Advanced Heart Disease Diagnostics: Leveraging NumPy and Pandas for Effective Exploratory Data Analysis

Introduction to Heart Disease

Heart disease remains a leading cause of death globally. Understanding its **diagnostics** is crucial for early detection and treatment. This presentation explores how **NumPy** and **Pandas** can enhance **exploratory data analysis** (EDA) in heart disease diagnostics, providing insights into patient data and outcomes.

Objective:

Our objective is to perform a comprehensive analysis of the dataset, including but not limited to:

- Descriptive statistics of the features.
- Data visualization to identify patterns and relationships.
- Correlation analysis to understand the relationships between features.
- Distribution of the target variable.
- Identification of potential risk factors for heart disease.

Import Libraries

```
[ ] # import all library that will be used in entire project
    import numpy as np                          #importing numpy
    import pandas as pd                         #imporing pandas
    import matplotlib.pyplot as plt             #imporitng matplotlib
    import seaborn as sns                       #importing seaborn
```

Dataset Loading

```
[ ] # Mount google drive for access the dataset of Heart Disease diagnostic
    from google.colab import drive
    drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

Dataset First View

```
[ ] HeartDisease_df.head()
```

Missing values/Null values

```
[ ] #Checking NULL Values
    HeartDisease_df.isnull().sum()
```

Understanding Heart Disease

Heart disease encompasses various **cardiovascular conditions** affecting the heart's structure and function. Key factors include **genetics**, lifestyle choices, and **environmental influences**. Recognizing these aspects is essential for effective diagnostics and treatment strategies.

Assign: | New Tab | What is Linux? | Op... | AdBlock is now inst... | All Bookmarks

Heart_Disease_Diagnostic_Analysis.ipynb ☆

File Edit View Insert Runtime Tools Help Last edited on 28 June

Comment Share

+ Code + Text

Connect Gemini

```
trestbps    0
chol        0
fbs         0
restecg     0
thalach     0
exang       0
oldpeak     0
slope       0
ca          0
thal        0
target      0
dtype: int64
```

There is no missing Values in our Dataset

```
[ ] #Dataset Info and check how many entries are there left after cleaning the data
    HeartDisease_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1025 entries, 0 to 1024
Data columns (total 14 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   age       1025 non-null   int64
 1   sex       1025 non-null   int64
 2   cp        1025 non-null   int64
 3   trestbps  1025 non-null   int64
 4   chol      1025 non-null   int64
 5   fbs       1025 non-null   int64
 6   restecg   1025 non-null   int64
 7   thalach   1025 non-null   int64
 8   exang     1025 non-null   int64
 9   oldpeak   1025 non-null   float64
 10  slope     1025 non-null   int64
 11  ca        1025 non-null   int64
 12  thal      1025 non-null   int64
 13  target    1025 non-null   int64
dtypes: float64(1), int64(13)
memory usage: 112.2 KB
```

```
[ ] #Counting the number of rows and column after cleaning the data set
    row, columns = HeartDisease_df.shape
    print(f"No of rows: {row}")
    print(f"No of columns: {columns}")
```

```
No of rows: 1025
No of columns: 14
```

```
[ ] #All Columns in the Dataset
```

Role of Data in Diagnostics

Data plays a pivotal role in diagnosing heart disease. By analyzing **clinical data**, healthcare professionals can identify patterns and risk factors. Effective use of **data analytics** tools like **NumPy** and **Pandas** can significantly improve diagnostic accuracy and patient outcomes.

Heart_Disease_Diagnostic_Analysis.ipynb


```
#All Columns in the Dataset
HeartDisease_df.columns
```

```
Index(['age', 'sex', 'cp', 'trestbps', 'chol', 'fbs', 'restecg', 'thalach',
       'exang', 'oldpeak', 'slope', 'ca', 'thal', 'target'],
      dtype='object')
```


There are fourteen features in Dataset

age: The person's age in years

sex: The person's sex (1 = male, 0 = female)

cp: The chest pain experienced (Value 1: typical angina, Value 2: atypical angina, Value 3: non-anginal pain, Value 4: asymptomatic)

trestbps: The person's resting blood pressure (mm Hg on admission to the hospital)

chol: The person's cholesterol measurement in mg/dl

fbs: The person's fasting blood sugar (> 120 mg/dl, 1 = true; 0 = false)

restecg: Resting electrocardiographic measurement (0 = normal, 1 = having ST-T wave abnormality, 2 = showing probable or definite left ventricular hypertrophy by Estes' criteria)

thalach: The person's maximum heart rate achieved

exang: Exercise induced angina (1 = yes; 0 = no)

oldpeak: ST depression induced by exercise relative to rest

slope: the slope of the peak exercise ST segment (Value 1: upsloping, Value 2: flat, Value 3: downsloping)

ca: The number of major vessels (0-3)

thal: A blood disorder called thalassemia (3 = normal; 6 = fixed defect; 7 = reversable defect)

target: Heart disease (0 = no, 1 = yes)

Percentage of people having Heart Disease

```
num=HeartDisease_df.groupby('target').size()
num
```

```
target
0    499
1    526
dtype: int64
```

```
#Converting Numerical Data into Categorical Data
```

Introduction to NumPy

NumPy is a powerful library for numerical computing in Python. It provides support for **arrays**, matrices, and a vast collection of mathematical functions. Utilizing NumPy allows for efficient data manipulation, essential for handling large datasets in heart disease diagnostics.

colab.research.google.com/drive/1Znu3NMjNiFkNEeI9p3P-8py37S_-InZQ

Assign: | New Tab | What is Linux? | Op... | AdBlock is now inst... | All Bookmarks

Heart_Disease_Diagnostic_Analysis.ipynb

File Edit View Insert Runtime Tools Help Last edited on 28 June

Comment Share

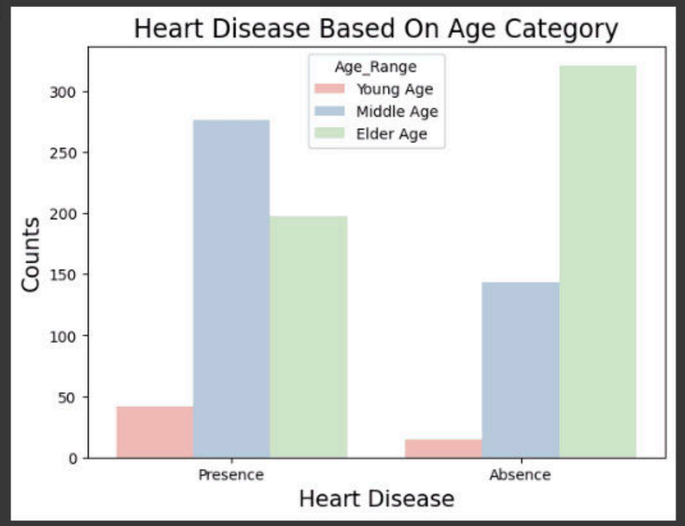
+ Code + Text

Connect Gemini

3	61	1	0	148	203	0	1	161	0	0.0	2	1	3	0	Absence	Male	Elder Age
4	62	0	0	138	294	1	1	106	0	1.9	1	3	2	0	Absence	Female	Elder Age

From Our Population chart we can say that Number Of Males are more in Middle Age Category and Females are more in Elder Age Category.

```
#Count Plot Creation of Heart Disease Based On Age Category using MatplotLib and Seaborn
plt.figure(figsize=(7,5))
hue_order=['Young Age', 'Middle Age', 'Elder Age']
sns.countplot(x='Heart_Disease', hue='Age_Range', data=HeartDisease_df, order=['Presence','Absence'], hue_order=hue_order, palette='Pastel1')
plt.title('Heart Disease Based On Age Category', fontsize=17)
plt.xlabel('Heart Disease', fontsize=15)
plt.ylabel('Counts', fontsize=15)
plt.show()
```



Middle Age People are most AFFECTED by Heart Disease and Elder Age People are mostly FREE from any kind of Heart Disease.

```
#Count Plot Creation of Heart Disease Based on Gender using MatplotLib and Seaborn
plt.figure(figsize=(7,5))
```

Introduction to Pandas

Pandas is a versatile library for data manipulation and analysis. It offers data structures like **DataFrames** that simplify data handling. With Pandas, we can perform complex data analyses, making it indispensable for exploring heart disease datasets.

Exploratory Data Analysis Techniques

Effective **exploratory data analysis** involves techniques such as data cleaning, visualization, and statistical analysis. Using **NumPy** and **Pandas**, we can uncover trends, correlations, and anomalies in heart disease data, leading to more informed clinical decisions.

Heart_Disease_Diagnostic_Analysis.ipynb

File Edit View Insert Runtime Tools Help Last edited on 28 June

Comment Share

+ Code + Text

Connect Gemini

```
[ ] # Converting the Heart disease dataframe into a CSV file
    HeartDisease_df.to_csv('Final Heart Disease data.csv')
```

Conclusion:

After conducting a thorough analysis of the dataset on heart disease diagnostics, several key insights have been derived:

1. **Demographic Insights:**
   - The dataset includes individuals across different age groups, with a notable proportion belonging to middle and elder age categories.
   - The gender distribution shows a mix of male and female individuals, indicating a diverse sample population.
2. **Clinical Parameters:**
   - Various clinical parameters such as blood pressure, cholesterol levels, fasting blood sugar, and maximum heart rate have been recorded for each individual.
   - The dataset provides a comprehensive overview of the health status of the individuals, allowing for a detailed analysis of potential risk factors for heart disease.
3. **Risk Factors Analysis:**
   - Exploration of risk factors such as age, gender, and chest pain type reveals potential correlations with the presence or absence of heart disease.
   - Further analysis of risk factors like exercise-induced angina, ST depression, and number of major vessels colored by fluoroscopy can provide deeper insights into the diagnostic process.
4. **Healthcare Implications:**
   - The findings from this analysis can contribute to better understanding and early detection of heart disease, enabling healthcare professionals to implement targeted prevention and intervention strategies.
   - Identifying individuals at higher risk of heart disease can lead to personalized healthcare approaches, including lifestyle modifications and medical interventions.
5. **Further Research Directions:**
   - Future research can focus on exploring additional features or incorporating external datasets to enhance the predictive power of the models.
   - Longitudinal studies and clinical trials may validate the findings and assess the effectiveness of predictive models in real-world healthcare settings.

In conclusion, the analysis of this heart disease diagnostic dataset provides valuable insights into the factors influencing the presence or absence of heart disease. By leveraging advanced analytics and predictive modeling techniques, we can strive towards improving diagnostic accuracy, patient outcomes, and overall cardiovascular health management.

Case Studies in Diagnostics

Real-world case studies demonstrate the power of **NumPy** and **Pandas** in heart disease diagnostics. By analyzing patient data, we can identify risk factors and improve treatment protocols, ultimately enhancing patient care and outcomes.

Conclusion and Future Directions

In conclusion, leveraging **NumPy** and **Pandas** for heart disease diagnostics significantly enhances exploratory data analysis. Future advancements in data analytics will further improve our understanding and treatment of heart disease, paving the way for better healthcare solutions.