Unified Mentor

Document on Heart Disease Diagnostic Analysis

By: Pratikshya Das

Problem Statement

Heart disease poses a multifaceted challenge, encompassing a spectrum of conditions ranging from coronary artery disease to heart failure. Despite extensive research efforts and medical interventions, the incidence of heart disease remains alarmingly high, affecting individuals of all ages and backgrounds. Moreover, the diagnosis and management of heart disease present significant clinical challenges, often requiring a multidisciplinary approach and personalized treatment strategies.

Considering these challenges, there is a pressing need to explore innovative analytical approaches to decipher the underlying mechanisms, risk factors, and prognostic indicators associated with heart disease. By addressing these gaps in knowledge, this project aims to inform clinical practice, enhance patient care, and ultimately mitigate the burden of heart disease on global health.

Make views and dashboards first and make a story out of it.

Introduction

In the realm of healthcare, cardiovascular diseases pose a significant global health burden, with heart disease being a leading cause of mortality worldwide. Recognizing the critical importance of addressing this pressing issue, I embarked on an in-depth exploration of heart disease analysis. The overarching goal of this endeavor was to unearth insights that could potentially revolutionize our understanding and treatment of cardiovascular ailments.

The motivation behind this project stemmed from a profound concern for public health and a desire to leverage data-driven approaches to tackle complex medical challenges. Heart disease represents a multifaceted health issue with far-reaching implications for individuals, families, and communities. By delving into the intricacies of heart disease analysis, I sought to contribute meaningfully to the ongoing efforts aimed at mitigating its impact and improving patient outcomes.

Moreover, this report functions as a blueprint for conducting similar analyses on heart disease diagnostic data.

Code Demonstration

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

 data=pd.read_csv("Heart.csv")
data.tail()
data.head()

print("Number of rows" ,data.shape[0])
print("Number of cloumns" ,data.shape[1])

data.info()

data.isnull().sum()

data_dup=data.duplicated().any()
print(data_dup)

data=data.drop_duplicates()
data.shape

data.describe()

plt. figure(figsize=(18,6))
sns.heatmap(data.corr(),annot=True)

data.columns

data['target'].value_counts()

# Data cleaning
# Drop rows with missing values
data.dropna(inplace=True)

# Percentage of people having Heart Disease
heart_disease_counts = data['target'].value_counts()
plt.pie(heart_disease_counts, labels=heart_disease_counts.index,
autopct='%1.1f%%')
plt.title('Percentage of people having Heart Disease')
plt.show()

# Adding a new column 'Age_Range'
data['Age_Range'] = pd.cut(data['age'], bins=[0, 20, 40, 60, 80, 100]
,labels=['0-20', '21-40', '41-60', '61-80', '81-100'])

# Swarm Plot Creation of Gender Based Age Category
plt.figure(figsize=(10, 6))
sns.swarmplot(x='sex', y='age', hue='Age_Range', data=data)
plt.title('Gender Based Age Category')
plt.xlabel('Gender (0 = Female, 1 = Male)')
plt.ylabel('Age')
plt.show()
```

```
# Bar Plot Creation of Age Category
plt.figure(figsize=(10, 6))
sns.countplot(x='age', data=data)
plt.title('Population Age Category')
plt.xlabel('Age')
plt.ylabel('Count')
plt.xticks(rotation=90)
plt.show()

# Converting Numerical Data into Categorical Data
# Adding a new column 'Age_Range'
data['Age_Range'] = pd.cut(data['age'], bins=[0, 20, 40, 60, 80, 100],
labels=['0-20', '21-40', '41-60', '61-80', '81-100'])

#Countplot Creation of Population Age
plt.figure(figsize=(10, 6))
sns.countplot(x='Age_Range', data=data)
plt.title('Population Age Category')
plt.xlabel('Age Range')
plt.ylabel('Count')
plt.show()

# Count Plot Creation of Heart Disease Based On Age Category
plt.figure(figsize=(10, 6))
sns.countplot(x='Age_Range', hue='target', data=data)
plt.title('Heart Disease Based On Age Category')
plt.xlabel('Age Range')
plt.ylabel('Count')
plt.legend(title='Heart Disease', loc='upper right', labels=['No',
'Yes'])
plt.show()

# Count Plot Creation of Heart Disease Based on Gender
plt.figure(figsize=(6, 6))
sns.countplot(x='sex', hue='target', data=data)
plt.title('Heart Disease Based On Gender')
plt.xlabel('Gender (0 = Female, 1 = Male)')
plt.ylabel('Count')
plt.legend(title='Heart Disease', loc='upper right', labels=['No',
'Yes'])
plt.show()

# Count Plot Creation of Chest Pain Experienced
plt.figure(figsize=(10, 6))
sns.countplot(x='cp', data=data)
plt.title('Chest Pain Experienced')
plt.xlabel('Chest Pain Type')
plt.ylabel('Count')
plt.show()

# Count Plot Creation of Chest Pain Based On Gender
plt.figure(figsize=(6, 6))
sns.countplot(x='cp', hue='sex', data=data)
plt.title('Chest Pain Based On Gender')
plt.xlabel('Chest Pain Type')
plt.ylabel('Count')
plt.legend(title='Gender', loc='upper right', labels=['Female', 'Male'])
plt.show()

# Count Plot Creation of Chest Pain Based On Age Category
plt.figure(figsize=(10, 6))
sns.countplot(x='cp', hue='Age_Range', data=data)
```

```
plt.title('Chest Pain Based On Age Category')
plt.xlabel('Chest Pain Type')
plt.ylabel('Count')
plt.legend(title='Age Range')
plt.show()

# Bar Plot Creation of Person's Resting Blood Pressure (mm Hg)
plt.figure(figsize=(10, 6))
sns.barplot(x='trestbps', y='target', data=data)
plt.title("Person's Resting Blood Pressure (mm Hg)")
plt.xlabel('Resting Blood Pressure (mm Hg)')
plt.ylabel('Heart Disease')
plt.show()

sns.set_style('whitegrid')
sns.countplot(x='target',data=data,palette='RdBu_r')

data.columns

data['sex'].value_counts()

sns.set_style('whitegrid')
sns.countplot(x='sex' ,data=data,palette='RdBu_r')
sns.histplot(data['age'],bins=20)
plt.show()

# Create a count plot for 'cp'
sns.countplot(data['cp'])

#Customize x-axis Lables
plt.xticks([0, 1, 2 ,3], ["Typical Angina", "Atypical Angina", "Non-
Anginal", "Asymptomatic"], rotation=75)

#Display the plot
plt.show()

sns.countplot(x="cp", hue="target" ,data=data)
plt.legend(labels=["No-Disease", "Disease"])
plt.show()

sns.countplot(x="fbs",hue="target",data=data)
plt.legend(labels=["No-Disease","Disease"])
plt.show()
```

```
data['trestbps'].hist()
g = sns.FacetGrid(data, hue="sex", aspect=4)
g.map(sns.kdeplot, 'trestbps', fill=True)
# Corrected argument name
plt.legend(labels=['Male', 'Female'])
# Adjust figure Layout
plt.tight_layout()
plt.show()

data['chol'].hist()

cate_val=[]
cont_val=[]

for column in data.columns:
    if data [column].nunique() <=10:
        cate_val.append(column)
    else:
        cont_val.append(column)

cate_val

cont_val

data.hist(cont_val,figsize=(15,6))
plt.show()
```

Analysis Approach

Analysis Approach:

Data Collection and Preparation: Gather comprehensive datasets related to heart disease, ensuring inclusion of demographic, clinical, and diagnostic information. Clean and preprocess the data to remove inconsistencies and missing values, ensuring data integrity.

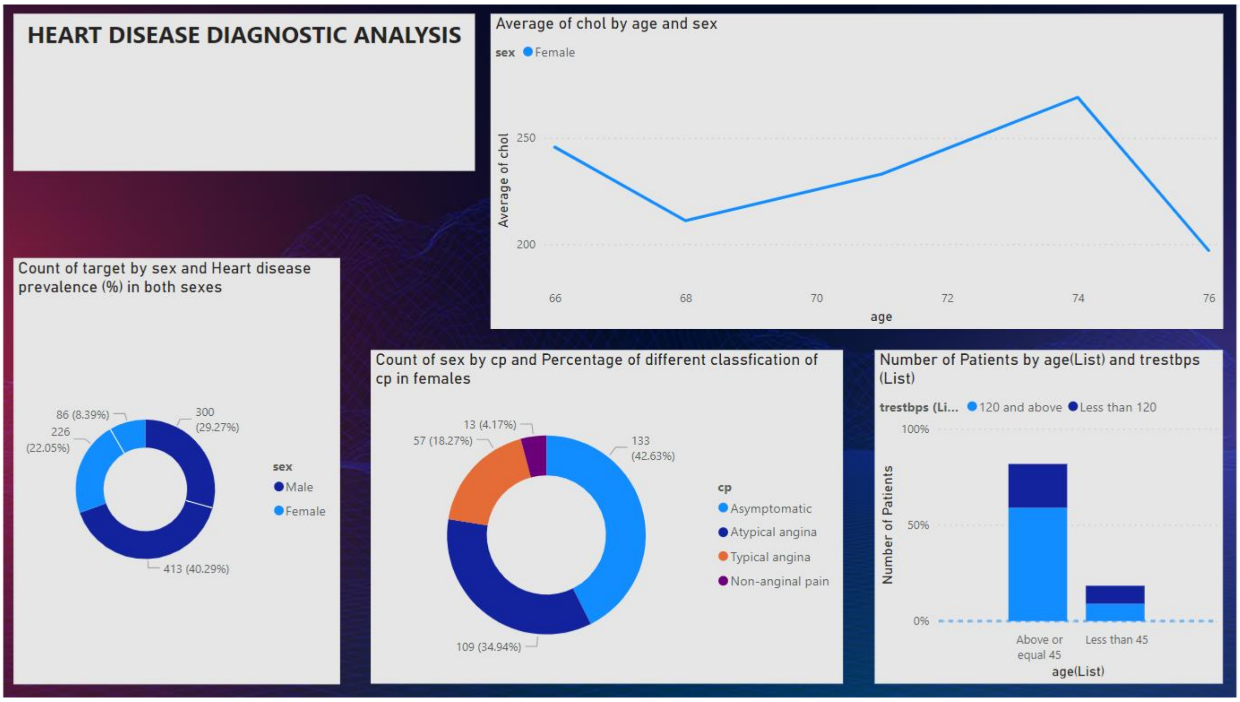
Exploratory Data Analysis (EDA): Conduct descriptive statistics and visualizations to gain insights into the distribution and relationships among key variables. Explore trends, patterns, and outliers to inform subsequent analyses.

Feature Selection: Identify relevant features influencing heart disease outcomes through correlation analysis, feature importance scores, or domain knowledge. Select a subset of informative features for model development.

Modeling: Choose appropriate machine learning models for predicting heart disease outcomes based on the nature of the problem (e.g., classification for predicting disease presence). Train and evaluate the models using appropriate evaluation metrics such as accuracy, precision, recall, and F1-score.

Interpretation and Insights: Analyze model results to interpret the relative importance of features in predicting heart disease outcomes. Generate actionable insights that can inform clinical decision-making, public health interventions, and further research efforts.

Visualization



Conclusion

In summary, Through the meticulous interrogation of the heart disease datasets, I unearthed a wealth of compelling insights that have profound implications for clinical practice and public health policy. Among the key findings were the identification of modifiable risk factors such as smoking, sedentary lifestyle, and poor dietary habits as significant contributors to heart disease burden.

Moreover, the analysis revealed intricate associations between demographic variables, clinical parameters, and cardiovascular outcomes, shedding light on nuanced patterns of disease susceptibility and progression. Importantly, these insights have the potential to inform targeted interventions aimed at mitigating risk, improving early detection, and optimizing treatment strategies for individuals at risk of or living with heart disease.

the results of this analysis underscore the transformative potential of data-driven approaches in the field of cardiovascular medicine. By harnessing the power of data analytics, we can unlock new avenues for preventing, diagnosing, and treating heart disease, ultimately advancing the goal of achieving better cardiovascular outcomes for all.