A célula já está em execução

Untitled8.ipynb\_

Arquivo

Editar

Ver

Inserir

Ambiente de execução

Ferramentas

Ajuda

ComentárioCompartilhar

[](https://accounts.google.com/SignOutOptions?hl=pt-BR&continue=https://colab.research.google.com/drive/1J2uxulKrwX6ZpzmDyeyswtFIUVSnPCd-)

Arquivos

..

sample\_data

Solte arquivos aqui para fazer upload no armazenamento da sessão

Disco

Disponível: 70.74 GB

CódigoTexto

RAM

Disco

Editar

[134]

0s

import pandas as pd



[135]

clientes = pd.read\_excel('/content/db\_clientes.xlsx')

estados = pd.read\_csv('/content/db\_estados\_brasileiros.csv', sep=';', encoding='latin-1')

idade = pd.read\_csv('/content/db\_idade\_clientes.csv', sep=';', encoding='latin-1')



[136]

4s

!pip install dataprep

from dataprep.eda import create\_report



Requirement already satisfied: dataprep in /usr/local/lib/python3.7/dist-packages (0.3.0)

Requirement already satisfied: scipy<2,>=1 in /usr/local/lib/python3.7/dist-packages (from dataprep) (1.4.1)

Requirement already satisfied: wordcloud<2.0,>=1.8 in /usr/local/lib/python3.7/dist-packages (from dataprep) (1.8.1)

Requirement already satisfied: aiohttp<4.0,>=3.6 in /usr/local/lib/python3.7/dist-packages (from dataprep) (3.7.4.post0)

Requirement already satisfied: dask[array,dataframe,delayed]<3.0,>=2.25 in /usr/local/lib/python3.7/dist-packages (from dataprep) (2.30.0)

Requirement already satisfied: numpy<2,>=1 in /usr/local/lib/python3.7/dist-packages (from dataprep) (1.19.5)

Requirement already satisfied: tqdm<5.0,>=4.48 in /usr/local/lib/python3.7/dist-packages (from dataprep) (4.62.0)

Requirement already satisfied: jsonpath-ng<2.0,>=1.5 in /usr/local/lib/python3.7/dist-packages (from dataprep) (1.5.3)

Requirement already satisfied: python-Levenshtein<0.13.0,>=0.12.2 in /usr/local/lib/python3.7/dist-packages (from dataprep) (0.12.2)

Requirement already satisfied: usaddress<0.6.0,>=0.5.10 in /usr/local/lib/python3.7/dist-packages (from dataprep) (0.5.10)

Requirement already satisfied: metaphone<0.7,>=0.6 in /usr/local/lib/python3.7/dist-packages (from dataprep) (0.6)

Requirement already satisfied: jinja2<3.0,>=2.11 in /usr/local/lib/python3.7/dist-packages (from dataprep) (2.11.3)

Requirement already satisfied: bottleneck<2.0,>=1.3 in /usr/local/lib/python3.7/dist-packages (from dataprep) (1.3.2)

Requirement already satisfied: bokeh<3,>=2 in /usr/local/lib/python3.7/dist-packages (from dataprep) (2.3.3)

Requirement already satisfied: nltk<4.0,>=3.5 in /usr/local/lib/python3.7/dist-packages (from dataprep) (3.6.2)

Requirement already satisfied: pandas<2.0,>=1.1 in /usr/local/lib/python3.7/dist-packages (from dataprep) (1.1.5)

Requirement already satisfied: regex<2021.0.0,>=2020.10.15 in /usr/local/lib/python3.7/dist-packages (from dataprep) (2020.11.13)

Requirement already satisfied: ipywidgets<8.0,>=7.5 in /usr/local/lib/python3.7/dist-packages (from dataprep) (7.6.3)

Requirement already satisfied: pydantic<2.0,>=1.6 in /usr/local/lib/python3.7/dist-packages (from dataprep) (1.8.2)

Requirement already satisfied: async-timeout<4.0,>=3.0 in /usr/local/lib/python3.7/dist-packages (from aiohttp<4.0,>=3.6->dataprep) (3.0.1)

Requirement already satisfied: multidict<7.0,>=4.5 in /usr/local/lib/python3.7/dist-packages (from aiohttp<4.0,>=3.6->dataprep) (5.1.0)

Requirement already satisfied: yarl<2.0,>=1.0 in /usr/local/lib/python3.7/dist-packages (from aiohttp<4.0,>=3.6->dataprep) (1.6.3)

Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.7/dist-packages (from aiohttp<4.0,>=3.6->dataprep) (21.2.0)

Requirement already satisfied: typing-extensions>=3.6.5 in /usr/local/lib/python3.7/dist-packages (from aiohttp<4.0,>=3.6->dataprep) (3.7.4.3)

Requirement already satisfied: chardet<5.0,>=2.0 in /usr/local/lib/python3.7/dist-packages (from aiohttp<4.0,>=3.6->dataprep) (3.0.4)

Requirement already satisfied: pillow>=7.1.0 in /usr/local/lib/python3.7/dist-packages (from bokeh<3,>=2->dataprep) (7.1.2)

Requirement already satisfied: PyYAML>=3.10 in /usr/local/lib/python3.7/dist-packages (from bokeh<3,>=2->dataprep) (3.13)

Requirement already satisfied: tornado>=5.1 in /usr/local/lib/python3.7/dist-packages (from bokeh<3,>=2->dataprep) (5.1.1)

Requirement already satisfied: python-dateutil>=2.1 in /usr/local/lib/python3.7/dist-packages (from bokeh<3,>=2->dataprep) (2.8.2)

Requirement already satisfied: packaging>=16.8 in /usr/local/lib/python3.7/dist-packages (from bokeh<3,>=2->dataprep) (21.0)

Requirement already satisfied: toolz>=0.8.2 in /usr/local/lib/python3.7/dist-packages (from dask[array,dataframe,delayed]<3.0,>=2.25->dataprep) (0.11.1)

Requirement already satisfied: cloudpickle>=0.2.2 in /usr/local/lib/python3.7/dist-packages (from dask[array,dataframe,delayed]<3.0,>=2.25->dataprep) (1.3.0)

Requirement already satisfied: fsspec>=0.6.0 in /usr/local/lib/python3.7/dist-packages (from dask[array,dataframe,delayed]<3.0,>=2.25->dataprep) (2021.8.1)

Requirement already satisfied: partd>=0.3.10 in /usr/local/lib/python3.7/dist-packages (from dask[array,dataframe,delayed]<3.0,>=2.25->dataprep) (1.2.0)

Requirement already satisfied: widgetsnbextension~=3.5.0 in /usr/local/lib/python3.7/dist-packages (from ipywidgets<8.0,>=7.5->dataprep) (3.5.1)

Requirement already satisfied: traitlets>=4.3.1 in /usr/local/lib/python3.7/dist-packages (from ipywidgets<8.0,>=7.5->dataprep) (5.0.5)

Requirement already satisfied: jupyterlab-widgets>=1.0.0 in /usr/local/lib/python3.7/dist-packages (from ipywidgets<8.0,>=7.5->dataprep) (1.0.0)

Requirement already satisfied: ipykernel>=4.5.1 in /usr/local/lib/python3.7/dist-packages (from ipywidgets<8.0,>=7.5->dataprep) (4.10.1)

Requirement already satisfied: ipython>=4.0.0 in /usr/local/lib/python3.7/dist-packages (from ipywidgets<8.0,>=7.5->dataprep) (5.5.0)

Requirement already satisfied: nbformat>=4.2.0 in /usr/local/lib/python3.7/dist-packages (from ipywidgets<8.0,>=7.5->dataprep) (5.1.3)

Requirement already satisfied: jupyter-client in /usr/local/lib/python3.7/dist-packages (from ipykernel>=4.5.1->ipywidgets<8.0,>=7.5->dataprep) (5.3.5)

Requirement already satisfied: pickleshare in /usr/local/lib/python3.7/dist-packages (from ipython>=4.0.0->ipywidgets<8.0,>=7.5->dataprep) (0.7.5)

Requirement already satisfied: simplegeneric>0.8 in /usr/local/lib/python3.7/dist-packages (from ipython>=4.0.0->ipywidgets<8.0,>=7.5->dataprep) (0.8.1)

Requirement already satisfied: pygments in /usr/local/lib/python3.7/dist-packages (from ipython>=4.0.0->ipywidgets<8.0,>=7.5->dataprep) (2.6.1)

Requirement already satisfied: prompt-toolkit<2.0.0,>=1.0.4 in /usr/local/lib/python3.7/dist-packages (from ipython>=4.0.0->ipywidgets<8.0,>=7.5->dataprep) (1.0.18)

Requirement already satisfied: setuptools>=18.5 in /usr/local/lib/python3.7/dist-packages (from ipython>=4.0.0->ipywidgets<8.0,>=7.5->dataprep) (57.4.0)

Requirement already satisfied: decorator in /usr/local/lib/python3.7/dist-packages (from ipython>=4.0.0->ipywidgets<8.0,>=7.5->dataprep) (4.4.2)

Requirement already satisfied: pexpect in /usr/local/lib/python3.7/dist-packages (from ipython>=4.0.0->ipywidgets<8.0,>=7.5->dataprep) (4.8.0)

Requirement already satisfied: MarkupSafe>=0.23 in /usr/local/lib/python3.7/dist-packages (from jinja2<3.0,>=2.11->dataprep) (2.0.1)

Requirement already satisfied: six in /usr/local/lib/python3.7/dist-packages (from jsonpath-ng<2.0,>=1.5->dataprep) (1.15.0)

Requirement already satisfied: ply in /usr/local/lib/python3.7/dist-packages (from jsonpath-ng<2.0,>=1.5->dataprep) (3.11)

Requirement already satisfied: jupyter-core in /usr/local/lib/python3.7/dist-packages (from nbformat>=4.2.0->ipywidgets<8.0,>=7.5->dataprep) (4.7.1)

Requirement already satisfied: ipython-genutils in /usr/local/lib/python3.7/dist-packages (from nbformat>=4.2.0->ipywidgets<8.0,>=7.5->dataprep) (0.2.0)

Requirement already satisfied: jsonschema!=2.5.0,>=2.4 in /usr/local/lib/python3.7/dist-packages (from nbformat>=4.2.0->ipywidgets<8.0,>=7.5->dataprep) (2.6.0)

Requirement already satisfied: click in /usr/local/lib/python3.7/dist-packages (from nltk<4.0,>=3.5->dataprep) (7.1.2)

Requirement already satisfied: joblib in /usr/local/lib/python3.7/dist-packages (from nltk<4.0,>=3.5->dataprep) (1.0.1)

Requirement already satisfied: pyparsing>=2.0.2 in /usr/local/lib/python3.7/dist-packages (from packaging>=16.8->bokeh<3,>=2->dataprep) (2.4.7)

Requirement already satisfied: pytz>=2017.2 in /usr/local/lib/python3.7/dist-packages (from pandas<2.0,>=1.1->dataprep) (2018.9)

Requirement already satisfied: locket in /usr/local/lib/python3.7/dist-packages (from partd>=0.3.10->dask[array,dataframe,delayed]<3.0,>=2.25->dataprep) (0.2.1)

Requirement already satisfied: wcwidth in /usr/local/lib/python3.7/dist-packages (from prompt-toolkit<2.0.0,>=1.0.4->ipython>=4.0.0->ipywidgets<8.0,>=7.5->dataprep) (0.2.5)

Requirement already satisfied: probableparsing in /usr/local/lib/python3.7/dist-packages (from usaddress<0.6.0,>=0.5.10->dataprep) (0.0.1)

Requirement already satisfied: future>=0.14 in /usr/local/lib/python3.7/dist-packages (from usaddress<0.6.0,>=0.5.10->dataprep) (0.16.0)

Requirement already satisfied: python-crfsuite>=0.7 in /usr/local/lib/python3.7/dist-packages (from usaddress<0.6.0,>=0.5.10->dataprep) (0.9.7)

Requirement already satisfied: notebook>=4.4.1 in /usr/local/lib/python3.7/dist-packages (from widgetsnbextension~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (5.3.1)

Requirement already satisfied: terminado>=0.8.1 in /usr/local/lib/python3.7/dist-packages (from notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (0.11.0)

Requirement already satisfied: nbconvert in /usr/local/lib/python3.7/dist-packages (from notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (5.6.1)

Requirement already satisfied: Send2Trash in /usr/local/lib/python3.7/dist-packages (from notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (1.8.0)

Requirement already satisfied: pyzmq>=13 in /usr/local/lib/python3.7/dist-packages (from jupyter-client->ipykernel>=4.5.1->ipywidgets<8.0,>=7.5->dataprep) (22.2.1)

Requirement already satisfied: ptyprocess in /usr/local/lib/python3.7/dist-packages (from terminado>=0.8.1->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (0.7.0)

Requirement already satisfied: matplotlib in /usr/local/lib/python3.7/dist-packages (from wordcloud<2.0,>=1.8->dataprep) (3.2.2)

Requirement already satisfied: idna>=2.0 in /usr/local/lib/python3.7/dist-packages (from yarl<2.0,>=1.0->aiohttp<4.0,>=3.6->dataprep) (2.10)

Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.7/dist-packages (from matplotlib->wordcloud<2.0,>=1.8->dataprep) (0.10.0)

Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.7/dist-packages (from matplotlib->wordcloud<2.0,>=1.8->dataprep) (1.3.1)

Requirement already satisfied: mistune<2,>=0.8.1 in /usr/local/lib/python3.7/dist-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (0.8.4)

Requirement already satisfied: bleach in /usr/local/lib/python3.7/dist-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (4.0.0)

Requirement already satisfied: testpath in /usr/local/lib/python3.7/dist-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (0.5.0)

Requirement already satisfied: entrypoints>=0.2.2 in /usr/local/lib/python3.7/dist-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (0.3)

Requirement already satisfied: pandocfilters>=1.4.1 in /usr/local/lib/python3.7/dist-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (1.4.3)

Requirement already satisfied: defusedxml in /usr/local/lib/python3.7/dist-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (0.7.1)

Requirement already satisfied: webencodings in /usr/local/lib/python3.7/dist-packages (from bleach->nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<8.0,>=7.5->dataprep) (0.5.1)

[137]

8s

create\_report(base\_completa)



[138]

0s

estados.isna().sum()



id\_estado 0

estado 0

sigla\_estado 0

pais 2

dtype: int64

[139]

0s

idade.isna().sum()



id\_cliente 0

idade 0

dtype: int64

[140]

0s

clientes.isna().sum()



id\_cliente 0

peso 0

colesterol 0

genero 0

id\_estado 0

dtype: int64

[141]

0s

estados.pais.fillna('Brasil',inplace=True)



[142]

0s

estados



[143]

0s

clientes



[144]

0s

from sklearn.preprocessing import LabelEncoder

LE = LabelEncoder()

clientes['genero\_number']=LE.fit\_transform(clientes['genero'])



[145]

0s

clientes



[146]

0s

clientes\_estados = pd.merge(clientes, estados, on="id\_estado")

clientes\_estados



[147]

0s

base\_completa = pd.merge(clientes\_estados, idade, on="id\_cliente").sort\_values('id\_cliente')



[149]

0s

base\_completa.duplicated().sum()



0

[154]

0s

base\_completa=base\_completa.reset\_index(drop=True)



[155]

0s

base\_completa.rename(columns={'id\_cliente':'cod\_cliente','genero\_number':'cod\_genero'}, inplace=True)



[156]

0s

base\_completa['sigla\_estado']=base\_completa.sigla\_estado.str.upper()



[157]

0s

base\_completa



[158]

0s

from sklearn.cluster import KMeans



[159]

0s

def calcular\_wcss(dados\_cliente):

  wcss=[]

  for k in range(1,11):

    kmeans=KMeans(n\_clusters=k)

    kmeans.fit(X=dados\_cliente)

    wcss.append(kmeans.inertia\_)

  return wcss



[175]

0s

dados\_cliente = base\_completa[['peso','colesterol','idade']]



[176]

0s

wcss\_cliente=calcular\_wcss(dados\_cliente)



[177]

0s

for i in range(len(wcss\_cliente)):

  print(f'O cluster {i} possui valor WCSS de: {wcss\_cliente[i]}')



O cluster 0 possui valor WCSS de: 1481446.4533820841

O cluster 1 possui valor WCSS de: 455970.5535834613

O cluster 2 possui valor WCSS de: 265021.2549377008

O cluster 3 possui valor WCSS de: 206738.26593357133

O cluster 4 possui valor WCSS de: 178433.04442735814

O cluster 5 possui valor WCSS de: 151039.13503247092

O cluster 6 possui valor WCSS de: 129576.49976750824

O cluster 7 possui valor WCSS de: 108337.29215751175

O cluster 8 possui valor WCSS de: 93442.30249957771

O cluster 9 possui valor WCSS de: 82799.70419933891

[178]

0s

import plotly.express as px

import plotly.offline as py

import plotly.graph\_objects as go



[179]

0s

grafico\_wcss = px.line (x=range(1,11),

                        y=wcss\_cliente

                        )

fig = go.Figure(grafico\_wcss)

fig.update\_layout(title='calculando o WCSS',

                  xaxis\_title='Número de clusters',

                  yaxis\_title='Valor do WCSS',

                  template='plotly\_white'

)

fig.show()



[184]

0s

kmeans\_cliente=KMeans(n\_clusters=3,random\_state=0)

base\_completa['cluster']=kmeans\_cliente.fit\_predict(dados\_cliente)



[185]

0s

base\_completa



[186]

0s

centroides\_clusters=kmeans\_cliente.cluster\_centers\_

centroides\_clusters



array([[110.31216931, 125.71957672, 43.54497354],

[182.26315789, 217.04093567, 42.78362573],

[141.80748663, 173.00534759, 42.03743316]])

[187]

1s

grafico = px.scatter (x=base\_completa['peso'],

                        y=base\_completa['colesterol'],

                        color=base\_completa['cluster'])

grafico\_centroide = px.scatter(x=centroides\_clusters[:,0],y=centroides\_clusters[:,1],size=[7,7,7])

grafico\_final = go.Figure(data=grafico.data+grafico\_centroide.data)

grafico\_final.update\_layout(title='Clusterização',

                  xaxis\_title='Peso',

                  yaxis\_title='Colesterol')

grafico\_final.show()



[189]

0s

base\_completa.loc[base\_completa['cluster']==2,'nome\_cluster']="Alto Risco"

base\_completa.loc[base\_completa['cluster']==1,'nome\_cluster']="Médio Risco"

base\_completa.loc[base\_completa['cluster']==0,'nome\_cluster']="Baixo Risco"



[190]

0s

base\_completa = base\_completa.reset\_index(drop=True)



[191]

0s

base\_completa.groupby(['nome\_cluster']).describe()



[191]

0s



check

0sconclusão: 19:51

axis, hint

Não foi possível conectar-se ao serviço reCAPTCHA. Verifique sua conexão com a Internet e atualize a página para ver um desafio reCAPTCHA.