

# EX-1 CHIPOTLE DATA SET

```
In [ ]: 1 import pandas as pd
        2 import matplotlib.pyplot as plt
        3 import numpy as np
```

```
In [2]: 1 # importing the dataset from the address
        2
        3 DF_CHIPO =pd.read_csv('C:\\Users\\GAYATRI\\Downloads\\chipotle.tsv' , del
```

```
In [3]: 1 # Assign it to the variable called CHIPO
        2
        3 DF_CHIPO
```

```
Out[3]:
```

	order_id	quantity	item_name	choice_description	item_price
0	1	1	Chips and Fresh Tomato Salsa	NaN	\$2.39
1	1	1	Izze	[Clementine]	\$3.39
2	1	1	Nantucket Nectar	[Apple]	\$3.39
3	1	1	Chips and Tomatillo-Green Chili Salsa	NaN	\$2.39
4	2	2	Chicken Bowl	[Tomatillo-Red Chili Salsa (Hot), [Black Beans...	\$16.98
...	...	...	...	...	...
4617	1833	1	Steak Burrito	[Fresh Tomato Salsa, [Rice, Black Beans, Sour ...	\$11.75
4618	1833	1	Steak Burrito	[Fresh Tomato Salsa, [Rice, Sour Cream, Cheese...	\$11.75
4619	1834	1	Chicken Salad Bowl	[Fresh Tomato Salsa, [Fajita Vegetables, Pinto...	\$11.25
4620	1834	1	Chicken Salad Bowl	[Fresh Tomato Salsa, [Fajita Vegetables, Lettu...	\$8.75
4621	1834	1	Chicken Salad Bowl	[Fresh Tomato Salsa, [Fajita Vegetables, Pinto...	\$8.75

4622 rows × 5 columns

```
In [8]: 1 # See the first 10 entries
        2
        3 DF_CHIPO.head(10)
```

```
Out[8]:
```

	order_id	quantity	item_name	choice_description	item_price
0	1	1	Chips and Fresh Tomato Salsa	NaN	\$2.39
1	1	1	Izze	[Clementine]	\$3.39
2	1	1	Nantucket Nectar	[Apple]	\$3.39
3	1	1	Chips and Tomatillo-Green Chili Salsa	NaN	\$2.39
4	2	2	Chicken Bowl	[Tomatillo-Red Chili Salsa (Hot), [Black Beans...	\$16.98
5	3	1	Chicken Bowl	[Fresh Tomato Salsa (Mild), [Rice, Cheese, Sou...	\$10.98
6	3	1	Side of Chips	NaN	\$1.69
7	4	1	Steak Burrito	[Tomatillo Red Chili Salsa, [Fajita Vegetables...	\$11.75
8	4	1	Steak Soft Tacos	[Tomatillo Green Chili Salsa, [Pinto Beans, Ch...	\$9.25
9	5	1	Steak Burrito	[Fresh Tomato Salsa, [Rice, Black Beans, Pinto...	\$9.25

```
In [23]: 1 # What is the number of observations in the dataset
        2
        3 print('The Shape of the DataFrame is : ',DF_CHIPO.shape); print('\n')
        4 print('Number of observations (rows) in DataFrame : ', DF_CHIPO.shape[0])
        5
```

The Shape of the DataFrame is : (4622, 5)

Number of observations (rows) in DataFrame : 4622

```
In [31]: 1 # What are the number of columns
        2 print('The number of columns in the DataFrame : ' ,DF_CHIPO.shape[1])
```

The number of columns in the DataFrame : 5

```
In [35]: 1 # Print the names of the columns
        2 print('The names of the columns in the DataFrame are : ', DF_CHIPO.columns)
```

The names of the columns in the DataFrame are : Index(['order\_id', 'quantity', 'item\_name', 'choice\_description', 'item\_price'], dtype='object')

```
In [26]: 1 # How is the dataset indexed
          2
          3 DF_CHIPO.index.name = 'Sr.No.'
          4
          5 DF_CHIPO
```

```
Out[26]:
```

	order_id	quantity	item_name	choice_description	item_price
Sr.No.					
0	1	1	Chips and Fresh Tomato Salsa	NaN	\$2.39
1	1	1	Izze	[Clementine]	\$3.39
2	1	1	Nantucket Nectar	[Apple]	\$3.39
3	1	1	Chips and Tomatillo-Green Chili Salsa	NaN	\$2.39
4	2	2	Chicken Bowl	[Tomatillo-Red Chili Salsa (Hot), [Black Beans...	\$16.98
...	...	...	...	...	...
4617	1833	1	Steak Burrito	[Fresh Tomato Salsa, [Rice, Black Beans, Sour ...	\$11.75
4618	1833	1	Steak Burrito	[Fresh Tomato Salsa, [Rice, Sour Cream, Cheese...	\$11.75
4619	1834	1	Chicken Salad Bowl	[Fresh Tomato Salsa, [Fajita Vegetables, Pinto...	\$11.25
4620	1834	1	Chicken Salad Bowl	[Fresh Tomato Salsa, [Fajita Vegetables, Lettu...	\$8.75
4621	1834	1	Chicken Salad Bowl	[Fresh Tomato Salsa, [Fajita Vegetables, Pinto...	\$8.75

4622 rows × 5 columns

In [89]:

```
1  # Which is the most ordered item ?
2  # For the most ordered item, how many items were ordered ?
3
4  ITEM_NAMES = DF_CHIPO['item_name']
5
6  print('ITEM_NAMES = ', ITEM_NAMES)
7
8  D= {}
9
10 for i in ITEM_NAMES:
11     if i in D :
12         D[i] +=1
13     else:
14         D[i] = 1
15
16 print('\n')
17 print("Dictionary : ") ; print(D); print('\n')
18 MAX_Value = max( D.values())
19
20 def GET_KEY(DICT):
21     for key in DICT:
22         if DICT[key] == max(DICT.values()) :
23             MAX_VALUE_KEY = key
24     return(MAX_VALUE_KEY)
25
26
27 KEY_WITH_MAX_VALUE = GET_KEY(D)
28
29 print(f'The {KEY_WITH_MAX_VALUE} is the item which has been ordered for
30       f'of times which is {MAX_Value} times ' )
```

```

ITEM_NAMES = Sr.No.
0           Chips and Fresh Tomato Salsa
1                               Izze
2                               Nantucket Nectar
3           Chips and Tomatillo-Green Chili Salsa
4           Chicken Bowl
...
4617          Steak Burrito
4618          Steak Burrito
4619          Chicken Salad Bowl
4620          Chicken Salad Bowl
4621          Chicken Salad Bowl
Name: item_name, Length: 4622, dtype: object

```

Dictionary :

```

{'Chips and Fresh Tomato Salsa': 110, 'Izze': 20, 'Nantucket Nectar': 27, 'Chips and Tomatillo-Green Chili Salsa': 31, 'Chicken Bowl': 726, 'Side of Chips': 101, 'Steak Burrito': 368, 'Steak Soft Tacos': 55, 'Chips and Guacamole': 479, 'Chicken Crispy Tacos': 47, 'Chicken Soft Tacos': 115, 'Chicken Burrito': 553, 'Canned Soda': 104, 'Barbacoa Burrito': 91, 'Carnitas Burrito': 59, 'Carnitas Bowl': 68, 'Bottled Water': 162, 'Chips and Tomatillo Green Chili Salsa': 43, 'Barbacoa Bowl': 66, 'Chips': 211, 'Chicken Salad Bowl': 110, 'Steak Bowl': 211, 'Barbacoa Soft Tacos': 25, 'Veggie Burrito': 95, 'Veggie Bowl': 85, 'Steak Crispy Tacos': 35, 'Chips and Tomatillo Red Chili Salsa': 48, 'Barbacoa Crispy Tacos': 11, 'Veggie Salad Bowl': 18, 'Chips and Roasted Chili-Corn Salsa': 18, 'Chips and Roasted Chili Corn Salsa': 22, 'Carnitas Soft Tacos': 40, 'Chicken Salad': 9, 'Canned Soft Drink': 301, 'Steak Salad Bowl': 29, '6 Pack Soft Drink': 54, 'Chips and Tomatillo-Red Chili Salsa': 20, 'Bowl': 2, 'Burrito': 6, 'Crispy Tacos': 2, 'Carnitas Crispy Tacos': 7, 'Steak Salad': 4, 'Chips and Mild Fresh Tomato Salsa': 1, 'Veggie Soft Tacos': 7, 'Carnitas Salad Bowl': 6, 'Barbacoa Salad Bowl': 10, 'Salad': 2, 'Veggie Crispy Tacos': 1, 'Veggie Salad': 6, 'Carnitas Salad': 1}

```

The Chicken Bowl is the item which has been ordered for the maximum number of times which is 726 times

```

In [75]: 1 # What was the most ordered item in the choice_description column?
          2
          3 from collections import Counter as C

```

```

In [40]: 1 # How many items were ordered in total
          2
          3 TOTAL_ORDERS = sum(DF_CHIPO['quantity'])
          4 print('Total Orders in the data set are :', TOTAL_ORDERS)

```

Total Orders in the data set are : 4972

```
In [60]: 1 # turn the item price in float
2
3 ITEM_PRICE = DF_CHIPO['item_price'].replace(['\$'], '', regex = True ).as
4 print(ITEM_PRICE)
```

```
0      2.39
1      3.39
2      3.39
3      2.39
4     16.98
...
4617   11.75
4618   11.75
4619   11.25
4620    8.75
4621    8.75
Name: item_price, Length: 4622, dtype: float64
```

```
In [67]: 1 # check the item_price type
2
3 DF_CHIPO['item_price']
4
5 x = type( DF_CHIPO['item_price'] )
6
7 print(f'the datatype of the column is {x}')
8
```

the datatype of the column is <class 'pandas.core.series.Series'>

```
In [76]: 1 # create a lambda function and change the type of item price
2
3 xx = DF_CHIPO['item_price'].map(lambda A : float(A.replace('$', ''))
4 # DF_CHIPO['item_price'] = DF_CHIPO['item_price'].apply(lambda x: float(x)
5
6 print(xx)
```

```
0      2.39
1      3.39
2      3.39
3      2.39
4     16.98
...
4617   11.75
4618   11.75
4619   11.25
4620    8.75
4621    8.75
Name: item_price, Length: 4622, dtype: float64
```

```
In [14]: 1 # check the item price column
2
3 a = type(DF_CHIPO['item_price'])
4 print(f' The type of the column is {a}')
5
```

The type of the column is <class 'pandas.core.series.Series'>

```
In [39]: 1 # How much was the revenue for the period in the dataset
2
3
4 DF_CHIPO['Amount'] = None # created a New Column with Null values
5
6 # Assigning the values to the Amount Column by column operations
7 DF_CHIPO['Amount'] = DF_CHIPO['quantity'].astype(float) * DF_CHIPO['item_
8
9 print(DF_CHIPO)
10
11 TOTAL_REVENUE = sum(DF_CHIPO['Amount'])
12
13 print('TOTAL_REVENUE = ', TOTAL_REVENUE)
```

	order_id	quantity	item_name \
0	1	1	Chips and Fresh Tomato Salsa
1	1	1	Izze
2	1	1	Nantucket Nectar
3	1	1	Chips and Tomatillo-Green Chili Salsa
4	2	2	Chicken Bowl
...	...	...	...
4617	1833	1	Steak Burrito
4618	1833	1	Steak Burrito
4619	1834	1	Chicken Salad Bowl
4620	1834	1	Chicken Salad Bowl
4621	1834	1	Chicken Salad Bowl

	choice_description	item_price	Amount
0	NaN	\$2.39	2.39
1	[Clementine]	\$3.39	3.39
2	[Apple]	\$3.39	3.39
3	NaN	\$2.39	2.39
4	[Tomatillo-Red Chili Salsa (Hot), [Black Beans...	\$16.98	33.96
...	...	...	...
4617	[Fresh Tomato Salsa, [Rice, Black Beans, Sour ...	\$11.75	11.75
4618	[Fresh Tomato Salsa, [Rice, Sour Cream, Cheese...	\$11.75	11.75
4619	[Fresh Tomato Salsa, [Fajita Vegetables, Pinto...	\$11.25	11.25
4620	[Fresh Tomato Salsa, [Fajita Vegetables, Lettu...	\$8.75	8.75
4621	[Fresh Tomato Salsa, [Fajita Vegetables, Pinto...	\$8.75	8.75

[4622 rows x 6 columns]  
TOTAL\_REVENUE = 39237.020000000055

```
In [45]: 1 # How many orders were made in this period
2
3 DF_CHIPO
4
5
6 TOTAL_ORDERS = sum( DF_CHIPO['quantity'] )
7
8 print( ' Total number of orders are ', TOTAL_ORDERS)
```

Total number of orders are 4972

```
In [61]: 1 # What is the average revenue per order
2
3 Average_Amount_for_order = (sum(DF_CHIPO['Amount']) / len(DF_CHIPO['Amount']
4
5 print('The average Amount of order is : $' , Average_Amount_for_order )
```

The average Amount of order is : \$ 8.489186499350943

```
In [74]: 1 # How many different items are sold
2
3 DF_CHIPO
4
5 UNIQUE_ITEMS = (DF_CHIPO['item_name'].unique())
6
7 print('Different Items ordered are : \n', UNIQUE_ITEMS )
```

Different Items ordered are :

```
['Chips and Fresh Tomato Salsa' 'Izze' 'Nantucket Nectar'
'Chips and Tomatillo-Green Chili Salsa' 'Chicken Bowl' 'Side of Chips'
'Steak Burrito' 'Steak Soft Tacos' 'Chips and Guacamole'
'Chicken Crispy Tacos' 'Chicken Soft Tacos' 'Chicken Burrito'
'Canned Soda' 'Barbacoa Burrito' 'Carnitas Burrito' 'Carnitas Bowl'
'Bottled Water' 'Chips and Tomatillo Green Chili Salsa' 'Barbacoa Bowl'
'Chips' 'Chicken Salad Bowl' 'Steak Bowl' 'Barbacoa Soft Tacos'
'Veggie Burrito' 'Veggie Bowl' 'Steak Crispy Tacos'
'Chips and Tomatillo Red Chili Salsa' 'Barbacoa Crispy Tacos'
'Veggie Salad Bowl' 'Chips and Roasted Chili-Corn Salsa'
'Chips and Roasted Chili Corn Salsa' 'Carnitas Soft Tacos'
'Chicken Salad' 'Canned Soft Drink' 'Steak Salad Bowl'
'6 Pack Soft Drink' 'Chips and Tomatillo-Red Chili Salsa' 'Bowl'
'Burrito' 'Crispy Tacos' 'Carnitas Crispy Tacos' 'Steak Salad'
'Chips and Mild Fresh Tomato Salsa' 'Veggie Soft Tacos'
'Carnitas Salad Bowl' 'Barbacoa Salad Bowl' 'Salad' 'Veggie Crispy Tacos'
'Veggie Salad' 'Carnitas Salad']
```

## EX- 2 HEART DISEASE DATASET



```
In [1]: 1 import pandas as pd
        2 import numpy as np
```

```
In [2]: 1
        2
        3 DF_HEART_DISEASE = pd.read_csv('C:\\Users\\GAYATRI\\Downloads\\heart_disea
        4
        5
        6 REINDEX = np.arange(1,1001)
        7 DF_HEART_DISEASE.reindex(REINDEX)
        8 DF_HEART_DISEASE.index.name = 'Sr.No.'
        9
        10 DF_HEART_DISEASE
```

Out[2]:

	Age	Gender	Cholesterol	Blood Pressure	Heart Rate	Smoking	Alcohol Intake	Exercise Hours	Family History	Diabetes
Sr.No.										
0	75	Female	228	119	66	Current	Heavy	1	No	
1	48	Male	204	165	62	Current	NaN	5	No	
2	53	Male	234	91	67	Never	Heavy	3	Yes	
3	69	Female	192	90	72	Current	NaN	4	No	Y
4	62	Female	172	163	93	Never	NaN	6	No	Y
...	...	...	...	...	...	...	...	...	...	...
995	56	Female	269	111	86	Never	Heavy	5	No	Y
996	78	Female	334	145	76	Never	NaN	6	No	
997	79	Male	151	179	81	Never	Moderate	4	Yes	
998	60	Female	326	151	68	Former	NaN	8	Yes	Y
999	53	Male	226	116	82	Current	NaN	6	No	

1000 rows × 16 columns



```

In [23]: 1 # How many columns and records are present in the DataFrame are there in
2
3
4 Column_Name_list = DF_HEART_DISEASE.columns
5 print('Following are the columns in the Dataset DF_HEART_DISEASE :\t ')
6 print('The Columns in the dataset are: \n ',Column_Name_list)
7 print('\n')
8 print('METHOD#1 ');
9 Columns_count = DF_HEART_DISEASE.shape[1]
10 Records_count = DF_HEART_DISEASE.shape[0]
11 print('The number of Columns in the Dataset are : ' , Columns_count);
12 print('The number of records in the Dataset are : ' , Records_count); print
13
14 print('METHOD#2')
15
16 Count_Columns = len(DF_HEART_DISEASE.columns)
17 Count_Records = len(DF_HEART_DISEASE)
18 print('The number of Columns in the Dataset are : ' , Count_Columns)
19 print('The number of records in the Dataset are : ' , Count_Records); print

```

Following are the columns in the Dataset DF\_HEART\_DISEASE :

The Columns in the dataset are:

```

Index(['Age', 'Gender', 'Cholesterol', 'Blood Pressure', 'Heart Rate',
       'Smoking', 'Alcohol Intake', 'Exercise Hours', 'Family History',
       'Diabetes', 'Obesity', 'Stress Level', 'Blood Sugar',
       'Exercise Induced Angina', 'Chest Pain Type', 'Heart Disease'],
      dtype='object')

```

METHOD#1

The number of Columns in the Dataset are : 16

The number of records in the Dataset are : 1000

METHOD#2

The number of Columns in the Dataset are : 16

The number of records in the Dataset are : 1000

```

In [28]: 1 # Average age of the Diabetes patients
2
3 AVERAGE_AGE = np.mean(DF_HEART_DISEASE['Age'])
4
5 print('The average age of the diabetes patient is ', AVERAGE_AGE)

```

The average age of the diabetes patient is 52.293

In [6]:

```
1  # Display the records of people who are Heavy consumers of alchohol and ar
2
3  Heavy_Drinkers_Current_Smokers = DF_HEART_DISEASE[ (DF_HEART_DISEASE['Alc
4                                                    (DF_HEART_DISEASE['Smok
5  print('Patients who are heavy drinkers and currently smoking are as follow
6  print(f' Hence , there are {Heavy_Drinkers_Current_Smokers.shape[0]} pati
7
```

Patients who are heavy drinkers and currently smoking are as follows:

Sr.No.	Age	Gender	Cholesterol	Blood Pressure	Heart Rate	Smoking
0	75	Female	228	119	66	Current
8	37	Female	317	137	66	Current
13	43	Male	155	169	82	Current
46	51	Male	251	170	71	Current
48	41	Male	268	95	70	Current
...	...	...	...	...	...	...
954	33	Female	198	178	67	Current
957	34	Male	186	134	82	Current
969	50	Female	281	120	92	Current
977	63	Female	187	178	69	Current
978	25	Female	305	100	76	Current

Sr.No.	Alcohol Intake	Exercise Hours	Family History	Diabetes	Obesity
0	Heavy	1	No	No	Yes
8	Heavy	3	No	Yes	Yes
13	Heavy	8	Yes	Yes	No
46	Heavy	3	No	Yes	No
48	Heavy	9	No	No	No
...	...	...	...	...	...
954	Heavy	3	No	No	Yes
957	Heavy	5	Yes	No	No
969	Heavy	7	No	Yes	Yes
977	Heavy	4	No	No	No
978	Heavy	2	Yes	No	No

Sr.No.	Stress Level	Blood Sugar	Exercise Induced Angina	Chest Pain Type
0	8	119	Yes	Atypical Angina
8	5	114	No	Non-anginal Pain
13	2	163	No	Typical Angina
46	4	111	No	Non-anginal Pain
48	4	176	No	Non-anginal Pain
...	...	...	...	...
954	3	97	No	Non-anginal Pain
957	1	132	No	Atypical Angina
969	6	106	Yes	Non-anginal Pain
977	5	128	No	Non-anginal Pain
978	9	107	No	Asymptomatic

Sr.No.	Heart Disease
0	1
8	0
13	0
46	1
48	0
...	...
954	0
957	0
969	0
977	0
978	0

[114 rows x 16 columns]

Hence , there are 114 patients who are heavy consumers of alchohol and are current smokers

```
In [5]: 1 # CATEGORIZE THE MEN IN 'NORMAL' , 'BORDERLINE' AND 'HIGH' CHOLESTROL LE
2 NORMAL_CHOLESTROL_MALE = DF_HEART_DISEASE[ (DF_HEART_DISEASE['Cholesterol']
3
4 BORDERLINE_CHOLESTROL_MALE = DF_HEART_DISEASE[ ( (DF_HEART_DISEASE['Choles
5
6 HIGH_CHOLESTROL_MALE = DF_HEART_DISEASE[ (DF_HEART_DISEASE['Cholesterol']
7
8
9 print('CATEGORIZING THE PATIENTS BASED ON THEIR CHOLESTROL LEVELS : \n');
10
11 print('Normal Cholestrol Levels are :\n') ;print(NORMAL_CHOLESTROL_MALE) ;
12 print('Borderline Cholestrol Levels are :\n') ;print(BORDERLINE_CHOLESTROL
13 print('High Cholestrol Levels are :\n') ;print(HIGH_CHOLESTROL_MALE) ; pr
```

CATEGORIZING THE PATIENTS BASED ON THEIR CHOLESTROL LEVELS :

Normal Cholestrol Levels are :

Sr.No.	Age	Gender	Cholesterol	Blood Pressure	Heart Rate	Smoking \
13	43	Male	155	169	82	Current
31	48	Male	179	117	61	Never
38	46	Male	192	139	61	Former
41	65	Male	189	143	70	Current
62	42	Male	166	119	97	Never
...	...	...	...	...	...	...
939	65	Male	151	107	64	Former
942	42	Male	156	111	95	Current
957	34	Male	186	134	82	Current
958	37	Male	162	130	78	Former
997	79	Male	151	179	81	Never

Alcohol Intake Exercise Hours Family History Diabetes Obesity \

```
In [16]: 1 # Find out the men with High risk of Heartattack i.e.  
2         # with High cholestrol  
3         # with Diabetese  
4         # with diagnosed heart disease  
5  
6 MALE_HEARTDISEASE_HIGH_CHOLESTROL = DF_HEART_DISEASE[(DF_HEART_DISEASE['He  
7                                                         (DF_HEART_DISEASE['CH  
8                                                         ( DF_HEART_DISEASE['C  
9                                                         (DF_HEART_DISEASE['D  
10                                                    ]  
11  
12 print('Men who have 3 health conditions: \n ' + 'a)Diabetes, \n b)High Cho  
13 print(MALE_HEARTDISEASE_HIGH_CHOLESTROL)  
14
```

Men who have 3 health conditions:

- a)Diabetes,
- b)High Cholestrol,
- c)Diagnosed Heart Disease are:

Sr.No.	Age	Gender	Cholesterol	Blood Pressure	Heart Rate	Smoking \
5	77	Male	309	110	73	Never
23	67	Male	287	117	64	Current
27	63	Male	273	103	64	Never
46	51	Male	251	170	71	Current
58	61	Male	261	156	97	Former
...	...	...	...	...	...	...
919	55	Male	300	121	99	Never
922	54	Male	294	168	63	Never
929	57	Male	320	99	83	Former
976	74	Male	255	112	99	Former
994	52	Male	248	159	76	Former

Sr.No.	Alcohol Intake	Exercise Hours	Family History	Diabetes	Obesity \
5	NaN	0	No	Yes	Yes
23	Moderate	0	Yes	Yes	Yes
27	NaN	6	Yes	Yes	Yes
46	Heavy	3	No	Yes	No
58	Heavy	2	Yes	Yes	Yes
...	...	...	...	...	...
919	Heavy	2	Yes	Yes	No
922	Heavy	0	No	Yes	Yes
929	Heavy	8	Yes	Yes	No
976	Moderate	8	Yes	Yes	No
994	Moderate	9	No	Yes	Yes

Sr.No.	Stress Level	Blood Sugar	Exercise Induced Angina	Chest Pain Type
5	4	122	Yes	Asymptomatic
23	1	112	Yes	Atypical Angina
27	9	104	No	Typical Angina
46	4	111	No	Non-anginal Pain
58	7	154	Yes	Atypical Angina
...	...	...	...	...
919	10	116	No	Atypical Angina
922	2	175	No	Asymptomatic
929	9	184	No	Asymptomatic
976	8	84	No	Non-anginal Pain
994	2	152	Yes	Asymptomatic

Sr.No.	Heart Disease
5	1
23	1
27	1
46	1
58	1
...	...
919	1

922	1
929	1
976	1
994	1

[76 rows x 16 columns]

In [25]:

```
1 # Who among men or women have higher rates of Diabetes,
2
3
4 DIABETESE_FEMALE = DF_HEART_DISEASE[ (DF_HEART_DISEASE['Gender'] == 'Female')]
5 DIABETESE_MALE = DF_HEART_DISEASE[ (DF_HEART_DISEASE['Gender'] == 'Male')]
6
7 print(f'Number of women who have diabetes are {DIABETESE_FEMALE.shape[0]}')
8 print(f'Number of men who have diabetes are {DIABETESE_MALE.shape[0]}')
9 print('\n')
10 if DIABETESE_FEMALE.shape[0] > DIABETESE_MALE.shape[0]:
11     print('in the dataset more women have diabetes than men')
12 else:
13     print('in the dataset more men have diabetes than women')
14
```

Number of women who have diabetes are 264

Number of men who have diabetes are 241

in the dataset more women have diabetes than men



In [38]:

```
1
2
3 NORMAL_BP_NORMAL_SUGAR_FEMALE = DF_HEART_DISEASE[( DF_HEART_DISEASE['Blood
4                                                     (DF_HEART_DISEASE['Blood
5                                                     (DF_HEART_DISEASE['Alcohol
6                                                     ((DF_HEART_DISEASE['Smok
7                                                     (DF_HEART_DISEASE['Heart
8                                                     (DF_HEART_DISEASE['Gender
9                                                     ]
10 print(f'Women who have HEART DISEASES with following health parameters :
11 print(f' Normal Blood Sugar Levels , Normal Blood Pressure, Moderate Alcoh
12 print(NORMAL_BP_NORMAL_SUGAR_FEMALE)
13
14 n = len(NORMAL_BP_NORMAL_SUGAR_FEMALE)
15 print(f'Total number of women with diagnosed HEART DISEASE with above fact
```

Women who have HEART DISEASES with following health parameters :

Normal Blood Sugar Levels , Normal Blood Pressure, Moderate Alcohol Intake , Never or former smoking

Sr.No.	Age	Gender	Cholesterol	Blood Pressure	Heart Rate	Smoking \
99	61	Female	248	100	91	Former
146	57	Female	248	112	90	Never
185	54	Female	238	102	63	Former
376	79	Female	331	117	93	Never
410	58	Female	349	96	62	Former
764	57	Female	288	104	71	Never
824	54	Female	313	108	79	Never
825	63	Female	284	120	71	Never

Sr.No.	Alcohol Intake	Exercise Hours	Family History	Diabetes	Obesity \
99	Moderate	3	Yes	Yes	No
146	Moderate	1	No	Yes	No
185	Moderate	9	Yes	Yes	No
376	Moderate	9	Yes	Yes	No
410	Moderate	7	No	Yes	Yes
764	Moderate	6	No	Yes	Yes
824	Moderate	7	No	No	Yes
825	Moderate	7	No	No	No

Sr.No.	Stress Level	Blood Sugar	Exercise Induced Angina	Chest Pain Type
99	3	110	No	Atypical Angina
146	2	90	No	Atypical Angina
185	1	85	No	Typical Angina
376	1	108	Yes	Atypical Angina
410	8	106	No	Asymptomatic
764	2	81	No	Typical Angina
824	8	122	No	Non-anginal Pain
825	2	74	No	Typical Angina

Sr.No.	Heart Disease
99	1
146	1
185	1
376	1
410	1
764	1
824	1
825	1

Total number of women with diagnosed HEART DISEASE with above factors are 8

```

In [54]: 1 # Give the records of the healthy people who
2 #       - exercise for 3 hrs ,
3 #       - do not have diabetes
4 #       - never smoked
5 #       - Have healthy heart Rate
6
7 x = DF_HEART_DISEASE[
8         (DF_HEART_DISEASE['Exercise Hours'] == 3) &
9         (DF_HEART_DISEASE['Diabetes'] == 'No' ) &
10        (DF_HEART_DISEASE['Smoking'] == 'Never' ) &
11        (DF_HEART_DISEASE['Heart Rate'] <=75 )
12
13        ]
14
15 x

```

Out[54]:

	Age	Gender	Cholesterol	Blood Pressure	Heart Rate	Smoking	Alcohol Intake	Exercise Hours	Family History	Diabetes
<b>Sr.No.</b>										
2	53	Male	234	91	67	Never	Heavy	3	Yes	
244	63	Male	250	154	69	Never	NaN	3	Yes	
661	48	Female	257	148	65	Never	Moderate	3	No	
706	75	Female	201	121	71	Never	NaN	3	No	
847	79	Male	158	110	73	Never	Moderate	3	No	
885	28	Male	283	139	65	Never	Moderate	3	No	

```
In [86]: 1 # What is the average Blood Pressure of the male patients with heart disease
2
3
4 DF1 = DF_HEART_DISEASE[
5         (DF_HEART_DISEASE['Heart Disease'] == 1) &
6         (DF_HEART_DISEASE['Diabetes'] == 'Yes')
7     ]
8
9 DF1
10
11 BP_MEAN_DIABETICS_HEARTDISEASE = np.mean(DF1['Blood Pressure'])
12 BP_MEAN_DIABETICS_HEARTDISEASE_ = round(BP_MEAN_DIABETICS_HEARTDISEASE , 2 )
13
14
15
16 print(f'Heart patients with Diabetese normally have an average of {BP_MEAN_DIABETICS_HEARTDISEASE_} Blood pressure')
17
```

Heart patients with Diabetese normally have an average of 135.35 Blood pressure

In [112]:

```
1  # Give records of Female Senior Citizen patients with no Diabetes and no
2  #and either smoked in the past or never smoked
3  # and Moderate or no alcohol intake
4
5  DF1 = DF_HEART_DISEASE[( DF_HEART_DISEASE['Gender'] == 'Female')&
6                          ( ( DF_HEART_DISEASE['Age']>=60) & ( DF_HEART_DISEASE['Diabetes'] == 'No') &
7                          ( DF_HEART_DISEASE['Heart Disease'] == 0 ) &
8                          ( (DF_HEART_DISEASE['Smoking'] == 'Never') | (DF_HEART_DISEASE['Cholesterol'] <= 200)&
9                          ( (DF_HEART_DISEASE['Alcohol Intake'] == 'Nan' )
10
11
12                          )
13
14  print(DF1); print('\n')
15
16  print(f'There are {DF1.shape[0]} senior citizen ladies who have \n\t - M
```

Sr.No.	Age	Gender	Cholesterol	Blood Pressure	Heart Rate	Smoking \
123	68	Female	185	158	78	Never
192	73	Female	184	104	60	Former
305	76	Female	159	96	65	Never
359	69	Female	188	179	97	Former
502	64	Female	191	178	75	Never
812	62	Female	198	133	93	Former

Sr.No.	Alcohol Intake	Exercise Hours	Family History	Diabetes	Obesity \
123	Moderate	0	No	No	Yes
192	Moderate	1	Yes	No	Yes
305	Moderate	1	No	No	No
359	Moderate	3	Yes	No	Yes
502	Moderate	9	Yes	No	No
812	Moderate	2	No	No	No

Sr.No.	Stress Level	Blood Sugar	Exercise Induced Angina	Chest Pain Type
123	9	90	Yes	Non-anginal Pain
192	7	196	No	Non-anginal Pain
305	2	142	No	Non-anginal Pain
359	1	169	Yes	Typical Angina
502	5	97	No	Asymptomatic
812	2	78	Yes	Atypical Angina

Sr.No.	Heart Disease
123	0
192	0
305	0
359	0
502	0
812	0

There are 6 senior citizen ladies with

- No diabetes
- No Heart disease
- not smoking currently
- having normal cholesterol level
- consuming moderate or no alcohol

In [ ]:

1