Assignment :- 2

QN 1 . **Find all unique values of columns: “name”, “mfr”, “vitamins” and store them in separate numpy arrays and print them.**

```
import pandas as pd
import numpy as np

# Provided dataset
file_path = 'LabAssignment_Day2/Dataset_Day2.csv'
df = pd.read_csv(file_path)

# Extract unique values from the specified columns
unique_names = np.unique(df['name'])
unique_mfrs = np.unique(df['mfr'])
unique_vitamins = np.unique(df['vitamins'])

#print
print("uniqnue names " ,unique_names)
print("unique_mfrs " ,unique_mfrs)
print("unique_vitamins " ,unique_vitamins)
```

```
uniqnue names  ['100% Bran' '100% Natural Bran' 'All-Bran' 'All-Bran
with Extra Fiber'
 'Almond Delight' 'Apple Cinnamon Cheerios' 'Apple Jacks' 'Basic 4'
 'Bran Chex' 'Bran Flakes' "Cap'n'Crunch" 'Cheerios'
 'Cinnamon Toast Crunch' 'Clusters' 'Cocoa Puffs' 'Corn Chex'
 'Corn Flakes' 'Corn Pops' 'Count Chocula' "Cracklin' Oat Bran"
 'Cream of Wheat (Quick)' 'Crispix' 'Crispy Wheat & Raisins' 'Double
Chex'
 'Froot Loops' 'Frosted Flakes' 'Frosted Mini-Wheats'
 'Fruit & Fibre Dates; Walnuts; and Oats' 'Fruitful Bran' 'Fruity
Pebbles'
 'Golden Crisp' 'Golden Grahams' 'Grape Nuts Flakes' 'Grape-Nuts'
 'Great Grains Pecan' 'Honey Graham Ohs' 'Honey Nut Cheerios' 'Honey-
comb'
 'Just Right Crunchy  Nuggets' 'Just Right Fruit & Nut' 'Kix' 'Life'
 'Lucky Charms' 'Maypo' 'Muesli Raisins; Dates; & Almonds'
 'Muesli Raisins; Peaches; & Pecans' 'Mueslix Crispy Blend'
 'Multi-Grain Cheerios' 'Nut&Honey Crunch' 'Nutri-Grain Almond-Raisin'
 'Nutri-grain Wheat' 'Oatmeal Raisin Crisp' 'Post Nat. Raisin Bran'
 'Product 19' 'Puffed Rice' 'Puffed Wheat' 'Quaker Oat Squares'
 'Quaker Oatmeal' 'Raisin Bran' 'Raisin Nut Bran' 'Raisin Squares'
 'Rice Chex' 'Rice Krispies' 'Shredded Wheat' "Shredded Wheat 'n'Bran"
 'Shredded Wheat spoon size' 'Smacks' 'Special K'
 'Strawberry Fruit Wheats' 'Total Corn Flakes' 'Total Raisin Bran'
 'Total Whole Grain' 'Triples' 'Trix' 'Wheat Chex' 'Wheaties'
 'Wheaties Honey Gold']
```

```
unique_mfrs  ['A' 'G' 'K' 'N' 'P' 'Q' 'R']
unique_vitamins  [  0  25 100]
```

Q2.Create a new dataframe with all columns for which,'sodium' is greater than 100 AND 'protein' is less than 3 Name this dataframe: df_HighSodLowProt Print this dataframe.

```
# create a new dataframe from the provided qn "df_HighSodLowProt"

df_HighSodLowProt = df[(df['sodium'] > 100) & (df['protein'] < 3)]

# print the dataframe
print(df_HighSodLowProt)
```

```
                          name mfr type  calories  protein  fat
sodium  \
4               Almond Delight   R    C       110        2    2
200
5       Apple Cinnamon Cheerios   G    C       110        2    2
180
6                 Apple Jacks   K    C       110        2    0
125
8                   Bran Chex   R    C        90        2    1
200
10                Cap'n'Crunch   Q    C       120        1    2
220
12        Cinnamon Toast Crunch   G    C       120        1    3
210
14                  Cocoa Puffs   G    C       110        1    1
180
15                    Corn Chex   R    C       110        2    0
280
16                  Corn Flakes   K    C       100        2    0
290
18                Count Chocula   G    C       110        1    1
180
21                      Crispix   K    C       110        2    0
220
22       Crispy Wheat & Raisins   G    C       100        2    1
140
23                  Double Chex   R    C       100        2    0
190
24                  Froot Loops   K    C       110        2    1
125
25                Frosted Flakes   K    C       110        1    0
200
29                Fruity Pebbles   P    C       110        1    1
135
31                Golden Grahams   G    C       110        1    1
280
```

```
35              Honey Graham Ohs   Q    C       120         1    2
220
37                    Honey-comb   P    C       110         1    0
180
38  Just Right Crunchy  Nuggets   K    C       110         2    1
170
40                            Kix   G    C       110         2    1
260
42                  Lucky Charms   G    C       110         2    1
180
47          Multi-Grain Cheerios   G    C       100         2    1
220
48              Nut&Honey Crunch   K    C       120         2    1
190
61                     Rice Chex   R    C       110         1    0
240
62                 Rice Krispies   K    C       110         2    0
290
69             Total Corn Flakes   G    C       110         2    1
200
72                       Triples   G    C       110         2    1
250
73                          Trix   G    C       110         1    1
140
76           Wheaties Honey Gold   G    C       110         2    1
200

    fiber  carbo  sugars  potass  vitamins  shelf  weight  cups
rating
4     1.0   14.0       8      -1        25      3     1.0  0.75
34.384843
5     1.5   10.5      10      70        25      1     1.0  0.75
29.509541
6     1.0   11.0      14      30        25      2     1.0  1.00
33.174094
8     4.0   15.0       6     125        25      1     1.0  0.67
49.120253
10    0.0   12.0      12      35        25      2     1.0  0.75
18.042851
12    0.0   13.0       9      45        25      2     1.0  0.75
19.823573
14    0.0   12.0      13      55        25      2     1.0  1.00
22.736446
15    0.0   22.0       3      25        25      1     1.0  1.00
41.445019
16    1.0   21.0       2      35        25      1     1.0  1.00
45.863324
18    0.0   12.0      13      65        25      2     1.0  1.00
22.396513
```

```
21    1.0   21.0        3       30         25       3      1.0  1.00
46.895644
22    2.0   11.0       10      120         25       3      1.0  0.75
36.176196
23    1.0   18.0        5       80         25       3      1.0  0.75
44.330856
24    1.0   11.0       13       30         25       2      1.0  1.00
32.207582
25    1.0   14.0       11       25         25       1      1.0  0.75
31.435973
29    0.0   13.0       12       25         25       2      1.0  0.75
28.025765
31    0.0   15.0        9       45         25       2      1.0  0.75
23.804043
35    1.0   12.0       11       45         25       2      1.0  1.00
21.871292
37    0.0   14.0       11       35         25       1      1.0  1.33
28.742414
38    1.0   17.0        6       60        100       3      1.0  1.00
36.523683
40    0.0   21.0        3       40         25       2      1.0  1.50
39.241114
42    0.0   12.0       12       55         25       2      1.0  1.00
26.734515
47    2.0   15.0        6       90         25       1      1.0  1.00
40.105965
48    0.0   15.0        9       40         25       2      1.0  0.67
29.924285
61    0.0   23.0        2       30         25       1      1.0  1.13
41.998933
62    0.0   22.0        3       35         25       1      1.0  1.00
40.560159
69    0.0   21.0        3       35        100       3      1.0  1.00
38.839746
72    0.0   21.0        3       60         25       3      1.0  0.75
39.106174
73    0.0   13.0       12       25         25       2      1.0  1.00
27.753301
76    1.0   16.0        8       60         25       1      1.0  0.75
36.187559
```

3. From the dataframe in 2. print the average 'calories' by 'mfr' and also print the 'mfr' with the highest average 'calories'.

```
# Calculate the average calories
average_calories_by_mfr = df_HighSodLowProt.groupby('mfr')
['calories'].mean()

# manufacturer with the highest average calories
```

```
highest_average_calories_mfr = average_calories_by_mfr.idxmax()
highest_avg_calories = average_calories_by_mfr.max()

print(average_calories_by_mfr)
print(highest_avg_calories)
```

```
mfr
G    109.230769
K    110.000000
P    110.000000
Q    120.000000
R    104.000000
Name: calories, dtype: float64
120.0
```