

Restaurant Setup Business Analysis Using Yelp Dataset

Sindhu Hegde*, Supriya Satyappanavar†, Shankar Setty‡

K. L. E. Technological University, Hubli, India

Email: *sindhuhegde56@gmail.com, †supriyasatyappanavar123@gmail.com, ‡shankar@bvb.edu

Abstract—In this paper, we address the issues associated with setting-up of a new restaurant business. To strategize a new restaurant business, we propose a restaurant business framework which comprises of 3 most important tasks, namely, high frequency attributes, most crowded day and location of the restaurant. First, we identify the features/attributes of the restaurants in which the customers are most interested in and provide those facilities and services to increase the profit. Next, we identify the day of the week when the restaurants are heavily crowded so that the best recipes and offers are made available on those days. Finally, since location has a profound effect on the success of a restaurant business, we consider location to be the most important to know the nearby restaurants and their facilities before coming up with a new restaurant business. The performance analysis of the proposed framework was carried out on the standard Yelp dataset. Thus, we found *credit card* to be the most preferred attribute, the most crowded day to be *Monday* and *Divey* to be the most desired ambience among the customers. We also demonstrate how the new restaurant can be setup by identifying the nearest restaurants and the services.

Keywords—Data Analysis, Restaurant business, Yelp dataset, kd-tree

I. INTRODUCTION

Restaurants have always played an essential role in the business, social, intellectual and artistic life of a thriving society. Restaurants are a significant part of many countries and occupy an important place in shaping the overall economy. In the United States, the restaurant industry makes 4% contribution to the Gross Domestic Product [1]. In addition, there has been a steady rate expansion of 2% to 4% in the restaurant industry over the past three decades [2]. According to the National Restaurant Association (NRA) 2017 Restaurant Industry Forecast [1], restaurant and food service sales are estimated to reach \$798.7 billion, with a gain of 4.3% from 2016(\$766 billion) as shown in Figure 1.

The restaurant industry employs 14.7 million individuals and remains the America's second largest private sector employer, providing jobs and careers for about 1 in 10 working employees. As shown in Figure 2, in the next decade, restaurants will add 1.7 million new positions, leading the country's economy in job creation.

With the rapid growth of restaurant business as predicted by NRA [1], the number of techies who crossed over from their IT jobs to set-up a restaurant business are humongous. With the advancement in technology, internet and social media are used as a powerful medium to run the restaurant business. Today, public opinion and expert reviews has taken over the internet for any business success or failure. There are various restaurant

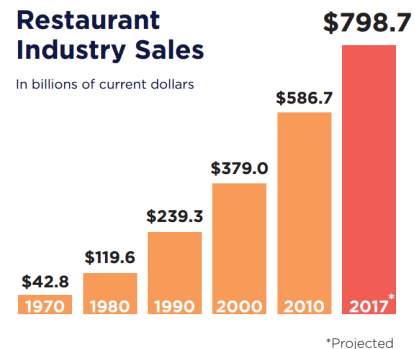


Fig. 1. Restaurant Industry Sales Forecast (Image courtesy [1])

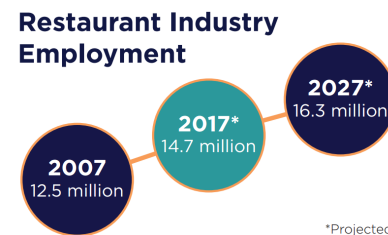


Fig. 2. Restaurant Industry Employment (Image courtesy [1])

review sites like Yelp, Zomato, Open Table, Zagat, Trip Advisor, Gayot, Dine [3] where customers can share their experiences about the food and the product quality. These review sites have helped transform the business strategies in restaurant industry by modernizing facilities like ambience, Wi-Fi, credit card, delivery and the most important, the location of the restaurant. Thus, in this paper, we address the problem of starting the restaurant business by strategizing the most suitable business location and providing most relevant facilities which would help to improve the customer dining experience and also increase the profit. We use the Yelp Dataset Challenge 2014 [4] to strategize the restaurant business by providing relevant services. Yelp dataset includes data from Phoenix, Las Vegas, Madison, Waterloo and Edinburgh and information about 42,153 businesses, 320,002 business attributes, 31,617 check-in sets and 1,125,458 reviews.

In this paper, we propose a solution to the new entrants into the restaurant business with the help of deep analysis of data which helps in interpreting the requirements of the customers. We identify the attributes such as *delivery*, *Wi-Fi*, *dessert*, *lunch*, *late night*, *dinner*, *breakfast*, *reservations*, *credit card*, *wheelchair accessible* whose frequency is quite high in the reviews. We estimate the capital investment over these

attributes which helps in improvising the restaurant business. We then determine the most crowded day of the week by exploring the customer reviews and use *review_count* with the assumption that the customers write the reviews on the same day they visit the restaurant. This analysis would help the restaurants to offer the best recipes and services they have, on those days. Also, when the queues are longer, the menu can feature only those items which are popular and can be prepared quickly which reduces the waiting time of the customers. These subtle changes can prove to be of great help to improve the customer experience, ultimately leading to increase in the profit.

Not every available location is suitable to open a restaurant. The success or failure of a restaurant business is largely dependent on the selection of the location. Apart from the physical site, the surrounding demographics also have a profound impact on the success of a location. In other words, location is a complex construct that encompasses geographic, demographic and psycho-graphic variables. Any variations in these variables might have significant impact on the location [2]. In order to address this problem, nearest restaurant search is done using kd-tree algorithm [5]. Since kd-tree is a k-dimension tree which is used in searches involving multi-dimensional search like range search and nearest neighbour search, we consider it to be the most efficient algorithm in geographically finding the nearby restaurants.

Towards this we make the following contributions:

- Proposing a three step framework for setting-up a restaurant business.
- Identifying the most relevant features/attributes and crowded days of the week, a restaurant is expected to have.
- Finding the nearest restaurants for setting-up the new restaurant business using kd-tree algorithm.
- Performing the experimental analysis of the proposed framework using the standard Yelp business dataset.

The rest of the paper is organized as follows: Section 2 presents the related work. Section 3 deals with methodology comprising of three tasks in setting-up new restaurant business. Experimental results are discussed in Section 4. Conclusions and future work is dealt in Section 5.

II. RELATED WORK

In this section, we provide a brief survey of the efforts that have been invested in information collection from the web with respect to restaurant data analysis using Yelp dataset.

In [6] authors, developed a system that predicts a business star from reviews. To determine the best forecast for choosing the restaurant category, a combination of three feature generation methods and four machine learning models were used from the Yelp dataset. By using top frequent words from the raw text reviews and top frequent words/adjectives from Parts of Speech (POS) analysis, the sack of words was collected and root mean square error was computed.

In [7] authors, the aim was to detect some of the effective facts about users, business and reviews. This was done

by executing some simple queries by considering customer reviews of business and other attributes from the Yelp dataset. The Yelp dataset comprises of the reviews written by the customers which are associated with the ratings. These ratings range from 1(negative) to 5(positive) stars. Data visualization and statistical analysis was used by the authors to determine the relationship between the success of the business and its geographic location, the business check-in information of the weekly and daily cycle and what makes a review useful.

In [8] authors, analyzed the customer reviews from the Yelp dataset and proposed how experience goods markets were affected by these reviews. To get the revenue data, the authors made use of Washington State Department of Revenue. The method presented three major findings: (1) Revenue increased by 5%–9% when the Yelp rating increased by one-star, (2) This impact was propelled by independent restaurants, (3) As the penetration of the Yelp dataset elevated, there was a decrease in the revenue share of chain restaurants. It was observed that when the information in the Yelp ratings was more, there was a large and a better response from the customers. Nevertheless, customers also reacted more firmly to the details that were more visible, implying that the presentation of the information also mattered.

In [9] authors, investigated the effects of restaurant environment on consumer behavior. The authors presented the relationship between restaurant environment and the behavioral intentions of the customers. The primary data was obtained by self-administered questionnaires. These questionnaires were distributed to patrons in selected restaurants as customers were waiting for their checks or as they were getting ready to leave after meals. The contents of questionnaires were divided into four different parts consisting of dimensions of restaurant atmospherics, participants' dining satisfaction level, participants' behavioral intentions and personal information such as participants' age, gender, job, income, educational level and nationality. The restaurants were chosen according to the Yelp dataset. After the survey, the frequency of each response was computed and summarized. The analysis of the data provided a general idea about consumers' attitude towards the restaurant environment.

In [10] authors, predicted the rating for a restaurant from the user's review histories and restaurant's statistics. The machine learning algorithm such as linear regression, random forest tree and latent factor model along with the sentiment analysis was used. The method was used to analyze the performance of each model and select the best model for predicting the ratings from the reviews. The Yelp dataset was used for training, validation and testing the models. The features like count of reviews, length of reviews, polarity of sentiment of review text, average rating star for each user, average rating star for each business were used as key for prediction of best model.

In [11] authors, described the latent subtopics from Yelp restaurant reviews by running an online Latent Dirichlet Allocation (LDA) algorithm. The goal was to identify what are the most influential factors that customers care while writing the reviews. The reviews were broken down into latent subtopics using online LDA, a generative probabilistic model for collections of discrete data. For all the hidden topics over the reviews, the stars were predicted and also

the information regarding the restaurant's peak hours were determined. Through these predictions, inference were made on how to improve the restaurant business.

However, based on our survey we observe that no emphasis was given on the best location to setup a restaurant business, and the attributes like credit card, wheel chair accessible which have the highest frequencies and the most appealing ambience for making profit.

III. METHODOLOGY

The proposed framework for restaurant business setup analysis is shown in Figure 3. Initially, we extract the restaurant business information from the Yelp dataset. To setup a restaurant business, we build a strategic plan consisting of 3 tasks:

- **Task 1:** We identify those attributes of a restaurant in which a customer is most interested in.
- **Task 2:** We identify the day of the week when the restaurants are heavily crowded.
- **Task 3:** We find the nearest restaurants, for a given location (latitude, longitude).

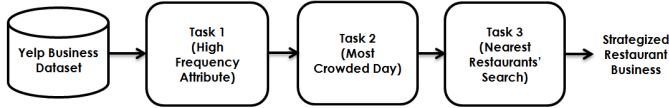


Fig. 3. Proposed framework for restaurant business setup analysis.

A. High Frequency Attribute

A restaurant has two types of attributes: Single valued attributes and Multi-valued attributes. Single valued attributes are those which have only one value, either *true* or *false* and multi-valued attributes are those which are further classified into single valued attributes. Single valued attributes are *delivery*, *dessert*, *lunch*, *late night*, *dinner*, *breakfast*, *reservations*, *credit card*, *Wi-Fi*, *wheelchair accessible* etc. which are either true or false. Multi-valued attribute includes ambience which is classified into key-value pairs like $\{romantic: true/false\}$, $\{intimate: true/false\}$, $\{classy: true/false\}$, $\{hipster: true/false\}$, $\{divey: true/false\}$, $\{touristry: true/false\}$, $\{trendy: true/false\}$, $\{upscale: true/false\}$ and $\{casual: true/false\}$.

Among these attributes, customers may not always require all of them. The goal is to identify those attributes which a customer is most interested in. This is achieved by inspecting the reviews and determining the high frequency attributes. This means that a maximum number of restaurants in the dataset have facilities with respect to high frequency attributes. This information helps to decide what facilities are most important and must be concentrated. It also helps to estimate how much capital should be invested on every attribute while setting up the restaurant business.

1) Why ambience is important and what ambience is most desired?: There are many factors that have a profound effect on the profitability of a restaurant. Some of these prime factors include the demand of the restaurant in the local area,

hospitality, quality of food, marketing strategies, management and much more. However, for long term benefits and to enhance the customers experience at the restaurant, one of the crucial factors that is frequently overlooked is the *ambience* of the restaurant [12]. According to studies, it is inferred that *ambience* has the power to improve the customer satisfaction and customer restitution. It also effects the time spent by the customers in the restaurant, the food they eat and the money they spend in the restaurant. Thus, we can say that, for a restaurant business to flourish, superior food, splendid service and an awe-inspiring ambience serve as the ingredients.

As a restaurant manager, if the atmosphere set is appropriate, the tranquility of the customers is enhanced and they enjoy being at the restaurant. When the customer satisfaction levels are high, their visit is high which leads to an increase in the profit. A wholesome and quality experience of the customers generates word of mouth marketing on behalf of the restaurant. Word of mouth is one of the most influential forms of marketing that has a considerable impact on the restaurant business [13]. For a restaurants long term profitability, word of mouth marketing undoubtedly has a favourable impact. Providing a complete and memorable dining experience might help the restaurant achieve an *ambience* enjoyed by the customers. To create or improve the restaurants ambience, we consider *romantic*, *intimate*, *classy*, *hipster*, *divey*, *touristry*, *trendy*, *upscale* and *casual* from the Yelp dataset since they help in achieving the perfect atmosphere which is a critical factor for long term profitability.

B. Most Crowded Day

In this task, we determine the most crowded day by exploring the reviews provided by the customers. We use business review counts with the assumption that people provide the review on the same day they visit the restaurant and determine the day on which the restaurant is most crowded. For the restaurant category, the review date is in the format of *dd/mm/yyyy* which is to be converted to the days of the week. The date from the reviews is extracted using Hue software [14]. Hue is a graphical user interface that is used to extract some or all the attributes from the Yelp dataset. Hue applications are delivered as a Web application by collecting them into a desktop-style environment. Once all the dates are obtained, the dates are converted into days. The total count for all the days is derived and the day with the maximum count is the day the restaurant is the most crowded.

C. Nearest Restaurant Search

Every available space is not the perfect space for setting up a new restaurant business. It is tough to find a suitable location to set up a restaurant than what it really appears to be. What emerges to be a flawless place - like a restaurant in the heart of downtown - may turn out to be dead duck. On the other hand, a small shoe shop in the run-down mill town may turn out to be huge success. Although, for a successful restaurant business, food and services are among the most important parameters, location can be just as crucial and can have a profound impact. From the Yelp dataset, the attributes *latitude* and *longitude* are considered to find the nearest restaurants.

In this task, nearest restaurant search is done to find the restaurants which are close to the business location. The

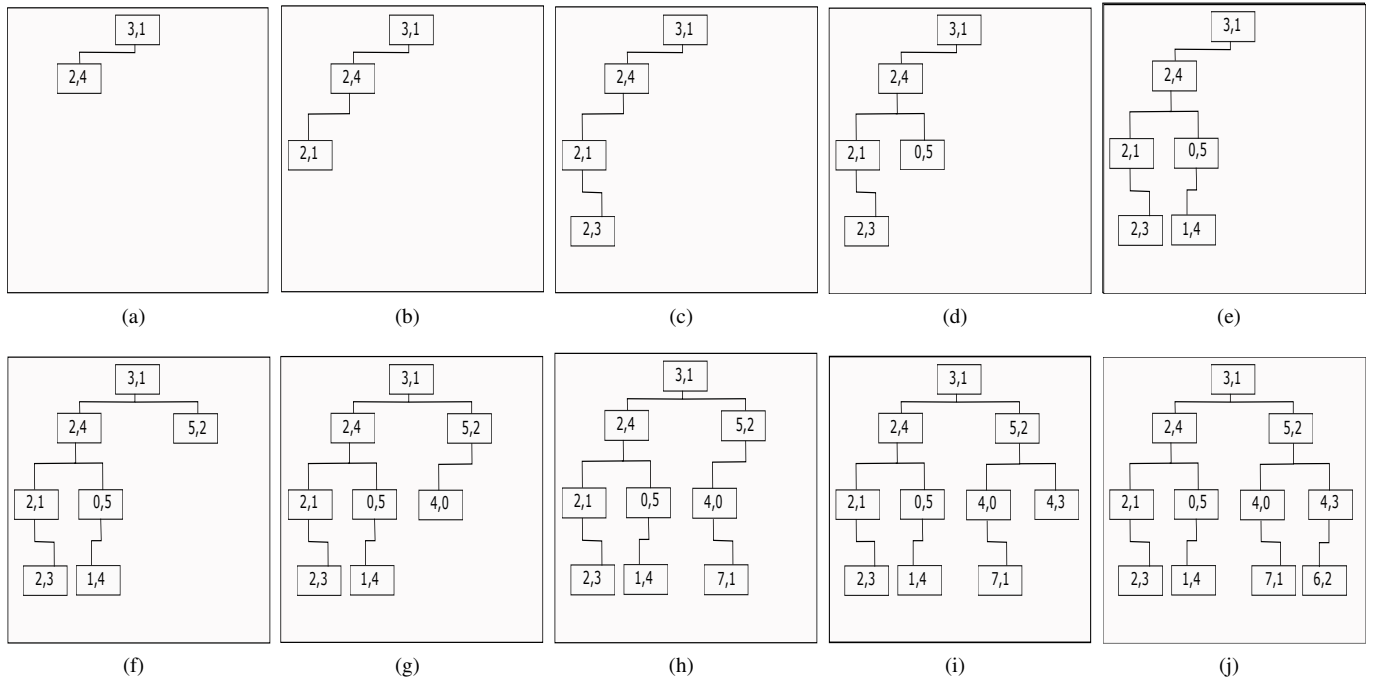


Fig. 4. (a-j) Steps involved in construction of kd-tree.

algorithm used to find the nearest restaurants is kd-tree algorithm [5]. It is known to be one of the most efficient algorithms for nearest neighbour search. The algorithm for building a kd-tree is a divide and conquer algorithm. The construction of kd-tree is similar to binary search tree (BST), where in the points which are less than the root node are stored in the left sub tree and the points greater than the root node are stored in the right sub tree. The main difference between the two is in kd-tree we split the data set of points by comparing with the k-dimension values, depending on the current depth in the tree [15]. The time complexity of the kd-tree algorithm is $O(n \log(n))$.

1) kd-tree Construction: To construct a kd-tree, a binary search tree is constructed by recursively bisecting the database of points. Each bisection is performed at the median of all data points in a chosen dimension [16].

Consider the set of points:

$\{(0, 5), (1, 4), (2, 1), (2, 3), (2, 4), (3, 1), (4, 3), (6, 1), (5, 2), (4, 0), (7, 1)\}$

Steps involved in construction of kd-tree:

- **Step 1:** The given points are sorted based on x -values and then split into two lists based on median x -value. The sorted list is $(0, 5), (1, 4), (2, 1), (2, 3), (2, 4), (3, 1), (4, 3), (4, 0), (5, 2), (6, 2), (7, 1)$. In this case the median is 3 i.e. $(3, 1)$. This will be the root node of kd-tree as shown in Figure 4(a). The two sublists are $(0, 5), (1, 4), (2, 1), (2, 3), (2, 4)$ and $(4, 3), (4, 0), (5, 2), (6, 2), (7, 1)$. The first list will be left subtree and second list will be right subtree.
- **Step 2:** The next step of the algorithm is to sort the sublists by y -values. The sorted left sublist (which contains all coordinates with $x < 3$) is $(2, 1), (2, 3),$

$(2, 4), (1, 4), (0, 5)$. The list is then split into two lists based on the median y -value. For the left sublist, the median value is 4 i.e. $(2, 4)$ which is inserted as left child of $(3, 1)$ as shown in Figure 4(a).

- **Step 3:** After splitting the list based on y -values, the sublists are sorted by x -values and the cycle repeats until we get a single point. Now, the two lists are $(2, 1), (2, 3)$ and $(0, 5), (1, 4)$. The median of the first list is $(2, 1)$ and the median of the second list $(0, 5)$ as shown in Figure 4(d). The median points $(2, 1)$ and $(0, 5)$ are made left and right child of $(2, 4)$ respectively.
- **Step 4:** After splitting the list based on x -values, the single point obtained is $(2, 3)$ from left-list and $(1, 4)$ from right-list. The y -value of point $(2, 3)$ is compared with the y -value of $(2, 1)$ to know the position of the point. Since $3 > 1$, the point $(2, 3)$ is made as the right child of $(2, 1)$ as shown in Figure 4(c). Similarly, $(1, 4)$ is inserted to left of $(0, 5)$ as shown in Figure 4(e).
- **Step 5:** Right sub tree of root node $(3, 1)$ is constructed in a similar manner by the process of splitting the list by x and y median values as shown in Figure 4(f)- 4(i). The final kd-tree obtained is as shown in Figure 4(j).

The next step is to search a node (location) from the constructed tree. Given a search point, we perform backtracking to find the n nearest neighbours.

Considering the *latitude* and *longitude* of business location from the Yelp dataset as the query (search) point (x, y) , the kd-tree algorithm is used to find the 10 nearest restaurants. Let (x_i, y_i) where $i = 1, \dots, 10$ be the latitude and longitude of the 10 nearest restaurants obtained using kd-tree algorithm. Euclidian distance is computed between the search point and

the 10 nearest neighbours to obtain the closest to farthest restaurants, given as,

$$D = \sqrt{(x - x_i)^2 + (y - y_i)^2} \quad (1)$$

This information is beneficial for a new entrant in the restaurant business to be aware of the facilities provided by various other restaurants in the locality and provide the most appealing facilities to satisfy the customers.

IV. RESULTS AND DISCUSSION

The proposed framework for restaurant business setup is tested with Yelp academic business dataset. The Yelp dataset consists of 1,125,458 reviews, information about 42,153 businesses, 320,002 business attributes and 31,617 check-in sets. Yelp academic business dataset includes features shown in Figure 5. In this paper, to address the problem of strategizing the restaurant business, we consider features like *latitude*, *longitude*, *attributes* and *review_count*, *attributes* and *categories*.

```
{
  "business_id": "encrypted business id",
  "name": "business name",
  "neighborhood": "hood name",
  "address": "full address",
  "city": "city",
  "state": "state -- if applicable --",
  "postal code": "postal code",
  "latitude": latitude,
  "longitude": longitude,
  "stars": star rating, rounded to half-stars,
  "review_count": number of reviews,
  "is_open": 0/1 (closed/open),
  "attributes": ["an array of strings: each array element is an attribute"],
  "categories": ["an array of strings of business categories"],
  "hours": ["an array of strings of business hours"],
  "type": "business"
}
```

Fig. 5. Features of the Yelp academic business dataset.

A. High Frequency Attribute

The most desired attribute in the hotel business data is found to be *credit card* as shown in Figure 6 which implies that majority of the restaurants in the Yelp dataset provide credit card facility. The possible reason for credit card to be the most desired attribute is, it is safer to carry credit cards rather than cash. Also, with the use of credit cards, a credit rating can be built which will be useful to apply for a loan. Attractive cash back incentives are offered by most of the credit cards every time the card is used. This can be as high as 10 percent of the total expenditure via the card. Various restaurants, hospitals, shopping malls, educational institutes, car-rental companies have agreements with credit card companies which benefits the customers to a great extent. The credit cards offering discounts are fascinating most of the time and can turn out to be a huge cash saver. These could be some of the reasons why most of the restaurants offer credit card facility and customers opt this facility. The second and third desired attributes are *wheelchair* and *waiter service*.

B. Most Crowded Day

We found that the most crowded day of the week is *Monday* as shown in Figure 7. This implies that the restaurant can

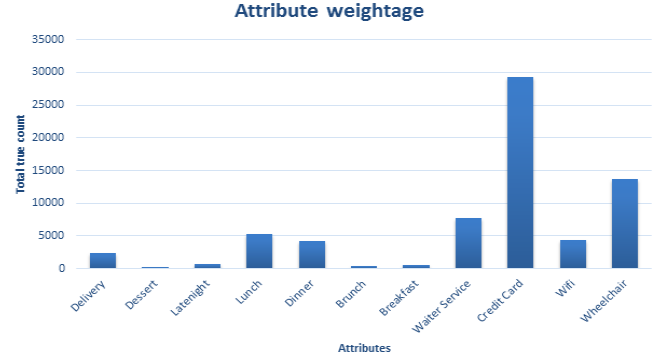


Fig. 6. High frequency attribute weightage derived from Yelp dataset.

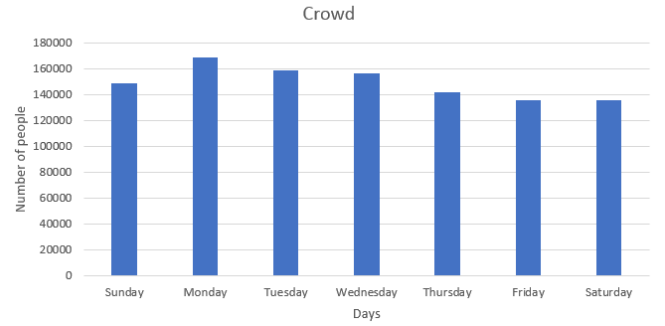


Fig. 7. Crowd distribution over the week derived from Yelp dataset.

present its best recipes on Monday. From the results obtained, it can be inferred that people from Phoenix, Waterloo, Las Vegas, Madison and Edinburgh prefer visiting restaurants after the weekend probably because weekends are spent staying home. Most people in these regions might work on Saturday and Sunday, due to which the holiday is suppose to be observed on Monday. Apart from the crowded day, *ambience* can also be considered as an important factor. Since ambience has a profound impact on customer satisfaction, special themes can be offered. The ambience that are taken into account are *romantic*, *intimate*, *classy*, *hipster*, *divey*, *touristry*, *trendy* and *upscale*. Of the seven listed ambience in the dataset, *Divey* ambience is found to be the most desired/most popular as shown in Figure 8. The restaurant managers can also think of some special offers that can be made available to attract more customers and increase the productivity of the business. The second and third crowded days are *Tuesday* and *Wednesday*. This implies most of the people in these localities might be working on weekends and taking off during weekdays. Also the second and third preferred ambience are *trendy* and *classy*.

C. Nearest Restaurant Search

In setting up a restaurant business, the highest frequency attribute, the information regarding the most crowded day and the ambience may not be sufficient. However, to setup a strategized business, location also plays a major role. Thus, to find the best location for a new restaurant setup, we strategize to obtain the 10 nearest restaurants from the search point (location) using kd-tree algorithm and the facilities provided by

TABLE I. NEAREST RESTAURANTS FOR SETTING-UP OF A NEW RESTAURANT AT LOCATION (43.05, 89.45) IN THE MADISON CITY.

Order	Restaurant Name	Latitude	Longitude	Distance	Wi-Fi	Accepts Credit Card	Parking lot	Wheelchair Accessible	Ambience	Category
01	EVP Coffee, S. Midvale Blvd	43.0543307	-89.450884	0.0044	Free	True	True	True	-	Food, Coffee and Tea
02	Chocolate Shoppe Ice Cream Co.	43.0546236	-89.4508969	0.0047	Free	True	True	True	-	Food, Desserts, Ice Cream and Frozen Yogurt
03	Luigi's Pizzeria	43.0548519	-89.450401	0.0048	No	True	True	True	Casual	Italian, Pizza
04	Yen Ching Restaurant	43.0380542	-89.4515708	0.0120	-	True	True	True	-	Chinese
05	Taqueria El Jalape	43.038073	-89.45202	0.0121	-	True	False	-	-	Mexican
06	Le Tigre Lounge	43.0374615	-89.4522307	0.0127	-	False	True	True	Divey	Dive Bars, Jazz and Blues, Nightlife and Bars
07	Coppertop Restaurant	43.036934	-89.4571154	0.0134	No	True	True	True	Casual	American, Diners
08	Feiler's Food and Cocktails	43.0364531	-89.4542172	0.0142	-	True	False	-	-	Seafood, American
09	Gates and Brovi	43.052916	-89.435789	0.0145	-	True	False	-	Casual	American
10	EVP Coffee, Mineral Point Rd	43.0605106	-89.4391741	0.0151	Free	True	False	-	-	Food, Coffee and Tea

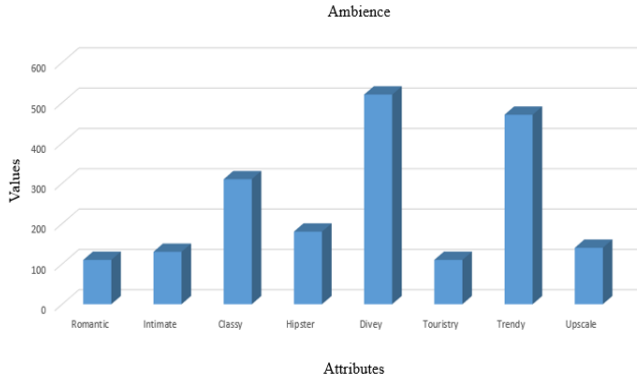


Fig. 8. Different ambiances and their preferences derived from Yelp dataset.

those restaurants. A new entrant to the restaurant business can consider this to understand how the location of the restaurant can affect the business. This information can be used to analyze how the restaurant can be made different in comparison with those in its proximity. For instance, if none of the restaurants nearby have the parking facility, then the upcoming restaurant can provide this facility. Since the restaurant concentrates on facilities that are considered important and are not provided by those nearby, it is more likely that the business flourishes. The process of the restaurant setup can be iterated until the best location with facilities and services maximizing the business and profit are obtained.

For example, consider the search point with *latitude* = 43.05 and *longitude* = 89.45 for setting-up a new restaurant business. Using kd-tree algorithm, the 10 nearest restaurants obtained surrounding the search point is shown in Figure 9. The facilities and other service details of the 10 nearest restaurants are shown in Table I. The analysis from Table I can be used to setup a new restaurant business at search point (43.05, 89.45). The most preferred attributes like *credit card*, *parking*, *wheelchair* and *Wi-Fi* which are provided in 10 neighbouring restaurants can be adopted in the new restaurant location. It can be observed that not all the 10 nearest restaurants are providing the expected ambience like *Divey*, *Trendy* and *Classy* which might be an influential factor to attract the customers and increase the profit in the new restaurant location. Also, the restaurant can be kept open on major days like Monday, Tuesday and Wednesday. In addition, *type of cuisine*

also plays a significant role in setting up a new restaurant. If most of the nearby restaurants follow a common type of cuisine, then it can be adopted along with the signature cuisine of the new restaurant. From the analysis of the Table I, it can be observed that *American* cuisine is adopted by 3 nearby restaurants which can be considered in setting up the new restaurant for maximizing the business and profit.

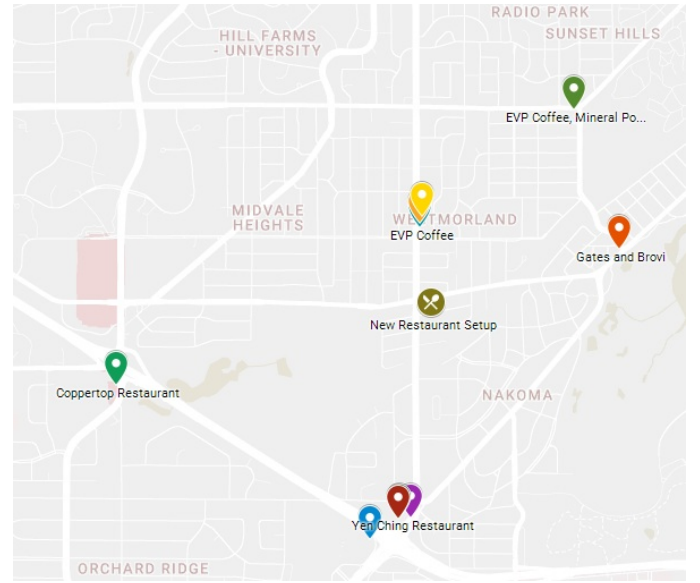


Fig. 9. Google Map view of the new restaurant setup in Madison city and the 10 nearest restaurants.

V. CONCLUSIONS

In this paper, we have addressed the problem of setting-up a restaurant business by proposing a restaurant business framework. From the experimental analyses, we observed *credit card* to be the most preferred attribute and Monday to be the most crowded day of the week. We also found that *Divey* to be the most desired ambience. Also, for a given location (43.05, 89.45), we found the 10 nearest restaurants to understand how the location can affect the business and to provide the services and facilities that are most desired. This framework can be effectively used to boost the performance of the restaurant which leads into maximizing the profit and business. As a future work, more analysis can be done to identify the best location for different search points.

REFERENCES

- [1] “National restaurant association,” http://www.restaurant.org/Downloads/PDFs/News-Research/2017_Restaurant_outlook_summary-FINAL.pdf, 2017.
- [2] H. Parsa, A. Gregory, and M. Terry, “Why do restaurants fail? part iii: An analysis of macro and micro factors,” *Emerging Aspects Redefining Tourism and Hospitality*, vol. 1, no. 1, pp. 16–25, 2010.
- [3] “Article: Find out what your patrons think 7 restaurant review sites owners can monitor,” <https://restaurantengine.com/restaurant-review-sites/>, 2016.
- [4] “Yelp dataset challenge,” https://www.yelp.com/dataset_challenge, 2014.
- [5] B. C. Ooi, K. J. McDonell, and R. Sacks-Davis, “Spatial kd-tree: An indexing mechanism for spatial databases,” in *IEEE COMPSAC*, vol. 87, 1987, p. 85.
- [6] M. Fan and M. Khademi, “Predicting a business star in yelp from its reviews text alone,” *arXiv preprint arXiv:1401.0864*, 2014.
- [7] T. Zhang and Y. Pan, “Yelp challenge project report,” 2014.
- [8] M. Luca, “Reviews, reputation, and revenue: The case of yelp.com,” *Com. Harvard Business School NOM Unit Working Paper*, no. 12-016, 2011.
- [9] Q. Jin, “A research proposal: The effects of restaurant environment on consumer behavior,” in *MBA Student Scholarship*. 36., 2015.
- [10] W. O. Mengqi Yu, Meng Xue, “Restaurants review star prediction for yelp dataset,” 2015.
- [11] E. J. James Huang, Stephanie Rogers, “Improving restaurants by extracting subtopics from yelp reviews,” 2013.
- [12] J. A. Fitzsimmons and G. B. Maurer, “A walk-through audit to improve restaurant performance,” *The Cornell Hotel and Restaurant Administration Quarterly*, vol. 31, no. 4, pp. 94–99, 1991.
- [13] D. Ribeiro Soriano, “Customers’ expectations factors in restaurants: The situation in spain,” *International Journal of Quality & Reliability Management*, vol. 19, no. 8/9, pp. 1055–1067, 2002.
- [14] “Yelp data analysis with hue,” <http://cloudera.github.com/hue>, 2013.
- [15] N. Sample, M. Haines, M. Arnold, and T. Purcell, “Optimizing search strategies in kd trees,” in *Fifth WSES/IEEE World Multiconference on Circuits, Systems, Communications & Computers (CSCC 2001)*, 2001.
- [16] R. F. Tobler, “The rkd-tree: An improved kd-tree for fast n-closest point queries in large point sets,” in *Proceedings of Computer Graphics International*, 2011.