

## ✓ Desafio:

Você trabalha em uma grande empresa de Cartão de Crédito e o diretor da empresa percebeu que o número de clientes que cancelam seus cartões tem aumentado significativamente, causando prejuízos enormes para a empresa


O que fazer para evitar isso? Como saber as pessoas que têm maior tendência a cancelar o cartão?

## O que temos:

Temos 1 base de dados com informações dos clientes, tanto clientes atuais quanto clientes que cancelaram o cartão

Referência: <https://www.kaggle.com/sakshigoyal7/credit-card-customers>

```
from google.colab import drive
drive.mount('/content/drive')
```

 Mounted at /content/drive

## ✓ Logica do Programa

- Passo 1: Importar a base de dados;
- Passo 2: Visualizar e tratar essa base de dados;
- Passo 3: Analisar a sua base de dados;
- Passo 4: Construir uma análise para identificar o motivo de cancelamento;
- Passo 5: Identificar qual o motivo ou os principais motivos dos clientes cancelarem o cartão;

```
import pandas as pd
tabela = pd.read_csv("/content/ClientesBanco.csv", encoding="latin1")
```

```
tabela= tabela.drop("CLIENTNUM", axis=1)
display(tabela)
```



	Categoria	Idade	Sexo	Dependentes	Educação	Estado Civil	Faixa Salarial Anual	Categoria Cartão	Meses como Cliente	Produtos Contratados	Inatividade 12m	Conta
0	Cliente	45	M	3	Ensino Médio	Casado	\$60K - \$80K	Blue	39	5	1	
1	Cliente	49	F	5	Ensino Superior	Solteiro	Less than \$40K	Blue	44	6	1	
2	Cliente	51	M	3	Ensino Superior	Casado	\$80K - \$120K	Blue	36	4	1	
3	Cliente	40	F	4	Ensino Médio	Não informado	Less than \$40K	Blue	34	3	4	
4	Cliente	40	M	3	Sem ensino formal	Casado	\$60K - \$80K	Blue	21	5	1	
...	...	...	...	...	...	...	...	...	...	...	...	...
10122	Cliente	50	M	2	Ensino Superior	Solteiro	\$40K - \$60K	Blue	40	3	2	
10123	Cancelado	41	M	2	Não informado	Divorciado	\$40K - \$60K	Blue	25	4	2	
10124	Cancelado	44	F	1	Ensino Médio	Casado	Less than \$40K	Blue	36	5	3	
10125	Cancelado	30	M	2	Ensino Superior	Não informado	\$40K - \$60K	Blue	36	4	3	
10126	Cancelado	43	F	2	Ensino Superior	Casado	Less than \$40K	Silver	25	6	2	

10127 rows × 20 columns

✓ Agora vamos tratar valores vazios e exibir um resumo das colunas da base de dados

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

```
display(tabela.describe().round(1))
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
Index: 10126 entries, 0 to 10126
```

```
Data columns (total 20 columns):
```

#	Column	Non-Null Count	Dtype
0	Categoria	10126 non-null	object
1	Idade	10126 non-null	int64
2	Sexo	10126 non-null	object
3	Dependentes	10126 non-null	int64
4	Educação	10126 non-null	object
5	Estado Civil	10126 non-null	object
6	Faixa Salarial Anual	10126 non-null	object
7	Categoria Cartão	10126 non-null	object
8	Meses como Cliente	10126 non-null	int64
9	Produtos Contratados	10126 non-null	int64
10	Inatividade 12m	10126 non-null	int64
11	Contatos 12m	10126 non-null	int64
12	Limite	10126 non-null	float64
13	Limite Consumido	10126 non-null	int64
14	Limite Disponível	10126 non-null	float64
15	Mudanças Transacoes_Q4_Q1	10126 non-null	float64
16	Valor Transacoes 12m	10126 non-null	int64
17	Qtde Transacoes 12m	10126 non-null	int64
18	Mudança Qtde Transações_Q4_Q1	10126 non-null	float64
19	Taxa de Utilização Cartão	10126 non-null	float64

```
dtypes: float64(5), int64(9), object(6)
```

```
memory usage: 1.6+ MB
```

```
None
```

	Idade	Dependentes	Meses como Cliente	Produtos Contratados	Inatividade 12m	Contatos 12m	Limite	Limite Consumido	Limite Disponível	Mudanças Transacoes_Q4_Q1	Trê
<b>count</b>	10126.0	10126.0	10126.0	10126.0	10126.0	10126.0	10126.0	10126.0	10126.0	10126.0	
<b>mean</b>	46.3	2.3	35.9	3.8	2.3	2.5	8632.0	1162.8	7469.2	108.7	
<b>std</b>	8.0	1.3	8.0	1.6	1.0	1.1	9089.2	815.0	9091.1	353.9	
<b>min</b>	26.0	0.0	13.0	1.0	0.0	0.0	1438.3	0.0	3.0	0.0	
<b>25%</b>	41.0	1.0	31.0	3.0	2.0	2.0	2555.0	358.0	1324.2	0.6	
<b>50%</b>	46.0	2.0	36.0	4.0	2.0	2.0	4549.0	1276.0	3473.0	0.7	
<b>75%</b>	52.0	3.0	40.0	5.0	3.0	3.0	11067.8	1784.0	9860.0	0.9	
<b>max</b>	73.0	5.0	56.0	6.0	6.0	6.0	34516.0	2517.0	34516.0	3397.0	

✓ Vamos avaliar como está a divisão entre Clientes x Cancelados

```
qtde_categoria = tabela["Categoria"].value_counts()
display(qtde_categoria)

qtde_categoria_perc = tabela["Categoria"].value_counts(normalize=True)
display(qtde_categoria_perc)
```

```
count
Categoria
Cliente    8499
Cancelado  1627

dtype: int64

proportion
```

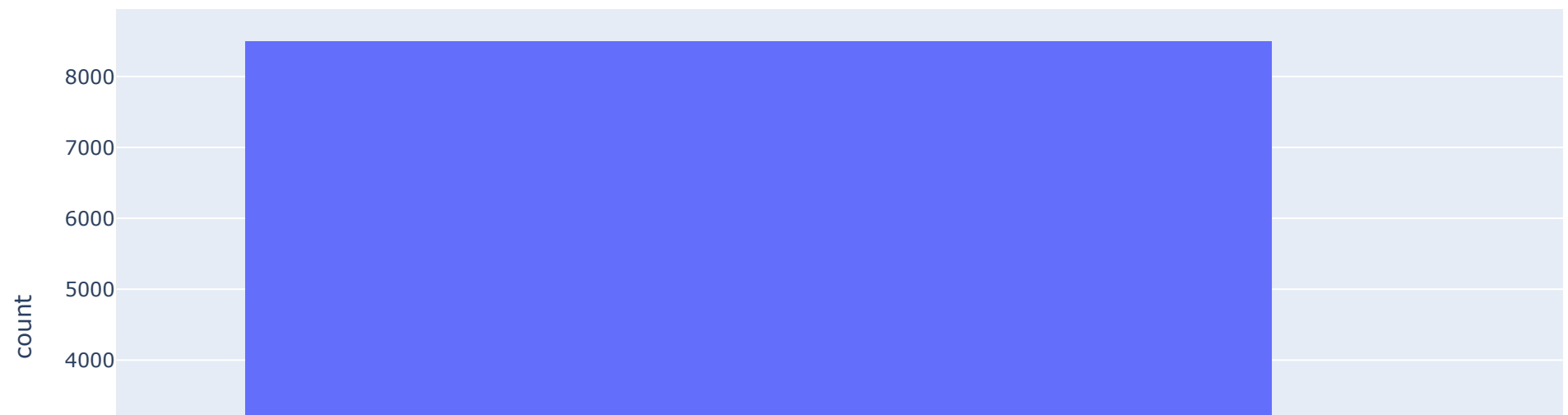
Categoria	
Cliente	0.839325
Cancelado	0.160675

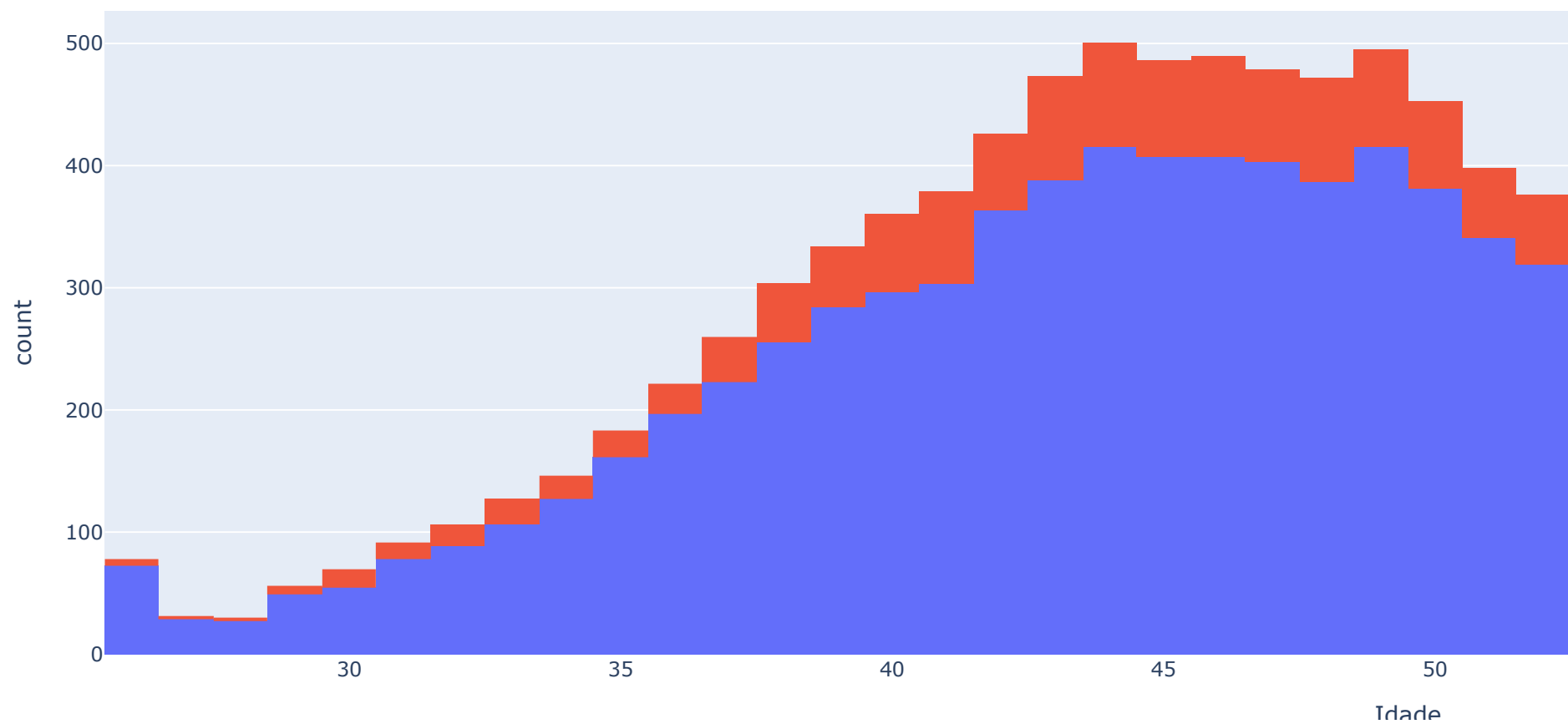
**dtype:** float64

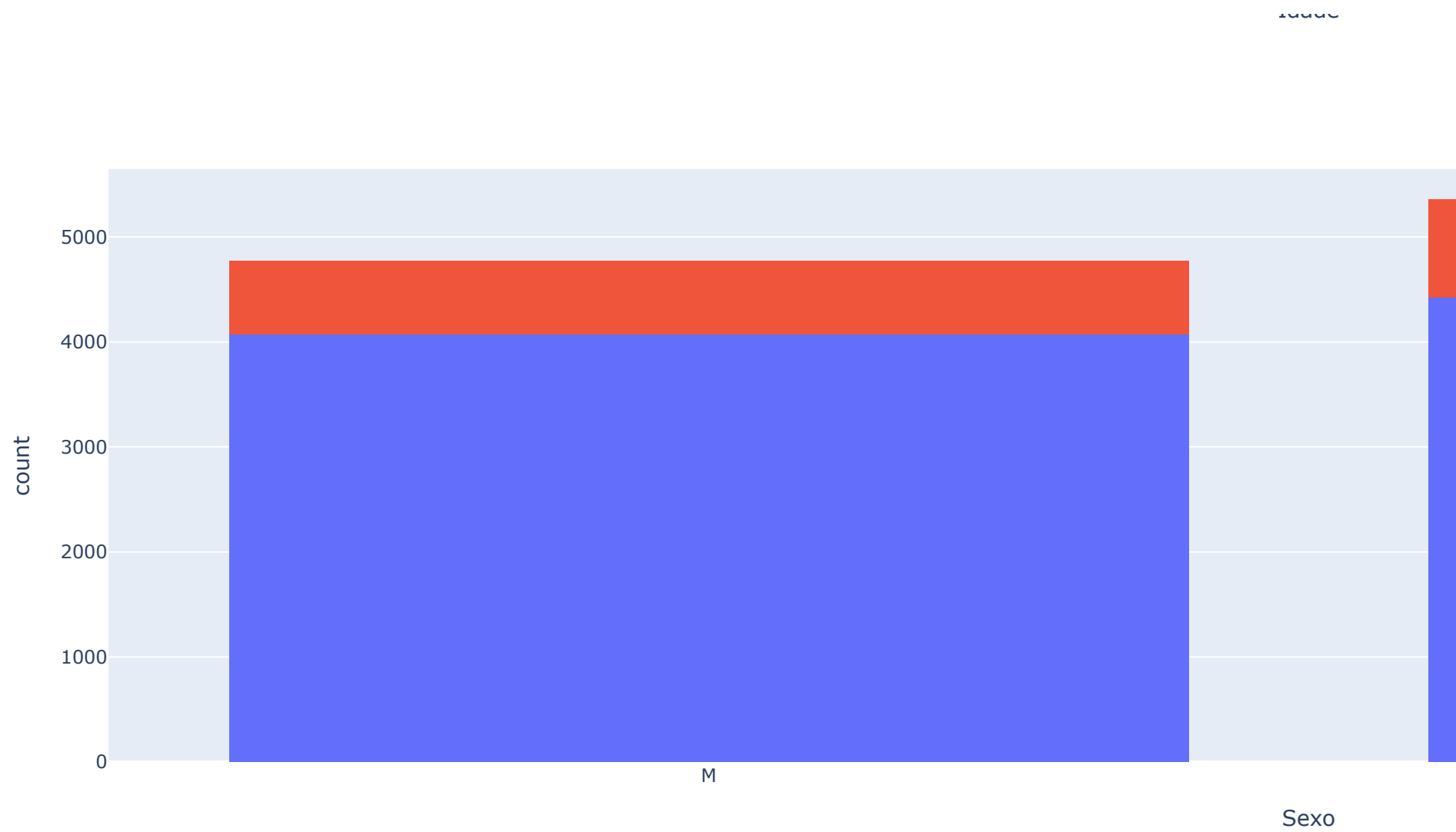
- ✓ Podemos olhar a comparação entre Clientes e Cancelados em cada uma das colunas da nossa base de dados, para ver se essa informações traz algum insight novo para a gente

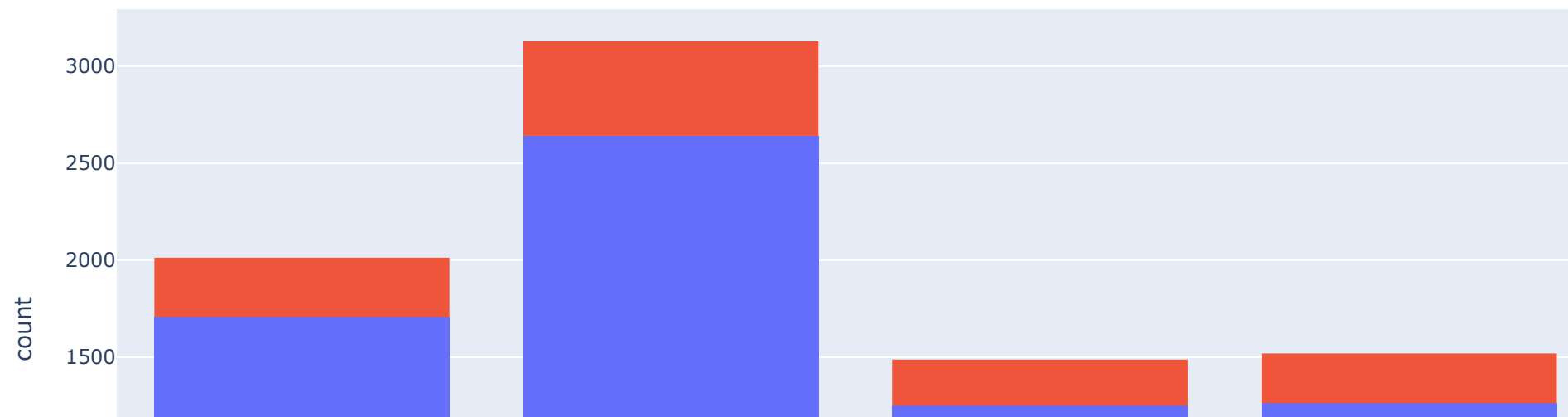
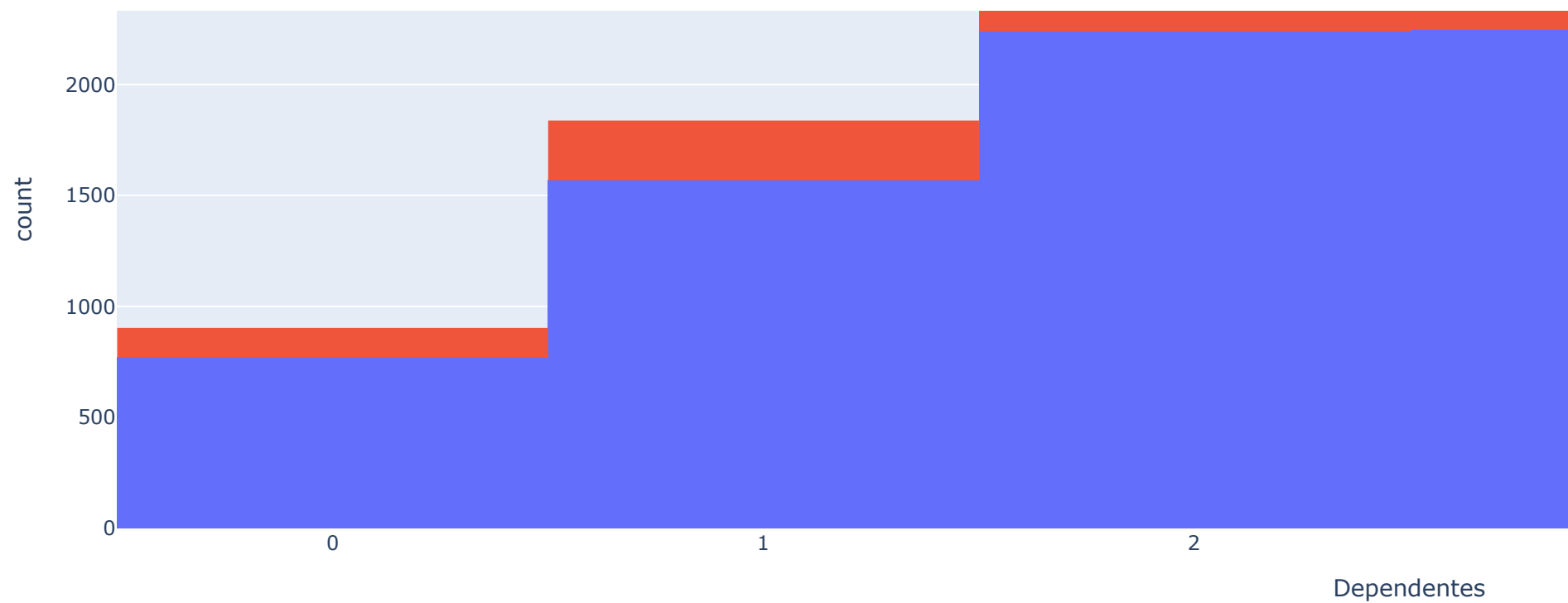
```
import plotly.express as px

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

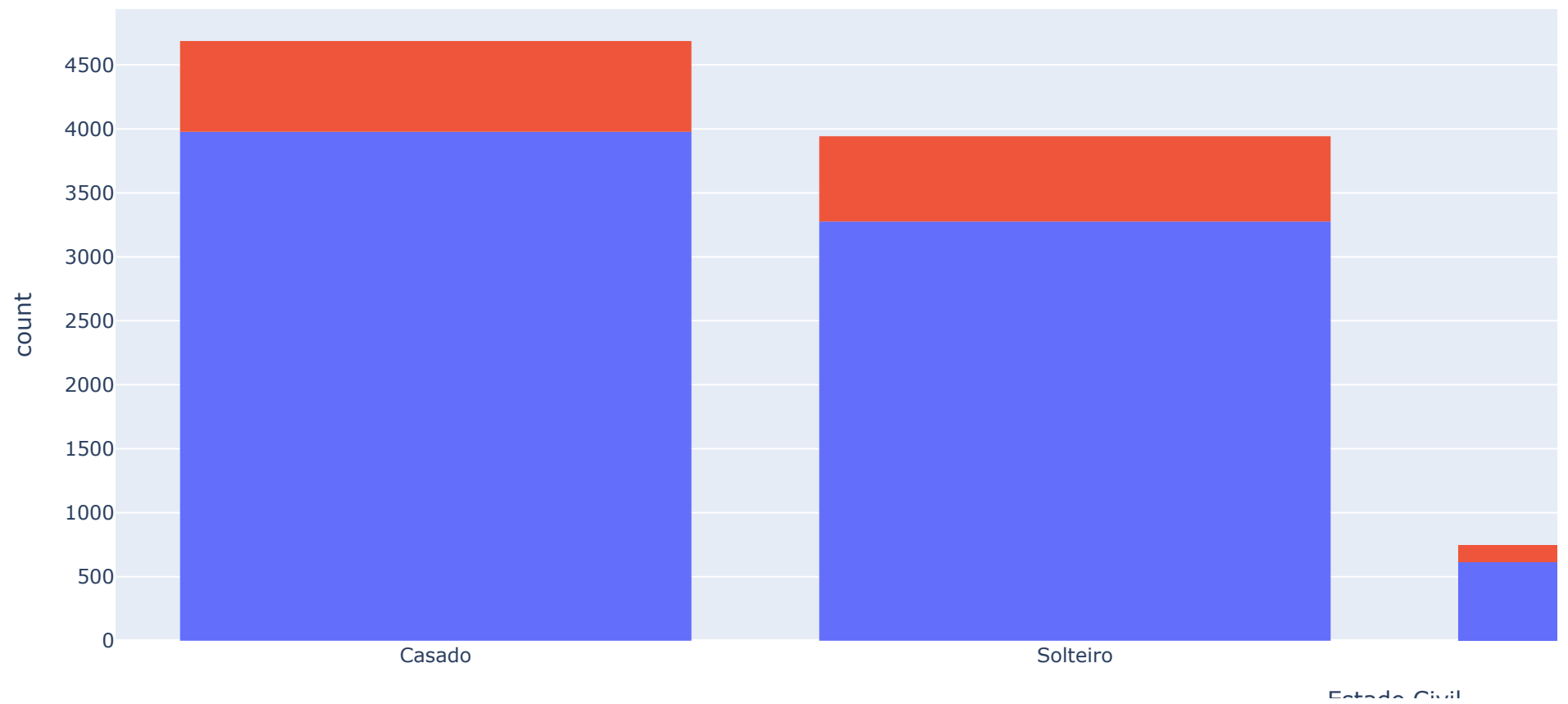
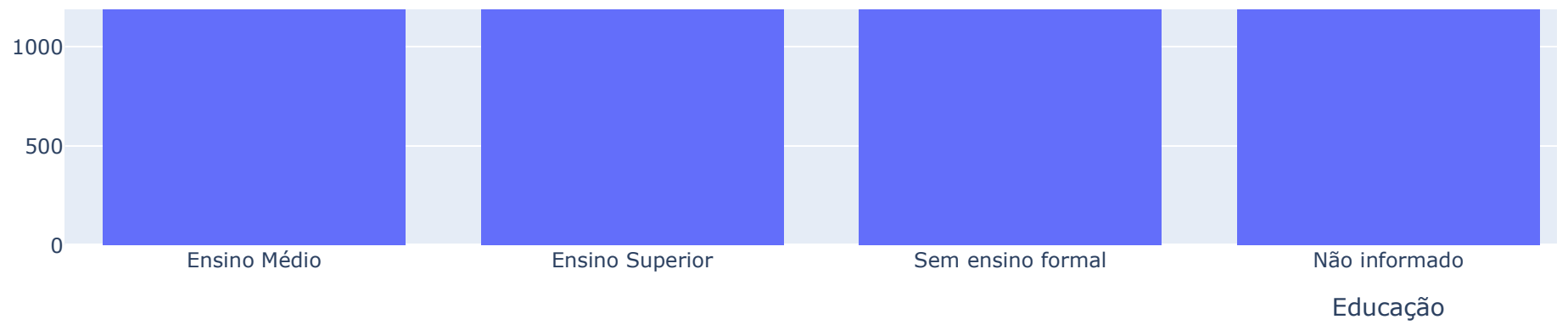


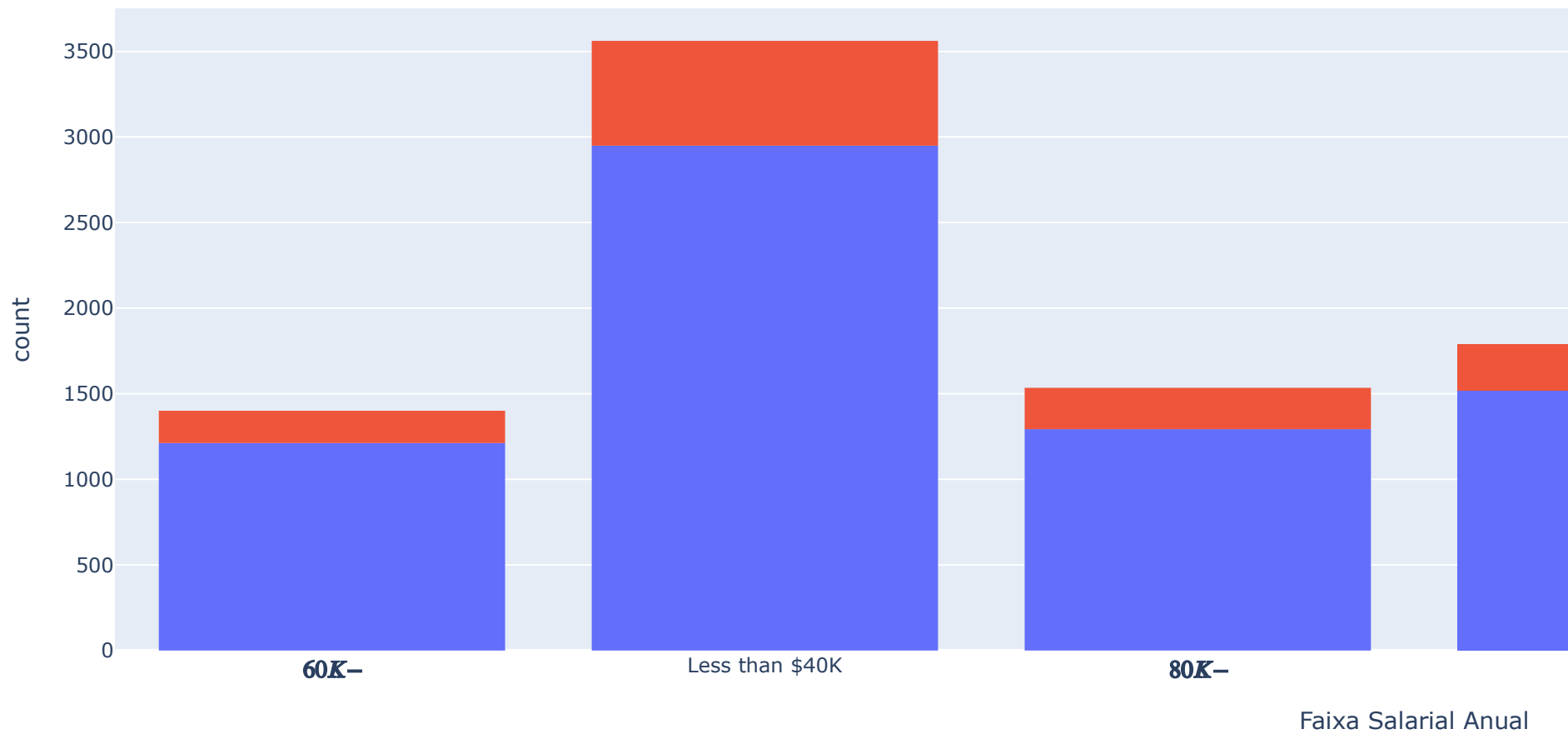


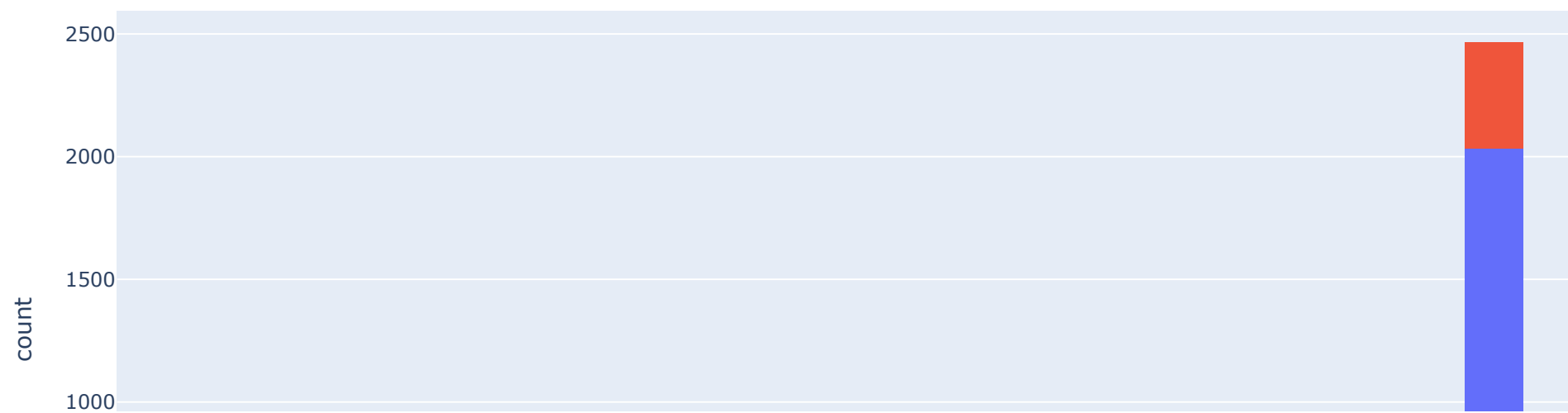
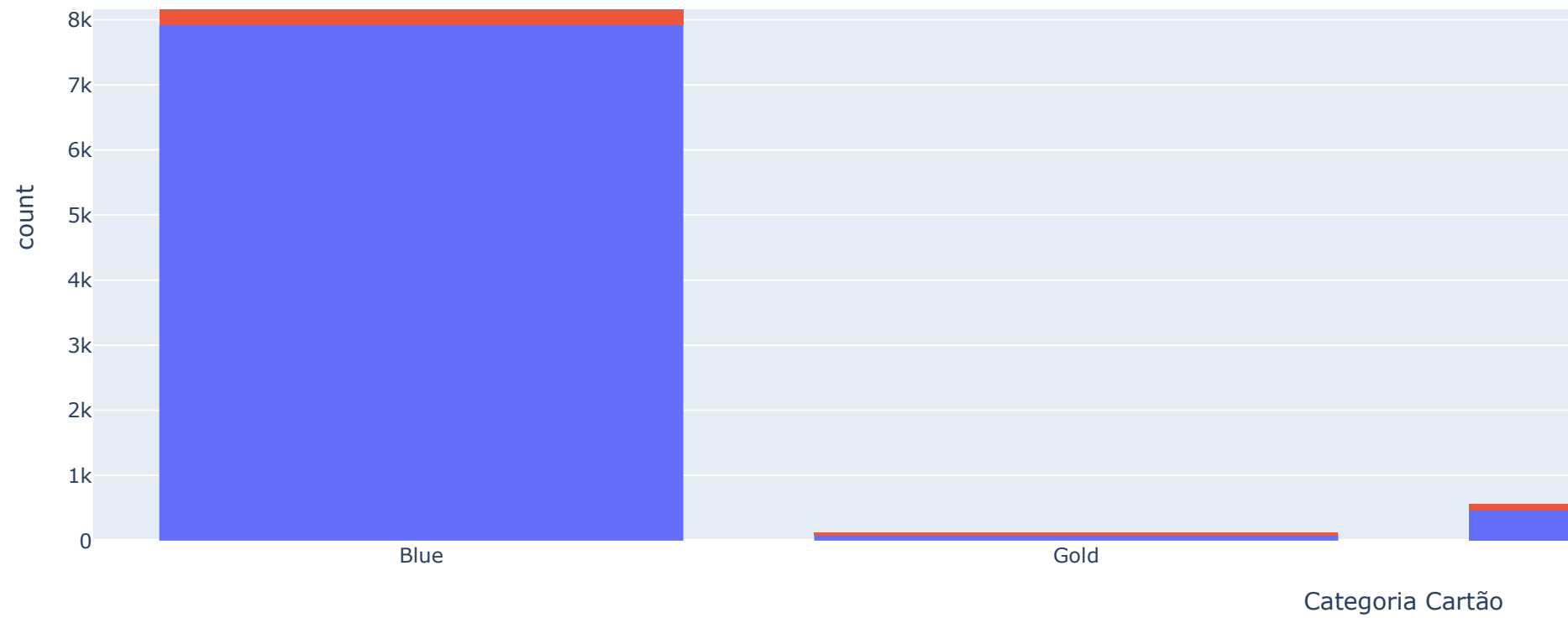


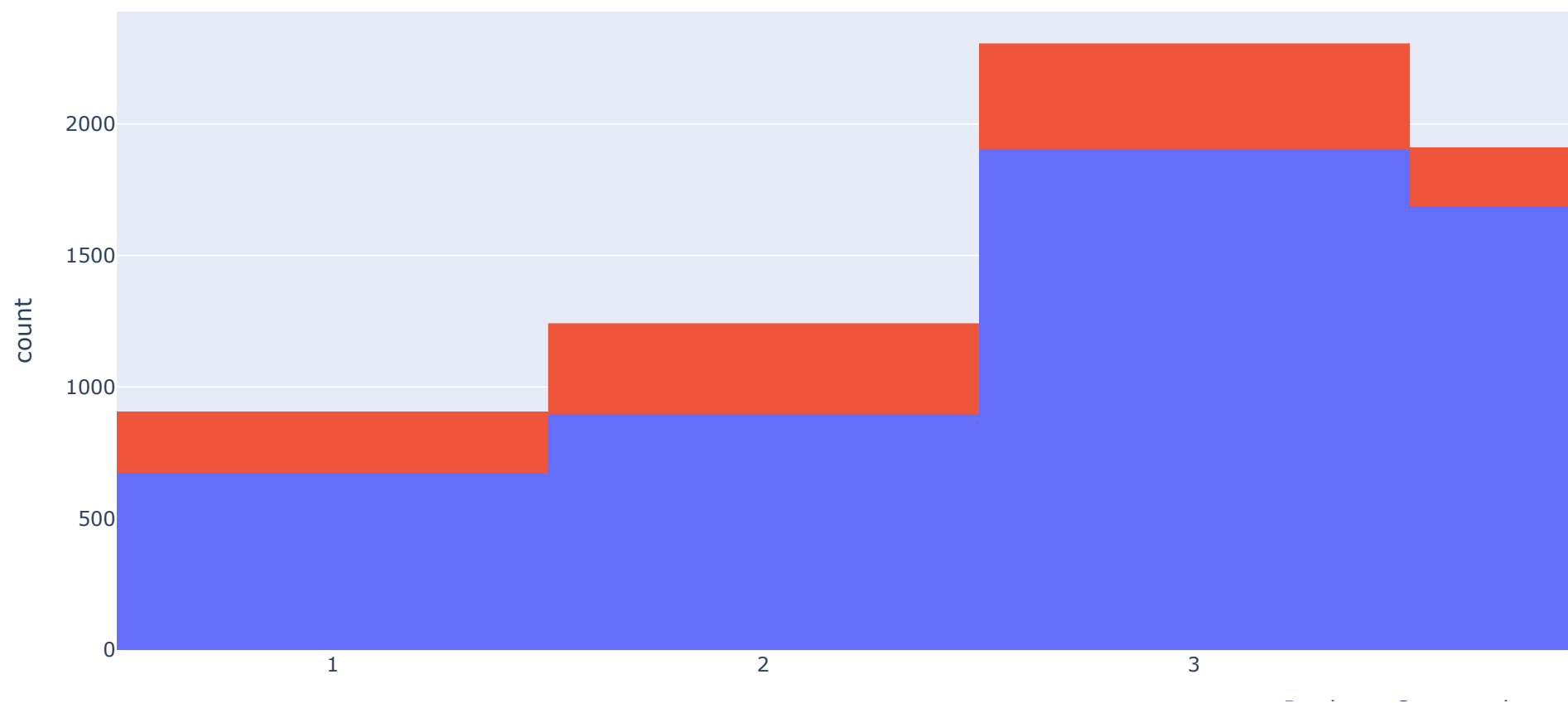
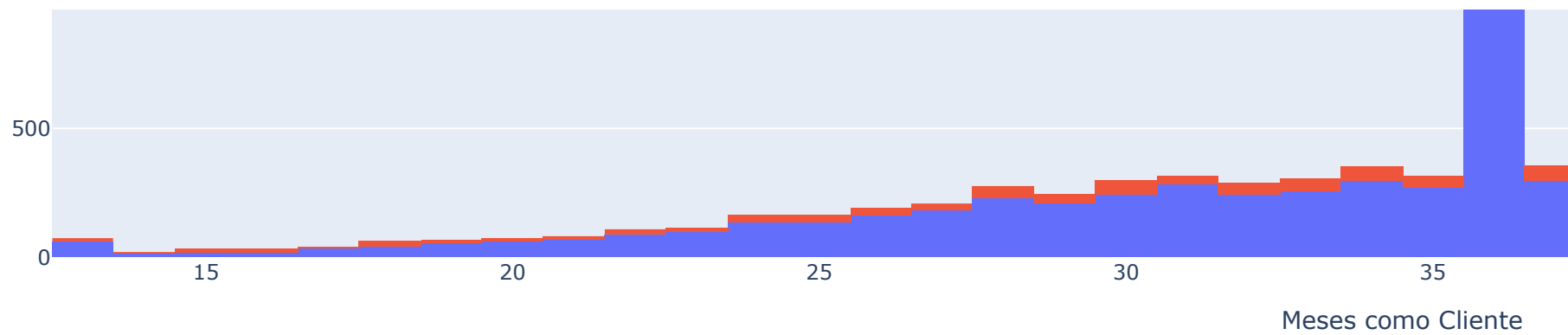




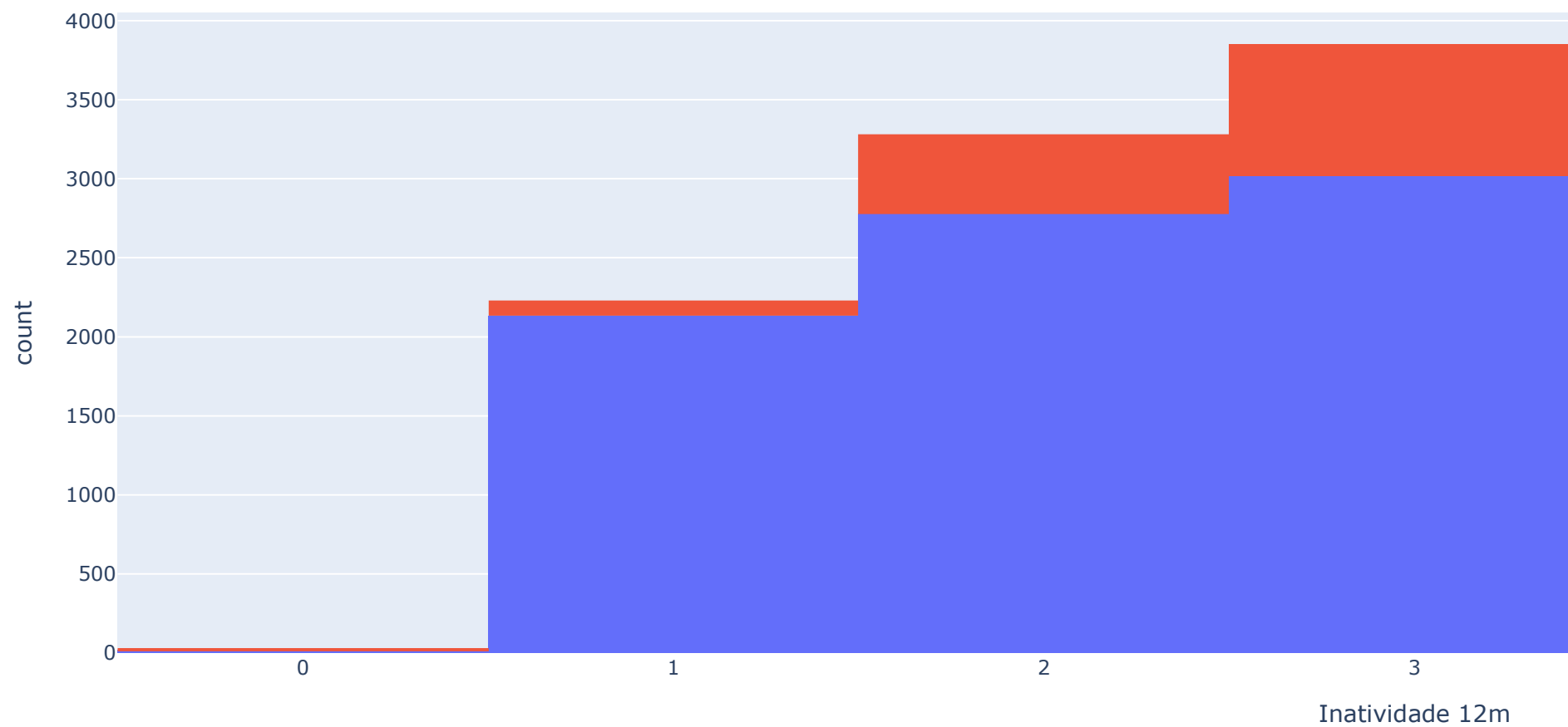


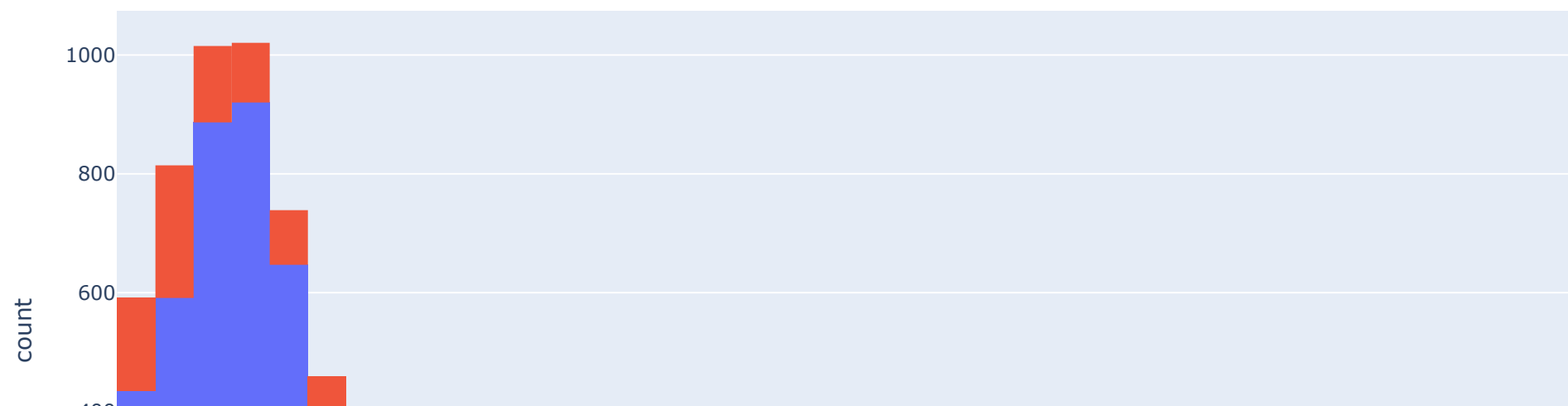
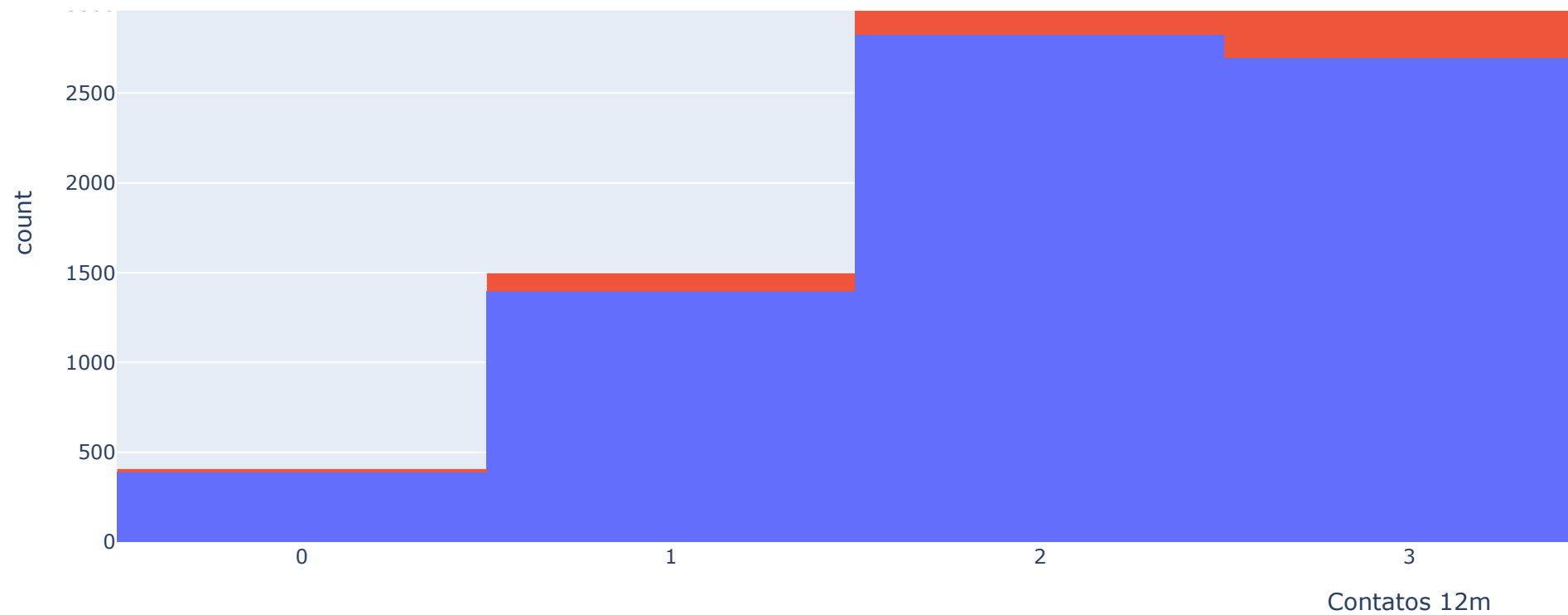


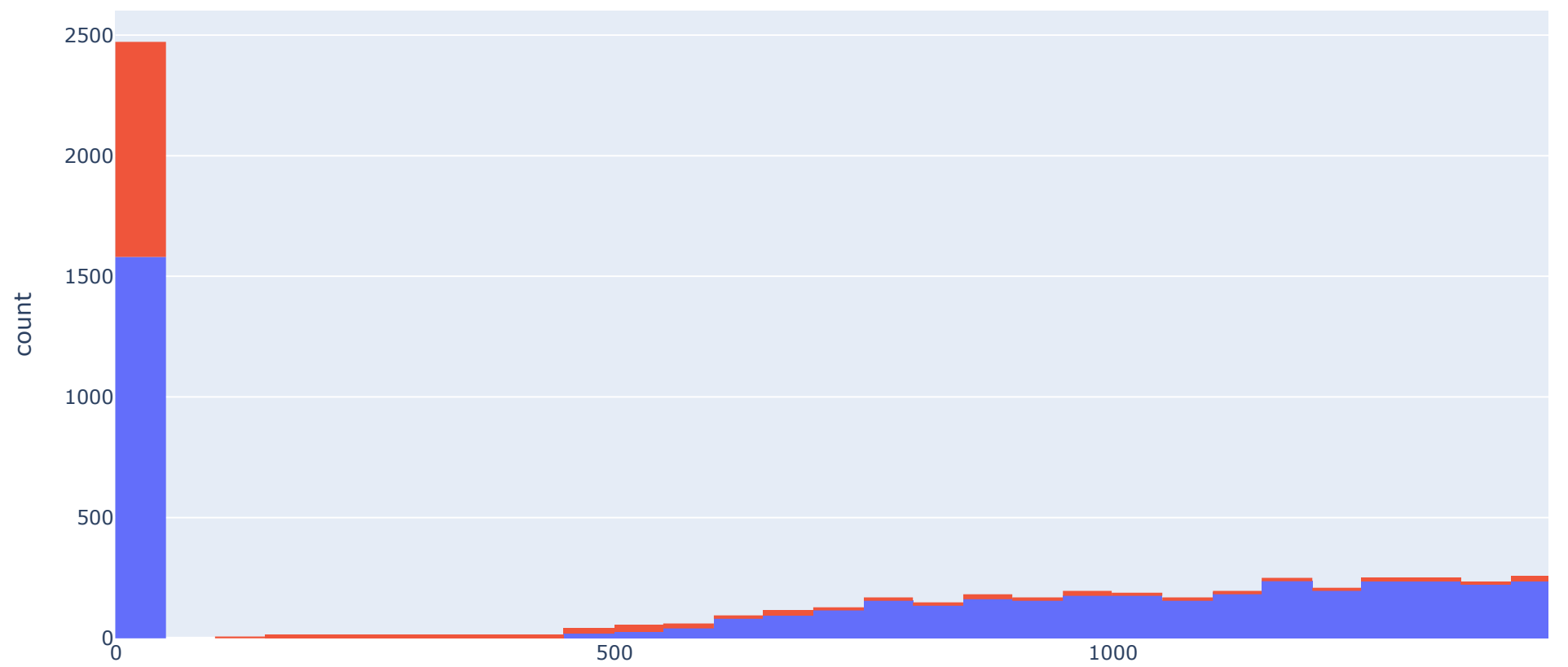
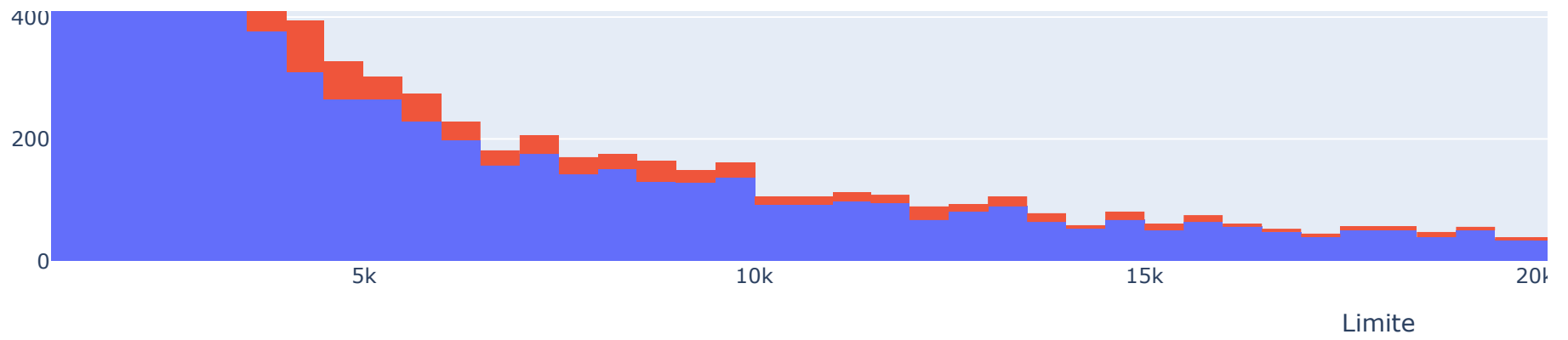




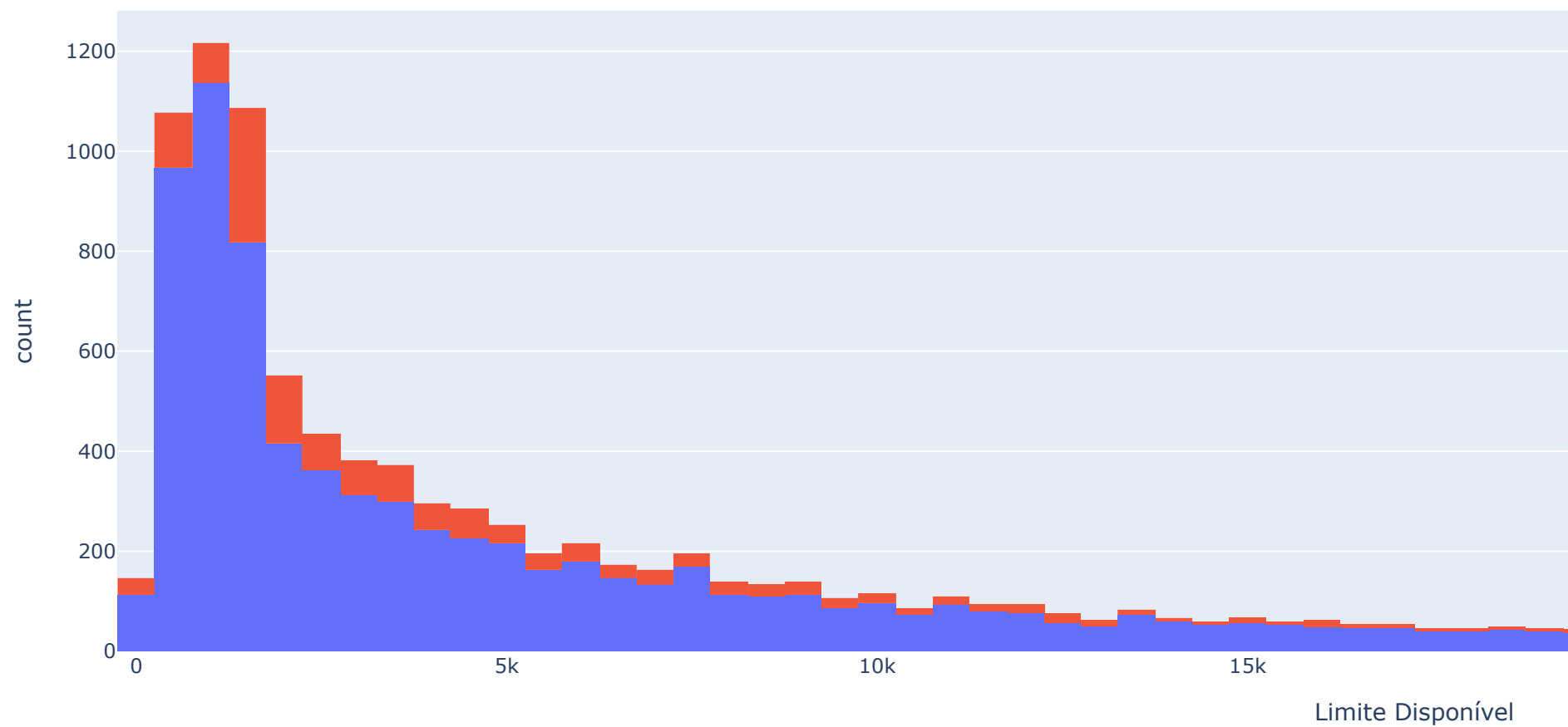
## Produtos Contratados



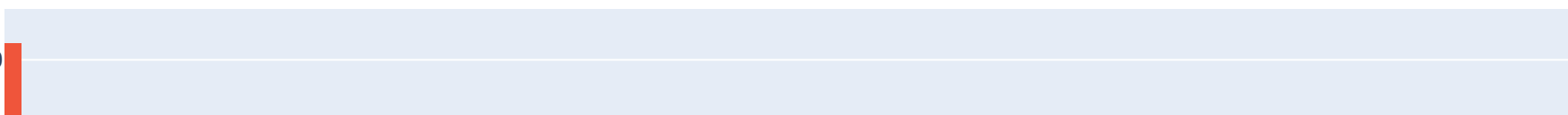




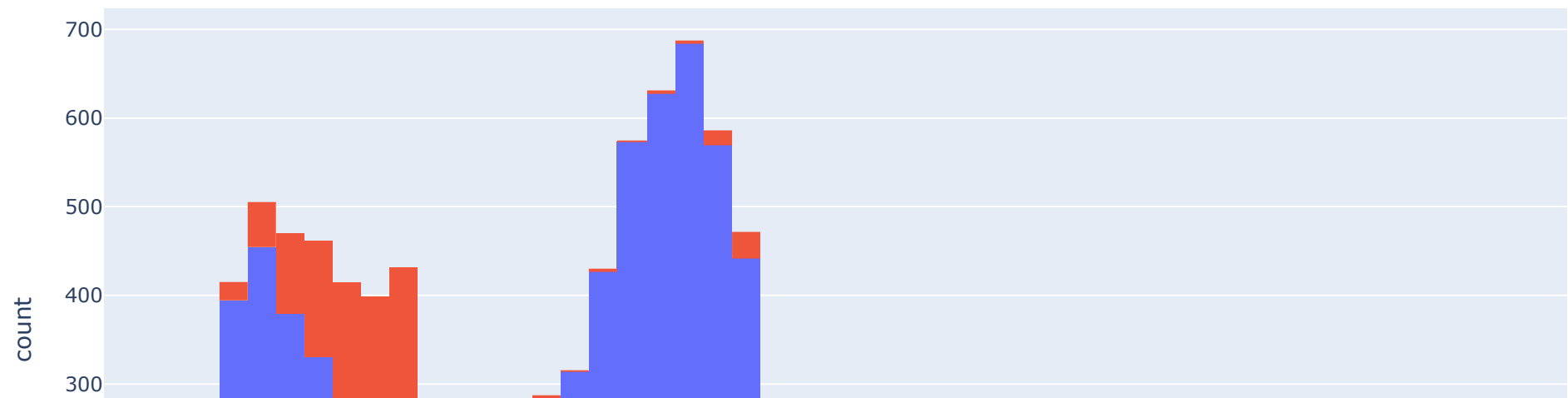
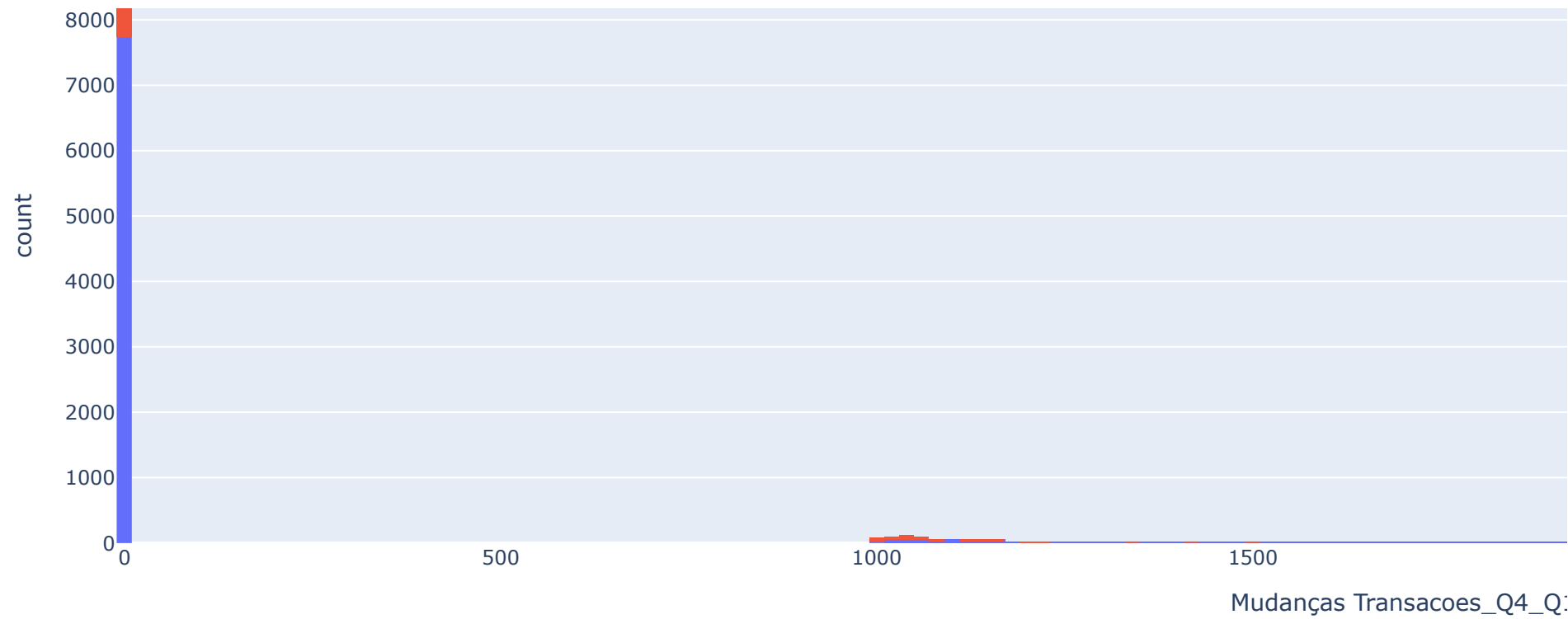
Limite Consumido

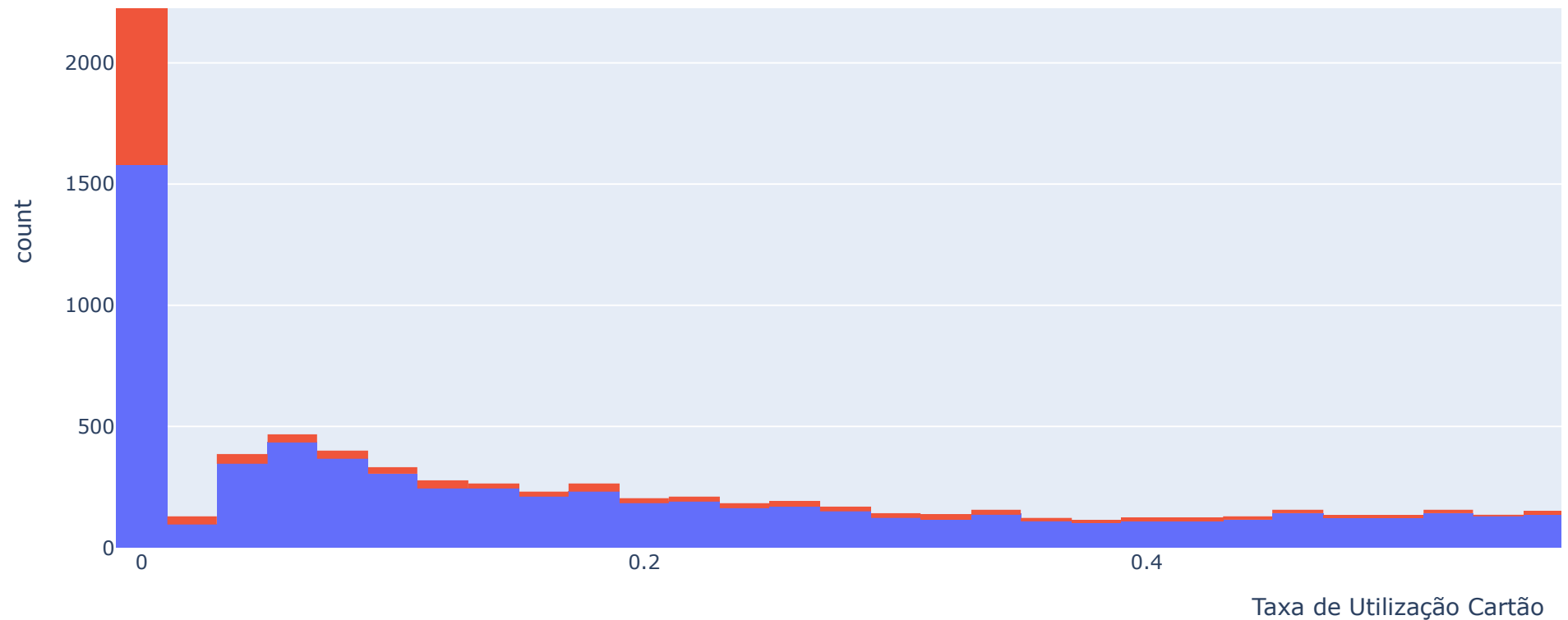


9000









## Informações retiradas da análise

- Me parece que quanto mais produtos contratados um cliente tem, menor a chance dele cancelar.
- E quanto mais transações e quanto maior o valor de transações, menor a chance dele cancelar.
- Quanto maior a quantidade de contatos que a pessoa teve que fazer, maior a chance dela cancelar.