

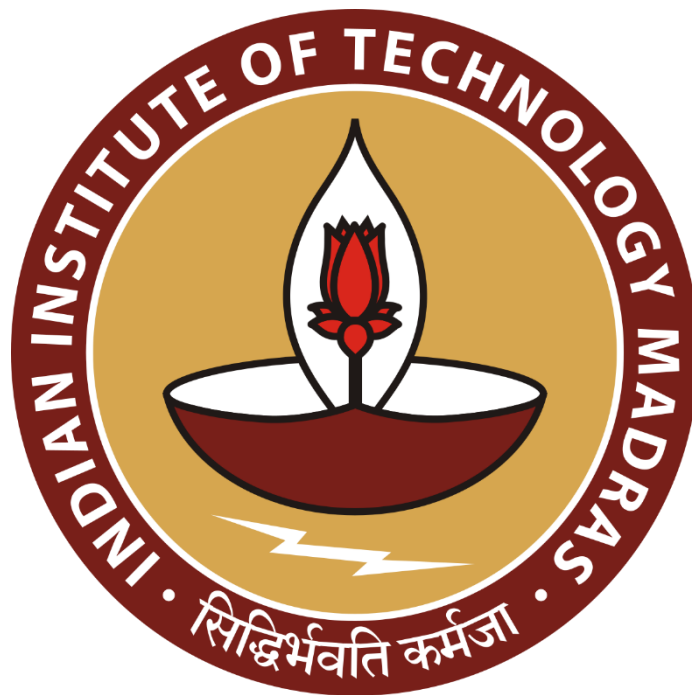
Improving Customer Acquisition at Gypsy Soul Holidays

A Mid Term report for the BDM capstone Project

Submitted by

Name: Aarav Aashish Mody

Roll number: 22F3002269



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

Contents

| | | |
|---|---|----|
| 1 | Executive Summary and Title | 2 |
| 2 | Proof of originality of the Data | 2 |
| 3 | Metadata | 3 |
| 4 | Descriptive Statistics | 5 |
| 5 | Detailed Explanation of Analysis Process/Method | 6 |
| 6 | Results and Findings | 7 |
| 7 | Appendix A | 10 |

1 Executive Summary and Title

The travel company "Gypsy Soul Holidays" is a startup located in one of Mumbai's busiest places, Kemps Corner. It is a B2C business, which means the company deals directly with customers in the travel segment. In a very competitive market, it tries to build a strong foothold. Since it is a start-up business, one of the major problems faced by the company is making the prospective clients believe in the ability and reliability of the company. Further, the firm has a challenge recruiting trained and experienced staff at their competitive salaries.

The approach that will be used for this project in overcoming such challenges will be by finding solutions by a data-driven approach using tools like Excel and Python-machine learning approach where data is analysed with an aim to extract insights that could be useful.

The travel company can reduce operational inefficiency and enhance the overall business performance by fastening up the recruitment process. By performing sales analysis, the company can identify when and what services lead to high margins and hence try to increase selling those services. It can also try and find solutions for off-season by reducing prices by a little percentage. The company can incentivise employees by giving bonuses for every client and moreover give discounts to regular customers.

The desired outcomes will include more efficient and effective customer acquisition strategies, better and faster hiring, and an improved market position. Combining data analytics into the decision-making process, Gypsy Soul Holidays will be able to address few problems and hence lay a strong base for future growth, and be at a better place to find competitive advantage in the travel industry.

2 Proof of Originality of the Data

Details:

Travel Firm – Gypsy Soul Holidays

Owner – Ms. Ashmi Dharia

Location – Mumbai

Letter from organisation -

<https://drive.google.com/file/d/1KpIUISf0L8FQWQT9QeunketD9YVfklQ/view?usp=sharing>

Short video interacting with the owner -

<https://drive.google.com/file/d/1xNGpEAMHRke6sWn2GesLAu2775ktYzry/view?usp=sharing>

Transcription for the short video – <https://drive.google.com/file/d/1LyX9-hQbRdLIorRIRhcNKPj1R3747/view?usp=sharing>

Images of the firm: Included in the link and also attached below - https://drive.google.com/drive/folders/1468T_gw389PsmWwmOI-5UsQEQ312YMSe?usp=drive_link

Dataset:

https://docs.google.com/spreadsheets/d/1i-u_sLOYgYrkfjQAN1SQggWATmLhb7l/edit?usp=sharing&ouid=105891692553322065596&rtpof=true&sd=true

BDM Project folder (All data stored in this folder) - <https://bit.ly/bdmproject>

The owner, Ms. Ashmi Dharia works in her office, the photo of the same is given below



(2.1) Office workspace

3 Metadata

The owner granted me access to the bookings data over the period of one year from April 2023.

It includes:

1. **Date:** Date of invoice.
2. **Invoice number:** It is used to track and reference specific transactions.
3. **Pax Name:** Name of the traveller.
4. **Pax:** Number of people.
5. **GST:** Traveller's GST number.
6. **Service:** Type of service provided like a ticket, or a complete package.
7. **Supplier:** Supplier or service provider
8. **Amount Received:** Amount received for the transaction.
9. **CGST, SGST, IGST:** These columns represent different types of taxes paid.
10. **Amt Paid:** The amount paid out to the supplier for the services provided.
11. **Margin:** The profit earned by Gypsy Soul Holidays before tax.

12. **Receipt Date:** The date on which the amount was received.

13. **Payment Date:** The date on which payment was made to the supplier.

14. **Supplier Inv No:** The supplier's invoice number, used for tracking and reference.

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q |
|----|-----------|-----------------|----------------------------|-----|-----------------|---------|-----------------------------|-----------|-------|-------|------|-----------|--------|-------------------------|---|-----------------|--------------|
| 1 | Date | Inv / CN No | Pax Name | Pax | GST | Service | Supplier | Amnt Recd | COST | SOET | IGST | Amnt Paid | Margin | Receipt date | Payment Date | Supplier Inv No | Supplier GST |
| 2 | 01-Apr-23 | GSH2023-2400001 | Jade (Zameer Moon) | 1 | 27AAGFJ4705R1ZU | Ticket | TBO | 72274 | 347 | 347 | | 5381 | 3850 | recd 81554 on 17 April | 3779 paid on 1 April | IW23241415 | |
| 3 | | | | | | | | | | | | 5381 | | | 5381-adjusted with excess payment made to TBO | IW23241775 | |
| 4 | | | Diya Singh and Pooja Joshi | 2 | | | | | | | | 13098 | | | 21450 paid on 10 April | IW2324094330 | |
| 5 | | | | | | | | | | | | 8352 | | | 37120 Paid on 14 Apr 23 | IW232404524 | |
| 6 | | | zameer Moon BOMMO | | | | | | | | | 5066 | | | | IW2324156046 | |
| 7 | | | zameer Moon AMODEL | | | | | | | | | 7863 | | | | IW2324171882 | |
| 8 | | | zameer Moon DELBOM | | | | | | | | | 7201 | | | | IW2324171702 | |
| 9 | | | Monica BOM DEL | | | | | | | | | 7089 | | | | IW2324172275 | |
| 10 | | | Monica DEL BOM | | | | | | | | | 10101 | | | | IW2324191640 | |
| 11 | | | | | | | | | | | | 67730 | | | | | |
| 12 | | | | | | | | | | | | | | | | | |
| 13 | 03-Apr-23 | GSH2023-2400002 | Ghungroo Shrivastava | 2 | 30BAUP83914C1Z9 | Ticket | Heat Flexi Holidays | 108700 | | | 900 | 99800 | 5000 | 50000 on 4 April | 99800 paid on 03 April | | |
| 14 | | | | | | | | | | | | | | 55700 on 5 April | | | |
| 15 | | | | | | | | | | | | | | | | | |
| 16 | 05-Apr-23 | GSH2023-2400003 | Wint Bekat | 4 | | Package | Nam Ho DMC / PawarSut Fores | 469030 | 11168 | 11168 | | 223581 | 63574 | 50000 on 14 Mar | 223581 on 05 April | | |
| 17 | | | | | TCS 9% | 23452 | TBO | 492482 | | | | 150429 | | 150000 on 15 Mar | 150429 on 17 March | IW22235331904 | |
| 18 | | | | | TOTAL | 492482 | Visa Lounge | | | | | 9110 | | 100000 on 24 Mar | 9110 paid on 24 April | VW 23-10038 | |
| 19 | | | | | | | | | | | | | | 192482 on 05 Apr | | | |
| 20 | | | | | | | | | | | | 383120 | | | | | |
| 21 | | | | | | | | | | | | | | | | | |
| 22 | 05-Apr-23 | GSH2023-2400004 | Amrita Vakil | 2 | | Package | Heat Flexi | 312240 | 7435 | 7435 | | 246750 | 50620 | 327852 on 05 April 2023 | 246750 on 05 April 2023 | | |
| 23 | | | | | TCS 9% | 19612 | | | | | | | | | | | |
| 24 | | | | | TOTAL | 327852 | | | | | | | | | | | |
| 25 | | | | | | | | | | | | | | | | | |
| 26 | 06-Apr-23 | GSH2023-2400005 | Prnya Kapadia | 3 | | Ticket | TBO | 77100 | 1080 | 1080 | | 75210 | 12000 | 224160 on 10 April 23 | 210000 paid on 10 April 23 | IW2324091130 | |
| 27 | | | | | | | | 147080 | | | | 134790 | | | | IW2324091587 | |
| 28 | | | | | | | | 234160 | | | | | | | | | |
| 29 | | | | | | | | | | | | | | | | | |
| 30 | 10-Apr-23 | GSH2023-2400006 | Anuran Modi | 4 | | Hotel | Omnia | 47755 | 171 | 171 | | 45813 | 1900 | 10-Apr-23 | 11-Apr-23 | RCS-181404 | |

(3.1) A snippet of the data provided by the firm

Data Cleaning

| Row Labels | Sum of Revenue | Row Labels | Sum of Revenue | Business type |
|--------------------|-------------------|-----------------|----------------|---------------|
| Car | 173968 | Car | 173968 | Others |
| Cars | 220167 | Cars | 220167 | Others |
| Cruise | 99320 | Cruise | 99320 | Others |
| Go City Pass | 28060 | Go City Pass | 28060 | Others |
| Guide | 4452 | Guide | 4452 | Others |
| Hotel | 1149485 | Insurance | 24339 | Others |
| Insurance | 24339 | Marhaba | 26639 | Others |
| Marhaba | 26639 | Meet and Assist | 62220 | Others |
| Meet and Assist | 62220 | Transport | 23906 | Others |
| Package | 6181659.9 | Visa | 283168 | Others |
| Ticket | 12789926 | Package | 6181659.9 | Top 3 |
| Tickets | 171339 | Ticket | 12961265 | Top 3 |
| Transport | 23906 | Hotel | 1149485 | Top 3 |
| Visa | 283168 | | | |
| Grand Total | 21238648.9 | | | |

| Sum of Revenue | Column Labels | | | | | | | | | | | | | | | | |
|--------------------|---------------|---------------|---------------|--------------|--------------|-------------|----------------|--------------|-----------------|--------------|------------------|-----------------|---------------|--------------|---------------|-------------------|--|
| Row Labels | Car | Cars | Cruise | Go City Pass | Guide | Hotel | Insurance | Marhaba | Meet and Assist | Package | Ticket | Tickets | Transport | Visa | Grand Total | | |
| January | | 44867 | | | | 27907 | 390 | | 62220 | 178040 | 811996 | | 23906 | | 1149326 | | |
| February | | 22446 | | | | | 492 | | | 1456547 | 657965 | | | 20176 | 2157626 | | |
| March | | | | | | 13041 | | | | 514761 | 380213 | | | 20120 | 928135 | | |
| April | 25687 | | | | | 74368 | | | | 1672162 | 1132173 | | | 17552 | 2921942 | | |
| May | | | | | | | 17998 | | | | 2637490 | | | 49870 | 2705358 | | |
| June | 18612 | | | | | 90348 | | | | 218293.9 | 596807 | | | 9000 | 933060.9 | | |
| July | 9992 | | | | | | | | | | 748852 | 171339 | | | 930183 | | |
| August | 41189 | | | | | 108556 | 1001 | | | | 1607647 | | | | 1758393 | | |
| September | 0 | | | | | 587071 | 4458 | 26639 | | | 1590936 | | | 64400 | 2273504 | | |
| October | 78488 | | 99320 | | | 42533 | | | | 560683 | 1557861 | | | 71580 | 2410465 | | |
| November | | 62138 | | 28060 | 4452 | 54601 | | | | 999368 | 689388 | | | 30470 | 1868477 | | |
| December | | 90716 | | | | 151060 | | | | 581805 | 378598 | | | | 1202179 | | |
| Grand Total | | 173968 | 220167 | 99320 | 28060 | 4452 | 1149485 | 24339 | 26639 | 62220 | 6181659.9 | 12789926 | 171339 | 23906 | 283168 | 21238648.9 | |

4 Descriptive Statistics

| | Pax | Amount Received | Revenue | Amt Paid | Margin |
|---------------------------|-------------|------------------------|----------------|-----------------|---------------|
| Sum | 288 | 21683956.9 | 21239464.9 | 8661194.78 | 2215368 |
| Mean | 2.102189781 | 112937.2755 | 109481.7778 | 46316.54963 | 11974.96216 |
| Standard Deviation | 1.491567827 | 160374.3121 | 155235.8585 | 85846.70753 | 31122.47982 |
| Median | 1 | 46550 | 47882.5 | 13010.33 | 3500 |
| Minimum | 1 | 374 | 0 | 0 | 100 |
| Maximum | 7 | 1122678 | 1069217 | 470997 | 343714 |

Variance Statistics

| | Variance |
|---------------------|-----------------|
| Pax | 2.22 |
| Amt Received | 25719919977.81 |
| CGST | 4471519.34 |
| SGST | 4471519.34 |
| IGST | 251022599.88 |
| Total GST | 32424865.63 |
| Revenue | 24098171762.98 |
| Amt Paid | 7369657193.97 |
| Margin | 968608750.43 |

Pax: A very low variance of 2.22 would imply that passenger counts are relatively stable across the dataset

Amt Received: Very high variance of 25,719,919,977.81 would suggest wide fluctuations in money received, possibly showing big differences in customer spending or variability in sales.

Margin: With an especially medium-sized variance of 968,608,750.43, it does indicate that there are differences in profit margins across the various transactions as different people opt for different services.

Skewness Statistics

| | Skewness |
|---------------------|-----------------|
| Pax | 1.279687 |
| Amt Received | 2.704709 |
| CGST | 4.517949 |
| SGST | 4.517949 |
| IGST | 2.640515 |
| Total GST | 5.571105 |
| Revenue | 2.673455 |
| Amt Paid | 3.021296 |
| Margin | 7.79246 |

Pax: Skew = 1.28, indicating that the passengers are reasonably right-skewed, meaning most trips carry few passengers and few trips have large groups.

Revenue & Amt Paid: Right-skewed (2.67 and 3.02, respectively), showing most of the sales are of smaller amounts; only a few are of high values.

Margin: Very right-skewed at 7.79, indicating that while most services yield lower margins, a few transactions can be very profitable.

Correlation Matrix

| | Pax | Amt Received | CGST | SGST | IGST | Total GST | Revenue | Amt Paid | Margin |
|--------------|-------|--------------|------|------|------|-----------|---------|----------|--------|
| Pax | 1.00 | 0.17 | 0.30 | 0.30 | 0.06 | 0.20 | 0.17 | 0.23 | 0.15 |
| Amt Received | 0.17 | 1.00 | 0.63 | 0.63 | 1.00 | 0.73 | 1.00 | 0.51 | 0.73 |
| CGST | 0.30 | 0.63 | 1.00 | 1.00 | | 1.00 | 0.62 | 0.28 | 0.95 |
| SGST | 0.30 | 0.63 | 1.00 | 1.00 | | 1.00 | 0.62 | 0.28 | 0.95 |
| IGST | -0.06 | 1.00 | | | 1.00 | 1.00 | 0.99 | 0.20 | 0.99 |
| Total GST | 0.20 | 0.73 | 1.00 | 1.00 | 1.00 | 1.00 | 0.72 | 0.22 | 0.96 |
| Revenue | 0.17 | 1.00 | 0.62 | 0.62 | 0.99 | 0.72 | 1.00 | 0.51 | 0.71 |
| Amt Paid | 0.23 | 0.51 | 0.28 | 0.28 | 0.20 | 0.22 | 0.51 | 1.00 | 0.13 |
| Margin | 0.15 | 0.73 | 0.95 | 0.95 | 0.99 | 0.96 | 0.71 | 0.13 | 1.00 |

Pax and Amt Received: Weak positive correlation (0.173). This means that the more passengers, the less proportional increase in revenue earned.

Amt Received and Total GST/Revenue: 0.73/0.99. That means very high correlation, showing that amount received is much related to total revenue and tax figures.

Revenue and Margin: 0.71—This positive correlation means that high revenue generally brings high margins.

Margin and IGST: The high correlation between margin and IGST goes on to prove that a few of the services generating very high margins are subjected to heavy taxation.

5 Detailed Explanation of Analysis Process/Method

12 months of data was cleaned, organised and analysed.

Firstly, the data maintained by the startup was a blend of clean and mix data, hence not in an organised way on which analysis could be done, so based on the data, what insights can be driven was found. The 12 months dataset was inconsistent structurally and hence needed an in-depth look into any discrepancies such as missing values, duplicate information, and inconsistent formatting. One of the key tasks was to line up the data by date, so that every entry would be in chronological order for analysis.

Next in line for cleaning, calculate the Total GST, which was the summation of CGST, SGST, and IGST values. The column gave a clear view of the tax added. Then, adding up the

GST to the amount received plus the margin gave the total revenue for each transaction. GST subtracted from this gave the net revenue.

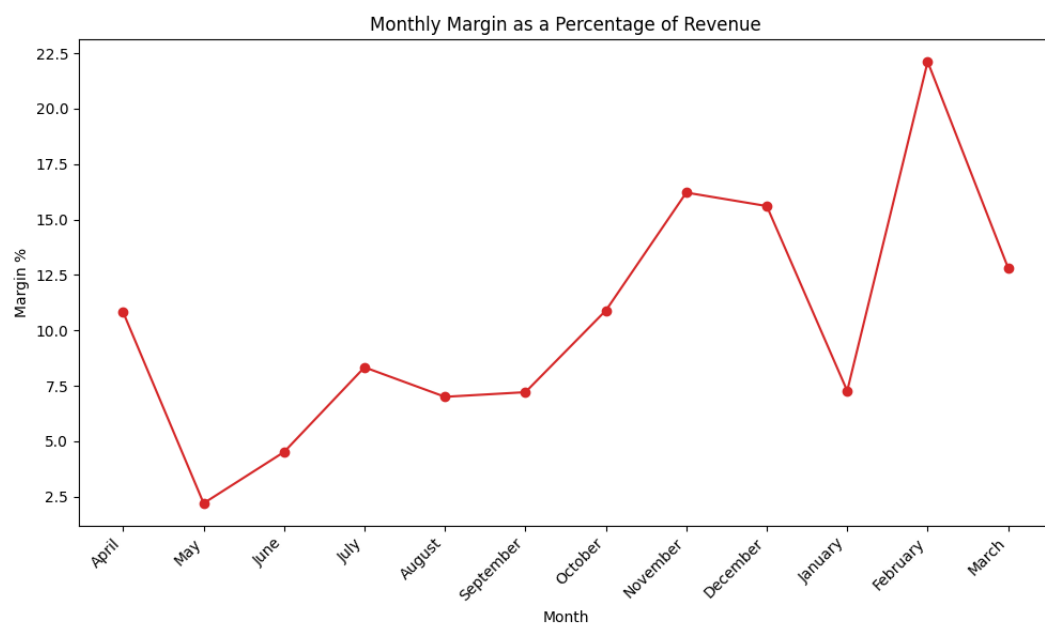
The data was then sorted by date. It was also categorized and separated according to the relevant month, making the data useful for reviewing and comparing monthly performance. This data was reviewed briefly and the key information was filtered out for further analysis.

The information was then summarized and rearranged using pivot tables to enable the extraction of any key insights and patterns to be observed across the various dimensions, such as over time, by service type. Pivot tables permit very powerful dynamic analysis whereby the view of data might be changed quickly, and underlying trends identified.

After, several visualizations were created to facilitate insight into trends and distributions within this data. Line chart was used to show trends in total revenue over the year, which helps us see how these metrics change month by month. Then the core business was analysed and it was found that there were 3 services which generated 96% of the revenue and hence from those services, a pie chart was created to find which service generates the most revenue. A bar chart was drawn to see which service accounts for maximum % of margin. Using these graphs, the peak season for the travel firm was found and the most profitable service was found. Using this data, solutions can be found to increase customer acquisition and hire workers.

The steps helped transform the data from a jumbled collection of records to a structured and insightful data. The cleaning and analysis of data was done using excel and python libraries like numpy, pandas, matplotlib and seaborn.

6 Results and Findings



(6.1) Monthly Margin as a Percentage of Revenue

Based on the graph a few things can be observed -

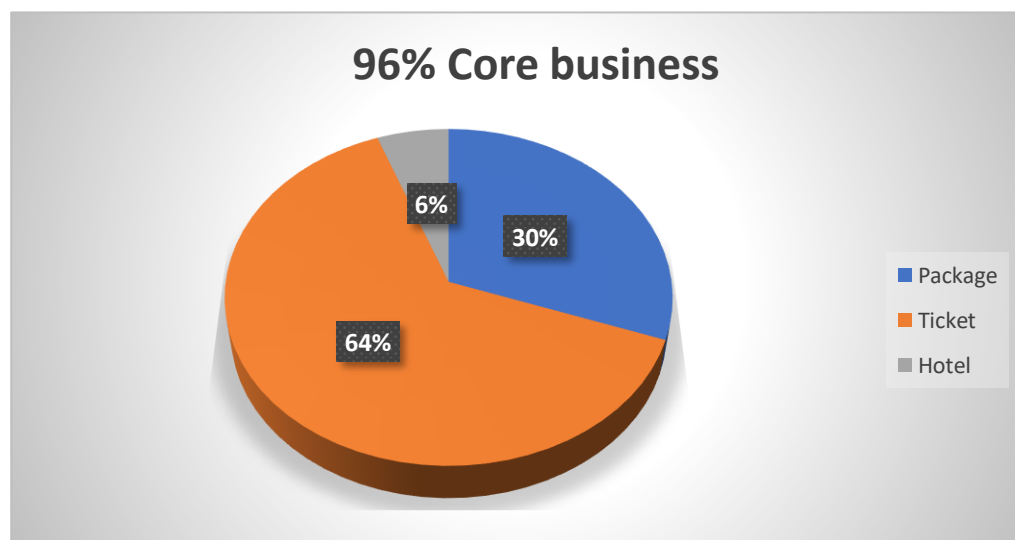
The revenue pattern is fluctuating throughout the whole year. This goes on to prove the variability in business performance, because tourism is very seasonal oriented.

July to November shows an increase in demand for travel as it is festival season in India hence many people prefer to travel during these months.

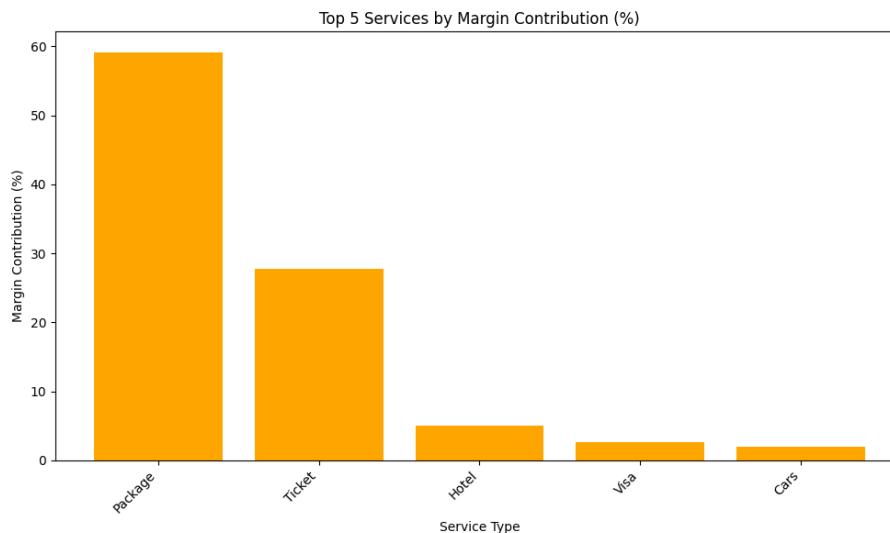
However, there is a sharp decline in December which may not be expected because of Christmas and New Year holidays but a reason for decline in sales may be because the prices in December are very high therefore people avoid travelling during this time.

By looking at these trends, the challenge of workforce hiring, coupled with customer acquisition, can be addressed by adopting a multi-dimensional approach. Given the fluctuating revenue pattern and seasonality of tourism, one approach would be to hire part-time or seasonal workers during peak demand periods such as February, July, and the festival season. Additional workload can be managed without long-term financial burdens of maintaining a full-time workforce throughout the year.

Additionally, using off-peak months such as June, September, and December to focus on discounted travel packages or early-bird deals for the peak season can also reap benefits. This may attract price-sensitive customers who do not prefer to travel during peak seasons because of high prices, thereby increasing sales during months that usually generate low revenue. Moreover, digital marketing campaigns based on demography, students and families, will be able to capture more customers during summer and festive holidays. It is through staff planning and seasonal marketing strategies the business can maximize its revenues and manage employees effectively.



(6.2) Core business' proportion to total revenue



(6.3) Top 5 services by Margin (%)

After filtering out the top 3 services which account for the major revenue, it is found that providing tickets account to maximum revenue which is 64% of the 96% revenue, followed by packages at 30%. This is because majority of the people take help in buying tickets at a cheaper rate without doing the hassle themselves and only a few percentages of people book the entire package for a trip.

The pie chart on sales distribution clearly shows that ticket sale forms the dominant aspect of this business. This then implies that ticket service can be one major pull factor for customers to the business since it pulls the highest proportion of clients.

However, turning to the profit margins, as high as ticket sales are, we note the margins are rather low compared to the other services, particularly package sales. Packages bring significantly higher margins, making them the most profitable service despite lower sales volume.

Given the fact that ticket sales are instrumental in attracting customers, a possible strategy would be to employ the loss leader model; the margins on ticket sales, while not being sold at a loss, could be further reduced to attract more customers. The meaning here is that more customers would go through the business funnel via ticket services.

The intention is to turn ticket buyers into package buyers after acquiring a customer. This could be done through tailor-made packages and, therefore, the sale of a package would capture a higher margin.

Each time a problem is with respect to how to persuade the employees to effectively make conversions of ticket sales to package sales and how to reduce the hassle of hiring in the workforce, an incentive system for the employees who effectively make such conversions in ticket-sales could be initiated. This will encourage employees if a share is given for the profit gained when their conversions succeed. The other benefit includes the improvement in the persuasion of the customer. In addition, such incentivization will improve employee retention without raising the base salary amount while, at the same time, they are also likely to attract new employees since it is hard to get such an opportunity in the industry.

APPENDIX A: Proof of originality



Date: 01 August 2024

To Whom It May Concern,

Subject: No Objection Certificate for Data Usage

This is to certify that I, Ashmi Dharia, the Proprietor of Gypsy Soul Holidays with the office located at D/G/9, Anand Nagar, ~~Forrest Street, Tarden~~, Mumbai 400026 have no objection to Aarav Mody using the data provided by us for the purpose of analysis.

The data in question may be used solely for analytical purposes. We understand that this data will be handled in accordance with applicable data protection and privacy regulations.

We affirm that there are no objections to the use of this data for the stated purpose and that all relevant permissions have been granted.

Should you have any further questions or require additional information, please feel free to contact us.

Sincerely,

Ashmi Dharia

Founder and Holiday Curator

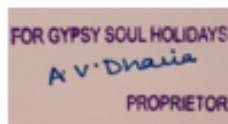
Gypsy Soul Holidays

D/G/9, Anand Nagar, ~~Forrest Street, Tarden~~, Mumbai 400026

+91 9820782087 // +91 8104260074

www.gypsypsoulholidays.com

GSTIN: 27ADYPD6214G1ZE



Address: D/G/9, Anand Nagar, ~~Forrest Street, Tarden~~, Mumbai 400026 // Mobile: + 91 9820782087 // +91 8104260074

BDM Project folder - <https://bit.ly/bdmproject>