

Machine Learning

Assignment Part-2

Name: Saideep Reddy Vurivakili
700#: 700755895

Question 1 & 2:

```
from google.colab import drive
drive.mount('/drive')
import numpy as np
import pandas as pd

#(1)Read the provided CSV file 'data.csv'
data_Manip = pd.read_csv('/drive/MyDrive/data.csv')
data_Manip.head()
```

Mounted at [/drive](#)

	Duration	Pulse	Maxpulse	Calories
0	60	110	130	409.1
1	60	117	145	479.0
2	60	103	135	340.0
3	45	109	175	282.4
4	45	117	148	406.0

```
#(2) Show the basic statistical description about the data.
data_Manip. head ()
```

	Duration	Pulse	Maxpulse	Calories
0	60	110	130	409.1
1	60	117	145	479.0
2	60	103	135	340.0
3	45	109	175	282.4
4	45	117	148	406.0

Explanation

We use data_manip.head() to show basic statistical description

Question 3:

```
#(3) Check if the data has null values.  
data_Manip.isnull().any()
```

Python

```
Duration    False  
Pulse       False  
Maxpulse    False  
Calories    True  
dtype: bool
```

Explanation:

We use `data_manip.isnull().any()` to check if there are any null values

Question 3a:

```
#(3a) Replace the null values with the mean  
column_means = data_Manip.mean()  
print(column_means)  
data_Manip = data_Manip.fillna(column_means)  
print(data_Manip.head(20))
```

```
Duration    63.846154  
Pulse       107.461538  
Maxpulse    134.047337  
Calories    375.790244  
dtype: float64
```

	Duration	Pulse	Maxpulse	Calories
0	60	110	130	409.100000
1	60	117	145	479.000000
2	60	103	135	340.000000
3	45	109	175	282.400000
4	45	117	148	406.000000
5	60	102	127	300.000000
6	60	110	136	374.000000
7	45	104	134	253.300000
8	30	109	133	195.100000
9	60	98	124	269.000000
10	60	103	147	329.300000
11	60	100	120	250.700000
12	60	106	128	345.300000
13	60	104	132	379.300000
14	60	98	123	275.000000
15	60	98	120	215.200000
16	60	100	120	300.000000
17	45	90	112	375.790244
18	60	103	123	323.000000
19	45	97	125	243.000000

Explanation:

We are replacing all the null values if there are any with the mean values

Question 4:

```
#(4) Select at least two columns and aggregate the data using: min, max, count, mean.  
res = data_Manip.agg({'Calories':['mean','min','max','count'],'Pulse':['mean','min','max','count']})  
print(res)
```

	Calories	Pulse
mean	375.790244	107.461538
min	50.300000	80.000000
max	1860.400000	159.000000
count	169.000000	169.000000

Explanation:

We select calories pulse columns and getting mean, min. Max, count

Question 5:

```
# (5) Filter the dataframe to select the rows with calories values between 500 and 1000  
filter_data_Manip1=data_Manip[(data_Manip['Calories'] > 500) & (data_Manip['Calories'] < 1000)]  
print(filter_data_Manip1)
```

	Duration	Pulse	Maxpulse	Calories
51	80	123	146	643.1
62	160	109	135	853.0
65	180	90	130	800.4
66	150	105	135	873.4
67	150	107	130	816.0
72	90	100	127	700.0
73	150	97	127	953.2
75	90	98	125	563.2
78	120	100	130	500.4
90	180	101	127	600.1
99	90	93	124	604.1
103	90	90	100	500.4
106	180	90	120	800.3
108	90	90	120	500.3

Explanation:

Here we are selecting the rows between 500 and 1000 calories values

Question 6:

```
#(6) Filter the dataframe to select the rows with calories values > 500 and pulse < 100.  
filter_data_Manip2 = data_Manip[(data_Manip['Calories'] > 500)&(data_Manip['Pulse'] < 100)]  
print (filter_data_Manip2)
```

	Duration	Pulse	Maxpulse	Calories
65	180	90	130	800.4
70	150	97	129	1115.0
73	150	97	127	953.2
75	90	98	125	563.2
99	90	93	124	604.1
103	90	90	100	500.4
106	180	90	120	800.3
108	90	90	120	500.3

[+ Code](#)[+ Markdown](#)

Explanation:

We are selecting rows with calorie values less than 500 and greater than 1000

Question 7:

```
#(7)Create a new "f_modified" dataframe that contains all the columns from dst_data except for  
#"Maxpulse".  
data_modified = data_Manip.loc[:, data_Manip.columns != 'Maxpulse']  
print(data_modified)
```

	Duration	Pulse	Calories
0	60	110	409.1
1	60	117	479.0
2	60	103	340.0
3	45	109	282.4
4	45	117	406.0
..
164	60	105	290.8
165	60	110	300.0
166	60	115	310.2
167	75	120	320.4
168	75	125	330.4

[169 rows x 3 columns]

Explanation:

We are creating a file except for maxpulse

Question 8:

```
#(8)Delete the "Maxpulse" column from the main dst_data dataframe
data_Manip.drop('Maxpulse', inplace=True, axis=1)
print(data_Manip.dtypes)
```

```
Duration    int64
Pulse       int64
Calories    float64
dtype: object
```

+ Code

+ Markdown

Explanation:

We delete the maxpulse column from the data frame

Question 9:

```
 #(9)Convert the datatype of Calories column to int datatype.
data_Manip["Calories"] = data_Manip["Calories"].astype(float).astype(int)
print(data_Manip.dtypes)
```

```
Duration    int64
Pulse       int64
Calories    int64
dtype: object
```

+ Code

+ Markdown

Explanation:

We are converting the calories column to int datatype

Github Link:

<https://github.com/saideep8/MachineLearning/tree/main>

