HEART DISEASE PREDICTION

OBJECTIVE:

TO PERFORM THE EXPLORATORY DATA ANALYSIS IN THE DATASET TO HELP LOOK AT DATA BEFORE MAKING ANY ASSUMPTIONS. IT CAN HELP IDENTIFY OBVIOUS ERRORS, AS WELL AS BETTER UNDERSTAND PATTERNS WITHIN DATA, DETECT OUTLIERS OR ANOMALOUS EVENTS, FIND INTERESTING RELATIONS AMONG THE VARIABLES IN THE DATASET AND INSIGHTS ABOUT THE FACTORS AFFECTING THE HEART ATTACK RISK.

ABOUT THE DATASET:

THIS DATASET IS USED TO PREDICT HEART DIESASE. PATIENTS ARE LIKELY TO BE DIAGNOSED WITH ANY CARDIOVASCULAR HEART DISEASES BASED ON THEIR MEDICAL ATTRIBUTES SUCH AS AGE, BLOOD PRESSURE, BMI, CHOLESTEROL,

TRIGLYCERIADS, SMOKING, ALCOHOL CONSUMPTION, STRESS LEVEL, DIABETICS, HEART

RATE etc.

EXPLORATORY ANALYSIS

```
In [108]: import numpy as np
          import pandas as pd
          import matplotlib.pyplot as  plt
          import seaborn as sb
          import warnings
          warnings.filterwarnings('ignore')
          %matplotlib inline
```

```
In [3]: data = pd.read_csv(r"Documents\heart_attack_prediction_dataset.csv")
```

```
In [4]: data
```

Out[4]:

	Patient ID	Age	Sex	Cholesterol	Blood Pressure	Heart Rate	Diabetes	Family History	Smoking	Obesity	...	Sedentary Hours Per Day	Income	BMI	Triglycerides	Physical Activity Days Per Week	S
0	BMW7812	67.0	Male	208	158/88	72	0	0	1	0	...	6.615801	261404	31.251233	286	0	
1	CZE1114	21.0	Male	389	165/93	98	1	1	1	1	...	4.963459	285768	27.194973	235	1	
2	BNB906	21.0	Female	324	174/99	72	1	0	0	0	...	9.463426	235282	28.176571	587	4	
3	JLN3497	84.0	Male	383	163/100	73	1	1	1	0	...	7.648981	125640	36.464704	378	3	
4	GFO8847	66.0	Male	318	91/88	93	1	1	1	1	...	1.514821	160555	21.809144	231	1	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
8758	MSV9918	60.0	Male	121	94/76	61	1	1	1	0	...	10.806373	235429	19.655895	67	7	
8759	GSV6764	28.0	Female	120	157/102	73	1	0	0	1	...	3.833636	217681	23.993866	617	4	
8760	XKA5925	47.0	Male	250	161/75	105	0	1	1	1	...	2.375214	36998	35.406146	527	4	
8761	EPE6801	36.0	Male	178	119/67	60	1	0	1	0	...	0.829184	299943	27.294820	114	2	
8762	ZWN9666	25.0	Female	356	136/67	75	1	1	0	0	...	9.805234	247538	32.914151	189	7	

```
In [6]: data.head()
```

Out[6]:

	Patient ID	Age	Sex	Cholesterol	Blood Pressure	Heart Rate	Diabetes	Family History	Smoking	Obesity	...	Sedentary Hours Per Day	Income	BMI	Triglycerides	Physical Activity Days Per Week	Slee Hour Pe Da
0	BMW7812	67.0	Male	208	158/88	72	0	0	1	0	...	6.615801	261404	31.251233	286	0	
1	CZE1114	21.0	Male	389	165/93	98	1	1	1	1	...	4.963459	285768	27.194973	235	1	
2	BNB906	21.0	Female	324	174/99	72	1	0	0	0	...	9.463426	235282	28.176571	587	4	
3	JLN3497	84.0	Male	383	163/100	73	1	1	1	0	...	7.648981	125640	36.464704	378	3	
4	GFO8847	66.0	Male	318	91/88	93	1	1	1	1	...	1.514821	160555	21.809144	231	1	

5 rows × 26 columns

In [8]:
```
data.tail()
```

Out[8]:

	Patient ID	Age	Sex	Cholesterol	Blood Pressure	Heart Rate	Diabetes	Family History	Smoking	Obesity	...	Sedentary Hours Per Day	Income	BMI	Triglycerides	Physical Activity Days Per Week	Sl Hc
8758	MSV9918	60.0	Male	121	94/76	61	1	1	1	0	...	10.886373	235420	19.655895	67	7	
8759	QSV6764	28.0	Female	120	157/102	73	1	0	0	1	...	3.833038	217881	23.993866	617	4	
8760	XKA592S	47.0	Male	250	161/75	105	0	1	1	1	...	2.375214	36998	35.406146	527	4	
8761	EPE6801	36.0	Male	178	119/67	60	1	0	1	0	...	0.029104	209943	27.294329	114	2	
8762	ZWN9666	25.0	Female	356	138/67	75	1	1	0	0	...	9.065234	247338	32.914151	550	7	

5 rows × 26 columns

In [40]:
```
data.shape
```

Out[40]: (8763, 22)

In [48]:
```
data.dtypes
```

Out[48]:
```
Age                             float64
Sex                              object
Cholesterol                       int64
Blood Pressure                   object
Heart Rate                        int64
Diabetes                          int64
Family History                    int64
Smoking                           int64
Obesity                           int64
Alcohol Consumption               int64
Exercise Hours Per Week         float64
Diet                             object
Previous Heart Problems           int64
Medication Use                    int64
Stress Level                      int64
Sedentary Hours Per Day         float64
Income                            int64
BMI                             float64
Triglycerides                     int64
Physical Activity Days Per Week   int64
Sleep Hours Per Day               int64
Heart Attack Risk                 int64
dtype: object
```

In [9]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8763 entries, 0 to 8762
Data columns (total 26 columns):
 #   Column                           Non-Null Count  Dtype
---  ------                           --------------  -----
 0   Patient ID                       8763 non-null   object
 1   Age                              8762 non-null   float64
 2   Sex                              8763 non-null   object
 3   Cholesterol                      8763 non-null   int64
 4   Blood Pressure                   8763 non-null   object
 5   Heart Rate                       8763 non-null   int64
 6   Diabetes                         8763 non-null   int64
 7   Family History                   8763 non-null   int64
 8   Smoking                          8763 non-null   int64
 9   Obesity                          8763 non-null   int64
 10  Alcohol Consumption              8763 non-null   int64
 11  Exercise Hours Per Week          8763 non-null   float64
 12  Diet                             8763 non-null   object
 13  Previous Heart Problems          8763 non-null   int64
 14  Medication Use                   8763 non-null   int64
 15  Stress Level                     8763 non-null   int64
 16  Sedentary Hours Per Day          8763 non-null   float64
 17  Income                           8763 non-null   int64
 18  BMI                              8763 non-null   float64
 19  Triglycerides                    8763 non-null   int64
 20  Physical Activity Days Per Week  8763 non-null   int64
 21  Sleep Hours Per Day              8763 non-null   int64
 22  Country                          8763 non-null   object
```

In [40]:
```
data.count()
```

Out[40]:
```
Age                                8763
Sex                                8763
Cholesterol                        8763
Blood Pressure                     8763
Heart Rate                         8763
Diabetes                           8763
Family History                     8763
Smoking                            8763
Obesity                            8763
Alcohol Consumption                8763
Exercise Hours Per Week            8763
Diet                               8763
Previous Heart Problems            8763
Medication Use                     8763
Stress Level                       8763
Sedentary Hours Per Day            8763
Income                             8763
BMI                                8763
Triglycerides                      8763
Physical Activity Days Per Week    8763
Sleep Hours Per Day                8763
Heart Attack Risk                  8763
dtype: int64
```

In [41]:
```
data.columns.tolist()
```

Out[41]:
```
['Age',
 'Sex',
 'Cholesterol',
 'Blood Pressure',
 'Heart Rate',
 'Diabetes',
 'Family History',
 'Smoking',
 'Obesity',
 'Alcohol Consumption',
 'Exercise Hours Per Week',
 'Diet',
 'Previous Heart Problems',
 'Medication Use',
 'Stress Level',
 'Sedentary Hours Per Day',
 'Income',
 'BMI',
 'Triglycerides',
 'Physical Activity Days Per Week',
 'Sleep Hours Per Day',
 'Heart Attack Risk']
```

In [10]:
```
data.nunique()
```

Out[10]:
```
Patient ID                         8763
Age                                  73
Sex                                   2
Cholesterol                         281
Blood Pressure                     3915
Heart Rate                           71
Diabetes                              2
Family History                        2
Smoking                               2
Obesity                               2
Alcohol Consumption                   2
Exercise Hours Per Week            8763
Diet                                  3
Previous Heart Problems               2
Medication Use                        2
Stress Level                         10
Sedentary Hours Per Day            8763
Income                             8615
BMI                                8763
Triglycerides                       771
Physical Activity Days Per Week       8
Sleep Hours Per Day                   7
Country                              20
Continent                             6
Hemisphere                            2
Heart Attack Risk                     2
dtype: int64
```

In [11]:
```
data.isnull().sum()
```

Out[11]:
```
Patient ID                        0
Age                               1
Sex                               0
Cholesterol                       0
Blood Pressure                    0
Heart Rate                        0
Diabetes                          0
Family History                    0
Smoking                           0
Obesity                           0
Alcohol Consumption               0
Exercise Hours Per Week           0
Diet                              0
Previous Heart Problems           0
Medication Use                    0
Stress Level                      0
Sedentary Hours Per Day           0
Income                            0
BMI                               0
Triglycerides                     0
Physical Activity Days Per Week   0
Sleep Hours Per Day               0
Country                           0
Continent                         0
Hemisphere                        0
Heart Attack Risk                 0
dtype: int64
```

In [68]:
```
age_mean = data['Age'].mean()
data['Age'].fillna(value =  age_mean, inplace  = True)
data['Age']
```

Out[68]:
```
0       67.0
1       21.0
2       21.0
3       84.0
4       66.0
        ...
8758    60.0
8759    28.0
8760    47.0
8761    36.0
8762    25.0
Name: Age, Length: 8763, dtype: float64
```

In [69]:
```
data.isnull().sum()
```

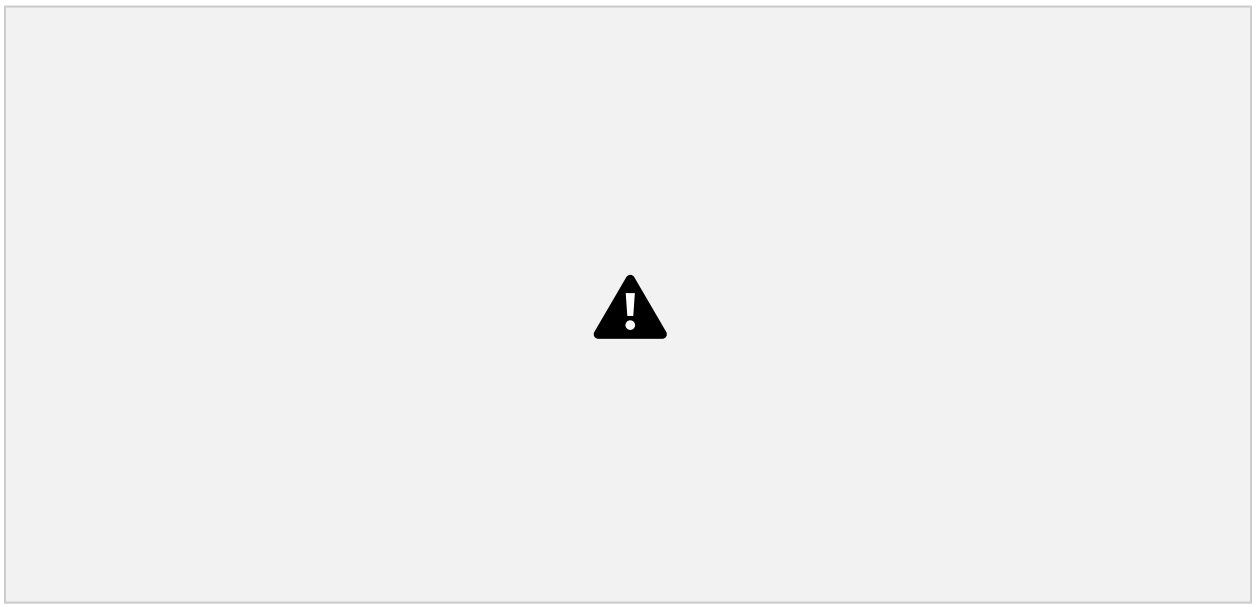
Out[69]:
```
Age                               0
Sex                               0
Cholesterol                       0
Blood Pressure                    0
Heart Rate                        0
Diabetes                          0
Family History                    0
Smoking                           0
Obesity                           0
Alcohol Consumption               0
Exercise Hours Per Week           0
Diet                              0
Previous Heart Problems           0
Medication Use                    0
Stress Level                      0
Sedentary Hours Per Day           0
Income                            0
BMI                               0
Triglycerides                     0
Physical Activity Days Per Week   0
Sleep Hours Per Day               0
Heart Attack Risk                 0
dtype: int64
```

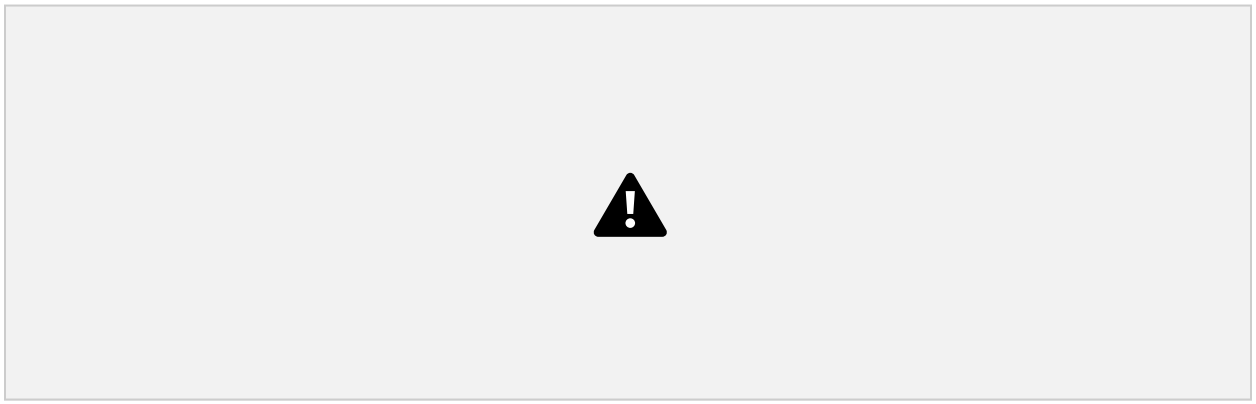
In [12]:
```
(data.isnull().sum()/(len(data)))*100
```

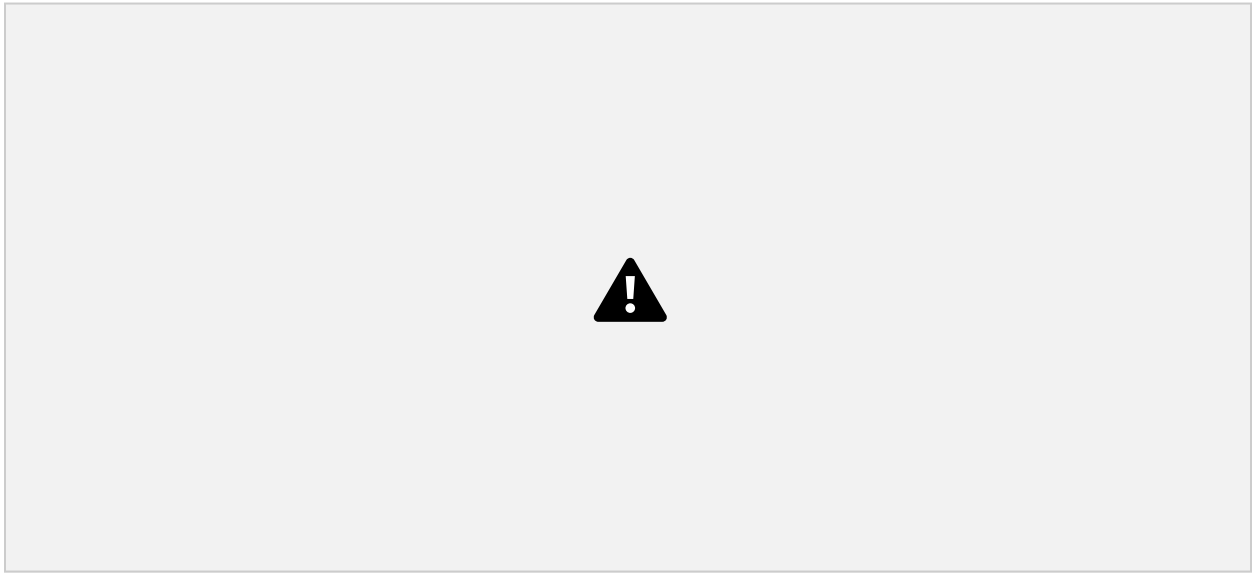
Out[12]:
```
Patient ID                        0.000000
Age                               0.011412
Sex                               0.000000
Cholesterol                       0.000000
Blood Pressure                    0.000000
Heart Rate                        0.000000
Diabetes                          0.000000
Family History                    0.000000
Smoking                           0.000000
Obesity                           0.000000
Alcohol Consumption               0.000000
Exercise Hours Per Week           0.000000
Diet                              0.000000
Previous Heart Problems           0.000000
Medication Use                    0.000000
Stress Level                      0.000000
Sedentary Hours Per Day           0.000000
Income                            0.000000
BMI                               0.000000
Triglycerides                     0.000000
Physical Activity Days Per Week   0.000000
Sleep Hours Per Day               0.000000
Country                           0.000000
Continent                         0.000000
Hemisphere                        0.000000
Heart Attack Risk                 0.000000
dtype: float64
```

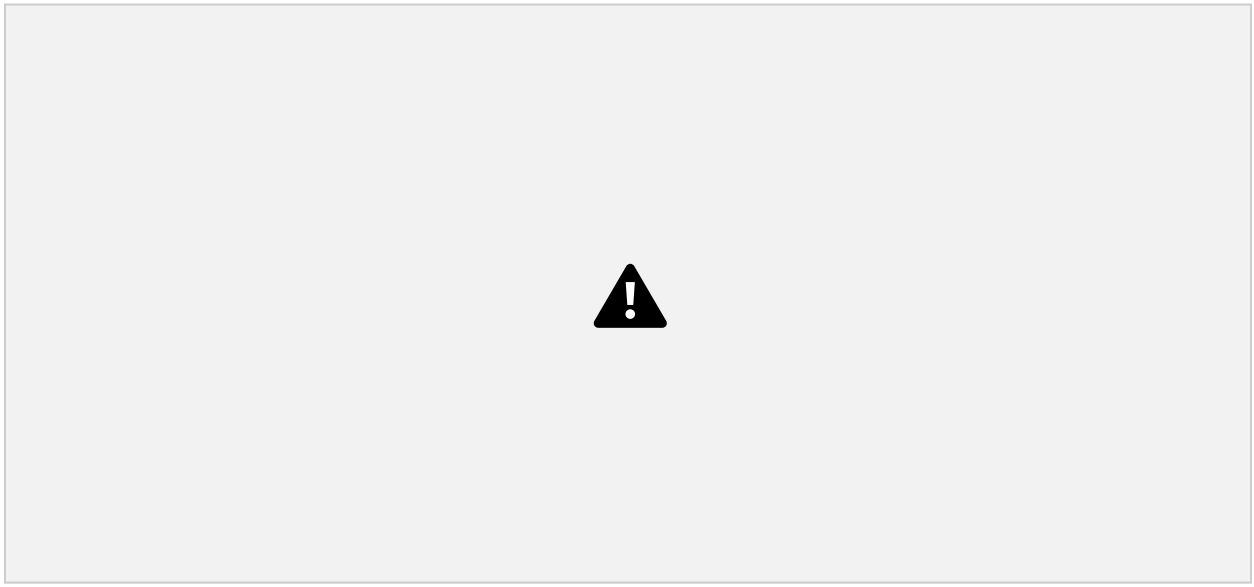
In [70]:
```
(data==0).sum()
```

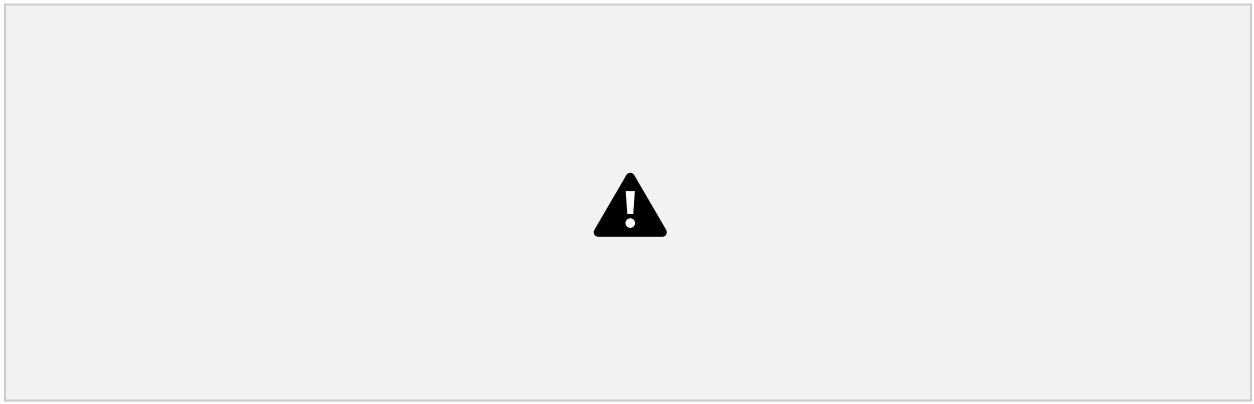
Out[70]:
```
Age                                  0
Sex                               2652
Cholesterol                          0
Blood Pressure                       0
Heart Rate                           0
Diabetes                          3047
Family History                    4443
Smoking                            904
Obesity                           4369
Alcohol Consumption               3522
Exercise Hours Per Week              0
Diet                                 0
Previous Heart Problems           4418
Medication Use                    4396
Stress Level                         0
Sedentary Hours Per Day              0
Income                               0
BMI                                  0
Triglycerides                        0
Physical Activity Days Per Week   1065
Sleep Hours Per Day                  0
Heart Attack Risk                 5624
dtype: int64
```

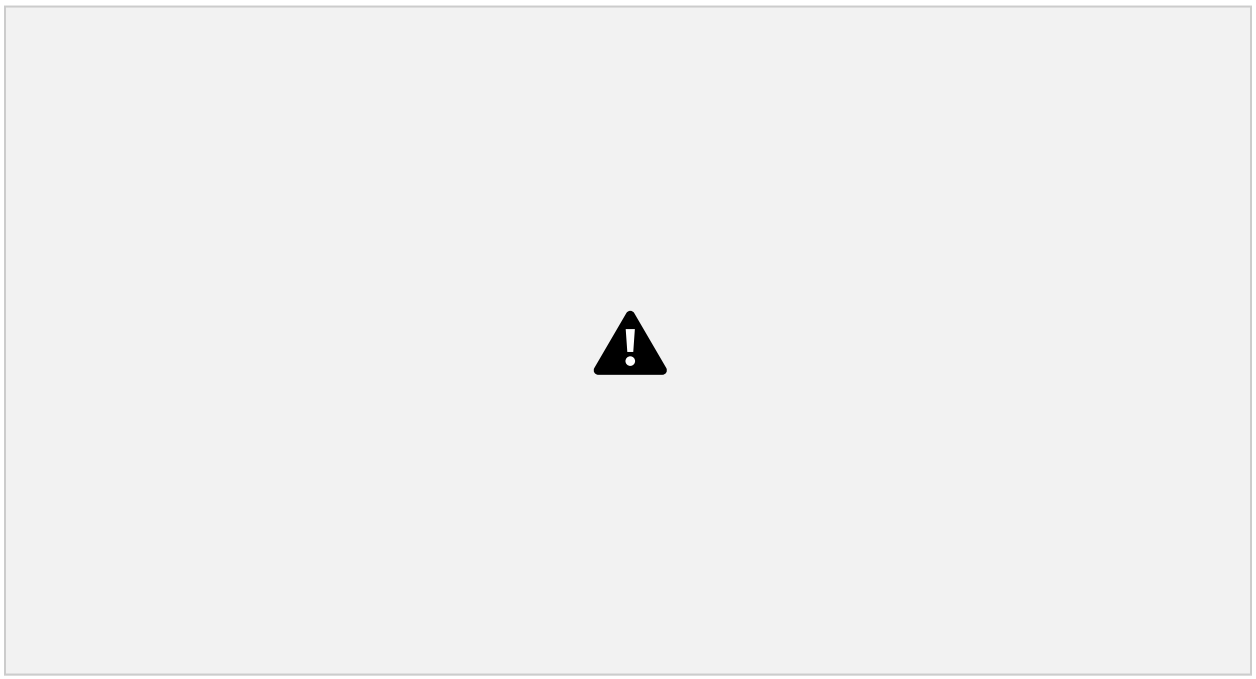












CONCLUSION

THE EXPLORATORY DATA ANALYSIS FOR THE DATASET IS PERFORMED TO UNDERSTAND THE DATASET, BASED ON THE HIGH FACTORS AFFECTING THE HEART ATTACK RISK, THE INDEPENDENT VARIABLES AND DEPENDENT VARIABLES, AND

CORRELATION BETWEEN THE VARIABLES.