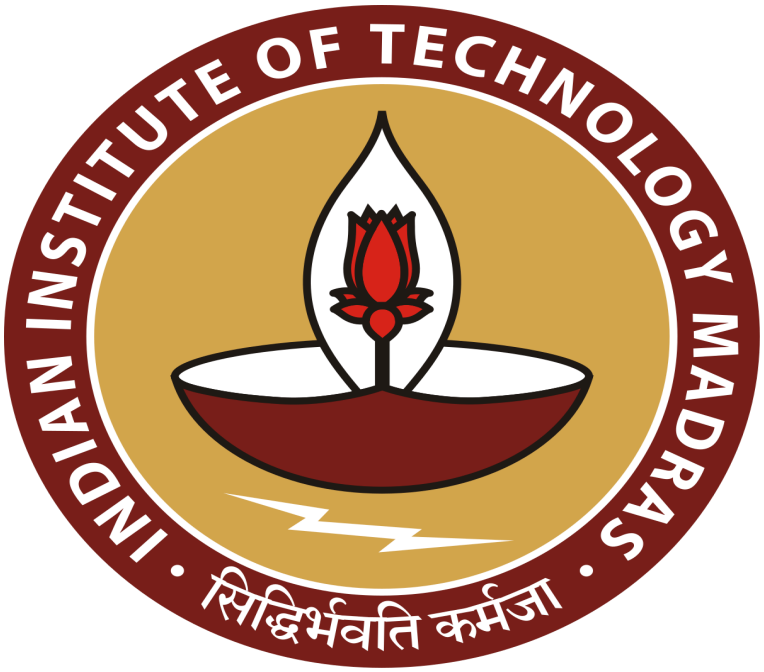
Concluding Report on Pre Market Segmentation Analysis for Pixie Hair Salon

The Final report for BDM capstone project

Submitted by

Name : Aditya Siddharth Jyoti
Roll Number : 22f3001792



IITM Online BS Degree Program,
Indian Institute of Technology, Madras, Chennai
Tamil Nadu, India, 600036

Contents

1. Executive Summary

After brainstorming for months on the data with limited attributes , I came up with the ideas of analyzing the most effective timelines to promote their contents on digital and print media based on Time-Series Analysis. In the report multiple parameters for Time-Series Analysis are used. Details about the best months , best seasons and best time are mentioned with adequate proof of analysis. For identifying the potential demography like Gender , a new feature is added based on the trends in availing the services combinations. This new feature unlocked extra possibilities to achieve CLV (Customer Lifetime Value) Analysis on the data. Using CLV , makes it clear which group of customers needs to be focused for growing the business or to keep a constant acceleration in the revenue. CLV also helped to solve the base parameter for the required demographic data , specially based on the gender.

Growth of the non-popular services is specially focused in the later part of the report. This helps to identify the potential in these ignored services and pave a way for extra revenue by including them in the promotional theme. Star services were already identified in the previous report which included Haircut for male and female , beard trimming and eyebrow threading . These star services will be used to compare the customer segment which are most likely to spend more on these star services and hence get a customer segmentation to focus on , based on these services.

Effect of locality is considered well in the end , which cannot be ignored to decide the best effective working periods. As the proposal mentioned that the surrounding included residential areas and universities , I had a personal research on the best suitable timings , which is mentioned properly in this report.

All the methods , inferences , proofs , reasonings and conclusions are mentioned properly in this report.

2. Detailed Explanation of Analysis Process/Method

2.1 Time-Series Analysis :-

Time series analysis is a way of analyzing a sequence of data points collected over an interval of time. Three types of analysis are present in this section.

1. **Revenue per year :-** The data provided had a column ‘Date’ in dd-mm-yyyy format. This indicated that I already had the daily basis data of the individual customers. The requirement was to convert this daily basis data into a consolidated yearly basis data. ‘Date’ and ‘Billed Cost’ were the core features which could help to infer the revenue generated per year. The individual billed cost column was summed over the resampling of the column ‘Date’ from date-by-date format to year-by-year format using the python libraries . Now using bar charts it will be very clear about the yearly revenue of the salon from the services provided to the customers.

2. **Revenue per Quarter and Month :-** The analysis on the basis of year was not enough to generate perfect inferences about the potential performance of revenue generation. Thus further investigation was required on the monthly and quarterly basis. Using the existing data , I created a new data-frame which included three columns :- Year , Month , Revenue. The same method which was used to generate revenue per year , was used here to generate revenue per month. This data now can be used to even further calculate the revenue of each quarter by grouping the data on the quarter basis. Simple bar chart won’t be enough for analyzing the monthly revenue of all three years . And hence the stacked bar chart comes into play . Each bar in the plot will represent the month and each stack of the bar will represent the Year. Y axis will represent the Revenue generated in each month.

3. **Gender Distribution of the most frequent customers :-** There are services like Haircut (Male) , Beard Trimming which can only be availed by the male customers and there are services like Haircut(Female) , Manicure , Pedicure etc , which can only be availed by the Female customers . I enquired with the salon about this segmentation and prepared a list for both the categories. Based

on this list I found out the potential gender distribution of the data . This included Male , Female , Couples (customers who availed both Male and Female services in a single visit) and Unknown (customers who availed only the services which can be availed by either male or female). After this division I calculated the frequency of top 50 customers who visited the salon frequently and investigated their gender distribution . A simple bar chart was enough to convey the most frequent Gender demography. After this frequency of gender distribution per month was calculated . The inferences were shown using a line chart.

2.2 Customer Lifetime Value Analysis :-

Customer lifetime value (CLV) is a business metric that measures how much a business can plan to earn from the average customer over the course of the relationship. The formula to calculate customer lifetime is mentioned below :-

Customer Value = Average Order Value x Average Purchase Frequency

Average Order Value: This is the typical amount a customer spends per purchase. You can find it by dividing your total revenue by the number of orders in a specific period.

Average Purchase Frequency: This is how often a customer makes a purchase within a given timeframe. You can calculate it by dividing the total number of orders by the total number of customers within that same period.

To calculate the CLV value I created a separate data frame which stored the CLV value of each and every customer prepared using the same formula as mentioned above. After calculating the CLV values I calculated the grouped CLV scores on the basis of gender . Top 5 CLV scores were also compared by their gender distribution using a simple bar chart.

2.3 Day and Night Preference division :-

I created a new column ‘shift’ in the data-frame , which helped me to classify the timings of the customer visit on the day and night basis.

Whenever a customer visits between 8:00 am to 8:00 pm , it is considered to be the 1st shift and for the rest of the timing , it is considered to be the second shift. Based on the shifts I calculated the frequency of customers visited in each of the individual shifts. Further I calculated the distribution of gender in each of the shifts . Later I also calculated the revenue generated by each shift for the past 2 years in total. Bar graphs and stacked bar graphs were enough to represent these classifications.

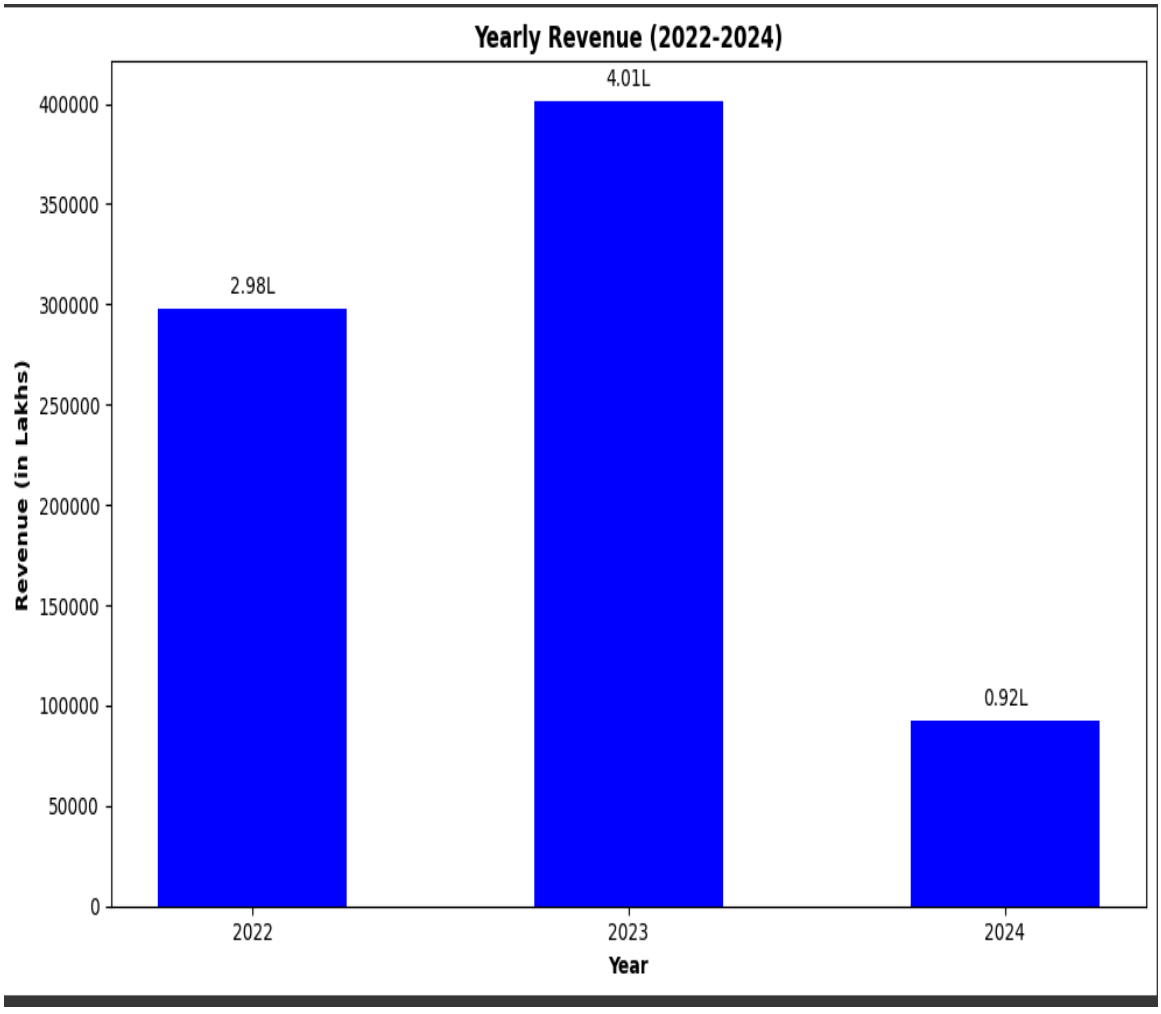
2.3 Checking the influence of locality on the salon :-

It is a general phenomena that locality will surely have some influence on the functioning of any shop / salon / company or organization. Hence I decided to think of the effects which locality can have on the salon. As pixie belongs to a beauty and personal care industry , people usually come to the salon when either they are in their free space or they are planning to visit somewhere in the vacations. People also visit the salon when they usually come back after a long break.

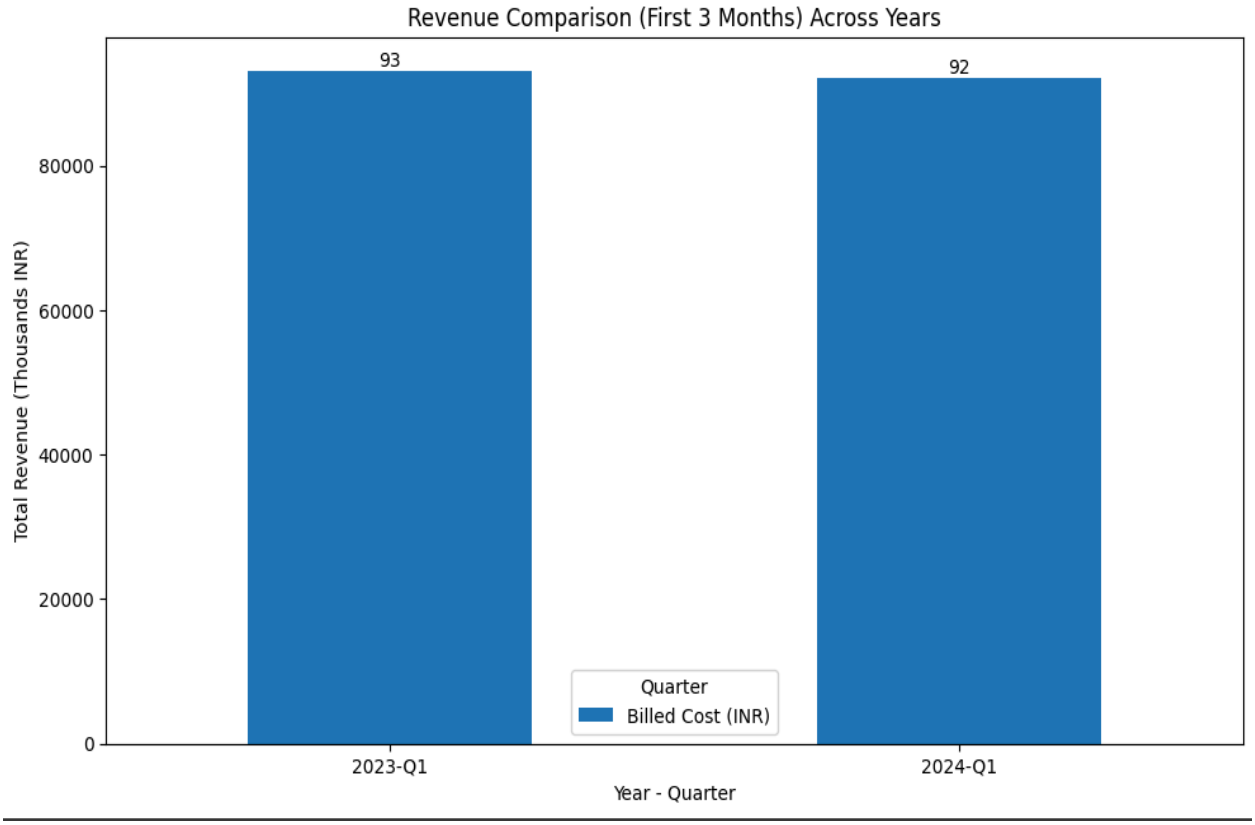
I started collecting the data from my local friends , seniors and neighbors about their vacations provided by the school / university / company.

3. Results and Findings

1. After analyzing the revenues per year , It was found that for the year 2022 (where data of first 2 months was not complete) the revenue generated was 2.9 Lakhs , for year 2023 (only year with a complete data present) the revenue shoot up by 1.03 Lakh and the total revenue was 4.01 Lakh. For the current year 2024 , the salon had already made 0.92 Lakh in the 1st quarter. This is almost 25% of the revenue generated in the year 2023. At this point of time , if we assume a linear and steady growth in upcoming months , it can be inferred that in this year the revenue will be the same as the revenue of the year 2023. We now need to investigate more about the quarterly trends now for further investigation.

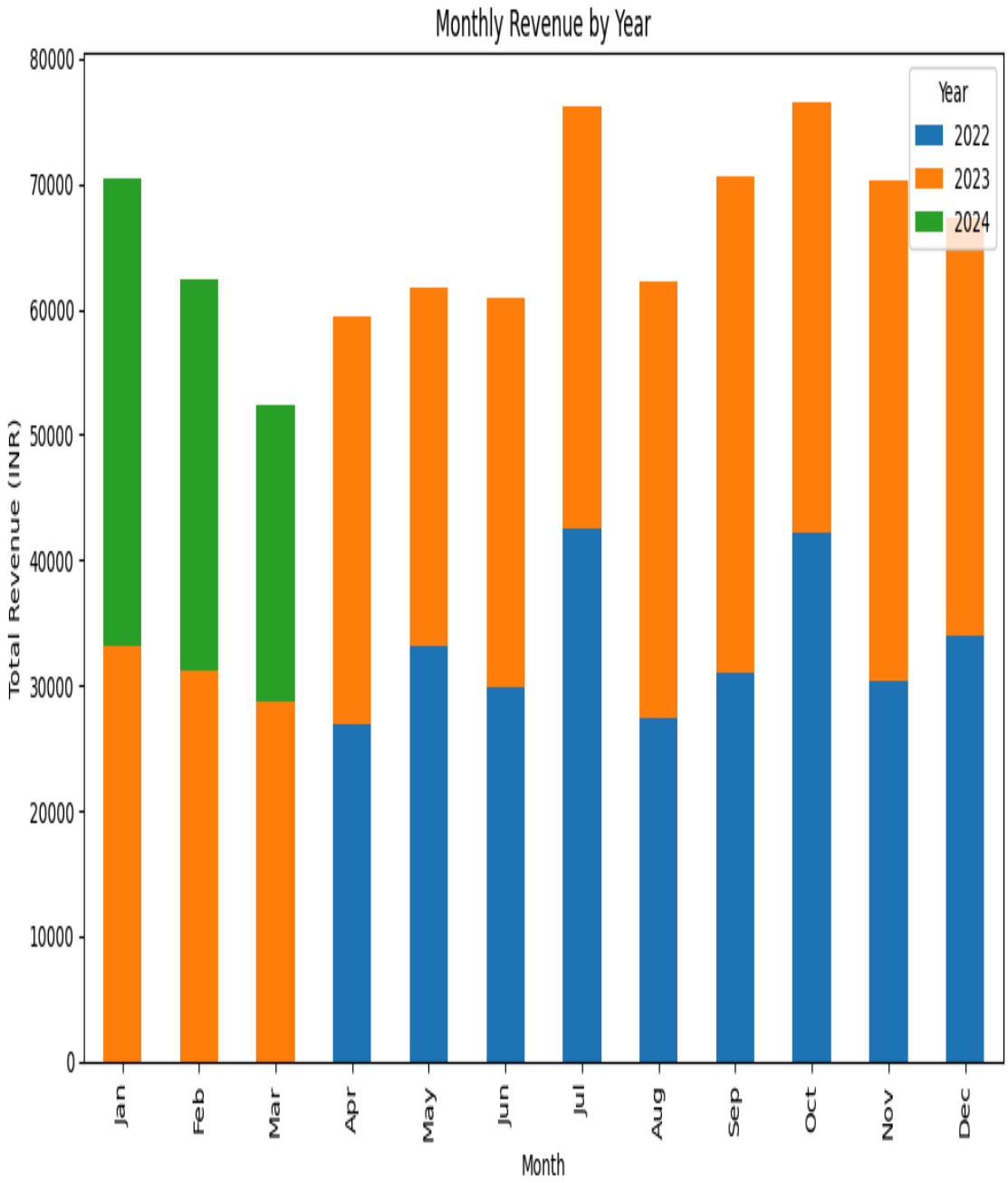


2. On comparing the 1st quarter months' combined revenue , it was found that due to missing data of 1st quarter of 2022 , we can only compare the other 2 years. The trend after comparing 1st quarter of 2023 and 2024 the revenue generated in 1st quarter of 2023 was 93 Thousand and in the 1st quarter of 2024 was 92 thousand. These two stats are almost equal in comparison , with a little decline in the revenue of the year 2024 as compared to that of the year 2023. Now we can infer that we need more analysis on individual month wise data to compare the revenue trends again.



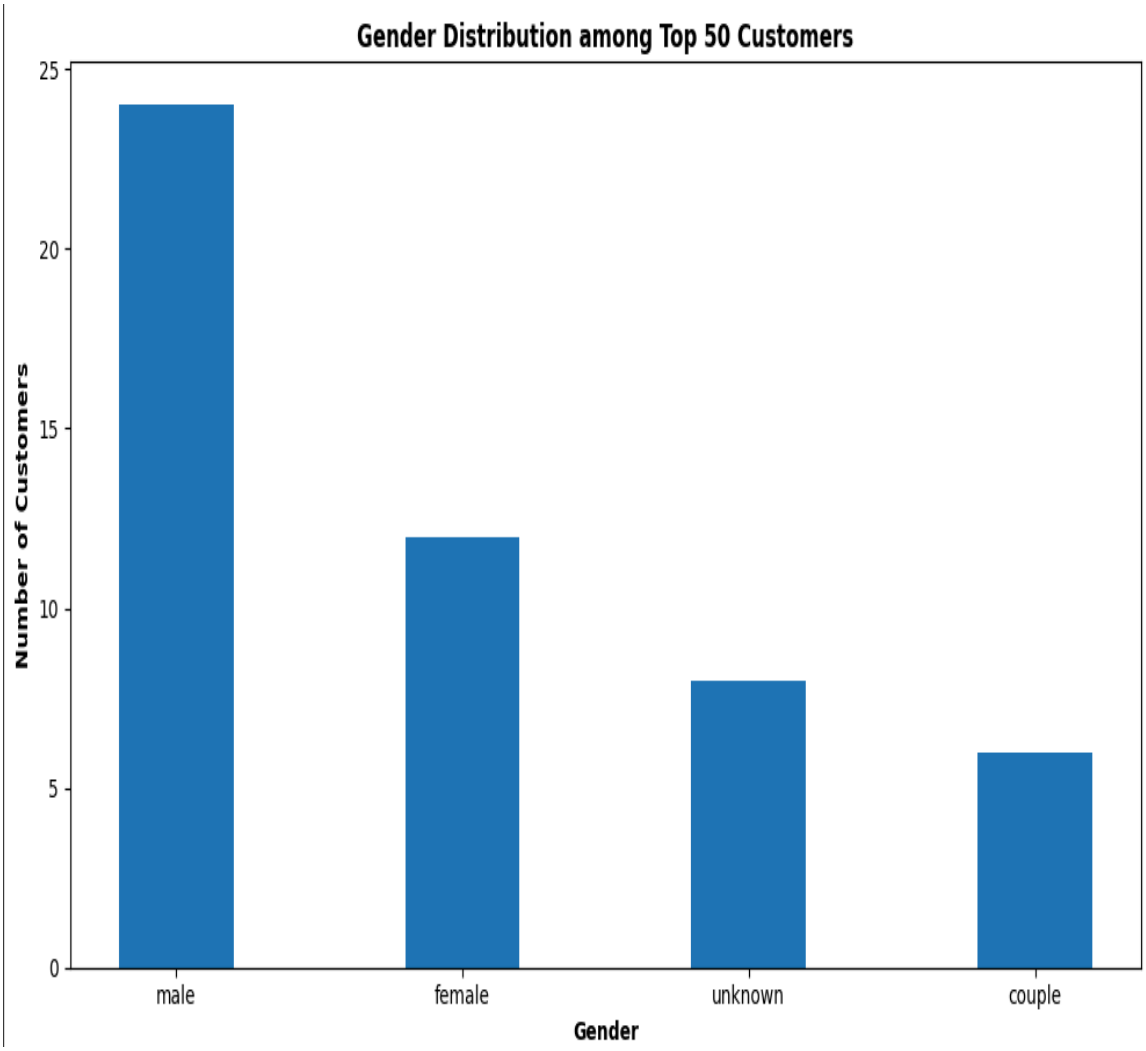
3. The final comparison in the monthly stacked chart shows a very clear trend in the behavior of revenue when the whole data was brought down on a month - to -month basis. The most profitable months for the services in the

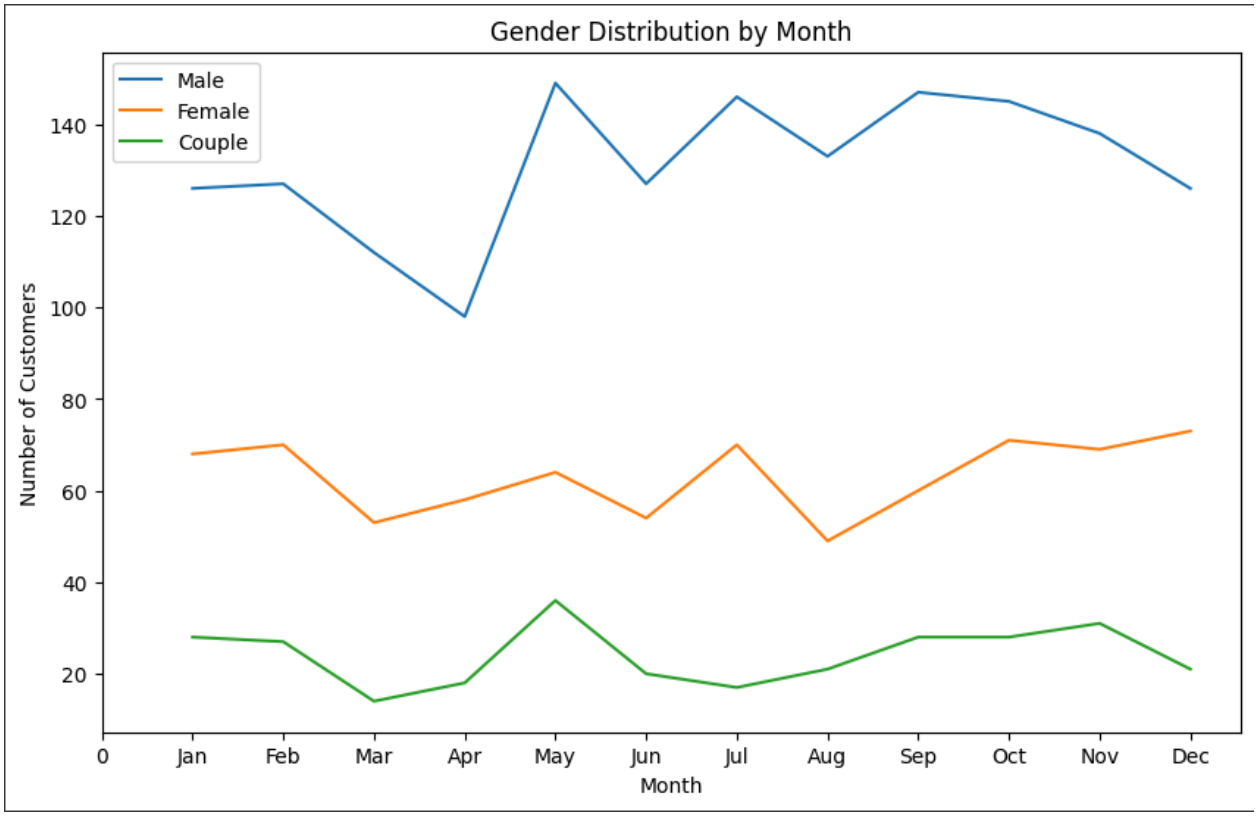
year 2022-2023 are July and october. According to calculations July and October on a combined basis generates 19% of the total revenue in the year. This is due to the fact that in July the Universities nearby reopens after a summer break . For October it can be said that the residential areas have working professionals and they get long holidays for Dushehra and Diwali in the month of October. Hence we can still expect a shoot of 19-20% of revenue in these months at the very minimum . And with proper advertisement strategy this revenue can be again maximized in this particular period.

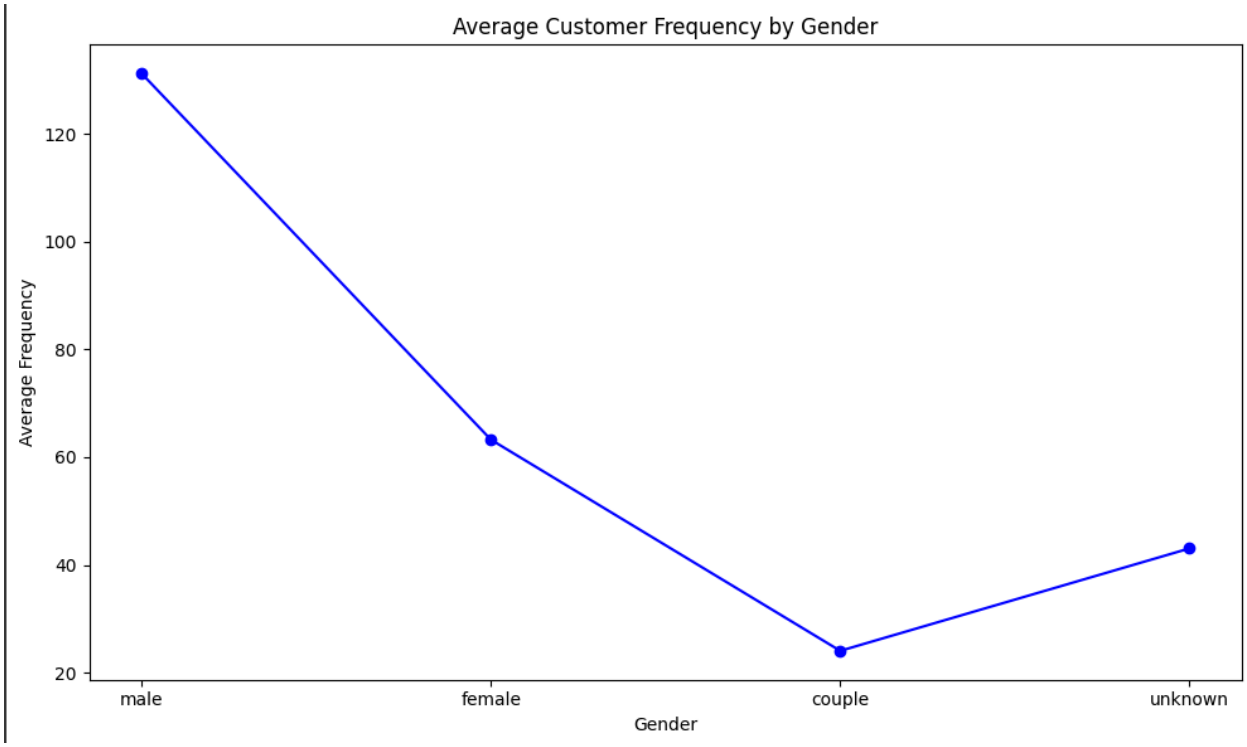


4. On comparing the gender distribution on the basis frequency the inference is direct that over the past two years male customers had been the most frequent over the past two years. Couples had been visiting the salon on very less frequency. The graph also confirms the previous claims that in the month of July and October a spike in revenue can be observed , as in the ‘Gender Distribution by Month’ curve , a spike in male and female customers can be observed in both of these months . According to the

calculations , the highest frequency of the male customer was noted 27 times in 2 years . Which accounts to be 1 per month roughly. When it comes to the average frequency of male , female and couples male customers have two times more frequency as compared to females.

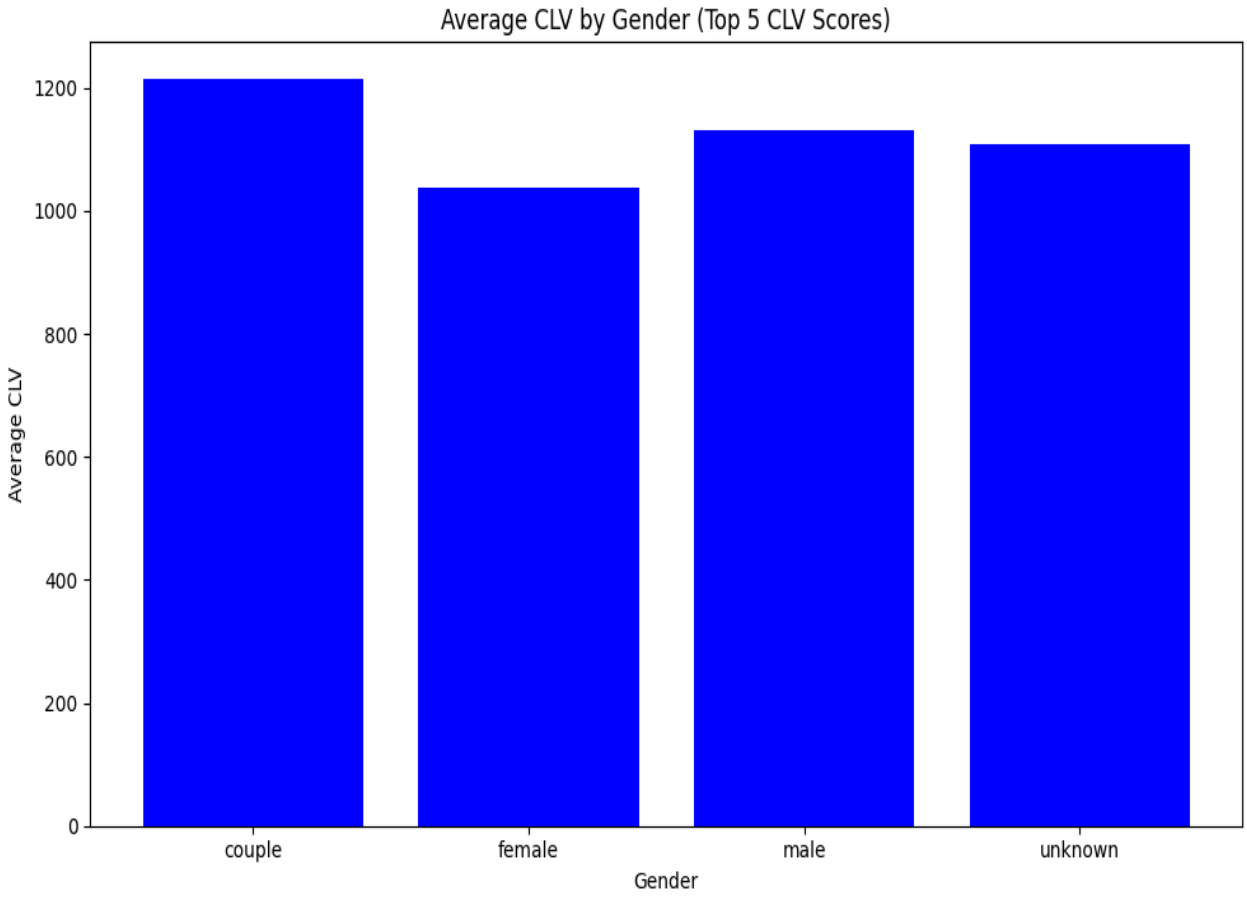



Gender Distribution by Month
Male
Female
Couple
Number of Customers
140
120
100
80
60
40
20
0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month

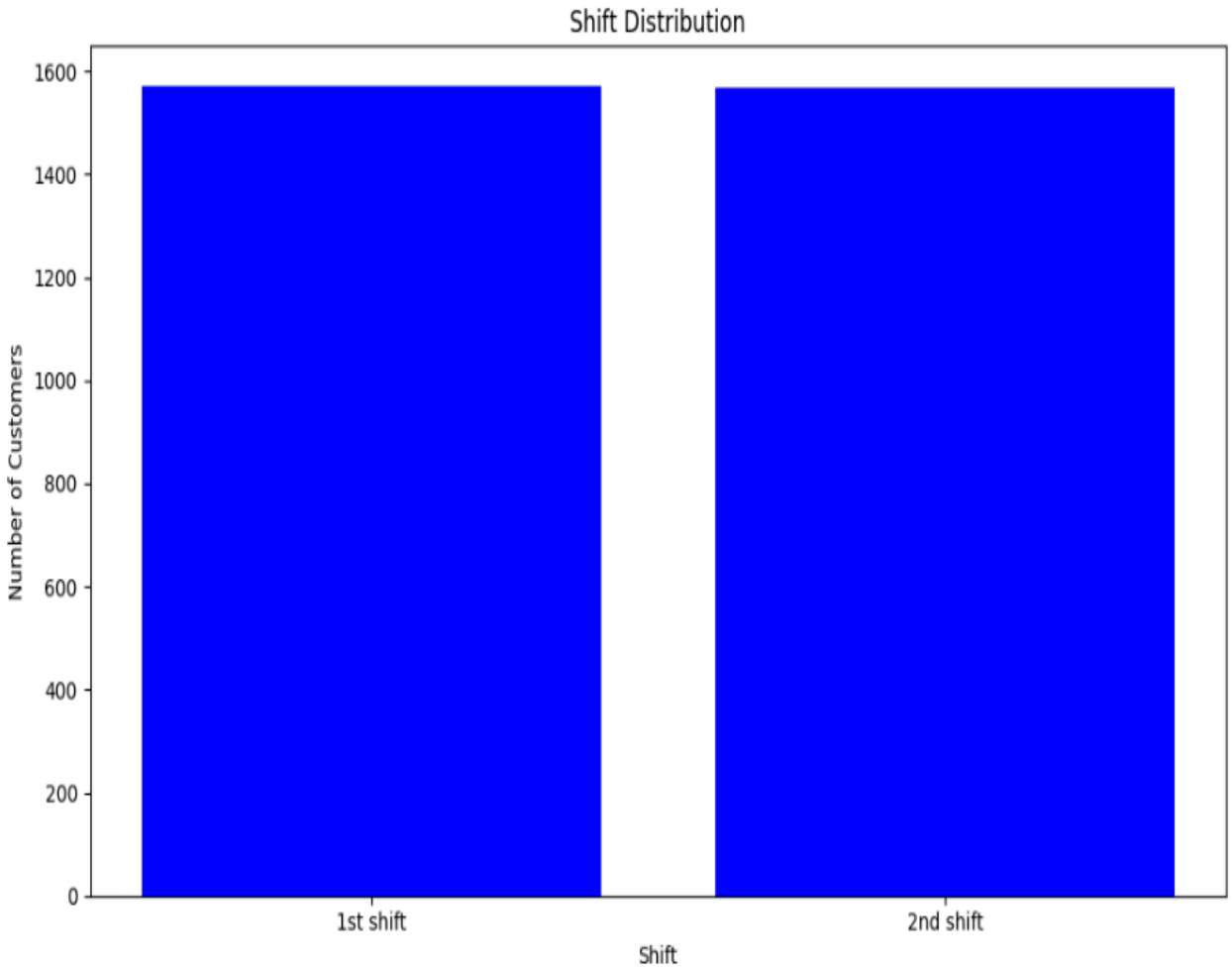

Average Customer Frequency by Gender
Average Frequency
120
100
80
60
40
20
male
female
couple
unknown
Gender

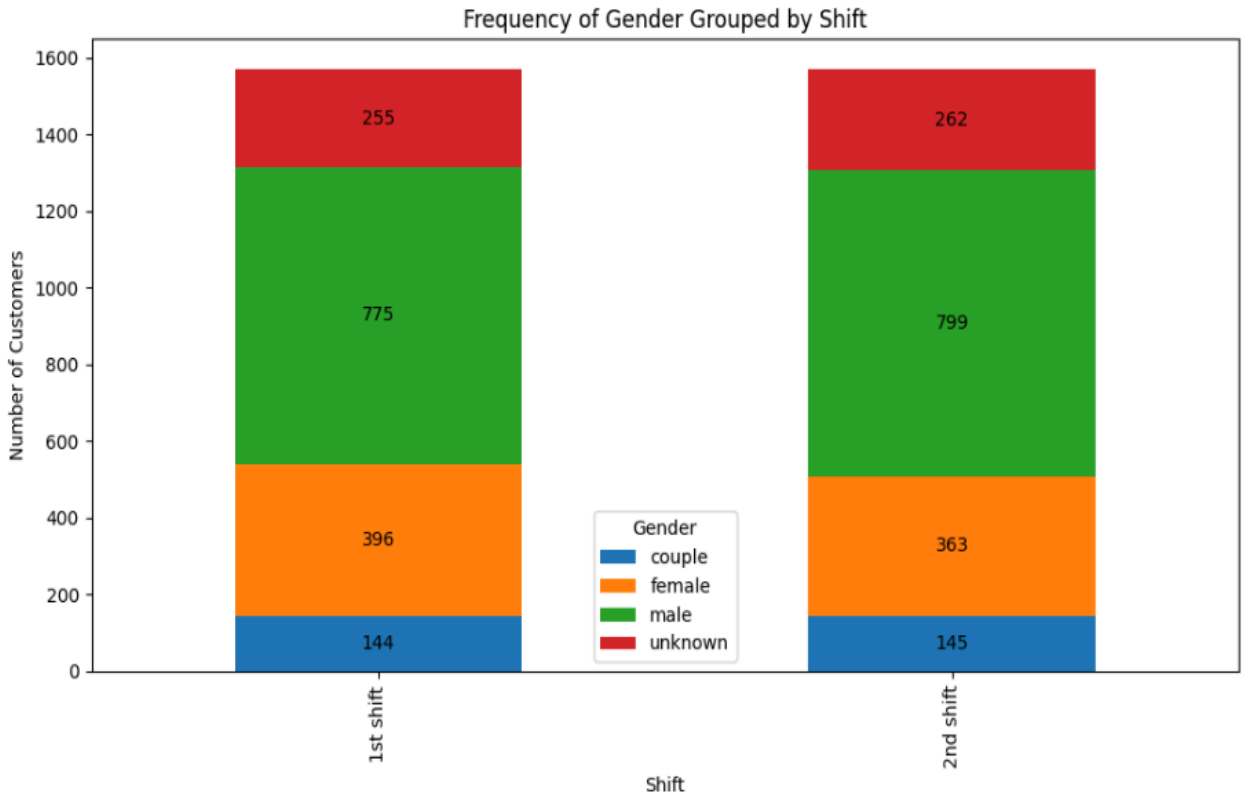
5. On running a proper CLV (Customer Lifetime Value) analysis on the provided data the results were in the favor of Female gender distribution. The CLV analysis shows that Females have the highest average CLV in the entire data. Followed by Couples , Unknown and Males respectively . The calculations show that the female category had `379.441986` CLV score , the couples category had `369.630701` CLV score , the unknown category had `344.896418` CLV score and the male category had `341.606865` CLV score. While the top 5 CLV scores were observed to be contributed more by the couples category as compared to the female category.

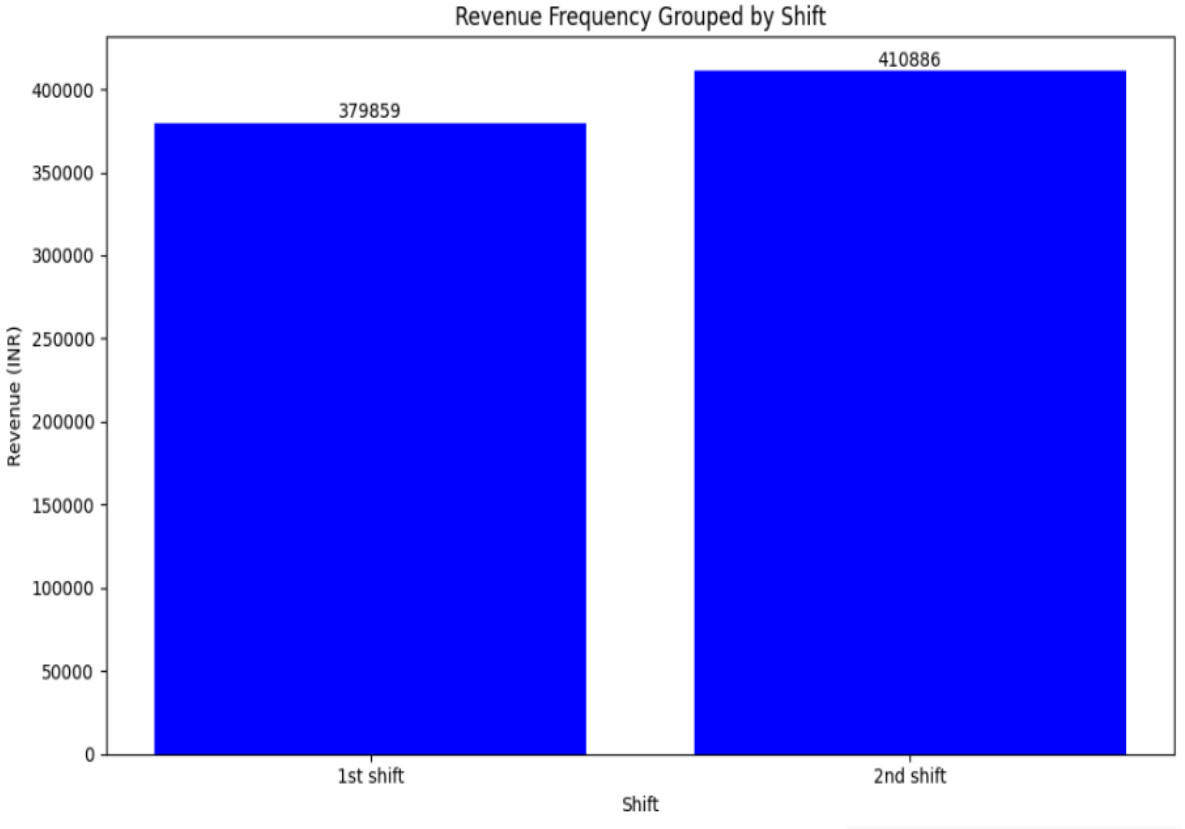
```
Gender
couple     369.630701
female     379.441986
male       341.606865
unknown    344.896418
Name: CLV, dtype: float64
```



6. On running analysis on the basis of Morning and night shift the inferences were very well distributed . None of the shifts was more dominating over others in terms of gender frequency distribution. Both the shifts had almost the same number of male , female , couples and unknown type customers in the past 2 years. Although when it comes to the total revenue generated by the salon in the past two years in both shifts , a slight increase of revenue in the 2nd shift can be observed. According to the calculations , the total revenue generated in the 1st shift in the past two years is 3.8 Lakhs , whereas total revenue generated in the 2nd shift in the past 2 years is 4.1 lakhs. A surge of 30 thousand in revenue can be observed in the 2nd shift. This might be due to the IT companies located within 2 KMs of the salon. The night shift workers are generally paid well and hence if they get the free time , they may visit the salon for refreshing , styling or personal healthcare.




Frequency of Gender Grouped by Shift
Number of Customers
1600
1400
1200
1000
800
600
400
200
0
255
775
396
144
262
799
363
145
Gender
couple
female
male
unknown
1st shift
2nd shift
Shift


Revenue Frequency Grouped by Shift
Revenue (INR)
400000
350000
300000
250000
200000
150000
100000
50000
0
379859
410886
1st shift
2nd shift
Shift

7. On comparing the locality , specially collecting information from the local schools , colleges and universities the common inference which could be obtained is that there is a general scheme of holidays on Saturday and Sunday for each and every week of the month. The schools have summer holidays in the month of April to May , which is accounted to be the longest holiday by them. The schools generally resume by the end of May. For the Universities , the longest break is in Summer vacation after the end of a semester from April to June , in which students generally go for summer Internships. The Universities generally resume their curriculum by the 1st week of July.

4. Interpretation of the results and Recommendations

With the following recommendations and results I would like to conclude my report.

1. The salon had made through the 1st quarter and had slightly less revenue as compared to the 1st quarter of the last year. Hence it is recommended to gear up the promotional materials and the methods for the advertisement to focus on the months and the timelines which can provide the maximum profit . Otherwise the salon's revenue will be naturally declining for this year. As they are going to burn money on the advertisement , they need to make sure of a good revenue to balance the burn.

2. Salon should focus more on advertising their materials in the university and school areas in the month of July . July is potentially a good profitable month due to the comeback of school and university students from a long break.

3. Salon should also focus on advertising more in the residential areas in the month of October as that is the most profitable month from residential people as they plan to go on a vacation in those months. They are likely to spend more on beautification before visiting on a trip.

4. The male and female sector are stable markets for the salon . As the frequency of the male and female customers are quite decent. But as per the analysis , couples and females are more likely to generate a very high margin business for the salon . So if the salon wants to advertise in a safe zone they can just target male and females individually. If the salon is ready to take some risk , then they can come up with some combos and discounts specially for couples to retain and increase the couple's visit frequency. In the long run females and couples are the most profitable customer segments.

5. Other than Haircuts , Beard trimming and eyebrow threading are the most profitable service segments. Salon can make discount strategies based on

this . One such example is :- If someone is taking a service worth more than INR 1000 , beard trimming or eyebrow threading should be a free service.

6. The revenue generated by the 2nd shift or after 8:00 PM can be maximized if the salon focuses on advertising their services in Industrial areas.