Behavioural Analysis of Factors Influencing Online Food Ordering and Its Relation to Obesity

by

S. M. Fazle Rabby Labib
19301049
Fahmida Zaman Achol
19101470
Md. Aqib Jawwad
19301017

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
School of Data and Sciences
Brac University
June 2023

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

S. M. Fazle Rabby Labib
19301049

Fahmida Zaman Achol
19101470

Md. Aqib Jawwad

Md. Aqib Jawwad
19301017

Approval

The thesis/project titled "Behavioural Analysis of Factors Influencing Online Food Ordering and Its Relation to Obesity" submitted by

1. S. M. Fazle Rabby Labib(19301049)
2. Fahmida Zaman Achol(19101470)
3. Md. Aqib Jawwad(19301017)

Of Spring, 2023 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on June 04, 2023.

Examining Committee:

Supervisor:
(Member)

Dr. Md. Golam Rabiul Alam
Professor
Department of Computer Science and Engineering
Brac University

Program Coordinator:
(Member)

Dr. Md. Golam Rabiul Alam
Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi, PhD
Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

The purpose of this research is to determine what influences people to order food online and the intensity of ordering as well as to see if there is any correlation between online food ordering and obesity among the people of Bangladesh. We conduct an online questionnaire survey analyzing data from 343 participants aged 16-30 residing in Bangladesh. We use Ordinary Least Squares, Decision Tree, and Random Forest to determine the factors influencing customers' decision to order food online. The most influential factors are impulsive decision (mean = 2.01, OLS coefficient = 0.90), convenience of ordering (mean = 2.98, OLS coefficient = 0.65), variety of options available (mean = 2.98, OLS coefficient = 0.15), and promotional offers and discounts (mean = 2.60, OLS coefficient = 0.33). Then, we conduct statistical analyses to determine the correlation between obesity and the intensity of online food ordering. We find a weak positive relationship (p-value = 2.43^{-7}, coefficient = 0.28). We classify frequent users of OFO by using classifiers such as Logistic Regression, Naive Bayes, Decision Tree Classifier, Random Forest Classifier, Gradient Boosting Machine and K-Nearest Neighbors, where we find Random Forest to perform the best with an accuracy of 81%. Customer segmentation using K-Means clustering reveals that infrequent orders perceive OFO services as expensive, while frequent customers find the food lacking in nutrition. Several Chi-Squared Tests are conducted, revealing significant associations. Firstly, tech savviness is strongly related to perception of OFO services as user-friendly ($p < 0.001$). Secondly, individuals leading busy lives tend to make impulsive decisions ($p < 0.001$). Lastly, affordability of OFO services correlates with ordering behavior driven by promotional offers and discounts ($p < 0.001$). These insights, along with the developed prediction models and customer segmentation, contribute to a deeper understanding of consumer behaviour in an increasingly digitalized food landscape.

Keywords: Online Food Ordering, Obesity, Statistical Analysis, Prediction Models, Customer Segmentation, Hypothesis Testing.

Table of Contents

List of Figures

List of Tables

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

BMI Body Mass Index

GBM Gradient Boosting Machines

KMO Kaiser-Meyer-Olkin

KNN K-Nearest Neighbour

MCQ Multiple-choice Questions

OFO Online Food Ordering

OLS Ordinary Least Squares

PCA Principal Component Analysis

Chapter 1

Introduction

In the era of technological advancement, a substantial amount of people are seeking solutions through the Internet. As a result, people are looking for the fastest and optimum solution in every part of life. Being a developing country, Bangladesh is quickly evolving through this technology and most of us are adapting modern features and one of the most common habits is ordering food from online platforms. Online food ordering is so popular on such a day that we have access to any food at any time by our doorsteps.

Online Food Ordering or OFO Services are making our life easier but due to most food consisting of saturated fat, salt and sugar, we are also seeing an increase in major health issues like high blood pressure, obesity, diabetes, cardiovascular diseases and even several types of cancer. The proportion of overweight and obesity among adults is rising in developed as well as developing countries, notably among young adults. The World Health Organization (WHO) defines obesity as an abnormal accumulation of fat in the body that harms one's health. It is a worldwide public health problem which is increasing rapidly as the rate of obesity has tripled since 1975 [1]. Obesity is determined by measuring the Body Mass Index (BMI), which is the ratio of a person's weight in kilograms to their height in meters.

Bangladesh is also witnessing the rise of obesity over 33 years, as the rate of obesity has doubled [2]. From the statistics of the Global Obesity Observatory [3], Bangladesh currently scores 7/10 as the National obesity risk factor which is significantly high based on obesity prevalence.

OFO is currently witnessing its peak because of the accessibility and faster services, which raises the question of its relation to obesity. Although, there exists few papers on what influences a user's decision to order food online, near to no research has been done on what encourages people to order frequently. We aim to fill that gap. Our purpose here is to find the factors that cause people to use online food ordering services and if the ordering frequency plays any part in causing obesity.

1.1 Research Problem

With the increase in the popularity of delivered goods, the food sector has seen a major boom as well. Especially during the time of COVID-19, when people had no choice but to resort to online food ordering services for the sake of their safety and convenience. What once was a physical trip to the restaurant, now can easily be done with just a few clicks on our phones. According to 2022 research by Allied Market Research [4], in 2020, the worldwide market for mobile food delivery apps was valued at $6,752.32 million. This rise in popularity can be seen in Bangladesh too as online food delivery services like foodpanda, HungryNaki and Pathao Food are getting popular all over the country.

Yet, it is still unknown what factors are influencing people to order food online. It could be to save time or to save money and order something with discounts. Maybe it is because of how convenient it is to order food online or maybe it is driven by the hedonic motivation to eat comfort food.

Along with the popularity of online food ordering services, there has been an increase in obesity rates among adults. According to WHO's World Obesity Day 2022 study [5], more than 1 billion people worldwide are obese and this number is still increasing. This rise might have some relation to the increase in online food ordering. According to National Health Service (NHS) [6], obesity is the result of consuming excessive amounts of food and not engaging in enough physical activity. When large quantities of calories, particularly from saturated fat, salt, and sugar, are consumed but not used through exercise or physical activity, the excess energy is stored as fat in the body. Excess amounts of fat will lead to obesity. Most of the online food ordering services are offering restaurants that sell food that is highly processed, and filled with large amounts of saturated fat, sugar and salt. Along with that none of these apps provide any sort of nutritional information about the foods that are being ordered.

As the rate of obesity increases, it is now considered a global epidemic. According to a 2016 study by Hruby et al. [7], being obese can lead to serious health issues such as stroke, heart disease, type 2 diabetes, specific types of cancer, and early death. Therefore, it has become severely important to figure out what factors are causing it so that we can take measures to stop it.

In order to do that we need a significant amount of data which will be analyzed over the course of a few months. A problem arises as there is little to no data available related to this problem worldwide let alone in Bangladesh.
As a result, the research is trying to address the following questions:

What factors are influencing users' decision to order food through online services and if the intensity of ordering has any correlation with the increase in obesity rates among people of Bangladesh?

1.2 Research Contributions

In this research, we created a dataset on Bangladeshi customers using OFO services. We find the factors that influence the use of OFO services. We try to see if there exists any correlation between BMI and orders per month. Moreover, we develop a prediction model that classifies frequent online food orders. We apply customer segmentation to divide the user base in two clusters and analyze the results. Finally we do some hypothesis testing to find relation between the factors. This kind of research is non existent in Bangladeshi context The main contributions are summarized as follows:

1. We create a dataset of Bangladeshi OFO service users which we consider to be an important contribution since this type of dataset is non existent specially in Bangladeshi context.

2. We use the coefficient and feature importance values of OLS, Decision Tree and Random Forest to mine the factors that influence the use of OFO services.

3. We do several correlation analysis using Pearson Correlation, Spearman's Rank Correlation and Point-biserial correlation between BMI and orders per month as well as between order history and orders per month. We show that however very weak, but there exists a correlation between BMI and orders per month.

4. We use classifiers such as Logistic Regression, Naive Bayes, Decision Tree, Random Forest, Gradient Boosting Machine and K-Nearest Neighbors to develop a model that successfully classifies frequent users of OFO services.

5. We apply customer segmentation using K-Means on our data and divide the users into two clusters, analyze the data and show our findings. We apply PCA to do some dimensionality reduction.

6. We use Chi-squared test of independence to do some hypothesis testing between some of the factors to determine relation between each other.

1.3 Thesis Organization

The rest of the report is organized in the following order:

In chapter 2, we provide some literature review. It contains how previous researches were conducted and their findings. In chapter 3, we provide the methodology behind our work. In chapter 4, we show our findings and finally in chapter 5, we wrap up by sharing some concluding remarks, acknowledging the limitations and the scope for future work.

Chapter 2

Literature Review

OFO is a form of e-commerce. It refers to the procedure where food is ordered online and is then prepared and delivered to the customer [8]. Customers use platforms to place their order from options of various restaurants, the order is then cooked in the restaurant or kitchen then a delivery person delivers that order to the customer. Throughout the whole process, the customer can track where their food is and an estimated time of how long it will take to get the delivery. These OFO platforms provide many benefits like having the options of multiple restaurants, no waiting line, discounts and also saving the frustrations and confusion that happen when calling a restaurant to order [9].

Obesity is an illness that happens due to disproportionate fat storage in the body which can affect health adversely [1]. According to National Health Service (NHS) [6], it is caused by eating excessively or without following a limit and having little to no movement. Consumption of high energy and especially saturated fat, salt and sugar and not burning that energy off leads to get stored by the body as fat. It can also be caused by genetics and medical conditions like hypothyroidism and Cushing's syndrome but mostly caused due to poor diet and lack of physical activity. A sharp rise in obesity prevalence has made this a global epidemic.

2.1 Related Works

Online food ordering has become really popular lately, which brings us to the question of what influences people to order food through online platforms.

The following 2020 paper by Kale et al. [10], the authors describe the customer perception regarding OFO alongside its significance. Though the idea of a food menu may be outdated, in the recent era, we have more options to select from more restaurants and more menus. In fact, it plays a huge impact on the dining business. A survey was conducted using a questionnaire that was sent to people aged 18-60. Data analysis is performed using various techniques such as percentage tool, weighted mean, standard deviation, and coefficient of variation. The study also examines the audience's preference for reading reviews and the role of online review portals, as well as food bloggers' reviews and social media marketing, in shaping customers' perception of online food ordering services. Finally, the authors conclude that although online food ordering services are vast and diverse, based on the data and research, it is shown that 50% of people prefer online food delivery services instead of going out because of the convenience and flexibility. Additionally, ease of payment, sales promotions, and diverse reviews from different platforms play a crucial role in more consumers changing their perception about online food ordering.

Alagoz et al. [11] in their, propose an analysis based on customer attitudes in terms of ordering food using the internet among Turkey's university students using the Technology Acceptance Model (TAM).. The hypothesis of this research work mentions that the easy online ordering process encourages students to choose the online food ordering system, the necessity of online food ordering process promotes students to select this way, the innovativeness of the students based on information technology tends to order online, students trust in e-retailers and external influences encourage them to rely on online food ordering. Okan University's Faculty of Administrative Sciences and Economics students are hard copy survey participants with 28 closed-ended questions consisting of several items for measuring ease, usefulness, innovativeness, trust, external influence and attitude etc. The response rate of 71% is tested by reviewing kurtosis and skewness. Moreover, Barlett's test with 0.000 sphericity and with a value greater than 0.6 KMO test is also used where Bartlett's test is declined. Furthermore, Cronbach's alpha coefficient is used and considered with satisfactory value. To test the hypothesis and find the relationship of each independent factor, the execution of multiple regression analysis is done by IBM SPSS version 19.

In this 2022 study by Jahidi et al. [12], the positivist approach is applied to find factors that affect the attitude towards online food retailing by involving respondents who often use online food ordering systems in Bandung, Indonesia. They carry out research by using a non-probability sampling method, as the population is not known and the sampling frame is not accessible. Questionnaires are used to collect data. Thirty experiments are used for an initial test to see whether the form of the survey is appropriate to obtain the study. The questionnaire consists of two sections, one with demographic data and another with the research variables. Then the Partial Least Square (PLS) algorithm is applied to analyze the data. The authors then run the model against multiple hypotheses. It is found that hedonic motivation, price saving orientation is not that significant but time-saving orientation, prior online purchases, and convenience motivation play a significant role when ordering online. They conclude their findings by mentioning that a larger sample would have given a better representation of the population and the motivations may vary across other nations because of differences in culture and the acceptance of technological advancements.

The authors of this 2020 research [13], study the factors that influence online consumers behaviour in Salem, Tamil Nadu. They use Percentage Analysis, Ranking Analysis, One sample t-Test and Factor Analysis to evaluate the factors. After analysis, they find some statistical information such as how most respondents are females and most online shoppers are below age 30. They also find most online shoppers are influenced by online advertisements or through their friends. The authors conclude that the most important factor for online shopping is saving time and transportation costs.

Some of the research included the changes that come to decision-making when there is a global pandemic like COVID-19.

This research [14], shows the factors that are increasing online shopping during the pandemic period. The study also shows how social distancing has brought a lot of change in consumer behaviour over the two years. The aim is to detect consumers adopting and exploring different platforms and payment systems over the years and how it pushes people to access things more online. The research uses quantitative methods to evaluate sixteen factors to determine the consumers' purpose to prefer buying things online during the outbreak. The author concludes that because of the pandemic online shopping has increased by 27% and based on user experience, it will continue to grow further. Lastly, The author mentions that future studies in the field of technological uptake, especially in the area of online purchasing, may benefit significantly from the exploratory part of this research.

The writers of this article [15], explain how individuals act when using OFO platforms, particularly during the COVID-19 epidemic. They combine the task-technology fit model, the theory of planned behaviour, alongside the technology of acceptance to comprehend the behavioural pattern. 1320 consumers were asked a series of demographic questions and their responses helped shape the data. The main objective is to develop a Structural Equation Model which determines the relationship between the attitudes, subjective norms, and behavioural control that affect consumer behaviour toward online food ordering. The urge to use the OFO app for ordering food online is strongly influenced by mindset, personal standards, as well as perceived behavioural control. In light of many circumstances, such as non-linearity, correlated variable independence, measurement inaccuracies, associated error circumstances and one or more implicit variable dependencies, this makes the approach more efficient and compatible. Therefore, this research provides the factors that are drawn towards consumption and production which lead to creating a sustainable business model where consumers are drawn more towards availing the food delivery service. The author also mentions how the service of the delivery application made the consumers order more through online platforms. This study's sole drawback is the necessity to test a wide range of delivery applications in order to analyze user behaviour and measure people's opinions on food delivery services. Users' preferences may differ from nation to country or even from area.

This study [16], looks at how customers' intentions to use online food delivery services are affected before and during the COVID-19 pandemic. Data is collected using Amazon's Mechanical Turk, and the chi-square test of homogeneity is used for verifying significant imbalance within frequency counts of socio-demographic factors comparing two groups. The results show that factors such as Perceived Ease of Use, Perceived Usefulness Trust, Price Saving Benefits, Time Saving Benefits, Perceived Severity, and Perceived Vulnerability have a positive influence on customers' intentions to use OFO services. However, Food Safety Risk Perception has a negative influence. The study finds that customers are more likely to adopt online food delivery applications if they find them useful. They suggest that future research could include constructs such as customer satisfaction, willingness to pay a premium, positive word-of-mouth, and revisit intention.

With the rise of obesity around the world, it is important to study what risks this medical condition brings to our daily lives.

Chatterjee et al [17]. with their 2020 study, try to identify risk factors connected with obesity and overweight with a machine learning approach. Considering social determining and behavioural factors, physiological risks have been identified, where an outcome is obesity and overweight, which increases the risk of other diseases. The primary goal of this study is not to highlight any risk predicting model but to review several machine learning methods and execution using health data. Analyzing already existing obesity-related data associated with machine learning algorithms helps to identify weight change-related risk factors, the nature of the data to construct the eCoach system, the necessary at the primary level of the existing datasets from ”UCI” and ”Kaggle”, finding efficiently performing classification and regression models are the goals of this work. The authors review the scientific literature from the year 2012 to 2019 from renowned platforms such as ”Google Scholar”, ”Scopus”, ”MDPI” etc. Besides ”Microsoft Excel”, ”EndNote”, ”DOAJ” and more to make article searching possible, searching is dependent on appropriate keywords and the number of selected articles for the final phase is 40. An epidemiological study shows how several groups of people's lifestyles lead to illness, its reason and its risks. Some related works are also included done by other authors such as different multivariate regression models, Multilayer Perceptron (MLP), Millennium Cohort Study (MCS), Mean Absolute Error (MAE) and many more models noted in the study by Chatterjee et al. By employing open datasets from the resources mentioned above, several machine learning models' performance in classification and regression analysis is measured. Chatterjee et al. find the riskiest factors of obesity through selected literature studies. Continuous and categorical in these two groups collected data are arranged. To filter and discard data samples, data mining is introduced. Furthermore, the language library for data processing is Python 3 where Spyder IDE is used for this experiment. The body consumption assessment is also shown with BMI and waist-hip ratio techniques. Using statistical, machine learning and data visualization methods, potential risk factors have been identified by the authors. In the end, the authors focus on designing, testing, evaluating and developing the eCoach system to achieve the goal.

In their 2003 paper [18], Prentice and colleagues review a number of studies on the energy density of food and its impact on energy consumption. They find that people have difficulty choosing foods with high energy density and maintaining energy balance. By analyzing popular fast food websites, the authors find that many fast food items have a high energy density, with an average energy density that is 65% higher than the British average diet and two times higher than preferred healthy diets. This can lead to weight gain and obesity. The authors identify factors such as eating diversity, inactive lifestyle, and environmental changes as contributing to obesity. They propose that energy density plays a key role in this process. To support this claim, the authors use both physiological experiments and published literature. They also conduct a series of feeding trials in Cambridge to examine the effects of manipulating the dietary energy density. They find that people tend to consume foods with high energy density, and that common fast foods contribute to excessive consumption. Children are particularly at risk due to their undeveloped cognitive dietary restraints. The authors suggest that fast food industries could lower the energy density of their products through low-fat and low-energy options, which would also be profitable for the industry. They also propose various other steps such as including healthy options on menus and reducing the fat content of fast foods in order to address obesity as a multifactorial disease.

In his research, Dave et al. [19], represents the relationship between adults' fast food consumption frequency and their attitude towards it throughout their article. This study primarily uses telephone surveys as a cross-sectional examination to assess people's eating habits and opinions. The survey is conducted among 530 adults of different classes of life. With the help of an 11-item, 4-dimensional scale, attitudes toward fast food are measured. Based on gender, age, and other variables, the frequency of fast food consumption is calculated. The target of the study is to represent the unhealthy perspective of fast foods and to reduce their consumption.

This research [20], examines the factors that influence the frequency of online food ordering for high-risk food. The authors collect data from a cross-cultural study of 686 young adults aged 20-39, and observe their behaviour for a minimum of six months. They use a T-test and a Chi-square test to analyze the significance of online food ordering and high-risk food consumption through online food ordering services. Multivariate logistic regression is used to determine the corresponding predictive factors. The study finds that young adults are influenced by social, environmental, and appealing advertisements, which leads to consuming food through online food ordering services. While the study emphasizes the case that encouraging positive influences could be advantageous even when the link between cause and effect has not been determined. The authors recommend that highlighting the strong influence of consuming high-risk food through food delivery services would help to encourage healthy food consumption among young adults.

The author here [21] examines the relationship between the type and frequency of online food ordering and obesity among university students. The study is conducted using a sample of 83 students, and data was collected on their weight, height, and food frequency. The data is analyzed using a chi-square test, and the results showed no correlation between the types and frequency of online food ordering and obesity. The authors suggest that other factors, such as food consumption and daily lifestyle habits, may play a role in the incidence of obesity among young adults. They also note that online fast food can contribute to obesity, as it is affordable, convenient and appealing to current tastes, leading to overconsumption. Women are more likely to order food online, but the portions are not large enough to be considered a full meal, leading to overconsumption. Additionally, a lack of physical activity among this group can also contribute to obesity over time.

In this paper [22], the authors try to correlate online food ordering and nutritional status. They use descriptive correlation with a cross-sectional approach and study 388 college students in Surabaya, Indonesia. They used the dependent variable's nutritional status and the intensity of online food ordering as their independent and dependent variables, respectively. The Food Frequency Questionnaire is used as the research instrument. To characterize the frequency distribution of the sample's age, gender, knowledge, intensity, and BMI, they use univariate analysis. Using Spearman Rho's test, a bivariate analysis is performed. The majority of the ordered meal is high in calories, fat, sodium, and unsaturated sugar according to the analysis of all the data. There is no connection between OFO and the nutritional state of college students.

In summary, we can see that recent studies related to this issue mostly talk about online shopping as a whole or only study customer behaviour in order to grow the OFO industry even more. The studies that have correlated the intensity of buying food through online mediums and obesity are mostly based in other countries. On top of that the datasets used on those researches are not available for future works. None of the studies have been specifically done with the people of Bangladesh. With our research, we want to tackle the issues from the Bangladeshi perspective.

Chapter 3

Methodology

In this chapter we talk about the the procedure of collecting data and the methodologies behind our experiments.

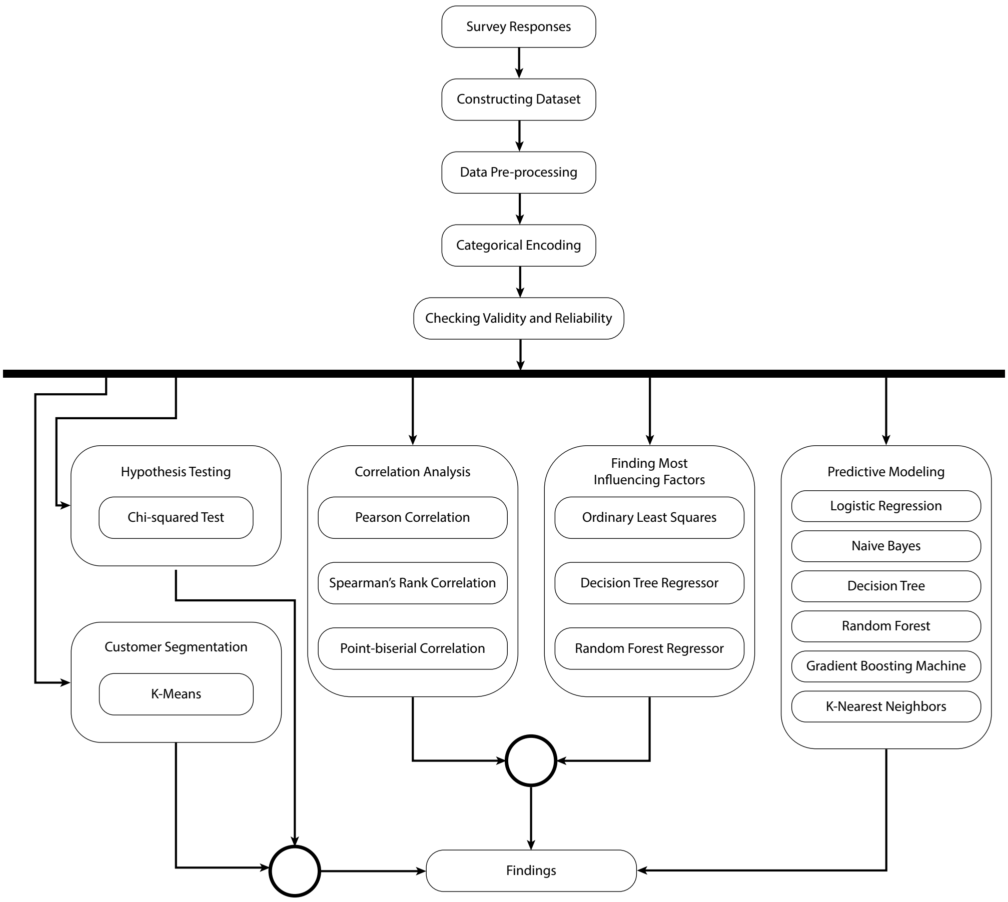

Figure 3.1: Top level overview of the proposed system.

In figure 3.1, we can see the top level overview of our proposed system. We start by collecting responses from the survey and constructing our dataset. We do data pre-processing and test for validity and reliability. Then we do correlation analysis, find factors that influence the use of OFO services, create a prediction model, apply customer segmentation on our data and do some hypothesis testing. Finally, we show our findings.

3.1 Data Collection

In order to investigate the factors that influence the decision to order food online among individuals in Bangladesh, a survey is conducted. The survey questions are divided into two parts. One is the demographic data and the other is questions related to online food ordering and the factors influencing it. There were in total 25 questions in the questionnaire. The survey is done anonymously to protect the respondent's privacy.

The survey is done through Google Forms and is designed to take the smallest amount of time possible. To accomplish that, we first share the form with our peers and take feedback from them. We incorporate the feedback and make it public. The questions ranged from text-based, multiple-choice questions (MCQ), check-boxes, and Likert scale questions. The survey questions are shared in the appendix.

Table 3.1: Questionnaire Description

Attributes	Data Type
Gender	Categorical Value
Age	Numeric Ranges
Weight	Numeric Value (Kilogram)
Height	Numeric Value (Feet and Inches)
Educational qualification	Categorical Value
Employment status	Categorical Value
Financial dependency	Categorical Value
Marital status	Categorical Value
Physical activity	Numeric Range: 1-7 (per week)
Preferred food ordering apps	Categorical Value
Duration of using OFO apps	Categorical Value
Orders per month	Numeric Value
Ordering time	Categorical Value
Types of food ordered	Categorical Value
Likes to try new technologies	Likert Scale
Makes impulsive decisions	Likert Scale
Too busy to cook	Likert Scale
Does not like to cook	Likert Scale
Ordering online is easy	Likert Scale
Many options to choose from	Likert Scale
Ordering is inexpensive	Likert Scale
Because of promo codes and discounts	Likert Scale
Food is delivered quickly	Likert Scale
Delivered food is safe and hygienic	Likert Scale
Delivered Food is nutritious	Likert Scale

Throughout the data collection period we receive several queries regarding why a respondent needs to login to their email to respond to our questions. We explain it to them that this is necessary to protect the survey from getting spam and random responses. Another concern we faced, is about the user being reluctant to share their personal information like height and weight we assured them that none of the answers are connected to their emails. We also made sure to not force anyone to share their information and only took responses if they gave it willingly.

3.2 Data pre-processing

After collecting the survey responses, the column names are modified for improved readability and ease of use. The height variable, which were initially recorded in feet and inches, get converted to meters for consistency. Using the converted height and weight variables, the BMI of each respondent is calculated. A new variable, BMI_CAT, is created to store the categorical values of BMI, including "Underweight", "Normal", "Overweight" and "Obese". In order to maintain the focus on the target population of individuals aged 16-30, any responses from individuals outside of this age range are removed, as about 98% of the data consists of individuals within this age group. The Likert scale data obtained from the questions related to various factors are encoded into ordinal numerical values for ease of analysis. This conversion allowed for a more accurate statistical analysis of the collected data. We remove any responses where orders per month is less than 1, since those information are not useful for our research problem. After pre-processing we end up with valid 331 responses which we use to conduct rest of our research.

Below, we use the 331 valid responses out of the 343 responses we receive, to show some statistical description of the dataset and do some exploratory data analysis:

Table 3.2: Distribution of Gender

Gender	Frequency	Percentage
Male	202	61.03%
Female	125	37.76%
Prefer not to say	4	1.21%

Based on our survey we see almost 331 people submitted their responses and most of the participants are male (61.03%). We have received 125 responses from female participants (37.76%) and also received 4 more responses from others (1.21%).

Table 3.3: Distribution of BMI

BMI	Frequency	Percentage
Underweight	25	7.6%
Normal	193	58.3%
Overweight	85	25.7%
Obese	28	8.5%

Analyzing the frequency distribution of BMI, most of the participants (58.3%) respond to having a normal BMI where the submission number is 193. The second largest submitted response (25.7%) for BMI is recorded to be overweight with 85 responses. Thirdly, almost 28 participants consider BMI as obese (8.5%). Moreover, 25 participants responded assuming to be underweight (7.6%).

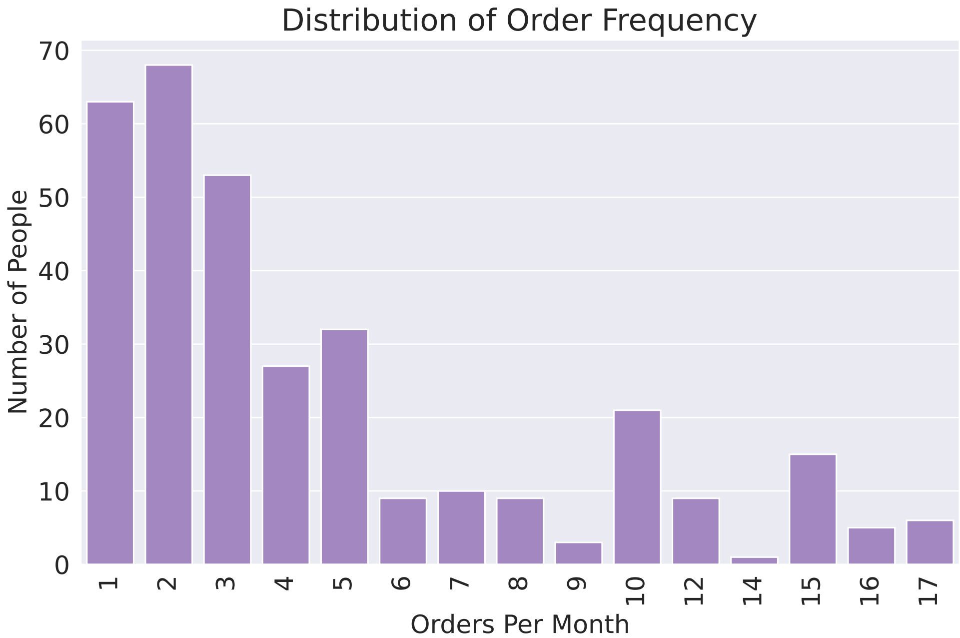

Figure 3.2: Distribution of Online Food Orders Per Month

From Figure 3.2 we can see the number of ordering food online per month by our survey participants. Here the range of ordering online is from 0 to 17. The highest number of ordering food in a month is 3 according to the chart. After that 2 and 1 are respectively the second and third highest numbers.

Table 3.4: Distribution of Educational Qualification

Education	Frequency	Percentage
Undergraduate	252	76.13%
Higher School Certificate (HSC)/A-Level	22	6.65%
Graduate/Post-Graduate/Phd	57	17.22%

Our online food ordering survey participants are mostly undergraduate-level students (76.13%) where the number is 252. We also have 57 responses from Graduate / Post-Graduate / Ph.D. students (17.22%). The least amount of responses (6.65%) we received from higher school certificate (HSC) / A-Level students with 22 responses.

Table 3.5: Distribution of Employment Status

Employment Status	Frequency	Percentage
Seeking opportunities currently	146	44.11%
Part-time	65	19.64%
Full-time	62	18.73%
Prefer not to say	58	17.52%

From table 3.5, we see that 146 participants are seeking career opportunities (44.11%). We also have 62 participants who are full-time employed (18.73%). A significant number of participants (17.52%) did not mention their employment status where the number is 58. Moreover, 65 participants are doing part-time jobs (19.64%).

Table 3.6: Distribution of Financial Dependency

Financial Dependency	Frequency	Percentage
Fully Dependent	142	42.90%
Partially Dependent	136	41.09%
Independent	53	16.01%

Inspecting the table 3.6, we see that 142 participants are fully dependent (42.90%). The number of partially dependent (41.09%) participants is 136. Lastly, 53 people responded to their status as independent (16.01%).

Table 3.7: Distribution of Marital Status

Marital Status	Frequency	Percentage
Single	304	91.84%
Married	18	5.44%
Prefer not to say	9	2.72%

Scanning table 3.7, we can see that 304 individuals are single (91. 84%). We also have 18 participants who are married (5.44%) and only 9 respondents did not mention their marital status (2.72%).

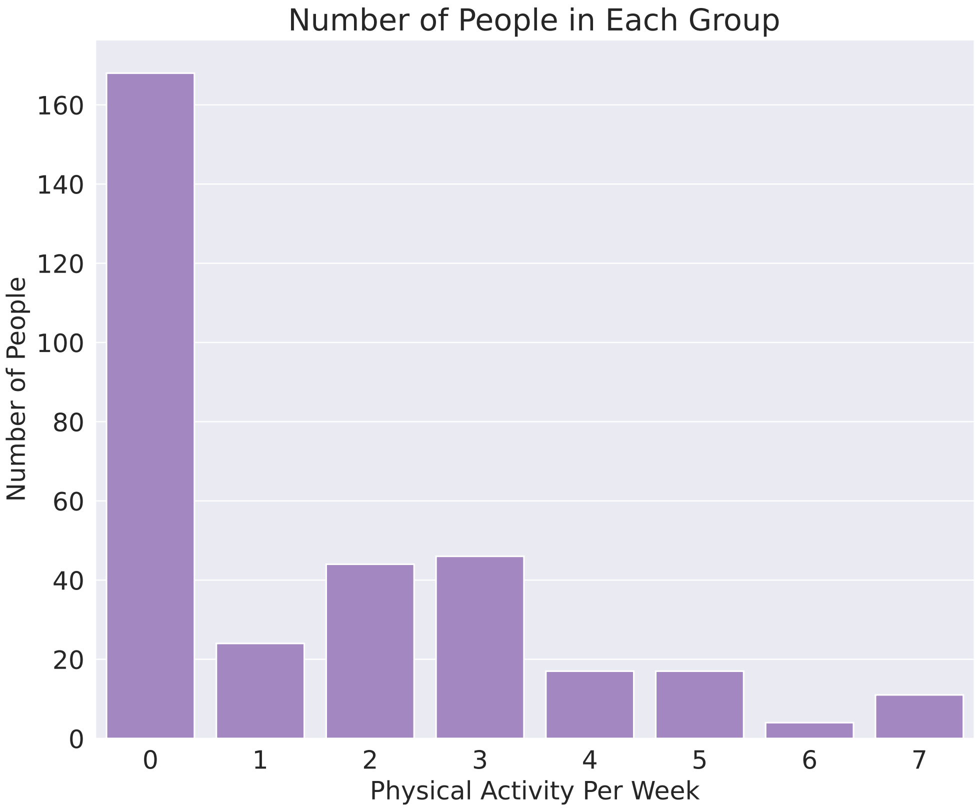

Figure 3.3: Distribution of Physical Activity

The figure 3.3 contains the distribution of the respondents' physical activity over a week. More than 160 participants responded that not performing any physical activity per week. Less than 60 participants responded that they are doing physical activities 2 to 3 times per week. Less than 40 participants voted 1 time engaging in physical activities and not more than 20 people responded 4 to 7 times per week of investing time per week for physical activities.

Table 3.8: Distribution of Online Food Ordering History

Ordering History	Frequency	Percentage
Less than 3 months	43	12.99%
3-6 months	21	6.34%
More than 6 months	262	80.66%

We can see that 262 participants are ordering food online for more than 6 months (80.66%). Moreover, we can notice that 43 participants ordered food online for 3 months (12.99%). Furthermore, 21 participants are ordering food online for 3 to 6 months (6.34%).

Table 3.9: Distribution of Online Food Ordering Time

Ordering Time	Frequency	Percentage
Breakfast	17	2.86%
Lunch	138	23.23%
Evening Snacks	227	38.22%
Dinner	176	29.63%
Midnight Snacks	36	6.06%

Table 3.9 gives us an idea of when people tend to order food. Firstly, most of the participants (38.22%) are ordering food for evening snacks and the number is 227. Secondly, 176 participants voted for online food ordering time as dinner (29.63%). Thirdly, 138 people responded to lunchtime for ordering food online (23.23%). Moreover, The number of ordering food online for midnight snacks (6.06%) is 36. Furthermore, 17 people selected breakfast for ordering food online (2.86%).

Table 3.10: Distribution of the Use of Online Food Ordering Apps

OFO Apps	Frequency	Percentage
foodpanda	302	59.57%
HungryNaki	55	10.85%
Pathao	136	26.82%
Others	14	2.76%

Most of the survey participants (59.57%) are using 'foodpanda' and the number is 302. The second highest (26.82%) used app by our survey participants is 'Pathao' where the number of users is 136. Thirdly, 55 people are using 'Hungrynaki' (10.85%). Lastly, we can see that 14 survey participants are using other apps (2.76%).

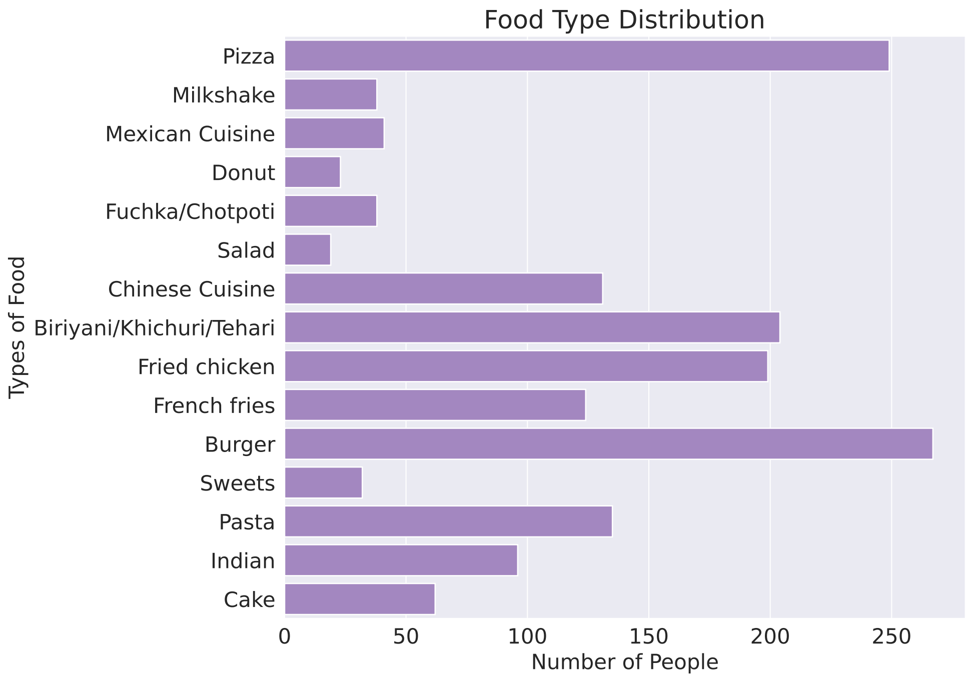

Figure 3.4: Distribution of type of foods ordered through online platforms

From figure 3.4, we can see that majority tend to order burgers (267) and pizzas (249) when ordering food online, which consist of high amount of saturated fat and salt. Salads (19) is among the lowest to be ordered. There were some food types with only one entry, we dropped those since those are negligible.

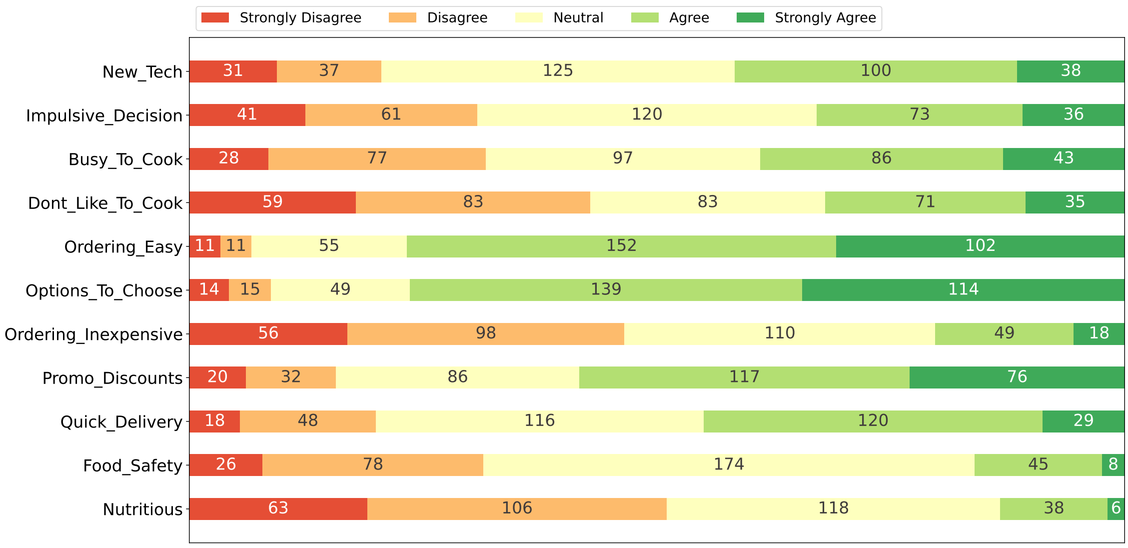

Figure 3.5: Distribution of Factors that Influence Online Food Orders

The figure 3.5 gives us an idea of the importance of each factors on a likert scale. Firstly, we found that 100 people agreed and 125 responded as neutral to being inclined to trying out new technologies. Secondly, 120 people voted neutral and 73 agreed to making impulsive decisions to order food online. Thirdly, 97 people responded neutral, 86 agreed and 77 disagreed with being too busy to cook food themselves and leaning towards OFO services. There is a match with 83 numbers of responses where each is allocated for disagree and neutral where the factor is people don't like to cook. A huge number of participants agreed that ordering food online is easy, where 152 people agreed and 102 strongly agreed. Again, lots of people agreed that they like having variety of options to choose their desired food from OFO platforms, where 139 people agreed and 114 strongly agreed. In the case of choosing to order food online due to being inexpensive, 110 participants responded as neutral and 98 people disagreed. 117 people agreed that promotional discounts convince them to order food online whereas 86 respond to neutral. A big portion of our survey participants agreed that ordering food online allows them to get their food quickly and the response number is 120, besides 116 people voted neutral. 174 of the participants stay neutral when they are asked if they think their food is safe and hygienic and 78 participants disagreed with it. Lastly, a very few survey participants believe that their online food ordered item is nutritious with 118 people responding as neutral, 106 disagreed and 63 participants strongly disagreed with the factor.

3.2.1 Dataset Validity and Reliability Test

Before applying all of the discussed methods we use Chronbach's Alpha [23], which is a way of assessing reliability by comparing the amount of shared variance, or covariance, among the items making up an instrument to the amount of overall variance. This is done to test the validity and reliability of our dataset. The idea is that if the instrument is reliable, there should be a great deal of covariance among the items relative to the variance.

After applying Chronbach's Alpha on our dataset we get the value of 0.89. According to this paper by Taber [24] a chronbach's alpha value higher than 0.70 is considered acceptable and above 0.80 and above is considered better. Therefore we can safely say that our dataset is valid and reliable.

3.2.2 Correlation Analysis

To examine the relationship between BMI (x) and the frequency of ordering food online (y), we apply several correlation analysis methods. These are Pearson Correlation [25], Spearman Rank Correlation [26] and Point-biserial Correlation [27].

Pearson correlation: This method measures the linear relationship between two continuous variables. The process involves calculating the means ($\bar{x}, \bar{y}$) and standard deviations of each variable, followed by computing the product of the deviations of each variable from its mean for each data point. The following equation is then used to calculate Pearson's correlation coefficient, r:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{3.1}$$

The resulting value of r ranges from -1 to 1, here -1 indicates a perfect negative linear relationship, 0 indicates no linear relationship, and 1 indicates a perfect positive linear relationship. To evaluate the significance of the correlation coefficient, a p-value is determined using the correlation coefficient, degrees of freedom (df = n-2) and a significance level of $\alpha = 0.05$ is used usually. The table of critical values is used to find the p-value.

The correlation coefficient is assumed as zero by the null hypothesis, implying that there is no correlation between the two variables. The alternative hypothesis is that the correlation coefficient is not zero, implying that there is a correlation between the two variables. If the p-value is less than α, we get enough evidence to accept the alternative hypothesis and reject the null hypothesis.

This method is applicable when both variables are continuous and assumes the data follow a normal distribution. It also assumes the observations are independent of each other.

Spearman's rank correlation coefficient: This is a non-parametric way of measuring correlation. With this we can measure how well the relationship between two variables can be described by using a monotonic function. It is similar to Pearson correlation but instead of measuring linear association, this evaluates rank order similarity between variables.

In this method each variable is assigned ranks based on their values. The smallest value receives a rank of 1 and the second smallest a rank of 2 and so on. Then it determines difference between ranks for each paired observation and these differences represent the change in the variables' relative positions. Each difference obtained in previous step is then squared to stop the positive and negative differences canceling each other out. Finally, the coefficient is calculated using the following formula:

$$\rho = 1 - \frac{(6 * sum\ of\ squared\ differences)}{n * n^{2-1}} \tag{3.2}$$

The results will range from -1 to 1, where a value of +1 indicates that as one variable increases, the other variable also increases consistently. On the other hand, a value of -1 means that as one variable increases, the other variable decreases consistently. A coefficient of 0 suggests no monotonic relationship between the variables.

This approach only works when the variables are ordinal or ranked. This method is also less sensitive to outlier data.

Point-biserial correlation: This approach measures the strength and direction of association between a continuous variable and a binary variable.

At first, the data is divided into two groups based on the binary variable. One group represents the presence of the characteristic while the other group represents the absence of it. Then, the mean value of the continuous variable is calculated for each group seperately. After that, standard deviation of the whole dataset is calculated, disregarding the grouping which represents the variablity of the continuous variables across the entire dataset. The mean of the group representing the absence of the characteristic is then subtracted from the mean of the group representing the presence of it. We divide the difference by the standard deviation. We then get the point-biserial correlation coefficient which quantifies the strength and direction of the relationship between the continuous and binary variable.
The coefficient value ranges from -1 to +1. A positive value indicates a positive relationship, meaning higher values of the continuous variable tend to be associated with the presence of the characteristic. A negative value indicates a negative relationship, where higher values of the continuous variable tend to be associated with the absence of the characteristic. A value of 0 indicates no relationship between the continuous variable and the binary variable.

This approach assumes that the continuous variable follow a normal distribution or at least approximate normality. Skewed or variable with outliers may affect the accuracy. The observations should also be independent of each other. Larger sample sizes tend to yield more accurate estimates.

For Pearson correlation we use the BMI with continuous values and orders per month which also consists of continuous values. When using Spearman's rank correlation coefficient we use the categorical values of BMI instead and for Point-biserial correlation, we divide the BMI categories in binary groups such as not obese and obese, not normal weight and normal weight and apply the method.

We also apply pearson and spearman correlation on physical activity and orders per month. Treating the values for physical activity as continuous for pearson and categorical for spearman.

3.3 Most Influential Factor Mining

To find factors that are influencing a person to use OFO, we use Ordinary Least Squares (OLS) [28], Decision Tree Regressor [29] and Random Forest Regressor [30].

Ordinary Least Squares: This is a linear least squares method that chooses unknown parameters in a linear regression. This method tries to find the line that fits the data points the best by reducing the total amount of error between the predicted values and the actual values of the dependent variable. The model takes the form of an equation:

$$y = b_0 + b_1x_1 + b_2x_2 + ... + b_nx_n \tag{3.3}$$

where y is the dependent variable, x_1, x_2, ..., x_n are the independent variables, and b_0, b_1, b_2, ..., b_n are the coefficients of the equation. The goal of OLS is to find the values of the coefficients that minimize the difference between the predicted values of y (based on the equation) and the actual values of y.

A system of linear equations are solved to find the coefficients. The matrix representation of this system is $X'Xb = X'y$, here X is the matrix of independent variables, b is the vector of coefficients and y is the vector of dependent variable.

These coefficients can tell us how strong and which way the independent variable affects the dependent variable. The one with the biggest coefficient is the most important factor in influencing the dependent variable.

Decision Tree: This method uses a tree-shaped model to make decisions and predict outcomes. The tree is created by splitting the data into smaller groups based on the values of the independent variables. This process is repeated recursively until we have subsets that can predict the consequences of each decision. At each internal node of the tree, a decision rule is applied to the data, and the tree branches accordingly. Once the tree is built, we used it to identify the most important variables for ordering food online by looking at the variables that are used in the decision rules at the top of the tree.

Random Forest: This is an ensemble method that merges the predictions of multiple decision trees. To use this method, we first divide the data into different subsets. Then, we train multiple decision trees on each subset. Each tree makes its own predictions based on the data it was trained on. Finally, we take the average of all the predictions from these trees to get the final result. Random forest is more robust to overfitting than a single decision tree, and it also allowed us to identify the most important variables by looking at the variables that are used in the decision rules of the majority of the trees.

The most important equation for Random Forest is the feature importance formula which is used to calculate the importance of each feature. The feature importance of a feature is calculated as the average decrease in impurity across all decision trees in the forest. The impurity can be calculated using the Gini Impurity.

$$Gini = 1 - \sum_{i=1}^{m} p_i^2 \tag{3.4}$$

We use the 11 factors which are liking to try new technologies, making impulsive decision, being too busy too cook, disliking towards cooking, finding ordering online to be easy, having variety of options, finding ordering to be inexpensive, because of promotional offers and discounts, quick delivery, food safety and nutritious value. We use orders per month as the dependent variable. We calculate mean and standard deviation of each factors. Then apply OLS, Decision Tree Regressor and Random Forest Regressor. After that, we use the r-squared value to measure the goodness of fit of our regression models. The r-squared value for OLS, Decision Tree and Random Forest are 0.64, 0.01 and 0.04. Which means OLS is a better approach for our data.

3.4 Classifying Frequent users of OFO services

To create classify frequent users of OFO services we use several methods such as Logistic Regression [31], Naive Bayes [32], Decision Tree Classifier [29], Random Forest Classifier [30], Gradient Boosting Machines (GBM) [33] and K-Nearest Neighbour (KNN) [34].

Logistic Regression: This is a commonly used statistical model for tasks that involve classifying things into two categories. The main aim is to predict the probability of an instance belonging to a specific class. Unlike linear regression, which predicts continuous values, it is specifically designed to predict the probability of an event happening. Instead of providing a direct numerical output, this method calculates the likelihood of an event occurring based on the input variables. This makes it suitable for tasks where we are interested in determining the likelihood or probability of a specific outcome.

Logistic Regression models the relationship between the predictor variables and the output class probabilities using a logistic function or sigmoid function. The sigmoid function is defined as:

$$S(z) = \frac{1}{1 + e^{-z}} \tag{3.5}$$

Here, z represents a linear combination of the input features and model parameters.

The model is trained by finding the optimal values for the model parameters which minimize the error between the predicted probabilities and the actual class labels in the training data. This is typically done using optimization algorithms such as Newton's method or Gradient Descent. The objective is to minimize the log loss of the predicted probabilities or maximize the likelihood of it.

Once the model is trained and the optimal parameters are obtained, we can use the model to make predictions on new, unseen data. The sigmoid function is applied to the linear combination of the input features and parameters to obtain the predicted probability. A common threshold (0.5) is then used to classify the instance into one of the two classes based on the predicted probability. If the predicted probability is above the threshold, it is classified as class 1; otherwise, it is classified as class 0.

Naive Bayes: This algorithm is built upon the Bayes' theorem and assumes that given the class label, the features are conditionally independent of each other.

This algorithm starts by calculating the prior probabilities of each class in the dataset. The prior probability of a class is the probability of encountering that class in the dataset without considering any feature information.

For each feature in the dataset, this method calculates the conditional probability of observing that feature given a specific class label. This is done by counting the occurrences of each feature in the training samples belonging to a particular class.

Once the prior probabilities and conditional probabilities are computed, Bayes' theorem is applied to calculate the posterior probability of each class given the observed features.

Finally, the classifier predicts the class label for a new, unseen sample by selecting the class with the highest posterior probability.
The algorithm for Decision Tree and Random Forest classifier works in the same as the regressor approach but the classifier is more suitable for categorical or discrete class labels.

Gradient Boosting Machines: This algorithm combines multiple weak predictive models (usually decision trees) to create a better predictive model. It is an ensemble learning method that sequentially builds an additive model by minimizing a predefined loss function using gradient descent.

A weak learner which usually is a decision tree with a small depth which is also called a decision stump, is fitted to the training data. The tree is trained to minimize the loss function with respect to the target variable.

The predictions of the weak learner are then subtracted from the true values of the target variable to obtain the residuals. The residuals represent the errors or the parts of the target variable that the model has not yet captured.

Another weak learner is then fitted to the residuals obtained from the previous step. The goal is to find a new model that can capture the remaining patterns in the data that the previous model missed.

The new weak learner is now added to the ensemble by combining it with the previous models. To combine the models, a weight is assigned to each weak learner, this is also known as the learning rate. The weight determines the contribution of each model to the final prediction.

The process of calculating residuals, fitting the next learner and updating of the model is repeated iteratively for a specified number of times or until a predefined stopping criterion is met. In each iteration, a new weak learner is fitted to the negative gradients of the ensemble model built so far.

The final prediction is gained by summing the predictions of all the weak learners with their respective weights. The model assigns higher weights to the more accurate learners, which allows it to give more importance to their predictions.

K-Nearest Neighbour: KNN is an algorithm that does not rely on any assumptions about the underlying data distribution. The fundamental idea behind KNN is to classify or predict the target variable of a new data point based on the majority vote or the average of the target variables of its K number of nearest neighbors in the feature space.

At first, the value of K is figured out, which represents the number of nearest neighbors to consider for classification or regression. The optimal value of K can be found using techniques such as cross-validation or grid search.

To classify or predict a new data point, the distance between that point and every other data points in the training data is calculated. The distance can be calculated using different metrics such as Manhattan distance and Euclidean distance.

The distances are then sorted in ascending order and the K data points are selected with the shortest distances as the nearest neighbors.

Then, the class label is determined that occurs most frequently among the K nearest neighbors. This can be done by counting the occurrences of each class label and selecting the one with the highest count.

Once the class label or predicted value has been determined, it is assigned as the output for the new data point.

We divide the orders per month variable into two category. One with less or equal to 5 orders per month which covers 56% of the sample and another with more than 5 orders per month which covers the rest of 44% of the sample. We then use this as the target label against all the features available in our dataset. We scale the data, select the best features using correlation heat map and implement the algorithms. We use uni variate methods such as chi-squared test and ANOVA F-value to do some feature selection.

Table 3.11: Results of feature selection

Chi-squared Test	ANOVA F-value
Gender	Gender
BMI	BMI
Employment Status	Employment Status
Financial Dependency	Financial Dependency
Physical Activity	Marital Status
Ordering History	Physical Activity
Impulsive Decision	Ordering History
Busy To Cook	Impulsive Decision
Dont Like To Cook	Busy To Cook
Ordering Easy	Dont Like To Cook
Options To Choose	Ordering Easy
Ordering Inexpensive	Options To Choose
Promo Discounts	Promo Discounts
Quick Delivery	Quick Delivery
Food Safety	Food Safety
Nutritious	Nutritious

From table 3.11, we can see that 15 out of the 16 selected features are common for both methods. Therefore, we use select these 15 to build our prediction models.

We use several evaluation metrics like accuracy score, precision, recall and F1-score to find the model with best results.

3.5 Customer Segmentation

Customer segmentation is a technique in data analysis to group customers based on their similarities and differences. One way of doing this is using K-Means clustering algorithm [35], which is an unsupervised machine learning algorithm that aims to divide a given dataset into K distinct clusters, where each data point goes to the cluster with the nearest mean value. Principal Component Analysis (PCA) [36] is a technique that can be used to reduce the dimension of the dataset.

K-Means: The algorithm works by assigning data points iteratively to the cluster with the closest centroid and then updating the centroids to be the average of the data points in each cluster. This process continues until the centroids no longer change significantly.

Number of clusters K is usually chosen by using elbow method or silhouette scores. Then the the centroids of the clusters are initialized. This is done randomly or by using some other method, such as k-means++. Each data point then gets assigned to the cluster with the closest centroid. The centroids get updated to be the average of the data points in each cluster. This is repeated until the centroids no longer change significantly.

Principal Component Analysis: This is a widely used statistical technique for dimensionality reduction and data analysis. It aims to find the directions or principal components in which the data varies the most and represents the data in a lower-dimensional space while preserving the essential information.

At first, dataset is standardized to have a unit variance and zero mean. This step ensures that all features are on a similar scale and prevents variables with large variances from dominating the analysis. Then the covariance matrix of the standardized data is computed. The covariance matrix indicates the relationships between different features and measures how they vary together. The value in the i^{th} row and j^{th} column of the covariance matrix represents the covariance between the i^{th} and j^{th} features. An eigendecomposition is performed on the covariance matrix to find its eigenvalues and corresponding eigenvectors. The amount of variance explained by each principal component is represented by the eigenvalues. The eigenvectors represent the directions of these components. The eigenvalues indicate the relative importance of each principal component. PCA typically sorts the eigenvalues in descending order and selects the top-k eigenvectors corresponding to the largest eigenvalues, where k is the desired number of dimensions in the reduced space. The selected eigenvectors form a new coordinate system. The original data is projected onto this new coordinate system to obtain the reduced-dimensional representation. This projection involves multiplying the standardized data matrix by the eigenvector matrix, resulting in a transformed dataset with reduced dimensions.

We use only the 11 factors we developed that influence OFO to do customer segmentation. We apply this using both Principal Component Analysis (PCA) and one without using it. Before applying PCA, we use Kaiser-Meyer-Olkin test (KMO) [37] and Bartlett's Test of Sphericity [38] to check if the data is suitable to apply PCA. The results of KMO test gives us 0.85 which makes it sufficient to perform factor analysis according to this paper [39]. The result of Bartlett's Test of Sphericity gives us a p-value of 1.06^{-228} which is lower than the α value of 0.05 and a high chi-squared value of 1301.43. Therefore, we can apply PCA on our data.

We use elbow method and silhouette score to find optimal number of clusters. From figures 3.6, 3.8, 3.7 and 3.9 we can see that the optimal number of clusters is 2.

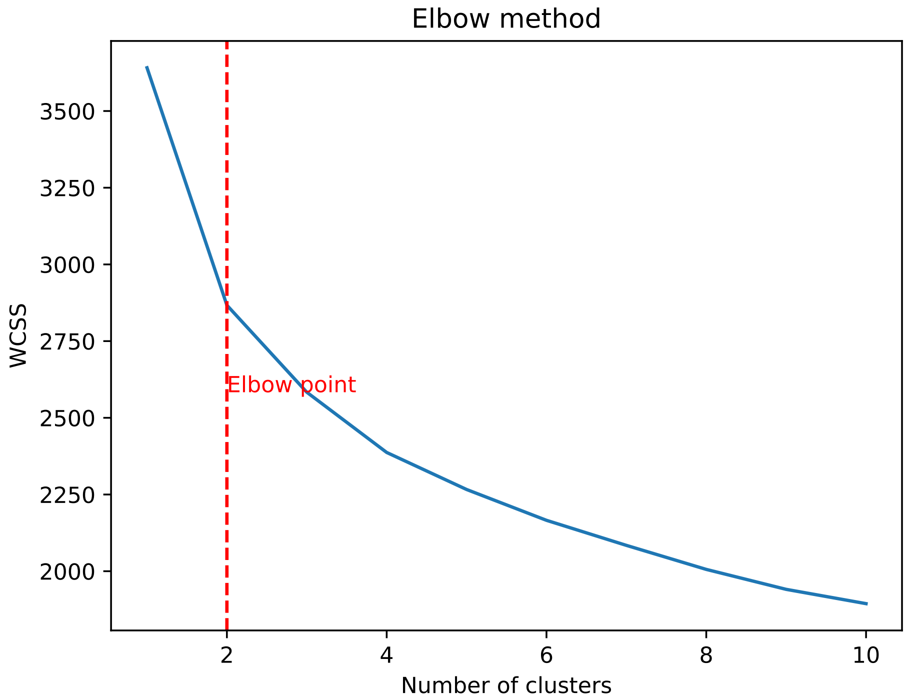

Figure 3.6: Elbow method for determining the optimal number of clusters.

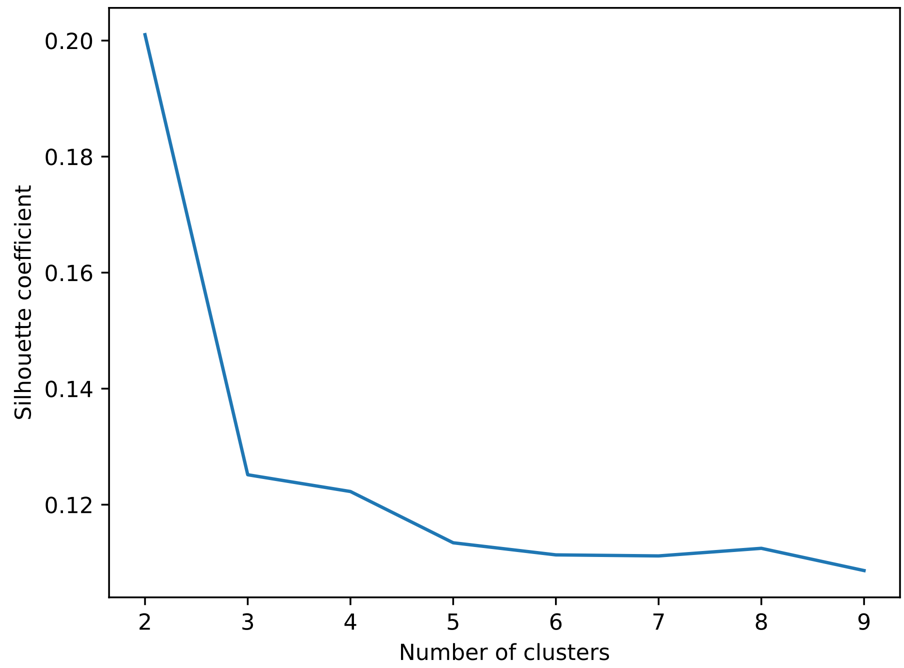

Figure 3.7: Silhouette plot for determining the optimal number of clusters.

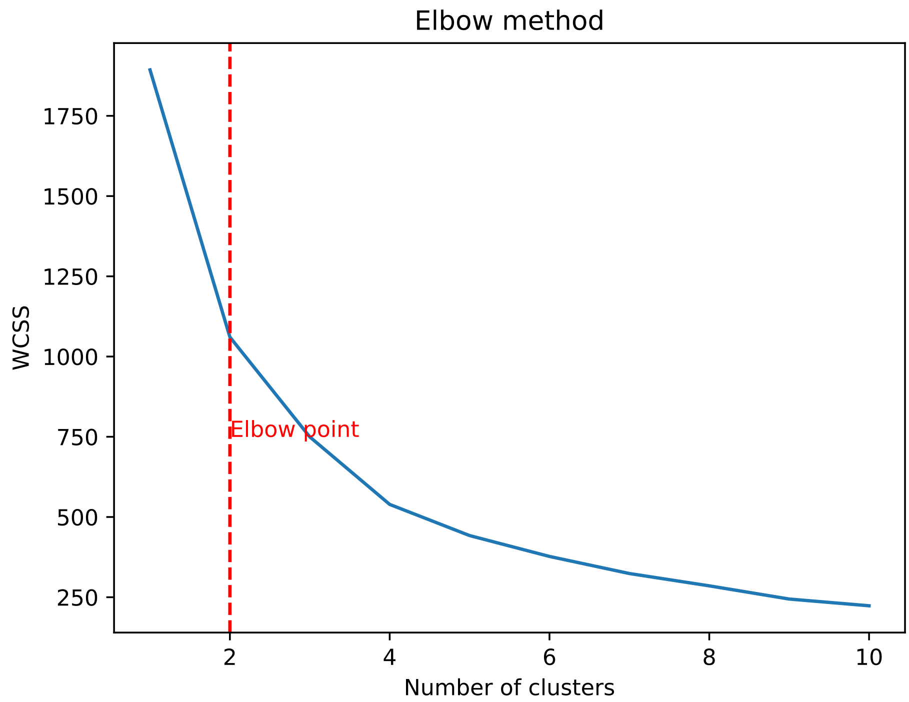

Figure 3.8: Elbow method for determining the optimal number of clusters with PCA.

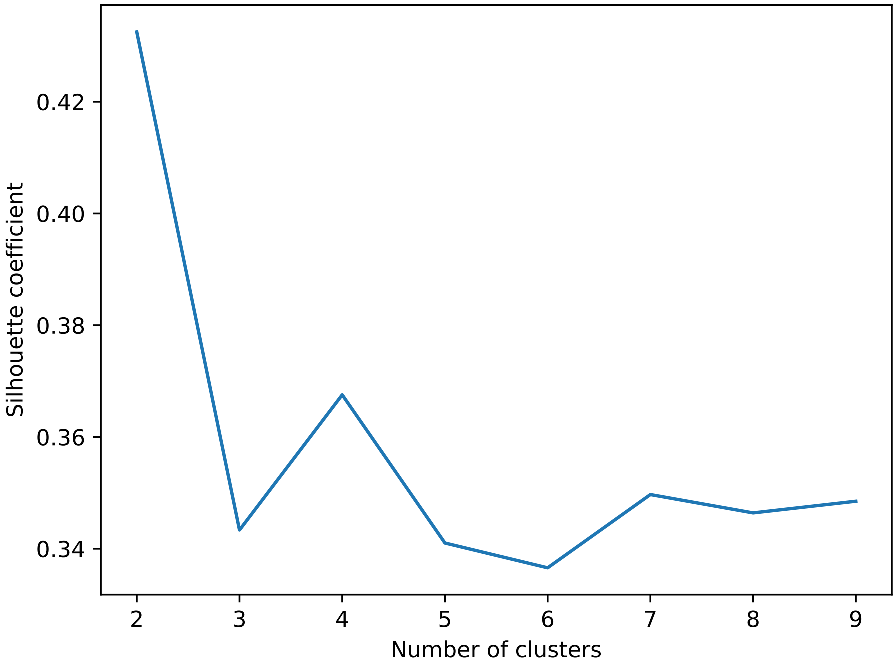

Figure 3.9: Silhouette plot for determining the optimal number of clusters with PCA.

Table 3.12: Comparison of results of evaluation metrics for K-Means clustering with and without PCA.

Metric	With PCA	Without PCA
Silhouette Coefficient ↑	0.43	0.21
Calinski-Harabasz Index ↑	258.27	88.82
Davies-Bouldin Index ↓	0.87	1.71

We use evaluation metrics such as the Silhouette Coefficient, Calinski-Harabasz Index [40] and Davies-Bouldin Index [41] to choose the one with best performance. In table 3.12, we can see that applying PCA gives us a comparitively a better evaluation scores on all of the three metrics.

3.6 Hypothesis Testing

In our study, we also conduct some hypothesis testing using the chi-squared test [42].

Chi-Squared Test: This test is used to determine if there is a significant difference between an observed distribution and a theoretical distribution. The test statistic is calculated using the following equation:

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \tag{3.6}$$

Where:
O = observed frequency
E = expected frequency

The degrees of freedom for the test is calculated as:

$$df = (number\, of\, rows - 1) * (number\, of\, columns - 1) \tag{3.7}$$

The p-value is then looked up in a chi-squared distribution table with the corresponding degrees of freedom.

If the calculated p-value is less than the chosen significance level (usually 0.05), then the null hypothesis (that there is no significant difference between the observed and theoretical distributions) is rejected, and it is concluded that there is a significant difference between the two.

We use the chi-squared test of independence as the factors have ordinal values. So, something like Pearson's correlation coefficient might not be the best measure of testing here as it assumes that the data is normally distributed and continuous.

We test the following hypotheses for our study,

People who are open to experimenting with new technology such as OFO services, may find it to be effortless to place an order. These platforms offer easy to use and intuitive inteface which simplifies the ordering process, thus allowing users to place orders quickly without any sort of hassle. Moreover, these applications often provide the option to track orders in real time, personalized recommendations. All of these features can enhance the overall ordering experience and make it more convenient for the users to use. Based on this argument we come up with the following hypothesis:

H1: There is a significant relationship between trying out new technology and thinking that using OFO service is easy.

People who have busy schedules may be more likely to make impulsive decisions when it comes to using OFO services. Due to their busy lifestyle, they may not have the energy to shop for ingredients or cook at home. Therefore, they may turn to quick and easy options such as ordering takeout. Based on this argument we come up with the following hypothesis:

H2: There is a significant relationship between making impulsive decisions when using OFO service and being too busy to cook.

People who do not like to cook may appreciate the variety of options that are available on the OFO platforms. For them, the thought of preparing meals at home can be unappealing or they might prefer to have a variety of options to choose from when ordering food. Based on this argument we come up with the following hypothesis:

H3: There is a significant relationship between people not liking to cook and finding there are many options to choose from when using OFO service.

Promotional offers or discounts can make the use of OFO services more cost-effective for users. These offers can take various forms such as coupons, loyalty programs, and special deals. These can be used to reduce the cost of items, delivery fees or sometimes even the entire order. By making use of these discounts, customers can save money on their orders. Therefore leading them to believe that ordering food online is inexpensive. Based on this argument we come up with the following hypothesis:

H4: There is a significant relationship between thinking OFO service is inexpensive and ordering food because of promo codes and discounts.

Chapter 4

Results and Discussions

In this chapter, we share our findings from the various experiments we have conducted.

4.1 Correlation Analysis

Table 4.1: Pearson and Spearman Correlation Test Results of BMI vs Orders Per Month

Test	P-value	Coefficient	Relationship
Pearson	2.43^{-7}	0.28	Weak Positive
Spearman	0.003	0.15	Very Weak Positive

The results from Pearson correlation and Spearman's rank correlation which was done between BMI and orders per month, gives us p-value of 2.43^{-7} and 0.003 respectively, which suggests that there exists a statistically significant positive correlation between the two variables since both these values are less than 0.05. The coefficients values are 0.28 and 0.15 respectively which suggests there exists a weak positive correlation and a very weak positive correlation [43]. Although this does not indicate that there is necessarily any causation here.

Table 4.2: Results of Point-biserial Correlation on BMI vs Orders Per Month

BMI	P-value	Coefficient	Relationship
Underweight	0.84	-0.01	Rejected
Normal	1.97^{-5}	-0.23	Weak Negative
Overweight	0.04	0.04	Very Weak Positive
Obese	1.13^{-10}	0.34	Weak Positive

The results of shown on table 4.2 suggests that due to insufficient p-value any relation between Underweight and orders per month is rejected, normal weight has a weak negative relationship, overweight has a very weak positive relationship and obese has a weak positive relationship.

Table 4.3: Results of Correlation analysis between physical activity and orders per month

Test	P-value	Coefficient	Relationship
Pearson	0.026	-0.12	Very Weak Negative
Spearman	0.085	-0.10	Rejected

As seen on table 4.3, we can see that there exists a very weak negative relation between physical activity and orders per month which suggests that people who tend to order more may exercise less. However, the results from spearman's test is rejected as the p-value is 0.085, which is greater than the α value of 0.05.

Table 4.4: Results of Point-biserial Correlation on Ordering History and Orders Per Month

Ordering History	P-value	Coefficient	Relationship
Less than 3 months	1.40^{-5}	-0.23	Weak Negative
3 to 6 months	0.66	0.02	Rejected
More than 6 months	<0.0001	0.18	Very Weak Positive

The results of correlation analysis between ordering history and orders per month as seen on table 4.4 shows that people who have been using OFO for more than 6 months tend to order more. The coefficient value of -0.23 suggests that there exists a weak relationship between ordering frequency and people who have been using this type of service for less than 3 months. However, due to the p-value (0.66) being higher than the α value of 0.05, any relationship between people who have been ordering for 3-6 months is rejected.

4.2 Most Influencing Factors

The results of the experiments which are done to find the most influencing factors are shown in table 4.5 gives us an idea on which factors influence a person to use OFO services.

Table 4.5: Influential Factors in Online Food Orders: Comparing Mean, Standard Deviation, OLS, Decision Tree, and Random Forest Values

Factors	Mean	SD	OLS	DT	RF
New Tech	2.23	1.09	0.39	0.01	0.22
Impulsive Decision	2.01	1.16	0.90	0.34	0.22
Busy To Cook	2.12	1.16	0.18	0.05	0.10
Don't Like To Cook	1.82	1.25	0.33	0.01	0.13
Ordering Easy	2.98	0.95	0.65	0.11	0.08
Options To Choose	2.98	1.03	0.15	0.01	0.06
Ordering Inexpensive	1.62	1.10	0.27	0.22	0.07
Promo Discounts	2.60	1.12	0.33	0.01	0.09
Quick Delivery	2.28	1.00	0.11	0.01	0.05
Food Safety	1.79	0.86	-0.40	0.01	0.05
Nutritious	1.45	0.98	-0.32	0.28	0.09

The mean values based on solely the responses from the respondents suggests that the most important factors in influencing people's decision to order food online are the convenience of ordering (mean score of 2.98), the variety of options available (mean score of 2.98), and promotional offers and discounts (mean score of 2.60).

The results of OLS show that the most influencing factors are making impulsive decision (coefficient value of 0.90), convenience of ordering (coefficient value of 0.65). On the other hand, food safety (coefficient value of -0.40) and nutritious value of the food (coefficient value of -0.36) have a negative impact.

4.3 Classifying Frequent users of OFO services

The results of the prediction model can be seen on table 4.6.

Table 4.6: Comparison of results among prediction models before feature selection

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.78	0.66	0.59	0.61
Naive Bayes	0.64	0.53	0.53	0.53
Decision Tree	0.63	0.45	0.45	0.45
Random Forest	0.81	0.78	0.59	0.60
Gradient Boosting Machine	0.78	0.66	0.59	0.61
K-Nearest Neighbor	0.78	0.67	0.64	0.65

We can see that out of 6 models, Random Forest performs the best with 81% accuracy. whereas Naive Bayes performed worst with 64% accuracy. Random Forest also has the best precision score of 0.78 which means it has a high level of precision in identifying positive instances. On the other hand, KNN correctly identified 64%

of the positive instances as seen on the Recall column.

Table 4.7: Comparison of results among prediction models after feature selection

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.81	0.74	0.61	0.63
Naive Bayes	0.67	0.55	0.55	0.55
Decision Tree	0.61	0.49	0.49	0.49
Random Forest	0.81	0.78	0.59	0.60
Gradient Boosting Machine	0.78	0.65	0.57	0.58
K-Nearest Neighbor	0.81	0.73	0.64	0.66

We can see on table 4.7, that after applying feature selection the performance of Logistic Regression, Naive Bayes and KNN improves. Although, performance of Decision Tree decreases and that of Random Forest and Gradient Boosting Machine stays the same as before. This can also be seen on figure 4.1.

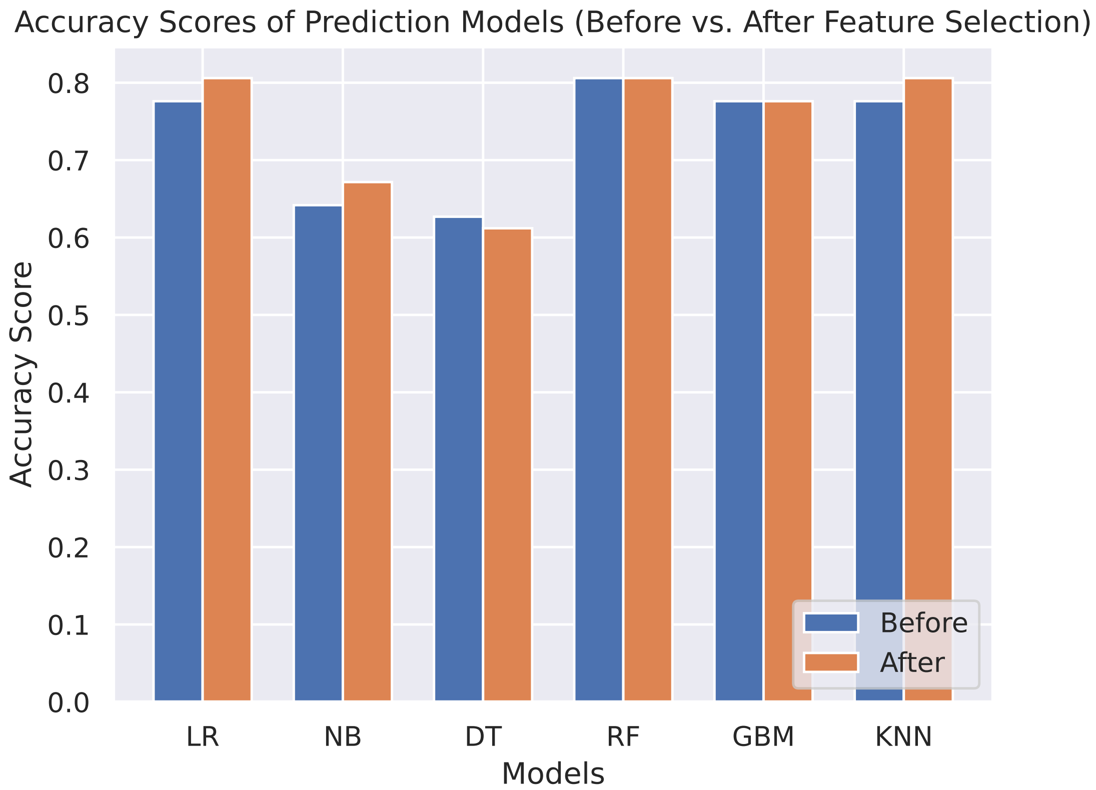

Figure 4.1: Comparison between results of prediction models before and after applying feature selection

4.4 Customer Segmentation

Table 4.8: Comparison between cluster means between the clusters created with and without PCA.

Factors	Cluster 0		Cluster 1	
	With PCA	Without PCA	With PCA	Without PCA
New Tech	1.45	1.67	2.47	2.48
Impulsive Decision	1.19	1.40	2.25	2.27
Busy To Cook	1.47	1.61	2.31	2.34
Dont Like To Cook	1.08	1.31	2.04	2.04
Ordering Easy	1.95	2.04	3.29	3.38
Options To Choose	1.75	1.91	3.35	3.44
Ordering Inexpensive	0.83	0.97	1.86	1.90
Promo Discounts	1.53	1.62	2.92	3.02
Quick Delivery	1.29	1.43	2.59	2.65
Food Safety	1.17	1.27	1.98	2.02
Nutritious	1.29	1.91	1.64	1.66

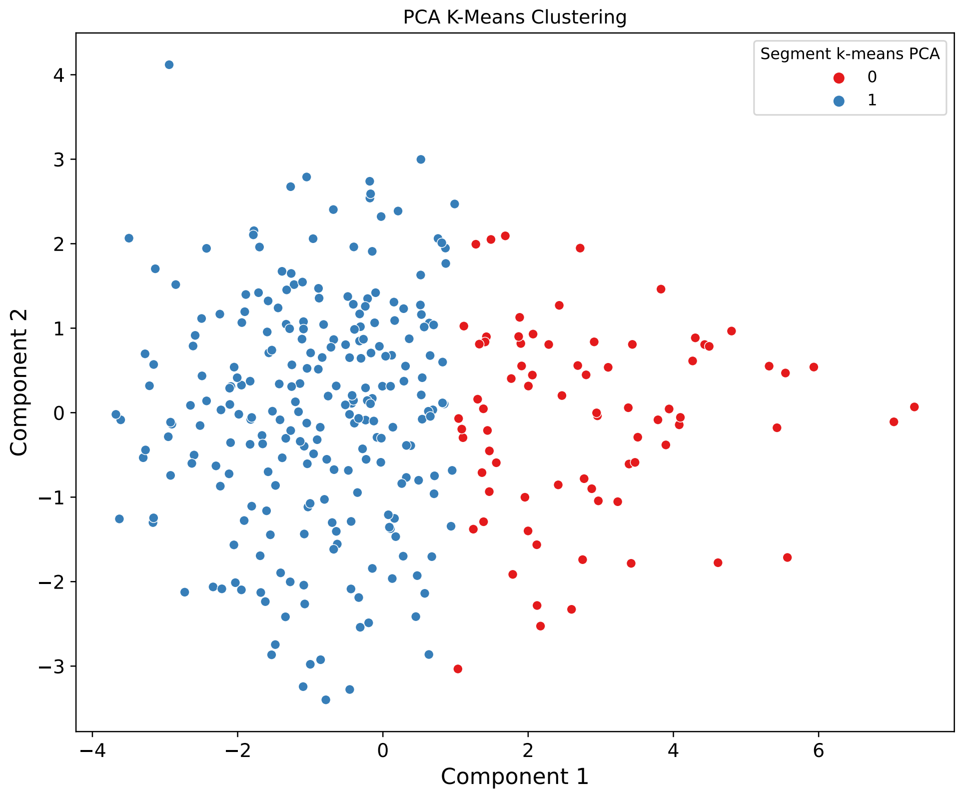

Figure 4.2: Scatter plot of clusters after using PCA

On table 4.8, we have the mean values of each cluster. We show the values of both with PCA applied and without it. On cluster 0 which contains respondents who tend to order less and cluster 1 represents the people who order comparatively more. Respondents in both clusters show the tendency to order food because it is convenient, there are variety of options to choose from and due to the availability of promotional offers and discounts. However, we can see that people of cluster 0 find food delivered by OFO services to not be inexpensive with the mean value for 'Ordering Inexpensive' as 0.83 and 0.97. Whereas, people belonging in cluster 1 tend to find the food from the OFO services to lack nutritious values with mean value for 'Nutritious' as 1.64 and 1.66.

4.5 Hypothesis Testing

Table 4.9: Results of Hypothesis Testing

Hypothesis	Relation	P-Value	Decision
H1	New Tech - Ordering Easy	< 0.001	Accepted
H2	Busy To Cook - Impulsive Decision	0.114	Rejected
H3	Options To Choose - Don't Like To Cook	< 0.001	Accepted
H4	Ordering Inexpensive - Promo Discounts	< 0.001	Accepted

The results of the hypothesis study point out that the first hypothesis, H1 is accepted and the null hypothesis is rejected as the p-value is <0.001, which is less than ($\alpha = 0.05$). This means that people who tend to try out new technologies also feel that ordering food online is really convenient and easy.

For the second hypothesis, H2 is rejected as the p-value (0.1143) is higher than the α (0.05). This means we could not find any significant relationship between being too busy to cook and making impulsive decisions when ordering food online.

The third hypothesis, H3 is accepted and the null hypothesis is rejected as the p-value (< 0.001) is lower than the α (0.05). This means there is a significant relationship between having many options to choose from when ordering food online and not liking to cook.

And the fourth hypothesis, H4 is accepted and the null hypothesis is rejected as the p-value (< 0.001) is less than the α (0.05). This means that there exists a significant relationship between finding ordering food online to be inexpensive and using promo codes and discounts when ordering food online.

4.6 Discussions

The results of the most influential factor mining supports some of the findings from previous works. The high coefficient value on impulsive decision supports [11] as they also find that hedonic motivation is one of the important factors. The positive impact of ease of convenience as found by [10], [11] is supported by our results as well. We found that food safety is a major concern as it has a negative impact when using OFO services, this is also supported by [16]. The results of correlation analysis between BMI and orders per month shows a weak positive correlation, but we would like to clarify that this does not necessarily mean that there exists any causation. This finding can be used in future researches to combine with other factors to see if any causality actually exists.

Chapter 5

Conclusion

As online food ordering platforms are rising in popularity, it has become one of the most common ways of purchasing food. This research tried to investigate the factors that influence customers to use online food ordering platforms in the context of Bangladesh. We were able to create a dataset containing the activities of OFO service users, which is non existent specially in Bangladeshi context. Our findings reveal that there exists a weak positive correlation between BMI and orders per month. It also shows that people who tend to order more, happen to be obese. However, it is important to note that a correlation does not necessarily imply causation. We found that making impulsive decision, convenience of ordering and the availability of promotional offers and discounts are the most important factors behind the frequency of ordering food online. We also were able to successfully classifiy frequent users of OFO services with 81% accuracy.

However, our research has its limitations. First of all, the sample size of our data is only 331. This means that our results may not be generalizable to the larger population. Additionally, our data was collected from people of ages 16-30, which means that it may not accurately reflect the characteristics and perspectives of the entire population. Moreover, our study was cross-sectional, which means that we cannot make causal inferences from our results. Future research could be done with even bigger data with more factors to consider. Another research opportunity is to include more factors such as advertisements, spending level to get better insights on this topic.

Despite the limitations, we believe that this study provides valuable insights into the online food ordering behaviour of individuals in Bangladesh. Given the scarcity of research in this area, particularly in the context of Bangladesh, we believe that our findings will make a significant contribution to future studies on online food ordering.

Bibliography

[1] *Obesity and overweight — who.int*, https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight.

[2] *Adult rates of overweight and obesity rise in bangladesh*, https://www.icddrb.org/dmdocuments/Bangladesh%20obesity%20release.pdf.

[3] *Bangladesh — data.worldobesity.org*, https://data.worldobesity.org/country/bangladesh-16/.

[4] *Food Delivery Mobile Application Market Statistics | 2030 — alliedmarketresearch.com*, https://www.alliedmarketresearch.com/food-delivery-mobile-application-market#:~:text=The%20global%20food%20delivery%20mobile,25%25%20from%202021%20to%202030..

[5] *World Obesity Day 2022 – Accelerating action to stop obesity — who.int*, https://www.who.int/news/item/04-03-2022-world-obesity-day-2022-accelerating-action-to-stop-obesity.

[6] NHS, *Obesity - Causes — nhs.uk*, https://www.nhs.uk/conditions/obesity/causes/.

[7] A. Hruby, J. E. Manson, L. Qi, *et al.*, "Determinants and consequences of obesity," *American Journal of Public Health*, vol. 106, no. 9, pp. 1656–1662, Sep. 2016. DOI: 10.2105/ajph.2016.303326. [Online]. Available: https://doi.org/10.2105/ajph.2016.303326.

[8] C. Li, M. Mirosa, and P. Bremer, "Review of online food delivery platforms and their impacts on sustainability," *Sustainability*, vol. 12, no. 14, p. 5528, Jul. 2020. DOI: 10.3390/su12145528. [Online]. Available: https://doi.org/10.3390/su12145528.

[9] *What Are The Benefits Of Online Ordering Food — theotherstream.com*, https://www.theotherstream.com/what-are-the-benefits-of-online-ordering-food/.

[10] G. Kale and S. Chourishi, "Title the customer perception on online food ordering and its significance," Jun. 2020.

[11] S. M. Alagoz and H. Hekimoglu, "A study on tam: Analysis of customer attitudes in online food ordering system," *Procedia - Social and Behavioral Sciences*, vol. 62, pp. 1138–1143, Oct. 2012. DOI: 10.1016/j.sbspro.2012.09.195. [Online]. Available: https://doi.org/10.1016/j.sbspro.2012.09.195.

[12] I. Jahidi, N. A. Ruyani, and D. P. Alamsyah, "The study of consumer behavior on online food ordering system (go-food) in the metropolitan city," Sep. 2022. DOI: 10.20944/preprints202209.0085.v1. [Online]. Available: https://doi.org/10.20944/preprints202209.0085.v1.

[13] R. Parameshwaran and K. Krishnasamy, "A study on factors influencing online consumer buying behavior," *International Journal of Scientific & Technology Research*, vol. 9, pp. 3620–3625, Feb. 2020.

[14] S. Anam, D. Golam Yazdani Showrav, M. A. Hassan, and A. Chakrabarty, "Factors influencing the rapid growth of online shopping during covid-19 pandemic time in dhaka city, bangladesh," vol. 20, pp. 1–13, Jul. 2021.

[15] C. Inthong, T. Champahom, S. Jomnonkwao, V. Chatpattananan, and V. Ratanavaraha, "Exploring factors affecting consumer behavioral intentions toward online food ordering in thailand," *Sustainability*, vol. 14, no. 14, p. 8493, Jul. 2022. DOI: 10.3390/su14148493. [Online]. Available: https://doi.org/10.3390/su14148493.

[16] C. Hong, H. (Choi, E.-K. (Choi, and H.-W. (Joung, "Factors affecting customer intention to use online food delivery services before and during the COVID-19 pandemic," *Journal of Hospitality and Tourism Management*, vol. 48, pp. 509–518, Sep. 2021. DOI: 10.1016/j.jhtm.2021.08.012. [Online]. Available: https://doi.org/10.1016/j.jhtm.2021.08.012.

[17] A. Chatterjee, M. W. Gerdes, and S. G. Martinez, "Identification of risk factors associated with obesity and overweight—a machine learning overview," *Sensors*, vol. 20, no. 9, p. 2734, May 2020. DOI: 10.3390/s20092734. [Online]. Available: https://doi.org/10.3390/s20092734.

[18] A. M. Prentice and S. A. Jebb, "Fast foods, energy density and obesity: A possible mechanistic link," *Obesity Reviews*, vol. 4, no. 4, pp. 187–194, Nov. 2003. DOI: 10.1046/j.1467-789x.2003.00117.x. [Online]. Available: https://doi.org/10.1046/j.1467-789x.2003.00117.x.

[19] J. M. Dave, L. C. An, R. W. Jeffery, and J. S. Ahluwalia, "Relationship of attitudes toward fast food and frequency of fast-food intake in adults," *Obesity*, vol. 17, no. 6, pp. 1164–1170, Jun. 2009. DOI: 10.1038/oby.2009.26. [Online]. Available: https://doi.org/10.1038/oby.2009.26.

[20] E. Martha, D. Ayubi, B. Besral, *et al.*, "Online food delivery services among young adults in depok: Factors affecting the frequency of online food ordering and consumption of high-risk food," Dec. 2021. DOI: 10.21203/rs.3.rs-1103144/v1. [Online]. Available: https://doi.org/10.21203/rs.3.rs-1103144/v1.

[21] L. A. H. Harahap, E. Aritonang, and Z. Lubis, "The relationship between type and frequency of online food ordering with obesity in students of medan area university," *Britain International of Exact Sciences (BIoEx) Journal*, vol. 2, no. 1, pp. 29–34, Jan. 2020. DOI: 10.33258/bioex.v2i1.109. [Online]. Available: https://doi.org/10.33258/bioex.v2i1.109.

[22] N. D. Kurniawati, S. N. Cahyaningsih, and A. S. Wahyudi, "The correlation between online food ordering and nutritional status among college students in surabaya," *Indonesian Journal of Community Health Nursing*, vol. 6, no. 2, p. 70, Nov. 2021. DOI: 10.20473/ijchn.v6i2.27520. [Online]. Available: https://doi.org/10.20473/ijchn.v6i2.27520.

[23] C. G. Forero, “Cronbach’s alpha,” in *Encyclopedia of Quality of Life and Well-Being Research*, A. C. Michalos, Ed. Dordrecht: Springer Netherlands, 2014, pp. 1357–1359, ISBN: 978-94-007-0753-5. DOI: 10.1007/978-94-007-0753-5_622. [Online]. Available: https://doi.org/10.1007/978-94-007-0753-5_622.

[24] K. S. Taber, “The use of cronbach’s alpha when developing and reporting research instruments in science education,” *Research in Science Education*, vol. 48, no. 6, pp. 1273–1296, 2017. DOI: 10.1007/s11165-016-9602-2.

[25] D. Freedman, R. Pisani, and R. Purves, “Statistics (international student edition),” *Pisani, R. Purves, 4th edn. WW Norton & Company, New York*, 2007.

[26] “Spearman rank correlation coefficient,” in *The Concise Encyclopedia of Statistics*. New York, NY: Springer New York, 2008, pp. 502–505, ISBN: 978-0-387-32833-1. DOI: 10.1007/978-0-387-32833-1_379. [Online]. Available: https://doi.org/10.1007/978-0-387-32833-1_379.

[27] L. J.M, *The expected value of a point-biserial (or similar) correlation*, 2008. [Online]. Available: https://www.rasch.org/rmt/rmt221e.htm.

[28] B. Zdaniuk, “Ordinary least-squares (ols) model,” in *Encyclopedia of Quality of Life and Well-Being Research*, A. C. Michalos, Ed. Dordrecht: Springer Netherlands, 2014, pp. 4515–4517, ISBN: 978-94-007-0753-5. DOI: 10.1007/978-94-007-0753-5_2008. [Online]. Available: https://doi.org/10.1007/978-94-007-0753-5_2008.

[29] J. Fürnkranz, “Decision tree,” in *Encyclopedia of Machine Learning*, C. Sammut and G. I. Webb, Eds. Boston, MA: Springer US, 2010, pp. 263–267, ISBN: 978-0-387-30164-8. DOI: 10.1007/978-0-387-30164-8_204. [Online]. Available: https://doi.org/10.1007/978-0-387-30164-8_204.

[30] T. K. Ho, “Random decision forests,” in *Proceedings of 3rd international conference on document analysis and recognition*, IEEE, vol. 1, 1995, pp. 278–282.

[31] D. R. Cox, “The regression analysis of binary sequences,” *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 20, no. 2, pp. 215–232, 1958.

[32] G. I. Webb, “Naïve bayes,” in *Encyclopedia of Machine Learning*, C. Sammut and G. I. Webb, Eds. Boston, MA: Springer US, 2010, pp. 713–714, ISBN: 978-0-387-30164-8. DOI: 10.1007/978-0-387-30164-8_576. [Online]. Available: https://doi.org/10.1007/978-0-387-30164-8_576.

[33] J. H. Friedman, “Greedy function approximation: A gradient boosting machine,” *Annals of statistics*, pp. 1189–1232, 2001.

[34] A. Mucherino, P. J. Papajorgji, and P. M. Pardalos, “K-nearest neighbor classification,” in *Data Mining in Agriculture*. New York, NY: Springer New York, 2009, pp. 83–106, ISBN: 978-0-387-88615-2. DOI: 10.1007/978-0-387-88615-2_4. [Online]. Available: https://doi.org/10.1007/978-0-387-88615-2_4.

[35] X. Jin and J. Han, “K-means clustering,” in *Encyclopedia of Machine Learning*, C. Sammut and G. I. Webb, Eds. Boston, MA: Springer US, 2010, pp. 563–564, ISBN: 978-0-387-30164-8. DOI: 10.1007/978-0-387-30164-8_425. [Online]. Available: https://doi.org/10.1007/978-0-387-30164-8_425.

[36] I. Jolliffe, "Principal component analysis," in *International Encyclopedia of Statistical Science*, Springer Berlin Heidelberg, 2011, pp. 1094–1096. DOI: 10.1007/978-3-642-04898-2_455. [Online]. Available: https://doi.org/10.1007/978-3-642-04898-2_455.

[37] H. F. Kaiser, "A second generation little jiffy," *Psychometrika*, vol. 35, no. 4, pp. 401–415, Dec. 1970. DOI: 10.1007/bf02291817. [Online]. Available: https://doi.org/10.1007/bf02291817.

[38] M. S. Bartlett, "Tests of significance in factor analysis.," *British journal of psychology*, 1950.

[39] A. Zbiciak and T. Markiewicz, "A new extraordinary means of appeal in the polish criminal procedure: The basic principles of a fair trial and a complaint against a cassatory judgment," en, *Access to Justice in Eastern Europe*, vol. 6, no. 2, pp. 1–18, Mar. 2023.

[40] C. T. Harabasz and M. Karoński, "A dendrite method for cluster analysis," in *Communications in Statistics*, 1, vol. 3, 1974, pp. 1–27.

[41] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE transactions on pattern analysis and machine intelligence*, no. 2, pp. 224–227, 1979.

[42] K. Pearson, "X. on the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 50, no. 302, pp. 157–175, Jul. 1900. DOI: 10.1080/14786440009463897. [Online]. Available: https://doi.org/10.1080/14786440009463897.

[43] Z. Jaadi, *Eveything you need to know about interpreting correlations*, Oct. 2019. [Online]. Available: https://towardsdatascience.com/eveything-you-need-to-know-about-interpreting-correlations-2c485841c0b8.

Survey Questions

Below we share the questions from the survey:

1. Please, choose your gender. (MCQ)
   (a) Male
   (b) Female
   (c) Prefer not to say
2. Please, choose your age. (MCQ)
   (a) 0 - 15 years old
   (b) 16 - 30 years old
   (c) 31 - 45 years old
   (d) 45+
3. Please, state your weight (in kg). (Text-Based)
4. Please, state your height (in feet and inches). (Text-Based)
5. Please, choose your educational qualification. (MCQ)
   (a) Primary Level
   (b) Secondary School Certificate (SSC)/O-Level
   (c) Higher School Certificate (HSC)/A-Level
   (d) Undergraduate
   (e) Graduate/Post-Graduate/Phd
6. Please, choose your employment status. (MCQ)
   (a) Full-time
   (b) Part-time
   (c) Seeking opportunities currently
   (d) Retired
   (e) Prefer not to say
7. Please, choose your financial dependency. (MCQ)
   (a) Independent

(b) Partially dependent
(c) Fully Dependent

8. Please, choose your marital status. (MCQ)
   (a) Single
   (b) Married
   (c) Divorced
   (d) Widowed
   (e) Prefer not to say
9. How many days per week do you exercise (on average)? (MCQ)
   (a) 0
   (b) 1
   (c) 2
   (d) 3
   (e) 4
   (f) 5
   (g) 7
10. Which online food ordering apps do you use? (Checkbox)
    (a) foodpanda
    (b) HungryNaki
    (c) Pathao
    (d) Other:
11. For how long have you been using online food ordering apps? (MCQ)
    (a) Less than 3 months
    (b) 3-6 months
    (c) More than 6 months
12. How many times per month do you order food online (on average)? (Text-Based)
13. When do you use online food ordering apps usually? (Checkbox)
    (a) Breakfast
    (b) Lunch
    (c) Evening Snacks
    (d) Dinner
    (e) Midnight snacks
14. What types of food do you order (Checkbox)

(a) Indian
(b) Chinese
(c) Mexican
(d) Fuchka/Chotpoti
(e) Burger
(f) Pizza
(g) Pasta
(h) Salad
(i) Fried chicken
(j) French fries
(k) Biriyani/Khichuri/Tehari
(l) Cake
(m) Donut
(n) Sweets
(o) Milkshake
(p) Other:

15. What are your reasons behind using online food ordering apps? (5 point Likert Scale)
Strongly Disagree to Strongly Agree

(a) I like to try new technologies
(b) I make impulsive decisions to order food
(c) I am too busy to cook
(d) I do not like to cook
(e) Ordering online is really easy
(f) There are different types of foods to choose from
(g) Ordering online is inexpensive
(h) Because of promo codes or discounts
(i) Food is delivered quickly
(j) Available food is very safe and hygienic
(k) Available food can cover daily nutritional requirements