

# Analysing Success Factors and Predicting Ratings for Upcoming Restaurants

<sup>1</sup>Abhijith S., <sup>2</sup>Animesh Khare, <sup>3</sup>Aniruddha Krishna Jha

<sup>1, 2, 3</sup>Department of Computer Science Engineering

PES University, Bangalore, India

<sup>1</sup>[abhijiths.bang@gmail.com](mailto:abhijiths.bang@gmail.com), <sup>2</sup>[animesh1.khare1@gmail.com](mailto:animesh1.khare1@gmail.com), <sup>3</sup>[aniruddhakj100@gmail.com](mailto:aniruddhakj100@gmail.com)

*Abstract—Predictive models are often used by investors to decide whether a budding business would be profitable in their domain. The success of a new restaurant can be similarly predicted based on past data relating to the location and services offered. We develop a predictor model to find out the popularity and success a new restaurant can generate with an extensive study. This study aims to offer huge insights on which factors determine the success of a new restaurant.*

**Keywords—Restaurant, Factors of Success, Predictor**

## I. INTRODUCTION

Greek philosopher Heraclitus once famously said that “change is the only constant in life.” This statement holds true in regards to our lifestyle and eating habits as well. With the boom of delivery services and takeouts in India, businesses like Zomato, Ola Eats, Swiggy have popped off. E-commerce as a whole has become the backbone for a lot of industries and the restaurants and even small shops feast off of it. Bengaluru, being an IT hub in India, invites a lot of diverse cultures to its abode. Thus, it has become a key destination for foodies, with all the variety available. Everyday new restaurants and dhabas pop-up in the city, thereby matching the increasing demand with increasing supply. However, the established restaurant chains find it much easier than their newly established counterparts in terms of competition. There are a lot of factors that contribute to this, the location, the kind of style a restaurant goes with, the popularity associated with the chain, the initial ratings etc. In this study, we will be talking about a few of those factors.

## II. LITERATURE REVIEW

Prior studies relating to this domain involve determining factors that affect the rating of a restaurant, how restaurant types and location affects the profit margin from a business standpoint, what are the most popular and most favored restaurant chains in particular cities and building recommender systems for suggesting good and better restaurants adjusted according to a user’s preference. A lot of studies have been done in this domain and regard, but most of them actually aim to proceed towards some sort of recommender system for the customers. In this regard, we look at a few studies carried out in different regions and draw insights from them

### A. ARIZONA

A 2015 study by Sahoo & Saunil<sup>[1]</sup> looks into finding the ideal conditions and practices needed by up and coming restaurants to minimise competition and maximise initial gains. The purpose of this research was to determine the factors that affect the rating of a restaurant, to find out what the successful businesses in the restaurant industry do differently that leads to higher customer satisfaction, translating to higher ratings and revenues. The data available on Yelp for the state of Arizona, USA was utilised to gather insights on the factors that lead to positive customer sentiments as reflected by their reviews and to train a model that can predict whether a restaurant will be successful. primarily conducted 2 different analyses.

A multiple linear regression analysis was carried out to find the impact of different types of

independent variables on the rating, the dependent variable, with the assumption that the data had negligible multicollinearity, possessed multivariate normality and was homoscedastic.

Furthermore they performed a Time Series Analysis on text-based user reviews to see whether the general sentiment, which is identified by an R shiny app through the analysis of patterns in reviews, varied depending on the time of the year. The key findings of these two analyses were that Fine Dining restaurants with a relaxed ambience have the highest chance of being successful and that festive seasons garner lower sentiments due to the increase in prices.

## B. KOREA

Jeong-Hee, Myoung Soo, Kyu-Suk in 2018<sup>[2]</sup> have suggested and developed research hypotheses for restaurant industries in South Korea. These hypotheses aim to find out moderating effects of restaurant types and locations, expected outcome being that the research would help start-up restaurant managers get useful insights to better manage their businesses. A research model was built and evaluated on IBM's SPSS 18.0 through multiple regressions.

The key findings state that for Korean style restaurants, ops management and customer relations management were positively associated with financial performance. Also, financial performance was negatively affected by external environments which included government regulations, labor shortages, high taxes, credit card fees, and high initial investments, etc. Management of employees turns out to be a crucial factor for success.

Customer Relationship Management (CRM) is one of the main positive factors affecting business performance. In addition, ops management strongly affects business performance.

## C. BANGALORE & CHENNAI

A 2020 research<sup>[3]</sup> carried out in a similar domain but different context provided new intuitions and insights. The motive behind this research was to find out the most popular restaurants in major cities, Bangalore and Chennai in this case, determine the popular cuisines savoured by people and use data

science to analyze datasets and thus assisting people in finding out where they can find their preferred cuisine, and also providing them with the maximum number of choices possible to serve their purpose. Here the authors have made their analysis on Zomato, a popular online food delivery app. An effort is made by the authors to analyze Zomato's food delivery process.

Python and its packages were used to complete and implement their research model namely Numpy, Pandas, Matplotlib, Scikit-learn and Seaborn. They have also implemented various algorithms like Linear regression, Logistic regression, Decision tree regression and Random forest regression to enhance their research. For example, they have used Linear regression to predict the number of restaurants based on the city of choice and so on.

Some of the conclusions of their research suggested that the most popular cuisines of Bangalore were North Indian, Chinese and South Indian dishes. Then, the most popular restaurants in Chennai were found to be Shri Krishna Bhavan, Hotel Pandian, and Sultan Biryani. Also, more importantly it was found that the accuracy of Bangalore dataset was 84% whereas Chennai's dataset was approximately 99% and that orders placed online was 62% and offline order was 29%. Thus from this research, the most popular cuisines, restaurants in Bangalore & Chennai were inferred and also a key finding suggested that online orders are by far outnumbering offline orders, which suggests that online food orders are becoming more of a commonplace in our society.

## III. PROPOSED PROBLEM STATEMENT

We are more interested in finding out the factors correlating to the success of a restaurant based on prior customer reviews, cost, ratings, cuisine style and such. The problem statement we have come up with can be safely divided into three parts.

- Analysing the factors that impact the ratings of a restaurant
- Building a predictor model to find out the projected ratings for a new restaurant
- Test and refine the said model to find out and improve upon its accuracy.

#### IV. METHODOLOGY

One of the most common approaches when analyzing user reviews in the current world is by looking for patterns in the textual reviews that users post. However, it is our belief that those studies lead to biased results due to their assumption that users post both good and bad reviews with equal dedication and effort<sup>[4]</sup>. Just on general observation, it can be noticed that people tend to be more vocal about the negatives of any experience than about the positives on any reviewing platform. The exact opposite is observed however when those platforms are linked to, or are themselves, a social media based platform as people then tend to put an extra effort into painting a more positive picture of their lives. But again, these businesses might get subjected to review bombing, where even bots are created just to give out negative reviews. To ensure such a bias does not creep into our study, we will be focusing on facilities provided by the restaurants, their prices, their location and certain other variables that can be verified as facts to see their impact on the average numerical rating. We will also be more focused towards how the numbers and numerical aspects direct the ratings of restaurants or restaurant chains.

The dataset that we will be working on for this study and analysis can be found at kaggle on the given link below :  
<https://www.kaggle.com/absin7/zomato-bangalore-dataset>.

It contains data about all the restaurants in the city of Bengaluru, India that are listed on their app. The attributes in the dataset provide information about the location, the type of service offered, the cuisine, the cost and a number of other things along with the average rating, number of raters and the descriptive reviews for each restaurant. It also contains a few variables that are irrelevant to the study we intend to do, which will be dropped to enable a cleaner, easier access to the data and variables that we are concerned with. One of the issues we will need to address is that the number of users giving the ratings vary drastically among different restaurants.

The intuition is that we need to analyse the ratings of every restaurant and the reviews as well. However there are some restaurants which might have a lot of customer reviews while some might have less customer reviews even if the ratings are similar. For example, consider a restaurant A, which has a rating of 4.8 on a 5-point scale, but it has been rated by only 10 people. Another restaurant B is rated 4.5 but it has been rated by 200 people. Does that mean restaurant B is more popular despite being worse? To remove such ambiguity, we need to do some sort of scaling to remove a high bias. An intuition that can be offered here is to supply weighted ratings so that all the restaurants stand at equal ground. This idea gives more importance to the average rating of restaurants with higher numbers of votes. This ensures that we look at our data without giving preference to a certain sample.

The first phase of the approach involves carrying out the exploratory data analysis (EDA) on the dataset to process it for suitable use. After trimming the unnecessary attributes, the dataset would become suitable for our study purpose.

After obtaining inferences from the EDA and visualization, we intend to build and test a model that can predict how successful a new restaurant can be in a particular region, providing a particular set of facilities, to enable businesses to tweak their plans in order to improve their chances of success.

We also aim to provide budding businesses with a projected rating that will allow them to make an informed decision on whether their business model and their plans for the restaurant needs to be refined. We will train and test the models on splits of the dataset, followed by refining and selecting the one that turns out to be the most accurate.

#### V. REFERENCES

- [1] PP Sahoo, MS Saunil, R Chandra, L Xiang, "Analyzing success in the Restaurant Industry", *wiki.smu.edu.sg*, 2015
- [2] Jeong-Hee Hwang, Kyu-Suk Chung, Myoung-Soo Kim, "A Study on the Success Factors for the Restaurant Service Industry: Moderating Effects of Restaurant Types and Locations",

*Asia-Pacific Journal of Business*, Vol. 9, No. 3, pp. 11-24, Sept. 2018, doi: 10.32599/apjb.9.3.201809.11

[3] Dr. D. Singh, S. Srivastav, S. Shree, S. Shrivastava, S. Sharma, “Analysis of Online Food Delivery Process by Zomato using Data Science”, *International Journal for Modern Trends in Science and Technology*, 6(8S), pp. 125-127, 2020, doi: 10.46501/IJMTSTCIET24

[4] Amin Mahmoudi, “Identifying Biased Users in Online Social Networks to Enhance the Accuracy of Sentiment Analysis: A Behaviour-Based Approach”, arXiv, 2021, arXiv: 2105.05950