

CALL CENTRE DATA ANALYSIS

Name: Prachi Goyal

Introduction:

This report presents an analysis of call center data, aimed at evaluating key metrics and providing actionable insights to optimize call handling, improve user satisfaction, and enhance call center performance. The data includes various parameters such as call charges, TalkTime, user spending, refund status, and more. Through detailed data exploration, visualizations, and performance analysis, this report highlights areas for potential improvements.

Objective:

The objective of this analysis is to explore and interpret the call center data effectively to identify areas of improvement in operations, ensure better user engagement, and optimize call handling for more efficient performance.

Task 1: Data Exploration

Task: Load the dataset and explore its structure. Handle missing values and provide summary statistics

Step 1: Load the dataset using Python (pandas)

Step 2: Check for missing values and handle them:

- Fill missing values for numerical columns with the median.
- Fill missing values for categorical columns with the mode

Solution: Missing Values per Column:

_id	0
user	0
chatStatus	8512
guru	0
guruName	0
gid	0
uid	0
consultationType	0
website	0
refundStatus	0
isWhitelistedUser	0
chatSeconds	8513
queue	0
freeCall	0
freeChat	0
createdAt	0
updatedAt	0
__v	0
statementEntryId	151
chatStartTime	18697
chatEndTime	14996
timeDuration	0
callChannel	19519
callivrType	19665

callStatus	19486
CallSid	19662
amount	0
astrologerCallStatus	11769
astrologerOnCallDuration	19662
astrologersEarnings	0
netAmount	0
region	19008
userCallStatus	18682
userOnCallDuration	0
RecordingUrl	23702
feedback	25411
hideHistory	8689
cardPosition	27317
Remedies	27925
offer	9249
refunds[0]	28022
source	11736
callSeconds	28025
complementaryCallOrganiser	28025
disconnectedBy	28018

Step 3: Provide summary statistics for key numeric columns

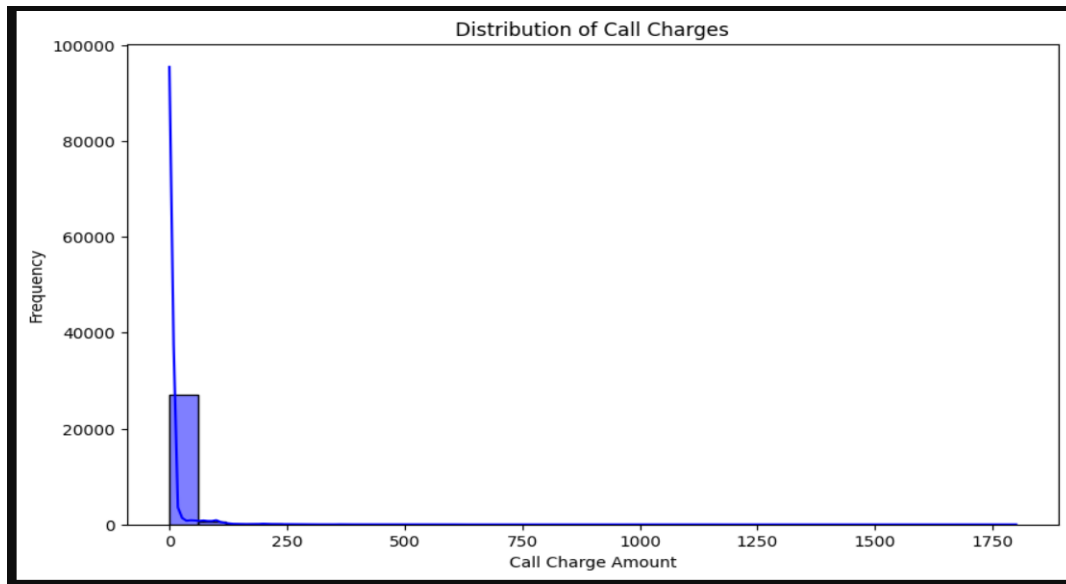
Solution:

Summary Statistics for Numeric Columns:

	amount	timeDuration
count	28027.000000	28027.000000
mean	7.637846	1.533177
std	50.648404	3.025836
min	0.000000	-0.166667
25%	0.000000	0.000000
50%	0.000000	0.100000
75%	0.000000	3.000000
max	1800.000000	67.000000

Step 4: Visualize the distribution of call charges using a histogram.

Solution:



Task 2: Call Center Performance Metrics

Task: Calculate metrics like average TalkTime, most common source of calls, total earnings, and the relationship between TalkTime and Charge.

Step 1: Calculate the average TalkTime for different call activities

Approach: In the data, there is no column of TalkTime so I used the timeDuration column because In the timeDuration column we have the data of the duration in which the talk has done and for call activities I used the callStatus by which we can easily figure out the analysis of the talktime for different call activities

Solution: Average TalkTime for Different Call Activities:

	callStatus	timeDuration
0	busy	0.000000
1	completed	4.727163
2	failed	0.015816
3	incomplete	0.277429
4	no-answer	0.000000

Step 2: Determine the most common source of calls

Solution: The most common source of calls is: Production

Step 3: Calculate total earnings for masters and spending for users

Approach: Assuming 'amount' is the total charge and 'astrologersEarnings' is earnings for masters and assuming 'netAmount' is the total amount spent by users

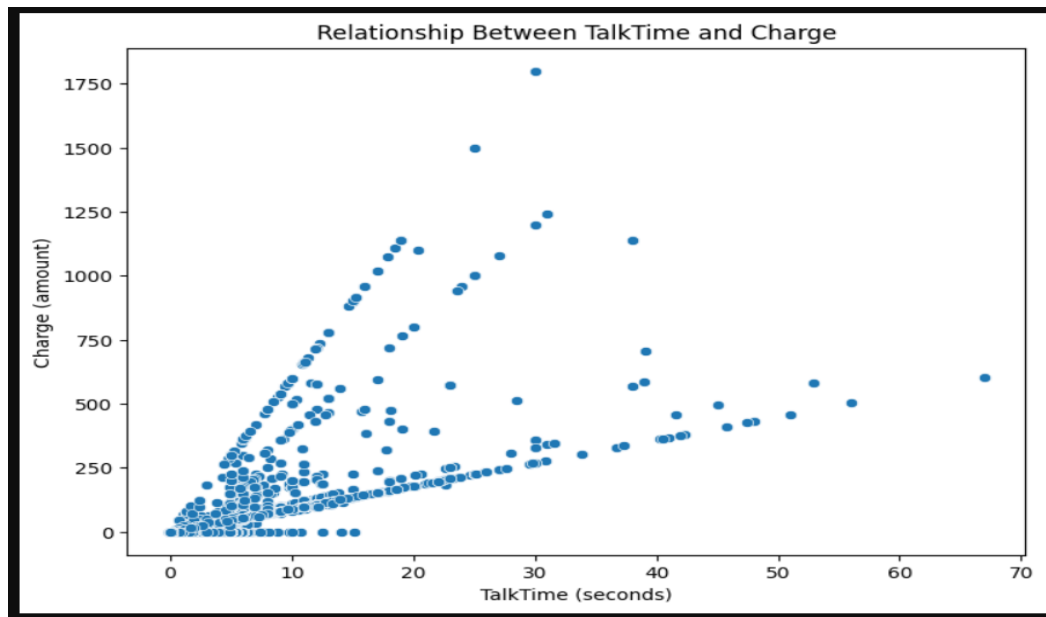
Solution: Total earnings for masters: 99146.57095000001

Total spending for users: 213987.31533333333

Step 4: Analyze the relationship between TalkTime and Charge.

Approach: Assuming 'TalkTime' is the timeDuration and 'charge' is 'amount'. To analyze the relationship, we can use correlation and scatter plot.

Solution: Correlation between TalkTime and Charge: 0.5992186910956641



Task 3: Call Handling Analysis

Task: Analyze the time it takes for calls to be connected, reasons for disconnection, and call handling patterns.

Step 1: Calculate the average time it takes for calls to be connected.

Approach: To calculate the average time it takes for calls to be connected (I used $\text{ConnectionDelay} = \text{astrologerOnCallDuration} - \text{userOnCallDuration}$) and after it I find the mean of ConnectionDelay

Solution: Average time it takes for calls to be connected: 21.226709958254542 seconds

Step 2: Identify the most common reason for call disconnection.

Approach: To Identify the most common reason for disconnection assuming 'callStatus' contains the reason. We can count the occurrences of each reason in the 'callStatus' column.

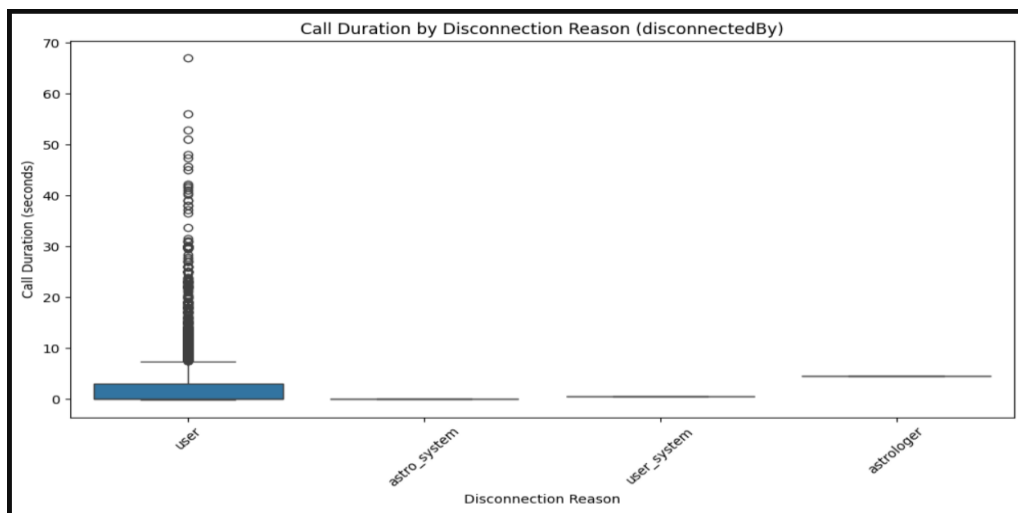
Solution: The most common reason for call disconnection: completed

Step 3: Analyze HangUpTime patterns.

Approach: Assuming 'HangUpTime' is the timeDuration and Use a boxplot to show the distribution of timeDuration by disconnectedBy

Solution:

```
count    28027.000000
mean      1.533177
std       3.025836
min       -0.166667
25%       0.000000
50%       0.100000
75%       3.000000
max       67.000000
```



Task 4: Order and Refund Analysis

Task: Analyze the order status distribution and refund status

Step 1: Determine the order status distribution.

Approach: Determine the order status distribution for calls. We assume 'callStatus' column contains the status of the call (completed, cancelled, in-progress, etc.)

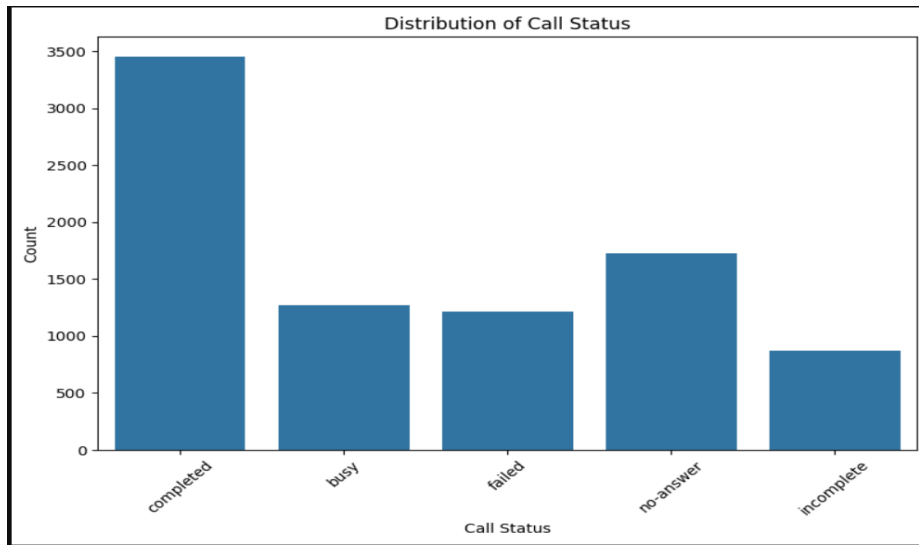
Solution:

Order Status Distribution for Calls:

```
completed    3453
no-answer    1729
busy         1270
```

failed 1214

incomplete 875



Step 2: Identify the refund status distribution.

Approach: There is no refund amount column in the data. We assume 'refundStatus' indicates the status of the refund (refunded, pending, etc.)

Solution:

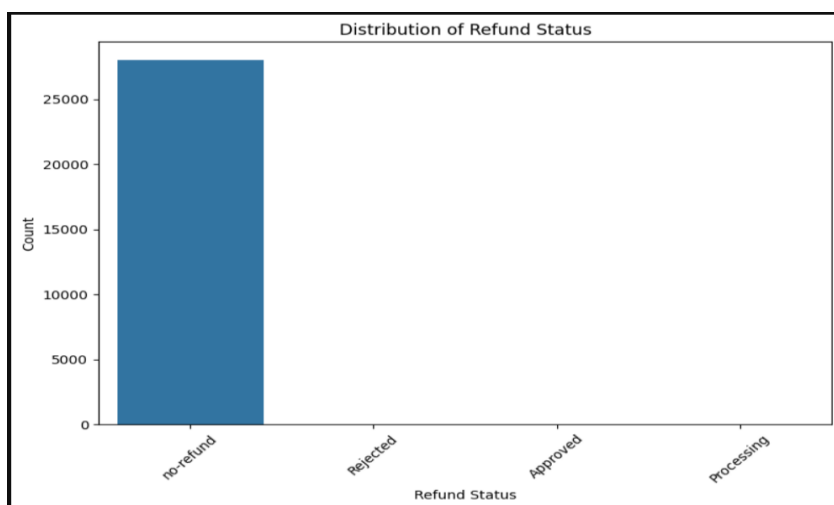
Refund Status Distribution:

no-refund 28022

Approved 3

Rejected 1

Processing 1



Task 5: Insights and Recommendations

Task: To find the Insights and give the recommendation.

Insights:

1. Average timeDuration for completed calls is 4.72 minutes.
2. The most common reason for disconnection is Call Completed.
3. Mostly the call is disconnected by user.
4. By the chart of call status Distribution we can see that the highest number of call status is completed which is 3453

Recommendation:

1. Consider offering promotions for users with longer TalkTime to increase engagement.
2. Increase the TalkTime to complete the call for the user satisfaction.

Task 6: Data Visualization

Additional Visualization:

1. Find the most common consultation type and the count of the most common consultation type.

Solution:

The most common consultation type is: Chat
It appears 19514 times in the dataset.

2. Find the name of the guru who organised the most complementary call

Solution:

The guru who organised the more complementary call is Tarot Bee.

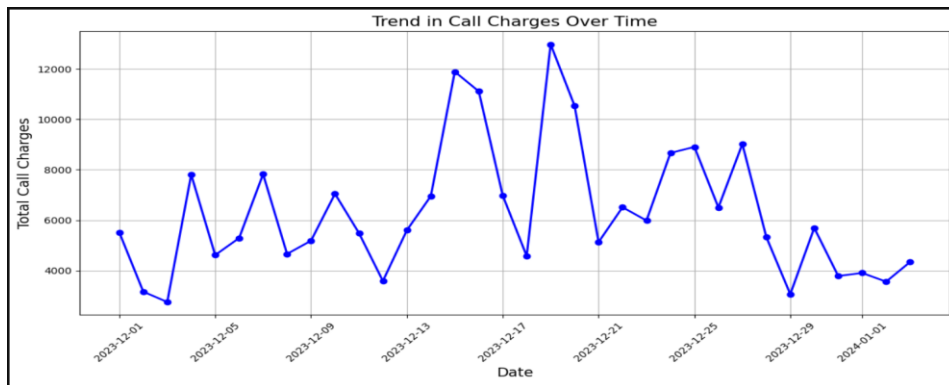
Task 7: (a) Additional Visualization

Task: Create a line chart showing the trend in call charges over time and create a scatter plot to analyze the relationship between TalkTime and UserSpend.

Step: Create a Line Chart Showing the Trend in Call Charges Over Time

Approach: Firstly, Ensure that the 'createdAt' column is in datetime format then Group the data by date then Aggregate the total call charges (amount) by date and Plot the line chart for the trend in call charges over time

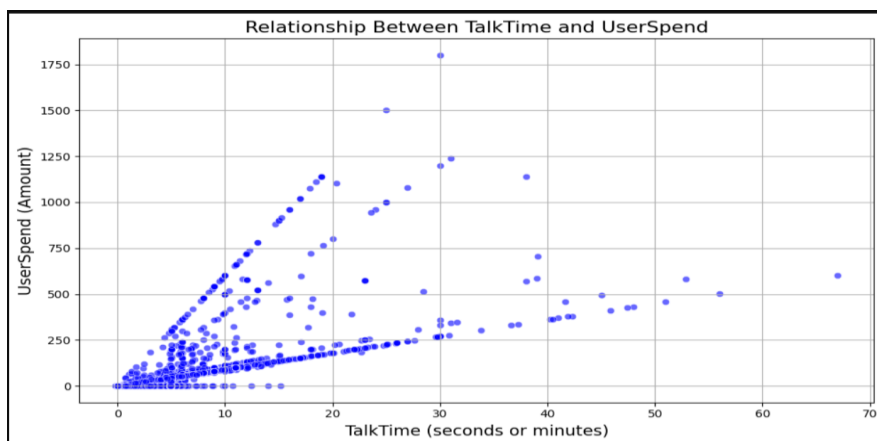
Solution:



Step: Create a scatter plot to analyze the relationship between TalkTime and UserSpend.

Approach: Assuming the TalkTime is timeDuration and the UserSpend is amount.

Solution:



Task 7: Interpretation

- **Histogram of Call Charges:**

The Distribution is heavily skewed to the right. This indicates that a majority of calls have lower charges, while a smaller number of calls have significantly higher charges. It suggest that the pricing strategy might favor lower call charges, possibly attracting a larger customer base with budget-friendly options.

The most frequent call charges appears to be around 250 units which indicates that a common call duration or a popular pricing plan.

- **Line Chart:**

The chart exhibits fluctuations in call charges over the given time period. There doesn't appear to be a clear, repeating seasonal pattern in the data. The peaks and troughs seem to occur randomly.

The variations in call charges could be influenced by external factors like holidays, promotions, or changes in service offerings.

- **Scatter Plot:**

The points in the scatter plot generally move upwards from left to right, indicating a positive relationship. There seems to be linear trend in the data, suggesting a linear relationship between talk time and user spend.

There are a few outliers present, which might represent unusual call patterns or pricing plan.

CONCLUSION

The analysis of the call center data has provided actionable insights into improving the overall efficiency of operations, increasing revenue, and enhancing the customer experience. By leveraging the power of data, the call center can make informed decisions, optimize its workforce, enhance customer satisfaction, and ultimately, improve profitability. Implementing these recommendations will create a more streamlined, data-driven call center environment that can better serve customers and increase operational efficiency.

Thank You Message

Thank You for Your Attention!

I sincerely appreciate your time and attention in reviewing this report. The analysis conducted here aims to provide valuable insights into the current state of call center operations and offer data-driven recommendations for improvement. I hope these insights will prove useful in optimizing performance and enhancing overall user satisfaction. Should you have any questions or require further details, please feel free to reach out. I look forward to the opportunity to discuss these findings in greater depth.

Thank you again for your time!

Best regards,

Prachi Goyal

goyalprachi2324@gmail.com

8359978903