


```
#passo a passo do projeto

#Passo 1: importar a base de dados
import pandas as pd
import numpy as np
import plotly.express as px

tabela = pd.read_csv('/content/cancelamentos.csv')
display(tabela)
```



	CustomerID	idade	sexo	tempo_como_cliente	frequencia_uso	ligacoes_callcenter	dias_atraso	assinatura	duracao_con
0	2.0	30.0	Female	39.0	14.0	5.0	18.0	Standard	/
1	3.0	65.0	Female	49.0	1.0	10.0	8.0	Basic	M
2	4.0	55.0	Female	14.0	4.0	6.0	18.0	Basic	Qu
3	5.0	58.0	Male	38.0	21.0	7.0	7.0	Standard	M
4	6.0	23.0	Male	32.0	20.0	5.0	8.0	Basic	M
...	
881661	449995.0	42.0	Male	54.0	15.0	1.0	3.0	Premium	/
881662	449996.0	25.0	Female	8.0	13.0	1.0	20.0	Premium	/
881663	449997.0	26.0	Male	35.0	27.0	1.0	5.0	Standard	Qu
881664	449998.0	28.0	Male	55.0	14.0	2.0	0.0	Standard	Qu
881665	449999.0	31.0	Male	48.0	20.0	1.0	14.0	Premium	Qu

881666 rows × 12 columns

```
#Passo 2: Tratamento dos Dados
#verificand dados vazios
display(tabela.info())

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 881666 entries, 0 to 881665
Data columns (total 12 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   CustomerID                           881664 non-null float64
1   idade                                881664 non-null float64
2   sexo                                  881664 non-null object
3   tempo_como_cliente                   881663 non-null float64
4   frequencia_uso                       881663 non-null float64
5   ligacoes_callcenter                 881664 non-null float64
6   dias_atraso                          881664 non-null float64
7   assinatura                           881661 non-null object
8   duracao_contrato                     881663 non-null object
9   total_gasto                          881664 non-null float64
10  meses_ultima_interacao               881664 non-null float64
11  cancelou                             881664 non-null float64
dtypes: float64(9), object(3)
memory usage: 80.7+ MB
None

#Passo 2 continuação :tratar valores vazios
tabela = tabela.dropna() # (DROPNA) este comando delata linhas vazias
display(tabela.info())
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 881659 entries, 0 to 881665
Data columns (total 12 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   CustomerID            881659 non-null float64
1   idade                 881659 non-null float64
2   sexo                  881659 non-null object
3   tempo_como_cliente    881659 non-null float64
4   frequencia_uso        881659 non-null float64
5   ligacoes_callcenter   881659 non-null float64
6   dias_atraso           881659 non-null float64
7   assinatura            881659 non-null object
8   duracao_contrato      881659 non-null object
9   total_gasto           881659 non-null float64
10  meses_ultima_interacao 881659 non-null float64
11  cancelou              881659 non-null float64
dtypes: float64(9), object(3)
memory usage: 87.4+ MB
None

# Passo 2 excluindo colunas
# linha excluir = axis =0
# coluna excluir = axis =1
tabela = pd.read_csv('/content/cancelamentos.csv')
tabela = tabela.drop("CustomerID", axis=1)
display(tabela)

      idade  sexo  tempo_como_cliente  frequencia_uso  ligacoes_callcenter  di
0      30.0  Female                39.0             14.0                 5.0
1      65.0  Female                49.0              1.0                10.0
2      55.0  Female                14.0              4.0                 6.0
3      58.0   Male                38.0             21.0                 7.0
4      23.0   Male                32.0             20.0                 5.0
...      ...      ...                ...             ...                 ...
881661  42.0   Male                54.0             15.0                 1.0
881662  25.0  Female                 8.0             13.0                 1.0
881663  26.0   Male                35.0             27.0                 1.0
881664  28.0   Male                55.0             14.0                 2.0
881665  31.0   Male                48.0             20.0                 1.0

881666 rows x 11 columns

# Passo 3: Analise Inicial dos dados
# quantos cancelou
display(tabela['cancelou'].value_counts(normalize=True))

1.0    0.567107
0.0    0.432893
Name: cancelou, dtype: float64

# 56% cancelou, 43% cancelou

# Passo 4: Analise mais Profunda dos dados
import plotly.express as px

for coluna in tabela.columns: # agora ele vai criar todos os graficos em relação ao cancelamento
    grafico = px.histogram(tabela, x=coluna, color="cancelou")
    grafico.show()
```

