Dataset Suicide cases by Occupations Cause by Depression

#import raw dataset of Suicide Cases Based on Occupation and display

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

from matplotlib import pyplot as plt

data = pd.read\_csv('capstone D6 deathByOccAge.csv')

df=pd.DataFrame(data)

print(df)

display(df.head())

# Check for null values

df.info()

print('\n')

#drop row having age\_group 29 and below. And also age\_group of 70 and above

df.drop([0,1,2,3,4,5,6,7,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57],axis=0, inplace=True)

df.drop([92,93,94,95,96,97,98,99,100,101,102,103,104,105,106,107,108,109,110,111,112,146,147,148,149,150,151,152,153,154,155,156,157,158,159,160,161,162,163,164,165],axis=0, inplace=True)

df.drop([201,202,203,204,205,206,207,208,209,210,211,212,213,214,215,216,217,218,219,220,221],axis=0, inplace=True)

df.drop([254,256,257,258,259,260,261,262,263,264,265,267,268,269,270,271,306,307,308,309,310,311,312,313,314],axis=0, inplace=True)

df.head()

print(df)

# Export current dataset to project file

df.to\_csv(r'C:\Users\ASUS\Desktop\utem\DataScience\capstone\dataCleanFinal\Suicide\_Occupation.csv', index = False)

Change age and gender to integer and remove occupation not working

import pandas as pd

import numpy as np

df = pd.read\_csv('Working.csv')

df["sex"] = df["sex"].astype('category')

df["sex"] = df["sex"].cat.codes

df["age\_group"] = df["age\_group"].astype('category')

df["age\_group"] = df["age\_group"].cat.codes

new = df[df.occupation != 'Not Working']

print(new)

new.to\_csv(r'D:\BITI2513 INTRODUCTION TO DATA SCIENCE\CAPSTONE PROJECT\Latest\_Working\_Dataset\_2.csv')