BIZ CONSULTING PROJECT REPORT

A. Data Wrangling

At first, meeting data was received from the S3 bucket in AWS , dumped by Project manager 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 using formulas after significant effort.

• 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 Pincode 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 Agrawal," so it was corrected to "Ad. Manish Agrawal."

• 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, 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.

• 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

These four tables were created in Excel using this format.

Thus Data is Ready to import in MYSQL

Analysis Using MYSQL

All Four tables were imported into MySQL.

Challenges faced:

• Importing tables into MySQL created a problem as Table 1 had duplicate Business_ID values Due to many instances of the same Business_Name. So, after assigning the primary key, it Gave a primary key-foreign key constraint failure error. To resolve this, duplicates were Removed from Table 1 based on the Business_Category, and then it was imported into MySQL.

• The remaining tables were successfully imported into MySQL.

• The date format of the Meeting Date and Calling Date columns in Table 1 was changed from VARCHAR to DATE, and the time format of the Meeting Time column was changed to TIME.

• The Login Date column was changed from VARCHAR to DATE format.

Thus After importing to MySQL we have run various query to Analyze data a follows

Q1.What is the demographic profile of the clients and how does it vary across districts ?

```
SELECT City, COUNT(*) AS Total_Clients, AVG(Sales_Amount) AS Average_Sales
FROM table_one b
JOIN table_Four t  ON b.Business_ID = t.Business_ID
GROUP BY City
ORDER BY Total_Clients DESC;
```

Result Grid | Filter Rows:

City	Total_Clients	Average_Sales
Nagpur	956	15125.683054
Bhopal	262	10883.206107
Chandrapur	182	13439.725275
Amravati	168	10719.500000
Akola	138	8778.094203
Bhandara	84	10689.166667
Aurangabad	84	13764.023810
Nashik	81	16862.592593
Yavatmal	49	16924.673469
Wardha	48	9766.250000
Chhindwara	18	10482.222222

Client is more from city Nagpur, Chandrapur from Maharashtra and Bhopal From Madhya Pradesh.

Q2.How the Biz have performed over the years. Give their detailed analysis year & month- wise ?

```
SELECT YEAR (t.Login_Date) AS year, MONTH(t.Login_Date) AS month,
SUM(t.Sales_Amount) AS total_sales
FROM table_Four t
GROUP BY YEAR(t.Login_Date), MONTH(t.Login_Date)
ORDER BY total_sales DESC;
```

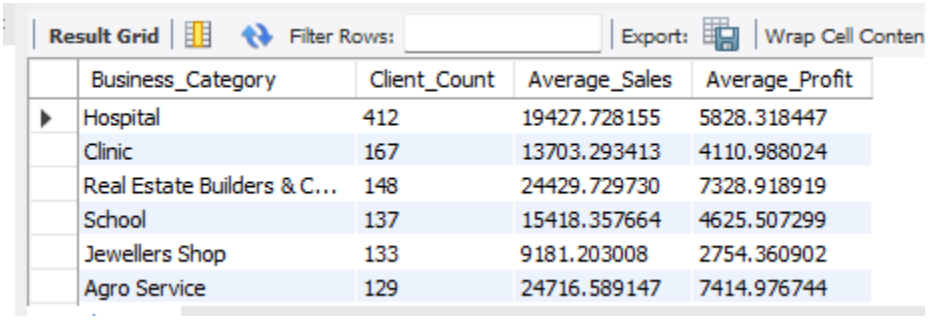
Result Grid | Filter Rows:

year	month	total_sales
2019	11	3874020.00
2019	12	3642823.00
2020	1	2059820.00
2020	2	1130044.00
2019	10	973696.00
2019	8	913796.00

Biz performed Highest in November 2019, December 2019 & January 2020

Q3.What are the most common types of clients and how do they differ in terms of usage and profitability?

```
SELECT Business_Category, COUNT(*) AS Client_Count, AVG(Sales_Amount) AS Average_Sales, AVG(profit) AS Average_Profit

FROM table_one b

JOIN table_three p ON b.Business_ID = p.Business_ID

JOIN table_Four  t ON b.Business_ID = t.Business_ID

GROUP BY Business_Category

ORDER BY  client_count DESC;
```



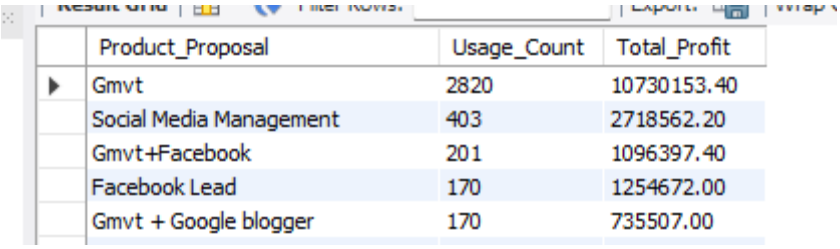
Result Grid | Filter Rows: | Export: | Wrap Cell Conten

Business_Category	Client_Count	Average_Sales	Average_Profit
Hospital	412	19427.728155	5828.318447
Clinic	167	13703.293413	4110.988024
Real Estate Builders & C...	148	24429.729730	7328.918919
School	137	15418.357664	4625.507299
Jewellers Shop	133	9181.203008	2754.360902
Agro Service	129	24716.589147	7414.976744

Most Common type of client is from Hospital, Real Estate and clinic which is gives most profit .

Q4.Which types of product are most frequently used by the clients and what is the overall profitability of the client need?

```
SELECT Product_Proposal, COUNT(*) AS Usage_Count, SUM(profit) AS Total_Profit

FROM table_three p

JOIN  table_four t ON p.Business_ID = t.Business_ID

GROUP BY Product_Proposal

ORDER BY Usage_Count DESC;
```



Product_Proposal	Usage_Count	Total_Profit
Gmvt	2820	10730153.40
Social Media Management	403	2718562.20
Gmvt+Facebook	201	1096397.40
Facebook Lead	170	1254672.00
Gmvt + Google blogger	170	735507.00

GMVT & social media management is most commonly used Product by client and it is most profitable product

Q5. What are the major expenses of the Biz and how can they be reduced to improve profitability?

```
SELECT Bi.Business_Name, SUM(td.expense) AS Total_Expenses, SUM(td.profit) AS Total_Profits
FROM table_one bi
JOIN table_four td ON bi.Business_ID = td.Business_ID
GROUP BY Bi.Business_Name
ORDER BY Total_Expenses DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell C

Business_Name	Total_Expenses	Total_Profits
Ayushman Hospital	388920.00	166680.00
Maitreya Developers	208250.00	89250.00
Farme	189728.00	81312.00
Nibe College Of Hotel Management	130403.00	55887.00
Gandhi Nursing Home	123900.00	53100.00

Ayushman Hospital ,Maitreya Developers ,Farme are the major expenses of the Biz.

Q6.What is the client portfolio and how does it vary across different purposes and client segments?

```
SELECT Business_Category, Product_Proposal, COUNT(*) AS Client_Count
FROM table_one b
JOIN table_three p ON b.Business_ID = p.Business_ID
GROUP BY Business_Category, Product_Proposal
ORDER BY Client_Count DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Cont

Business_Category	Product_Proposal	Client_Count
Hospital	Gmvt	517
Clinic	Gmvt	443
Clinic	No Data	372
Clinic	Gmvt+Facebook	350
Hospital	Gmvt+Facebook	250

Maximum client of hospital using GMVT.

Q7. How can the Biz improve its customer service and satisfaction levels?

```
SELECT Meeting_Status, COUNT(*) AS meeting_count
FROM table_two
GROUP BY Meeting_Status
ORDER BY meeting_count DESC;
```

Result Grid | Filter Rows:

Meeting_Status	meeting_count
Confirm	15305
No Data	2695
Call and Go	2051
Confirm	68
Call And Go	3

Biz should focous on confirm, call and go to improve its customer service and satisfaction levels.

Q8.Can the Biz introduce new products or services to attract more customers and increase profitability?

```
SELECT Business_Category, COUNT(DISTINCT Product_Proposal) AS Product_Types
FROM Table_Three
GROUP BY Business_Category
ORDER BY Product_Types desc;
```

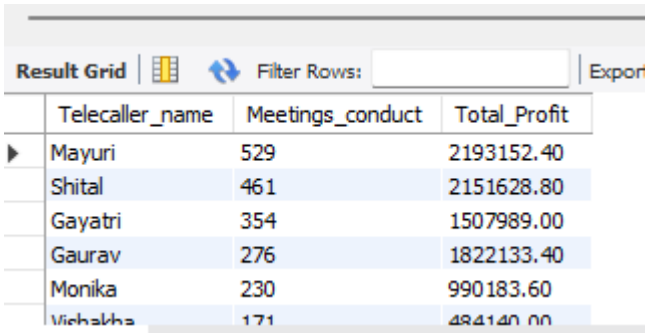
Result Grid | Filter Rows: | Export:

Business_Category	Product_Types
Clinic	18
Hospital	17
Coaching Classes	16
Real Estate Agency	13
Makeup Aademy	12

Biz should introduce new product or services related to Clinic, hospital, coaching classes to attract more customers and increase profitability.

Q9.How are telecallers role in the sales.?

```
SELECT Telecaller_name, COUNT(*) AS Meetings_conduct, SUM(profit) AS Total_Profit
FROM table_two m
JOIN  table_four t ON m.Business_ID = t.Business_ID
GROUP BY Telecaller_name
order by Meetings_conduct desc;
```



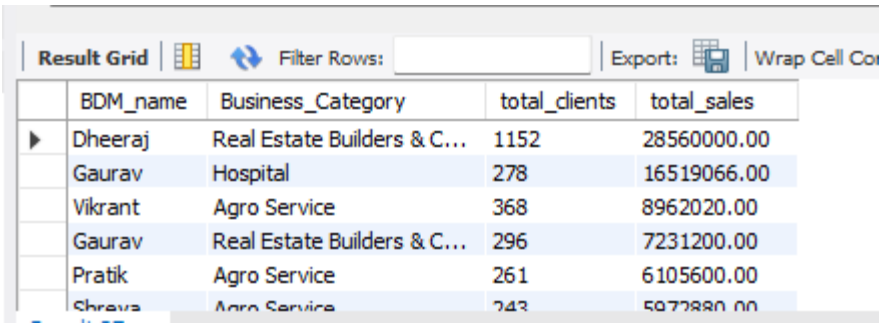
Result Grid | Filter Rows: | Export

Telecaller_name	Meetings_conduct	Total_Profit
Mayuri	529	2193152.40
Shital	461	2151628.80
Gayatri	354	1507989.00
Gaurav	276	1822133.40
Monika	230	990183.60
Vishakha	171	484140.00

Mayuri ,Shital ,Gayatri are the top 3 tellecaller who makes most meeting and profit

Q10.What is BDM's indivisual performance with various segments of client?

```
SELECT m.BDM_name, p.Business_Category, COUNT(m.Business_ID) AS total_clients,
SUM(t.Sales_Amount) AS total_sales
FROM table_two m
JOIN table_three p ON m.Business_ID = p.Business_ID
LEFT JOIN table_four t ON m.Business_ID = t.Business_ID
GROUP BY m.BDM_name, p.Business_Category
ORDER BY total_sales DESC;
```



Result Grid | Filter Rows: | Export: | Wrap Cell Co

BDM_name	Business_Category	total_clients	total_sales
Dheeraj	Real Estate Builders & C...	1152	28560000.00
Gaurav	Hospital	278	16519066.00
Vikrant	Agro Service	368	8962020.00
Gaurav	Real Estate Builders & C...	296	7231200.00
Pratik	Agro Service	261	6105600.00
Shreya	Agro Service	243	5972880.00

Dheeraj have Highest no of client in Real Estate Builders & Construction Category.

Q11.How many businesses retain with same or different product?

```
SELECT COUNT(b.Business_ID) AS total_businesses, p.Product_Proposal
FROM table_one b
JOIN table_three p ON b.Business_ID = p.Business_ID
GROUP BY p.Product_Proposal
order by total_businesses desc;
```

Result Grid | Filter Rows: | Expo

total_businesses	Product_Proposal
6415	Gmvt
4682	Gmvt+Facebook
4273	Gmvt+ Social Media Management
2477	No Data
983	Social Media Management
393	vps

Most of business retain with GMVT And GMVT + Facebook

Q12.Which is best selling prodcut and category?

```
SELECT p.Product_Proposal, p.Business_Category, SUM(t.Sales_Amount) AS total_sales
FROM table_three p
JOIN table_four  t ON p.Business_ID = t.Business_ID
GROUP BY p.Product_Proposal, p.Business_Category
ORDER BY total_sales DESC
LIMIT 1;
```

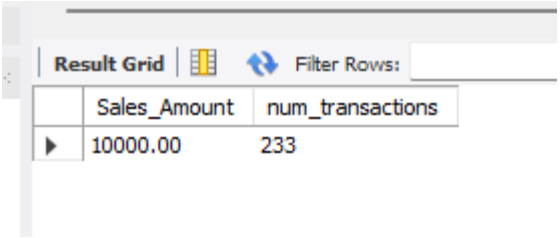
Result Grid | Filter Rows: | Export: | Wr

Product_Proposal	Business_Category	total_sales
Gmvt	Hospital	5137674.00

Gmvt is the Best selling product in Hospital

Q13.What is popular selling amount ?

```
SELECT t.Sales_Amount, COUNT(*) AS num_transactions
FROM table_four t
GROUP BY t.Sales_Amount
ORDER BY num_transactions DESC
LIMIT 1;
```

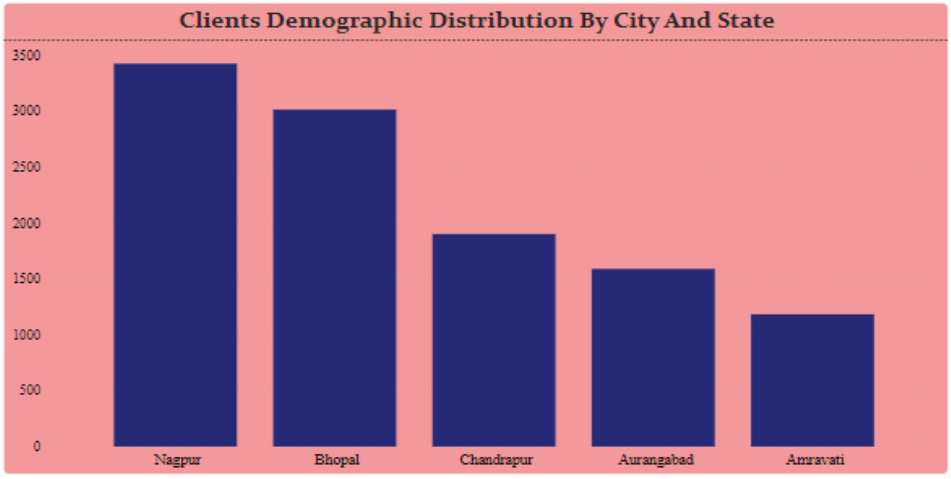


Maximum Amount of transcation is 10000 which is 233.

Power BI Dashboard

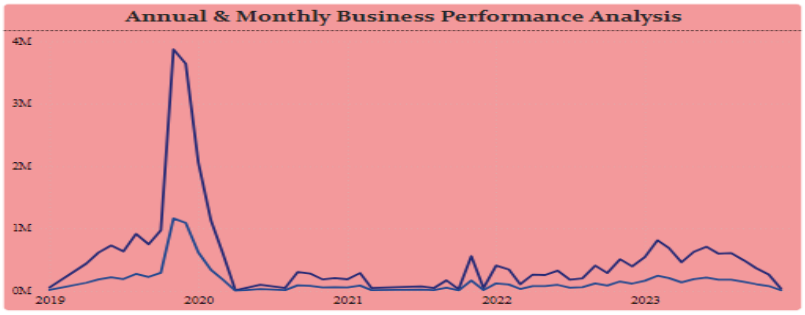
A Comprehensive report was created to facilitate in-depth analysis once the data was successfully loaded Into Power BI. Various measures were subsequently developed within Power BI to enhance the Effectiveness of the data analysis.

- **Demographic Profile of the Clients:**



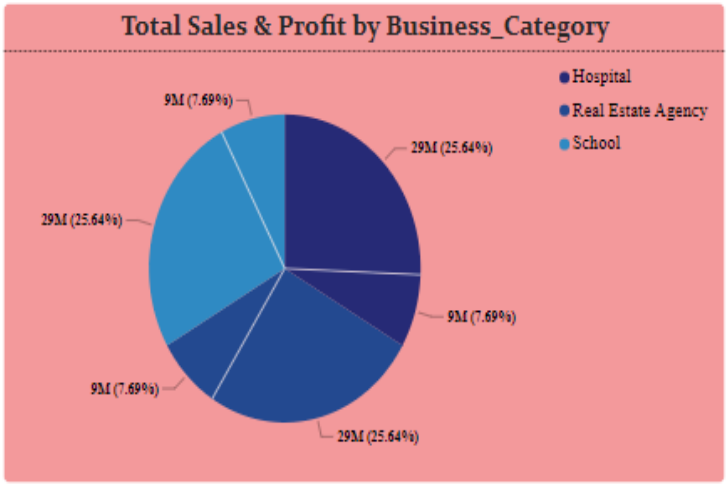
Analysis: Highest Number of Client are from city Nagpur, Chandrapur from Maharashtra and Bhopal From Madhya Pradesh.

- **Annual & Monthly Performance Analysis:**



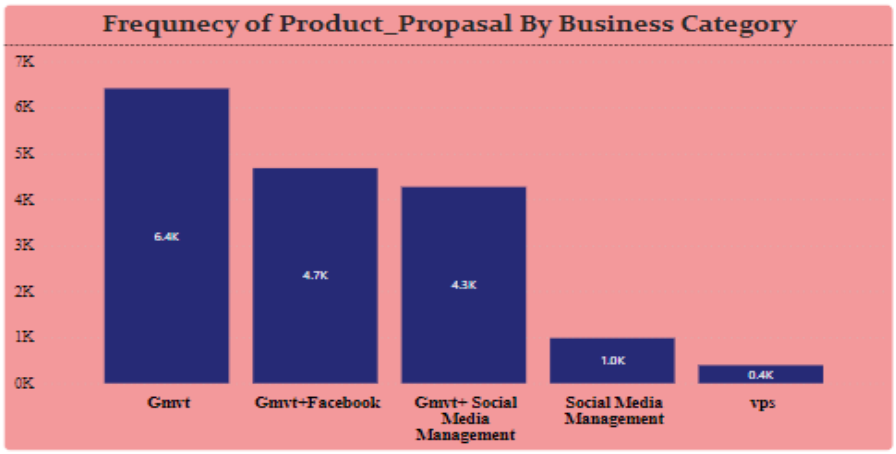
Analysis: Biz performed Highest in November 2019, December 2019 & January 2020

- **Business Analysis:**



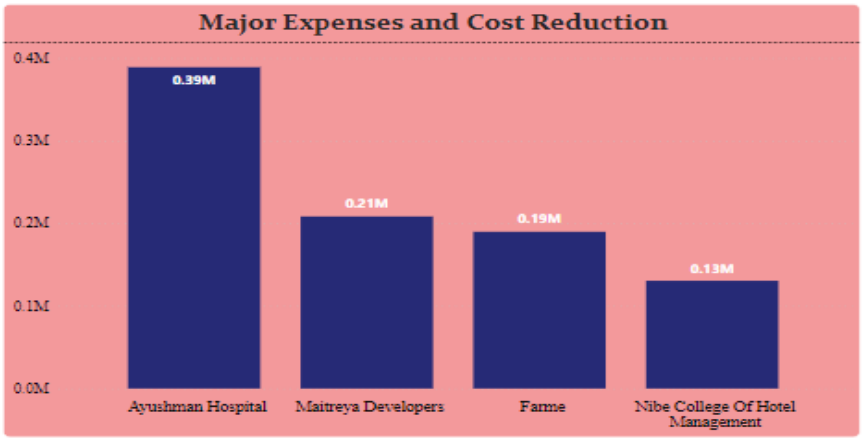
Analysis: Most Common type of client is from Hospital, Real Estate and School which gives most profit .

- **Product Analysis:**



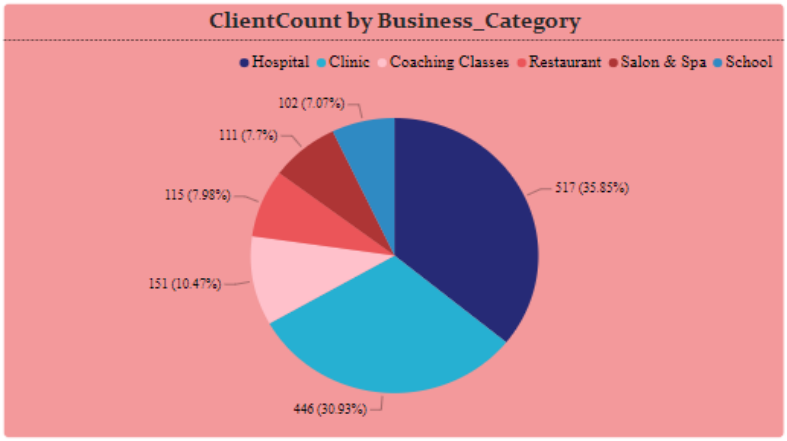
Analysis: GMVT is most commonly used Product by client and it is most profitable product

- **Company Major Expenses:**



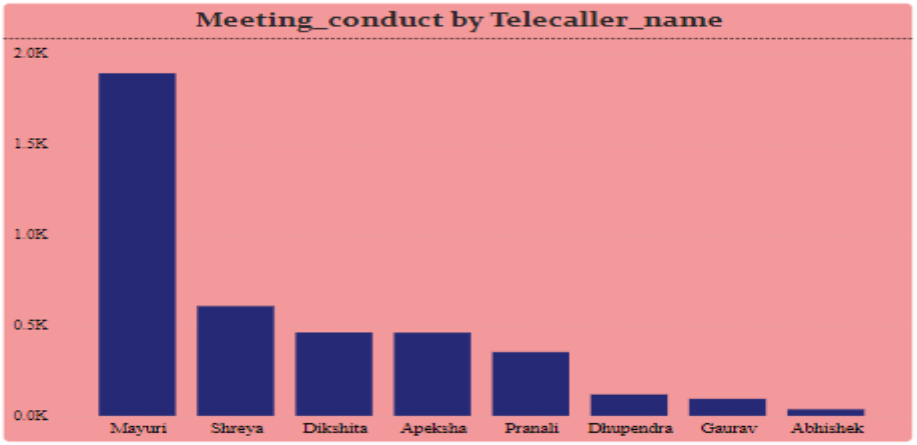
Analysis: Ayushman Hospital ,Maitreya Developers , Farme are the major expenses of the Biz. so need to focus on this Business to find why expense is higher in this Businesses.

- **Client Count By Business Category :**



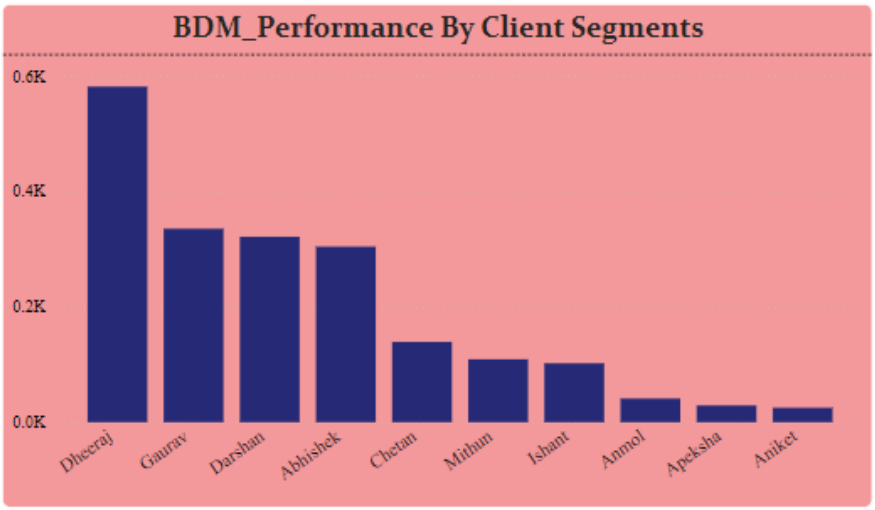
Analysis: Thus we get to Know that Hospital (517), Clinic(446) and Coaching Classes(151) Are the top 3 Category have most Client .

- **Telecaller Performance:**



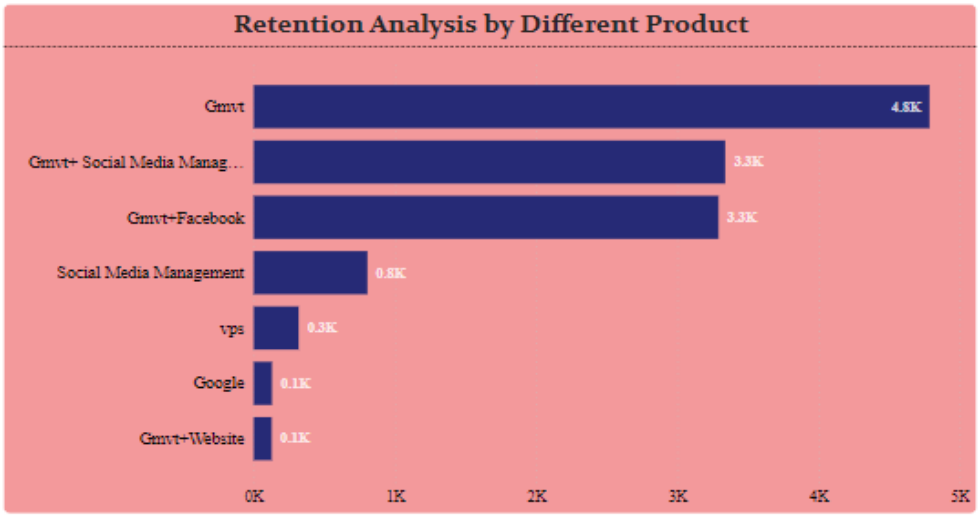
Analysis: Mayuri ,Shreya ,Dikshita are the top 3 tellecaller who makes most meeting and profit

- **BDM Performance:**



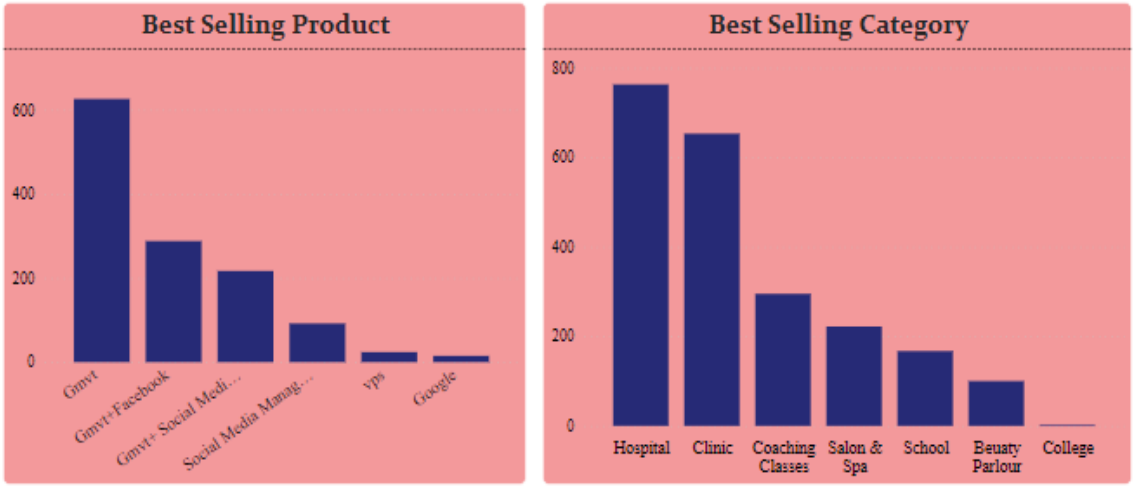
Analysis: Dheeraj have Highest no of client Followed By Gaurav & Darshan.

- **Retention Analysis:**



Analysis: Most of business retain with Gmvt Followed By Gmvt+Social Media Management & Gmvt+Facebook.

- **Best selling product and category:**



Analysis: Gmvt is the Best selling product And Hospital is the Best Selling Category.