Indian Restaurants - Zomato

Divya Hebballi
Department of Computer Science and
Mathematics, CITY, University of
London, UK
Divya.Hebballi@city.ac.uk

Abstract— **The dining out behaviour of consumers in India has seen a marked transformation in recent years. The food services industry is one of the most promising sectors of investment. In this project we explore the interests of the Indian consumer in cuisines, popular dishes and their expectations from restaurants. We do this by understanding the factors affecting the prices of food in restaurants. We use decision tree and random forests to understand the various factors influencing the prices such as city, rating, cuisine and services provided by the restaurant. Our findings show that in general, people prefer dishes to be in a reasonable price range of 100 to 400 rupees. The famous tourist cities and the cities that are ranked among higher tiers tend to have higher food prices. We explore the popular cuisines served in restaurants with the consumer preferences and understand their impact on restaurant food prices.**

I. Introduction

The Indian food service market is one of the most vibrantly growing markets that has seen exceptional growth during the past decade. It has been forecasted to reach USD 95.75 billion by 2025 .The reasons for this are evolving and changing lifestyles, increase of youth population, disposable income, commuting and travelling, retail space and other factors which indicate that more Indians are choosing to eat out, thus creating vast opportunities [1].

The food services industry has become one of the most promising sectors of investment. Online platforms in India have had crucial role in building a great restaurant ecosystem. Zomato is one such foodtech platform which provides information about restaurants, menus, user-reviews along with home-delivery options and dine out choices. The restaurants take advantage of Zomato's intense local advertising platform to expand their delivery business, improve the efficiency of their table reservations and get connected to more customers.

The motive is to explore and understand the factors influencing consumer behaviour while eating out. According to Clark and Wood [2] and Pettijohn et al. [3] the quality of food and its worth are the most significant factors considered by customers while choosing the right restaurant. However the food choice model suggests that the price considerations of food is a more dominating attribute [4].This project aims to study the various factors influencing foodprices in Indian restaurants.

II. Analytical Questions and Data

The dataset represents information collected by Zomato from restaurants in the India. It contains 17 Columns and 224,520 Rows.

1) How does rating, location, cuisine & services provided by the restaurant affect the cost of two people dining out?

   Dataset characteristics: the dataset has attributes such as ratings, rating counts, addresses, location coordinates, cuisine and three service columns provided by the restaurant which makes it suitable choice for predicting the cost.

2) What are the top restaurant chains in India?

   The attribute containing names of the restaurants makes its convenient to check the most frequently occurring restaurant names to know the most widespread restaurant chains in the country.

3) How do online delivery and table reservation from Zomato affect the restaurants popularity among the consumers?

   The attributes such as online delivery, table reservation, rating and rating count make it easier to understand the popularity of restaurant among the consumers.

4) What are the cuisines and popular foods that restaurants serve, and how can the restaurants understand consumer interests in order to tailor their menus to earn more revenue?

   The attributes, such as cuisines, tell us the most frequent cuisine menu the Indian restaurants follow, and the attribute famous foods tell us the consumer preferences in these restaurants. There have been insights derived from the literature [1]. According to Tharavath, Gunasekar and Gupta, personality of individual, expectations, demographic factors such as age group, type of city of residence, annual income and occupation, give useful insights into the preferences of Indian consumers while dining out.

III. Analysis

A. Data-Preparation and derivation

1. Deleting redundant columns - dropped columns such as 'unnamed:0', 'sno', 'zomato_url', 'telephone', 'famous_food' which are not helpful in the modelling process .This was done based on available domain knowledge. The famous food column has been reused again for

the word cloud creation; it has not been used in the data-modelling as it would bring bias for having a lot of missing values (170,000) which was almost three quarters of dataset.

2. Dropping duplicate records.

3. Dropping the records with missing values as they were negligible in comparison to the size of the dataset.

4. Renaming the columns- renaming the columns such as ‘cost_for_two’ as ‘cost’.

5. Cleaning individual columns:

   a) Dropping the rows in rating columns which had the unique values such as ’NEW’ and ‘Nove’. Removing those rows that gave a better result in prediction of data compared to imputation of values.

   b) Removing commas, full stops in ‘Cost’ column and turning the column into a datatype float.

   c) Turning ‘name’ column to camel-case.

6. Label Encoding- Label encoding is done for all the categorical columns such as ‘Cuisines’, ‘City’, ‘Area’, ‘OnlineDelivery’, ‘Name’, Table_Reservation’, and ‘Delivery_Only’.

7. Heatmap is plotted to find the correlation between the attributes. ‘City’, ‘Area’, ‘Address’ and ‘coordinates’ are correlated so we will use only the ‘city’ column in our model to avoid multicollinearity. The columns ‘rating’ and ‘rating_count’ however remain the same.

   We have used Pearson’s coefficient to plot the heatmap.

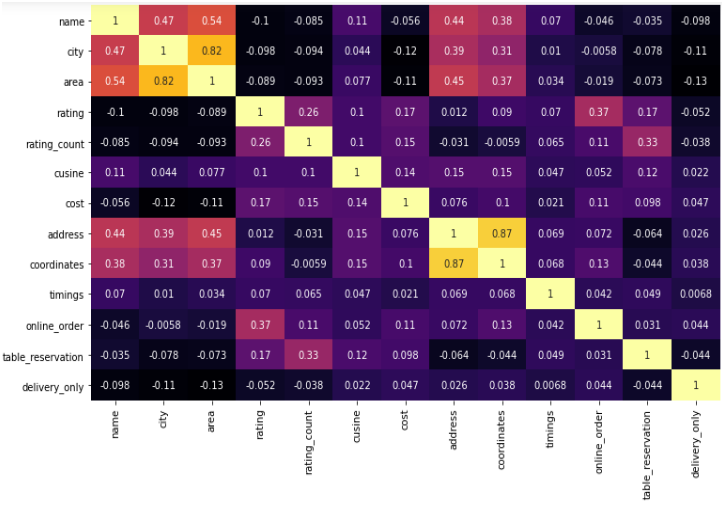

Fig. 1. Correlation matrix between all attributes

8. Plotting boxplots and eliminating the outliers for columns ‘rating count’ and ‘cost’ as we would be using it into our model.The columns such as name, timings and area are not used in the model to avoid Multicollinearity.

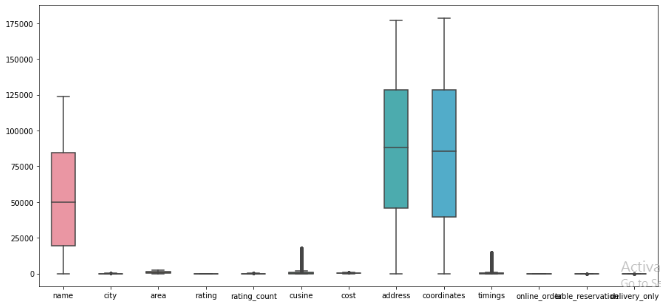

Fig. 2. Box plots for all the attributes

B. Construction of Model

1. The dataset was split into train and test with 30% of the data as test and 70% as the train.

2. The predictor columns were - City, Rating, Rating_Count, Cuisine, Online_Order, Table_Reservation and Delivery_Only.

3. Using variance inflation factor to check for Multicollinearity. However as none of the attributes had value more than five we did not make any changes.

4. Output column is the cost which is a continuous variable.

5. We use the models decision tree and random forests regressors and train them on our train dataset.

6. We predict the output on the test dataset.

C. Validation of Results

We are predicting the cost on the test set and checking the model efficiency using R-squared values.

TABLE I. PERFORMANCE OF DECISION TREES AND RANDOM FORESTS

Models	**Test data R-squared**	**Train data R-squared**
Decision Trees	0.9153	0.9347
Random forests	0.9356	0.9230

We can see that our R-squared value is above 90% for both the models. We can account for 90% of the changeability of the ‘cost’ attribute to be explained by both the model .The performance metrics are encouraging and it seems like they are good fitting models. Random forest consists of multiple single decision trees based on random sample of the training data and are typically more accurate than single decision tree.

IV. Findings, Reflections and Further Work

In understanding the cost of dining for two people in a restaurant, we tried to see the how the rating, location and services affect food prices. The chart below shows the relation between the rating and cost of food.

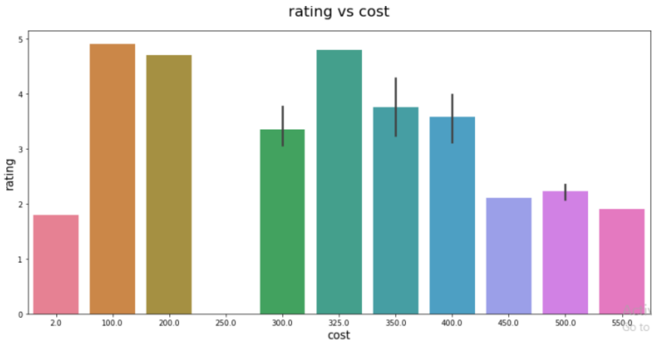

Fig. 3. Relation between Food Rating and Cost

The choice of food for the consumer is often influenced by the price of the food and it's perceived worth [4]. The chart above suggests that people in general prefer foods priced in the range of 100 to 400 rupees for a dish and tend to rate the dish according to their expectations and perceived worth of food.

The influence of location or the cities tend to have an impact on the cost. The below chart suggests that food tends to be more pricier in cities that are famous tourist destinations, which explains the higher cost of food in Manali, Goa, Gangtok, Rishikesh etc. The literature in [1] also suggests that the types of cities, i.e., Tier-1 like Mumbai, Pune tend to influence the cuisine choices and the cost of food.

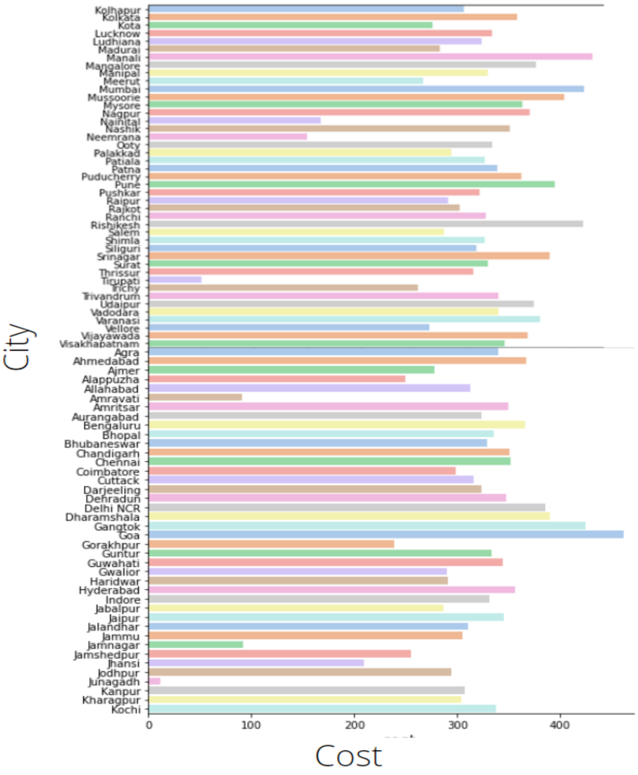

Fig. 4. Relation between City and Cost

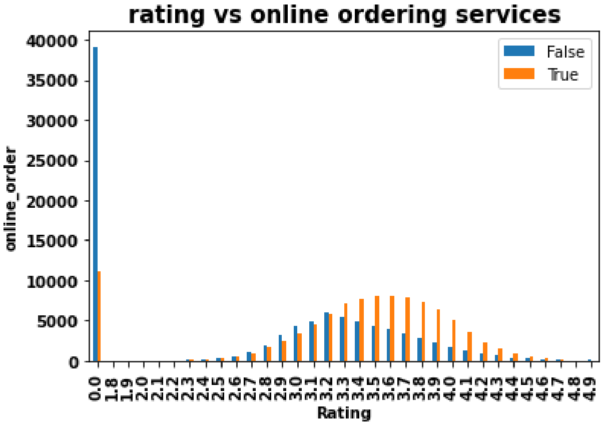

Fig. 5. Relation between Rating and online_order

The restaurants with either low or high ratings have not yet transitioned to online services, however the new restaurants with zero rating and the restaurants with the rating in the range of 2.9 to 4.2 have transitioned in order to expand their delivery business, and get connected to more customers.

Restaurants however have not used the features of table booking and delivery services as much as the online order feature in the Zomato app. It can help them connect to customers more easily or it may be the case that the feature has not really been useful to them and they do not use it.

The cuisines that were commonly observed in many restaurants were North-Indian, South-Indian, Chinese, Italian and Fast food, as illustrated below.

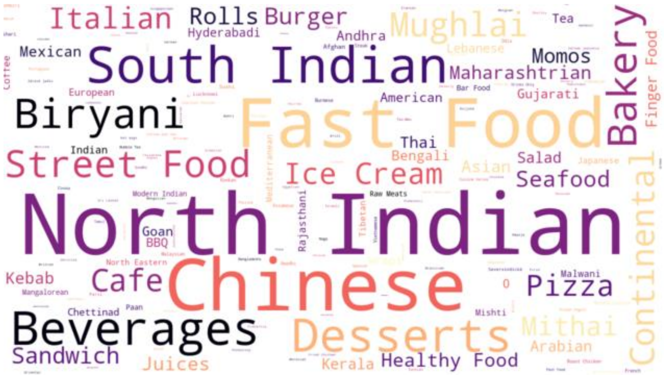

Fig. 6. Cuisines

According to [1], consumer expectations when dining out influences cuisine choice. Consumers who expect food variety prefer Indian, Chinese and Italian cuisines. Different personality traits influence people's cuisine choices. People who rate their personality trait as thrifty are less likely to prefer Indian cuisine. Consumers who prefer Indian cuisine are less price sensitive, placing greater importance on food quality [5].That probably is the reason why Indian cuisine is popularly served in restaurants.

The word-cloud (below) for popular food indicated that restaurant food favourites were pizza, pasta, burgers, noodles, butter chicken and biryani.

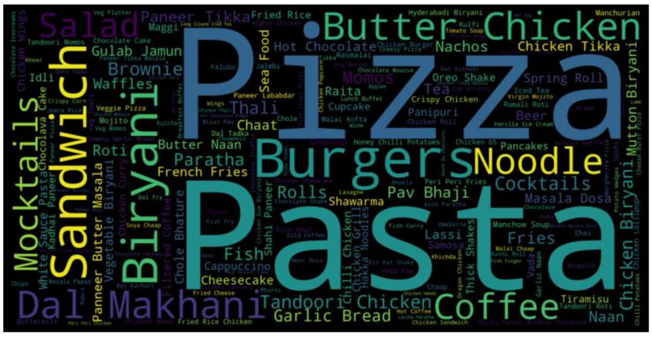

Fig. 7. Popular food served in restaurants

This may be due to the fact [1] that younger consumers prefer Italian cuisine like Pizza, Pasta etc. However, older consumers prefer Indian cuisine for dining out. These results show an interesting transition of food choices between different age groups in India. The insights derived help restaurants serving particular cuisines to tailor their menus and marketing activities. This result however is limited to a few restaurants as the data for almost half the dataset was missing, so having more records will give more accurate results.

Our dataset shows the top 50 restaurant chains in India as below.

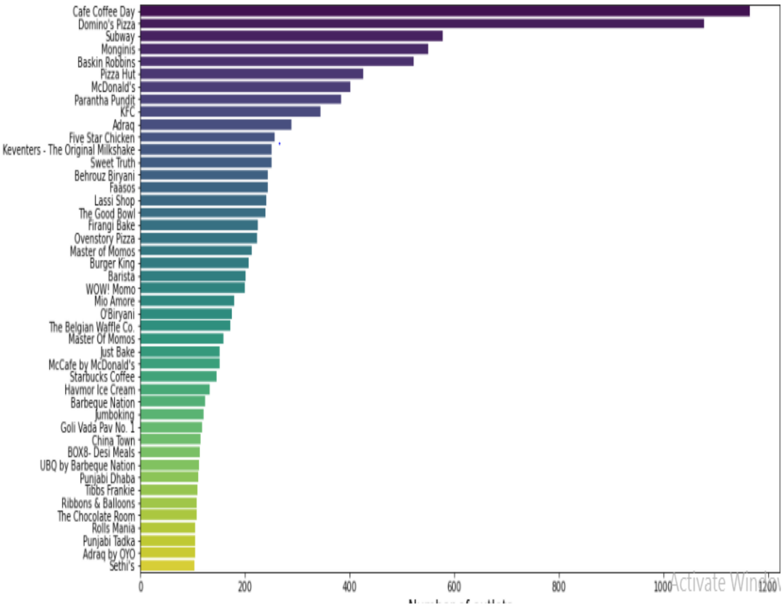

Fig.8. Top 50 restaurant chains in India

That might also explain the popularity of pizza as Domino's Pizza is ranked second in the top 50 restaurant chains in India.

The dataset chosen did have many observations and had covered many popular cities, although it does not really cover a lot of places in India, so this analysis is restricted to the cities whose records were available to us through the dataset. For future work it would be interesting to have more attributes and records covering the popular foods, demographic features, and personality traits in order to fully understand consumer preferences.

REFERENCES

[1] Vishnu Tharavath, Dr. Deepak Gupta and Dr. Sangeetha Gunasekar "Factors Influencing Choice of Cuisines While Indian Consumers Eat Out", *International Conference on Data Management, Analytics and Innovation* (ICDMAI) Zeal Education Society, Pune, India, Feb 24-26, 2017.

[2] Mona A. Clark, Roy C. Wood, (1998) "Consumer loyalty in the restaurant industry a preliminary exploration of the issues", *International Journal of Contemporary Hospitality Management*, Vol. 10 Iss: 4, pp.139 - 144R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[3] Pettijohn, L. S., Pettijohn, C. E., Luke, R. H. (1997). "An evaluation of fast food restaurant satisfaction" determinants, competitive comparisons and impact on future patronage".*Journal of Restaurant & Foodservice Marketing*, 2(3), 3-20. K. Elissa, "Title of paper if known," unpublished.

[4] T. Furst, M. Connors, C. A. Bisogni, J. Sobal, and L. W. Falk, "Food Choice : A Conceptual Model of the Process," *Appetite*, vol. 26, pp. 247–265, 1996.

[5] Li, Jian-rong and Yun-hwa P. Hsieh. 2004. "Traditional Chinese Food Technology and Cuisine." *Asia Pacific Journal of Clinical Nutrition* 13(January):147–55.

TABLE II. LIST OF WORD COUNTS

Section	Word Counts
Abstract	149
Introduction	224
Analytical Questions and Data	253
Analysis	500
Findings, Reflection and Future Work	577