Online Food Delivery Analysis in Bengaluru region

THESIS SUBMITTED TO

Symbiosis Institute of Geoinformatics

FOR PARTIAL FULFILLMENT OF M.SC DEGREE

By

Ashwin Gadge-22070243020

Juned Manyar-22070243023

Ajinkya Kadam-22070243024

Batch 2022-2024

Symbiosis Institute of Geoinformatics

Symbiosis International (Deemed University)

5th Floor, Atur Centre, Gokhale Cross Road,

Model Colony, 411016

Table of Contents

Abstract:

This report presents a comprehensive analysis of the online food delivery ecosystem in the dynamic city of Bangalore, where the burgeoning restaurant industry has adapted to the fast-paced urban lifestyle. Leveraging advanced data analysis techniques such as Random Forest for feature importance and geospatial analysis using ArcGIS, we delve into the intricate dynamics of this rapidly evolving landscape.

In this study, we investigate customer behaviour, including patterns of order placement, delivery preferences, and the crucial impact of location on service quality and customer satisfaction. By plotting customer location points on maps, we uncover insights into hotspot areas, clustering of restaurants, and the influence of proximity on customer choices.

The findings of this research are invaluable to both the online food delivery industry and the restaurant sector in Bangalore. As convenience and culinary experiences continue to evolve, this study offers a deep understanding of the shifts in consumer behaviour, technology integration, and the changing face of dining in the Silicon Valley of India.

Introduction:

In recent years, the landscape of the restaurant industry has undergone a remarkable transformation, particularly in bustling metropolises like Bangalore. The relentless pace of urban life, coupled with the ever-increasing demands of work and daily commitments, has left people with little time to engage in the traditional culinary pursuits of grocery shopping and home-cooked meals. As a result, the phenomenon of online food delivery has emerged as a lifeline for busy individuals and families, enabling them to savor their favourite restaurant dishes from the comfort of their homes.

Start-up's such as Swiggy, Zomato, Dunzo etc. are huge choice of medium for delivering food at their door steps. These start-ups become a medium to deliver food and restaurants becomes the place to prepare food demanded by the consumers.

In the vibrant city of Bangalore, the significance of online food delivery cannot be overstated. It has become an integral part of the urban fabric, with delivery captains crisscrossing the city at a frenetic pace to fulfil the gastronomic desires of the residents. These dedicated delivery captains play a pivotal role in the process, receiving order requests from customers and ensuring that the entire transaction, from the restaurant kitchen to the customer's doorstep, is seamlessly executed.

A crucial aspect of this dynamic ecosystem is the customer feedback mechanism. In the digital age, customers wield the power of rating and reviewing their entire dining experience. They rate the delivery service for its timeliness, accuracy, and professionalism, while also providing a separate rating for the food quality. These ratings, in turn, have a profound impact on the choices of other customers. A high rating for a restaurant's food quality can lead to a surge in demand for its offerings, while positive feedback on delivery service can create a loyal following for a particular online platform.

Understanding the intricacies of this evolving online food delivery landscape in Bangalore is of paramount importance. To this end, our study leverages cutting-edge data analysis techniques, including Random Forest for feature importance, and utilizes ArcGIS for geospatial analysis. We delve into the nuanced world of customer behaviour, examining patterns of order placement, delivery preferences, and the impact of location on service quality and customer satisfaction. By plotting customer location points on maps, we unveil insights into hotspot areas, clustering of restaurants, and how customer choices are influenced by proximity to their favourite eateries.

2.1 Objectives:

1. Collect latitude and longitude coordinates of consumers to map their locations within Bangalore.
2. Determine the factors influencing the adoption and continued use of online food delivery services in Bangalore.
3. Investigate patterns of customer order placement and preferences in terms of cuisine, restaurant choices, and delivery service providers.

4. Examine how proximity to restaurants and delivery hubs affects service quality, customer satisfaction, and the frequency of service use.
5. Utilize consumer behaviour data to build the model, helping online platforms improve customer retention strategies.

2.2 Problem Statement:

1. As the online food delivery industry flourishes in Bangalore, the primary challenge is aligning food quality and service with what customers expect to keep them consistently satisfied.
2. With increasing competition, the key task is not only attracting but also retaining loyal customers by understanding what motivates them to order repeatedly.
3. Efficiently planning where and how deliveries are made using geospatial insights is vital for operational efficiency.
4. Data analysis is crucial in making informed decisions, benefiting both food delivery providers and customers. The main issue revolves around using data to make smart choices in the online food delivery ecosystem.

Methodology:

The study employed a multi-faceted methodology to comprehensively analyse consumer behaviour in the online food delivery industry in Bangalore, India. Data was collected through a survey encompassing 55 variables categorized into demographics, purchase decision factors, and restaurant preferences. Extensive Exploratory Data Analysis (EDA) was conducted, exploring respondent attributes, influential features, deterrent conditions, cancellation events, food-related importance, and likelihood of repeat usage. Additionally, correlation analysis and geospatial assessments were performed, utilizing ARCGIS for clustered location and hotspot analysis. The Random Forest algorithm was selected for modelling, leveraging its robustness in handling complex relationships. Post-modelling, feature importance analysis was conducted to discern the pivotal factors impacting consumer behaviour. The study culminated in actionable insights for service providers, emphasizing the significance of factors such as convenience, speed, restaurant variety, and discounts, while also highlighting the need for location-specific training for food delivery personnel and targeted marketing strategies. These findings offer a roadmap for industry players to enhance customer satisfaction and retention, ultimately fostering competitiveness in this burgeoning market.

3.1 Dataset Description:

The dataset utilized in this study encompasses a wide range of data with approximately 55 variables, which are systematically organized into four main categories:

1. Demographics of Consumers: This category likely includes information about the survey respondents' personal attributes, such as age, gender, income, and other demographic details, enabling the analysis of how these factors influence their online food delivery choices.
2. Overall/General Purchase Decision: This section of the dataset is expected to include data pertaining to consumers' general attitudes, perceptions, and decision-making criteria when it comes to using online food delivery services. It may involve variables related to convenience, preferences, and opinions about the service.
3. Time of Delivery Influencing the Purchase Decision: This category might encompass data that describes how the timing of food delivery impacts consumers' choices. It could include variables related to delivery speed, punctuality, and time-saving factors.
4. Rating of Restaurant Influencing the Purchase Decision: This part of the dataset likely contains information about how the quality, ratings, and other attributes of the restaurants offering online delivery services affect consumer decisions. It might include variables related to restaurant ratings, menu options, and food quality.

Data collection was accomplished through a survey administered to consumers in Bangalore, India. The primary objective of this survey was to gather valuable insights into consumer preferences, behaviour, and the factors that drive their choices when using online food delivery services. The dataset serves as the foundation for the subsequent analysis, helping to draw meaningful conclusions and actionable insights for improving the online food delivery industry in the region.

3.2 Exploratory Data Analysis (EDA):

The Exploratory Data Analysis (EDA) encompassed a comprehensive examination of various facets pertaining to consumer behaviour in online food delivery services. The analysis commenced with an in-depth assessment of respondent attributes, employing descriptive statistics and visualizations to delineate the demographic profile of users. Subsequently, a focused exploration into the determinants influencing respondents' adoption of online food delivery revealed pivotal factors encompassing ease, convenience, time-saving, restaurant choices, payment options, and promotional offers. Concurrently, the study elucidated deterrents to adoption, identifying concerns including cost considerations, delayed deliveries, and a predilection for self-prepared meals.

Additionally, events precipitating order cancellations were scrutinized to glean insights into potential pain points in the delivery process. The investigation extended to evaluating the relative significance of various food-related elements in shaping consumer preferences within the online food delivery domain. Furthermore, an appraisal of the propensity of respondents to revisit online delivery services based on prior experiences was conducted.

Correlation analysis afforded a deeper understanding of the interrelationships between distinct features, shedding light on the complex dynamics influencing consumer choices. Moreover, a Geo-spatial Analysis leveraging ARCGIS was instrumental in visually mapping consumer locations, facilitating the identification of high-demand clusters. These insights culminated in a nuanced understanding of consumer behaviour, providing a foundation for strategic enhancements in the online food delivery industry, particularly in the metropolitan context of Bangalore, India.

3.3 Model Selection:

The **Random Forest** algorithm was judiciously selected as the modelling technique for this study, owing to its remarkable proficiency in managing intricate relationships within the data. This algorithm stands out for its capacity to yield precise predictions, especially in tasks involving classification. The rationale behind this choice lies in Random Forest's innate capability to harness the collective wisdom of multiple decision trees, each offering its unique perspective on the data. By aggregating these individual insights, Random Forest constructs a robust and versatile predictive model. This characteristic is particularly advantageous in scenarios where the underlying data may exhibit intricate interactions and non-linear dependencies. Moreover, Random Forest exhibits a commendable resilience to overfitting, which is instrumental in ensuring the model's generalizability to unseen data.

Furthermore, the algorithm's innate capacity to handle a diverse array of input variables, be they categorical or continuous, renders it well-suited for the multifaceted nature of the dataset under consideration. Its adaptability to accommodate a wide spectrum of feature types enhances its applicability in scenarios characterized by diverse data attributes. By harnessing the strengths of Random Forest, this study is poised to derive accurate and reliable predictions, thereby fortifying the analytical foundation of the research outcomes.

3.4 Feature Importance:

Following the training of the Random Forest model, a comprehensive feature importance analysis was undertaken to discern the pivotal factors exerting the most substantial influence on the prediction of consumer behaviour within the realm of online food delivery. This critical phase of the analytical process aimed to distil the multitude of input variables into a select set of influential determinants. Through an intricate examination of the model's internal workings, it became evident which features played the most prominent roles in shaping the predictive outcomes. This exercise is instrumental in illuminating the underlying dynamics of consumer preferences and behaviour, enabling a nuanced understanding of the factors that hold the greatest sway in this domain.

The outcomes of this feature importance analysis furnish valuable insights for strategic decision-making. By pinpointing the key drivers of consumer behaviour, businesses and stakeholders can strategically allocate resources and efforts towards optimizing the most influential aspects of their online food delivery services. Additionally, this knowledge empowers the formulation of targeted interventions and enhancements that are tailored to cater to the specific needs and preferences of the consumer base. In essence, the feature importance analysis serves as a compass, guiding the strategic trajectory of endeavours aimed at enhancing the online food delivery experience for consumers.

Results:

4.1 Data Visualization:

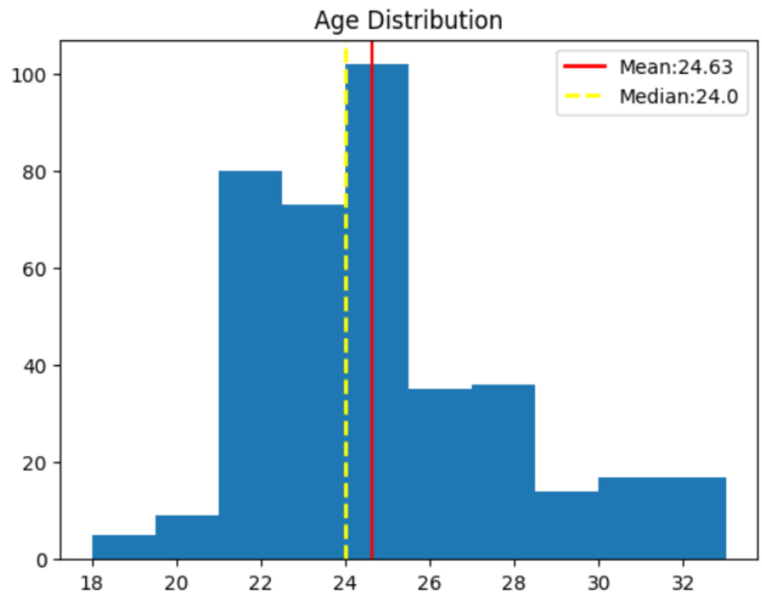

Fig 1. Age Distribution

In the provided code and description, a histogram of respondent ages is presented. The red vertical line represents the mean age (24.63), and the yellow dashed line represents the median age (24). This visual representation indicates that the majority of respondents fall within the 21-25 age group.

Key Inference:

The age distribution is right-skewed, with a mean slightly higher than the median, suggesting the presence of outliers or a skewed tail towards older ages.

The most common age group among respondents is 21-25.

The age range of respondents varies from 18 to 33 years, with 18 being the youngest and 33 the oldest.

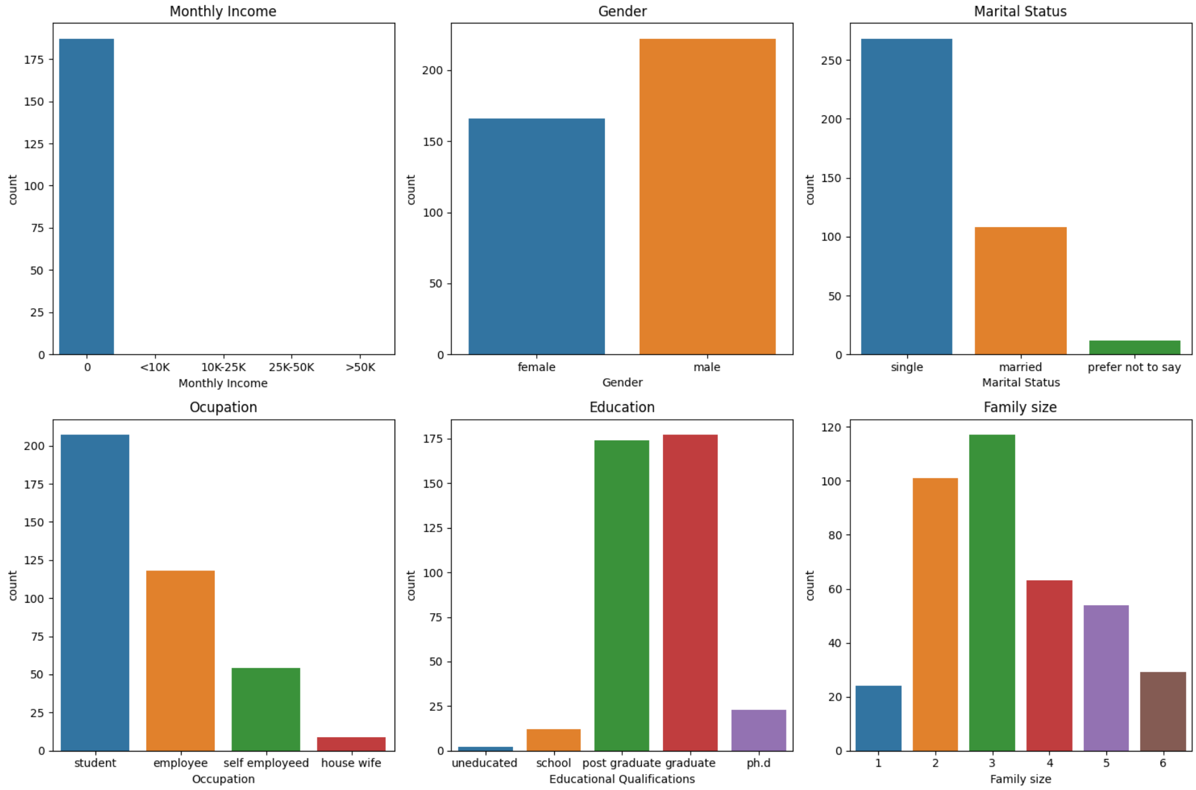

Fig 2.

Findings from the survey data reveal the following insights:

1. Income Distribution: The majority of respondents reported having no income, indicating a significant portion of the survey population may be students or individuals not currently employed.
2. Gender Distribution: There are more male respondents than female respondents in the survey. This gender distribution could be influenced by various factors, including the demographics of the survey's target population.
3. Marital Status: Most of the respondents are single, suggesting that the survey predominantly captures the perspectives of unmarried individuals.
4. Occupational Diversity: The most common types of occupation among the respondents are students, followed closely by employees. This indicates a diverse demographic, including students and working professionals.
5. Educational Background: The survey participants primarily hold graduate and post-graduate degrees, signifying a well-educated sample group.
6. Family Size: The majority of respondents reported a family size of three, indicating the prevalence of small-sized families within the survey population.

These findings provide valuable demographic insights into the characteristics of the survey participants, which can be useful in tailoring and targeting online food delivery services to specific segments of the population.

What Features Influence Respondents Using Online Food Delivery?

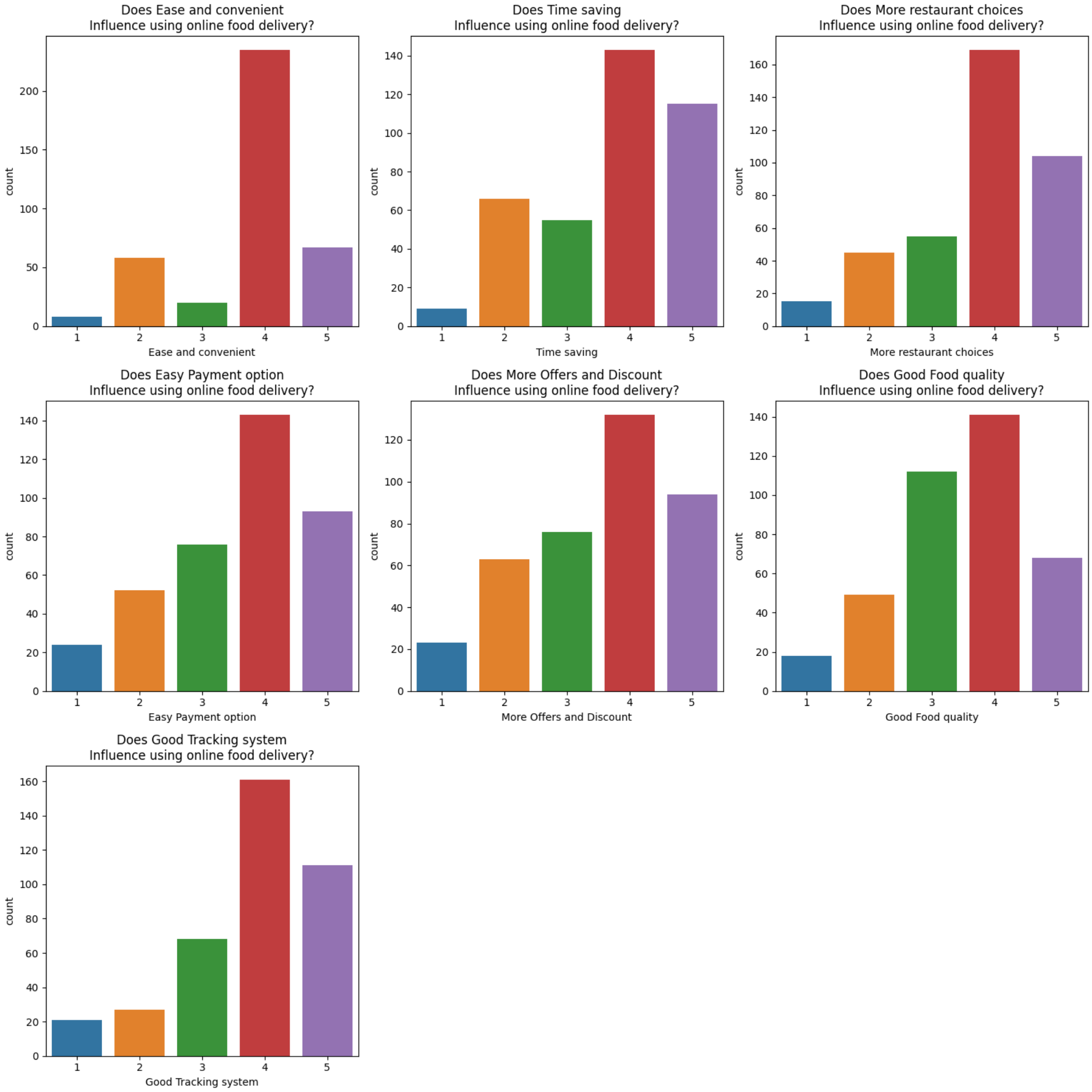

Fig 3.

Based on the survey responses, it is evident that respondents are most influenced to use online food delivery when they agree or strongly agree with the provided statements. Specifically, when respondents rated factors such as ease of application usage, time-saving, restaurant choices, and discounts with scores of 4 (Agree) or 5 (Strongly agree), it indicates a strong positive influence on their decision to use online food delivery services.

This implies that improving these aspects of the online food delivery service can potentially lead to an increase in customer adoption and usage. Therefore, focusing on enhancing the user interface for easier application usage, optimizing delivery times for increased efficiency, expanding restaurant options, and offering attractive discounts and offers could be effective strategies for influencing and retaining users of online food delivery services. These improvements align with the preferences and priorities expressed by the survey respondents, which can lead to a more satisfied and loyal customer base.

What Conditions Keep Respondents from Using Online Food Delivery?

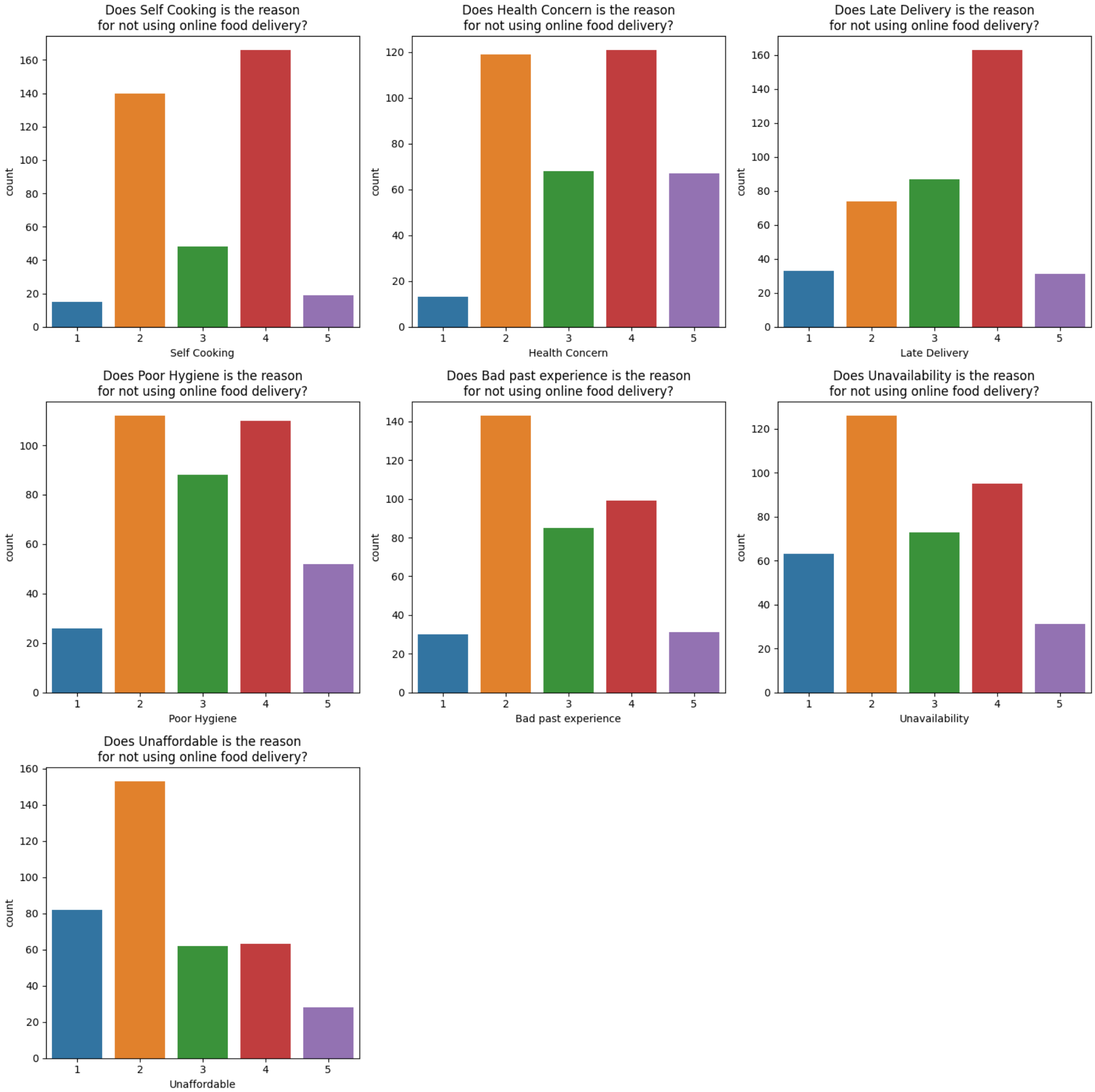

Fig 4.

Finding:

1. Most of respondents answered 4(Agree) for Self-Cooking is the reason they don't use online delivery
2. For the question about Health Concern as the reason, 4(Agree) and 2(Disagree) almost equal
3. Late Delivery is very clear become the reason why is respondent not using online delivery

4. Many respondents do not care about hygiene, it can be seen from the question about 'Poor Hygiene', many answered 2(Disagree) and the number is almost the same as those who answered 4(Agree)
5. Bad Past Experience, Unavailability and Unaffordable are not the reason for not using online delivery for most respondents.

What Event Make Respondents Cancel the Online Food Delivery?

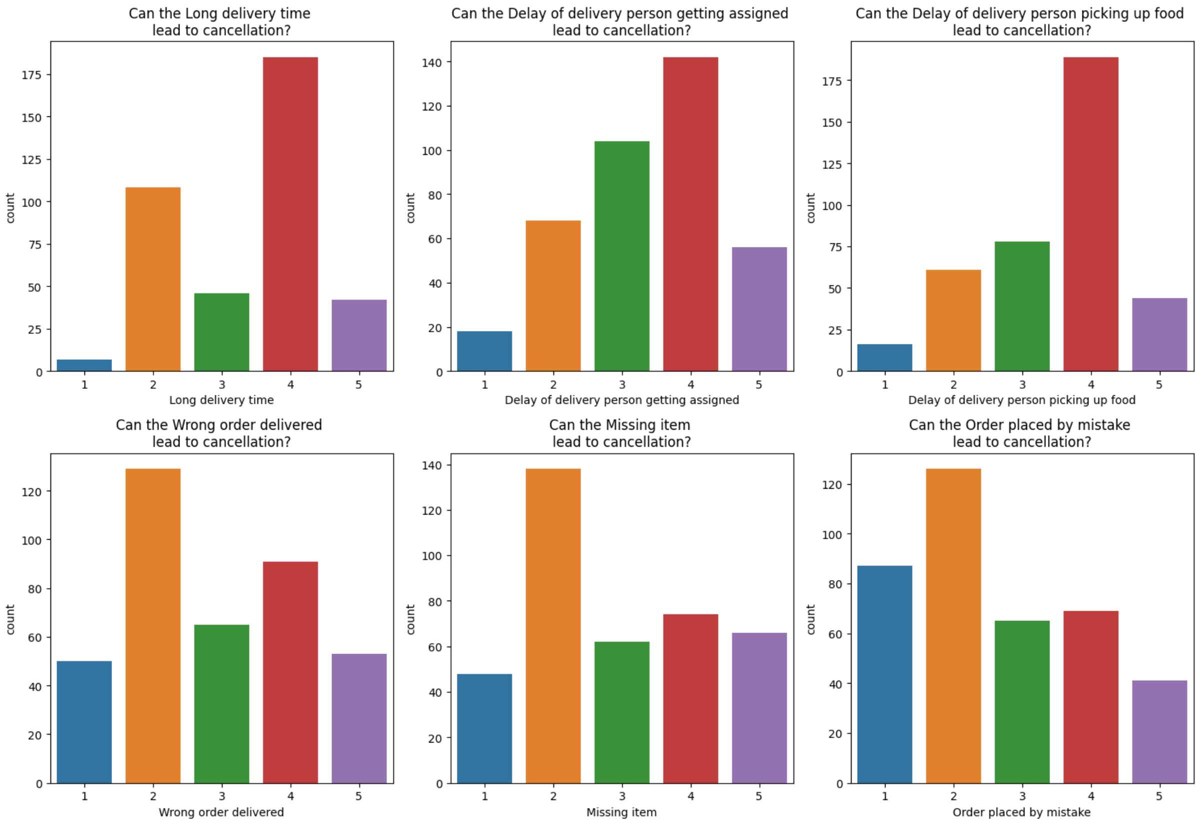

Fig 5.

Based on the survey findings, it is evident that time-related factors are the primary reasons for respondents cancelling their online food delivery orders. These factors include the length of time for delivery, the duration it takes for the delivery person to be assigned, and the time it takes for the delivery person to pick up the food. In contrast, issues related to errors in ordering and delivery do not significantly contribute to order cancellations.

Is This Food-Related Thing Important?

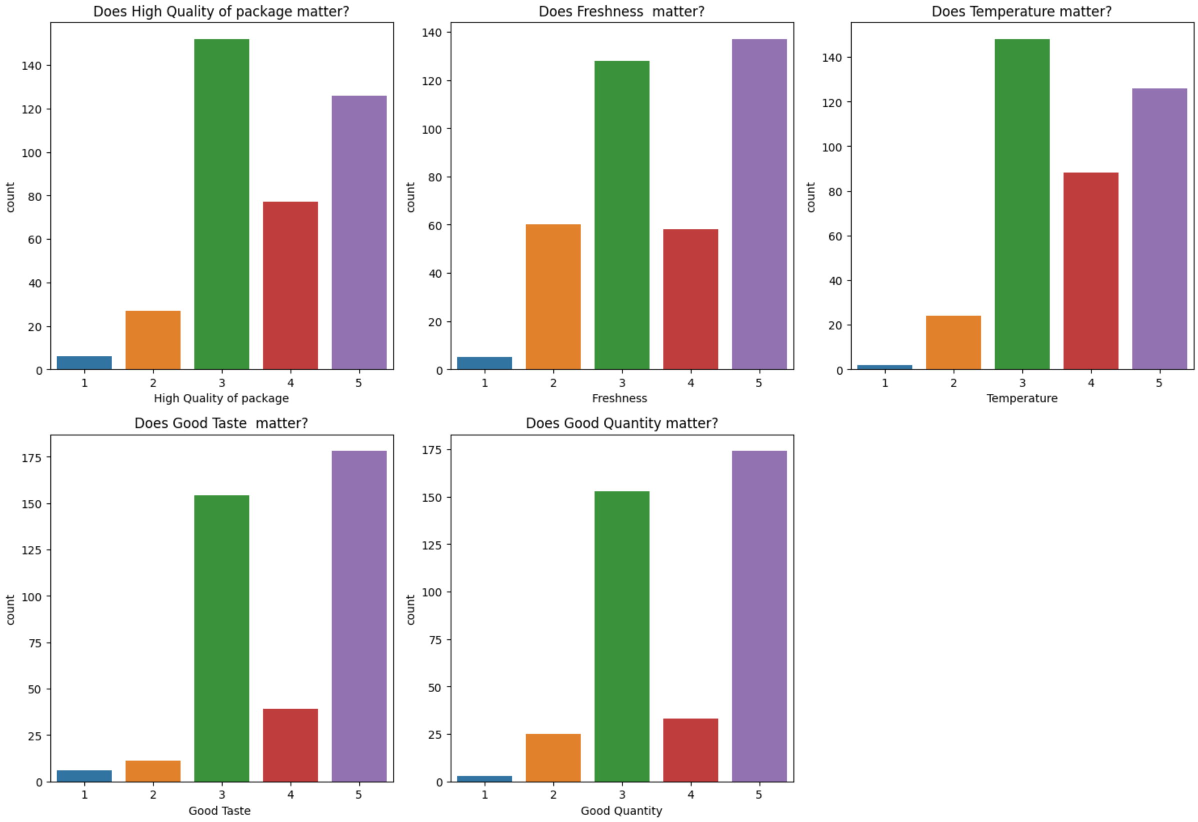

Fig 6.

Findings:

1. A significant portion of the respondents indicated that factors such as the packaging and food temperature were of lesser concern when it came to their food delivery experience.
2. The data analysis revealed that a majority of respondents prioritize the freshness of the delivered food.
3. A delicious and enjoyable meal is a crucial determinant of customer satisfaction and potential repeat orders.
4. Quantity of food was also highlighted as a key factor for many respondents.

The Number of Respondents Who Will Use Online Food Delivery Again ?

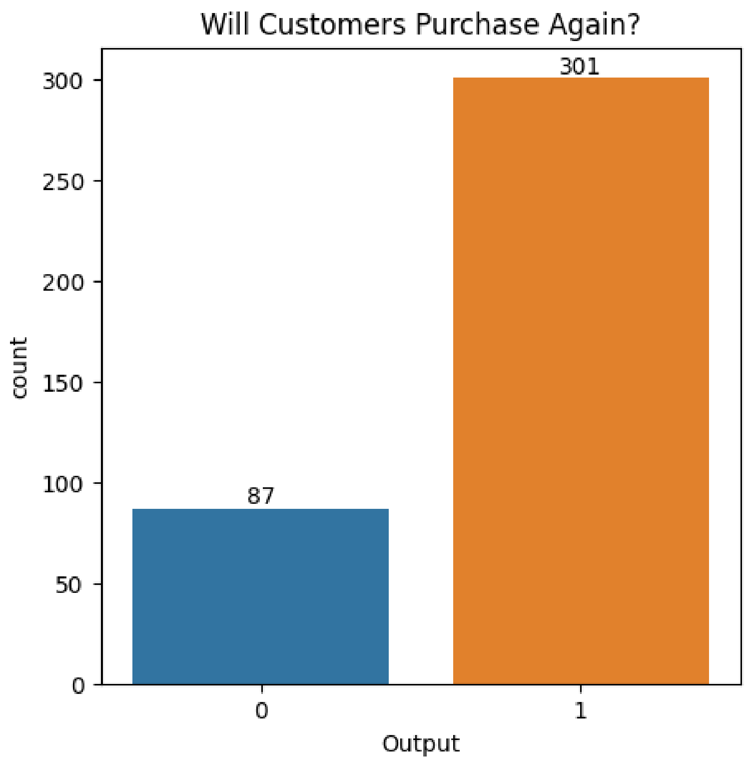

Fig 7.

From the graph, we know that 87(22%) respondents will not purchase again and it's quite high.

Features Correlation

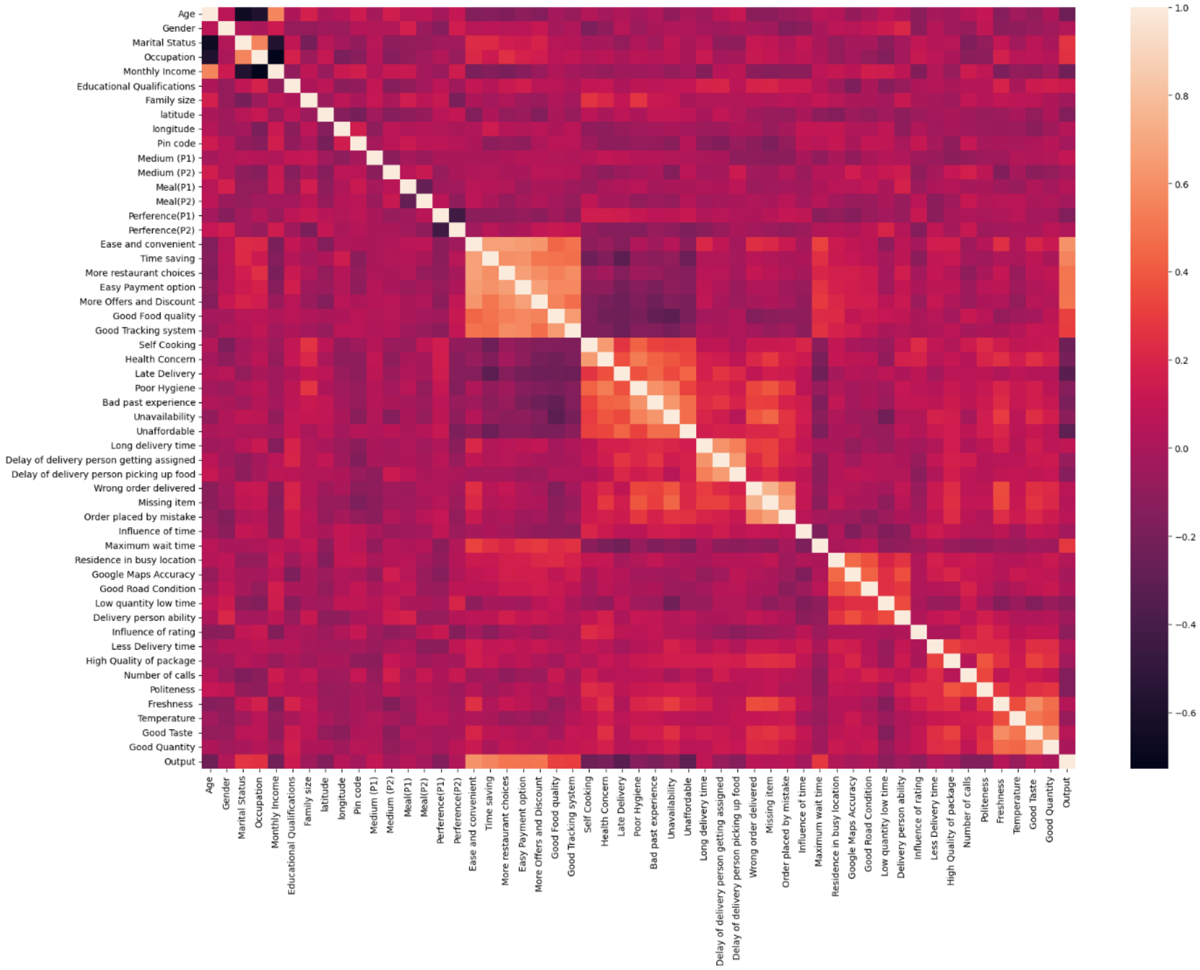

Fig 8.

We can see features that we put into one group has a high correlation with each other, and if we see the correlation for 'Output' we can see that 'Output' has a positive correlation with 'Influence Group' and has a negative correlation with 'Not Purchase Group'.

What Features Influence Customers to Use Online Food Delivery?

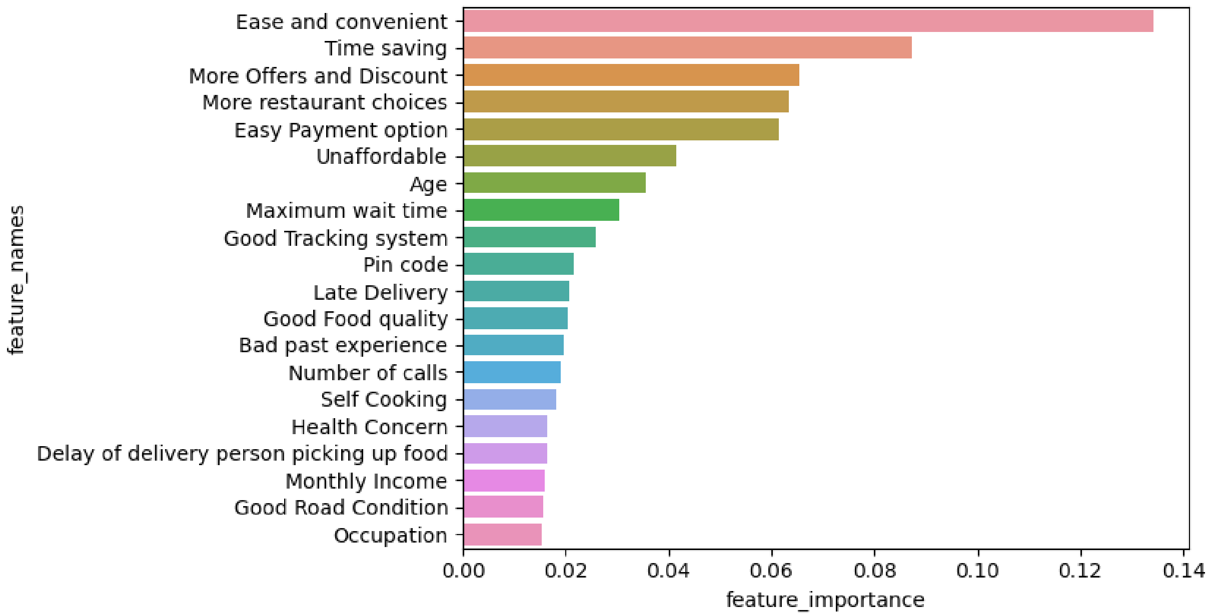

Fig 9.

Findings:

1. The higher the value they give for Ease and convenient, Time Saving, more restaurant choices, Easy payment options, and More offers and discounts, users tend to use online food delivery again
2. The older the user and the higher the monthly income, the more likely they are not to use online food delivery anymore
3. The higher the value they give for Unaffordable, Late Delivery and Self Cooking, the users are less likely to use online food delivery anymore.

4.2 Geospatial Analysis:

Geo Pointers:

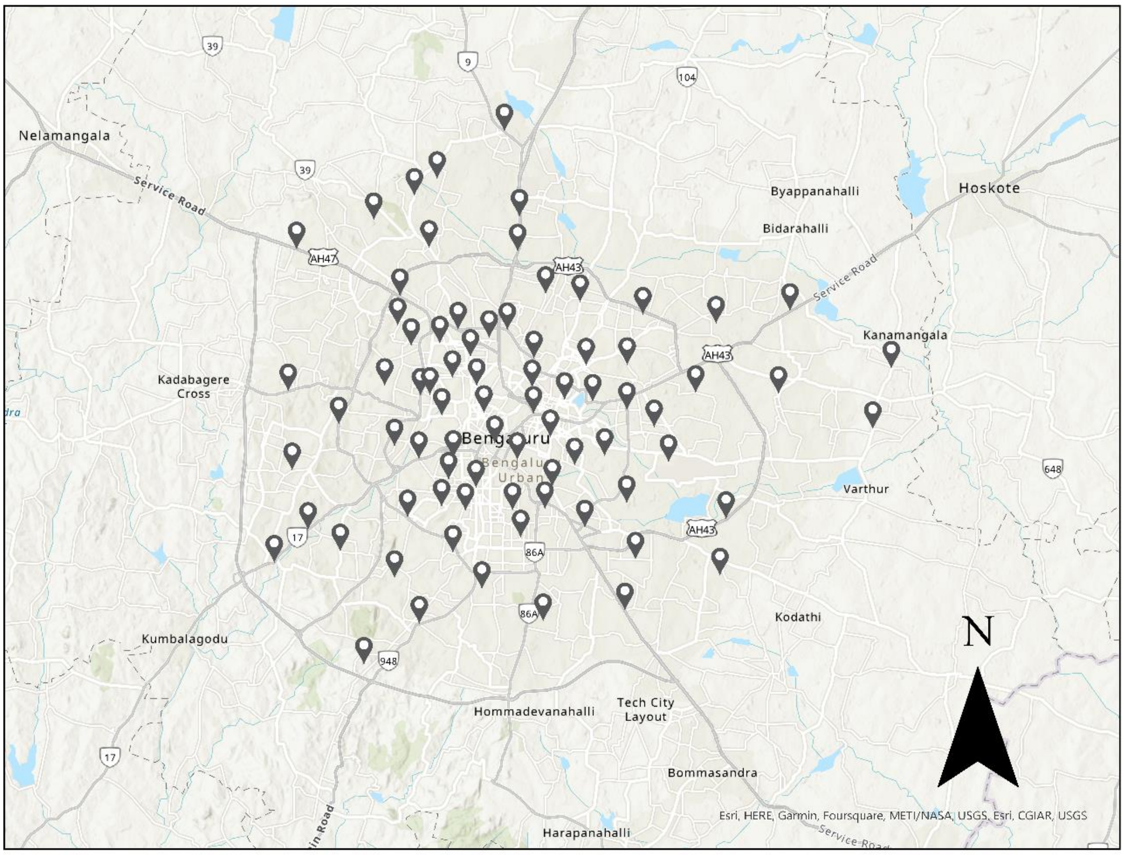

Fig 10.

We have considered the Age band of 18-40 to get the maximum pointers for this plot. As you can see, we have collected data across all the places inside the Bangalore Urban and also the outskirts of the city which makes the data more reliable for its findings. Most of the data are concentrated much on the Bangalore Urban(inside the city)

Clustered Location:

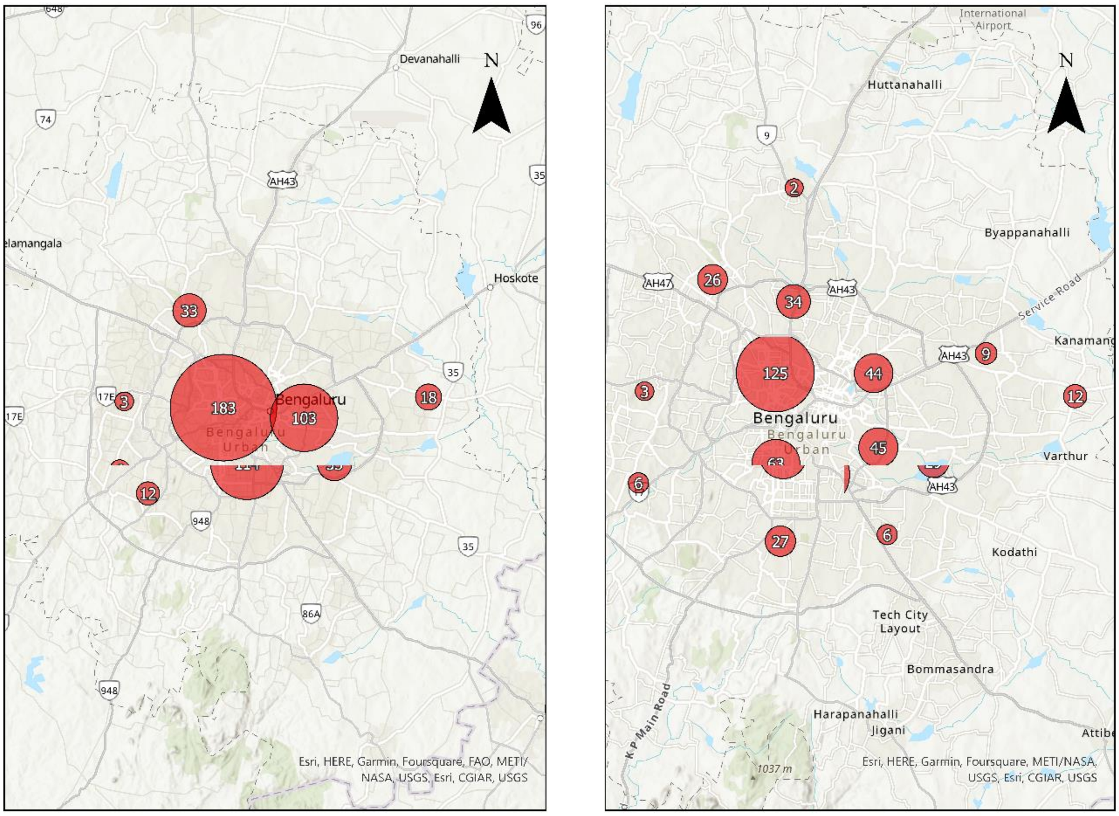

Fig 11.

Most of the data are concentrated in the heart of city- which includes places like Gandhinagar, Chikkapete, VVPuram, Rajajinagar. The next highest data cluster is recorded in South Bangalore where JP Nagar, Madiwala and Koramangala are located. We have also collected nearly 18 responses from Whitefield which is located in the outskirts of Bangalore.

Hotspot Analysis:

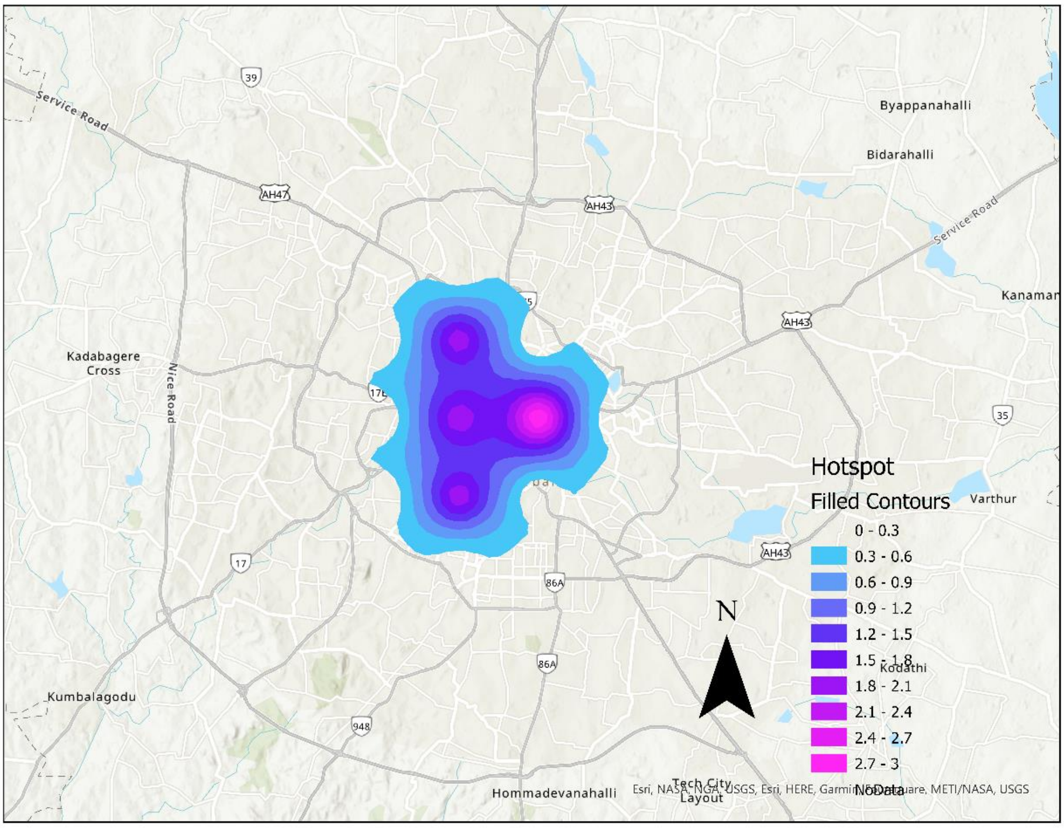

Fig 12.

A prominent finding is that the primary customer base for online food delivery services is concentrated in the central areas of Bangalore. This spatial insight indicates a dense clustering of demand around the city's core, suggesting that food delivery platforms should strategically focus their resources and services in these central zones. Understanding this geographic distribution is crucial for optimizing delivery routes, service coverage, and marketing efforts, ultimately leading to more efficient and targeted service provision. These findings not only enhance our understanding of customer behaviour but also provide actionable insights for service providers in catering to the core customer base in the heart of Bangalore.

Model : Random Forest

Random Forest is an ensemble learning algorithm that can be used for both classification and regression tasks. It works by constructing multiple decision trees during the training phase and outputting the mode of the classes (in classification) or the mean prediction (in regression) of the individual trees.
Here's an explanation of why Random Forest is used and how it works:

1. Ensemble Learning:
   Random Forest belongs to the ensemble learning family of algorithms. Ensemble learning combines the predictions of multiple base models (in this case, decision trees) to produce a more robust and accurate prediction. It leverages the wisdom of the crowd, where the collective decision of multiple models tends to be more accurate than that of any individual model.
2. Reduced Overfitting:
   Random Forest helps reduce overfitting, which is a common problem in machine learning. Overfitting occurs when a model learns the training data too well, including noise or random fluctuations, and fails to generalize well to new, unseen data. By aggregating the predictions of multiple trees, Random Forest tends to be more robust against overfitting compared to a single decision tree.
3. Handling High-Dimensional Data:
   Random Forest can handle a large number of input features, including both categorical and numerical features. It is robust to noisy or irrelevant features, which can be a challenge for some other algorithms.
4. Feature Importance:
   Random Forest provides a measure of feature importance, indicating which features have the most influence on the model's predictions. This can be valuable for understanding the underlying factors that drive the outcome and for feature selection.
5. Handling Non-Linear Relationships:
   Random Forest can capture non-linear relationships in the data. This is because each decision tree in the forest is constructed based on a random subset of features and data points, allowing it to model complex relationships.

```
Confusion Matrix:
 [[14  3]
 [ 1 60]]

Accuracy Score:
 0.9487179487179487

Classification Report:
               precision    recall  f1-score   support

           0       0.93      0.82      0.87        17
           1       0.95      0.98      0.97        61

    accuracy                           0.95        78
   macro avg       0.94      0.90      0.92        78
weighted avg       0.95      0.95      0.95        78
```

The output from the Random Forest Classifier (RFC) model provides valuable insights into the model's ability to predict whether a consumer will buy a meal again or not. The key metrics to consider in the context of this prediction are:

1. Accuracy Score (0.9487):The model achieved an accuracy score of approximately 94.87%. This indicates that the model correctly predicts whether a consumer will buy the meal again in nearly 95% of cases. In the context of predicting consumer behaviour, a high accuracy score is a positive indicator.
2. Precision and Recall: The precision and recall values for both classes (buy again or not) are also high, with values around 0.88 and 0.97. These metrics provide insights into the model's ability to make correct predictions (precision) and to identify all relevant instances (recall) in each category. In the context of predicting whether a consumer will buy the meal again, these high values suggest that the model performs well for both positive and negative outcomes.
3. F1-Score:The F1-scores, which are approximately 0.88 and 0.97 for the two classes, provide a balanced measure of the model's performance in terms of precision and recall. These scores indicate that the model is effective at finding the right balance between making accurate predictions and capturing most relevant instances.

In the context of predicting whether a consumer will buy the meal again, this RFC model appears to be a reliable tool for making such predictions. The high accuracy and robust precision and recall values reflect the model's ability to distinguish between consumers who are likely to reorder and those who are not. However, for a comprehensive evaluation, it's important to consider factors such as the distribution of classes, potential class imbalances, and the specific requirements of the application to make informed decisions regarding the model's performance.

Conclusion:

This study offers a comprehensive understanding of the intricate dynamics underpinning consumer behaviour within the online food delivery sector, specifically within the burgeoning market of Bangalore, India. The meticulous examination encompassing Exploratory Data Analysis (EDA), Geo-spatial Analysis, and the application of Random Forest modelling has yielded a trove of invaluable insights. These insights, in turn, furnish a strategic roadmap for industry stakeholders and service providers seeking to fortify customer satisfaction and retention rates. Through a judicious emphasis on the salient features delineated in the analytical findings, businesses can recalibrate their operational paradigms to align with the discerning preferences of their clientele. This adaptive approach not only augments competitiveness but also engenders a heightened resonance with the ever-evolving expectations of the consumer base. Thus, this study not only elucidates the current landscape but also equips industry players with the requisite intelligence to navigate and thrive in the dynamic milieu of the online food delivery ecosystem. It stands as a testament to the pivotal role that data-driven insights play in steering businesses towards sustained growth and relevance in an increasingly competitive market landscape.

The analysis of the online food delivery data in metropolitan cities like Bangalore has provided valuable insights into consumer behavior and preferences. The rise in demand for online delivery services can be attributed to factors such as ease and convenience, time-saving benefits, a wide range of restaurant choices, easy payment options, and attractive offers and discounts. These findings underscore the importance of continuously improving the service to enhance customer satisfaction and retention.

Demographically, the majority of respondents reported no income, with a higher representation of male respondents. Most respondents were single, and students were the most common occupation, followed by employees. Additionally, graduate and post-graduate education levels were prevalent among respondents, and families with a size of three were the most common.

The analysis also revealed that time-related factors, such as delivery time and the time taken by the delivery person, were significant contributors to order cancellations. However, errors in ordering or delivery were not major factors influencing cancellations.

The Random Forest Classifier (RFC) model proved to be highly effective in predicting whether a consumer would buy a meal again. With an impressive accuracy score of approximately 94.87% and high precision and recall values, the model demonstrates its reliability in distinguishing between potential repeat customers and those less likely to reorder.

Based on these findings, actionable data-driven solutions have been proposed. These include continuous improvement of service, a focus on maximizing delivery speed, targeted training for delivery personnel, expanding restaurant choices, implementing offers and discounts, and launching marketing campaigns to engage self-cooking individuals.

Future Work:

1. Keep improving the service so it's more ease and more convenient to increase the customer
2. Maximize speed in food delivery, because many customers still using online food delivery because they think it is Time-saving, and many customers are employees so they don't have much time during break
3. Train food deliverymen to know more about the areas where customers are so that it will reduce late delivery, too many calls the customers, and other things that customers don't like
4. Add more restaurant choices
5. Give more offers and discounts so you can reach the customer who said its Unaffordable
6. Have a marketing campaign to influence the self-cooking person to use online food delivery at least they use it once a week