

Biz Consulting Company Sales Performance

A. Data Wrangling

At first, meeting data was received from the S3 bucket in AWS , dumped by Client and informed by our Manager Amit Makode and then extracted to an Excel sheet.

The data from the Excel sheet was analyzed, and the Pin code, City, and State were separated from the address column, creating new columns for these names.

Special symbols were removed from the data to eliminate anomalies, especially in Business names and addresses, and the meeting data was properly structured.

All columns were checked for any anomalies, and the meeting date and calling date were arranged in sequence-wise order when found different.

Challenges Faced:

- Extracting the Pin code from the address was difficult as its location varied, but it was eventually separated out through the flash fill and regular expression
- The dates were not in the proper format, and there was a considerable difference, for example, the meeting date being 22-05-2022 and the calling date being 26-06-2023. Since the calling date should always be earlier than the meeting date, it was corrected.
- The main challenge was structuring the data into the proper format, which required a significant amount of time and effort.
- There were many spelling mistakes throughout the meeting data, which were corrected as required.

Imputation:

Values were imputed column-wise as follows:

- Blank calling date and meeting date values were imputed using the mode.
- Blank meeting time values were imputed using the mode.
- Blank Tele caller name and BDM name values were imputed by considering the values in the cells above and below and imputing the appropriate one.
- The Map column had three values: Unverified, Done, and Not Done.
- The Meeting Status was distributed among "call and go" and "Confirm."
- Remaining blank values in the Business Category were filled according to the business names.
- Blank values for city, state, and Pinc ode were imputed according to the local address.
- Remaining blank values in the Meeting table were replaced with "No Data."

Secondly, the login data received from the manager was dumped in the S3 bucket in AWS and then extracted to an Excel sheet.

Again, the Pin code, City, and State were removed from the Address column, and new columns were created for them. Anomalies in all columns were checked and corrected, and then the login data was structured into the required format.

Challenges Faced:

- Extracting the Pin code from the address was difficult as its location varied, but it was eventually separated out using formulas after significant effort.
- The dates were not in the proper format, with some being in dd-mm-yyyy and others in dd/mm/yyyy, so they were converted to dd-mm-yyyy.
- The main challenge was faced in structuring the data into the proper format, which required a significant amount of time and effort.
- There were many spelling mistakes throughout the meeting data, which were corrected as required.

Imputation:

Values were imputed column-wise as follows:

- Blank login date values were imputed using the mode.
- Blank values for city, state, and Pin code were imputed according to the local address.
- Blank Tele caller name and BDM name values were imputed by considering the values in the cells above and below and imputing the appropriate one.
- Remaining blank values in the Business Category were filled according to the business names.
- The Tele caller name and BDM name were in one column, so they were extracted into two columns named Tele caller and BDM, and blank values were down filled in Excel.
- Two columns, Expense and Profit, were created from the Total Sales amount, which were 70 percent and 30 percent of the Total Sales, respectively.

Data Structuring:

- Checking the data from two tables revealed a data imbalance, so blank spaces were replaced with "No data" to balance the data.
- Blank Space from Contact Person is replaced by Unknown.

Challenges faced:

- For many cities, the state was written incorrectly, so it was corrected.
- In some cases, "No data" was written between names, e.g., "Ad. ManNo dataish Aggrawal," so it was corrected to "Ad. Manish Aggrawal."
- Some dates were found to be in the future, beyond 2024, while the meeting date was in the past, so they were corrected according to the meeting date.
- The meeting time was in the 12-hour format, which created issues, so it was converted to the 24-hour format.

Final Check:

- After balancing the data, a Business ID was created to connect both tables, making it common to both tables for creating a database. For this, a 4-alphabet temporary ID was created, which included the first letter of the Business name, the first letter of the Business Category, the first letter of the BDM name, and the first letter of the City name from the meeting table.
- We have to link all tables in MySQL for relational database thus we have to make primary key to connect all this Table in MySQL
- After creating the temporary ID, a unique ID was created from all the data by assigning a row number to that unique ID, and thus, using VLOOKUP, a permanent Unique ID was created as the Business ID.
- Again, using VLOOKUP, the Business ID from the meeting table was assigned to the Login Table.
- After analyzing the Business Category and Product Category columns, many categories were found to be the same and repeated. To normalize them into a single category for better analysis, many categories were reduced to a minimum number of categories, and the same was done for the Product Proposal column.

After discussing with the Manager, it was decided to create four tables from the two datasets. The four tables are described below

- Table 1: Business_ID (PK) Business_Name Contact_Person Address PinCode City State GST_Number
- Table 2: Business_ID (FK) Telecaller BDM Calling_Date Meeting_Date Meeting_Time Meeting_Status
- Table 3: Business_ID (FK) Business_Category Map Product_Proposal
- Table 4: Business_ID (FK) Login_Date Sales_Amount Advanced_Amount GST_Amount Payment_Mode

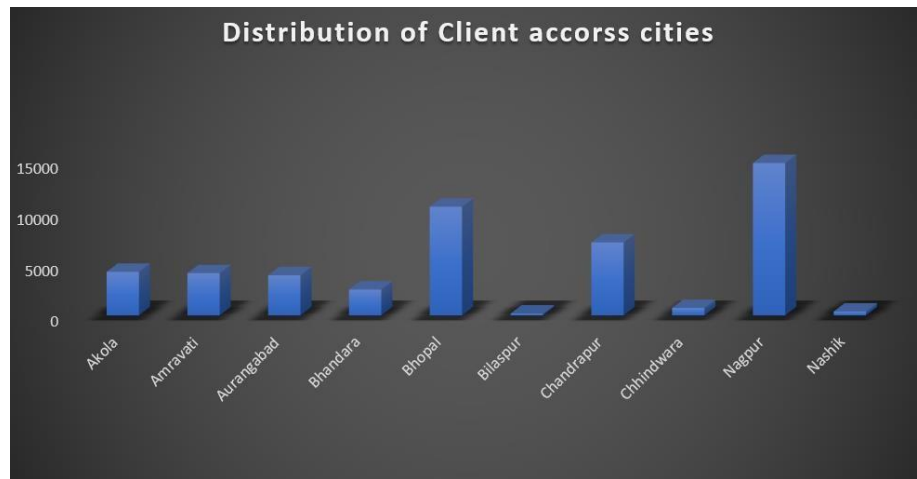
Thus as discuss by Manager we GST is 18 % of total sales amount so we have created and by formulas we imputed values.

These four tables were created in Excel using this format.

B. Analysis Using Power BI

All Four tables were imported into Excel and two column in table 4 is added after discussing with manager Expense and Profit which is 70 % of Total Sales and 30 % of Total Sales respectively and then merged with each other as we cannot establish connection between foreign keys in Excel.

1.What is the demographic profile of the clients and how does it vary across districts?



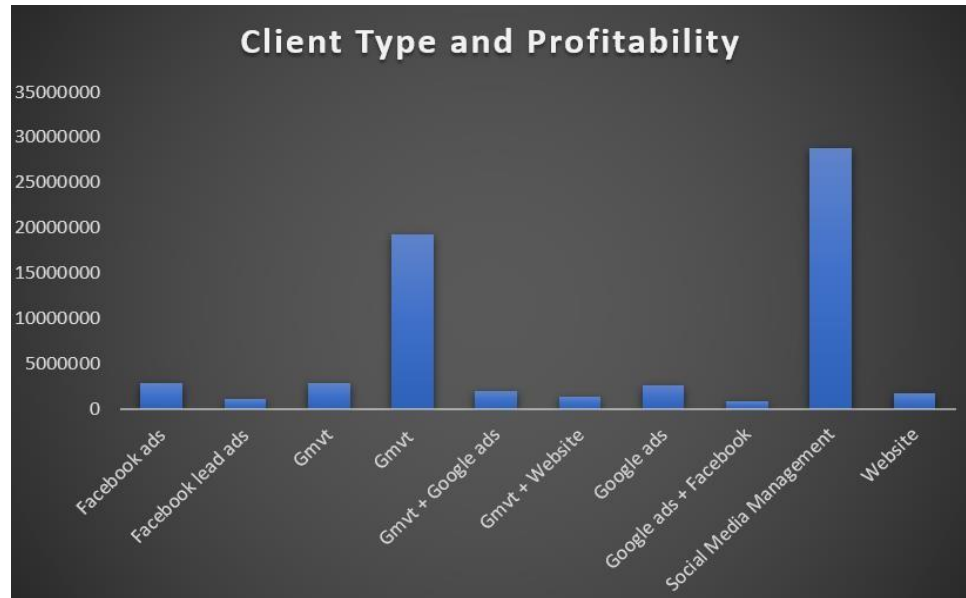
Clients are more prevalent in Nagpur, Bhopal, and Chandrapur.

2.How the Biz have performed over the years. Give their detailed analysis year & month-wise.



Sales peaked in 2019, followed by 2023.

3.What are the most common types of clients and how do they differ in terms of usage and profitability?



The most profitable products are GMVT, Social Media Management and Google Ads

4. Which types of product are most frequently used by the clients and what is the overall profitability of the client need?



The most used product is GMVT, Social Media Management

5. What are the major expenses of the Biz and how can they be reduced to improve profitability?



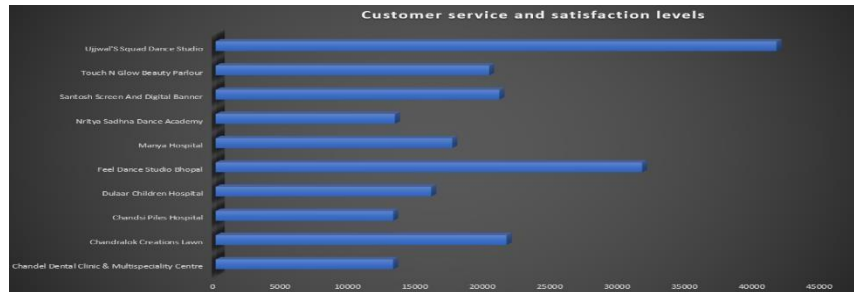
Major expenses are from Ayushman Hospital, Maitreya Developers, and Farme. We need to investigate why expenses are higher for these businesses.

6. What is the client portfolio and how does it vary across different purposes and client segments?



Clinics, hospitals, and educational classes are the most profitable categories.

7. How can the Biz improve its customer service and satisfaction levels?



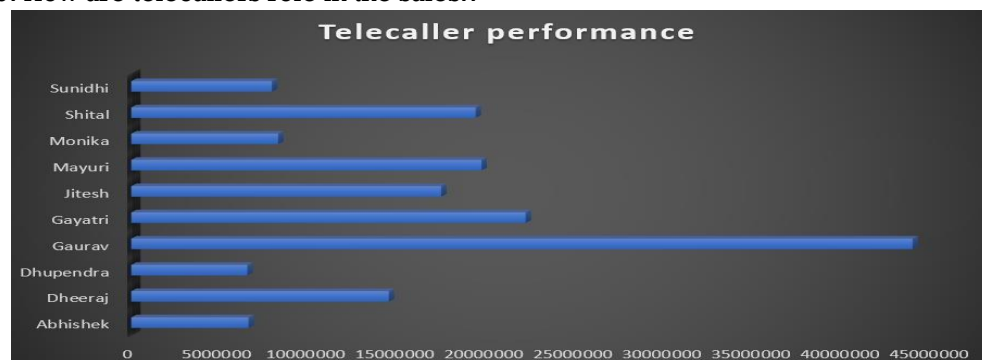
Bizz should focus on Ujjwal'S Squad Dance Studio, Feel Dance Studio Bhopal, Chandralok Creations Lawn on this business to improve customer service and satisfaction levels

8.Can the Biz introduce new products or services to attract more customers and increase profitability?



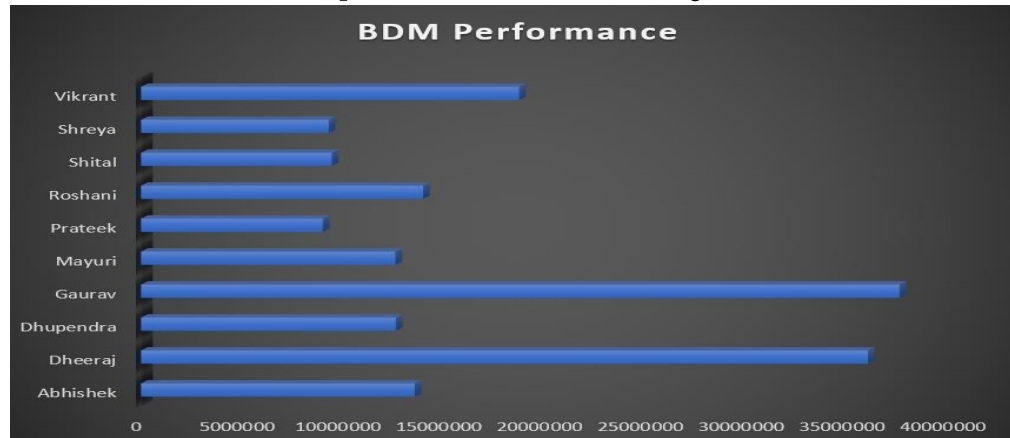
If Bizz wants to introduce new products, they should consider those related to hospitals, clinics, and educational classes.

9. How are telecallers role in the sales.?



Mayuri has scheduled the maximum number of meetings, followed by Shital and Gayatri.

10. What is BDM's individual performance with various segments of client.



Gaurav is the most effective Business Development Manager, followed by Mayuri and Roshni.

11. How many businesses retain with same or different product?



Maitreya Developers, Farme, and Pallavi Jewelers have different retention products, while Chandsi Piles Hospital, Khanch Ghar, and Metro Dental Care businesses have the same product retention.

12. Which is best selling products and category?



GMVT is the most sold product, and hospitals are the trending category.

13. What is popular selling amount?



10,000 is the most popular selling amount, followed by 11,800.