# import libraries

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sb

# loading the data

Dataframe = pd.read\_csv("C:\\Users\\unic\\Downloads\\archive (18)\\winequalityN.csv")

# show rows and columns

Dataframe.head()

# getting info.

Dataframe.info()

Dataframe.describe()

# null value check

Dataframe.isnull().sum()

# plot pairplot

sb.pairplot(Dataframe)

#show graph

plt.show()

#plot histogram

Dataframe.hist(bins=20,figsize=(10,10))

#plot showing

plt.show()

#plt.figure(figsize=[15,6])

#plt.bar(['residual sugar'],['chlorides'])

#plt.xlabel('quality')

#plt.ylabel('alcohol')

#plt.show()

# correlation by visualization

plt.figure(figsize=[18,7])

# plot correlation

sb.heatmap(Dataframe.corr(),annot=True)

plt.show()

colm = []

# loop for columns

for i in range(len(Dataframe.corr().keys())):

# loop for rows

# for j in range(j):

#if abs(Dataframe.corr().iloc[i,j]) > 0.7:

colm = Dataframe.corr().columns[i]

# drop column

new\_df = Dataframe.drop('total sulfur dioxide',axis = 1)

new\_df.update(new\_df.fillna(new\_df.mean()))

# no of categorical columns

cat = new\_df.select\_dtypes(include='O')

# create dummies of categorical columns

df\_dummies = pd.get\_dummies(new\_df,drop\_first = True)

print(df\_dummies)

df\_dummies['best quality']=[1 if x>=7 else 0 for x in Dataframe.quality]

print(df\_dummies)