Untitled

December 9, 2023

1 Python Insights - Analisando Dados com Python

1.0.1 Case - Cancelamento de Clientes

Você foi contratado por uma empresa com mais de 800 mil clientes para um projeto de Dados. Recentemente a empresa percebeu que da sua base total 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.

[1]:
```
import pandas as pd

tabela = pd.read_csv("cancelamentos.csv")
tabela = tabela.drop("CustomerID", axis=1)
display(tabela)
```

```
        idade    sexo  tempo_como_cliente  frequencia_uso  \
0        30.0  Female                39.0            14.0
1        65.0  Female                49.0             1.0
2        55.0  Female                14.0             4.0
3        58.0    Male                38.0            21.0
4        23.0    Male                32.0            20.0
…        …     …                   …               …
881661   42.0    Male                54.0            15.0
881662   25.0  Female                 8.0            13.0
881663   26.0    Male                35.0            27.0
881664   28.0    Male                55.0            14.0
881665   31.0    Male                48.0            20.0

        ligacoes_callcenter  dias_atraso assinatura duracao_contrato  \
0                       5.0         18.0   Standard           Annual
1                      10.0          8.0      Basic          Monthly
2                       6.0         18.0      Basic        Quarterly
3                       7.0          7.0   Standard          Monthly
4                       5.0          8.0      Basic          Monthly
…                      …           …         …                …
881661                  1.0          3.0    Premium           Annual
881662                  1.0         20.0    Premium           Annual
```

```
881663                  1.0          5.0   Standard        Quarterly
881664                  2.0          0.0   Standard        Quarterly
881665                  1.0         14.0    Premium        Quarterly

        total_gasto  meses_ultima_interacao  cancelou
0            932.00                    17.0       1.0
1            557.00                     6.0       1.0
2            185.00                     3.0       1.0
3            396.00                    29.0       1.0
4            617.00                    20.0       1.0
…               …                       …         …
881661       716.38                     8.0       0.0
881662       745.38                     2.0       0.0
881663       977.31                     9.0       0.0
881664       602.55                     2.0       0.0
881665       567.77                    21.0       0.0

[881666 rows x 11 columns]
```

[4]:
```
# Indentificando e removendo valores vazios
display(tabela.info())
tabela = tabela.dropna()
display(tabela.info)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 881666 entries, 0 to 881665
Data columns (total 11 columns):
 #   Column                  Non-Null Count   Dtype
---  ------                  --------------   -----
 0   idade                   881664 non-null  float64
 1   sexo                    881664 non-null  object
 2   tempo_como_cliente      881663 non-null  float64
 3   frequencia_uso          881663 non-null  float64
 4   ligacoes_callcenter     881664 non-null  float64
 5   dias_atraso             881664 non-null  float64
 6   assinatura              881661 non-null  object
 7   duracao_contrato        881663 non-null  object
 8   total_gasto             881664 non-null  float64
 9   meses_ultima_interacao  881664 non-null  float64
 10  cancelou                881664 non-null  float64
dtypes: float64(8), object(3)
memory usage: 74.0+ MB

None

<bound method DataFrame.info of         idade    sexo  tempo_como_cliente ␣
 ↪frequencia_uso  \
0        30.0  Female                39.0            14.0
1        65.0  Female                49.0             1.0
```

```
2         55.0  Female                14.0             4.0
3         58.0    Male                38.0            21.0
4         23.0    Male                32.0            20.0
…          …     …                   …                …
881661    42.0    Male                54.0            15.0
881662    25.0  Female                 8.0            13.0
881663    26.0    Male                35.0            27.0
881664    28.0    Male                55.0            14.0
881665    31.0    Male                48.0            20.0

        ligacoes_callcenter  dias_atraso assinatura duracao_contrato  \
0                       5.0         18.0   Standard           Annual
1                      10.0          8.0      Basic          Monthly
2                       6.0         18.0      Basic        Quarterly
3                       7.0          7.0   Standard          Monthly
4                       5.0          8.0      Basic          Monthly
…                       …           …         …                …
881661                  1.0          3.0    Premium           Annual
881662                  1.0         20.0    Premium           Annual
881663                  1.0          5.0   Standard        Quarterly
881664                  2.0          0.0   Standard        Quarterly
881665                  1.0         14.0    Premium        Quarterly

        total_gasto  meses_ultima_interacao  cancelou
0            932.00                    17.0       1.0
1            557.00                     6.0       1.0
2            185.00                     3.0       1.0
3            396.00                    29.0       1.0
4            617.00                    20.0       1.0
…             …                        …          …
881661       716.38                     8.0       0.0
881662       745.38                     2.0       0.0
881663       977.31                     9.0       0.0
881664       602.55                     2.0       0.0
881665       567.77                    21.0       0.0

[881659 rows x 11 columns]>
```

[5]:
```
# qunatas pessoas cancelaram e não cancelaram
display(tabela["cancelou"].value_counts())
display(tabela["cancelou"].value_counts(normalize=True).map("{:.1%}".format))
```

```
cancelou
1.0    499993
0.0    381666
Name: count, dtype: int64

cancelou
1.0    56.7%
```

```
0.0    43.3%
Name: proportion, dtype: object
```

[6]:
```
# Verificando o Cancelamento por Contrato
display(tabela["duracao_contrato"].value_counts(normalize=True))
display(tabela["duracao_contrato"].value_counts())
```

```
duracao_contrato
Annual       0.401964
Quarterly    0.400448
Monthly      0.197588
Name: proportion, dtype: float64
```

```
duracao_contrato
Annual       354395
Quarterly    353059
Monthly      174205
Name: count, dtype: int64
```

[7]:
```
# analisando o contrato mensal
display(tabela.groupby("duracao_contrato").mean(numeric_only=True))
# descobrimos aqui que a média de cancelamentos é 1, ou seja, praticamente␣
 ↪todos os contratos mensais cancelaram (ou todos)
```

```
                     idade  tempo_como_cliente  frequencia_uso  \
duracao_contrato
Annual           38.842165           31.446186       15.880213
Monthly          41.552407           30.538555       15.499274
Quarterly        38.830938           31.419916       15.886662

                 ligacoes_callcenter  dias_atraso  total_gasto  \
duracao_contrato
Annual                      3.263401    12.465156   651.697738
Monthly                     4.985649    15.007267   550.616435
Quarterly                   3.265245    12.460863   651.427783

                 meses_ultima_interacao  cancelou
duracao_contrato
Annual                        14.236107  0.460760
Monthly                       15.478012  1.000000
Quarterly                     14.234544  0.460255
```

[8]:
```
# então descobrimos que contrato mensal é ruim, vamos tirar ele e continuar␣
 ↪analisando
tabela = tabela[tabela["duracao_contrato"]!="Monthly"]
display(tabela)
display(tabela["cancelou"].value_counts())
display(tabela["cancelou"].value_counts(normalize=True).map("{:.1%}".format))
```

```
        idade    sexo  tempo_como_cliente  frequencia_uso  \
```

```
0         30.0  Female                39.0            14.0
2         55.0  Female                14.0             4.0
5         51.0    Male                33.0            25.0
6         58.0  Female                49.0            12.0
7         55.0  Female                37.0             8.0
…          …     …                     …               …
881661    42.0    Male                54.0            15.0
881662    25.0  Female                 8.0            13.0
881663    26.0    Male                35.0            27.0
881664    28.0    Male                55.0            14.0
881665    31.0    Male                48.0            20.0

        ligacoes_callcenter  dias_atraso assinatura duracao_contrato  \
0                       5.0         18.0   Standard           Annual
2                       6.0         18.0      Basic        Quarterly
5                       9.0         26.0    Premium           Annual
6                       3.0         16.0   Standard        Quarterly
7                       4.0         15.0    Premium           Annual
…                        …           …         …                …
881661                  1.0          3.0    Premium           Annual
881662                  1.0         20.0    Premium           Annual
881663                  1.0          5.0   Standard        Quarterly
881664                  2.0          0.0   Standard        Quarterly
881665                  1.0         14.0    Premium        Quarterly

        total_gasto  meses_ultima_interacao  cancelou
0            932.00                    17.0       1.0
2            185.00                     3.0       1.0
5            129.00                     8.0       1.0
6            821.00                    24.0       1.0
7            445.00                    30.0       1.0
…             …                         …         …
881661       716.38                     8.0       0.0
881662       745.38                     2.0       0.0
881663       977.31                     9.0       0.0
881664       602.55                     2.0       0.0
881665       567.77                    21.0       0.0

[707454 rows x 11 columns]

cancelou
0.0    381666
1.0    325788
Name: count, dtype: int64

cancelou
0.0    53.9%
1.0    46.1%
Name: proportion, dtype: object
```

[9]:
```
# chegamos agora em menos da metade de pessoas cancelando, mas ainda temos␣
 ↪muitas pessoas ai, vamos continuar analisando
display(tabela["assinatura"].value_counts(normalize=True))
display(tabela.groupby("assinatura").mean(numeric_only=True))
# vemos que assinatura é quase 1/3, 1/3, 1/3
# e que os cancelamentos são na média bem parecidos, então fica difícil tirar␣
 ↪alguma conclusão da média, vamos precisar ir mais a fundo
```

```
assinatura
Standard    0.339648
Premium     0.338138
Basic       0.322215
Name: proportion, dtype: float64
```

assinatura	idade	tempo_como_cliente	frequencia_uso	ligacoes_callcenter	dias_atraso	total_gasto	meses_ultima_interacao	cancelou
Basic	38.904813	32.316031	15.876921	3.310021	12.507054	648.642614	14.240814	0.475188
Premium	38.817814	30.977869	15.889673	3.235886	12.433427	653.337633	14.231150	0.452338
Standard	38.790478	31.048621	15.883393	3.249275	12.450690	652.566793	14.234280	0.454714

[10]:
```
# vamos criar gráfico, porque só com números tá difícil de visualizar
import plotly.express as px

for coluna in tabela.columns:
    grafico = px.histogram(tabela, x=coluna, color="cancelou")
    grafico.show()
```

[11]:
```
# com os graficos a gente consegue descobrir muita coisa:
# dias atraso acima de 20 dias, 100% cancela
# ligações call center acima de 5 todo mundo cancela

tabela = tabela[tabela["ligacoes_callcenter"]<5]
tabela = tabela[tabela["dias_atraso"]<=20]
display(tabela)
display(tabela["cancelou"].value_counts())
display(tabela["cancelou"].value_counts(normalize=True).map("{:.1%}".format))
```

```
# se resolvermos isso, já caímos para 18% de cancelamento
# é claro que 100% é utópico, mas com isso já temos as principais causas (ou talvez 3 das principais):
# - forma de contrato mensal
# - necessidade de ligações no call center
# - atraso no pagamento
```

```
        idade    sexo  tempo_como_cliente  frequencia_uso  \
6        58.0  Female                49.0            12.0
7        55.0  Female                37.0             8.0
9        64.0  Female                 3.0            25.0
13       48.0  Female                35.0            25.0
19       42.0    Male                15.0            16.0
…         …     …                   …               …
881661   42.0    Male                54.0            15.0
881662   25.0  Female                 8.0            13.0
881663   26.0    Male                35.0            27.0
881664   28.0    Male                55.0            14.0
881665   31.0    Male                48.0            20.0

        ligacoes_callcenter  dias_atraso assinatura duracao_contrato  \
6                       3.0         16.0   Standard        Quarterly
7                       4.0         15.0    Premium           Annual
9                       2.0         11.0   Standard        Quarterly
13                      1.0         13.0      Basic           Annual
19                      2.0         14.0    Premium        Quarterly
…                      …           …         …                …
881661                  1.0          3.0    Premium           Annual
881662                  1.0         20.0    Premium           Annual
881663                  1.0          5.0   Standard        Quarterly
881664                  2.0          0.0   Standard        Quarterly
881665                  1.0         14.0    Premium        Quarterly

        total_gasto  meses_ultima_interacao  cancelou
6            821.00                    24.0       1.0
7            445.00                    30.0       1.0
9            415.00                    29.0       1.0
13           518.00                    17.0       1.0
19           262.00                    16.0       1.0
…             …                       …         …
881661       716.38                     8.0       0.0
881662       745.38                     2.0       0.0
881663       977.31                     9.0       0.0
881664       602.55                     2.0       0.0
881665       567.77                    21.0       0.0

[464479 rows x 11 columns]
```

```
cancelou
0.0    379032
1.0     85447
Name: count, dtype: int64
```

```
cancelou
0.0    81.6%
1.0    18.4%
Name: proportion, dtype: object
```

2 Conclusão análise gráfica

Nós começamos o problema com um taxa de cancelamento de 56,7%. Após o primeiro tratamento conseguimos diminuir um pouco e atingimos 46,1%. No final com ajuda dos gráficos conseguimos ajustar nossa base de dados e chegamos ao percentual de 18,4% em cancelamentos.