

Unified Mentor

Heart Rate Diagnosis Data Analysis

By Tahir Malik

Problem Statement

Health is real wealth in the pandemic time we all realized the brute effects of covid-19 on all irrespective of any status. You are required to analyze this health and medical data for better future preparation.

Do ETL: Extract- Transform and Load data from the heart disease diagnostic database You can perform EDA through python. The database extracts various information such as heart disease rates, heart disease by gender, and by age. You can even compare attributes of the data set to extract necessary information. Make the necessary dashboard with the best you can extract from the data.

Use various visualization and features and make the best dashboard.

Find key metrics and factors and show the meaningful relationships between attributes. Do your own research and come up with your findings.

Introduction

In the wake of the COVID-19 pandemic, the importance of health has been underscored globally. As communities grapple with the repercussions of the virus, it has become increasingly evident that health is indeed the true wealth of individuals and societies alike. Amidst this realization, it is crucial to not only focus on addressing immediate health crises but also to understand and prepare for other prevalent health concerns.

Heart disease remains one of the leading causes of mortality worldwide, affecting individuals across all demographics regardless of socioeconomic status. By leveraging data from a heart disease diagnostic database, this report aims to conduct an in-depth analysis of heart disease rates, distribution by gender and age, and explore various attributes associated with heart health.

Through the process of Extract-Transform-Load (ETL) and exploratory data analysis (EDA) using Python, we will uncover key insights into the prevalence and characteristics of heart disease. By employing various visualization techniques and creating an interactive dashboard, we will present these findings in a comprehensive and actionable manner.

This report endeavours to identify significant metrics and factors related to heart disease, elucidating meaningful relationships between attributes within the dataset. By amalgamating data analysis with real-world insights, we strive to contribute to the collective understanding of heart health and pave the way for informed decision-making and future preparedness in healthcare.

Python Code

Data Processing

```
# importing python libraries
```

```
import pandas as pd
```

```
# importing and reading data stored in heart-disease-dataset.csv
```

```
dataFrame = pd.read_csv('./csv/heart-disease-dataset.csv')
```

```
# printing dataset overview stored in dataframe
```

```
dataFrame
```

```
# changing column names
```

```
dataFrame.columns = ['Age', 'Sex', 'Chest Pain Type', 'Resting BP',  
'Serum Cholestorol (mg/dl)', 'Fasting Blood Sugar > 120 mg/dl', 'Resting  
Electro-Cardiographic Results', 'Max Heart Rate Achieved', 'Exercise  
Induced Angina', 'Old-Peak', 'Slope', 'No. of Major Vessels  
(fluoroscopy)', 'Thalium Scintigraphy', 'Presense of Heart Disease']
```

```
# printing dataset overview stored in dataframe
```

```
dataFrame
```

```
# checking brief statistical summary of dataset
```

```
dataFrame.describe()
```

```
# checking NaN values by each column
```

```
dataFrame.isnull().sum()
```

Data Cleaning

There is not any NaN values present in our dataset so we do not need to do perform data cleaning

Data Analysis

```
# finding heart disease prevalence rate throughout dataset
```

```
dataFrame.shape
```

```
count = 0
```

```
for i in range(len(dataFrame)):
```

```
    if dataFrame['Presense of Heart Disease'][i] == 1:
```

```
        count += 1
```

```

print(f"Heart Patient: {count}")
print(f"Non Heart Patient: {len(dataFrame) - count}")
print(f"Total Patients: {526 + 499}")
heart_disease_prevalence_rate = ((count/len(dataFrame)) * 100)
heart_disease_prevalence_rate = round(heart_disease_prevalence_rate, 1)
print(f"Heart Disease Prevalence Rate: {heart_disease_prevalence_rate}%")

```

```

# finding average age throughout dataset

average_age = dataFrame['Age'].mean()
print(f"Average age: {average_age.astype(int)} years")

```

```

# finding average resting blood pressure throughout dataset

average_resting_blood_pressure = dataFrame['Resting BP'].mean()
print(f"Average resting blood pressure: {average_resting_blood_pressure.astype(int)} BPM")

```

```

# finding average serum cholesterol level throughout dataset

average_serum_cholesterol = dataFrame['Serum Cholestoral (mg/dl)'].mean()
print(f"Average serum cholesterol: {average_serum_cholesterol.astype(int)} mg/dl")

```

```

# finding percentage of individuals with high fasting blood sugar throughout dataset

count = 0
for i in range(len(dataFrame)):
    if dataFrame['Fasting Blood Sugar > 120 mg/dl'][i] == 1:
        count += 1

high_fasting_blood_sugar_rate = ((count/len(dataFrame)) * 100)
high_fasting_blood_sugar_rate = round(high_fasting_blood_sugar_rate, 1)
print(f"High fasting blood sugar rate: {high_fasting_blood_sugar_rate}%")

```

```

# finding max heart rate achieved by male & female throughout dataset

female_count = 0
for i in range(len(dataFrame)):
    if dataFrame['Sex'][i] == 0: # sex == 0 female, 1 male
        female_count += 1

print(f"Females in dataset: {female_count}")
print(f"Males in dataset: {len(dataFrame) - female_count}")

```

```
max_heart_rate = dataframe.groupby('Sex')['Max Heart Rate  
Achieved'].max()  
print(f"Max Heart Rate (female): {max_heart_rate[0]} BPM")  
print(f"Max Heart Rate (male): {max_heart_rate[1]} BPM")
```

```
# finding percentage of individuals with Exercise-Induced Angina
```

```
count = 0  
for i in range(len(dataFrame)):  
    if dataFrame['Exercise Induced Angina'][i] == 1:  
        count += 1  
  
exercise_induced_angina_rate = ((count/len(dataFrame)) * 100)  
exercise_induced_angina_rate = round(exercise_induced_angina_rate, 1)  
print(f"Exercise Induced Angina Rate: {exercise_induced_angina_rate} %")
```

```
# finding average ST depression throughout dataset
```

```
average_st_depression = dataFrame['Old-Peak'].mean()  
print(f"Average ST depression: {average_st_depression.astype(int)}")
```

```
# finding percentage of individuals with abnormal resting  
electrocardiographic results
```

```
count = 0  
for i in range(len(dataFrame)):  
    if dataFrame['Resting Electro-Cardiographic Results'][i] == 1 or  
dataFrame['Resting Electro-Cardiographic Results'][i] == 2:  
        count += 1  
  
abnormal_resting_electrocardiographic_rate = ((count/len(dataFrame)) *  
100)  
abnormal_resting_electrocardiographic_rate =  
round(abnormal_resting_electrocardiographic_rate, 1)  
print(f"Abnormal Resting Electrocardiographic Rate:  
{abnormal_resting_electrocardiographic_rate} %")
```

```
# creating new column 'Gender' on the basis of sex column's 0 & 1 values
```

```
dataFrame['Gender'] = dataFrame['Sex'].apply(lambda x: 'Male' if x == 1  
else 'Female')
```

Insights

- Heart Patient: 526
- Non-Heart Patient: 499
- Total Patients: 1025
- Heart Disease Prevalence Rate: 51.3 %
- Average age in dataset: 54 years
- Average resting blood pressure: 131 BPM
- Average serum cholesterol: 246 mg/dl
- High fasting blood sugar rate: 14.9 %
- Females in dataset: 312
- Males in dataset: 713
- Max Heart Rate (female): 192 BPM
- Max Heart Rate (male): 202 BPM
- Exercise Induced Angina Rate: 33.7 %
- Average ST depression: 1
- Abnormal Resting Electrocardiographic Rate: 51.5 %

Conclusion

Heart Disease Prevalence: The prevalence of heart disease in the dataset is relatively high at 51.3%, indicating that heart disease is a significant health concern among the patients included in the study.

Demographic Distribution: The dataset consists of 312 females and 713 males, suggesting a higher representation of males. This may indicate either a higher prevalence of heart disease among males or a higher proportion of males seeking medical attention for heart-related issues.

Age and Heart Disease: The average age of patients in the dataset is 54 years, indicating that heart disease affects individuals across a wide age range. However, it would be beneficial to analyze the age distribution further to understand if there are specific age groups more susceptible to heart disease.

Resting Blood Pressure and Serum Cholesterol: The average resting blood pressure and serum cholesterol levels are 131 BPM and 246 mg/dl, respectively. Elevated levels of both indicators are associated with an increased risk of heart disease, highlighting the importance of monitoring these parameters for early detection and prevention.

High Fasting Blood Sugar: A concerning finding is the high fasting blood sugar rate of 14.9%. Elevated blood sugar levels are associated with diabetes, which is a significant risk factor for heart disease. This suggests a potential need for diabetes screening and management among the patient population.

Max Heart Rate: The maximum heart rate recorded for males (202 BPM) is higher than that for females (192 BPM). While individual variation exists, this may reflect physiological differences between genders.

Exercise Induced Angina and Abnormal Resting Electrocardiographic Rate: The rates of exercise-induced angina (33.7%) and abnormal resting electrocardiographic findings (51.5%) are both relatively high. These findings indicate that a significant portion of the patient population experiences symptoms during physical activity and may have abnormal heart rhythms at rest, further emphasizing the prevalence and severity of heart disease in the dataset.