

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

print('''
                                \nWELCOME TO THE HEART DISEASE DATA ANALYSIS-SOURCE
                                CODE\n''')

print('\nload the csv file into the database.')
fp = "C:/Users/aash1/OneDrive/Desktop/project 2/Heart Disease data.csv"
data = pd.read_csv(fp)

#Exploratory Data Analysis
print(data.head())

print(data.dropna())           #Handles missing values
print(data.shape)             #indicates the number of rows and columns
print(data.index)

#Gender Distribution
gc= data['sex'].value_counts()
gp=gc/len(data)*100

print('\n\nTo observe the gender distribution in our dataset, we use the count function
divided by the length
of the entire population.')
print(gp)

print('\n\nThis insight underscores the importance of gender-specific analysis in medical
studies to ensure that the conclusions drawn are applicable and accurate for both genders.

Males: The majority of the patients in the dataset are male, of the total sample. This
indicates a
higher representation of males in the study population.
Females: The remaining of the patients are female. This indicates a lower representation of
females compared to males.
''')

print('
                                THALACH (MAXIMUM HEART RATE ACHIEVED')
print('\n\nThe mean of maximum heart rate found from the above dataset is:', data[
'thalach'].mean())

print('Based on the dataset of individuals aged 40-80, the mean maximum heart rate of 149 bpm
appears to be within a healthy range.
For older adults (60-80 years), this value is particularly fitting, aligning well with expected
maximum heart rates
for this age group. For the younger segment (40-50 years), while slightly lower than the
highest expected values,
it still falls within a normal and healthy range, especially for moderate physical activity.
This indicates that
the cardiovascular health of the individuals in this dataset is generally good, with heart rate
responses that are
typical for their age.')

print('\n
                                RESTING ECOCARDIOGRAPHIC RESULTS')
print(data.groupby('restecg')[['slope']].count())

print('This analysis helps understand how the 'restecg' (resting electrocardiographic
results)
categories are distributed in the dataset and ensures that there are sufficient
non-null 'slope' (the slope of the peak exercise ST segment) values to perform
further analysis or modeling based on these categories.

1. The majority of the records have 'restecg' values of 0 or 1.
2.A small number of records have 'restecg' value 2.
''')

```

```

print('count(restecg)', data['restecg'].value_counts())
print('\nObservation:
considering that the ideal restecg falls under the category of 0, a significant number of
patients (497) come
under this category, while only 15 patients exhibit probable or definite left ventricular
hypertrophy (2).
This distribution indicates that most patients have typical or mildly abnormal(513) ECG
results'')

print('\nComparison between Resting electrocardiographic results and Maximum heart rate
achieved can be shown as:')

print(data.groupby('thalach')[['restecg']].count())

print('\nObservations: Most Common Values: Certain maximum heart rate values, such as 96 bpm,
are more frequently
observed in the dataset, indicating they might be common peak heart rates for individuals in
the study.
Heart Rate Variability: The data shows a wide range of maximum heart rates, from 71 bpm to 202
bpm, reflecting
the different levels of cardiovascular fitness and exertion levels among participants.
Potential Clusters: Clusters of common heart rate values (like around 95-96 bpm) can indicate
typical
cardiovascular responses in the studied population.'')

print('\n
                                SORTING DATA')

# print(data.sort_index())                # Sort the DataFrame by its row index and print the
result
print(data.sort_index(ascending=False))    # Sort the DataFrame by its row index and print the
result in desc order

# Sort the DataFrame by the 'age' column in ascending order and print the result
print(data.sort_values(by='age'))
print('\nBy sorting by age, we can compare how other variables (such as thalach, chol, etc.)
vary with age.
This can help identify correlations or trends, such as whether certain health metrics are
better or worse at
specific ages.

It allows us to observe and analyze trends and patterns related to age. For example, we can
easily identify the
youngest and oldest individuals in the dataset.')

print('
                                \nST Depression: Comparative study of mean, median and mode
')

print('mean: ', data['oldpeak'].mean())
print('median: ', data['oldpeak'].median())
print('mode: ', data['oldpeak'].mode())

print('
Mean of ST Depression (1.07):
Impact: The average ST depression of 1.07 suggests that, on average, patients experience mild
to moderate
ischemia during exercise. This level of ST depression could indicate that many patients may
experience some
degree of cardiac stress, which may lead to symptoms like chest pain or discomfort during
physical exertion.

Median of ST Depression (0.8):
Impact: A median value of 0.8 indicates that half of the patients experience less than 0.8
units of ST depression.
This lower median value compared to the mean suggests that many patients have relatively minor
ischemic changes,
which might not lead to severe symptoms but could still indicate underlying heart disease that

```

requires monitoring and management.

Mode of ST Depression (0.0):

Impact: The most frequently occurring value being 0.0 shows that a significant number of patients do not experience any ST depression during exercise. These patients are less likely to have exercise-induced ischemia, indicating better cardiac health during physical activity compared to those with higher ST depression values.'')

```
print('\n\n                                CHEST PAIN TYPE(0-3)''')
```

```
print('\n\nAge wise observation of Chest pain type (0-3) :-  
(0 being the highest and 3 being the lowest cp)''')
```

```
print('\nmode: ', data['cp'].value_counts())
```

```
print(''
```

Type 0 (Typical Angina): 497 occurrences

Interpretation: This is the most common type of chest pain in the dataset, indicating that typical angina is the predominant chest pain experienced by the patients.

Type 2 (Non-anginal Pain): 284 occurrences

Interpretation: The second most common chest pain type is non-anginal pain, which suggests that a significant number of patients experience chest pain that does not fit the typical characteristics of angina.

Type 3 (Asymptomatic): 77 occurrences

Interpretation: Asymptomatic chest pain, or no chest pain, is the least common in this dataset.'')

```
custom=['#7592ee', '#fc77ec']
```

#plotting the count plot using the given code

```
plt.figure(figsize=(10, 6))  
sns.countplot(data=data, x='cp', hue='sex', palette=custom)  
plt.title('Frequency of Chest Pain Types by Gender')  
plt.xlabel('Chest Pain Type')  
plt.ylabel('Frequency')  
plt.legend(title='Gender', labels=['Female', 'Male'])  
plt.show()
```

```
print('\n\n                                DATA CORRELATION:
```

we may apply the data.corr() function on our dataset to analyse the type of correlation that exists between every attribute against every other attribute. This may give us a brief summarisation and the correlation coefficients of every single health factor between each other.

These coefficients range from -1 to 1, where:

1 indicates a perfect positive correlation.

-1 indicates a perfect negative correlation.

0 indicates no correlation.'')

```
print(data.corr())
```

```
print('\n\n                                OBSERVATIONS
```

1. Age:

Negatively correlated with thalach (-0.390): As age increases, the maximum heart rate tends to decrease.

Positively correlated with ca (0.272): Older individuals tend to have a higher number of major vessels colored by fluoroscopy.

Negatively correlated with target (-0.229): Younger individuals are more likely to have heart disease.

2. Sex:

Negatively correlated with target (-0.280): Males (1) are less likely to have heart disease compared to females (0).

Positively correlated with thal (0.198): Males are more likely to have certain types of thalassemia.

3. Chest Pain Type (cp):

Positively correlated with target (0.435): Higher chest pain type values are associated with a higher

likelihood of heart disease.

Negatively correlated with exang (-0.402): Exercise-induced angina is less common in individuals with higher chest pain type values.

4. Resting Blood Pressure (trestbps):

Positively correlated with ca (0.105): Higher resting blood pressure is associated with a higher number of

major vessels colored by fluoroscopy.

Negatively correlated with target (-0.139): Individuals with lower resting blood pressure are more likely to have heart disease.

5. Serum Cholesterol (chol):

Positively correlated with age (0.220): Older individuals tend to have higher serum cholesterol.

Negatively correlated with target (-0.100): Lower cholesterol levels are slightly associated with the presence of heart disease.

6. Maximum Heart Rate Achieved (thalach):

Negatively correlated with age (-0.390): Younger individuals tend to have higher maximum heart rates.

Positively correlated with target (0.423): Higher maximum heart rates are associated with the presence of heart disease.

7. Exercise-Induced Angina (exang):

Negatively correlated with target (-0.438): Absence of exercise-induced angina is associated with the presence of heart disease.

8. ST Depression (oldpeak):

Positively correlated with target (-0.438): Lower ST depression values are associated with the presence of heart disease.

''')

```
print(''                                     FASTING BLOOD SUGAR ''')
```

```
print('count(fbs): ', data['fbs'].value_counts())
```

```
print('' Observation:
```

The count of fasting blood sugar (fbs) levels in the dataset reveals the following:

Fasting Blood Sugar <= 120 mg/dl (fbs = 0): The majority of patients, approximately 872, have fasting blood sugar levels less than or equal to 120 mg/dl. This suggests that most patients in the dataset do not have elevated fasting blood sugar levels.

Fasting Blood Sugar > 120 mg/dl (fbs = 1): A smaller portion of the patients, around 153, have fasting blood

sugar levels greater than 120 mg/dl, indicating a subset of the population with higher blood sugar levels, potentially at risk for diabetes or other metabolic disorders.'')

```
print('\n                                EXERCISE INDUCED ANGINA ''')
print('count(exang)', data['exang'].value_counts())
```

```
print(''Observation:
The count of exercise-induced angina (exang) in the dataset reveals the following:
```

No Exercise-Induced Angina (exang = 0): The majority of patients, approximately 680, do not experience angina induced by exercise. This indicates that for most patients, physical exertion does not trigger angina.

Exercise-Induced Angina (exang = 1): A substantial number of patients, around 345, do experience angina when they exercise. This highlights a significant subset of the population for whom physical activity may provoke chest pain. ''')

#Conclusion

```
print('\n\n                                CONCLUSION
After analysing the heart disease data and its various parameters, we have compared and analysed trends and patterns among various factors which affect heart health and how common or uncommon it is among different comparative values. After a sought through analysis, the observations are as follows along with actionable reccomendations on some areas based upon the analysis.'')
```

```
print("""
Key Observations:
1. The majority of patients in the dataset are male.
2. The mean maximum heart rate (149 bpm) is within a healthy range for the studied age group.
3. Most patients have normal or mildly abnormal ECG results.
4. Higher chest pain type values are associated with a higher likelihood of heart disease.
5. There is a significant correlation between certain variables like age, maximum heart rate, and the presence of heart disease.
```

Recommendable Actions:

Stress Management and Meditation: Promote stress management techniques and meditation to help young individuals cope with stress, which can be a risk factor for heart disease. Regular meditation can improve heart health by reducing stress, lowering blood pressure, and improving overall mental well-being.

Regular Screening for Heart Disease:

Age 50 and above: Since the prevalence of heart disease significantly increases after the age of 50, it is recommended that individuals in this age group undergo regular cardiovascular screenings. Early detection can lead to better management and treatment outcomes.

1.2. Lifestyle Interventions:

Cholesterol and Blood Pressure Management: For individuals, especially those aged 50 and above, maintaining a healthy diet low in saturated fats, regular physical activity, and possibly medication can help manage cholesterol levels and blood pressure.

Weight Management and Smoking Cessation: Encourage maintaining a healthy weight and quitting smoking, as these factors significantly impact cardiovascular health.

1.3. Gender-Specific Health Programs:

Male-Focused Interventions: Given the higher prevalence of heart disease in males, targeted health programs and awareness campaigns for men can be beneficial.

1.4. Exercise Recommendations:

Moderate Physical Activity: Promote moderate physical activity tailored to age and health status to improve cardiovascular health. Regular exercise can help maintain a healthier maximum heart rate and reduce the risk of heart disease.

Healthy Lifestyle Habits: Encourage young individuals (ages 30-50) to adopt healthy lifestyle habits early on, including a balanced diet, regular exercise, and avoiding smoking, to prevent the onset of heart disease later in life.

These recommendations, limitations, and future research suggestions aim to enhance the understanding and management of heart disease, ultimately leading to better health outcomes across all age groups."")