
HEART
FAILURE DATA
ANALYSIS

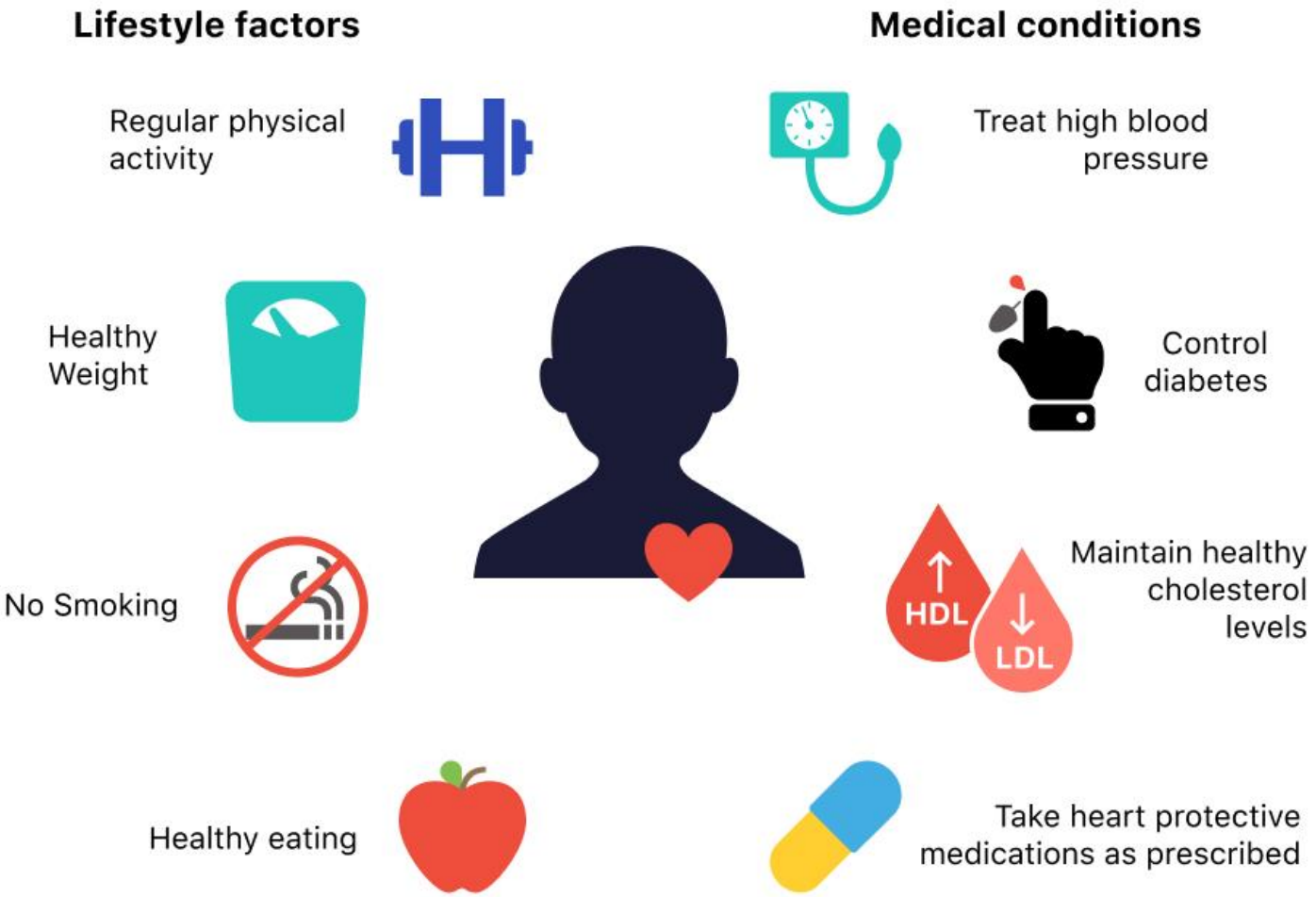
Team Members

Kaushik Chaudhari (845)

Abhishek Khose (830)

Amogh Dandwate (829)

Ways to Reduce Risk of Developing Heart Failure



LIBRARIES USED

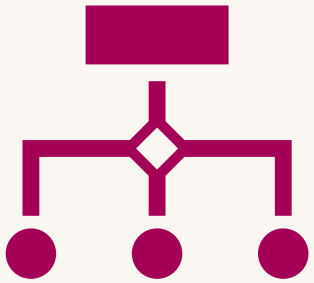
- Pandas
- Numpy
- Matplotlib
- seaborn

Pandas

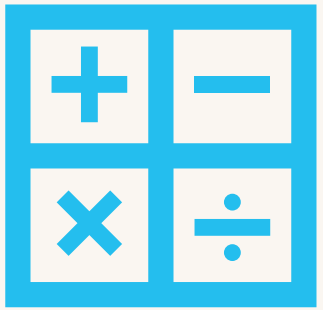
Pandas is an open-source Python package that is most widely used for data science/data analysis and machine learning tasks.

Pandas makes it simple to do many of the time consuming, repetitive tasks associated with working with data.

Numpy



NumPy is a Python library used for working with arrays.



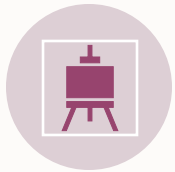
It also has functions for working in domain of linear algebra, fourier transform, and matrices.

Matplotlib

- Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib makes easy things easy and hard things possible.

Create publication quality plots.

Make interactive figures that can zoom, pan, update.



Customize visual style and layout.

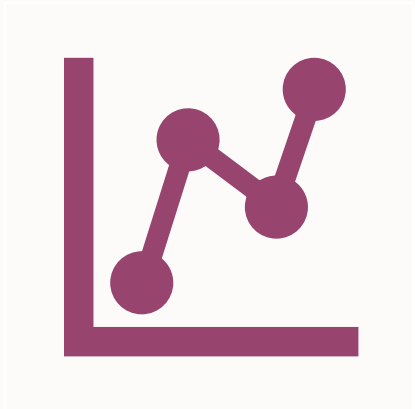
Export to many file formats.

Embed in JupyterLab and Graphical User Interfaces.

Use a rich array of third-party packages built on Matplotlib.

Seaborn

- Seaborn is an open-source Python library built on top of matplotlib. It is used for data visualization and exploratory data analysis. Seaborn works easily with data frames and the Pandas library. The graphs created can also be customized easily



Graphs make it easier to explain your data to non-technical people.

Visually attractive graphs can make presentations and reports much more appealing to the reader.

Code

- https://www.kaggle.com/kaushikchaudhari/heart-failure-anayalsis

KAUSHIK CHAUDHARI · 3H AGO · 6 VIEWS

Heart Failure Anayalsis

Python · Heart Failure Prediction

Notebook Data Logs Comments (0)

Run

30.8s

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

%matplotlib inline
```

```
# Lets import the data first
df=pd.read_csv('/kaggle/input/heart-failure-clinical-data/heart_failure_clinical_records_dataset.csv')
df.head()
```

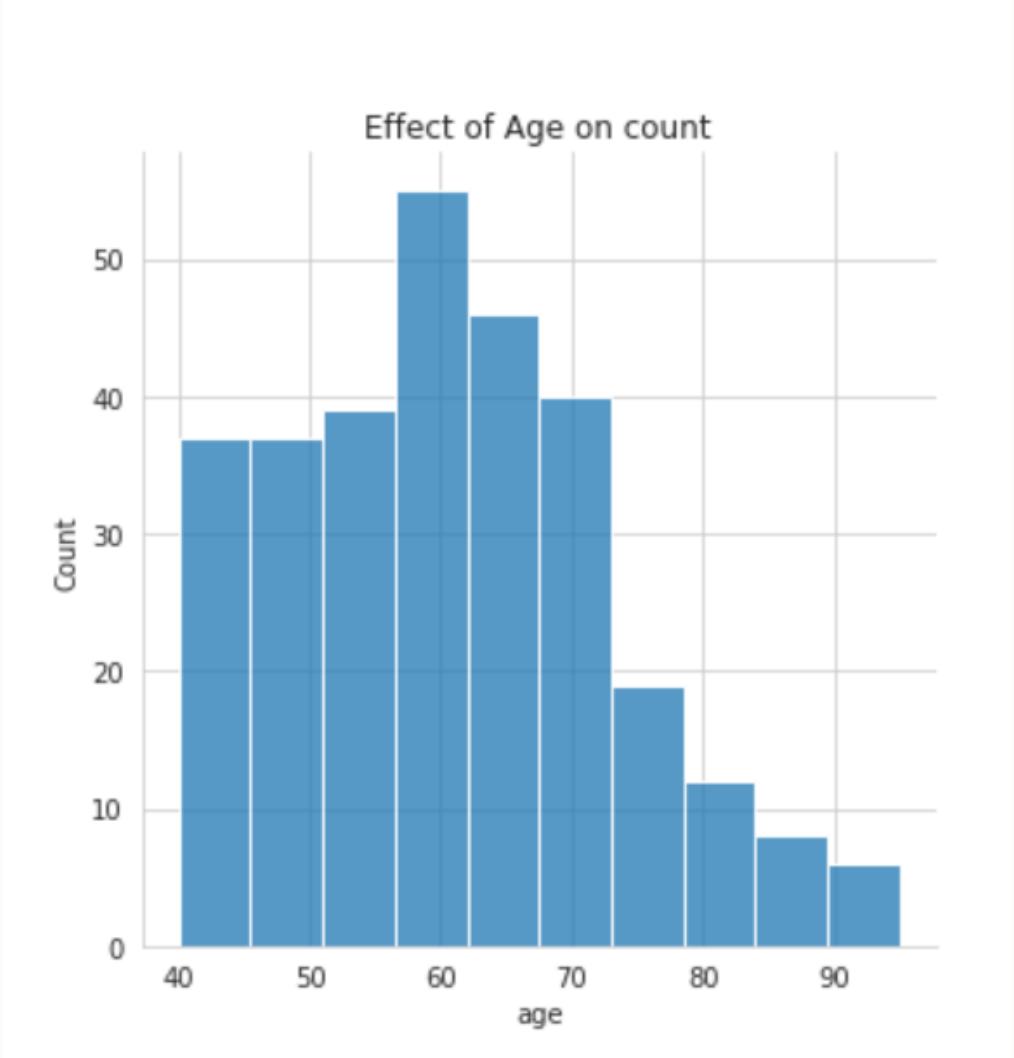
```
RangeIndex: 299 entries, 0 to 298
Data columns (total 13 columns):
 #   Column                    Non-Null Count  Dtype  
---  ------                    --------------  -----  
 0   age                       299 non-null    float64
 1   anaemia                   299 non-null    int64  
 2   creatinine_phosphokinase  299 non-null    int64  
 3   diabetes                  299 non-null    int64  
 4   ejection_fraction         299 non-null    int64  
 5   high_blood_pressure       299 non-null    int64  
 6   platelets                 299 non-null    float64
 7   serum_creatinine          299 non-null    float64
 8   serum_sodium              299 non-null    int64  
 9   sex                       299 non-null    int64  
 10  smoking                   299 non-null    int64  
 11  time                      299 non-null    int64  
 12  DEATH_EVENT               299 non-null    int64  
```

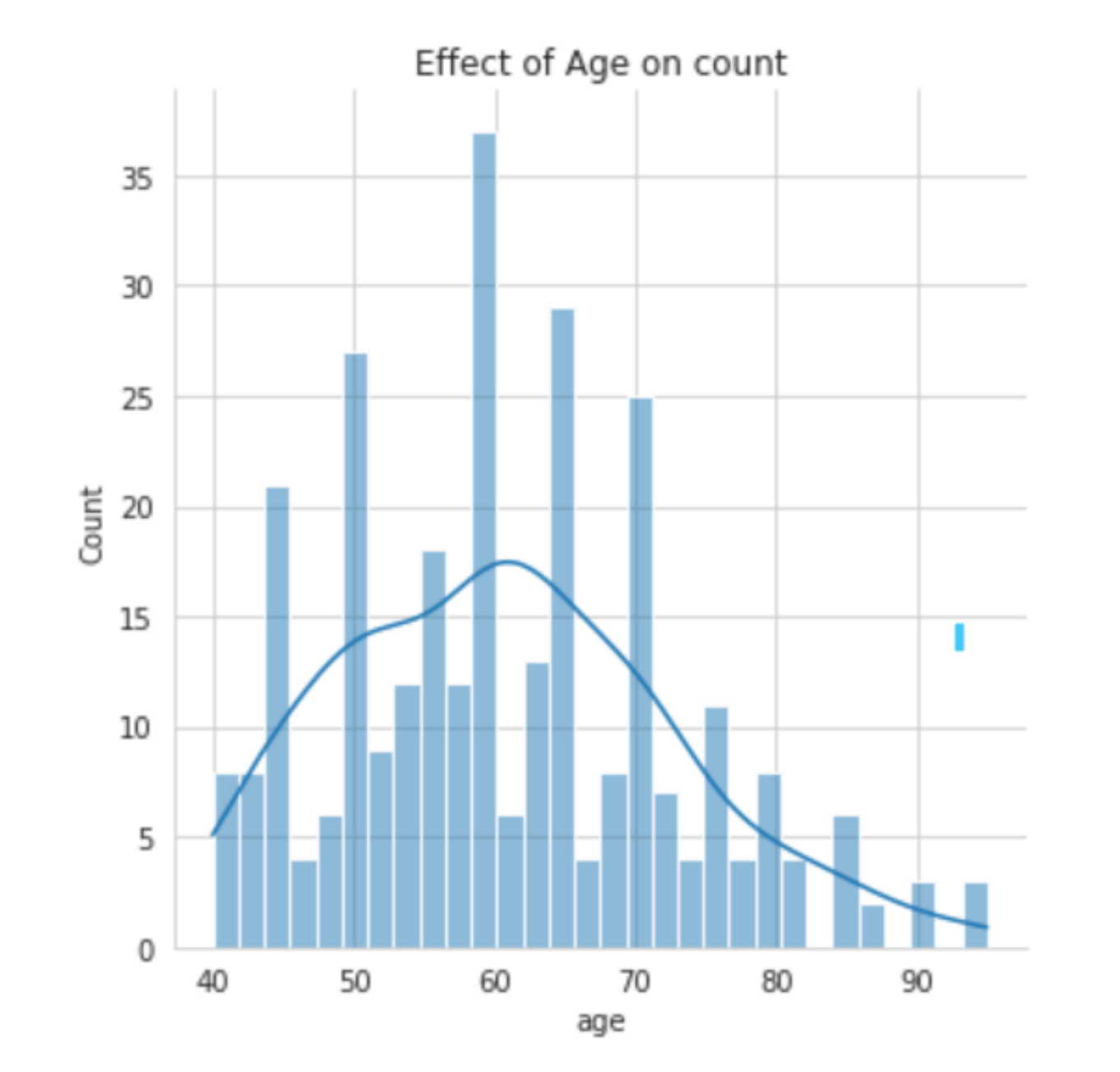
```
df_eda = pd.DataFrame()
df_eda["Age"] = df["age"]
df_eda["Anaemia"] = np.where(df["anaemia"] < 1, "No", "Yes")
df_eda["Creatinine_phosphokinase"] = df["creatinine_phosphokinase"]
df_eda["Diabetes"] = np.where(df["diabetes"] < 1, "No", "Yes")
df_eda["Ejection_fraction"] = df["ejection_fraction"]
df_eda["High_blood_pressure"] = np.where(df["high_blood_pressure"] < 1, "No", "Yes")
df_eda["Platelets"] = df["platelets"]
df_eda["Serum_creatinine"] = df["serum_creatinine"]
df_eda["Serum_sodium"] = df["serum_sodium"]
df_eda["Sex"] = np.where(df["sex"] < 1, "Female", "Male")
df_eda["Smoking"] = np.where(df["smoking"] < 1, "No", "Yes")
df_eda["Time"] = df["time"]
df_eda["Death_event"] = np.where(df["DEATH_EVENT"] < 1, "No", "Yes")
df_eda.head()
```

Out[3]:

	age	anaemia	creatinine_phosphokinase	diabetes	ejection_fraction	high_blood_pressure	platelets	serum_creatinine	serum_sodium	sex	si
0	75.0	0	582	0	20	1	265000.00	1.9	130	1	0
1	55.0	0	7861	0	38	0	263358.03	1.1	136	1	0
2	65.0	0	146	0	20	0	162000.00	1.3	129	1	1
3	50.0	1	111	0	20	0	210000.00	1.9	137	1	0
4	65.0	1	160	1	20	0	327000.00	2.7	116	0	0

Graphical Output




Effect of Age on count
Count
0
5
10
15
20
25
30
35
40
50
60
70
80
90
age

Thank You