Customer Cancellation Profile Analysis

About:

This project aims to analyze a database to determine the reasons for customer cancellations.

Proposal:

The project proposal is to extract information from a **.csv** format database to conduct an analysis and identify the reasons for customer cancellations. Based on this analysis, effective solutions will be proposed to reduce the number of cancellations.

Repository Structure:

- **data:** Here you will find the **.csv** file containing the data used for the analysis.
- **img:** This is where the screenshots obtained during the analysis can be found.
- **notebook:** This directory contains the **Jupyter notebook** where the analysis was conducted and the results obtained.
- **readme_translated:** This repository contains the **PDF** with the report translated into English.

Language Used:

Libraries Used:

Methodology:

Initially, we used the **Pandas** library to import and read the database.

	CustomerID	idade	sexo	tempo_como_cliente	frequencia_uso	ligacoes_callcenter	dias_atraso	assinatura	duracao_contrato	total_gasto	meses_ultima_interacao	cancelou
0	2.0	30.0	Female	39.0	14.0	5.0	18.0	Standard	Annual	932.0	17.0	1.0
1	3.0	65.0	Female	49.0	1.0	10.0	8.0	Basic	Monthly	557.0	6.0	1.0
2	4.0	55.0	Female	14.0	4.0	6.0	18.0	Basic	Quarterly	185.0	3.0	1.0
3	5.0	58.0	Male	38.0	21.0	7.0	7.0	Standard	Monthly	396.0	29.0	1.0
4	6.0	23.0	Male	32.0	20.0	5.0	8.0	Basic	Monthly	617.0	20.0	1.0
5	8.0	51.0	Male	33.0	25.0	9.0	26.0	Premium	Annual	129.0	8.0	1.0
6	9.0	58.0	Female	49.0	12.0	3.0	16.0	Standard	Quarterly	821.0	24.0	1.0
7	10.0	55.0	Female	37.0	8.0	4.0	15.0	Premium	Annual	445.0	30.0	1.0
8	11.0	39.0	Male	12.0	5.0	7.0	4.0	Standard	Quarterly	969.0	13.0	1.0
9	12.0	64.0	Female	3.0	25.0	2.0	11.0	Standard	Quarterly	415.0	29.0	1.0

```
<class 'pandas.core.frame.DataFrame'>
Index: 881659 entries, 0 to 881665
Data columns (total 11 columns):
 #   Column                  Non-Null Count   Dtype
---  ------                  --------------   -----
 0   idade                   881659 non-null  float64
 1   sexo                    881659 non-null  object
 2   tempo_como_cliente      881659 non-null  float64
 3   frequencia_uso          881659 non-null  float64
 4   ligacoes_callcenter     881659 non-null  float64
 5   dias_atraso             881659 non-null  float64
 6   assinatura              881659 non-null  object
 7   duracao_contrato        881659 non-null  object
 8   total_gasto             881659 non-null  float64
 9   meses_ultima_interacao  881659 non-null  float64
 10  cancelou                881659 non-null  float64
dtypes: float64(8), object(3)
memory usage: 80.7+ MB
```

The next step was to process and clean the data. We used the **.info** method to check the data types we were working with. We removed rows containing null values and the **"CustomerID"** column.

Next, we performed an analysis to determine the cancellation rate. We used the **.value_counts()** method to obtain the cancellation count and then applied the **.apply** method with a lambda function to format the values and calculate the percentages. Boolean data types were used to identify **"Cancelled" (1)** and **"Not Cancelled" (0).**

```
cancelou
0.0    379032
1.0     52121
Name: count, dtype: int64

cancelou
0.0    87.91%
1.0    12.09%
Name: proportion, dtype: object
```

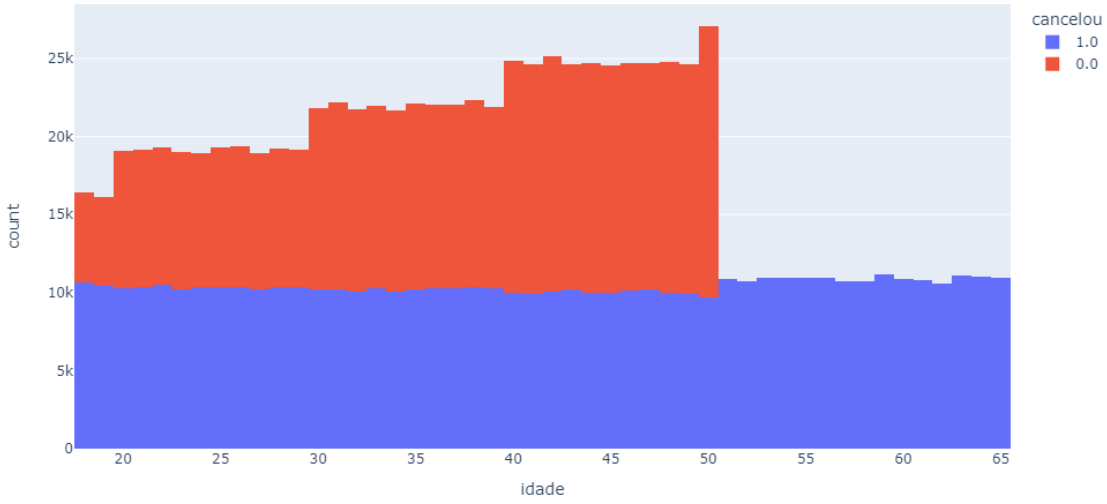
We found that there was a total cancellation rate of **56.71%.**

To aid in the analysis, we created graphs for better visualization. The **Plotly** library was used with a **for** loop.

Analysis:

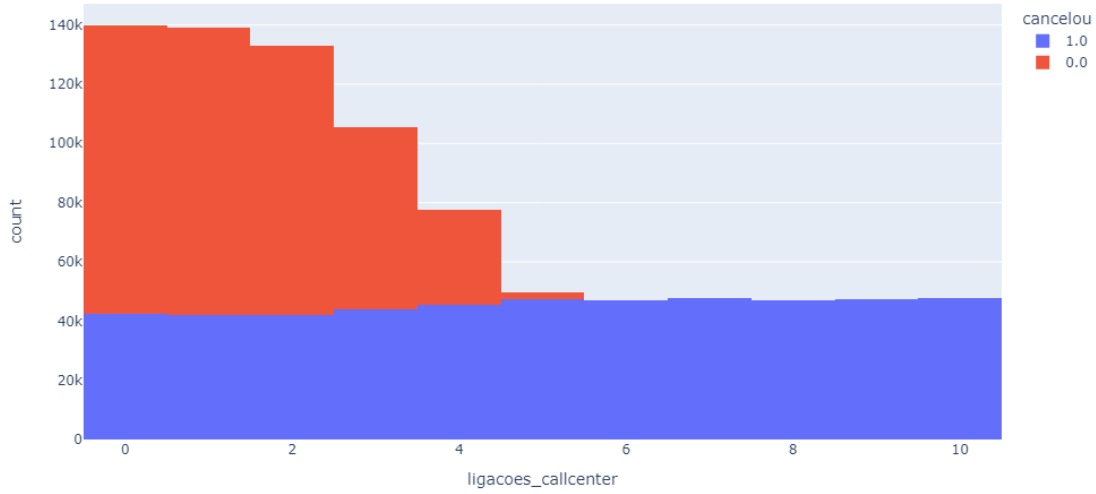
Based on the generated graphs, the analysis highlighted four factors that may be related to customer cancellations:

- Age:



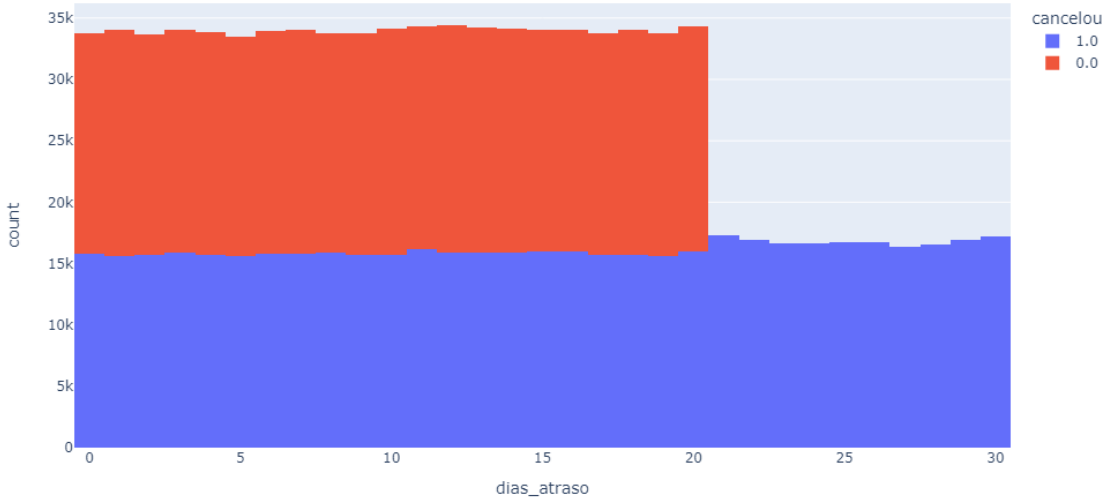
Customers above the age of 50 have a higher cancellation rate, suggesting a specific challenge in retaining this age group.

- Call Center Calls:



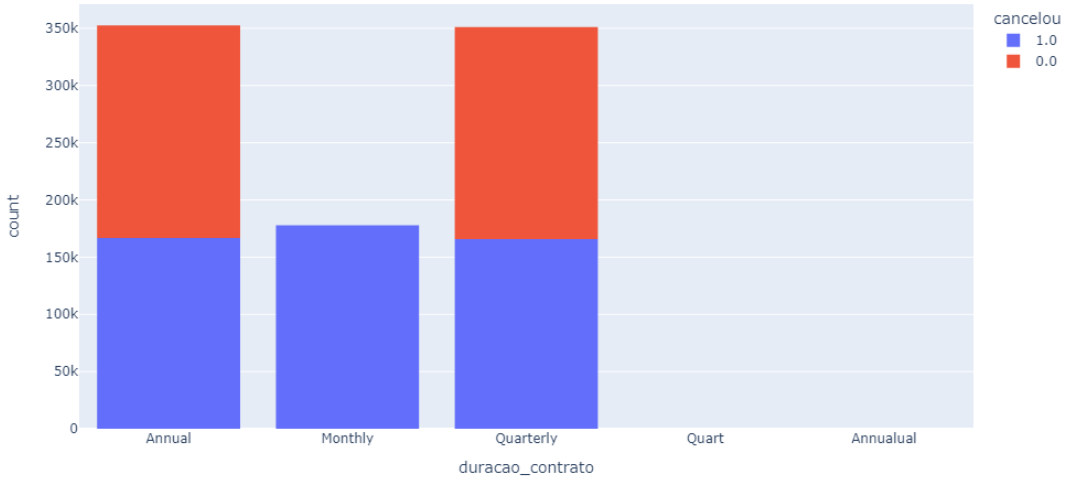
It was observed that customers who made more than 5 calls to the call center are more likely to cancel their subscriptions. This may suggest that customer dissatisfaction or difficulties when contacting support are leading to cancellations.

- Days of Delay:



Customers who accumulate more than 20 days of payment delay are inclined to cancel their subscriptions. This pattern suggests that delinquency may be related to an increase in customer cancellations.

- Contract Duration:



Customers who opt for monthly plans have a stronger tendency to cancel compared to annual and quarterly plans. This may indicate that customers with shorter plans are more likely to discontinue the service.

Conclusion:

After analysis, it was found that customers aged over 50 exhibit a significant cancellation rate. Additionally, customers who make more than 5 calls to the call center are more likely to cancel. Furthermore, those who accumulate more than 20 days of payment delay show a higher propensity to cancel their subscriptions. Lastly, customers who choose monthly plans exhibit a stronger trend toward cancellation.

It is possible to propose some solutions to reduce the number of cancellations and improve customer retention, including:

- Ensure that services and products are user-friendly and easy to understand, especially for older customers, by providing intuitive interfaces, accessible technical support, and clear communication materials.

- Enhance customer service at the call center by seeking to resolve customer issues and inquiries more quickly and efficiently.

- Implement measures to reduce delinquency, such as payment reminders and installment options.

- Offer incentives for customers who opt for long-term contracts, such as discounts or exclusive benefits to increase loyalty.

- Conduct satisfaction surveys to better understand customer needs and expectations, aiming to improve the quality of services provided.