

# Python Insights - Analisando dados com Python

## Case - Cancelamento alto de clientes

Você foi contratado por uma empresa com mais de 50mil clientes para um projeto de dados. Recentemente a empresa percebeu que da sua base de clientes, a maioria são clientes inativos, ou seja, que já cancelaram o serviço.

Precisando melhorar seus resultados, ela quer conseguir entender os principais motivos desses cancelamentos e quais as ações mais eficientes para reduzir esse número.

In [103...

```
"""Script de Análise de Dados

Script modelo para análise de dados utilizando bibliotecas populares como pandas, n

Objetivo: Nesse exemplo de uso, analisamos a base de dados de clientes de uma empre

Passo a passo do projeto
Passo 1: Abrir a base de dados (Importar a base de dados)
Passo 2: Visualizar a base de dados (Que tipo de dados ela me informa? Como entende
Passo 3: Corrigir os problemas da base de dados (tratamento de dados)
Passo 4: Análise inicial (Quais clientes cancelaram?)
Passo 5: Análise detalhada (causa do cancelamento dos clientes, como cada coluna im
"""

#Passo 1: Abrir a base de dados (Importar a base de dados)
import pandas as pd

tabela = pd.read_csv("cancelamentos.csv")

display(tabela)
# Passo 2: Visualizar a base de dados
# Que tipo de dados ela me informa?
# Como entender essas informações?
# Quais os erros/problemas da base de dados?

tabela = tabela.drop(columns="CustomerID")
display(tabela)
# A coluna "CustomerID" não agrega valor para a análise, então vamos removê-la, inf
```

	CustomerID	idade	sexo	tempo_como_cliente	frequencia_uso	ligacoes_callcenter
0	349936.0	23.0	Male	13.0	22.0	2.0
1	100634.0	49.0	Male	55.0	16.0	3.0
2	301263.0	30.0	Male	7.0	1.0	0.0
3	119358.0	26.0	Male	40.0	5.0	3.0
4	130955.0	27.0	Female	17.0	30.0	5.0
...	...	...	...	...	...	...
49995	195680.0	62.0	Female	35.0	7.0	2.0
49996	43477.0	36.0	Male	43.0	21.0	2.0
49997	169273.0	55.0	Male	42.0	8.0	1.0
49998	310693.0	40.0	Female	14.0	19.0	1.0
49999	251915.0	64.0	Male	41.0	29.0	5.0

50000 rows × 12 columns

	idade	sexo	tempo_como_cliente	frequencia_uso	ligacoes_callcenter	dias_atraso
0	23.0	Male	13.0	22.0	2.0	1.0
1	49.0	Male	55.0	16.0	3.0	6.0
2	30.0	Male	7.0	1.0	0.0	8.0
3	26.0	Male	40.0	5.0	3.0	8.0
4	27.0	Female	17.0	30.0	5.0	6.0
...	...	...	...	...	...	...
49995	62.0	Female	35.0	7.0	2.0	8.0
49996	36.0	Male	43.0	21.0	2.0	30.0
49997	55.0	Male	42.0	8.0	1.0	12.0
49998	40.0	Female	14.0	19.0	1.0	17.0
49999	64.0	Male	41.0	29.0	5.0	9.0

50000 rows × 11 columns

In [104...

```
# Passo 3: Corrigir os problemas da base de dados (tratamento de dados)
display(tabela.info())
# Valores em formatos errados: -> Não há valores em formatos errados
# Informações faltando: -> Existem 3 clientes com a informação "sexo" faltando, e 2
# Numa base de dados com 50 mil clientes, esses valores faltando não impactam na an
```

```
tabela = tabela.dropna()
display(tabela.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50000 entries, 0 to 49999
Data columns (total 11 columns):
#   Column                Non-Null Count  Dtype
---  -
0   idade                 50000 non-null  float64
1   sexo                 49997 non-null  object
2   tempo_como_cliente   49998 non-null  float64
3   frequencia_uso       50000 non-null  float64
4   ligacoes_callcenter 50000 non-null  float64
5   dias_atraso         50000 non-null  float64
6   assinatura           50000 non-null  object
7   duracao_contrato     50000 non-null  object
8   total_gasto          50000 non-null  float64
9   meses_ultima_interacao 50000 non-null  float64
10  cancelou             50000 non-null  float64
```

```
dtypes: float64(8), object(3)
```

```
memory usage: 4.2+ MB
```

```
None
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 49996 entries, 0 to 49999
Data columns (total 11 columns):
#   Column                Non-Null Count  Dtype
---  -
0   idade                 49996 non-null  float64
1   sexo                 49996 non-null  object
2   tempo_como_cliente   49996 non-null  float64
3   frequencia_uso       49996 non-null  float64
4   ligacoes_callcenter 49996 non-null  float64
5   dias_atraso         49996 non-null  float64
6   assinatura           49996 non-null  object
7   duracao_contrato     49996 non-null  object
8   total_gasto          49996 non-null  float64
9   meses_ultima_interacao 49996 non-null  float64
10  cancelou             49996 non-null  float64
```

```
dtypes: float64(8), object(3)
```

```
memory usage: 4.6+ MB
```

```
None
```

In [105...

```
#Passo 4: Análise inicial (Quais clientes cancelaram?)
# O cliente que cancelou o serviço está representado na coluna "cancelamento", onde
display(tabela["cancelou"].value_counts()) # Quantidade de clientes que cancelaram

display(tabela["cancelou"].value_counts(normalize=True)) # Porcentagem de clientes
```

```
cancelou
```

```
1.0    28393
```

```
0.0    21603
```

```
Name: count, dtype: int64
```

```
cancelou
```

```
1.0    0.567905
```

```
0.0    0.432095
```

```
Name: proportion, dtype: float64
```

In [106...

```
# Passo 5: Análise detalhada (causa do cancelamento dos clientes, como cada coluna
import plotly.express as px
# Criar os gráficos de cada coluna em relação ao cancelamento
for coluna in tabela.columns:
    if coluna != "cancelou":
        grafico = px.histogram(tabela, x=coluna, color="cancelou", title=f"Análise
        # Exibir o gráfico
        grafico.show()
```

## Relatório Final

Após analisar os gráficos, podemos tirar algumas conclusões sobre os motivos de cancelamento dos clientes. Essas conclusões podem ser usadas para melhorar os serviços oferecidos e reduzir a taxa de cancelamento no futuro.

### **Todos os clientes que tinham plano mensal cancelaram o serviço.**

a. A empresa pode considerar oferecer planos anuais com descontos para incentivar os clientes a permanecerem por mais tempo.

### **Clientes acima de 50 anos têm uma taxa de cancelamento mais alta.**

a. A empresa pode focar suas ações de marketing em indivíduos mais jovens. b. Rever se a interface do aplicativo do serviço está amigável para todas as faixas etárias.

### **Clientes que ligaram para o suporte técnico mais de 4 vezes têm uma taxa de cancelamento mais alta.**

a. A empresa deve investir em melhorar a qualidade do serviço para reduzir a necessidade de suporte técnico. b. Otimizar o atendimento ao cliente para resolver em até 2 ligações. Se houver uma terceira ligação, oferecer um canal direto com um especialista.

### **Clientes que atrasam o pagamento por mais de 20 dias têm uma taxa de cancelamento mais alta.**

a. Implementar lembretes de pagamento em caso de atrasos.

b. Atrasos maiores que 10 dias devem acionar um contato proativo do time financeiro para entender o motivo do atraso.

c. Oferecer opções facilitação do pagamento para clientes inadimplentes.

In [ ]:

```
#Taxa de cancelamento anterior: 56%
#duracao_contrato
print("duração de contrato")
condicao1 = tabela["duracao_contrato"] != "Monthly"
tabela1 = tabela[condicao1]
display(tabela1["cancelou"].value_counts(normalize=True))

#idade
```

```

print("idade")
condicao2 = tabela["idade"] <= 50
tabela2 = tabela[condicao2]
display(tabela2["cancelou"].value_counts(normalize=True))

#ligacoes_callcenter
print("ligações callcenter")
condicao3 = tabela["ligacoes_callcenter"] <= 4
tabela3 = tabela[condicao3]
display(tabela3["cancelou"].value_counts(normalize=True))

#dias_atraso
print("dias de atraso")
condicao4 = tabela["dias_atraso"] <= 20
tabela4 = tabela[condicao4]
display(tabela4["cancelou"].value_counts(normalize=True))

#duracao_contrato + idade
print("duração de contrato + idade")
tabela5 = tabela[condicao1 & condicao2]
display(tabela5["cancelou"].value_counts(normalize=True))

#duracao_contrato + ligacoes_callcenter
print("duração de contrato + ligações callcenter")
tabela6 = tabela[condicao1 & condicao3]
display(tabela6["cancelou"].value_counts(normalize=True))

#duracao_contrato + dias_atraso
print("duração de contrato + dias de atraso")
tabela7 = tabela[condicao1 & condicao4]
display(tabela7["cancelou"].value_counts(normalize=True))

#idade + ligacoes_callcenter
print("idade + ligações callcenter")
tabela8 = tabela[condicao2 & condicao3]
display(tabela8["cancelou"].value_counts(normalize=True))

#idade + dias_atraso
print("idade + dias de atraso")
tabela9 = tabela[condicao2 & condicao4]
display(tabela9["cancelou"].value_counts(normalize=True))

#ligacoes_callcenter + dias_atraso
print("ligações callcenter + dias de atraso")
tabela10 = tabela[condicao3 & condicao4]
display(tabela10["cancelou"].value_counts(normalize=True))

#duracao_contrato + ligacoes_callcenter + dias_atraso
print("duração de contrato + ligações callcenter + dias de atraso")
tabela11 = tabela[condicao1 & condicao3 & condicao4]
display(tabela11["cancelou"].value_counts(normalize=True))

#duracao_contrato + idade + dias_atraso
print("duração de contrato + idade + dias de atraso")
tabela12 = tabela[condicao1 & condicao2 & condicao4]
display(tabela12["cancelou"].value_counts(normalize=True))

```

```

#duracao_contrato + idade + ligacoes_callcenter
print("duração de contrato + idade + ligações callcenter")
tabela13 = tabela[condicao1 & condicao2 & condicao3]
display(tabela13["cancelou"].value_counts(normalize=True))

#idade + ligacoes_callcenter + dias_atraso
print("idade + ligações callcenter + dias de atraso")
tabela14 = tabela[condicao2 & condicao3 & condicao4]
display(tabela14["cancelou"].value_counts(normalize=True))

#duracao_contrato + idade + ligacoes_callcenter + dias_atraso
print("duração de contrato + idade + ligações callcenter + dias de atraso")
tabela15 = tabela[condicao1 & condicao2 & condicao3 & condicao4]
display(tabela15["cancelou"].value_counts(normalize=True))

# Implementando toda essas soluções, conseguimos reduzir a taxa de cancelamento de

```

```

duração de contrato
cancelou
0.0    0.538567
1.0    0.461433
Name: proportion, dtype: float64
idade
cancelou
0.0    0.529096
1.0    0.470904
Name: proportion, dtype: float64
ligações callcenter
cancelou
0.0    0.636134
1.0    0.363866
Name: proportion, dtype: float64
dias de atraso
cancelou
0.0    0.53539
1.0    0.46461
Name: proportion, dtype: float64
duração de contrato + idade
cancelou
0.0    0.635121
1.0    0.364879
Name: proportion, dtype: float64
duração de contrato + ligações callcenter
cancelou
0.0    0.736394
1.0    0.263606
Name: proportion, dtype: float64
duração de contrato + dias de atraso
cancelou
0.0    0.641115
1.0    0.358885
Name: proportion, dtype: float64
idade + ligações callcenter

```

```
cancelou
0.0    0.7259
1.0    0.2741
Name: proportion, dtype: float64
idade + dias de atraso
cancelou
0.0    0.631519
1.0    0.368481
Name: proportion, dtype: float64
ligações callcenter + dias de atraso
cancelou
0.0    0.730375
1.0    0.269625
Name: proportion, dtype: float64
duração de contrato + ligações callcenter + dias de atraso
cancelou
0.0    0.816462
1.0    0.183538
Name: proportion, dtype: float64
duração de contrato + idade + dias de atraso
cancelou
0.0    0.729264
1.0    0.270736
Name: proportion, dtype: float64
duração de contrato + idade + ligações callcenter
cancelou
0.0    0.813026
1.0    0.186974
Name: proportion, dtype: float64
idade + ligações callcenter + dias de atraso
cancelou
0.0    0.807607
1.0    0.192393
Name: proportion, dtype: float64
duração de contrato + idade + ligações callcenter + dias de atraso
cancelou
0.0    0.877963
1.0    0.122037
Name: proportion, dtype: float64
```