Biz Consulting Company Sales Performance

A. Data Wrangling

Initially, meeting data was received from the S3 bucket in AWS, provided by the client and informed by our Manager Amit Makode. The data was extracted to an Excel sheet for analysis.

The data from the Excel sheet was analyzed, and the pin code, city, and state were separated from the address column into new columns. 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 anomalies, and the meeting date and calling date were arranged in sequence when found differently.

Challenges Faced:

- Extracting the pin code from the address was difficult as its location varied, but it was eventually separated out using Python formulas.

```
import re
def clean_address(Contact_address)
if not isinstance(Contact_address, str):
        # If the input is not a string, return it as is
        return Contact_address
   # Remove pincode
    address = re.sub(r'\d{6}$', '', Contact_address).strip()
# Remove city and state
    address = re.sub(r', [^,]+, [^,]+$', '', Contact_address).strip()
 return address
df1['Contact_address1'] = df1['Contact_address'].astype(str).apply(clean_address)
```

- The dates were not in the proper format, and there was a considerable difference between the meeting date and the calling date. Since the calling date should be earlier than the meeting date, it was corrected.
- The main challenge was structuring the data into the proper format, which required significant 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 between "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 pin code 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 extracted from the S3 bucket in AWS 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 significant 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 filled down 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 spaces from Contact Person were 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 had to link all tables in MySQL for the relational database, so we had to make a primary key to connect all these tables 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 discussion 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, Expense and Profit

As discussed with the Manager, the GST is 18% of the total sales amount in table 4 , so we created and imputed values using formulas. These four tables were created in Excel using this format. Thus, the data is ready to import into MySQL.

B. 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.
- Again after discussing with Manager and Client we have to create Expense and Profit column which is 70% and 30% of Total Sales Respectively so again using formulas and created Two Columns as required

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

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

```
SELECT b.City, b.State, COUNT(DISTINCT b.Business_ID) AS num_clients
FROM business_info b
GROUP BY b.City, b.State
ORDER BY num_clients DESC;
```

City	State	num_clients
Nagpur	Maharashtra	3430
Bhopal	Madhya Pradesh	3020
Chandrapur	Maharashtra	1913
Aurangabad	Maharashtra	1603
Amravati	Maharashtra	1201
Bhandara	Maharashtra	926
Akola	Maharashtra	924
Chhindwara	Madhya Pradesh	326
Bilaspur	Chhattisgarh	158
Nashik	Maharashtra	60
Wardha	Maharashtra	47
Yavatmal	Maharashtra	41
Ujjain	Madhya Pradesh	33
Nanded	Maharashtra	28
Betul	Madhya Pradesh	13

thus we get to know that Client is more from Nagpur ,Bhopal from Madhya Pradesh and Chandrapur from Maharashtr

2.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 transaction_data t
GROUP BY YEAR(t.Login_Date), MONTH(t.Login_Date)
ORDER BY Total_sales;
```

year	month	total_sales
2019	11	3874020
2019	12	3633383
2020	1	2059820
2020	2	1130044
2019	10	973696
2019	8	918796
2023	2	812107
2019	9	748760
2023	6	730741
2019	6	728310
2023	3	688987
2019	7	634380
2023	8	633627

Biz performed in most in November 2019, December 2019 followed by January 2020

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

```
SELECT Category_type, COUNT(*) AS Client_Count, AVG(Sales_Amount) AS
Average_Sales, AVG(profit) AS Average_Profit
FROM business_info b
JOIN product_info p ON b.Business_ID = p.Business_ID
JOIN transcation_data t ON b.Business_ID = t.Business_ID
GROUP BY Category_type
ORDER BY client_count DESC;
```

Category_type	Client_Count	Average_Sales	Average_Profit
Hospital	628	18145.6210	5443.6783
Clinic	460	12471.4217	3741.4261
Real Estate	329	19243.3131	5772.9878
Educational Classes	315	13792.2063	4137.6635
Banquet Hall	193	15968.5130	4790.5648
School	145	15152.2414	4545.7034
Jewellery Shop	133	9181.2030	2754.3609
Agro Product	129	24716.5891	7414.9767
College	117	22482.0427	6744.6154
Enterprises	90	15307.0889	4592.2000
Restaurant	90	15570.6667	4673.0000

Most Common type of client is from Hospital, Clinic and Educational classes and Client which gives most profit Is from Agency and Tourism.

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

```
SELECT New_Product, COUNT(*) AS Usage_Count, SUM(profit) AS Total_Profit
FROM product_info p
JOIN transaction_data t ON p.Business_ID = t.Business_ID
GROUP BY New_Product
ORDER BY Usage_Count DESC;
```

New_Product	num_clients	total_sales
Gmvt	4289	26059590
Gmvt + Facebook Lead	2713	813520
Gmvt + Social Media	2161	199100
No Data	2017	803565
Gmvt+Social Media	1096	85200
Social Media Management	813	11203885
Gmvt	310	4745742
Vps	301	26260
Gmvt + Social Media + Facebook Lead	232	34160
Gmvt + Google ads	159	2472700
Google ads	140	2227505
Gmvt + Website	91	1873520
Gmvt + Social Media + Website	74	NULL

Gmvt product is most used by client having highest usage and most profitable product followed by Gmvt+Facebook lead

5. 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
business_info bi
JOIN
transaction_data td ON bi.Business_ID = td.Business_ID
GROUP BY
bi.Business_Name
ORDER BY
Total_Expenses DESC;
```

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
Indian Institute Of Fire Engineering	115080.00	49320.00
Central India Public School	106361.50	45583.50
Sky Agency	99890.00	42810.00
Mahalaxmi Devlopers	86800.00	37200.00
Lotus Garden And Celebration	86730.00	37170.00

Major Expense is from Ayushman Hospital, Maitreya Developers ,Farme , we should focus on Meeting Confirmation of Customer and should reduced cost by searching alternative in market.

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

```
SELECT Category_type, New_Product, COUNT(*) AS Client_Count
FROM business_info b
JOIN product_info p ON b.Business_ID = p.Business_ID
GROUP BY Category_type, New_Product
Order by Client_Count Desc;
```

Category_type	New_Product	Client_Count
Clinic	Gmvt	1109
Hospital	Gmvt	957
Clinic	Gmvt + Facebook Lead	679
Clinic	No Data	536
Hospital	Gmvt + Facebook Lead	361
Hospital	No Data	309
Educational Classes	Gmvt + Facebook lead	282
Clinic	Gmvt + Social Media	275
Educational Classes	Gmvt	271
Hospital	Social Media Management	221
Hospital	Gmvt + Social Media	189
Educational Classes	Gmvt + Social Media	160
Clinic	Social Media Management	146

Maximum Client is of Clinic using GMVt product followed by Hospital then Education Classes

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

```
SELECT
  b.Business_Name,
 m.BDM_name,
  m.Telecaller_name,
  m.Meeting_Status,
 DATEDIFF(m.meeting_date, m.calling_date) AS days_to_meeting,
 t.Sales_Amount,
 t.profit
FROM
business_info b
JOIN
meeting_details m ON b.Business_ID = m.Business_ID
LEFT JOIN
transcation_data t ON b.Business_ID = t.Business_ID
ORDER BY
days_to_meeting DESC;
```

Business_Name	Telecaller_name	BDM_name	Meeting_Status	New_Product	Days_Taken_For_Meeting	Sales_Amou
Nritya Sadhna Dance Acade...	Nisha	Peter	No Data	Gmvt + Facebook Lead	2055	NULL
Tulip Family Salon & Spa	Shruti	Peter	No Data	Gmvt + Facebook Lead	2053	NULL
Chandralok Creations Lawn	Nisha	Vishal	Call and Go	Gmvt + Facebook Lead	1972	NULL
Chandralok Creations Lawn	Nisha	Vishal	Call and Go	Gmvt+Social Media	1972	NULL
Mahabali's Fitness First	Nisha	Raj	Call and Go	Gmvt + Facebook Lead	1972	NULL
Chandsi Piles Hospital	Apeksha	Suraj	Confirm	Gmvt + Facebook lead + Go...	1764	NULL
Chandsi Piles Hospital	Apeksha	Suraj	Confirm	Gmvt + Facebook lead +Go...	1764	NULL
Chandsi Piles Hospital	Apeksha	Suraj	Confirm	Social Media Management	1764	NULL
Chandsi Piles Hospital	Apeksha	Suraj	Confirm	Gmvt + Facebook ads	1764	NULL
Maharashtra Engineering W...	Apeksha	Vikrant	Confirm	Gmvt + Facebook lead +Go...	1764	NULL
Artecstasy	Shruti	Peter	Call and Go	Gmvt + Facebook Lead	1732	NULL
Baba Mathar Dev Health Clinic	Nisha	Peter	No Data	Gmvt + Facebook Lead	1724	NULL
Chandel Dental Clinic & Multi...	Nisha	Peter	No Data	Gmvt + Facebook Lead	1724	NULL

Biz should concentrate on above mentioned business for good customer services.

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

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

Category_type	Product_Types
Clinic	53
Hospital	49
Educational Classes	48
Real Estate	43
Beauty Parlour	29
Restaurant	28
School	27
Tour And Travel	26
Interior Decor	25
Cloth Store	24
Salon	24
Hardware And Ply Store	23
Co-Curricular Classes	22
College	22
Jewellery Shop	21

Biz should introduce new product related to Clinic , Hospital, Educational Classes for more profit

9.How are telecallers role in the sales.

```
SELECT
    md.Telecaller_name,
    pi.Category_type,
    COUNT(*) AS Sales_Count,
    SUM(td.Sales_Amount) AS Total_Amount
FROM
    meeting_details md
JOIN
    product_info pi ON md.Business_ID = pi.Business_ID
LEFT JOIN
    transcation_data td ON md.Business_ID = td.Business_ID
WHERE
    md.Meeting_Status = 'Confirm'
GROUP BY
    md.Telecaller_name,
    pi.Category_type
order by Total_Amount desc;
```

Telecaller_name	Category_type	Sales_Count	Total_Amount
Gaurav	Real Estate	1171	28767400
Jitesh	Agro Product	484	11925760
Gaurav	Hospital	441	11735080
Dheeraj	Real Estate	288	7140000
Dheeraj	Hospital	71	6227200
Sunidhi	Agro Product	244	5987580
Gayatri	Agro Product	257	5975880
Mayuri	Banquet Hall	290	4409096
Gayatri	Real Estate	346	3801000
Priti	Real Estate	148	3570000
Dhupendra	Hospital	74	3480040
Varsha	Agro Product	124	3104160
Diksita	Agro Product	122	3018020
Shreya	Agro Product	121	2981440
Shital	Hospital	455	2976616

Gavrav makes most of profit followed by Jitesh and Dheeraj

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

```
SELECT m.BDM_name, p.Category_type, COUNT(m.Business_ID) AS num_clients,
SUM(t.Sales_Amount) AS total_sales
FROM meeting_details m
JOIN product_info p ON m.Business_ID = p.Business_ID
LEFT JOIN transaction_data t ON m.Business_ID = t.Business_ID
GROUP BY m.BDM_name, p.Category_type
ORDER BY total_sales DESC;
```

BDM_name	Category_type	num_clients	total_sales
Dheeraj	Real Estate	1186	28697440
Gaurav	Hospital	405	18883506
Vikrant	Agro Product	370	8962020
Gaurav	Real Estate	302	7283400
Prateek	Agro Product	261	6105600
Shreya	Agro Product	243	5972880
Dheeraj	Hospital	583	5183080
Vikrant	Real Estate	560	4070220
Praful	Real Estate	154	3698100
Gaurav	College	120	3586280
Jeshana	Real Estate	159	3254700

Dheeraj have most number of client and makes most sales in Real Estate same as Gavrav in Hospital and Vikrant in AgroProduct

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

```
SELECT COUNT(b.Business_ID) AS num_businesses, p.New_Product
FROM business_info b
JOIN product_info p ON b.Business_ID = p.Business_ID
GROUP BY p.New_product;
```

num_businesses	New_Product
5779	Gmvt
3816	Gmvt + Facebook Lead
2682	Gmvt + Social Media
2496	No Data
1442	Gmvt+Social Media
1011	Social Media Management
371	Vps
362	Gmvt
351	Gmvt + Social Media + Facebook Lead
202	Gmvt + Google ads
156	Google ads

Most of business is retained by GMVT product followed by Gmvt+Facebook lead

12.Which is best selling product and category.

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

New_Product	Business_Category	total_sales
Social Media Management	Real Estate Builders & Construction	3570000
Social Media Management	Agriculture	2710400
Gmvt	Hospital	2309920
		230

Social Media Management is best selling product followed by GMVT

13.What is popular selling amount

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

Sales_Amount	num_transactions
10000	233

Maximum Amount of transaction is of 10000 which is of 233.

C. Analysis Using POWER BI

Thus we connect MySQL to Power Bi for importing Data of four tables

Challenges Faced :

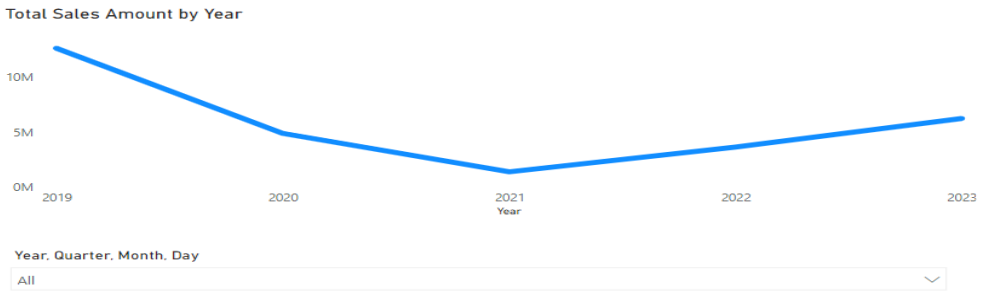
1. Connector problem (error shown requires one or more connector to establish connection thus after installing required connector problem resolved.
2. Relationship establishing Table is not mutually connected with each other , all 3 table is only connected by Table 1 using Business ID so mutually establishing relationship between Table2,Table3 and Table4

   Thus after this solving required Problem Statement.

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

Thus, here we get to Know Most of Clients is from Nagpur, Chandrapur and Bhopal.

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

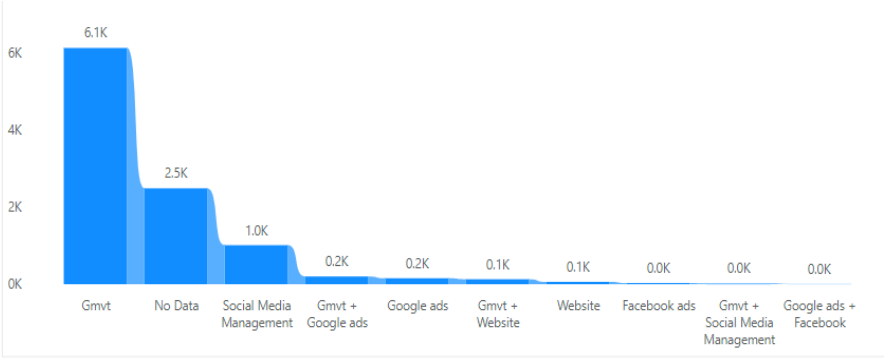


Thus Bizz have most sales in 2019 followed by 2023.

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

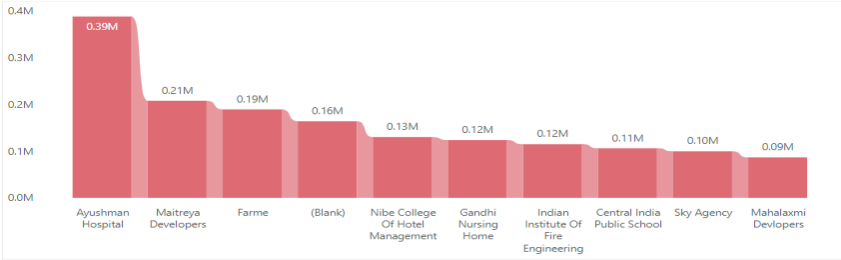
Most of the common type of Client are from Hospital ,Clinic and Educational Classes and Most profitable product is GMVT , Social Media Management and Google Ads .

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



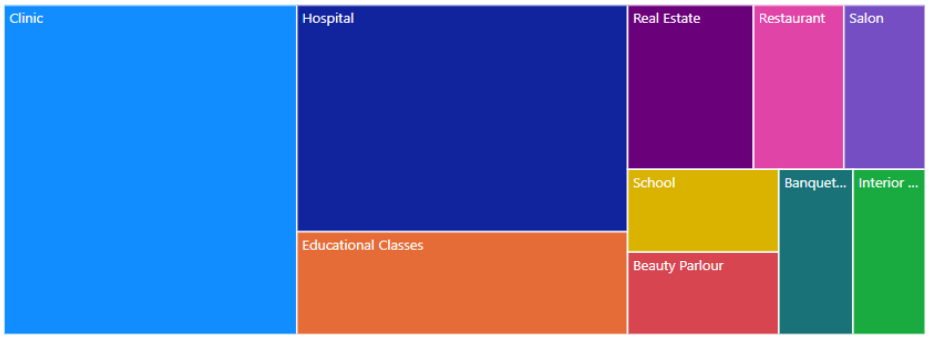
Most frequently used product are Gmvt, Social Media Management and Gmvt+Google Ads.

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



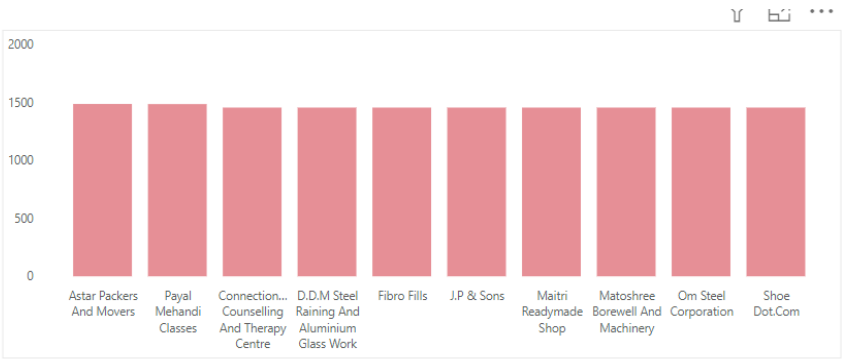
Major Expense is from Ayushman Hospital , Maitreya Developers ,Farme so we need to focus on this Business to find why expense is higher in this Businesses.

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



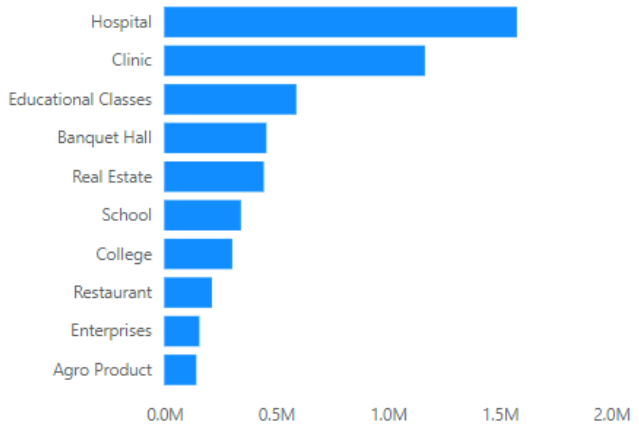
Thus we get to Know that Clinic , Hospital and Educational Classes have most profit in category.

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



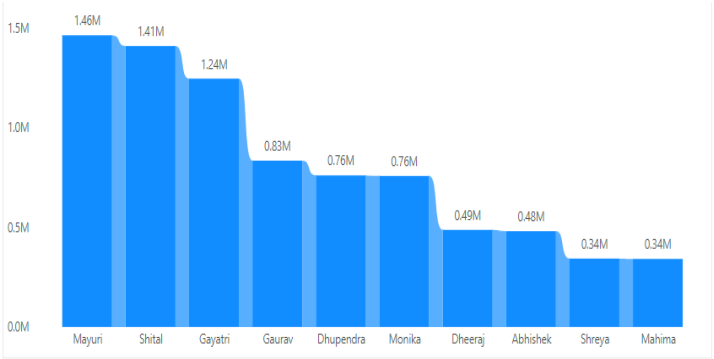
The significant disparity in call and meeting volumes between Astar Packer and Mover, Payal Mehandi Classes, and Connection - Counseling and Therapy Center suggests that the bizz should focus on customer service and satisfaction for these ventures.

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



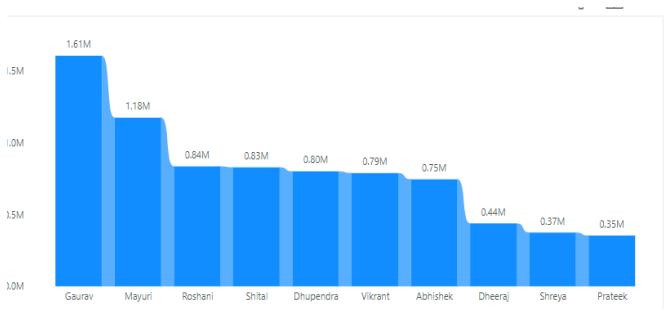
Thus Bizz should introduce product related to Hospital ,Clinic and educational Classes.

9.How are telecallers role in the sales.



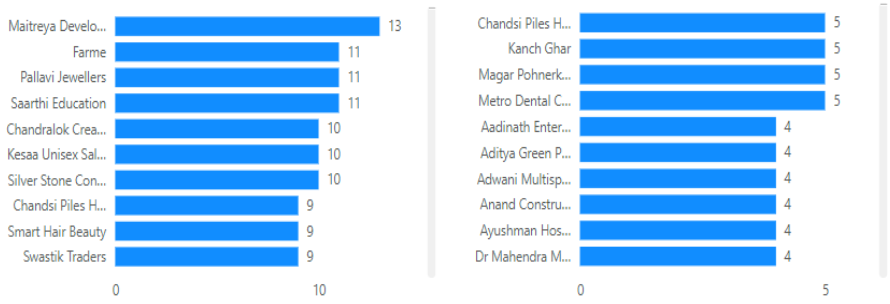
Mayuri has scheduled the maximum number of meetings followed by Shital and Gayatri

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



Gavrav is most effective Business Development Manager followed by Mayuri and Roshni .

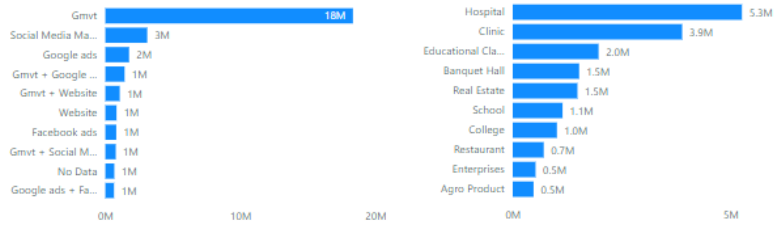
11. How many businesses retain with same or different product.



Maitreya Developer , Farme, Pallavi Jewelers have different retention product and Chandsi Piles Hospital , Khanch ghar, Metro Dental care

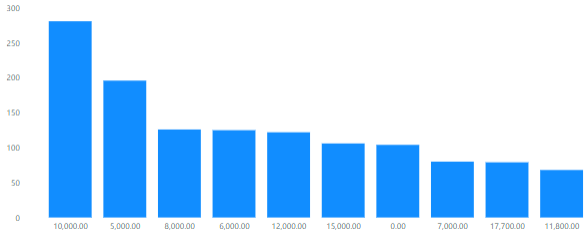
businesses are of same product retention.

12.Which is best selling products and category ?



GMVT is most Selling Product and in Category Hospital is Trending one.

13.What is popular selling amount?



10000 is most popular selling amount.