Name : Kshitij V Darwhekar

Roll No: TETB19

Sub: Soft Computitng

Batch: B2

Experiment 1 : Experimental Data Analysis: Perform following operations on any open dataset available in Python/Kaggle

```python
import numpy as np
import pandas as pd

from google.colab import drive
drive.mount('/content/drive/')
```

```python
#data = open('ML/penguins_size','r')
```

    Mounted at /content/drive/

```python
database = pd.read_csv('/content/drive/MyDrive/ML/penguins_size.csv')
```

```python
database.head()
```

	species	island	culmen_length_mm	culmen_depth_mm	flipper_length_mm	body_mas
0	Adelie	Torgersen	39.1	18.7	181.0	375
1	Adelie	Torgersen	39.5	17.4	186.0	380
2	Adelie	Torgersen	40.3	18.0	195.0	325
3	Adelie	Torgersen	NaN	NaN	NaN	N
4	Adelie	Torgersen	36.7	19.3	193.0	345

```python
database.head(10)
```

	species	island	culmen_length_mm	culmen_depth_mm	flipper_length_mm	body_mas
0	Adelie	Torgersen	39.1	18.7	181.0	375
1	Adelie	Torgersen	39.5	17.4	186.0	380
2	Adelie	Torgersen	40.3	18.0	195.0	325
3	Adelie	Torgersen	NaN	NaN	NaN	N
4	Adelie	Torgersen	36.7	19.3	193.0	345
5	Adelie	Torgersen	39.3	20.6	190.0	365
6	Adelie	Torgersen	38.9	17.8	181.0	362

```
database.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 344 entries, 0 to 343
Data columns (total 7 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   species            344 non-null    object
 1   island             344 non-null    object
 2   culmen_length_mm   342 non-null    float64
 3   culmen_depth_mm    342 non-null    float64
 4   flipper_length_mm  342 non-null    float64
 5   body_mass_g        342 non-null    float64
 6   sex                334 non-null    object
dtypes: float64(4), object(3)
memory usage: 18.9+ KB
```

```
print(database.isnull().sum())
```

```
species             0
island              0
culmen_length_mm    2
culmen_depth_mm     2
flipper_length_mm   2
body_mass_g         2
sex                10
dtype: int64
```

```
database = database.dropna()
database.head()
```

	species	island	culmen_length_mm	culmen_depth_mm	flipper_length_mm	body_mas
0	Adelie	Torgersen	39.1	18.7	181.0	375
1	Adelie	Torgersen	39.5	17.4	186.0	380
2	Adelie	Torgersen	40.3	18.0	195.0	325
4	Adelie	Torgersen	36.7	19.3	193.0	345
5	Adelie	Torgersen	39.3	20.6	190.0	365

```
len(database)
```

    334

```
len(database.columns)
```

    7

```
database.loc[(database['sex'] != 'FEMALE')& (database['sex'] != 'MALE')]
```

	species	island	culmen_length_mm	culmen_depth_mm	flipper_length_mm	body_mas
336	Gentoo	Biscoe	44.5	15.7	217.0	487

```
database['culmen_depth_mm'].fillna((database['culmen_depth_mm'].mean()), inplace=True)
database['flipper_length_mm'].fillna((database['flipper_length_mm'].mean()), inplace=True)
database['body_mass_g'].fillna((database['body_mass_g'].mean()), inplace=True)
database['culmen_length_mm'].fillna((database['culmen_length_mm'].mean()), inplace=True)
database['sex'].fillna((database['sex'].value_counts().index[0]), inplace=True)

database.reset_index()
database.head()
```

	species	island	culmen_length_mm	culmen_depth_mm	flipper_length_mm	body_mas
0	Adelie	Torgersen	39.1	18.7	181.0	375
1	Adelie	Torgersen	39.5	17.4	186.0	380
2	Adelie	Torgersen	40.3	18.0	195.0	325
4	Adelie	Torgersen	36.7	19.3	193.0	345
5	Adelie	Torgersen	39.3	20.6	190.0	365

```
col_new = ['new_species','new_island','new_culmen_length_mm','new_culmen_depth_mm','new_fl
database.columns = col_new
col_new
```

    ['new_species',
     'new_island',
     'new_culmen_length_mm',
     'new_culmen_depth_mm',
     'new_flipper_length',
     'new_body_mass_g',
     'new_sex']

```
database.head()
```

	new_species	new_island	new_culmen_length_mm	new_culmen_depth_mm	new_flipper_l
0	Adelie	Torgersen	39.1	18.7	
1	Adelie	Torgersen	39.5	17.4	
2	Adelie	Torgersen	40.3	18.0	

```
database_new = database.drop(['new_island','new_culmen_length_mm','new_flipper_length'],ax
database.head()
```

	new_species	new_island	new_culmen_length_mm	new_culmen_depth_mm	new_flipper_l
0	Adelie	Torgersen	39.1	18.7	
1	Adelie	Torgersen	39.5	17.4	
2	Adelie	Torgersen	40.3	18.0	
4	Adelie	Torgersen	36.7	19.3	
5	Adelie	Torgersen	39.3	20.6	

```
database_new.head()
```

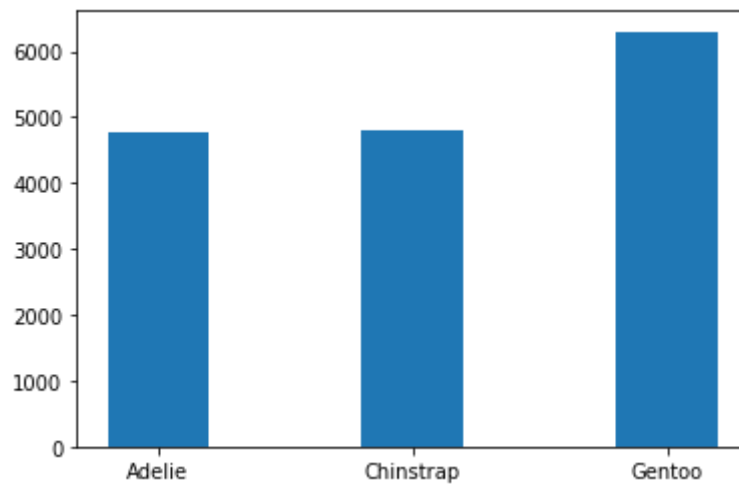|	new_species	new_culmen_depth_mm	new_body_mass_g	new_sex
0	Adelie	18.7	3750.0	MALE
1	Adelie	17.4	3800.0	FEMALE
2	Adelie	18.0	3250.0	FEMALE
4	Adelie	19.3	3450.0	FEMALE
5	Adelie	20.6	3650.0	MALE

```
database_new["islands"] = "Torgersen"
database_new.head()
```

	new_species	new_culmen_depth_mm	new_body_mass_g	new_sex	islands
0	Adelie	18.7	3750.0	MALE	Torgersen
1	Adelie	17.4	3800.0	FEMALE	Torgersen
2	Adelie	18.0	3250.0	FEMALE	Torgersen
4	Adelie	19.3	3450.0	FEMALE	Torgersen
5	Adelie	20.6	3650.0	MALE	Torgersen

```
import matplotlib.pyplot as plt
import seaborn as sns
```

```
X = database['new_species']
Y = database['new_body_mass_g']
plt.bar(X,Y,width = 0.4)
```

<BarContainer object of 334 artists>

Code   Text

✓ 0s   completed at 1:41 PM   ● ✕

```
X = database['new_species']
Y = database['new_body_mass_g']
plt.bar(X,Y,width = 0.4)
```

<BarContainer object of 334 artists>