

KGiSL Institute of Technology

NAAN MUDHALVAN

Project title :

Public Health Awareness

TEAM MEMBERS:

* Mersheena.F
* Dhiviya Shree.K
* Navin.c
* Surya.S P P

PROJECT DESCRIPTION :

**PHASE 3 : Analysing On The Public Health Awareness DataSet**

OBJECTIVE :

STEPS :

in google colab notebook:

* Mount the google drive
* Load the DataSet to the Google ColabNotebook
* Process the DataSet
* Cleansing
* Accuracy

**#MOUNT:**

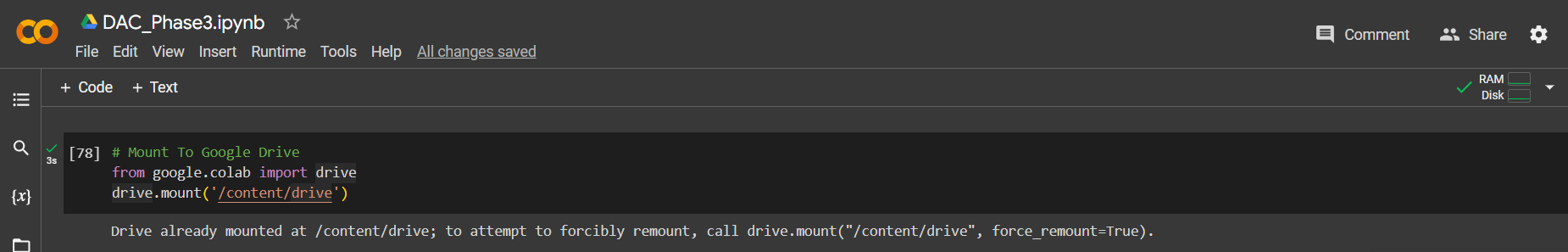
CODE :

# Mount To Google Drive

from google.colab import drive

drive.mount('/content/drive')

OUTPUT :



**#LOAD:**

CODE :

# Load The DataSet

import pandas as pd

import matplotlib.pyplot as plt

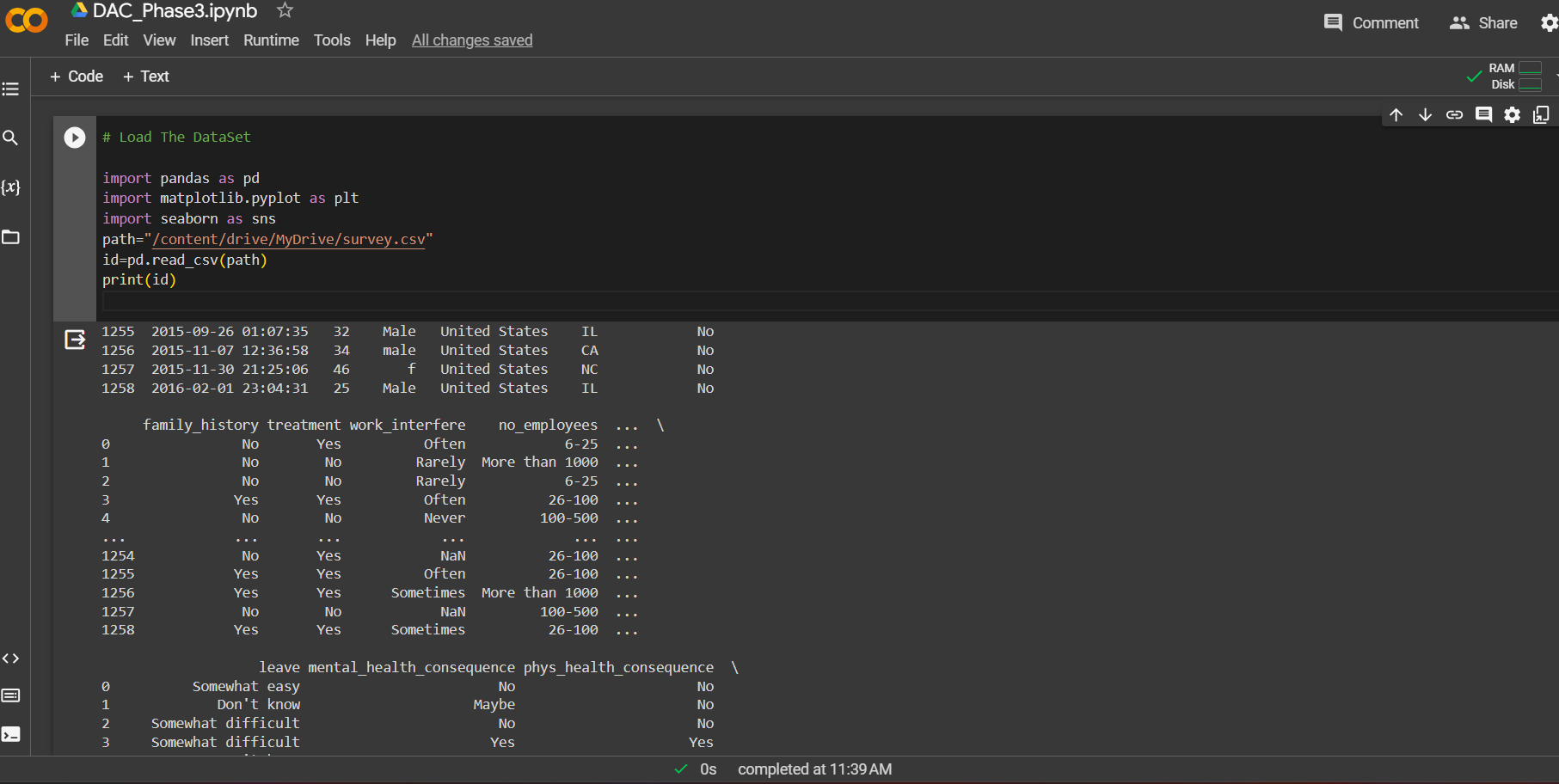
import seaborn as sns

path="/content/drive/MyDrive/survey.csv"

id=pd.read\_csv(path)

print(id)

OUTPUT:

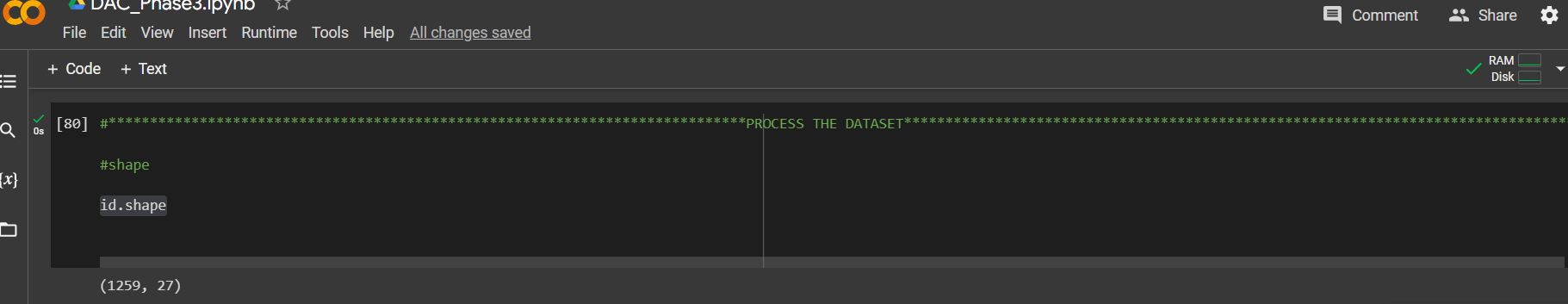


**#PROCESS & CLEANSING THE DATASET:**

CODE:

id.shape

OUTPUT:

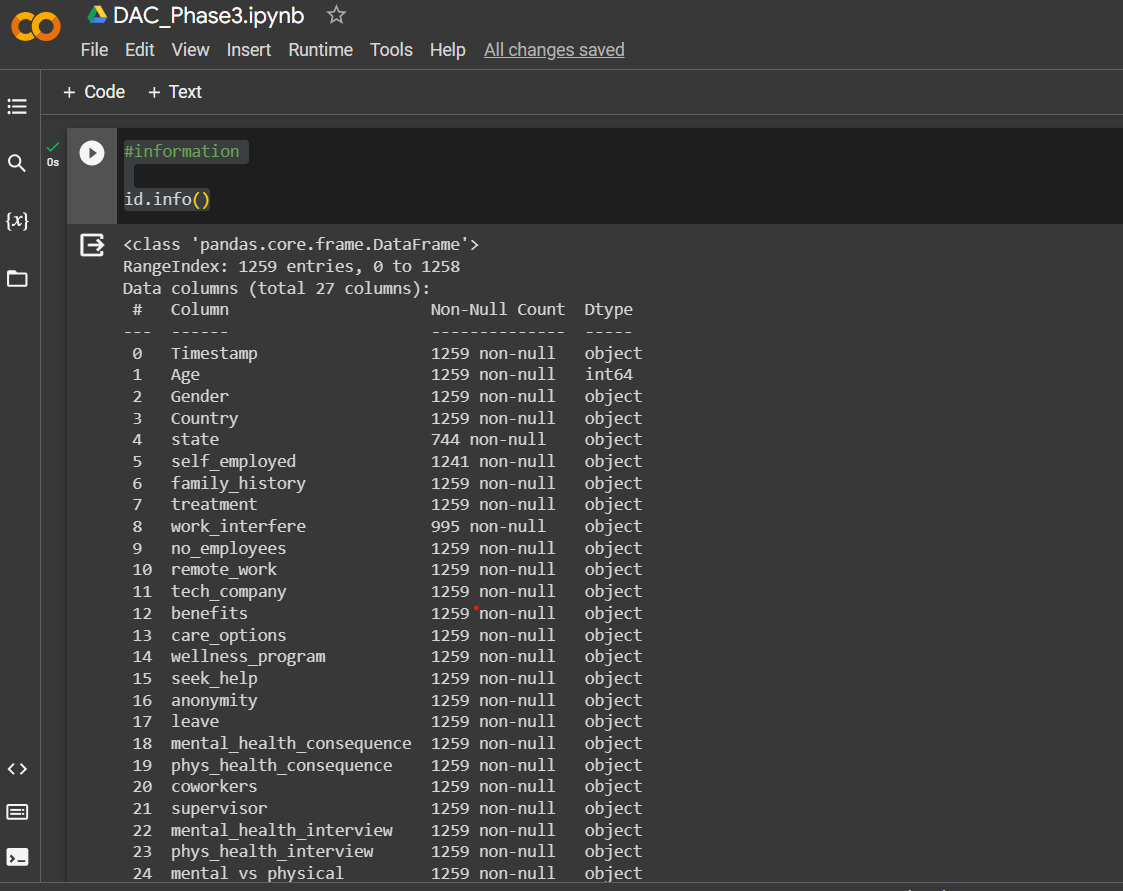


CODE :

#information

id.info()

OUTPUT:

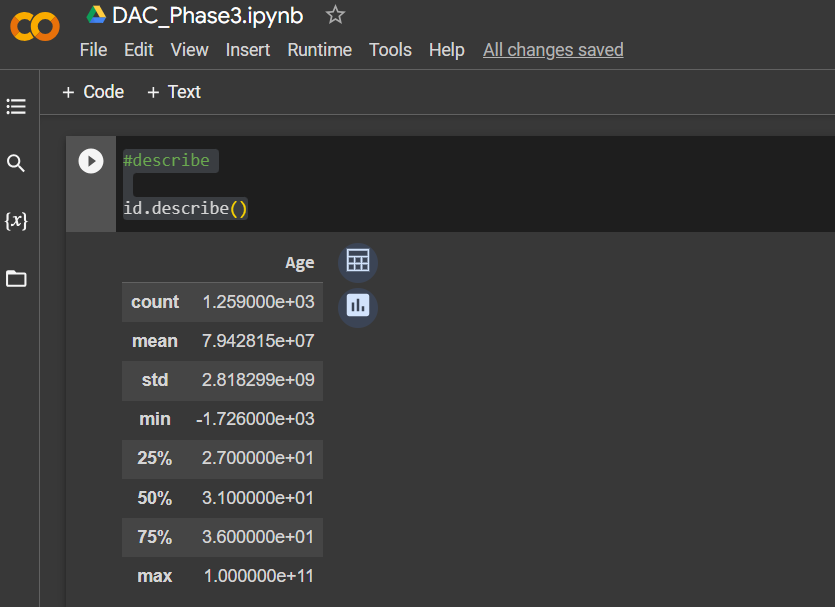


CODE :

#describe

id.describe()

OUTPUT:



CODE :

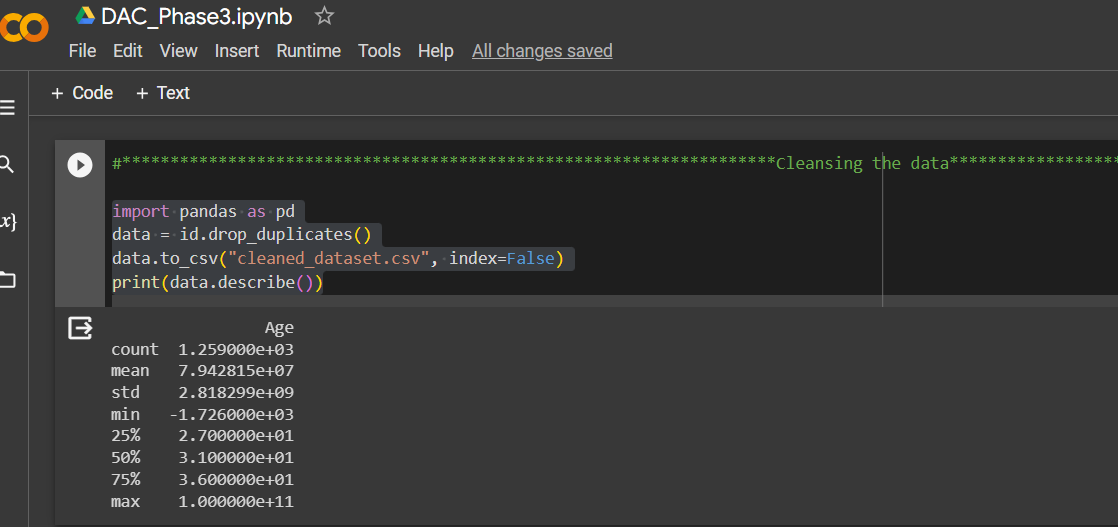
import pandas as pd

data = id.drop\_duplicates()

data.to\_csv("cleaned\_dataset.csv", index=False)

print(data.describe())

OUTPUT:



CODE :

#PLOTTING:

import seaborn as sns

import matplotlib.pyplot as plt

sns.countplot(x="Age",data=dt)

plt.figure(figsize=(10,6))

plt.show()

OUTPUT:



**#ACCURACY :**

CODE :

#logistic regression

from sklearn.datasets import make\_classification

from sklearn.model\_selection import train\_test\_split

from sklearn.metrics import accuracy\_score

from sklearn.linear\_model import LogisticRegression

nb\_samples = 1000

x, y = make\_classification(n\_samples=nb\_samples, n\_features=2, n\_informative=2, n\_redundant=0, n\_clusters\_per\_class=1)

xtrain, xtest, ytrain, ytest = train\_test\_split(x, y, test\_size=0.2, random\_state=42)

model = LogisticRegression()

model.fit(xtrain, ytrain)

#Accuracy of the data

print(accuracy\_score(ytest, model.predict(xtest)))

OUTPUT :

