



escola  
britânica de  
artes criativas  
& tecnologia

# Python para análise de dados



# ANÁLISE DE DADOS



## GUIA DA AULA 4



# Visualize dados

- Definição
- Visualizações categóricas
- Visualizações numéricas



Acompanhe aqui  
os temas que  
serão tratados  
na videoaula



## 1. Definição

Os dados estão prontos, vamos criar diversas visualizações para correlacionar variáveis explicativas com a variável resposta para buscar entender qual fator leva um cliente a inadimplência. E para isso, vamos sempre comparar a base com todos os clientes com a base de adimplentes e inadimplentes. Começamos, então, importando os pacotes de visualização e separando os clientes adimplentes e inadimplentes

```
In [ ]: import seaborn as sns
import matplotlib.pyplot as plt
sns.set_style("whitegrid")
```

```
In [ ]: df_adimplente = df[df['default'] == 0]
```

```
In [ ]: df_inadimplente = df[df['default'] == 1]
```



## 2. Visualizações categóricas

Nesta seção, vamos visualizar a relação entre a variável resposta *default* com os atributos categóricos.

In [ ]:

```
df.select_dtypes('object').head(n=5)
```

### Escolaridade

In [ ]:

```
coluna = 'escolaridade' titulos =  
[  
    'Escolaridade dos Clientes', 'Escolaridade dos Clientes  
    Adimplentes', 'Escolaridade dos Clientes Inadimplentes'  
]  
  
eixo = 0  
max_y = 0  
max = df.select_dtypes('object').describe()[coluna]['freq'] * 1.1  
  
figura, eixos = plt.subplots(1,3, figsize=(20, 5), sharex=True)
```



## 2. Visualizações categóricas

```
for dataframe in [df, df_adimplente, df_inadimplente]:

    df_to_plot = dataframe[coluna].value_counts().to_frame() df_to_plot.rename(columns={coluna:
    'frequencia_absoluta'}, inplace=True) df_to_plot[coluna] = df_to_plot.index
    df_to_plot.sort_values(by=[coluna], inplace=True) df_to_plot.sort_values(by=[coluna])

    f = sns.barplot(
        x=df_to_plot[coluna],
        y=df_to_plot['frequencia_absoluta'], ax=eixos[eixo]
    )
    f.set(
        title=titulos[eixo],
        xlabel=coluna.capitalize(),
        ylabel='Frequência Absoluta'
    )
    f.set_xticklabels(labels=f.get_xticklabels(), rotation=90)

    _, max_y_f = f.get_ylim()
    max_y = max_y_f if max_y_f > max_y else max_y
    f.set_ylim=(0, max_y))

    eixo += 1

figura.show()
```



## 2. Visualizações categóricas

### Salário anual

```
In [ ]: coluna = 'salario_anual' titulos = [  
        'Salário Anual dos Clientes',  
        'Salário Anual dos Clientes Adimplentes', 'Salário Anual  
        dos Clientes Inadimplentes'  
    ]  
  
    eixo = 0  
    max_y = 0  
    figura, eixos = plt.subplots(1,3, figsize=(20, 5), sharex=True)  
  
    for dataframe in [df, df_adimplente, df_inadimplente]: df_to_plot =  
  
        dataframe[coluna].value_counts().to_frame()  
        df_to_plot.rename(columns={coluna: 'frequencia_absoluta'}, inplace=True) df_to_plot[coluna] =  
        df_to_plot.index df_to_plot.reset_index(inplace=True, drop=True)  
        df_to_plot.sort_values(by=[coluna], inplace=True)  
  
    f = sns.barplot(  
        x=df_to_plot[coluna],  
        y=df_to_plot['frequencia_absoluta'], ax=eixos[eixo]
```



## 2. Visualizações categóricas

### Salário anual

```
)  
f.set(  
    title=titulos[eixo],  
    xlabel=coluna.capitalize(),  
    ylabel='Frequência Absoluta'  
)  
f.set_xticklabels(labels=f.get_xticklabels(), rotation=90)  
_, max_y_f = f.get_ylim()  
max_y = max_y_f if max_y_f > max_y else max_y  
f.set(ylim=(0, max_y))  
  
eixo += 1  
  
figura.show()
```





### 3. Visualizações numéricas

Nesta seção, vamos visualizar a relação entre a variável resposta default com os atributos numéricos.

In [ ]:

```
df.drop(['id', 'default'], axis=1).select_dtypes('number').head(n=5)
```

#### Quantidade de Transações nos Últimos 12 Meses

In [ ]:

```
coluna = 'qtd_transacoes_12m' titulos =  
[  
    'Qtd. de Transações no Último Ano',  
    'Qtd. de Transações no Último Ano de Adimplentes', 'Qtd. de  
    Transações no Último Ano de Inadimplentes'  
]
```



### 3. Visualizações numéricas

```
eixo = 0
max_y = 0
figura, eixos = plt.subplots(1,3, figsize=(20, 5), sharex=True)

for dataframe in [df, df_adimplente, df_inadimplente]:

    f = sns.histplot(x=coluna, data=dataframe, stat='count', ax=eixos[eixo]) f.set(
        title=titulos[eixo],
        xlabel=coluna.capitalize(),
        ylabel='Frequência Absoluta'
    )

    _, max_y_f = f.get_ylim()
    max_y = max_y_f if max_y_f > max_y else max_y
    f.set(ylim=(0, max_y))

    eixo += 1

figura.show()
```



### 3. Visualizações numéricas

```
        xlabel=coluna.capitalize(),  
        ylabel='Frequência Absoluta'  
    )  
  
    _, max_y_f = f.get_ylim()  
    max_y = max_y_f if max_y_f > max_y else max_y  
    f.set_ylim(0, max_y)  
  
    eixo += 1  
  
figura.show()
```



### 3. Visualizações numéricas

Valor de Transações nos Últimos 12 Meses x Quantidade de Transações nos Últimos 12 Meses

In [ ]:

```
f = sns.relplot(
    x='valor_transacoes_12m',
    y='qtd_transacoes_12m', data=df,
    hue='default'
)

_ = f.set(
    title='Relação entre Valor e Quantidade de Transações no Último Ano', xlabel='Valor das
    Transações no Último Ano',
    ylabel='Quantidade das Transações no Último Ano'
)
```

