```
#from google.colab import files
#data_uploaded = files.upload()
#this commands are used in linux

#import io
#import pandas as pd
#!wget https://github.com/devilsboy/DataPreprocessing/blob/main/Data.csv can also
try this to get the Data.csv file
#df = pd.read_csv(io.BytesIO(data_uploaded['Data.csv']))
#df
#used in linux

#this is used in windows
import io
import pandas as pd
df = pd.read_csv('Data.csv')
df

#1a
df.describe()

df.shape

df.head(3)

#2
import pandas as pd
import numpy as np
#method 1- Filling every missing values with 0
print("\n\n Every missing value Replaced with '0':")
print("-------------------------------------------------")
print(df.fillna(0))

#method2 - Dropping rows having missing value
print("\n\n Dropping Rows with Missing values:")
print(df.dropna())

#method 3 - Replacing missing values with the mean
df.fillna(df.mean())

from sklearn.preprocessing import LabelEncoder
LabelEncoder = LabelEncoder()
df['Country']=LabelEncoder.fit_transform(df['Country'])
df

#3a
from sklearn.preprocessing import OneHotEncoder
enc = OneHotEncoder(handle_unknown='ignore')
enc_df=pd.DataFrame(enc.fit_transform(df[['Country']]).toarray())
enc_df
```

```
df = df.join(enc_df)
df

#3b
from sklearn.preprocessing import LabelEncoder
LabelEncoder = LabelEncoder()
df['Purchased']=LabelEncoder.fit_transform(df['Purchased'])
df
```