

[NOTEBOOK](#), [PPT](#)

Name	Manish Thakur
Section	K21UW
Roll No	RK21UWB56
Reg No	12005905
Subject	INT 353
Submmitted to	Shivangini Gupta

Ques1: Domain Topic Knowledge.

Ans: I have selected the **Zomato dataset** from kaggle which is an online food industry that acts as a mediator between the restaurants and the people. Further, it is planning and also started working on the concepts of cloud kitchen. This concept can really bring a change in the online food industry.

There are few key **concepts and terminologies** are:

1. Restaurant Data: This dataset primarily contains information about various restaurants, including their names, locations, cuisines, pricing, and ratings.

2. Cuisine: The type of food a restaurant specializes in. It is an essential aspect of restaurant selection.

3. Table Booking: The ability for customers to reserve a table in advance, which is common in fine dining establishments.

4. Online Delivery: The option for customers to order food online for delivery to their location.

5. Rating and Reviews: Aggregate rating and customer reviews are crucial for assessing the quality of a restaurant.

6. Price Range: An indicator of the restaurant's affordability, with higher values indicating higher prices.

From a data analyst's perspective with a focus on the business side of the online food industry, there are several significant **Challenges** that companies like Zomato and similar platforms face:

1. Quality Assurance of Data: Ensuring the accuracy and reliability of data is crucial. Inaccurate restaurant information, incorrect menu items, or outdated prices can lead to customer dissatisfaction and loss of trust.

2. Data Privacy and Security: Handling customer data, including their preferences and order history, requires strict adherence to data privacy regulations. Any data breaches can result in severe legal and reputational consequences.

3. Competitive Landscape: The online food delivery industry is highly competitive, with numerous players vying for market share. Understanding market dynamics and staying ahead of competitors is a constant challenge.

4. Customer Retention: Acquiring new customers is essential, but retaining existing ones is equally important. Analyzing customer behavior and preferences can help in designing loyalty programs and personalized experiences.

5. Logistics and Delivery Optimization: Efficiently managing the logistics of food delivery, including delivery time, route optimization, and driver allocation, is a constant challenge. Delayed deliveries or incorrect orders can lead to customer dissatisfaction.

6. Restaurant Partnerships: Building and maintaining relationships with restaurants is crucial. Negotiating contracts, managing commission rates, and ensuring a diverse and high-quality restaurant portfolio are ongoing challenges.

7. Pricing Strategies: Determining the right pricing strategy that balances profitability with affordability for customers can be complex. Dynamic pricing and promotions require data-driven decision-making.

8. User Experience: Creating a seamless and user-friendly app or website interface is vital. Analyzing user behavior and feedback to continuously improve the user experience is an ongoing challenge.

9. Regulatory Compliance: Navigating the complex regulatory landscape, including food safety regulations and labor laws, can be a significant challenge for online food delivery platforms, especially when operating in multiple countries.

10. Sustainability: Increasing awareness of environmental concerns requires companies to address sustainability issues, such as packaging waste and carbon emissions from delivery. Implementing sustainable practices while maintaining profitability is a growing challenge.

11. Market Expansion: Scaling the business into new geographic regions involves understanding local preferences, regulations, and competition. It requires extensive market research and adaptation of strategies.

12. Customer Feedback and Reviews: Analyzing and acting upon customer feedback and reviews is crucial for quality control and improving the overall service. Handling negative feedback effectively is a challenge.

13. Supply Chain Disruptions: Events like natural disasters or pandemics can disrupt the supply chain, affecting the availability of ingredients and delivery operations. Developing contingency plans is essential.

14. Technology Upgrades: Keeping the platform up to date with the latest technology trends, such as AI and machine learning for recommendation systems and chatbots for customer support, is an ongoing challenge.

15. Profitability: Achieving profitability in the online food delivery industry can be challenging due to high competition, marketing costs, and operational expenses. Optimizing costs while maintaining service quality is a constant focus.

16. Crisis Management: Handling unexpected crises, such as food safety issues or public relations crises, requires a robust crisis management plan and the ability to analyze and respond to the situation promptly.

Addressing these challenges requires a data-driven approach, leveraging analytics to make informed decisions, improve operations, enhance customer experiences, and adapt to the dynamic nature of the online food delivery industry. Companies that can effectively navigate these challenges are better positioned for long-term success in this rapidly evolving market.

Basically, Zomato dataset offers valuable insights into the restaurant industry, including restaurant details, customer ratings, and location information. Analyzing this data can provide insights into consumer behavior, restaurant popularity, and market trends, while also posing challenges related to data quality and complex analysis techniques specific to the food service domain.

Ques2: Information about the dataset.

Ans: Here, I have done a concise explanation for all columns of the dataset for better understanding of the data.

1. Restaurant ID: A unique identifier for each restaurant in the dataset.
2. Restaurant Name: The name of the restaurant.

3. Country Code: A code representing the country where the restaurant is located.
4. City: The city where the restaurant is situated.
5. Address: The street address of the restaurant.
6. Locality: The specific locality or neighborhood where the restaurant is located.
7. Locality Verbose: A more detailed description of the restaurant's locality.
8. Longitude and Latitude: Geographic coordinates of the restaurant's location.
9. Cuisines: The type of cuisine the restaurant offers.
10. Average Cost for Two: The average cost for two people dining at the restaurant.
11. Currency: The currency used for pricing in that country.
12. Has Table Booking: Indicates whether the restaurant accepts table reservations (Yes/No).

13. Has Online Delivery: Indicates whether the restaurant offers online food delivery (Yes/No).

14. Is Delivering Now: Indicates whether the restaurant is currently delivering food (Yes/No).

15. Switch to Order Menu: A menu option for switching to the online ordering menu.

16. Price Range: A numeric value indicating the price range of the restaurant's menu.

17. Aggregate Rating: The overall rating of the restaurant based on customer reviews and feedback.

18. Rating Color: The color associated with the restaurant's rating, often used for visualization.

19. Rating Text: A text description of the restaurant's ratings, for example: Excellent, Good, Average, Poor.

20. Votes: The total number of votes or reviews received by the restaurant.

Ques3: Why you choose that dataset?

Ans: There is a special reason for choosing this dataset. As I've always been curious about this food industry and

how it works. Basically, like the revenue and their growth pace. Because according to me, expanding more as a food business is one of the most difficult things. And here, I finally get the chance to learn more about this industry and dig into their working and plans. Let's see how much I learn about this business after completing this project.

Ques4: Questions for Analysis.

Ans:

1. What is the total number of restaurants in the dataset?
2. Which countries are represented in the dataset, and how many restaurants are in each country?
3. How many unique cities are covered by the dataset?
4. What is the average cost for two people dining across all restaurants?
5. Which currency is the most commonly used in the dataset?
6. What are the top 5 cities with the most restaurants?
7. Can you identify the restaurant with the highest latitude and longitude values? Where is it located?
8. How many restaurants have "online delivery" available?
9. Which city has the highest number of restaurants offering online delivery?
10. What are the top 5 most popular cuisines in the dataset?

11. What is the distribution of restaurants based on their price range?
12. Are there any correlations between the price range and the average cost for two people?
13. What is the distribution of restaurant ratings (Aggregate Rating)?
14. Is there a correlation between the number of votes and the restaurant's rating?
15. How many restaurants have a rating of "Excellent"?
16. What is the most common rating color in the dataset?
17. How many restaurants accept table bookings?
18. Do restaurants that accept table bookings tend to have higher ratings?
19. What percentage of restaurants offer online delivery?
20. Is there a relationship between online delivery availability and a restaurant's rating?
21. Which locality has the highest number of restaurants?
22. Are there restaurants in the dataset that share the same address?
23. How many restaurants are located within a specific geographic radius of a given location?
24. What is the average cost for two people dining for each cuisine type?
25. Are certain cuisines associated with higher price ranges on average?

Hypothesis Testing

Now, to perform hypothesis testing we need to check our dataset. In our scenario, we aim to investigate whether there is a statistically significant difference in the average cost for two people dining between restaurants that offer online delivery and those that do not. The choice of a two-sample t-test is appropriate for comparing the means of two independent groups, providing insights into potential variations in dining costs.

The t-statistic measures the difference between the means of the two groups relative to the spread of data. A higher absolute t-statistic indicates a more substantial difference. The p-value, on the other hand, quantifies the probability of observing a t-statistic as extreme as the computed one, assuming the null hypothesis is true. A significance level, often set at 0.05, is then used to determine whether the observed difference is statistically significant.

Report Insights:

The analysis focuses on discerning whether there exists a significant discrepancy in the average cost for two people dining between restaurants offering online delivery and those that do not. By delving into this aspect, we aim to uncover potential patterns or disparities that could offer valuable insights for stakeholders. This investigation could shed light on the economic implications and consumer

preferences related to online delivery services in the restaurant industry.

Limitations:

1. Representativeness: The validity of our findings relies on the assumption that the dataset is representative of the broader population. Deviations from representativeness could impact the generalizability of our conclusions.

2. Unconsidered Factors: The dataset may not encompass all factors influencing dining costs. Variables such as location, restaurant size, or specific menu offerings might contribute to variations that are not captured in our analysis.

Conclusion:

The study found that there is a strong positive correlation between the price range and the average cost for two people. This means that as the price range increases, the average cost for two people also tends to increase. Upon performing the two-sample t-test and assessing the results, we can draw conclusions about the presence or absence of a significant difference in the average cost for two people dining between restaurants with and without online delivery services. If

the p-value is below the chosen significance level (commonly 0.05), we would reject the null hypothesis, signifying a substantial difference. Conversely, a p-value exceeding the significance level would lead to the retention of the null hypothesis, suggesting no significant difference. This is not surprising, as restaurants in higher price ranges typically offer more expensive menu items and a more upscale dining experience.

The study also found that restaurants that accept table bookings tend to have higher ratings than those that don't. This suggests that customers may perceive restaurants that accept table bookings as being more upscale and offering a better dining experience. Restaurant owners who want to improve their ratings may want to consider offering table bookings, especially if they are targeting a higher-end clientele.

Finally, the study found that there is no significant relationship between online delivery availability and a restaurant's rating. This suggests that customers do not necessarily value online delivery more than other factors such as food quality, service, and atmosphere when choosing a restaurant. Restaurant owners should focus on providing a great overall dining experience, regardless of whether or not they offer online delivery.

How to Improve the Study

The study could be improved in a number of ways. First, the researchers could use a larger and more diverse dataset of restaurants. This would make the results more generalizable to all restaurants. Second, the researchers could consider additional variables that may influence customer ratings, such as food quality, service, atmosphere, and location. Third, the researchers could use more sophisticated statistical methods to control for the effects of other variables.

Overall, the study provides some valuable insights into the relationship between price, table booking availability, online delivery availability, and customer ratings. Restaurant owners can use these insights to improve their operations and attract more customers.

THANK
YOU