Bharatiya Vidya Bhavan's

SARDAR PATEL INSTITUTE OF TECHNOLOGY

Advanced Data Visualization

Experiment no. 3

Submitted To

Prof. Pranav Nerurkar

Submitted By

Name : Dhanashree Mahesh Otari

Class : Btech Computer Engineering

Batch : A

UID : 2021300086

1. Dataset

Can view the dataset here

https://www.kaggle.com/datasets/redwankarimsony/heart-disease-data

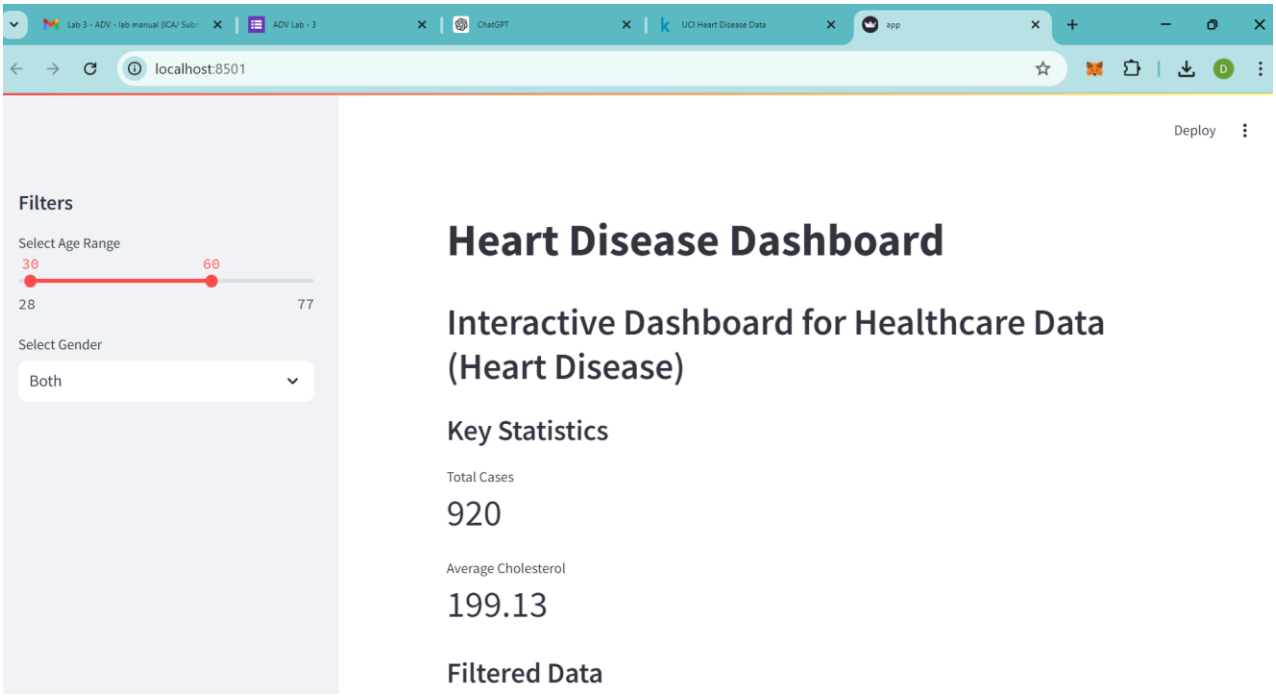
2. Description

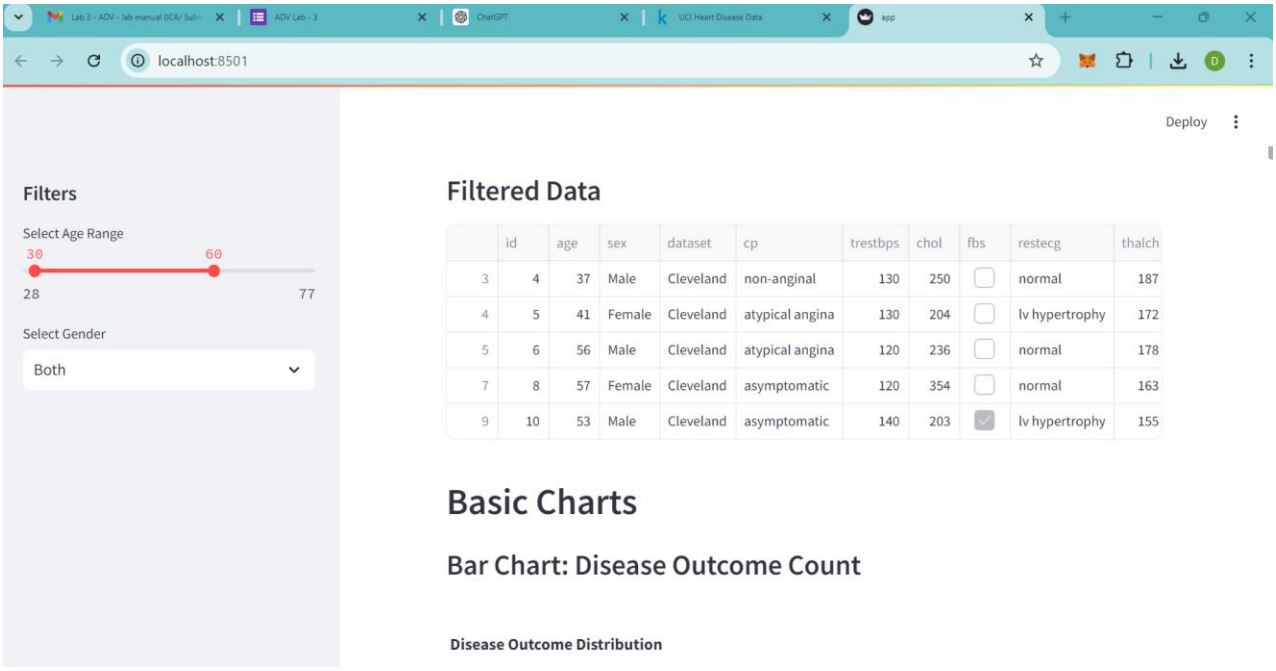
This is a multivariate type of dataset which means providing or involving a variety of separate mathematical or statistical variables, multivariate numerical data analysis. It is composed of 14 attributes which are age, sex, chest pain type, resting blood pressure, serum cholesterol, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise-induced angina, oldpeak — ST depression induced by exercise relative to rest, the slope of the peak exercise ST segment, number of major vessels and Thalassemia. This database includes 76 attributes, but all published studies relate to the use of a subset of 14 of them. The Cleveland database is the only one used by ML researchers to date. One of the major tasks on this dataset is to predict based on the given attributes of a patient that whether that particular person has heart disease or not and other is the experimental task to diagnose and find out various insights from this dataset which could help in understanding the problem more.

3. Metadata

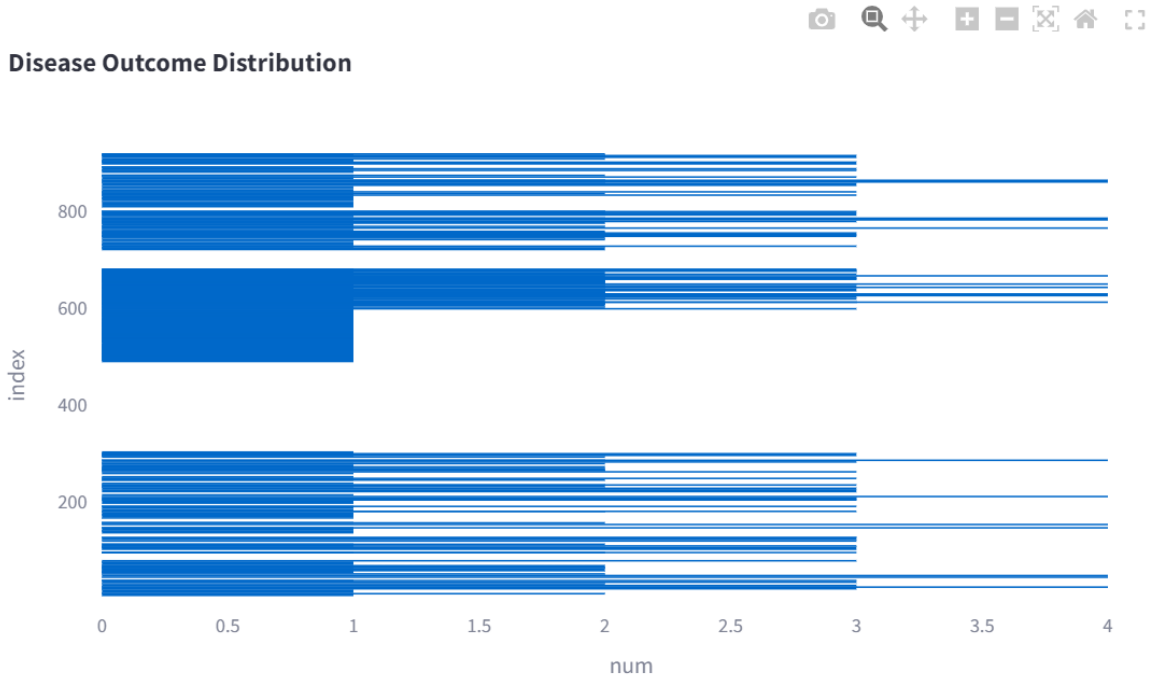
1. id (Unique id for each patient)
2. age (Age of the patient in years)
3. origin (place of study)
4. sex (Male/Female)
5. cp chest pain type ([typical angina, atypical angina, non-anginal, asymptomatic])
6. trestbps resting blood pressure (resting blood pressure (in mm Hg on admission to the hospital))
7. chol (serum cholesterol in mg/dl)
8. fbs (if fasting blood sugar > 120 mg/dl)
9. restecg (resting electrocardiographic results)
   -- Values: [normal, stt abnormality, lv hypertrophy]
10. thalach: maximum heart rate achieved
11. exang: exercise-induced angina (True/ False)
12. oldpeak: ST depression induced by exercise relative to rest
13. slope: the slope of the peak exercise ST segment
14. ca: number of major vessels (0-3) colored by fluoroscopy
15. thal: [normal; fixed defect; reversible defect]
16. num: the predicted attribute

4. Visualizations and Observations



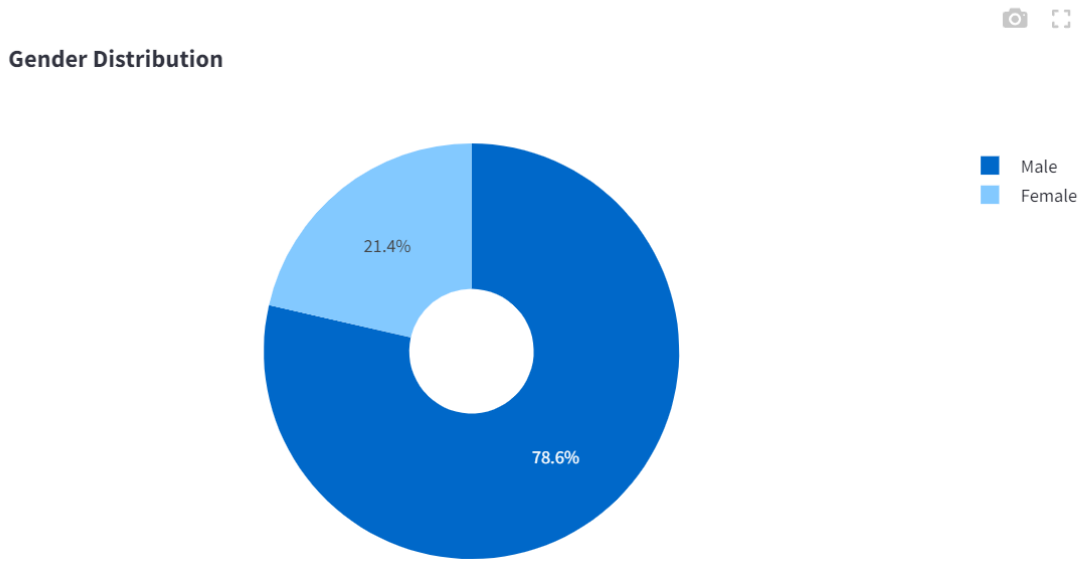


	id	age	sex	dataset	cp	trestbps	chol	fbs	restecg	thalch
3	4	37	Male	Cleveland	non-anginal	130	250	☐	normal	187
4	5	41	Female	Cleveland	atypical angina	130	204	☐	lv hypertrophy	172
5	6	56	Male	Cleveland	atypical angina	120	236	☐	normal	178
7	8	57	Female	Cleveland	asymptomatic	120	354	☐	normal	163
9	10	53	Male	Cleveland	asymptomatic	140	203	☑	lv hypertrophy	155

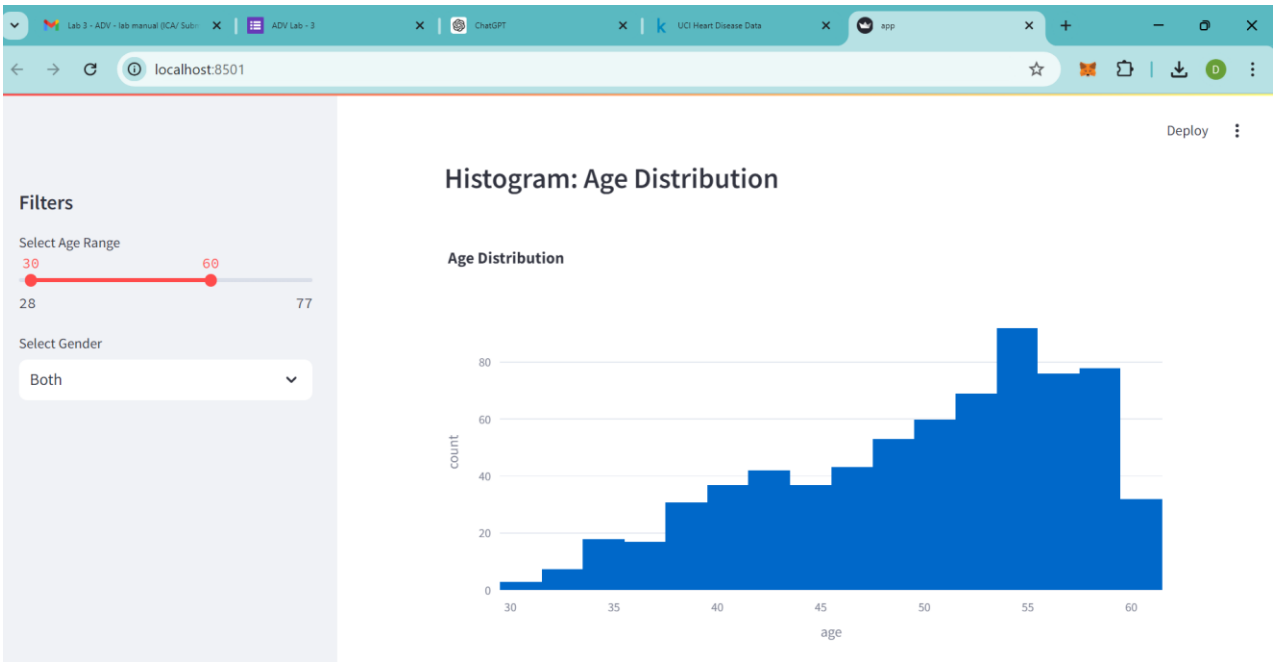


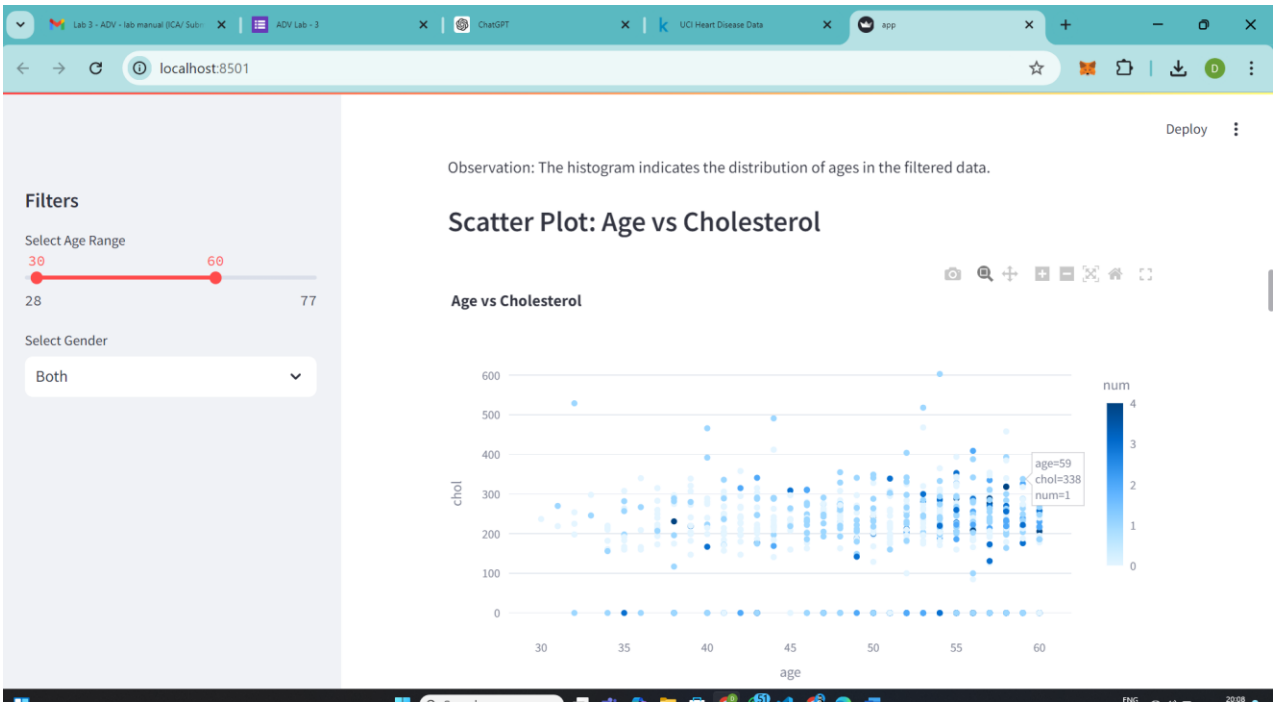
Observation: The bar chart shows the distribution of heart disease outcomes.

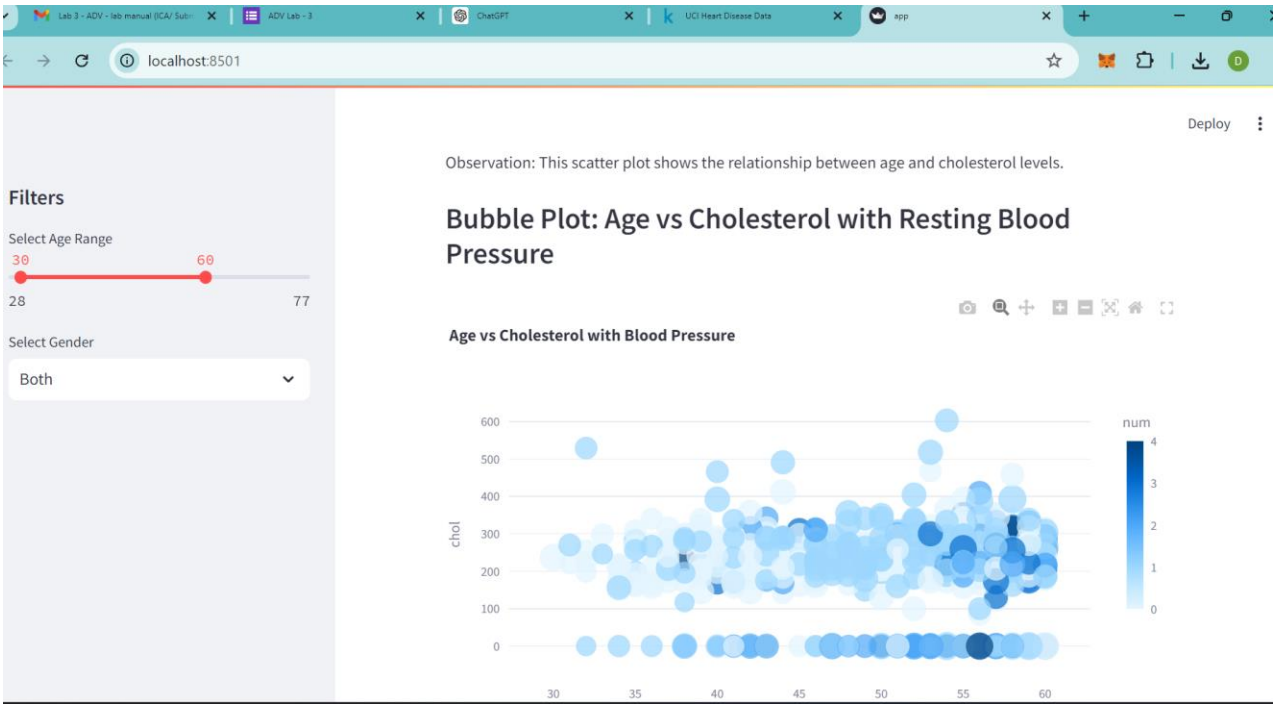
Pie Chart: Gender Distribution

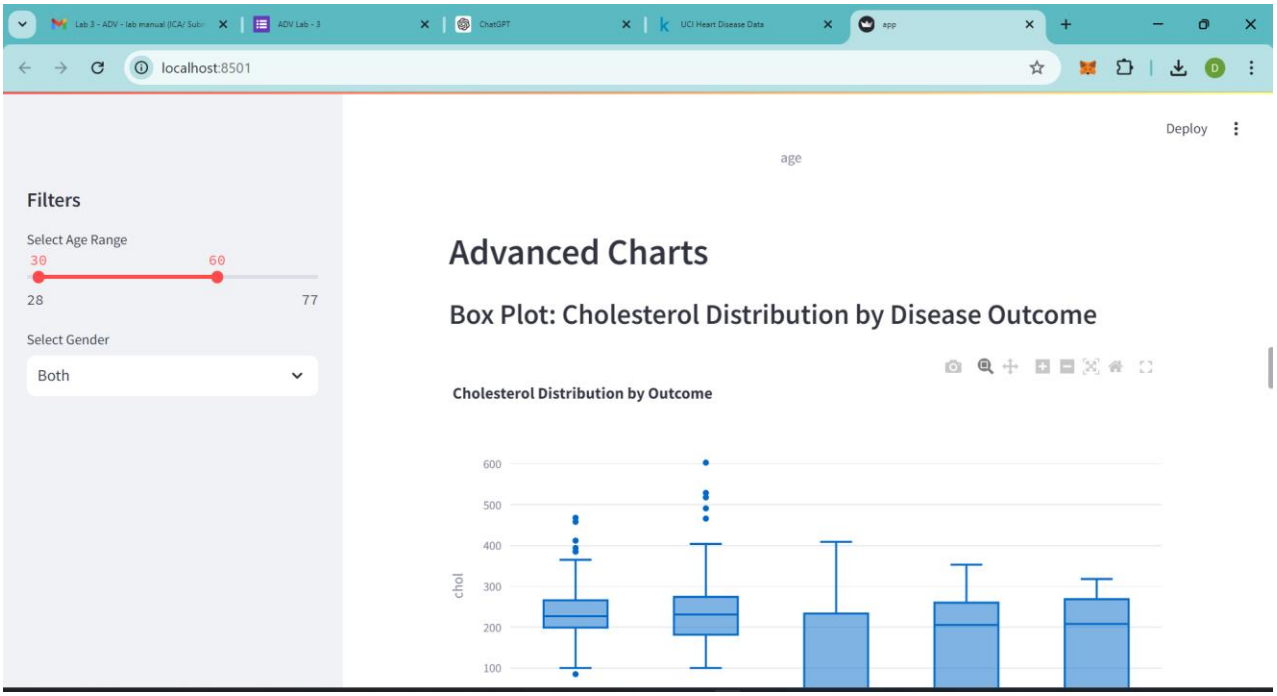


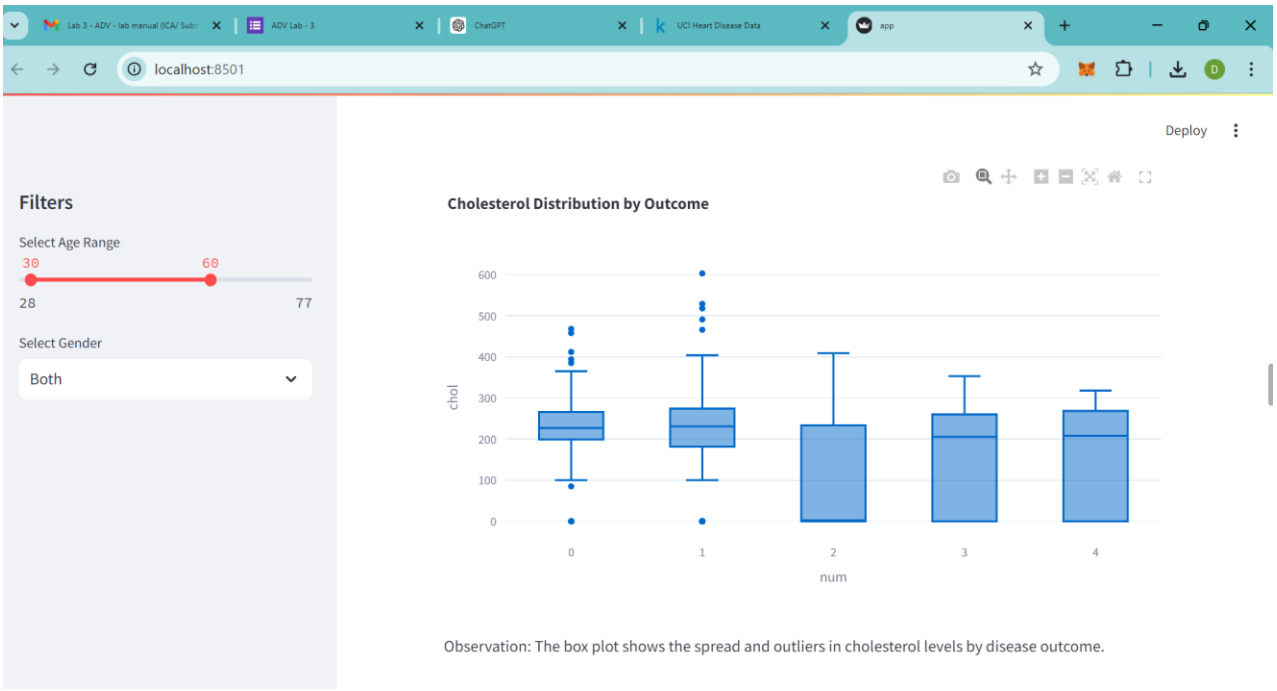
Observation: This pie chart shows the gender distribution of patients.

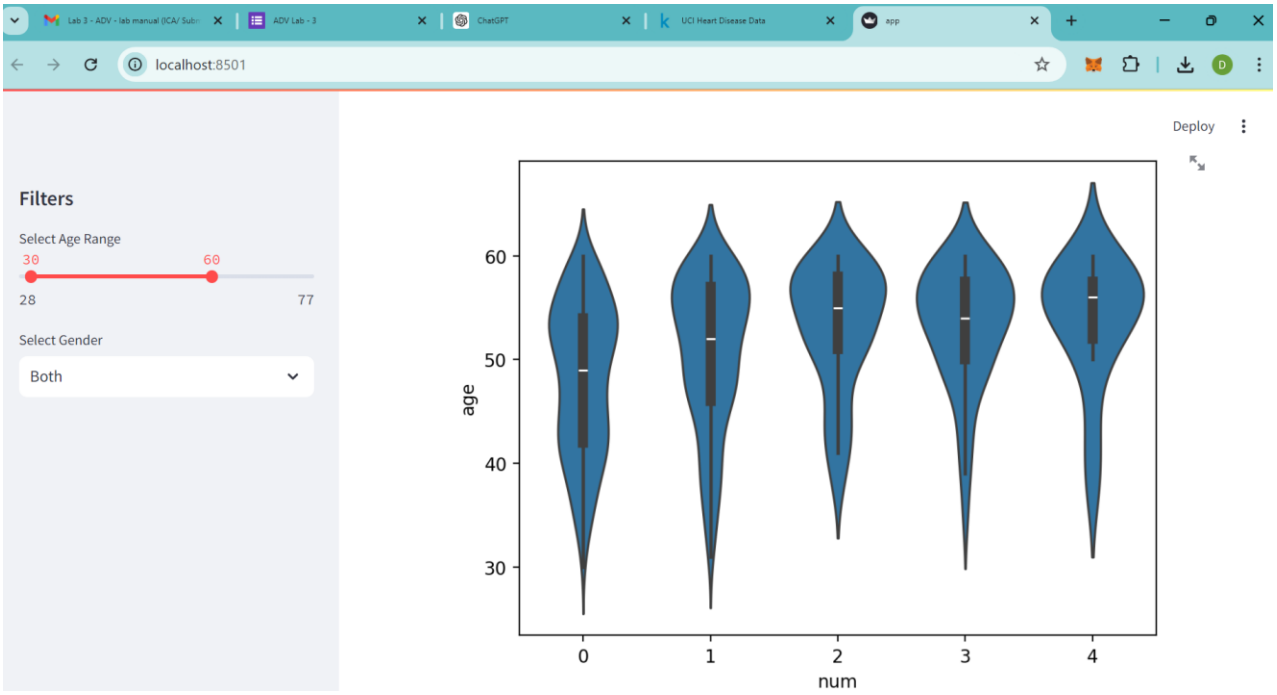

localhost:8501
Deploy
Filters
Select Age Range
30
60
28
77
Select Gender
Both
Histogram: Age Distribution
Age Distribution
count
age


localhost:8501
Deploy
Observation: The histogram indicates the distribution of ages in the filtered data.
Filters
Select Age Range
30
60
28
77
Select Gender
Both
Scatter Plot: Age vs Cholesterol
Age vs Cholesterol
chol
age
num
age=59
chol=338
num=1

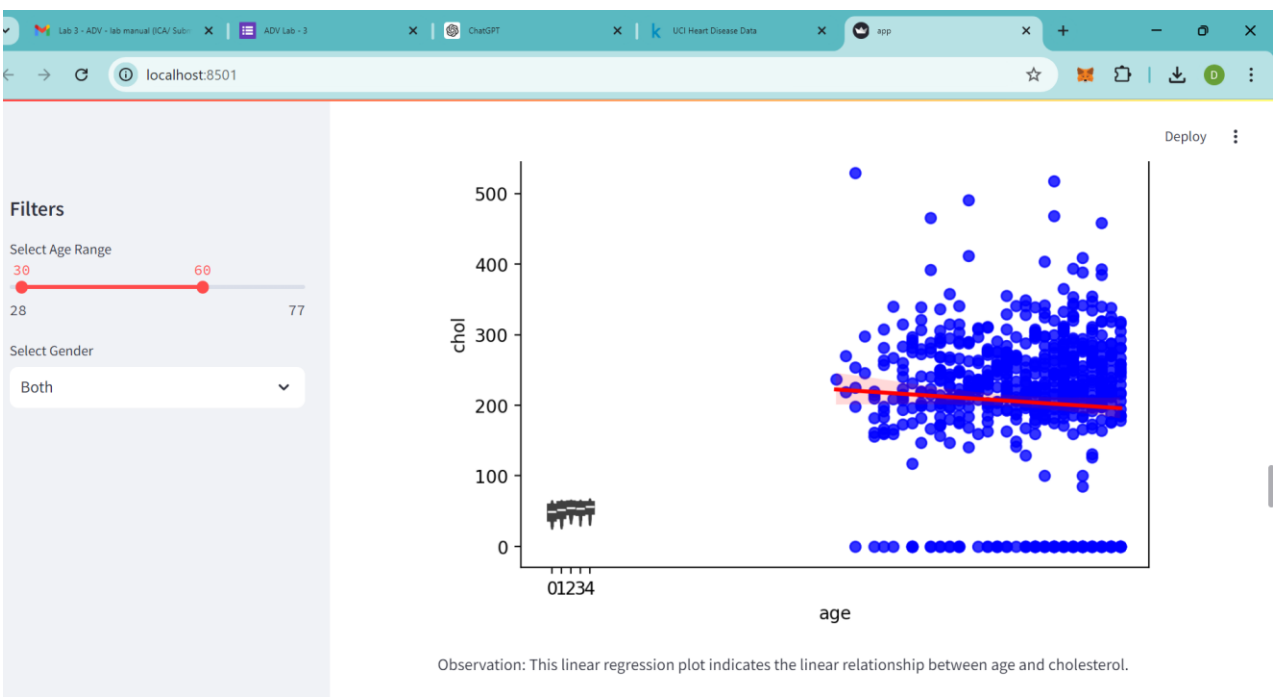

localhost:8501
Deploy
Observation: This scatter plot shows the relationship between age and cholesterol levels.
Filters
Select Age Range
30
60
28
77
Select Gender
Both
Bubble Plot: Age vs Cholesterol with Resting Blood Pressure
Age vs Cholesterol with Blood Pressure
chol
num
600
500
400
300
200
100
0
4
3
2
1
0
30
35
40
45
50
55
60

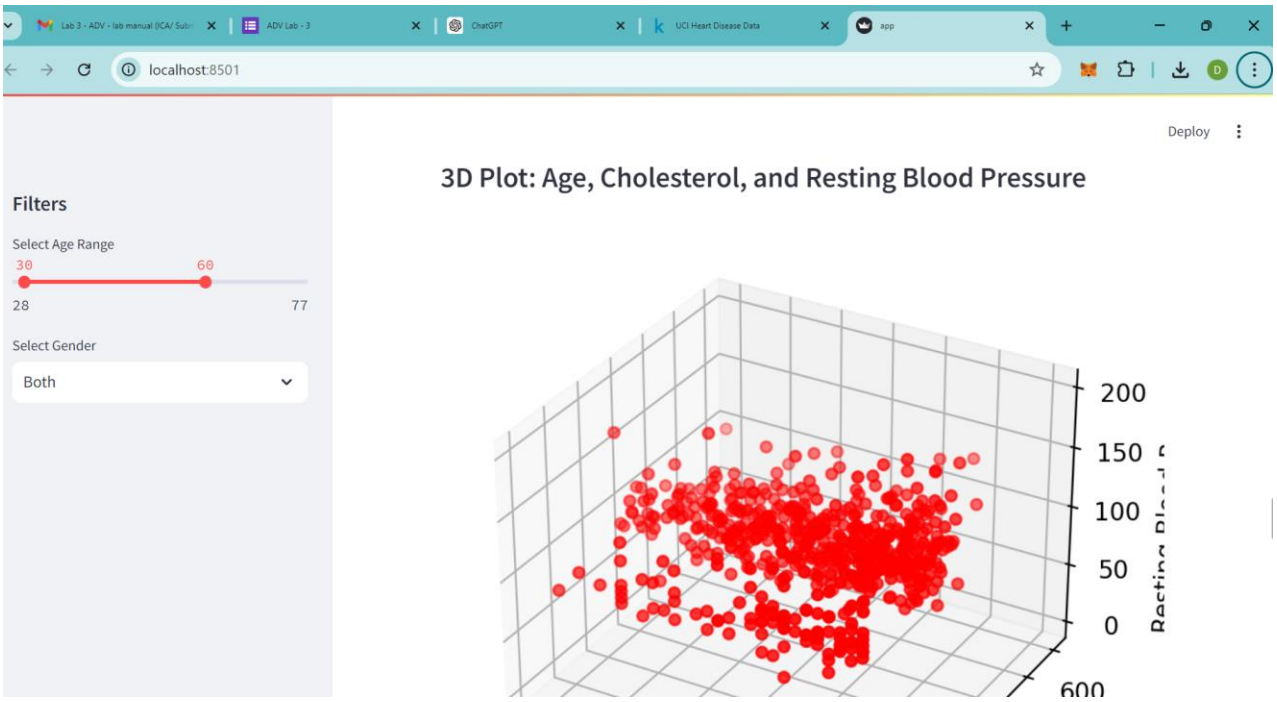

localhost:8501
Deploy
age
Filters
Select Age Range
30
60
28
77
Select Gender
Both
Advanced Charts
Box Plot: Cholesterol Distribution by Disease Outcome
Cholesterol Distribution by Outcome
chol
600
500
400
300
200
100

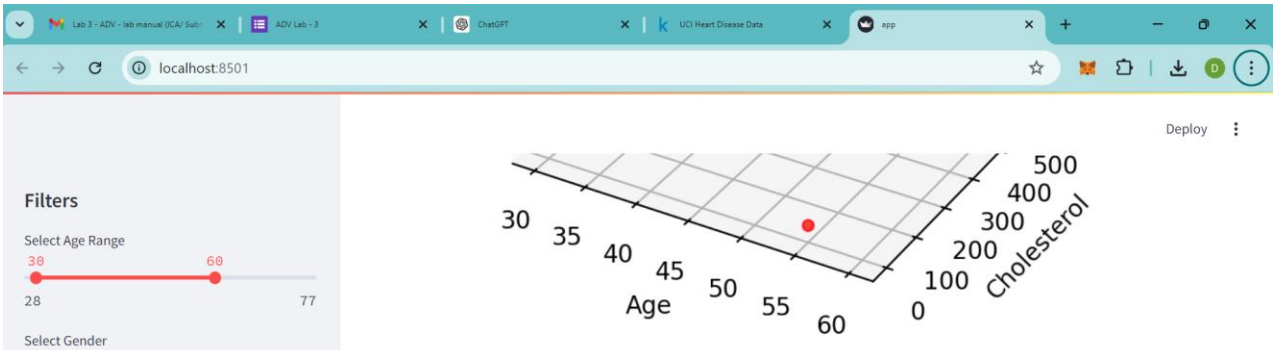

localhost:8501
Deploy
Filters
Select Age Range
30
60
28
77
Select Gender
Both
Cholesterol Distribution by Outcome
chol
num
Observation: The box plot shows the spread and outliers in cholesterol levels by disease outcome.


localhost:8501
Deploy
Filters
Select Age Range
30
60
28
77
Select Gender
Both
age
num


localhost:8501
Deploy
Filters
Select Age Range
30
60
28
77
Select Gender
Both
Observation: The violin plot visualizes the age distribution for patients with and without heart disease.
Linear Regression Plot: Age vs Cholesterol
chol
600
500
400
300
200
100

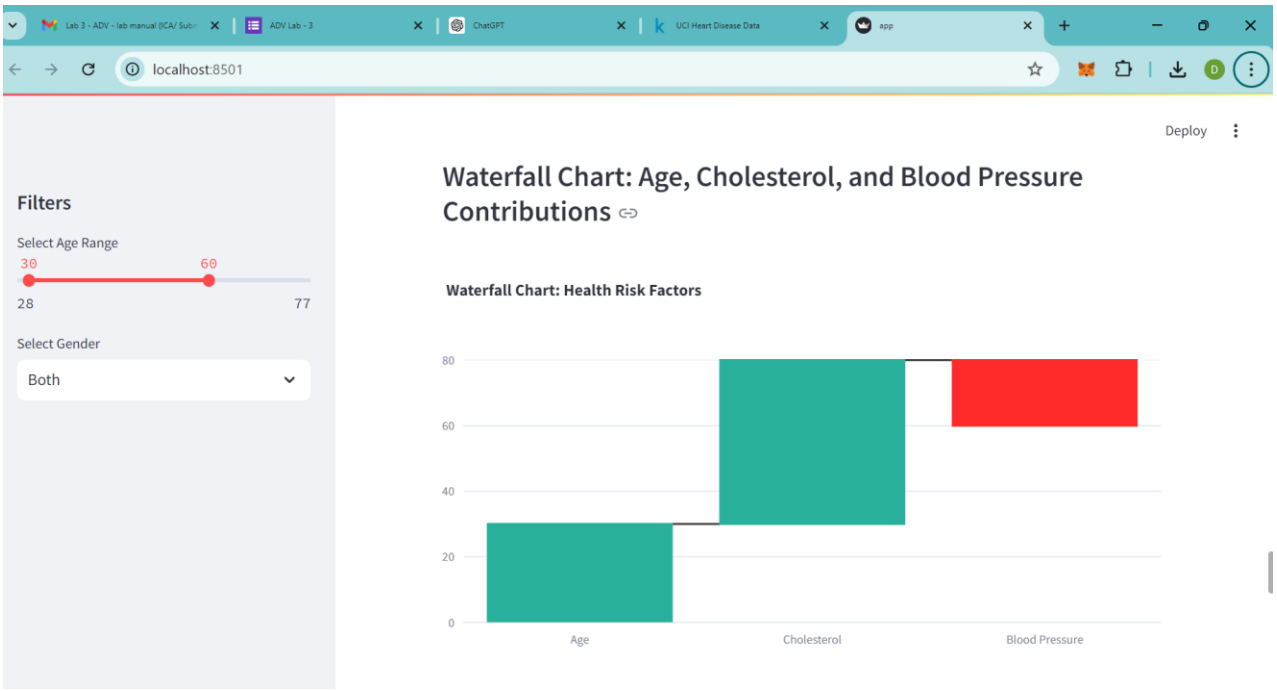

localhost:8501
Deploy
Filters
Select Age Range
30
60
28
77
Select Gender
Both
chol
500
400
300
200
100
0
01234
age
Observation: This linear regression plot indicates the linear relationship between age and cholesterol.

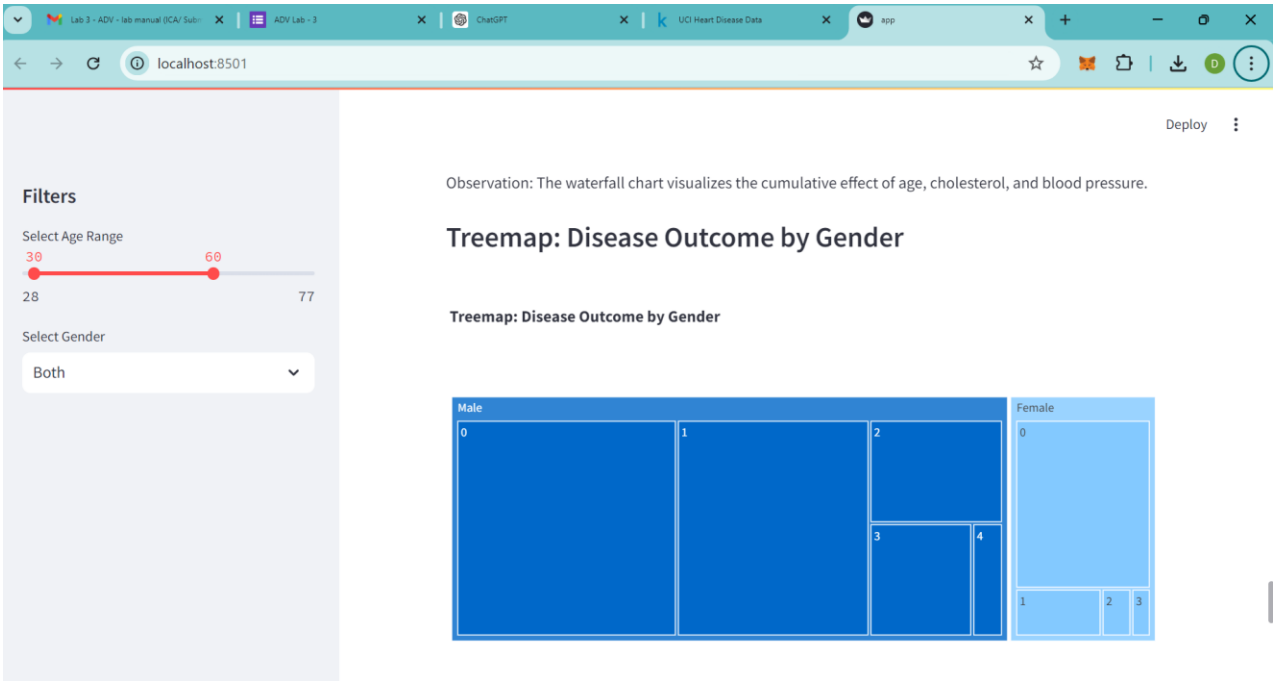


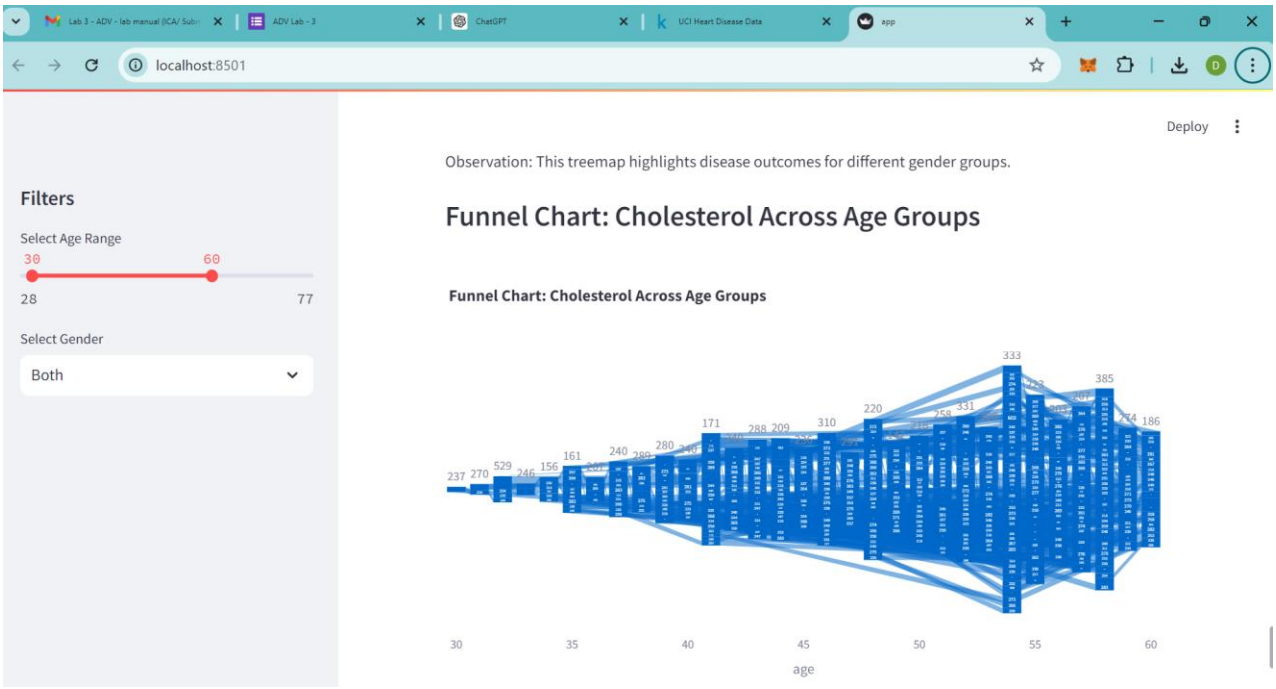


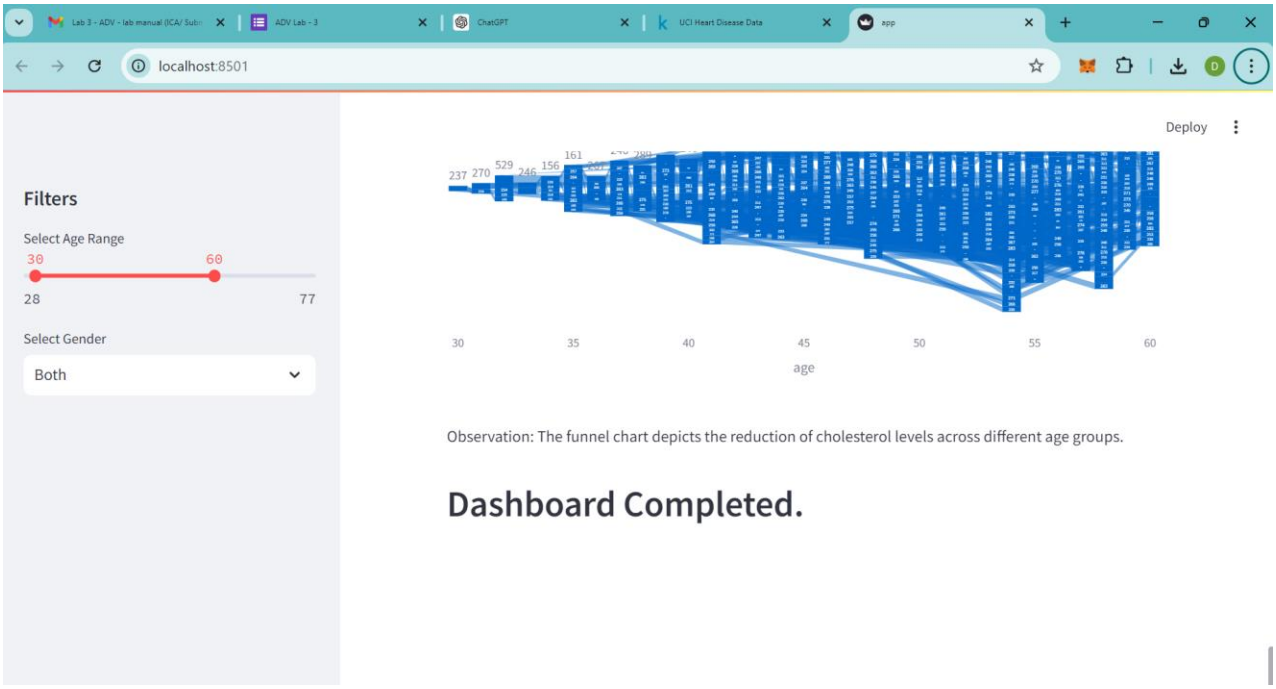
Observation: The 3D chart helps visualize the relationship between age, cholesterol, and blood pressure.

Waterfall Chart: Age, Cholesterol, and Blood Pressure Contributions


localhost:8501
Deploy
Filters
Select Age Range
30
60
28
77
Select Gender
Both
Waterfall Chart: Age, Cholesterol, and Blood Pressure Contributions
Waterfall Chart: Health Risk Factors
80
60
40
20
0
Age
Cholesterol
Blood Pressure


localhost:8501
Deploy
Filters
Select Age Range
30
60
28
77
Select Gender
Both
Observation: The waterfall chart visualizes the cumulative effect of age, cholesterol, and blood pressure.
Treemap: Disease Outcome by Gender
Treemap: Disease Outcome by Gender
Male
0
1
2
3
4
Female
0
1
2
3





5. Source Code

https://github.com/dmo-27/ADV_experiments/blob/main/app.py