to be implemented and the accuracy is yet to be evaluated. The proposed system serves as the framework for the development of an application, which estimates suitability of locations for the development of new restaurant units.

References

1. Raul, N., Shah, Y., Devganiya, M.: Restaurant revenue prediction using machine learning. *Int. J. Eng. Sci.* **6**(4), 91–94 (2016)
2. Yang, Y., Roehl, W.S., Huang, J.-H.: Understanding and projecting the restaurantscape: the influence of neighborhood socio-demographic characteristics on restaurant location. *Int. J. Hosp. Manag.* **67**, 33–45 (2017)
3. Mahalingam, S., Jain, B., Sahay, M.: Role of physical environment (Dinescape factors) influencing customers' revisiting intention to restaurants. In: *International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (2016)
4. Tarver, E.: How to Choose the Best Restaurant Location for Your Business. Online document
5. Unknown Author.: Restaurant market analysis—a guide. <http://www.menutek.com/assets/restaurant-market-analysis.pdf>
6. Karen, E., Spaeder.: Online documentation on How to Find the Best Location. <https://www.entrepreneur.com/article/73784>
7. Lasek, A., Cercone, N., Saunders, J.: Restaurant sales and customer demand forecasting: literature survey and categorization of methods. In: Leon-Garcia, A., et al. (Eds.) *Smart City 360°. SmartCity 360 2016, SmartCity 360 2015. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, vol. 166. Springer, Cham (2016)