**Importing Libraries**

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

**Importing Dataset**

data = pd.read\_csv('/content/Depression.csv')

data = data.sort\_values(['year'], ascending = True)

print(data.shape)

data.head(5)

**Count**

data.count

**splitting the data**

x=data.iloc[:, :-1].values

y=data.iloc[:,-1].values

**training and testing sets**

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

x\_train, x\_test, y\_train, y\_test = train\_test\_split(x, y, test\_size = 0.25, random\_state =0)

x\_train

y\_train

**Cleaning the data**

ata['Posts']=data['Posts'].astype(str)

data["Posts"][data["Posts"] == "[]"] = np.nan

data["Posts"] = data["Posts"].fillna(0)

data.Posts=data.Posts.str.replace("[", "")

data.Posts=data.Posts.str.replace("]", "")

data.Posts=data.Posts.str.replace("'", "")

data["Posts"] = data["Posts"].str.split(",")

data=data.explode('Posts')

data

**Data Visualization** Box Plot

sns.boxplot(data['country'],data['age'])

**Using the population field it can sum**

data.groupby('population').sum()

**Data Visualization** Heatmap

f, ax = plt.subplots(figsize = (4, 3))

corr = data.corr()

sns.heatmap(corr,annot=True, mask = np.zeros\_like(corr, dtype = np.bool),

            cmap = sns.diverging\_palette(3, 3, as\_cmap = True), square = True, ax = ax)

**Renaming the columns**

data.rename({'sex' : 'gender', 'suicides\_no' : 'suicides'}, inplace = True, axis = 1)

data.columns

data.info()

data.describe()

data.isnull().sum()

data['country'].value\_counts().count()

**Data Visualization** different countries distribution in the dataset

data['country'].value\_counts(normalize = True)

data['country'].value\_counts(dropna = False).plot.bar(color = 'blue', figsize = (24, 8))

plt.title('Distribution of 8 coutries in suicides')

plt.xlabel('country')

plt.ylabel('count')

plt.show()

data['year'].value\_counts().count()

data['year'].value\_counts(normalize = True)

data['year'].value\_counts(dropna = False,).plot.bar(color = 'magenta', figsize = (8, 6))

plt.title('Distribution of suicides from the year 2000 to 2016')

plt.xlabel('year')

plt.ylabel('count')

plt.show()

**Encoding**

from sklearn.preprocessing import LabelEncoder

# creating an encoder

le = LabelEncoder()

data['gender'] = le.fit\_transform(data['gender'])

data['gender'].value\_counts()

**Data Visualization of bar plot**

data['gender'].value\_counts(normalize = True)

data['gender'].value\_counts(dropna = True).plot.bar(color = 'black', figsize = (4, 3))

plt.title('Distribution of 8 coutries in suicides')

plt.xlabel('gender')

plt.ylabel('count')

plt.show()

**Data Visualization in distplot**

sns.distplot(data['year'])

**Data Visualization in joint plot**

  sns.jointplot(data['suicides'], data['year'],kind="reg")

sns.jointplot(data['suicides'], data['year'],kind="hex")

**Grouping**

data.groupby('age')['Posts'].sum()

Ascending order

data.groupby('gender')['Posts'].sum().sort\_values(ascending = True)

Descending order

data.groupby('gender')['Posts'].sum().sort\_values(ascending = False)

data.groupby('gender')['suicides'].sum()

data.groupby('Posts')['gender'].sum().head(10)

data.groupby('country')['population'].sum().head(8)

data.groupby('Posts')['country'].sum().head(10)

**NaN values**

data['population'] = data['population'].fillna(data['population'].median())

data['population'].isnull().any()

**Total population with the suicides survey is committed**

data['population'].sum()

**Average population**

Avg\_pop = data['population'].mean()

print(Avg\_pop)

**Total number of suicides committed in the 8 countries from 1985 to 2016**

data['suicides'].sum()