VISVESVARAYA TECHNOLOGICAL UNIVERSITY

"Jnana Sangama", Belagavi – 590 018

An Internship Report on

"HEART FAILURE"

Submitted in partial fulfillment for the award of degree of

Bachelor of Engineering

in

Computer Science Engineering

Submitted By

Mohammed Anas (4AD18CS040)

Internship Carried Out at

AUDAZ VENTURES PVT. LTD.

INTERNAL GUIDE

ATME College of Engineering
Mysuru

EXTERNAL GUIDE

Mr. Vikrant Kumar
L&D Manager
AUDAZ Venture PVT. LTD.

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

ATME COLLEGE OF ENGINEERING

13th KM Stone, Mysuru – Bannur Road, Mysuru – 570028

Phone: 0821 2954081

Website: www.atme.in

2022 – 2023

ATME College of Engineering

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the internship report titled **"HEART FAILURE"** is carried out by **Mohammed Anas** bearing USN **4AD18CS040** in partial fulfillment of the requirements for the award of **Bachelor of Engineering in Computer Science and Engineering** of **Visvesvaraya Technological University, Belagavi** during the year **2022-2023**.

Signature of the Internal Guide	**Signature of HOD**	**Signature of the External Guide**
Dept. of CSE ATME	**Dr. Puttegowda** Professor & HOD Dept. of CSE ATME	**Mr. Vikrant Kumar** L&D Manager AUDAZ Ventures Private Limited

DECLARATION

I hereby declare that I have completed my four weeks Internship at **"AUDAZ VENTURES PVT. LTD.",** from 21-08-2022 to 17-09-2022 under the guidance of the internal guide. I have declared that I have worked with full dedication during these four weeks of Internship in partial fulfillment for the award of the degree of **Bachelor of Engineering in Computer Science and Engineering** from **Visvesvaraya Technological University, Belagavi** during the year **2022 – 23**

Mohammed Anas (4AD18CS040)

ACKNOWLEDGEMENT

The success and outcome of this internship required a lot of guidance and assistance from many people, and I am extremely privileged to have got this all along the completion of my internship. I thank to all those who have rendered their cherished advice and services towards the completion of the internship.

I wish to express my deep sense of acknowledgement and gratitude to my Internal guide Department of Computer Science and Engineering, for the suggestions and encouragement throughout the making of the internship.

I wish to express my deep sense of acknowledgement and gratitude to my External guide **Mr. Vikrant Kumar,** L&D Manager, Audaz Ventures Private Limited, for the suggestions and encouragement throughout the making of the internship.

I wish to express my deep sense of acknowledgement and gratitude to my Trainers **Mr. Azhar**, **Mr. Mohammed Nawfal** and **Mr. Mohammed Sadiq** for their training and constant support.

I am highly indebted to **Dr. Puttegowda**, Head of Department, Computer science and Engineering, for his kind consents and wholehearted cooperation.

I would like to thank our Principal **Dr. L Basavaraj**, for his encouragement and providing an excellent working environment.

I thank all the lecturers of the dept. for their cooperation and providing with the facilities to carry out the seminar work. I would also express my thanks to all technical and non-technical staff of the Computer Science and Engineering department who have directly or indirectly cooperated with me.

Finally, I would like to express my gratitude to my parents and friends who always stood by me encouraging in all my endeavors.

Mohammed Anas (4AD18CS040)

INDEX

EXECUTIVE SUMMARY

The internship report highlights the major works carried out in terms of academic and non-academic perspectives. The scope of this document is to identify and describe the analysis carried out, project completed, experience gained and focuses on the achievements as an intern.

I was working with AUDAZ VENTURES PVT. LTD. to complete my internship. I found myself rather lucky by getting the chance to work in such an environment that AUDAZ provided and got introduced to some of the new terms, new technologies and languages.

The project that I worked in certainly helped me by increasing my practical knowledge depth. The research and analysis projects were particularly helpful in widening my views regarding data analysis tools and data science. Besides there were some vital lessons which will obviously help in future jobs.

This report intended to describe the analysis of the Heart Failure Prediction. The analysis provides the data to analyze and predict which patients are likely to have heart problems.

The Chapters One to Six provide the introduction to the project and data analytics. The Chapter Seven explains the data used and data processing for analysis. Finally, in the Chapter Eight to ten the analysis and conclusions based on the project analysis.

INTRODUCTION

In medical field, it is seen that irrespective of age, people experience heart problems due to unhealthy lifestyle.

Medical field is considered to treat patients and to provide a better diet and lifestyle. As each sector of the medical is growing, data is building up day by day, we need to keep the record of the data which can be helpful for the analytics and evaluation. Now we don't have data in gigabyte or terabyte but in petabyte and zettabyte and this data cannot be handled with the day-to-day software such as Excel or MATLAB. Therefore, in this report we will be dealing with large data sets with the high-level programming language 'Python'.

The main goal of this project is to aggregate and analyze the data collected from the different data sources available on the internet. This project mainly focuses on the usage of the Python programming language and its data analytics libraries in the field of medical field. This language has not only it's application in the field of just analyzing the data but also for the prediction of the upcoming scenarios.

The purpose of using this specific language is due to its versatility, vast libraries (Pandas, NumPy, Matplotlib, etc.), speed limitations, and ease of learning. We will be analyzing large housing prices datasets in this project which cannot be easily analyzed in other tools as compared to Python. Python does not have its limitation to only data analytics but also in many other fields such as Artificial intelligence, Machine learning, and many more.

COMPANY PROFILE

Audaz Ventures Private Limited

CIN: U80903DL2020PTC365560

Audaz Ventures Private Limited, with the headquarter in New Delhi and corporate office in Bengaluru was established in 2020. It is a service - based company which provides services and solutions to 62 institutes across India. It has its presence in more than 12 cities across India.

Its services include

- Software solutions
- Digital marketing
- SAP software
- ERP software
- Blockchain software
- Placement related training to the Engineering Graduates and many more
- Placement opportunities to students

Thanks & Regards

Rahul Oberoi

Director

08882264073
09034444243
info@audazlearning.com
C2C/2-100, JANAKPURI,
NEW DELHI - 110058

OBJECTIVES AND SCOPE OF THE STUDY

4.1 Objectives of the study

- To study the data analytics concepts of data gathering, data cleaning, EDA, and result interpretation.
- To study the what is causing heart problems.

4.2 Scope of Study

- The aim of this study is to analyze the data generated by Heart Failure Prediction.
- The scope of the study is limited to the data available is limited as more and patients data is updated every minute.

THEORETICAL BACKGROUND

5.1 What are Data Analytics?

Companies around the globe generate vast volumes of data daily, in the form of log files, web servers, transactional data, and various customer-related data. In addition to this, social media websites also generate enormous amounts of data. Companies ideally need to use all their generated data to derive value out of it and make impactful business decisions. Data analytics is used to drive this purpose.

Data analytics is the process of exploring and analyzing large datasets to find hidden patterns, unseen trends, discover correlations, and derive valuable insights to make business predictions. It improves the speed and efficiency of your business. Businesses use many modern tools and technologies to perform data analytics.

5.2 Ways to Use Data Analytics

- **Improved Decision Making**

  Data Analytics eliminates guesswork and manual tasks. Be it choosing the right content, planning marketing campaigns, or developing products. Organizations can use the insights they gain from data analytics to make informed decisions. Thus, leading to better outcomes and customer satisfaction.

- **Better Customer Service**

  Data analytics allows you to tailor customer service according to their needs. It also provides personalization and builds stronger relationships with customers. Analyzed data can reveal information about customers' interests, concerns, and more. It helps you give better recommendations for products and services.

- **Efficient Operations**

  With the help of data analytics, you can streamline your processes, save money, and boost production. With an improved understanding of what your audience wants, you spend lesser time creating ads and content that aren't in line with audience's interests.

- **Effective Marketing**

  Data analytics gives you valuable insights into how your campaigns are performing. This helps in fine-tuning them for optimal outcomes. Additionally, you can also find potential customers who are most likely to interact with a campaign and convert into leads.

5.3 Steps Involved in Data Analytics

There are a few steps that are involved in the data analytics lifecycle. Below are the steps that you can take to solve your problems.

- **STEP 1: Identify**

  Identifying or understanding the business problems, defining the organizational goals, and planning a lucrative solution is the first step in the analytics process.
- **STEP 2: Collection**

  Collection or Data collection, is need to collect transactional business data and customer-related information from the past few years to address the problems your business is facing.
- **STEP 3: Clean**

  Clean or data clean, is the data you collect will often be disorderly, messy, and contain unwanted missing values. Such data is not suitable or relevant for performing data analysis. Hence, you need to clean the data to remove unwanted, redundant, and missing values to make it ready for analysis.
- **STEP 4: Analyze**

  Analyze or Data exploration and analysis, is you gather the right data, the next vital step is to execute exploratory data analysis. It can be used as data visualization and business intelligence tools, data mining techniques, and predictive modeling to analyze, visualize, and predict future outcomes from this data. Applying these methods can tell you the impact and relationship of a certain feature as compared to other variables.
- **STEP 5: Interpret the results**:

  Interpret or Interpret of the results, is the final step is to interpret the results and validate if the outcomes meet your expectations. You can find out hidden patterns and future trends. This helps to gain insights that supports with appropriate data-driven decision making.

5.4 Data Analytics Tools

Here are 7 data analytics tools, including a couple of programming languages that can help you perform analytics better.

- **Python** is an object-oriented open-source programming language. It supports a range of libraries for data manipulation, data visualization, and data modeling.

- **R** is an open-source programming language majorly used for numerical and statistical analysis. It provides a range of libraries for data analysis and visualization.

- **Tableau** is a simplified data visualization and analytics tool. This helps you create a variety of visualizations to present the data interactively, build reports, and dashboards to showcase insights and trends.

- **Power BI** is a business intelligence tool that has an easy ‘drag and drop functionality. It supports multiple data sources with features that visually appeal to data. Power BI supports features that help you ask questions to your data and get immediate insights.

- **QlikView** offers interactive analytics with in-memory storage technology to analyze vast volumes of data and use data discoveries to support decision making. It provides social data discovery and interactive guided analytics. It can manipulate colossal data sets instantly with accuracy.

- **Apache Spark** is an open-source data analytics engine that processes data in real-time and carry out sophisticated analytics using SQL queries and machine learning algorithm.

5.5 Data Analytics Applications

Data analytics is used in almost every sector of business, below are a few of them:

- **Retail** helps retailers understand their customer needs and buying habits to predict trends, recommend new products, and boost their business. They optimize the supply chain, and retail operations at every step of the customer journey.

- **Healthcare** industries analyze patient data to provide lifesaving diagnoses and treatment options. Data analytics help in discovering new drug development methods as well.

- **Manufacturing** sectors can discover new cost-saving opportunities. They can solve complex supply chain issues, labor constraints, and equipment breakdowns.

- **Banking sector** uses analytics to find out probable loan defaulters and customer churn out rate. It also helps in detecting fraudulent transactions immediately.

- **Logistics** companies use data analytics to develop new business models and optimize routes. This, in turn, ensures that the delivery reaches on time in a cost-efficient manner.

RESEARCH METHODOLOGY

This research is a descriptive analytic, in which we are analyzing the real estate prices in California. It is based on primary data and inferences derived from it.

6.1 Types of Research in Data Analytics

- **Predictive Analytics**

  It turns the data into valuable, actionable information. predictive analytics uses data to determine the probable outcome of an event or a likelihood of a situation occurring and holds a variety of statistical techniques from modeling, machine, learning, data mining, and game theory that analyze current and historical facts to make predictions about a future event.

- **Descriptive Analytics**

  It looks at data and analyze past event for insight as to how to approach future events. It looks at the past performance and understands the performance by mining historical data to understand the cause of success or failure in the past. Almost all management reporting such as sales, marketing, operations, and finance uses this type of analysis. The descriptive model quantifies relationships in data in a way that is often used to classify customers or prospects into groups.

- **Prescriptive Analytics**

  It automatically synthesizes big data, mathematical science, business rule, and machine learning to make a prediction and then suggests a decision option to take advantage of the prediction. Prescriptive analytics goes beyond predicting future outcomes by also suggesting action benefit from the predictions and showing the decision maker the implication of each decision option. Prescriptive Analytics not only anticipates what will happen and when to happen but also why it will happen.

- **Diagnostic Analytics**

  It generally uses historical data over other data to answer any question or for the solution of any problem. We try to find any dependency and pattern in the historical data of the particular problem.

ANALYSIS AND INTERPRETATION OF DATA

Before analyzing and visualization we need the raw data, and this raw data can be gathered from different open-source data websites available on the internet. This data will be in raw CSV form, it may be individual subunits of data or the complete dataset containing all the groups of files.

7.1 STEP 1: EXTRACT THE DATA

```
df = pd.read_csv("heart.csv")
df
```

	Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngina	Oldpeak	ST_Slope	HeartDisease
0	40	M	ATA	140	289	0	Normal	172	N	0.0	Up	0
1	49	F	NAP	160	180	0	Normal	156	N	1.0	Flat	1
2	37	M	ATA	130	283	0	ST	98	N	0.0	Up	0
3	48	F	ASY	138	214	0	Normal	108	Y	1.5	Flat	1
4	54	M	NAP	150	195	0	Normal	122	N	0.0	Up	0
...	...	...	...	...	...	...	...	...	...	...	...	...
913	45	M	TA	110	264	0	Normal	132	N	1.2	Flat	1
914	68	M	ASY	144	193	1	Normal	141	N	3.4	Flat	1
915	57	M	ASY	130	131	0	Normal	115	Y	1.2	Flat	1
916	57	F	ATA	130	236	0	LVH	174	N	0.0	Flat	1
917	38	M	NAP	138	175	0	Normal	173	N	0.0	Up	0

918 rows × 12 columns

- **Age:** Shows the age of patients
- **Sex:** Shows the gender of patients
- **ChestPainType:** Shows what is the cause
- **RestingBP:** Shows patients normal blood pressure
- **FastingBS:** Shows patients blood sugar readings
- **Cholesterol:** Shows patients cholesterol level
- **RestingECG:** Shows patents normal ECG level
- **MaxHR:** Shows patients HR
- **ExerciseAngina:** Shows patients data whether they exercise or not
- **Oldpeak:** Shows patients data on peak
- **ST_Slope:** Shows patients medical history
- **HeartDisease:** Shows whether the patients has heart disease or not

7.2 STEP 2: DATA CLEANING

```
1 df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 918 entries, 0 to 917
Data columns (total 12 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   Age             918 non-null    int64
 1   Sex             918 non-null    object
 2   ChestPainType   918 non-null    object
 3   RestingBP       918 non-null    int64
 4   Cholesterol     918 non-null    int64
 5   FastingBS       918 non-null    int64
 6   RestingECG      918 non-null    object
 7   MaxHR           918 non-null    int64
 8   ExerciseAngina  918 non-null    object
 9   Oldpeak         918 non-null    float64
 10  ST_Slope        918 non-null    object
 11  HeartDisease    918 non-null    int64
dtypes: float64(1), int64(6), object(5)
memory usage: 86.2+ KB
```

info() method prints information about the DataFrame

The information contains the number of columns, column labels, column data types, memory usage, range index, and the number of cells in each column (non-null values)

```
1 df.describe()
```

	Age	RestingBP	Cholesterol	FastingBS	MaxHR	Oldpeak	HeartDisease
count	918.000000	918.000000	918.000000	918.000000	918.000000	918.000000	918.000000
mean	53.510893	132.396514	198.799564	0.233115	136.809368	0.887364	0.553377
std	9.432617	18.514154	109.384145	0.423046	25.460334	1.066570	0.497414
min	28.000000	0.000000	0.000000	0.000000	60.000000	-2.600000	0.000000
25%	47.000000	120.000000	173.250000	0.000000	120.000000	0.000000	0.000000
50%	54.000000	130.000000	223.000000	0.000000	138.000000	0.600000	1.000000
75%	60.000000	140.000000	267.000000	0.000000	156.000000	1.500000	1.000000
max	77.000000	200.000000	603.000000	1.000000	202.000000	6.200000	1.000000

describe() method returns description of the data in the Data Frame.
If the Data Frame contains numerical data, the description contains this information for each column: count - The number of not-empty values. mean - The average (mean) value. std - The standard deviation.

```
1 df.columns
```

```
Index(['Age', 'Sex', 'ChestPainType', 'RestingBP', 'Cholesterol', 'FastingBS',
       'RestingECG', 'MaxHR', 'ExerciseAngina', 'Oldpeak', 'ST_Slope',
       'HeartDisease'],
      dtype='object')
```

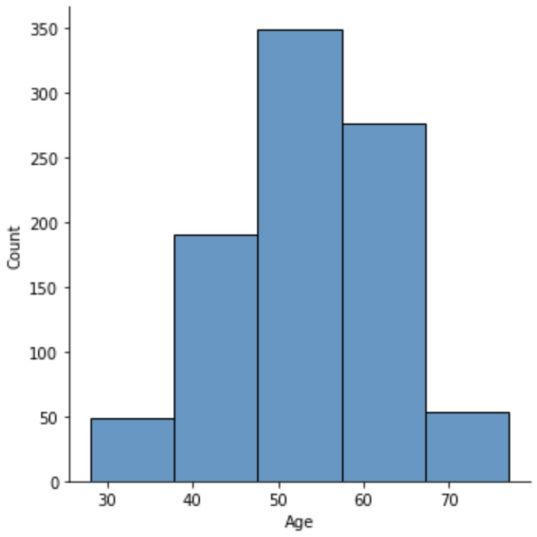
columns displays the data in two-dimensional size-mutable, potentially heterogeneous tabular data structure with labelled axes (rows and columns)

7.3 STEP 3: VIZUALIZAION OF DATA

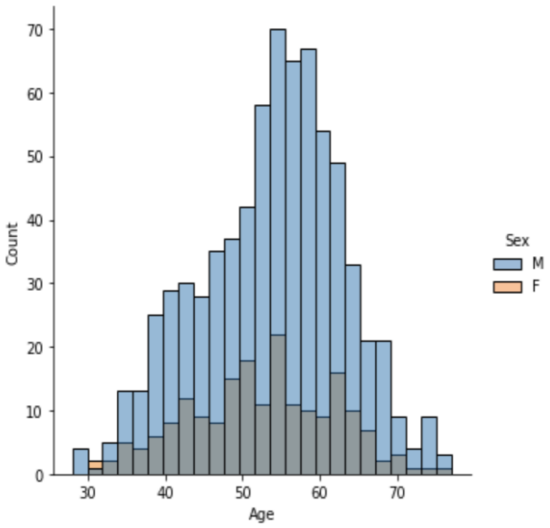
DISTPLOT

```
#Age is normaly distributed
plt.figure(figsize=(16,8))
sns.displot(df,x="Age",bins=5)
plt.show()
plt.figure(figsize=(16,8))
sns.displot(df,x="Age",hue="Sex",bins=25)
plt.show()
```

```
<Figure size 1152x576 with 0 Axes>
```



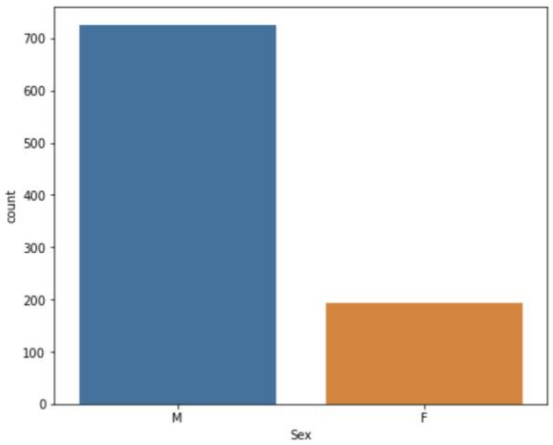

```
<Figure size 1152x576 with 0 Axes>
```

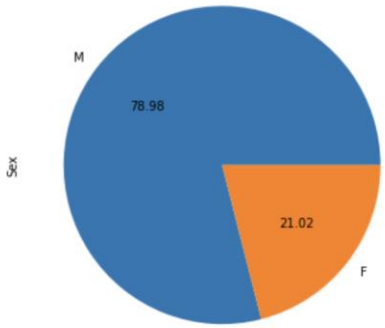


Graph is plotted using Distplot

COUNTPLOT

```
# Aproxx 79% is male and 21 % is female
plt.figure(figsize=(16,6))
plt.subplot(1,2,1)
sns.countplot(df['Sex'])
plt.subplot(1,2,2)
df['Sex'].value_counts().plot(kind='pie',autopct="%.2f")
plt.show()
```

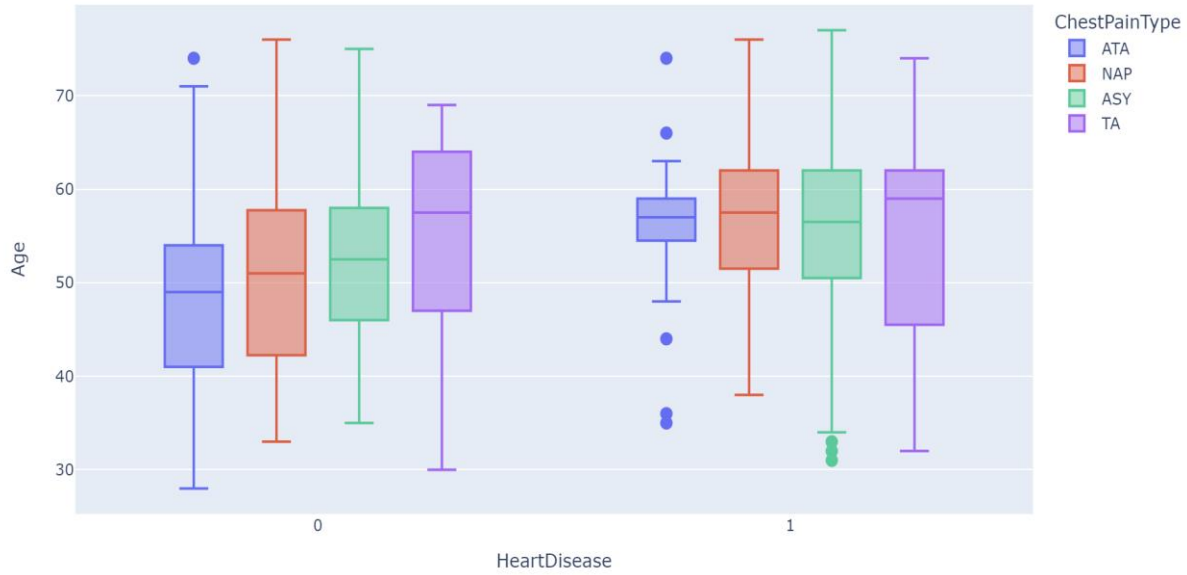




Categories genders consisting of both Bar graph and Pie chart

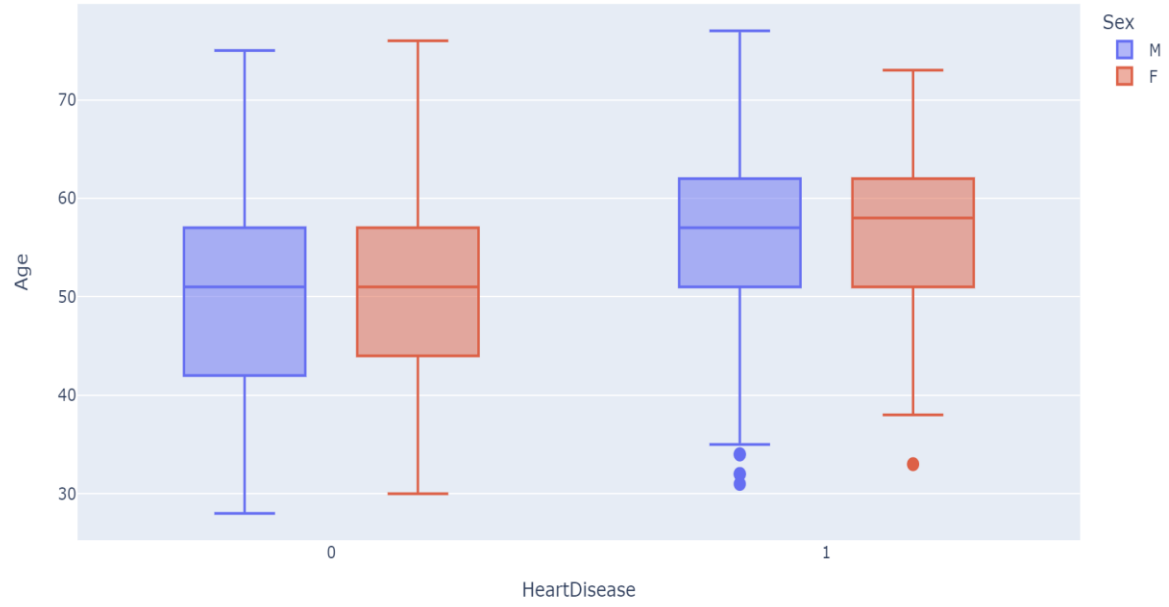
BOXPLOT

```
# Boxplot allows you to examine the distribution of data. One box plot appears for each attribute element.
# outlinear present in ASY chest pain
fig = px.box(df,"HeartDisease","Age", color="ChestPainType")
fig.update_traces(marker_size=10)
fig.show()
```



Box plot based on Heart disease Vs Age with Chest Pain Type as a reference

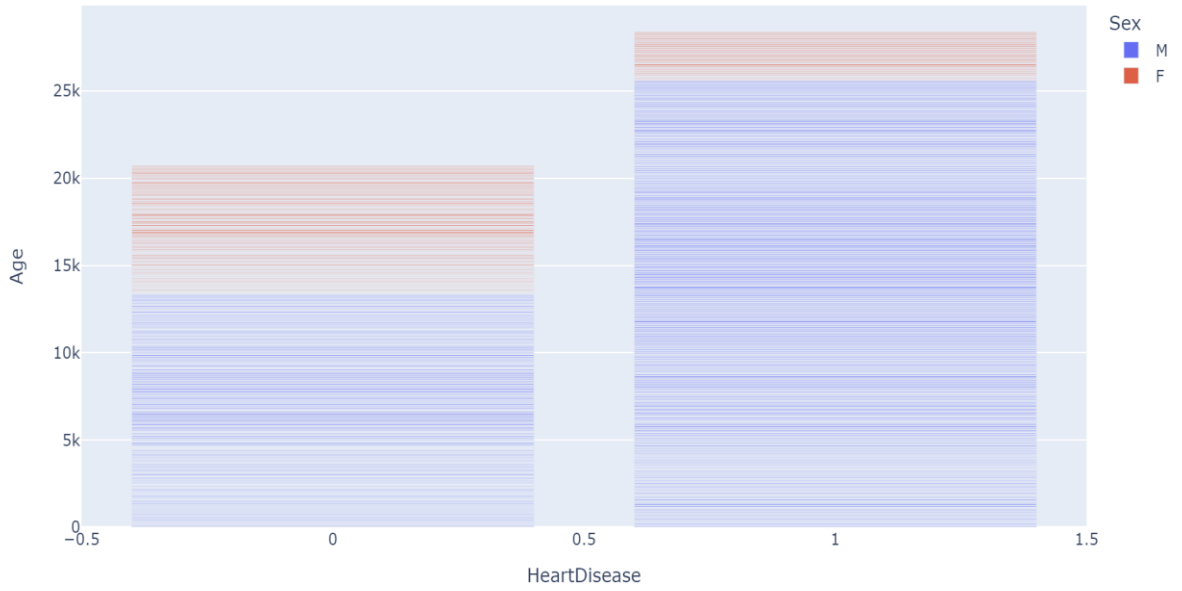
```
fig = px.box(df,"HeartDisease","Age", color="Sex")
fig.update_traces(marker_size=10)
fig.show()
```



Box plot based on Heart disease Vs Age with Sex as a reference

BARGRAPH

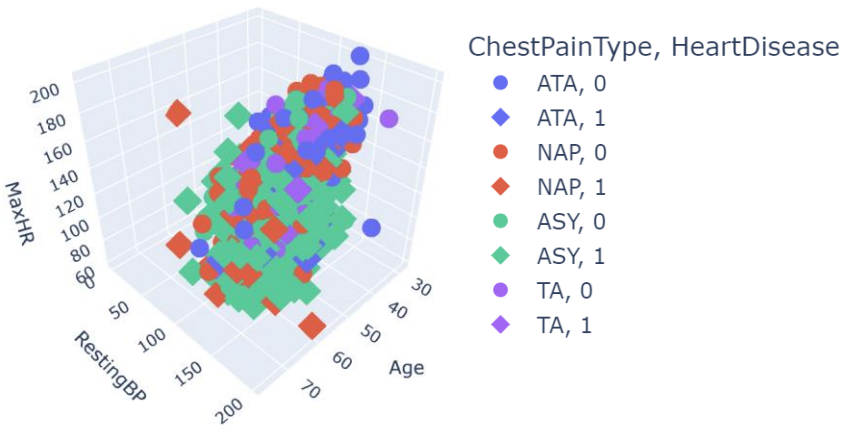
```
# Bargraph is a chart that represents categorical data with rectangular bars with heights proportional value representation
fig  = px.bar(df,'HeartDisease','Age',color='Sex')
fig.show()
```



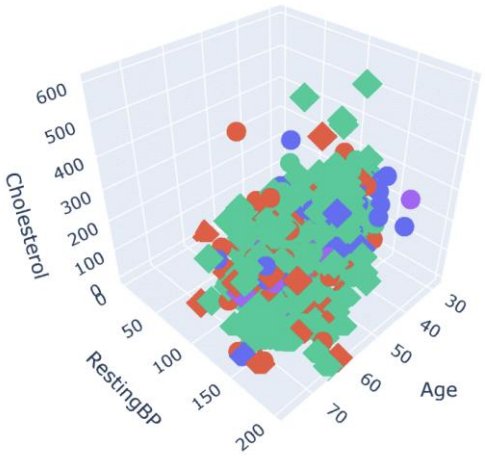
Shows which gender is likely to have heart disease

3D GRAPH

```
#3D plot with Age ,RestingBP,MaxHR
fig = px.scatter_3d(df, x='Age', y='RestingBP', z='MaxHR',color='ChestPainType',symbol='HeartDisease')
fig.show()
#3D plot with Age ,RestingBP, Cholesterol,
fig = px.scatter_3d(df, x='Age', y='RestingBP', z='Cholesterol',color='ChestPainType',symbol='HeartDisease')
fig.show()
```



3D graph based on Age, RestingBP and MaxHR



3D graph based on Age, RestingBP and Cholesterol

FINDINGS

Some of the conclusions that we get from this analysis are:

- It shows which gender has more chances of having heart related diseases
- It shows what is causing the heart problem

SUGGESTIONS / RECOMMENDATIONS

The good health of any human being depends on their individual lifestyle. Here are few suggestions which is purely based on subjective & objective data analysis.

The availability of a particular healthy lifestyles such as diet and exercise should be encouraged by medical professionals to patients which provides with a overall good health and reduces heart problems.

CONCLUSION

The objective of this analysis was to extract useful information for the prediction of Heart Failure based on the Heart Failure Prediction Data.

This study analyzed the reasons for heart disease. The analysis review suggested which gender has more heart problems.

LIMITATIONS

In every research undertaken there are certain unavoidable limitations. This research too has the same. This includes the fact that the data is of a certain only for a certain time, the current trends of data might be not accessible as it is difficult to acquire required information unless new data is made available.

BILBLIOGRAPHY

- https://www.kaggle.com/datasets/andrewmvd/heart-failure-clinical-data
- https://www.simplilearn.com/tutorials/data-analytics-tutorial/what-is-data-analytics
- https://matplotlib.org/stable/index.html
- https://pandas.pydata.org/
- https://seaborn.pydata.org/
- https://www.geeksforgeeks.org/