5-2022

Understanding Customer Behaviour in Restaurants based on Data Mining Prediction Technique

Abdulrahman Yousef AlShamsi
aya8777@rit.edu

RIT

Understanding Customer Behaviour in Restaurants based on Data Mining Prediction Technique

by

Abdulrahman Yousef AlShamsi

A Capstone Submitted in Partial Fulfilment of the Requirements for the Degree of Master of Science in Professional Studies:

Data Analytics

Department of Graduate Programs & Research

Rochester Institute of Technology

RIT Dubai

May 2022

RIT

Master of Science in Professional Studies:

Data Analytics

Graduate Capstone Approval

Student Name: Abdulrahman AlShamsi

Graduate Capstone Title: Understanding Customer Behaviour in Restaurants based on Data Mining Prediction Technique

Graduate Capstone Committee:

Name: **Dr. Sanjay Modak** **Date:**

Chair of committee

Name: **Dr. Ehsan Warriach** **Date:**

Member of committee

ACKNOWLEDGMENTS

I would like to express my thanks and gratitude for RIT university to offer me the chance of studying my master's degree, and to all my colleagues were great members with me during my studying. Moreover, Dr. Mick was such a great professor who taught me the fundamental of R programming language. In addition, to my mentor Dr. Ehsan Warriach for his guidance during my capstone project as well as Dr. Sanjay for being a great assistant and advisor.

ABSTRACT

Customer's behavior varies from person to person based on their segmentations, while understanding these differences is one of the key elements of success in food and beverage sector. By understanding customer's behaviors, restaurant's owners will be able to identify their targeted customers and will give a clear insight on their menu products. Additionally, it will allow them to target their marketing campaigns, increase the revenue and optimize the cost. Artificial intelligence applications in this field have a huge positive impact in operations of food and beverage sector and depending on of this technology will change the way of restaurant's management. Data mining prediction model is a tool that can be used by business's stakeholders to determine and predict the most attributes that can affect their customer behavior. Therefore, the current research finds better solutions to enhance business decision-making by the use of AI and data analytics which will help in understanding the consumer's behaviors.

Key words: Customer behavior, Artificial intelligence, Supervised machine learning algorithms, Prediction model, Decision trees, Logistic regression, Random Forest.

LIST OF FIGURES

LIST OF TABLES

TABLE OF CONTENTS

CHAPTER 1

1.1. Background of the Problem

The restaurant business has rapidly progressed in the past years, as modern methods of business processes paved the way for expansion and dynamic development in the foodservice market in UAE (Symons, 2013). In 2015, the foodservice and beverage industry's market size was $7.2 billion, the restaurant industry in UAE is among the fastest-growing sectors (2017). The restaurants in UAE are looking for effective strategies and targeted marketing activities to attract customers. The study of Kurnia et al. (2019) stated that competition is on the rise because restaurants are competing to understand consumer behaviour and adopt modern ways of insightful business decision-making. To remain in the competition, restaurants and other food businesses should explore the benefits of artificial intelligence and data analytics for marketing and promotional activities. Consumer behaviour analysis became an interesting area to explore because of its role in helping the stakeholders to have insightful business decisions and stay in the competition. Ping-Ho et al. (2010) stated that in restaurant businesses, the data and information regarding their orders, preferences of certain menu items, buying behaviour of certain cuisine is useful in understanding why and what customers order in a restaurant. In addition, more researches have been published to investigate the process behind consumer behaviours and their attitude towards purchasing a service or order particular food types in restaurants. Le and Liaw (2017) stated that till the 1960s, the restaurants' stakeholders viewed that customers make rational selections when they decide to buy a product or service, the choice was mostly linked to their satisfaction. This changed after 2008 when customers became cautious while buying products or services. Trifu and Ivan (2014) identified that the growth of technology and the internet has influenced customer decisions due to

a variety of choices; this has complicated the understanding of customer behaviour for many researchers. Although technology may have complicated understanding the process of customer behaviour, the development in data mining tools techniques has provided a range of solutions to understand consumer behaviour.

1.2. Statement of the problem

Understanding consumer purchasing behaviour and identify their needs is a complex problem facing restaurant's owners. Lack of understanding consumer behavior will lead to untargeted marketing strategy and inappropriate decision making. It can be considered as one of the major problems in retail businesses as well as food and beverage sector. According to Stankevich et al. (2017), understanding consumer behaviour helps businesses in developing targeted promotional and marketing strategies to compete and survive in the foodservice market. Exploring of new trends and changes in consumer's needs and purchasing behaviour in advance will have a positive impact on the restaurant sales and brand awareness. Moreover, the adoption of new technologies and artificial intelligence in restaurants' businesses will enhance the way of the management and it will give insight vision based on customers demand. With the help of data mining prediction technique, restaurants and other foodservice businesses can make effective decisions for sales strategy, promotional, and marketing activities after analysing sales patterns and consumer buying behavior (Kurnia, 2019).

1.3. Project Goals

The project targets to implement a prediction model which will identify the factors that can affect the customers to buy the bestselling item which is Maldives Frozen Zeros (such a special type of homemade ice cream) by using a large dataset based on transaction data in a restaurant that has several branches in the UAE. It is expected that the proposed model will contribute to the field in terms of advanced results and improved performance of the prediction model compared to the previous applied models. The goals of the proposed project are mentioned below:

- To recommend promotional offers and marketing activities based on data mining prediction model for restaurants and coffee shops in UAE.
- To enhance restaurant's operation activities by exploring most selling products based on customers' needs which will lead to generate more revenue and optimize the cost.
- To implement classification prediction model using machine learning algorithms and data analytics tools such as R for analysing consumer behaviour in a restaurant and coffee shops in the UAE.
- To evaluate the accuracy of the proposed model and compare it with previous models in terms of performance.
- To contribute in supporting food and beverage industry by promoting the use of machine learning techniques in restaurants, coffee shops, hotels and other foodservice businesses.

1.4. Methodology

The CRISP-DM methodology is a standard process used for data mining projects that provides a complete picture of the project's life cycle. It is divided into six major steps as shown in Figure 1

which are Business understanding, Data understating, Data preparation, Modelling, Evaluation and Deployment. The study will use coffee shop transactions database with relevant information that will be used to achieve the project goals and objectives. Project data will be collected from "Absolute Zero" coffee shop which has six branches in the UAE. The data will go through several stages including data cleaning by removing the missing values, outliers and unnecessary columns. The study's primary focus will be on the orders taken from customers, customers details, names of the menu items that were ordered in one transaction and a complete list of items in the menu including a variety of foods and drinks. These data will be used to build the model by identifying the association rules between menu items bought at the same time and determine the occurrence pattern in the given dataset.

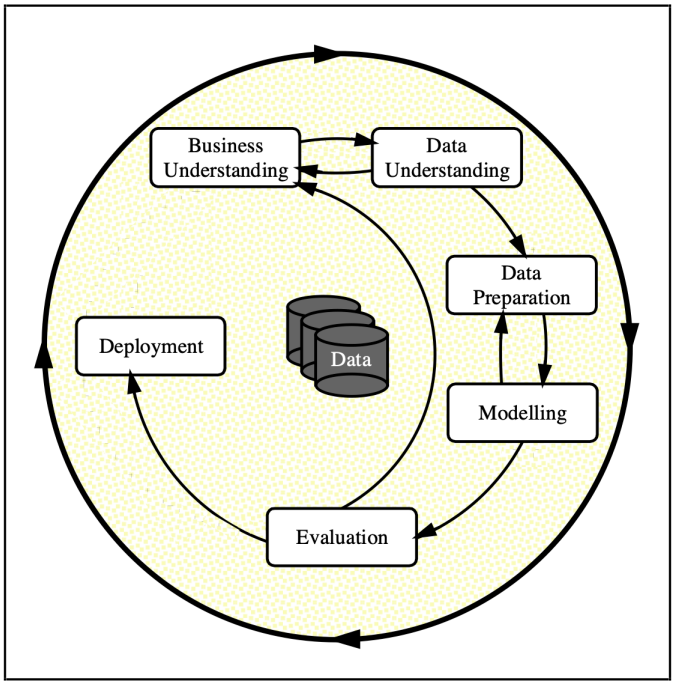

Figure 1: Steps of CRISP-DM Process Model. (Wirth & Hipp, 2000)

Step 1: Business understanding

The first step will focus on understanding the proposed project objectives and goals from the business perspective. This information will be converted to initial project plan by determining the data mining goals and techniques to achieve the project goals. (Wirth & Hipp, 2000)

Step 2: Data understanding

This step will start by collecting the required data that will be used in this project and exploring the data features in order to verify the quality of the data. This step is essential to prepare the project plan and determining the data mining goals and techniques. (Wirth & Hipp, 2000)

Step 3: Data preparation

After understanding the data, preparation and preprocessing processes will take place in order to convert the data to the final dataset which will be used in building the model. The preparation and preprocessing steps will include several steps starting with cleansing the data by removing the missing values and unnecessary columns, attributes selection based on the project requirements and further preprocessing steps to prepare it for the modelling stage.

Step 4: Modelling

The prepared data will be used to build the model, several data mining techniques will be built to get the most appropriate prediction model to identify the most factors that influence the customers to buy the bestselling product in the coffee shop.

Step 5: Evaluation

In this step, the built model will be evaluated and compared with different models by implementing the available machine learning algorithms to get most accurate model which will be deployed. In the evaluation stage, it is important to verify and ensure that the results will be aligned and fulfil the project objective and goals.

Step 6: Deployment

Deploying the project will be the final step in the data mining project. The outcome of the project will be presented to the business owner in order to use it and get the expected benefits out of it. Usually, the deployment phase will be carried out by the business owner after meeting their satisfaction. (Wirth & Hipp, 2000)

Data Analysis Method and Analytical Tool

The proposed project will be based on data mining prediction model to identify the customers purchasing behaviour. The data analytics technique that will be used is supervised classification technique which will be used to analyze the transactional data and the customers details in order to identify the most attributes that influence the customer's decision. However, the data visualization and analysis will be performed by using data analytics tool which is R for analyzing the restaurant transactional data.

1.5. Limitations of the study

While working on the project I faced some limitations of my study:

- Due to the pandemic, there was a change in the behavior of the customers due to the lockdown in the country. Due to the quarantine, customers shift from dining in the restaurants to buying online as a takeaway, which increases the sales of restaurants who offer food delivery services. This was reflected as a fluctuation in the sales.
- There are a lot of trends and changes in interests in the food and beverage sector. Customers desires and cravings can differ with seasons, weather, new products and trends. For instance, the craving of hot drinks can increase in Winter, while it could be decreased with Summer season and hot temperature; and this point can also show a difference in the data.
- The attributes in customer details were limited which can negatively impact the predictive model.

CHAPTER 2 - LITERATURE REVIEW

The approach marks the application of AI, among other methods used in the analysis of customer behaviour for business insights and decision-making (Kaur & Kang, 2016). According to Kaur and Kang (2016), periodic mining is proposed to be used in the study which is a prediction data mining technique used to understand the dynamic of the generation process by examining the changes in the discovered patterns. Eshlaghy and Alinejad (2011) used Artificial Neural Networks (ANN) to analyse the customer behaviour while selecting restaurants based on factors divided into 4 clusters, the results of this study shows that this technique is useful for marketing field and it has a potential to recognize the factors that affecting customer's behaviour. The authors Momtaz et al. (2013) used the k-means clustering technique and proposed RFM (Recency-Frequency-Monetary) based model to analyse consumer behaviour, the model did not provide significant results as no difference was observed between a valuable customer and ones who had just left the restaurant. According to Stankevich et al. (2017), understanding consumer behavior helps businesses in developing targeted promotional and marketing strategies to compete and survive in the foodservice market. The authors Abdi and Abolmakarem (2018) proposed a Customer Behavior Mining Framework (CBMF), a two-stage process utilizing both classification and clustering techniques. The model predicted attractiveness in new customers and identified three levels of customers. To analyse customer behaviour the purchasing order or transaction information is significant. Transaction data are increasingly used as research variables to gain business insights from consumers' transactions and orders. According to the study of Yrjölä et al. (2019), the customer experience is highly important for the research on customer services and consumer management. This study has analyzed the consumer perceived value of the food services through

the traditional models. Further, the study of Tuncer, Unusan and Cobanoglu (2020) have determined the service quality and behavioral intentions of the customers, this study has analyzed the structural models and combined the service quality. The study has stated that managerial practices affect the service quality and customer experiences in Turkish restaurants. One of the uses of data mining techniques in restaurants is to predict the satisfaction of customers. A study by Tama used C4.5 and REANN data mining techniques to determine customer satisfaction (2015). The C4.5 model showed that the strongest predictor of customer satisfaction was customer behavior. In another study of Madani and Alshraideh (2021) using the C4.5 model for data mining, machine learning techniques are employed to predict customer purchasing decisions in online food delivery sector. This is accomplished by evaluating a dataset including a number of metrics important to online meal delivery services, as well as a dataset pertaining to customer purchasing experiences. A comparison of three prediction models was presented in order to determine which model is the most appropriate and accurate. CART and C4.5 decision trees, a random forest, and a rule-based classifier were employed in this research. The four models all performed admirably in terms of forecasting purchasing decisions, but the C4.5 decision tree outperformed the others with a 91.67 percent accuracy. Halim et al. (2019) used prediction model to analyze customer's behavior regarding consumption of certain menu items in R software for a café restaurant in Surabaya. A demand for an Indonesian food menu was identified compared to other menu types like Chinese. Furthermore, these data mining tools extract customer information available on the website to business owners. These owners can use the information as recommendations for the business to improve its activities. However, the study of De Kervenoael et al. (2015) stated that business owners while conducting market basket analysis to identify their consumer behaviors are

required to be aware of the issues that exist in the environment of multi-store. The first issue that arises is that purchasing patterns of the consumer is temporary. For instance, demand for seasonal food items. Further, another problem is related to the patterns of the relationship among the subset of stores. Therefore, to overcome these issues prediction model is required to be developed to understand consumer purchasing behaviors. However, different studies on the purchasing patterns of customer's suggested that data-mining tools deliver the most reliable information.

While some data mining techniques are used to predict how customers could react to food and services provided in restaurants, other techniques are used to build customer profiles. Customer profiling is achieved by an in-depth analysis of guest demographics and lifestyle characteristics (Kasavana, 2010). The customer profile could help determine what would translate to a fine-dining experience for the customer. In this case, content analysis and sentiment analysis are two mining approaches that could be used to explore the emotional intent of words left on review sites. Harba et al. (2021) used sentiment analysis to determine what customers considered a fine dining experience in a restaurant in Bucharest. The data mining analysis showed the quality of the dishes served and the quality of the service as some of the predictive factors for a fine-dining experience. Fernandes et al. (2021) also applied sentiment analysis to study the key performance indicator of restaurants based on a review obtained from the site TripAdvisor. The results show that sentiment analysis and other mining techniques can be used to reveal key factors for customer relationship management in restaurants.

Other predictive techniques used to understand customer behavior are based on common data mining and machine learning processes. For instance, Lasarati et al. (2012) established that a step-wise logistic regression model was 73.39 percent accurate in classifying customers for a student-operated restaurant. Another model used to classify customers is the behavioral scoring model, based on the profile scoring model, which restaurants can use to predict the customers with a high value to the business (Chong & Lee, 2017). These models collectively help ensure that restaurants concentrate their resources in the right areas.

Understanding customer behavior through predictive data mining techniques helps shape the marketing strategies of a business. Kashani et al. (2017) used a decision tree and the Quest Algorithm to characterize customers based on the orders they had submitted to a restaurant. The results showed that the example restaurant had customers who considered healthy, voluminous, and free-living food consumption behaviors. Cheng et al. (2021) also applied the decision tree analysis to determine which customers would have a high level of repeat patronage intention. The results showed that female customers aged 18 to 34 and 45 and above would exhibit high patronage intention. These results can be used to adjust the marketing approaches of restaurants by concentrating on the groups that could bring more return on investments to the business.

As a result from the literature review, it is concluded that the data mining prediction techniques has the potential to enhance food and beverage sector by analyze and determine customer's buying behavior. The below are some of the important observations from the literature review:

1. Artificial Neural Networks (ANN) has a potential to recognize the factors that affect customer's behaviour which will be useful for marketing studies.
2. The C4.5 decision tree predicts purchasing decisions with the highest accuracy of 91.67%, and it might be considered one of the most important predictive techniques for restaurant customer behavior.
3. By using the data mining technique, customer profiling can be identified by analyzing the demographics and lifestyle characteristics of the customers.
4. Predictive models help the restaurants to optimize and utilize their resources in the right areas based on the appropriate customer segmentation.

CHAPTER 3 - PROJECT DESCRIPTION

In this project, the aim is to get the most factors that affected and influence the customer's decision to order the bestselling product in the shop. Therefore, our target to get the most accurate model by following several steps. The first stage is collecting the data. I collected my data from Absolute Zero coffee shop which provided me with two types of data which are the transactional data and customers details data. The second stage was preprocessing stage, in which we clean the data, and choose the attributes that we need to work on. In addition, the third stage is data exploration and visualization, in which we explore the data and insights, and create graphs of these insights. Finally, the modeling stage, where we use the machine learning algorithms to find the best model and compare them based on the model's performance.

3.1. Data Collection

Absolute Zero coffee shop will provide the data source to be used in this project. The data will be collected from different sources to gain more details about the menu items, customer details and transaction database for the last 4 years. The data will be retrieved from point-of-sale (POS) system to get the transaction log with the menu items that order in each transaction. Another source of data will be Koinz system which is mobile application that integrated with the POS system to get the demographic information about the customers. As the customer register in Koinz, their data will be automatically reflected in the POS system and it will capture any purchasing activities done by the customer.

3.2. Data Information

We have used two datasets which are POS transaction data and the customers data. POS transaction dataset composed of 72,107 records and 6 attributes (Table 1 includes the name of the attributes,

description, and type). However, the customers details dataset composed of 6,600 records and 14 attributes (Table 2 includes the name of the attributes, description, and type).

#	Attribute	Description	Type
1	Month	Date of purchasing	Date
2	customer phone	Customer’s phone number	character
3	Product 1	First product purchased	character
4	Product 2	Second product purchased	character
5	Product 3	Third product purchased	character
6	branch name	Branch name	character

Table 1: POS transaction dataset dictionary

#	Attribute	Description	Type
1	Phone Number	Customer’s phone number	character
2	Customer Name	Customer’s name	character
3	User Email	Customer’s email	character
4	Age	Customer’s age	numeric
5	Age group	Customer’s age group	character
6	City	Customer's address	character
7	Branch	Branch name	character
8	Gender	Gender of the customer	character
9	Education	level of education	character
10	Occupation	Customer's occupation	character
11	Monthly Income	Customer's Monthly Income	character
12	No of visits	Number of visits annually	numeric
13	Order Type	Represents the type of order	character
14	Following Instagram account	If the customer is following the account of the coffee shop	character

Table 2: Customers details dataset dictionary

CHAPTER 4 - PROJECT ANALYSIS

4.1. Exploratory Data Analysis

The first step in data exploration is to import several libraries in R in order to explore and visualize the data. We have explored the numerical and categorical columns as well as identifying the missing data.

The following table contains the summary statistics of all numeric columns in our data.

variable	mean	min	max	standard deviation	mode
Age	29.48	11	87	6.78	27.25625
id_col	14,599.50	1	29,198	8,428.78	14,599.33333
monthday	4.82	1	31	7.31	1.00000
No of visits	21.67	1	44	12.41	13.35317

Table 3: Summary statistics of numerical columns

The following table shows the frequency and level percentage of all categorical columns in our data without the product column analysis (separate table will be provided for the products analysis).

variable	level	n	percentage
Age group	Adult	40,431	46.16
Age group	Children	84	0.1
Age group	Teenagers	1,098	1.25
Age group	Youth	45,981	52.49
branch name	AbsoluteZERO JUM	3,537	4.04

branch name	Ajman	11,355	12.96
branch name	Al Sufouh	9,225	10.53
branch name	Al Warqa Branch	10,863	12.4
branch name	ALWarqa'a Branch	5,352	6.11
branch name	AZ Events	114	0.13
branch name	Dubai Mall	69	0.08
branch name	Events - HCT POS	15	0.02
branch name	Events - Winter Garden POS	48	0.05
branch name	Fujairah	21	0.02
branch name	Nad Al Hamar Avenues	22,719	25.94
branch name	PZ Pizza - Dubai	57	0.07
branch name	PZ Pizza SHJ	456	0.52
branch name	Ras Al Khaimah	1,305	1.49
branch name	Sharjah Muwaileh	22,458	25.64
City	Abu Dhabi	27,111	30.95
City	Ajman	17,850	20.38
City	Dubai	13,548	15.47
City	Sharjah	29,085	33.2
day_cat	third1	74,496	85.05
day_cat	third2	6,945	7.93
day_cat	third3	6,153	7.02
Education	Bachelor	42,630	48.67
Education	High school	12,057	13.76
Education	Master	24,774	28.28
Education	PHD	7,566	8.64
Education	School	567	0.65
Following instagram account	No	29,586	33.78
Following instagram account	Yes	58,008	66.22
Gender	Female	43,305	49.44
Gender	Male	44,289	50.56
month	April	5,631	6.43
month	August	8,757	10
month	December	5,511	6.29
month	February	6,117	6.98
month	January	6,516	7.44

month	July	7,725	8.82
month	June	6,354	7.25
month	March	6,345	7.24
month	May	5,265	6.01
month	November	7,713	8.81
month	October	12,939	14.77
month	September	8,721	9.96
Monthly Income	10,000 to 20,000	15,516	17.71
Monthly Income	20,000 to 30,000	36,699	41.9
Monthly Income	5,000 to 10,000	396	0.45
Monthly Income	Less than 5,000	1,797	2.05
Monthly Income	More than 30,000	33,186	37.89
Occupation	Business	11,862	13.54
Occupation	Engineering	10,215	11.66
Occupation	Finance	14,619	16.69
Occupation	HR	13,455	15.36
Occupation	IT	11,349	12.96
Occupation	Law	13,470	15.38
Occupation	Student	12,624	14.41
Order Type	Dine in	20,529	23.44
Order Type	Home Delivery	31,023	35.42
Order Type	Takeaway	36,042	41.15
weekday	Friday	13,902	15.87
weekday	Monday	16,095	18.37
weekday	Saturday	8,250	9.42
weekday	Sunday	9,027	10.31
weekday	Thursday	20,424	23.32
weekday	Tuesday	8,151	9.31
weekday	Wednesday	11,745	13.41

Table 4: Summary statistics of categorical data

Based on the above table we noticed the following observations:

1. The majority of the customers from youth age group with 52.49%
2. The best branch in sales is Nad Al Hammer branch with 25.94%

3. The majority of the customers are Male gender with 50.56%
4. The best month in sales is October with 14.77%
5. The best day in sales is Thursday with 23.32%

We illustrate in the below plot the top frequent purchased products, the best seller product based on the below plot is Maldives Frozen Zeros with 7148 records. However, we have 41,638 records of missing data in product column. Product column is the result of combining product 1 (1st product purchased), product 2 (2nd product purchased) and product 3 (3rd product purchased) in the same transaction. Therefore, the majority of the transactions has single or two products.

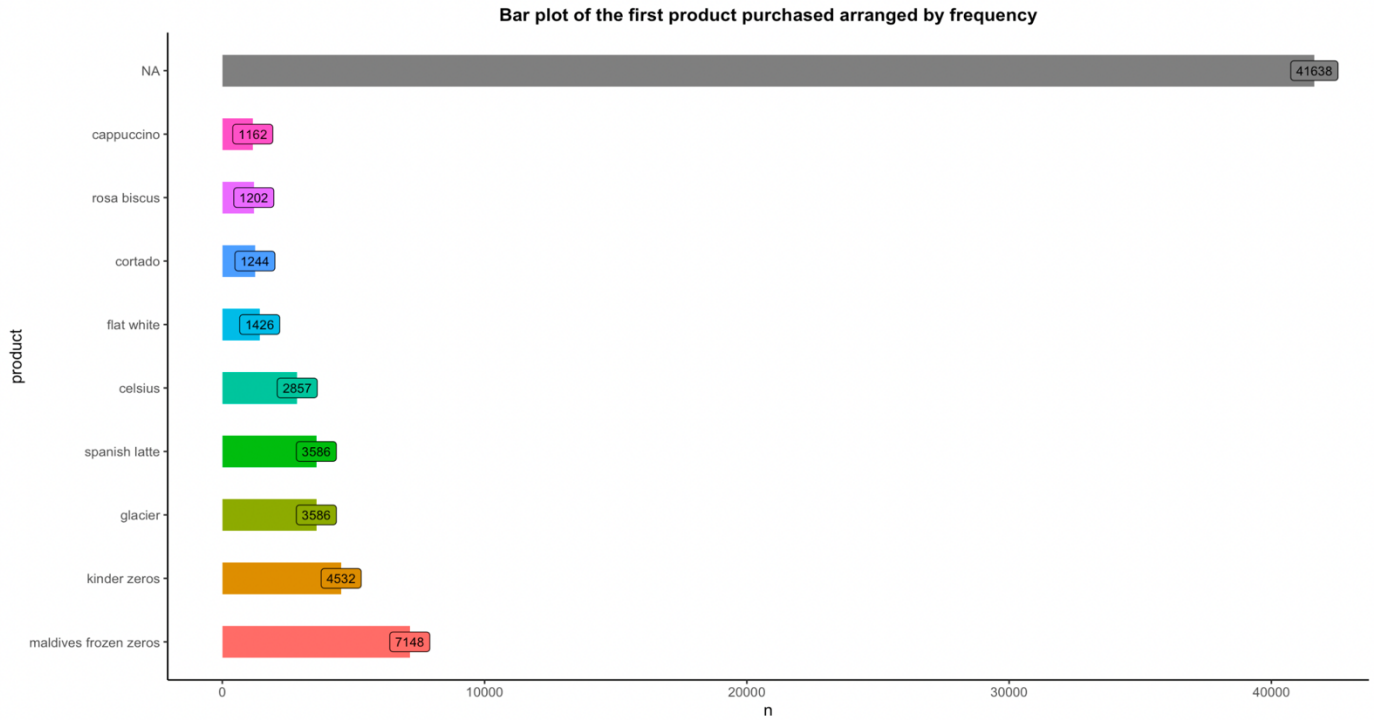

Figure 2: Bar plot of the first product purchased arranged by frequency

4.2. Data Cleaning

After exploration the data, we have gone through data cleaning steps in order to prepare them for further processing as per the following steps:

1. We have removed the duplicated column which is Branch column.
2. We have removed product column which was representing the product order in each transaction (First, Second and Third purchased product in the single transaction).
3. We have removed the date column as we discretize the day to 3 bins, 1-10,11-20, and 21-31 to test which month third is associated with the bestselling product.
4. We have removed the missing data from the dataset.
5. We have changed the name of value column to Maldives FZ
6. We have converted Maldives FZ column to Yes (if the purchased product is Maldives Frozen Zeros) and No (if the purchased product is not Maldives Frozen Zeros) to use it for the prediction model.

4.3. Data Visualization

The following histograms show the distribution of all numerical columns in our data.

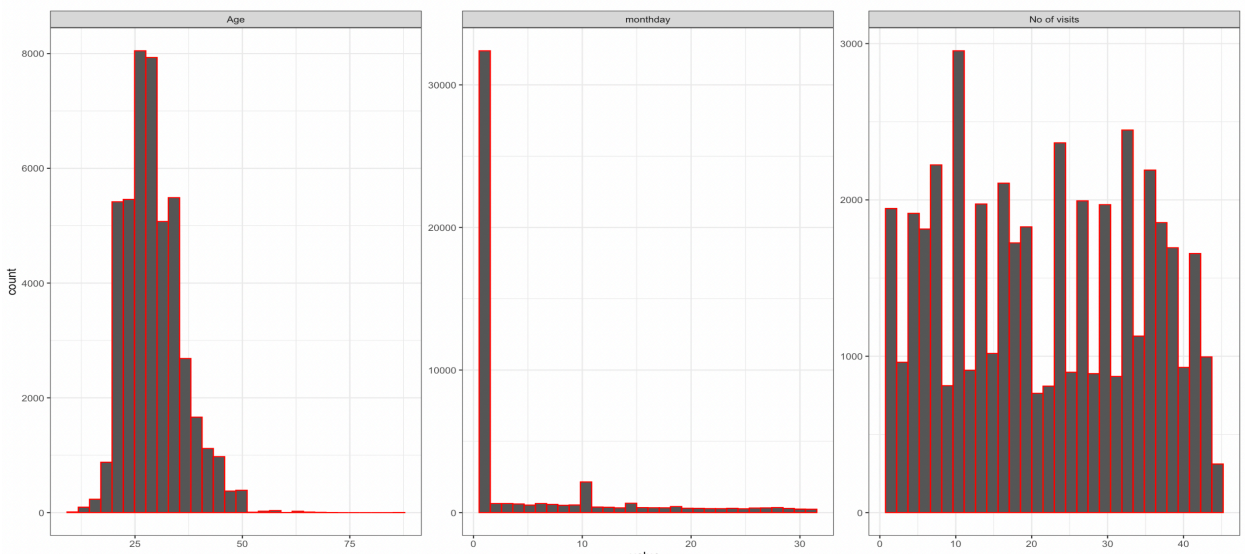

Figure 3: Histograms show the numerical columns

As shown in the above histogram, the age of most of the customers are from 25 to 30 years. In addition, about 3000 customers has around 10 visits annually. With regards to the month day, it's obvious from the histogram that the first day of the month has the greatest number of orders by more than 3000 records.

The following bar plots show the levels' frequency of each categorical column in our data.

Figure 4: Bar plots show the frequency of each categorical column

In the crosstab below, we focused on the age group and its effect on buying the most selling products. The four age groups are Children, Teenagers, Youth and Adults while the most selling products are Maldives frozen zeros, Kinder zeros, Glacier and Spanish latte. We can notice that the highest number of orders in all the products are from Youth except Spanish latte which is from Adults. In addition, we notice that the orders for Kinder zeros from Youth is almost doubled number of orders from Adults.

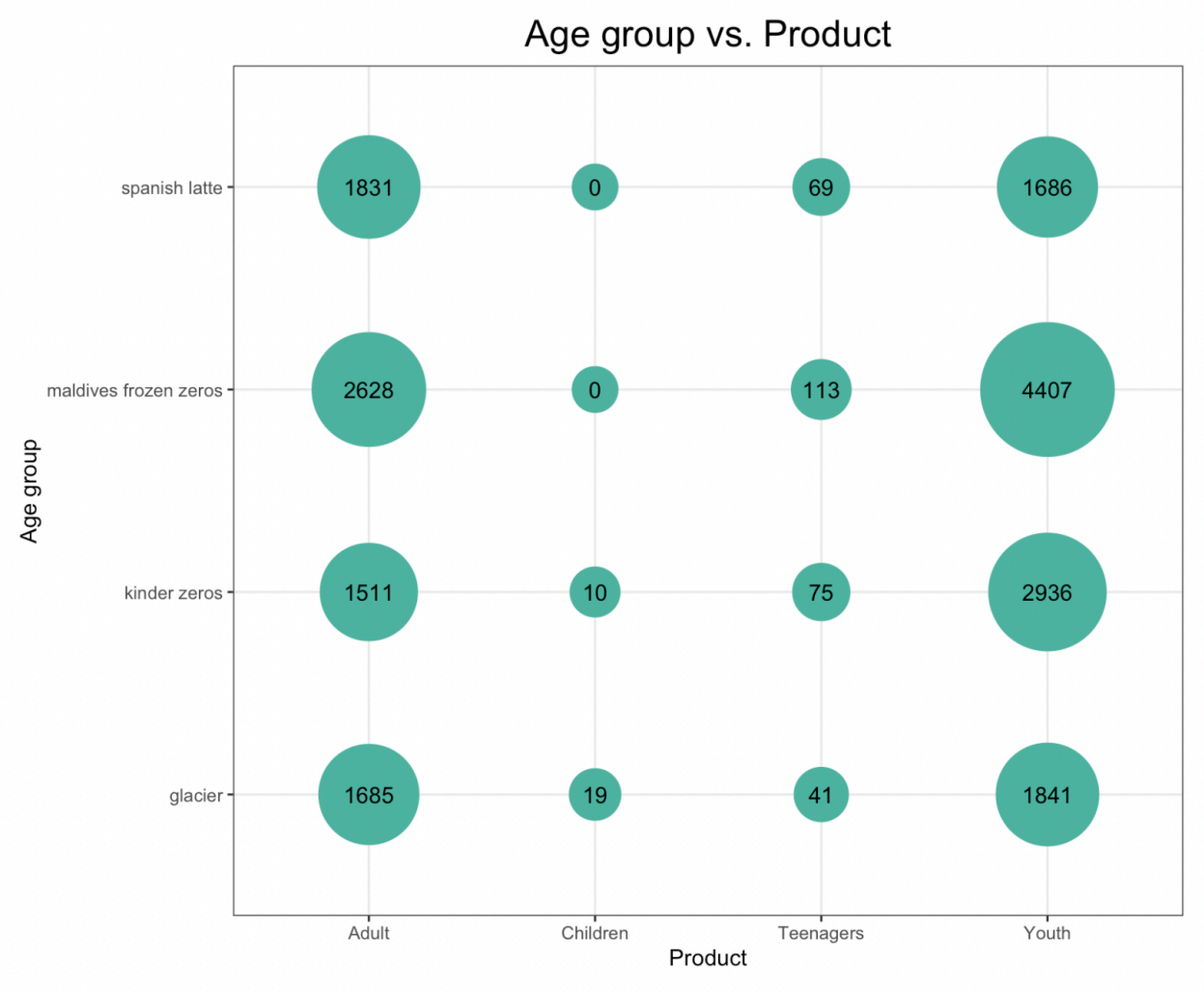

Figure 5: Crosstab for Age group vs. Product

As shown in the below bar plot, we notice that the heights number orders in October with almost 13,000 records. On the other hand, we notice that during the summer season and specially in May the number of orders has been reduced to less than half of the orders.

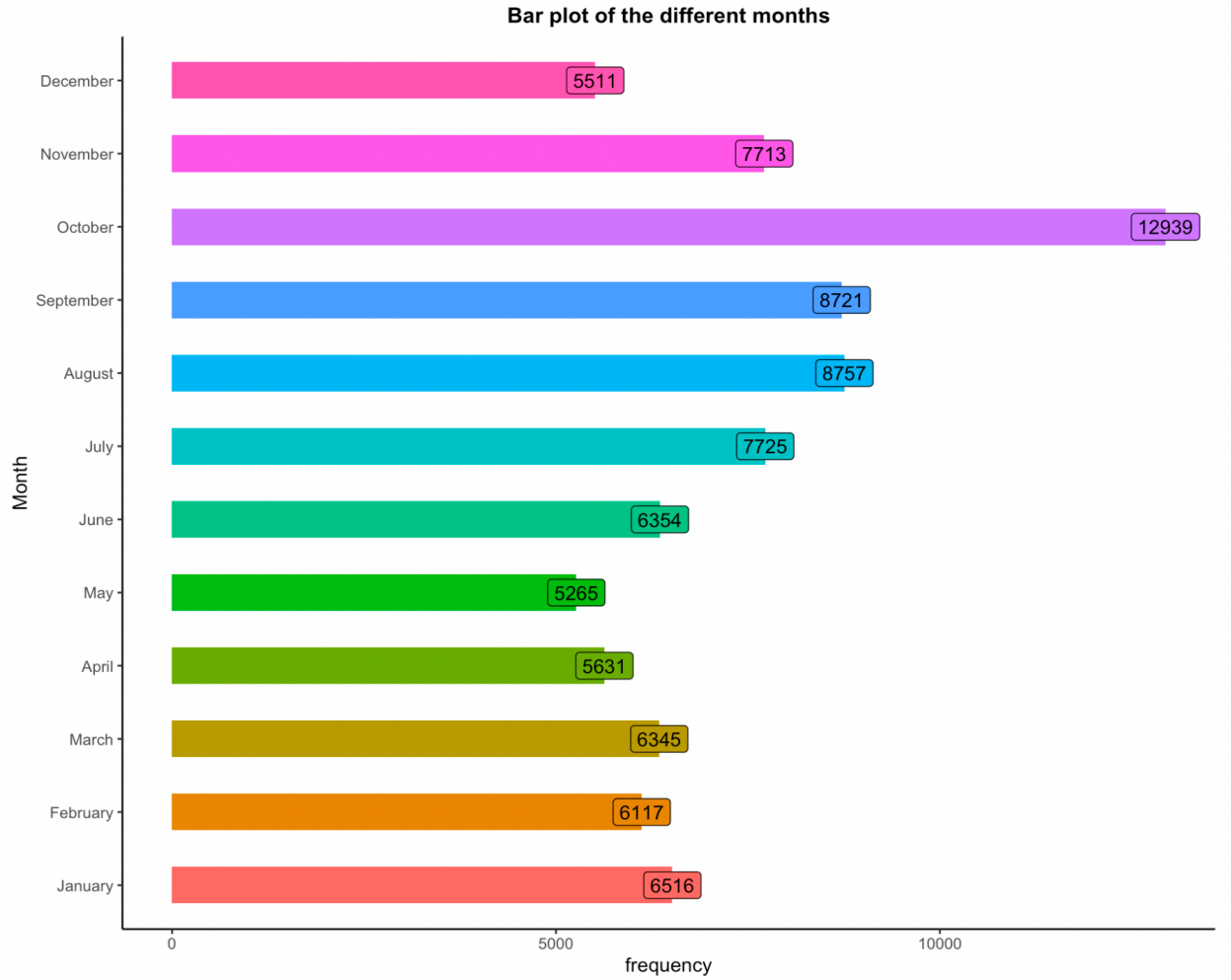

Figure 6: Bar plot of the different months

4.4. Results – Exploratory Data Analysis

We can observe from the above exploratory and visualization plots that many insights can be gathered. From the above exploratory and visualization plots, we can see that there are different insights can be captured. The majority of the customers are from Youth then Adults. Additionally, the most selling product is Maldives Frozen Zeros and the heights number of orders are during October. Most of the orders are Takeaway rather than dine in or home delivery. The reason behind that might be during Covid-19 pandemic there were restrictions in dine in restaurant which increase the takeaway orders. Lastly, the majority of the customer's monthly income between 20 to 30 thousand.

4.5. Model Building

Building the model stage will consist of several steps starting with splitting the data into training and testing data, building different machine learning algorithms and comparing different machine learning algorithm to identify the best results. We have selected three different algorithms to compare the model's performance as below.

- **Decision Trees**: It is a supervised learning algorithm that can be used to solve classification and regression problems. The purpose of employing a Decision Trees is to develop a training model that can predict the target variable's class or value based on learning simple decision rules inferred from the training data.
- **Logistic regression**: It is widely used in building binary classification models along with building correlations between the predictors and the outcomes. Logistic regression model is simple, easy to train and gives an accurate binary classification.

- **Random forest**: It's a Supervised Machine Learning Algorithm that's commonly used to solve classification and regression problems. It's a classifier that averages the results of several decision trees applied to various subsets of a dataset to improve prediction accuracy. The random forest collects predictions from each tree and predicts the ultimate output based on the majority votes of projections, rather than relying on a single decision tree.

4.6. Comparison of Different Models

After building the model, we will be comparing the models based on different parameters. Confusion matrix is an easy-to-understand crosstab of actual and expected class values. It contains the total number of observations in each category. From the confusion matrix we can calculate several ratios that can be used as well for the comparison.

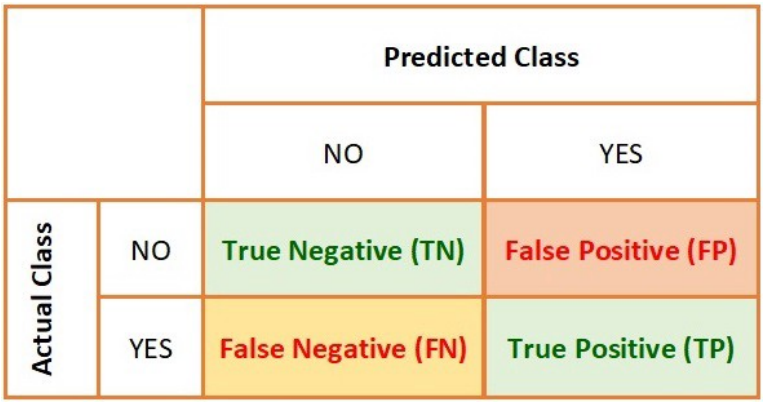

Figure 7: Confusion Matrix sample (Toshniwal, 2020)

	Accuracy	**Sensitivity**	**Specificity**	**Pos. pred. value**	**Neg. pred. value**
Formula	$\frac{(TP + TN)}{(TP + FP + FN + TN)}$	$\frac{TP}{TP + FN}$	$\frac{TN}{TN + FP}$	$\frac{TP}{TP + FP}$	$\frac{TN}{TN + FN}$

Table 5: Comparison ratios formula

- **Accuracy:** Ratio of correct predictions to total predictions.
- **Sensitivity/Recall:** Ratio of true positives to total (actual) positives in the data.
- **Specificity:** Ratio of true negatives to total negatives in the data.
- **Positive predictive value:** It is the ratio of truly tested as positive to all those who had positive test results.
- **Negative predictive value:** It is the ratio of truly tested as negative to all those who had negative test results. (Sharp, 2021)

Decision Tree:

		Actual	
Prediction		**No**	**Yes**
	No	7466	1034
	Yes	2236	753

Table 6: Decision Trees confusion matrix

	Accuracy	**Sensitivity**	**Specificity**	**Pos. pred. value**	**Neg. pred. value**	**Positive Class**
Results	71.54%	25.19%	87.84%	42.14%	76.95%	Maldives FZ

Table 7: Decision Trees comparison ratios

Based on the above results, decision tree model has an accuracy of almost 72% which is somehow good. The model has low sensitivity ratio by 25% which means the ratio of predicting the positive cases correctly is low. However, the specificity ratio is high with almost 88%. Moreover, the model has 42% of positive predictive value which is the ratio of truly tested as positive to all those who had positive test results while the model has almost 77% of the negative predictive value which is the ratio of truly tested as negative to all those who had negative test results.

Logistic regression:

		Actual	
Prediction		**No**	**Yes**
	No	6334	791
	Yes	3368	996

Table 8: Logistic regression confusion matrix

	Accuracy	**Sensitivity**	**Specificity**	**Pos. pred. value**	**Neg. pred. value**	**Positive Class**
Results	63.80%	22.82%	88.90%	55.74%	65.29%	Maldives FZ

Table 9: Logistic regression comparison ratios

Based on the above results, logistic regression has lower accuracy than decision tree model with almost 64%. It has lower sensitivity ratio as well by almost 23% and similar specificity ratio by 89%. However, the model has better positive predictive value than decision tree model with almost 56% while it has lower negative predictive value by 65%.

Random forest:

Prediction	Actual: No	Actual: Yes
No	8910	1426
Yes	792	361

Table 10: Random forest confusion matrix

	Accuracy	Sensitivity	Specificity	Pos. pred. value	Neg. pred. value	Positive Class
Results	80.69%	31.31%	86.20%	20.20%	91.84%	Maldives FZ

Table 11: Random forest comparison ratios

The final model we have built is random forest, which has better accuracy than the previous models with almost 81%. The model has better Sensitivity and Specificity ratio than other models by 31% and 86% respectively. However, it has lower Positive predictive value than other models with 20% and better Negative predictive value than other models with almost 92%.

In summary, we have identified the best model in accuracy which is Random forest model with almost 81%. Moreover, it was noticed that the model has high Specificity value as well as Negative predictive value which means that the model is performing well in predicting the negative results. However, Sensitivity and Specificity are inversely proportional to each other. So, when we increase Sensitivity, Specificity decreases, and vice versa.

CHAPTER 5 - CONCLUSION

5.1. Conclusion

Based on the customers segmentation, buying behavior will differ from person to person and recognizing these distinctions is one of the most important aspects of success in the food and beverage industry. Restaurant operators will be able to identify their target clients and provide a clear insight into their menu offerings by understanding customer behaviors. It will also enable them to target their marketing activities, improve income, and save costs. Based on this study, we came up with different outputs and conclusions that can be used by restaurants owners to get more knowledge about their customers interests. I have focused on my study on analyzing the transactional data and the customers details data for Absolute Zero Coffee Shop to identify the factors that can affect the customers to buy the bestselling item which is Maldives Frozen Zeros (such a special type of homemade ice cream). Several supervised classification prediction models where built which are decision tree, logistic regression and random forest. Comparative analysis was done to compare the performance of each model by measuring the accuracy and different comparison ratios such as sensitivity and specificity. As a result, we have noticed that Random Forest model has the best accuracy in prediction with almost 81%. However, the study will require more customers details to be incorporated for further modeling in order to get better performance results.

5.2. Recommendations

Based on the random forest model, we have identified the most predictors that can affect the customers behavior which also can help the business owner to target his marketing based on it. We identified the most 9 important predictors which are associated with 75% of the model importance.

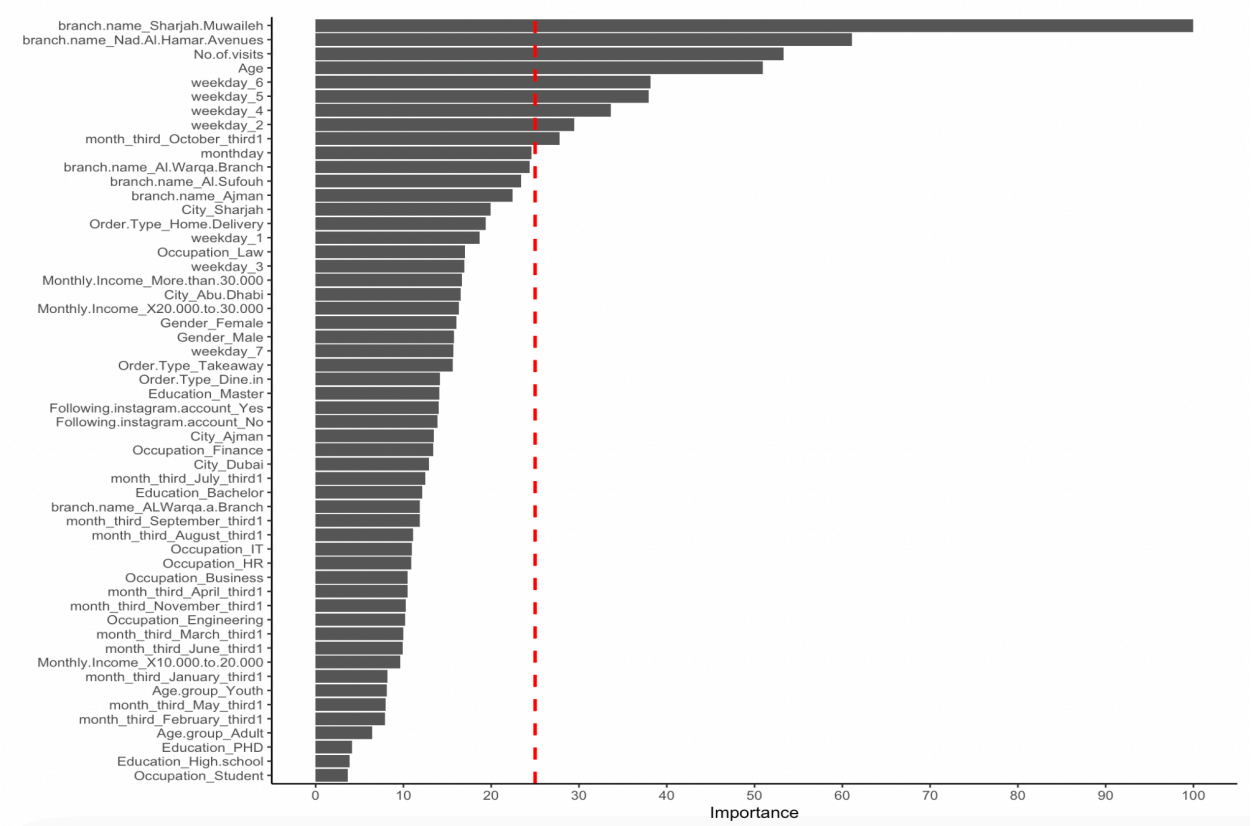

Figure 8: Variable importance

From the above plot, the below suggestions can be given to business owner to enhance their operation and marketing strategy:

- The customers of Sharjah and Nad AlHammer branches are the most customers who are buying the best-selling product. So, special promotions and loyalty program can be promoted to this geographical location in order to increase customers retention rate.
- Customer's ages play a big role in predicting the customers behavior. As it was visualized earlier that most of the customers are from Youth and adults. Hence, targeted promotions can be used to get the attention of this group of customers.
- The sales of the coffee shop during the weekend are more than the weekdays. So, instagram promotions and online delivery offers can work efficiently as the culture in UAE to have a family and friends gathering during the weekend.
- With reference to the months sales, it was noticed that the first third of October has the most sales during the year which can be the reason of that the change in weather where the winter season started with the beginning of October.

5.3. Future Work

For the future work, I would like to work with different types of restaurants in order to enhance the knowledge of understanding customers behaviors. This will enhance the decision of the business owners as they will rely on analytical results from real data of their restaurants which will result in customized marketing campaign to their specific customers based on their needs.

REFERENCES

[1] Puri-Mirza, A. (2020). *Food and beverage market size in the UAE by category 2015*. Retrieved from Statista.com: https://www.statista.com/statistics/719491/uae-food-and-beverage-market-size-by-category/

[2] Grau, G. R. (2017). *Market Basket Analysisin Retail.* UNIVERSITAT POLITÈCNICA DE CATALUNYA (UPC).

[3] Toshniwal, R. (2020, Jan 9). *How to select Performance Metrics for Classification Models*. Retrieved from Medium: https://medium.com/analytics-vidhya/how-to-select-performance-metrics-for-classification-models-c847fe6b1ea3

[4] Sharp, E. (2021, Nov 12). *Sensitivity, Specificity, PPV and NPV*. Retrieved from Geeky Medics: https://geekymedics.com/sensitivity-specificity-ppv-and-npv/

[5] Alshraideh, H., & Madani, B. (2021). *PREDICTING CONSUMER PURCHASING DECISIONS IN THE ONLINE FOOD DELIVERY INDUSTRY.* Sharjah: Department of Industrial Engineering, American University of Sharjah, Sharjah, UAE.

[6] Yrjölä, M., Rintamäki, T., Saarijärvi, H., Joensuu, J., & Kulkarni, G. (2019). A customer value perspective to service experiences in restaurants. *Retailing and Consumer Services*, 91-101.

[7] Tuncer , I., Unusan , C., & Cobanoglu, C. (2020). Service Quality, Perceived Value and Customer Satisfaction on Behavioral Intention in Restaurants: An Integrated Structural Model. *Journal of Quality Assurance in Hospitality & Tourism*, 447-475.

[8] Karina Kusuma, H., Halim, S., & Felecia. (2019). Business Intelligence for Designing Restaurant Marketing Strategy: A Case Study. *The Fifth Information Systems International*

Conference 2019 (pp. 121-131). Surabaya: Industrial Engineering, Petra Christian University.

[9] Kervenoael, R., Yanık, S., Bozkaya, B., Palmer, M., & Hallsworth, A. (2015). Trading-up on unmet expectations? Evaluating consumers' expectations in online premium grocery shopping logistics. *A Leading Journal of Supply Chain Management*, 83-104.

[10] Kaur, M., & Kang, S. (2016). Market Basket Analysis: Identify the changing trends of market data using association rule mining. *International Conference on Computational Modeling and Security (CMS 2016)* (pp. 78 – 85). Sangrur: International Conference on Computational Modeling and Security (CMS 2016).

[11] Eshlaghy, A. T., & Alinejad, S. (2011). Classification of Customers' behavior in Selection of the Restaurant with use of Neural Network. *European Journal of Economics, Finance and Administrative Sciences*.

[12] Momtaz, N. J., Alizadeh, S., & Vaghefi, M. S. (2013). A new model for assessment fast food customer behavior case study An Iranian fast-food restaurant. *British Food Journal*, 601-613.

[13] Stankevich, A. (2017). Explaining the Consumer Decision-Making Process: Critical Literature Review. *Journal of International Business Research and Marketing*, Volume 2, Issue 6.

[14] Abdi, F., & Abolmakarem, S. (2018). Customer Behavior Mining Framework (CBMF) using clustering and classification techniques. *Journal of Industrial Engineering International*.

[15] Wirth, R., & Hipp, J. (2000). *CRISP-DM: Towards a Standard Process Model for Data Mining.* Germany.

[16] Kurnia, Y. (2019). Study of application of data mining market basket analysis for knowing sales pattern (association of items) at the O! Fish restaurant using apriori algorithm. *Journal of Physics: Conference Series*.

[17] Symons, M. (2013). The rise of the restaurant and the fate of hospitality. *International Journal of Contemporary Hospitality Management*.

[18] *Food and beverage market size in the UAE by category 2015* . (2017, April). Retrieved from Statista: https://www.statista.com/statistics/719491/uae-food-and-beverage-market-size-by-category/

[19] Ting, P., Pan, S., & Chou, S. S. (2010). Finding ideal menu items assortments: An empirical application of market basket analysis. *Cornell Hospitality Quarterly*.

[20] Le, T. M., & Liaw, S.-Y. (2017). Effects of Pros and Cons of Applying Big Data Analytics to Consumers' Responses in an E-Commerce Context.

[21] TRIFU, M. R., & IVAN, M. L. (2014). Big Data: present and future. *Database Systems Journal*.

[22] TAMA, B. A. (2015). DATA MINING FOR PREDICTING CUSTOMER SATISFACTION IN FAST-FOOD RESTAURANT. *Journal of Theoretical and Applied Information Technology*, Vol. 75.

[23] Kasavana, M. (2010, June 1). *Mining Restaurant Data: Know your customer.* Retrieved from Hospitality Upgrade: https://www.hospitalityupgrade.com/_magazine/magazine_Detail.asp/?ID=522

[24] Harba, J.-N., Tigu, G., & Davidescu, A. (2021). Exploring Consumer Emotions in Pre-Pandemic and Pandemic Times. A Sentiment Analysis of Perceptions in the Fine-Dining Restaurant Industry in Bucharest, Romania. *International Journal of Environmental Research and Public Health*.

[25] Fernandes, E., Moro, S., Cortez, P., Batista, F., & Ribeiro, R. (2021). A data-driven approach to measure restaurant performance by combining online reviews with historical sales data. *International Journal of Hospitality Management*.

[26] Larasati, A., Slevitch, L., & DeYong , C. (2012). The Application of Neural Network and Logistics Regression Models on Predicting Customer Satisfaction in a Student-Operated Restaurant. *Procedia - Social and Behavioral Sciences*.

[27] Chong , Y., & Lee, G. (2017). Grouping hotel restaurant customers based on a behavioral scoring model : An exploratory study. *International Journal of Tourism and Hospitality Research*.

[28] Kashani, F., & Shahmirzaloo, Z. (2017). Developing Marketing Strategies Using Customer Relationship Management and Data Mining (Case Study: Perperook Chain Restaurants). *Marketing and Management of Innovations*.

[29] Cheng, Y.-S., Kuo, N.-T., Chang, K.-C., & Wu, H.-T. (2021). Using Data Mining Methods to Predict Repeat Patronage Intention in the Restaurant Industry. *Journal of Quality Assurance in Hospitality & Tourism* .