National College of Ireland

Project Submission Sheet – 2020/2021

Student Name:	Dhwani Dharmesh Hingu , Dhanshree Bauskar, Anurag Ratnaparkhe ………………………………………………………………………………………………
Student ID:	19216742, 19230460, 19229992 ………………………………………………………………………………………………
Programme:	…………MSc. Data Analytics……… **Year:** …Jan21-22……
Module:	…………Domain Applications of Predictive Analysis……………………
Lecturer:	……Mr. Vikas Sahni………………………………
Submission Due Date:	…………20-08-2021……………………………………………………
Project Title:	………Predictive Analysis of Bangalore Restaurants…..…
Word Count:	……………………………2964………………………………………

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the references section. Students are encouraged to use the Harvard Referencing Standard supplied by the Library. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action. Students may be required to undergo a viva (oral examination) if there is suspicion about the validity of their submitted work.

Signature:	Dhwani Dharmesh Hingu , Dhanshree Bauskar, Anurag Ratnaparkhe ………………………………………………………………………………………………………
Date:	…………………………20-08-2021………………………………………………

PLEASE READ THE FOLLOWING INSTRUCTIONS:

1. Please attach a completed copy of this sheet to each project (including multiple copies).
2. Projects should be submitted to your Programme Coordinator.
3. **You must ensure that you retain a HARD COPY of ALL projects**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. Please do not bind projects or place in covers unless specifically requested.
4. You must ensure that all projects are submitted to your Programme Coordinator on or before the required submission date. **Late submissions will incur penalties.**
5. All projects must be submitted and passed in order to successfully complete the year. **Any project/assignment not submitted will be marked as a fail.**

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Predictive Analysis of Bangalore Restaurants

Dhwani Dharmesh Hingu
Dept. of Computing
National College of Ireland
Dublin, Ireland
x19216742@student.ncirl.ie

Dhanshree Bauskar
Dept. of Computing
National College of Ireland
Dublin, Ireland
x19230460@student.ncirl.ie

Anurag Ratnaparkhe
Dept. of Computing
National College of Ireland
Dublin, Ireland
x19229992@student.ncirl.ie

Abstract—Today, various businesses are rising and involving in review sites like Zomato, Yelp, Amazon, on which people can write their reviews and rating about the particular place they visited. This rating is generally measured on the scale of 1 to 5. Nowadays, restaurant business is one of the profitable business, with assurance of food quality, ambiance, quality of service, the major part is to choose a location where the restaurant will grow. This paper intends to suggest the appropriate place for setting up a restaurant by using Zomato Bangalore dataset having 17 features. Investing in a restaurant is a successful business as it provides good returns on investments if the business plan is correctly executed. During Python analysis, we have performed data manipulation, predictive Analysis, word cloud, Quantitative Analysis and Qualitative Analysis keeping the Business perspective in mind. All the Findings are demonstrated using appropriate visualizations for better understanding.Along with the programming, interactive dashboards were also created which allows stakeholders to make better business decisions.

Index Terms—Random Forest, Predictive Analysis, Qualitative assessment, Quantitative assessment

I. INTRODUCTION

Restaurants have always been an important part of business, social and cultural life. In India restaurant industry is growing at the rate of 7% annually. On an average, Indians eat lesser than 2 times a month in comparison with 40 times in Singapore. With the small rise in this number can make huge market potential for restaurants in India [1]. In this paper, impact of the location of any restaurant on success of restaurant a solution is proposed to help the entrants to make a decision of the perfect location to open a new restaurant. The analysis is performed on the dataset of Zomato Bangalore as Bangalore is believed as the IT hub of India. With the increase in IT organisations, working professionals from different city or state are migrating towards technologically advanced cities like Bangalore and so the demand for restaurants with good quality food is also escalating.

The objective of this study is to :

- Predicting best locations for opening a restaurant business.
- In depth analysis of historic data and predicting futuristic trends.

Studies have shown that to expand a business ideal location is very important. In this paper, restaurant business has been chosen and an analysis has been performed as to how location can affect the success of a restaurant. Based on the reviews of customers on existing eateries, the prime location will be predicted.

II. HYPOTHESIS

Null Hypothesis: "There is a correlation between the success of a restaurant and its location."

Alternate Hypothesis: "There is No correlation between the success of a restaurant and its location."

III. LITERATURE REVIEW

A competitive research [2] was found on the importance of a restaurant's location for its success using YELP! Dataset. The author of this paper experimented by parameterizing the location in order to check whether the locations have impact on the success of any restaurants. It was observed that the author got 0.81 correlation ratio. However this paper was motivation for this project as it was helpful to understand that locations have a huge impact on the success of the restaurants and parameterizing will give good results.

There are various sources where people can read reviews regarding restaurants and choose whether where to go or what to eat from the various alternatives available. However, owing to the large amount of evaluations, it is nearly hard for consumers to read all of them and get the information they need and on the other hand from business perspective the author of this paper [3] addresses that sometime the reviews regarding the restaurants can be biased and cannot help to predict the success of the restaurant. So Linear Regression approach in Machine Learning have been used along with three feature generation approach in order to find best prediction results. They have created bag of words of top frequent words on user generated user ratings. This paper was helpful to understand that solely user rating of restaurants cannot help to predict success of restaurant for the new entrant in the restaurant business market, however but it can be useful to understand what user's want from the particular restaurant based the positive and negative comments.

The research by [4] on "Predicting Restaurant Rating using Machine Learning and comparison of Regression Models" demand the need for predicting restaurant rating. Author have used machine learning approach, the model is constructed using a variety of regression techniques, with the most efficient algorithm being chosen. This paper's findings assist new restaurants in determining their menu, cuisine, theme,

cost, demographic location, and other factors, resulting in increased sales. However, by author's research it was helpful to comprehend that regression techniques can be used for prediction.

The author of [5] proposed an approach to suggest suitable location for new restaurant launch using yelp dataset. They have used support vector machine, decision tree and linear regression for predicting locations. To summarize, author demonstrated that utilizing 75 factors and the SVM algorithm on data obtained from YELP, they were able to predict business's rating also the they avoided using a set of complicated variable combinations in order to make the model basic and accurate. So, based on author's comparison research, they discovered that SVM is the most accurate for their yelp dataset. Therefore, this research was helpful to apprehend that SVM gave promising results so it can be used.

IV. METHODOLOGY

The CRISP-DM methodology was employed for the implementation of the predictive analysis along with some minor changes in the methodology to better cater the needs of the study, the justification for that is CRISP-DM illustrates and employs the use of Business Understanding phase as opposed to KDD methodology which majorly focuses on technical aspect of the project.
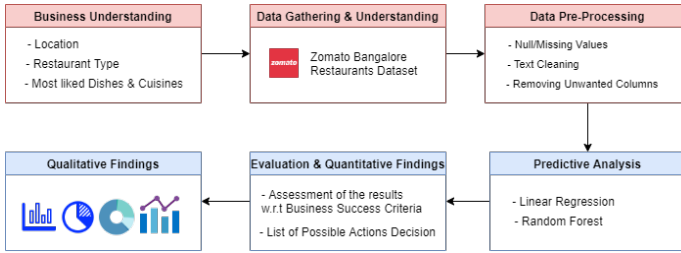

Fig. 1. Project Architecture

Figure: 1 illustrates the Methodology employed for this predictive analysis.

A. Business Understanding

The Food Industry is one of the oldest and largest Business market in the world, with new restaurants opening every day. In the cities like Bangalore where the competition of food restaurants is already very high, the implementation of the predictive analysis can highly benefit and help a new restaurant business to succeed. In this Predictive analysis study, we are predicting the best locations for a food restaurant to consider for opening the business, this is because locations can be one of the key factors for a business to grow as it directly affects the number of customers a business can expect which directly corelates to profit made by the business. For this we have analyzed the existing best restaurant in the area for the competition analysis along with dishes most liked, ratings and the approximate cost for two people.

B. Data Understanding

The dataset extracted from kaggle contains not only the the name and location of a particular restaurant but also how customers are reacting to it i.e. its reviews and ratings. A study conducted by [3] revealed the importance of rating and reviews of the customer on the restaurant which directly affects its growth, but the dataset used had very few parameters as opposed to the dataset used in this analyses which has a variety of parameters to accurately predict the best location for a new business. The dataset contains 23,310 rows of data which includes name and address of the restaurant, along with information like does the restaurant accepts online order or not, does they allow pre bookings of the dinner table, rating and votes by the customers, the location which the restaurant is based in, the type of the restaurant, the dishes liked by the customer in that particular restaurant, what cuisines do they serve, the approximate cost of two people in INR for a meal and the reviews posted by the customer.

C. Data Pre-processing

The raw dataset provided by zomato and kaggle was in a CSV format, and includes special characters, spaces in between words, trailing spaces etc. for the reviews field, and was cleaned in order to prepare it to feed to the predictive algorithm. As the reviews are mostly in big chunks of sentences in English, and thus cannot be interpreted by the predictive algorithm directly, the tokenization approach was used which refers to breaking down the big chunks of English sentences into smaller collection of key words, which can be fed to the algorithm for sentiment analysis. The term sentiment analysis refers to understanding the words used in a sentence and interpreting whether it is used in positive or negative context. The Dataset also contained some null and missing values which were processed. The columns like name and address of the restaurants were then dropped as it did not contribute to the predictive values. The other fields which were in text format were encoded so as to convert them to numeric format. The Dataset was then split in training and testing sets in the ratio of 70:30 respectively in order to perform validation and evaluation on the trained predictive model.

D. Modelling

After reviewing the related work done in the field it was observed that the most famous combination of predictive models used were the linear regression models so as to achieve a base line accuracy for the predictions and an advance model like random forest to achieve a benchmark accuracy for the predictive analysis. After splitting the dataset into training and testing sets , the models was trained on the training set, followed by passing the testing variables to predict the location on these independent variables.
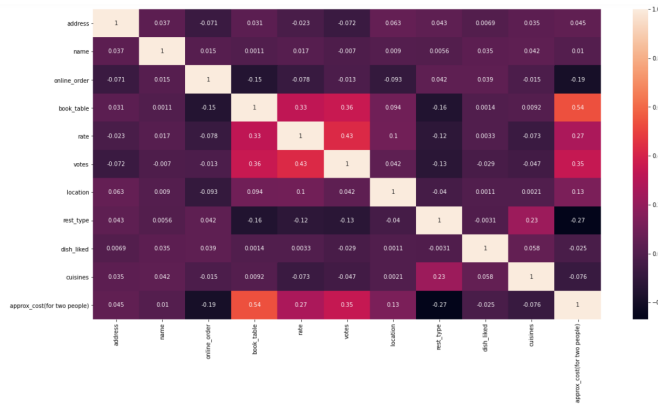
Fig. 2. Correlation Matrix

Fig. 4. Random Forest

A linear regression model uses a linear approach for establishing a relationship between a dependent variable also known as the label, and either one or multiple set of independent variables. The Figure 2 illustrates the correlation matrix between the independent variable in this case which is 'location' and multiple independent variables. The correlation matrix quantifies the range at which the different independent values affect the predicted value of the dependent variable i.e. location. The Figure 3 illustrates the graphical representation of predicted points and actual points (where blue lines demonstrates the predicted points and orange demonstrates the predicted points) by passing the values predicted by the model and the values in the actual test dataset. By analyzing the figure it can be inferred that there is still some room for improvement, which will allow the stakeholders to best evaluate the best locations for the new business.
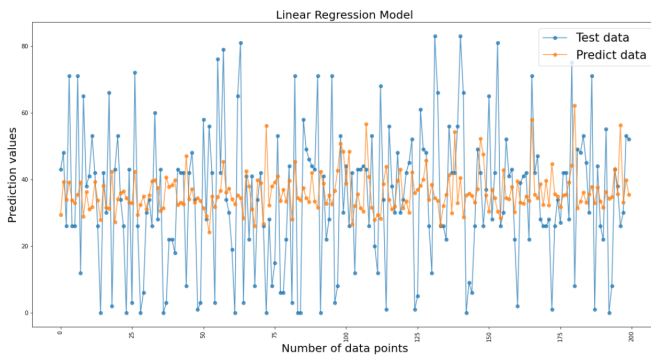

Fig. 3. Linear Regression

A Random Forest predictive model are type of an ensemble learning method and can predict both regression and classification values. The Random Forest methods are often used as "Black Box". models in all kinds of business to analyze and predict data as it generates reasonably higher accurate predictions and requires very less hyper parameters tuning.

The Figure 4 illustrates the graph where the values predicted by random forest model and the actual values from the test dataset are plotted(the blue lines demonstrates the actual points and orange demonstrates the predicted points) and from which
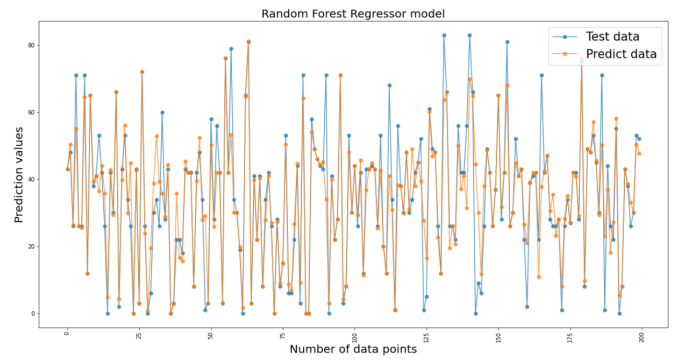
it can be easily inferred that random forest model provided very accurate and reliable predictions which then can be analyzed by the stakeholders to make an accurate business decisions.

E. Evaluation and Quantitative Analysis

After applying the Evaluation method, the accuracy which is the percentage of correct predictions made by the model against the actual values, of 80 percentage were calculated, which infers that the business stakeholders can reliably make the business decisions by analyzing the predictions made by the random forest predictive model.
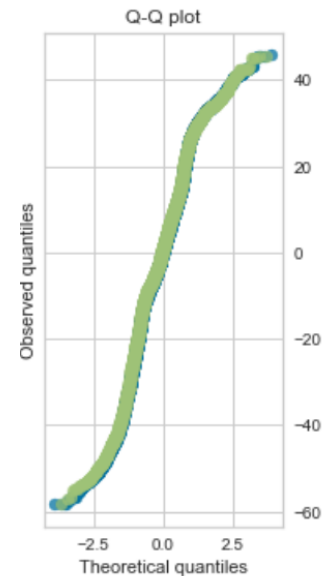

Fig. 5. QQ-Pllot

The Figure:5 illustrating the QQ-Plot (Quantile - Qunatile) plot, represents the distribution of two sets of data and creates a 45 degree line on the plot and the points falls on the reference line to clarify the distribution of two sets of data.

V. QUALITATIVE ANALYSIS

This Section covers the data presentation, where data is analyzed in order to get key business insights. The Past trends can be analyzed from the data presentation so as to get key insights in terms of business that what went well for the already established businesses and also future trends can also be inferred so as to inform the stakeholders to help make better business decisions.
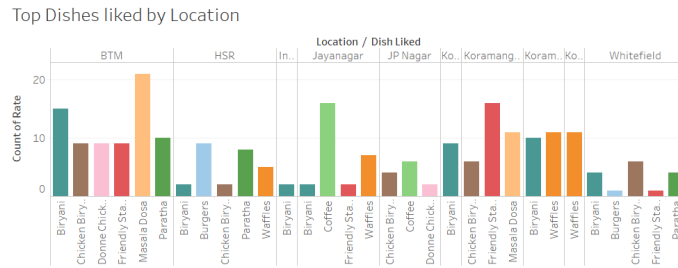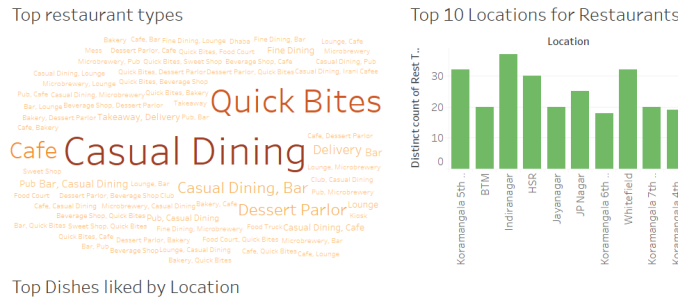

Fig. 6. Dashboard 1

The Figure: 6 created using tableau [6] demonstrates the best attributes in the restaurant business, the first sheet illustrates the most number of restaurants based on its type of food provided, for e.g. 'Casual dining' are the most abundant restaurants in city by far followed by 'Quick Bites' and cafes. This information gives the stakeholders what type of restaurant they should focus on, or what is most in demand. The second sheet demonstrates the top 10 locations for opening a new restaurant business, but these places also have the highest competition as well, 'Indira Nagar' and 'Kormangla' are one of the most sought after locations in the restaurant business, but also has the highest competition. The Third Sheet illustrates the most famous and frequently bought food dishes based on different locations, i.e. 'Masala Dosa' is most famous in BTM area and 'Burgers' and 'Parathas' are most famous in HSR area.
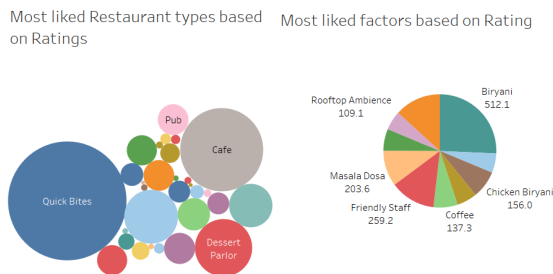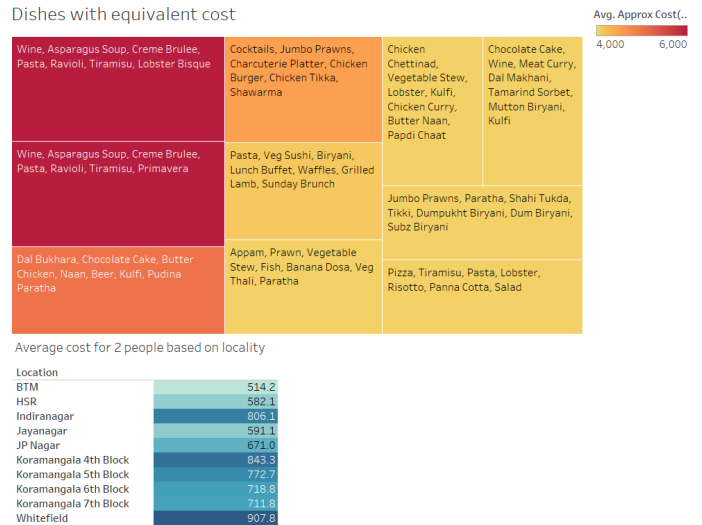

Fig. 7. Dashboard 2

Fig. 8. Dashboard 3

The Figure 7 illustrates the best attributes of the business based on the user reviews and ratings, like the first sheet illustrates the top restaurant types in the banglore city i.e. what type of food chains people of the city are liking and enjoying the most. The 'casual dining' type of businesses are the most liked, followed by 'Quick Bites' which is foods like sandwiches and snacks which is then trailed by cafes, Bars and dessert parlors. The second sheet illustrates the most liked aspects of a restaurant, for e.g. Biryani by far is most liked dish along with masala dosa and chicken Biryani, people are also liking the rooftop ambience in a restaurant along with a friendly and professional staff.

The Figure: 8 sheet 1 groups the cuisines which costs the approximately the same for two people, in which wine, Asparagus soup group are the most expensive and pizza, pasta etc are the cheapest food. The sheet 2 illustrates the approximate cost in INR for two people based on locality, which insights that 'whitefield', 'Kormangla 4th block' and 'Indira Nagar' are the top 3 most expensive locations to dine in the Bangalore city.

The Figure: 9 illustrates the number of current restaurant businesses accepting online orders and the number of restaurant which still does not accept online orders.

VI. KEY BUSINESS INSIGHTS

- With passage of time more and more restaurants are accepting online orders in order to increase the food order counts, and new restuarant should follow this trend so as not to lag behind the competetion.
- Casual Dining type restaurats are most abundant, Hoewever the Quick Bites are most liked by the customers, Hence demand of this kind of business is more and competetion is less.
- 'Indira Nagar' and 'Whitefield' are the most famous places to open a restaurant i.e. has the highest customer potential, However the competetion i.e. the number of

restaurants are also the most in these places, but Opening a less abundant type of restaurant like 'Quick Bites' in these locations could be benefecial.

- Along with good food, customers are also looking for and liking a friendly and professional staff, also attributes like rooftop dining and good ambience could attract more customers.
- Cuisines like coffee, quick bites and local foods like masala dosa are always in demand and also yield a good profitability.

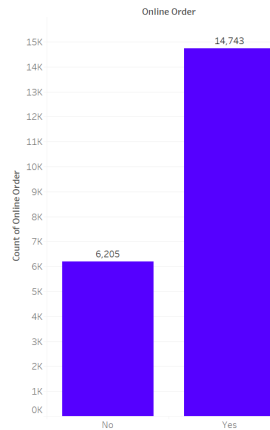Restaurant accepting online orders

Fig. 9. Counts of Online Orders

VII. CONCLUSION

From the above visualizations we can infer the top 10 best locations should an investor select while setting up the restaurants. With the help of visualization, we have seen the top 10 restaurants in Bangalore area had a closer look at their cuisines and for capturing the attention of people/customers investor should include some of the dishes which are liked the most by these top restaurants. With the help of reviews given by the customers for existing eateries, we have analyzed the most liked cuisine, and the type of restaurant. If we consider our analysis we can help set up a restaurant based on the amount to invest and which location to invest in and the investor can choose the most liked dishes as analysed in this study. Thus the Alternative Hypothesis that "There is No correlation between the success of a restaurant and its location", can be discarded and the Null hypothesis can be approved i.e. Yes, There is a corelation between the success of a restaurant and its location.

REFERENCES

[1] "Restaurant industry - SMERGERS industry watch," Smergers.com. [Online]. Available: https://www.smergers.com/industry-watch/indian-restaurant-industry/. [Accessed: 19-Aug-2021].

[2] A. S. M. Tayeen, A. Mtibaa, and S. Misra, "Location, location, location!: Quantifying the true impact of location on business reviews using a Yelp dataset," in Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2019, pp. 1081–1088.

[3] Text Alone, "Predicting a business' star in yelp from its reviews'," Arxiv.org. [Online]. Available: https://arxiv.org/ftp/arxiv/papers/1401/1401.0864.pdf. [Accessed: 20-Aug-2021].

[4] J. Priya, "Predicting Restaurant Rating using Machine Learning and comparison of Regression Models," in 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), 2020, pp. 1–5.

[5] I. F. Shihab, M. M. Oishi, S. Islam, K. Banik, and H. Arif, "A machine learning approach to suggest ideal geographical location for new restaurant establishment," in 2018 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), 2018, pp. 1–5.

[6] "Tableau: Business Intelligence and Analytics Software," Tableau.com. [Online]. Available: https://www.tableau.com/. [Accessed: 20-Aug-2021].