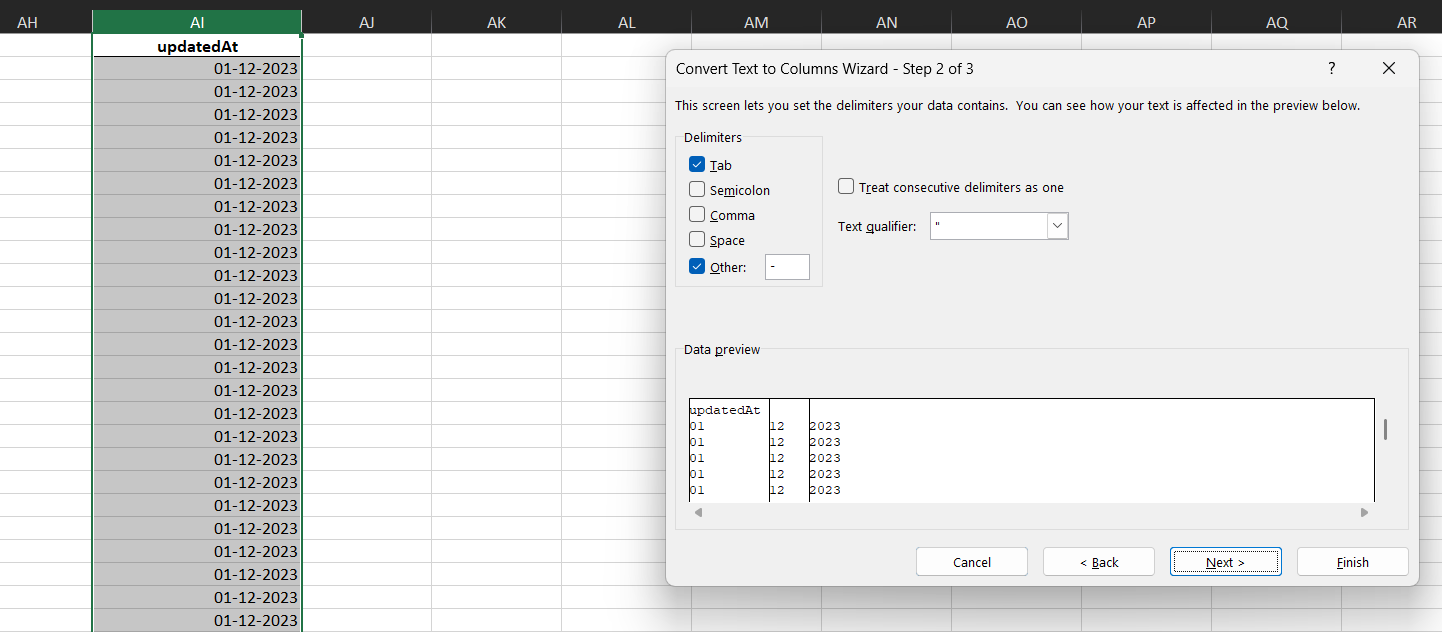
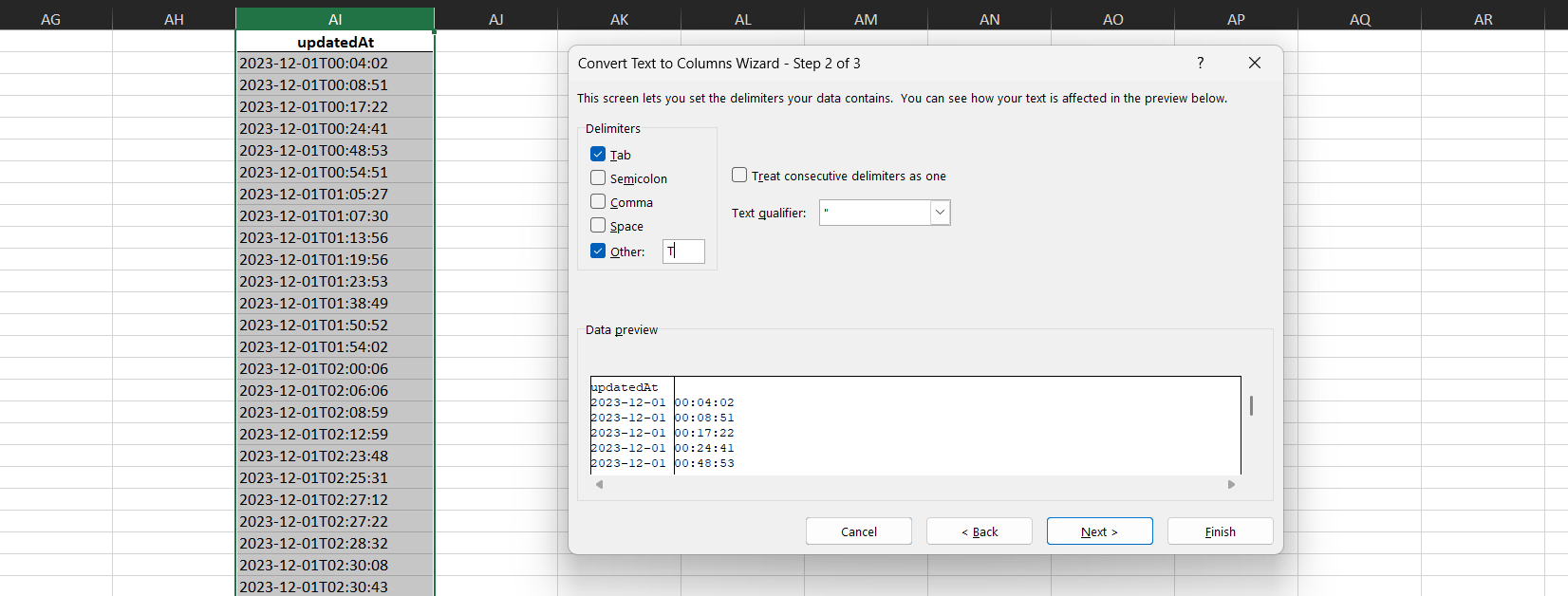
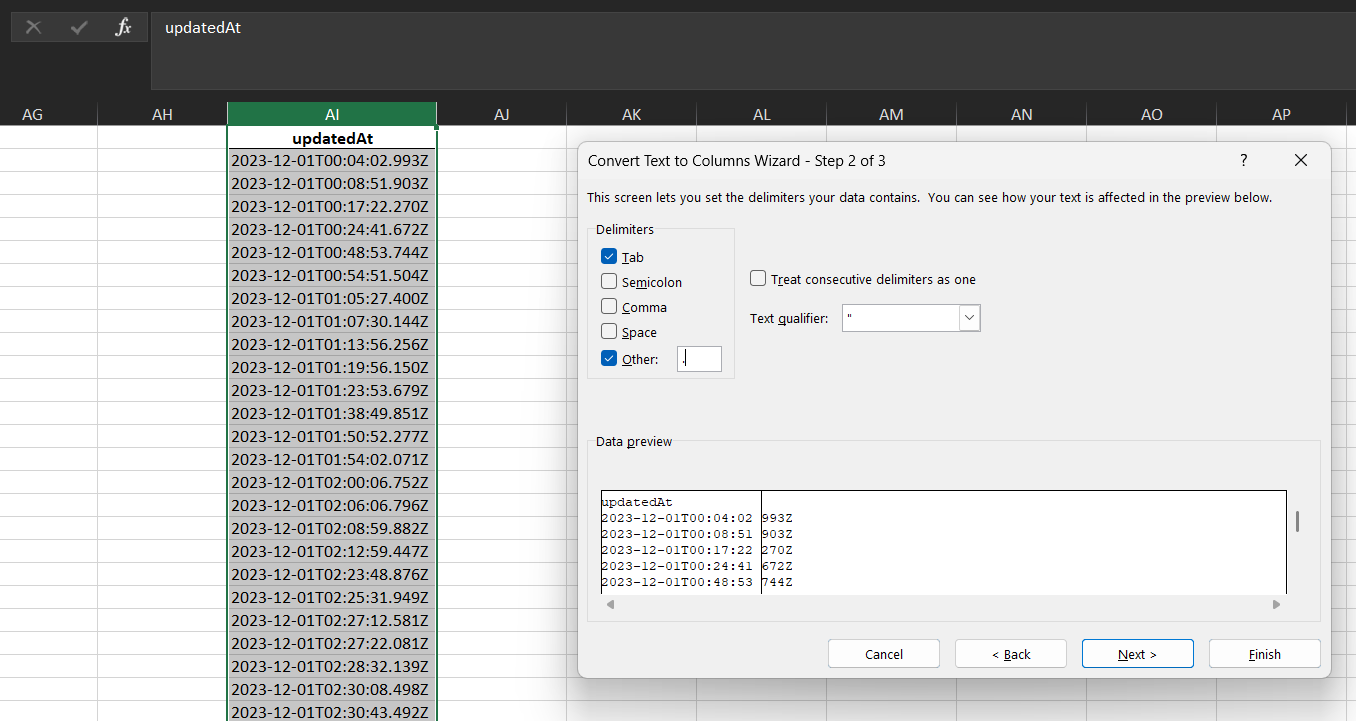
# **Tasks**

**Learners have to develop a dashboard to support the answers to the following questions.**

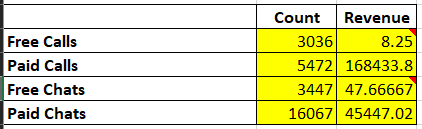
**Objective Questions**:

1. **What is the total no. of tables present in the data?***Single Table*
2. **What is the total no. of attributes present in the data?**  
   *Total 36 Attributes*
3. **The data consists of some inconsistent and missing values, so ensure that the data used for further analysis is cleaned.**



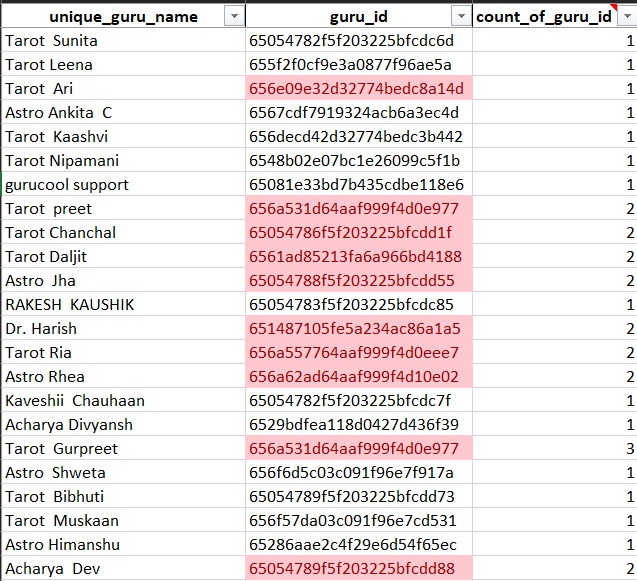
**---- DATA CLEANING ----**

* Rename "public\_live\_data" to "Public Live Data" to maintain consistency
* Removing "isWhiteListUser" column because all values are FALSE thus reducing attributes
* Removing "queue" column because all values are FALSE thus reducing attributes
* Removing "\_\_v" column because all values are 0 thus reducing attributes
* Removing "callChannel" column because all values are Normal thus reducing attributes
* Removing "callvrType" column because all values are Exotel thus reducing attributes
* There are 145 calls that didn't register any CallSid and were failed transactions so we can delete those data points
* There are some extra spaces in between text in guruName so we can use formula =IFERROR(REPLACE($D2, FIND(" ", $D2), 2, " "), $D2) to remove those spaces OR =TRIM($D2) then copy only values into the guruName column



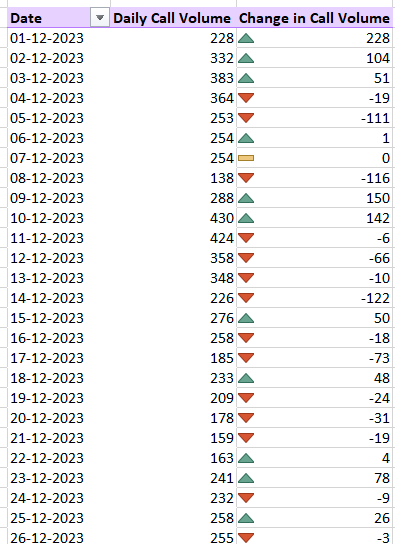
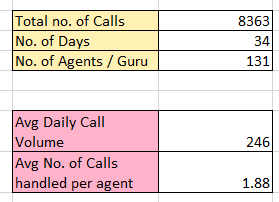
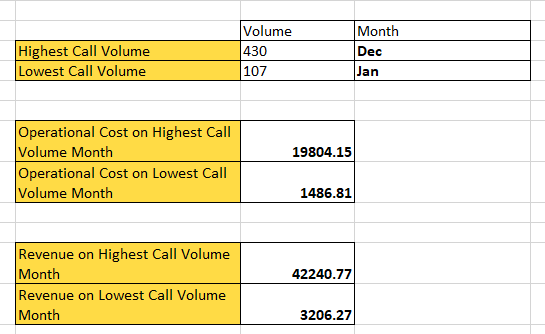
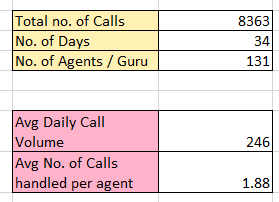
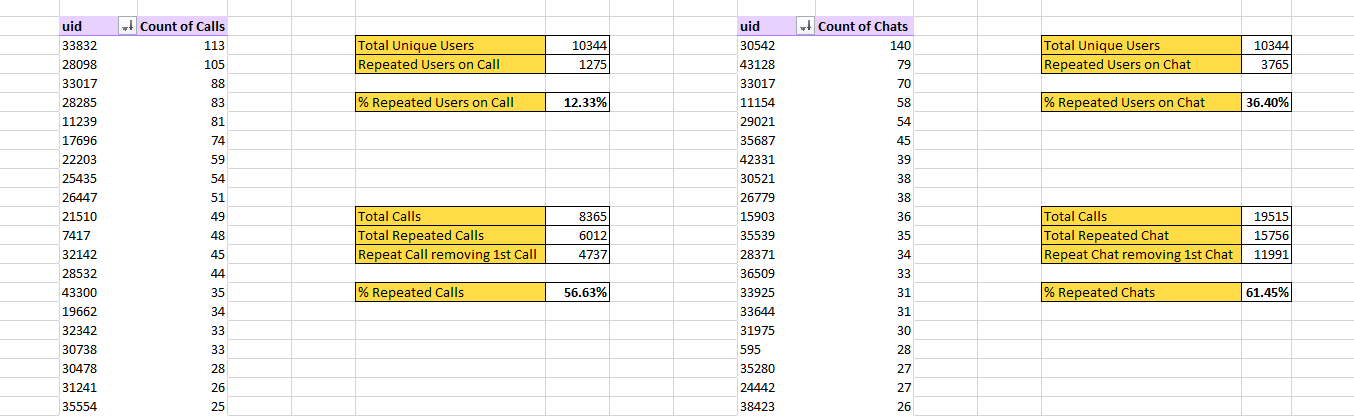
**---- DATA MANIPULATION ----**

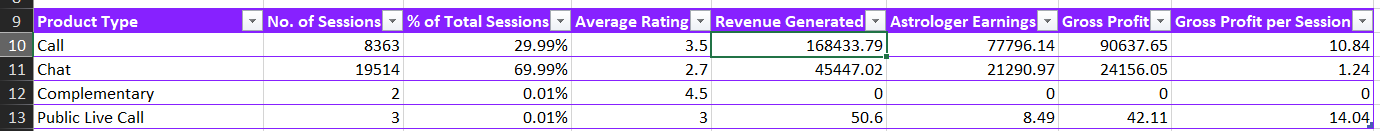
* There are 149 guru\_name and less guru\_id what should be done to eliminate the wrong or double mapped data
* Datetime data in "updatedAt", "chatStartTime" and "chatEndTime" is not in correct format. Following steps are taken to make them correct
  + using text to column based on delimiter "." to move extra 4 characters in next column and then remove them
  + using text to column based on delimiter "T" to separate date and time value
* To make "updateAt" column of right date format we first removed time part from text-to-column, then again split them using delimiter "-" to get integer values and then at last using DATE() function to built the correct date
* To make "chatStartTime" and "chatEndTime" column of right date format we first remove date part from tex-to-column, then again split them using delimiter ":" to get integer value and the at last using TIME() function to built the correct time
* There exist one record 65799142858de4867bd59634 that has free call but has charged customer so we have made amount, astrologersEarnings and netAmount 0 to maintain data consistency
* There exist two records 656da36e2d32774beda9c2ad, 656ca85d2d32774bed835e53 that has free chat but has charged customer so we have made amount, astrologersEarnings and netAmount 0 to maintain data consistency

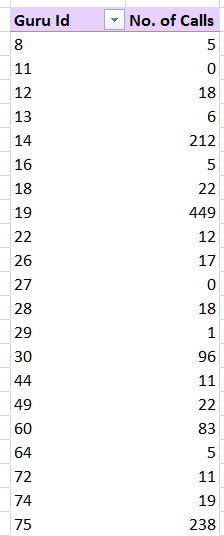
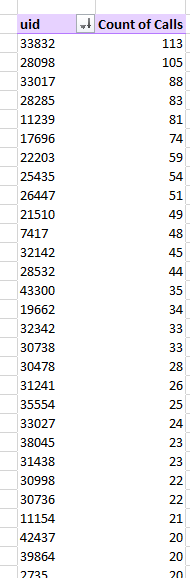
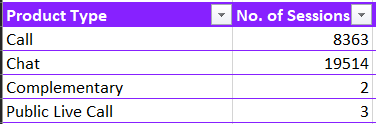


**---- DATA DUPLICACY ----**

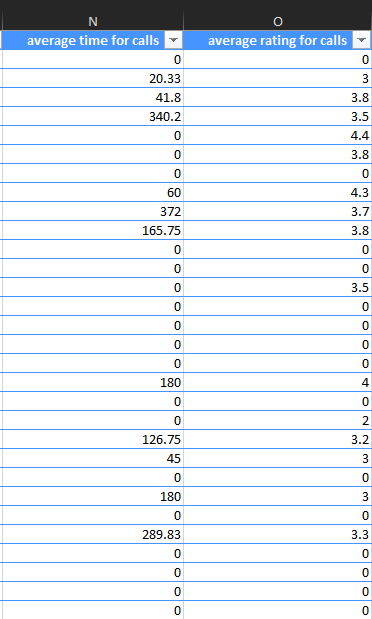
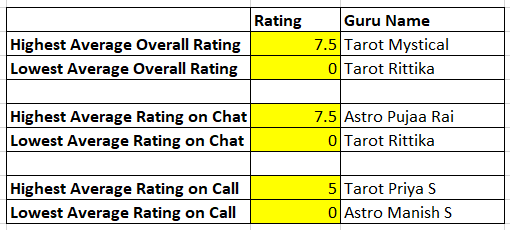
* There are 149 guru names and 131 guru id that is a signal that data may contain duplicates or false values
* Duplicates can be found by first talking unique guru names and mapping them with guru id using VLOOKUP function.
* Then using formula =COUNTIF($B$2:$B2, $B2) to identify duplicate values where first value will be marked first and next duplicates as 2, 3, 4....
* 65054782f5f203225bfcdc70 guru id is mapped to three different values but guru name is same i.e. "Astro Dr Balkrihna" which also exists as "Dr. Balkrisna" and "Dr Balkrisna" so we changed all occurrence to "Astro Dr Balkrihna"
* 65054789f5f203225bfcdd85 guru id is mapped to two different values but guru name is same i.e. "Astro Dr Shrey" which also exists as "Dr. shrey" so we changed all occurrence to "Astro Dr Shrey"
* 6561ad85213fa6a966bd4188 guru id is mapped to two different values but guru name is same i.e. "Tarot Daljit" which also exists as "Daljit Kaur" so we changed all occurrence to "Tarot Daljit"
* 6561ad85213fa6a966bd4188 guru id is mapped to two different values but guru name is same i.e. "Tarot Chanchal" which also exists as "Tarot Chanchal Singh" so we changed all occurrence to "Tarot Chanchal"
* 656a531d64aaf999f4d0e977 guru id is mapped to three different values but guru name is same i.e. "Tarot Gurpreet" which also exists as "Tarot Preet" and "Tarot Gurpreet Kaur" so we changed all occurrence to "Tarot Gurpreet"
* 65054788f5f203225bfcdd55 mapped to two values "Astro Jha Guruji" and "Astro Jha" merged into "Astro Jha"
* 656a62ad64aaf999f4d10e02 mapped to two values "Astro Rhea" and "Tarot Rhea" merged into "Tarot Rhea"
* 652868d02c4f29e6d54f502c mapped to two values "Astro Aditya Jhha" and "Astro Aditya" merged into "Astro Aditya"
* 65054789f5f203225bfcdd88 mapped to two values "Astro Acharya Dev" and "Acharya Dev" merged into "Acharya Dev"
* 651487105fe5a234ac86a1a5 mapped to two values "Astro Dr Harish" and "Dr. Harish" merged into "Astro Dr Harish"
* 656a557764aaf999f4d0eee7 mapped to two values "Tarrot Ria" and "Tarot Ria" merged into "Tarot Ria"
* 656e09e32d32774bedc8a14d mapped to two values "Tarot Ari" and "Tarot Aritra C" merged into "Tarot Aritra C"
* 658ee7e7883c1bbae7e7dec9 mapped to two values "Astro Akash" and "Aham T" merged into "Astro Akash"
* 654957a81a642edaf98c6c9c mapped to two values "Tarot Rupika" and "Tarot Gargi" merged into "Tarot Rupika" because it has more data points than previous one
* 65054788f5f203225bfcdd5b mapped to two values "Astro Prashant" and "Astro Vishwajeet" merged into "Astro Vishwajeet" because it has more data points than previous one
* 65054782f5f203225bfcdc64 mapped to two values "Himanshu Pandey" and "Swami Chandreshwaranand G
* " merged into "Himanshu Pandey" because it has more data points than previous one

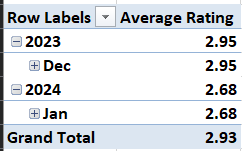
1. **What is the change in daily call volume day by day and also find the average daily call volume.**   
   In **time\_series\_analysis** sheet using pivot table we can find the daily call volumes reflected in below table as  
     
   first calculate volume of call i.e. (count of CallSid) on each day and then subtracting it with previous day’s volume  
     
     
     
   also average daily call volume can be found by   
     
     
   using following formulas  
   =SUM(time\_series\_detailed\_analysis[Total Calls])  
   =COUNT(time\_series\_detailed\_analysis[Date])  
   =COUNT(guru\_detailed\_analysis[gid])
2. **Which months experienced the highest and lowest call volumes?**
3. **What is the total operational cost for that month?**For both these question solution is given in **time\_series\_analysis** sheet as   
     
     
     
   using following formulas   
     
   for max min call volume  
     
   =MAX(time\_series\_detailed\_analysis[Total Calls])  
   =MIN(time\_series\_detailed\_analysis[Total Calls])  
     
   for month related to max min call volume  
     
   =INDEX(time\_series\_detailed\_analysis[Month], MATCH($G39, time\_series\_detailed\_analysis[Total Calls], 0))  
   =INDEX(time\_series\_detailed\_analysis[Month], MATCH($G40, time\_series\_detailed\_analysis[Total Calls], 0))  
     
   for operational cost and total revenue in these months  
     
   =SUMIF(time\_series\_detailed\_analysis[Month], $H39, time\_series\_detailed\_analysis[Total Astrologer Earning])  
   =SUMIF(time\_series\_detailed\_analysis[Month], $H39, time\_series\_detailed\_analysis[Total Revenue])
4. **What is the average number of calls handled per agent per day?**  
   In **time\_series\_analysis** sheet following solution was derived   
     
     
   using following formulas  
   =SUM(time\_series\_detailed\_analysis[Total Calls]) to count total calls  
   =COUNT(time\_series\_detailed\_analysis[Date]) to count total days  
   =COUNT(guru\_detailed\_analysis[gid]) to count unique guru  
   =($L$38 / $L$39) to get average total call volume  
   =($L$43 / $L$40) to get average calls handled by each guru
5. **How many repeat callers are there, and what percentage of total calls do they represent?**  
   In **user\_analysis** sheet we can find the following data that gives user repeat callers as **1275** and repeat call percentage as **56.63%** (about 4737 repeat calls)  
     
     
   using pivot table and following formulas  
   =COUNT(T3:$T$10346) to get total no. of users  
   =COUNTIF($U$3:$U$10346, ">1") to get users with more than one call  
   =($Y$4/$Y$3) to get percentage of repeated callers from all users  
     
   =SUM($U$3:$U$10346) to get sum of all calls made  
   =SUMIFS($U$3:$U$10346, $U$3:$U$10346, ">1") to get sum of calls where count of callId is more than 1 i.e. repeated call  
   =$Y$13-$Y$4 to exclude 1st call from repeated calls   
   =($Y$13 / $Y$12) to find percentage of repeat calls
6. **What are the total sales generated by the call centre for each product category?**

In **product\_platform\_analysis** sheet there is a table that showcase the following data which was calculated using formula from base table   
  
=ROUND(SUMIFS(base\_table[netAmount], base\_table[consultationType], platform\_product\_analysis!$A10), 2) ****

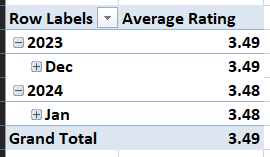
1. **How many calls were made for each user ID and guru ID?**Under the **guru\_analysis** sheet and the **user\_analysis** sheet, pivot tables are present that will show the number of calls for each user and each guru  
     
     
     
     
   For total no. of calls i.e. 8363 between them we have used it in **product\_platform\_analysis** sheet derived from base table using formula   
     
   =COUNTIFS(base\_table[consultationType], platform\_product\_analysis!$A10)  
     
   

|  |  |
| --- | --- |
| Correlation | **0.516725** |

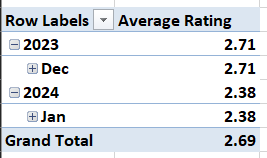
1. **What is the correlation between call duration and customer satisfaction?**  
   In **user\_analysis** sheet we have used correlation formula to find correlation between call duration and customer satisfaction  
     
     
     
   Formula is   
   =CORREL(user\_detailed\_analysis[average time for calls], user\_detailed\_analysis[average rating for calls])
2. **Which guru has the highest and lowest customer satisfaction scores?**  
   
3. **What is the average customer satisfaction score by month?**



**Overall Average Rating**

****

**Average Ratings on Call**

****

**Average Ratings on Call**

1. **How many categorical columns are there in the data? [Search about categorical and continuous data, and try to answer this question]**  
   **19 columns**

* \_id
* user
* chatStatus
* guru
* guruName
* consultationType
* website
* refundStatus
* updatedAt
* statementEntryId
* chatStartTime
* chatEndTime
* callChannel
* callIvrType
* callStatus
* CallSid
* astrologerCallStatus
* region
* userCallStatus

**Subjective Question:**

1. **Should the investment be used to hire more agents, improve training programs, or upgrade call center technology?**

**Hire more agents – 40%**

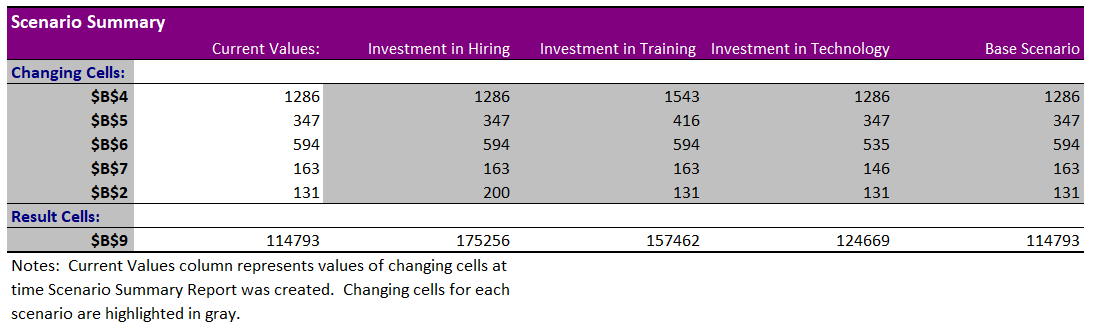
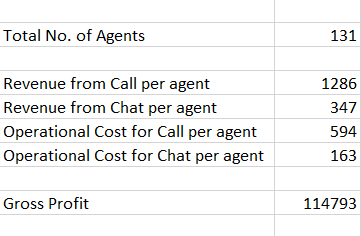
* Increase in revenue as more agents will eventually lead to handling of more calls and chats
* Reduced workload on existing agents during peak times
* Reduction in drop-off rate as peak time call handling is balanced

**Improve Training – 40%**

* This will enhance the average call on time duration, which has a strong correlation with gross profit
* Increase in customer satisfaction score, which will increase the retention rate of customers

**Upgrade Call Centre Technology – 20%**

* First upgrading and automating the chat system since it has a high failure, incomplete rate and lower satisfaction score
* Second invest in call scheduling system to efficiently manage peak load times

**Using the scenario manager, we can find a variation of gross profit** **Recommendation**

* Investing 40% in hiring new agents, 40% in training agents, and 20% in technology improvement and upgradation

1. **What are the potential risks of each investment option (hiring, training, technology upgrades), and how can they be mitigated?**

|  |  |  |
| --- | --- | --- |
| **Investment Domains** | **Risks** | **Mitigation** |
| **Hiring New Agents** | a. Create competition among existing agents  b. Required a lot of time and long process to find the right agent | a. Using platforms that help in hiring to reduce cost  b. Train existing agents to improve performance |
| **Training Programs** | a. Requires additional investment to develop modules  b. Correct program and modules are essential as per experience and level of agents | a. Developing in-house training systems where experience agents train new agents  b. Monthly or Weekly Q&A or consultation session of low performing agents with experienced agents |
| **Upgrading Call Centre Technology** | a. Additional Investment that will reduce operational expense  b. Refining the system using AI will require additional research and more budget | a. Initially not implementing AI just developing a basic system to check if the system sustains in the ecosystem |

1. **How does AstroSage's call center performance compare to AstroGuru's average call volume, customer satisfaction, and agent performance?**

**Will you use any aggregation function or a visualization here to solve the problem?**

By using aggregate functions like SUMIFS, AVERAGEIFS, COUNTIFS and CORREL we can first aggregate data on daily basis which will give us a baseline to compare both the companies

Like here Gross Profit and average call duration over time is used to find dependency of profits on call duration.

**Insights:**

* As we can see through this combo bar and line chart, gross profit is high when the average call duration on time is more
* The correlation coefficient between them is 0.714, which indicates high dependency

**Recommendations:**

* We should focus on increasing the average call duration time

1. **How can the call center improve its handling of peak call periods to ensure high customer satisfaction?**

**Mention the functionality you will use for giving the suggestions, will it be any aggregated function or a visualization?**

For this answer we have used a powerful aggregation and summarizing tool that is pivot tables. Using date in row field we have made three others columns that count total users per day, count of call volume and count of chat volume that day. And summarized them using visualization to find affect of peak users on chat and call

**Insights:**

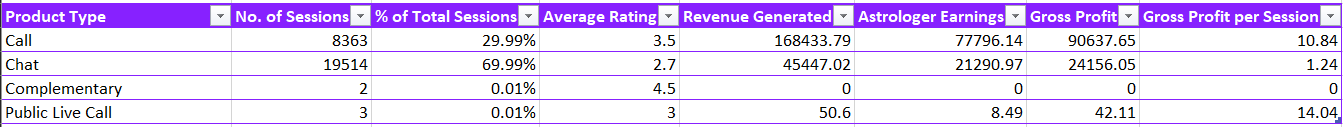
* There was a rise in active users from 1st December 2023 till 15th December 2023 after which active users have been declining
* When active users increase, call volume drops and chat volume increases, indicating that agents are unable to handle calls and prefer chats

**Recommendations:**

* Hiring more agents will enable to absorb fluctuations on peak time
* Chats have a high failure rate so we need to invest in chat systems or use AI

1. **Based on historical data, what strategic initiatives should be prioritized to improve efficiency and customer satisfaction?**

**Consultation type analysis table**



Here we can find the customers that have provided ratings by selecting rating in one row of pivot table and the count of customers on another one then grouping them in range of 0-1, 2-3, 4-5, 6-8 which will give us the count of customers in each bucket

**Insights:**

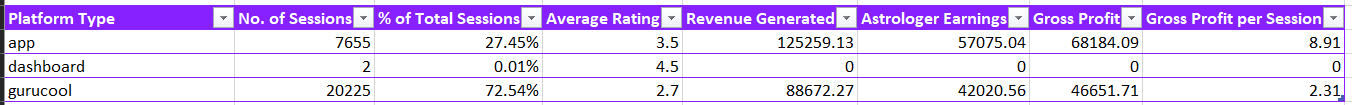
* People have given very low ratings to calls a compare to calls and we don’t have any rating in 0-8 bucket in calls
* Low ratings in calls is due to a high failed and incomplete rate. In calls our agents need to perform better to get at least one rating in higher bucket

**Recommendations:**

* We need to invest in chat automation and AI systems to actively engage customers on chat that will decrease failed and incomplete rate
* Invest in training programs of agents to get better call ratings

1. **What can be the key factors contributing to high customer satisfaction scores, and how can these be leveraged to improve overall performance?**

**What is the basis for the suggestions? And mention how you decided if the satisfaction score affects the ratings.**

****

**Insights:**

* Customer satisfaction is higher on calls as compared to chats because more customer falls in the bucket of 2-3 and 4-5 in calls
* As we can see in the second visualization, there is higher revenue on higher customer satisfaction in the 2400-2699 bucket, the 3000-3299 bucket, and 3300-3599 bucket of average call on duration
* Even though there are less sessions on the App they generate more revenue than gurucool or dashboard

**Recommendations:**

* We should invest in the training and development of agents so that they can connect better with customers delivering higher average call on duration value
* Strategically, moving customers from different sources to the App since it has high revenue per session
* Developing other websites to improve customer service satisfaction

1. **How should the call center balance the workload among agents to ensure optimal performance and avoid burnout?**

About this question, we would use the analysis used previously in question number 3 and adding following suggestions  
**- Hiring new agents** will reduce the load on existing agents and improve the drop-off rate since more calls can be managed   
- **Automating the chat system using AI** that can suggest specific timing when the agent will be available and suggest available timing to the agent based on user activity and availability  
- **Providing training to agents** so that they can connect with customers at a deeper level and improve ratings as well as call duration per call

1. **What new technologies or tools could be implemented to enhance call centre operations and customer service?**

- **Improving the existing chat** system to decrease chat failure and incomplete status. Additionally using AI assistance to handle incomplete status that will give more time to agents   
- **User active time tracker** This will help us to understand availability of user on various time intervals in a day. So we can strategically place calls sessions na chats during active time

1. **What metrics should be included in the final dashboard to view call center performance and guide investment decisions comprehensively?**

The following are some of the **metrics that can be considered** for investment decisions

* Customer retention rate on call and on chats (recurring customers means higher revenue and ratings)
* Average on-call duration for customers (high correlation with gross profit)
* Ratings for Agent (validate performance of agents)
* Daily Active users (directly affects call and chat volume)

1. **How would you allocate a 1 crore rupee investment to optimize operational efficiency, enhance customer satisfaction, and boost profitability, and what analysis-based recommendations would you offer to support this?**

**[you have to give bullet pointers to answer this question]**

* **Investing 20% in chat automation and call scheduling software** since chats have a high failure rate and incomplete rate, and calls are low during peak times
* **Investing 40% in agent training** to improve customer satisfaction and improve agents' skills, and call duration time
* **Investing 40% in hiring new agents** to increase the volume of calls and chats and reduce the load on current agents by handling more calls and chats

**Ensure that you put the slicers for choosing the country and year to observe the dashboard since the management will have a long discussion that can go on for weeks.**

**Note: The dashboard would be more interactive and user-friendly, allowing management to explore data in detail and make informed decisions.**