COM618 Data Science – Week 3 Lab Workbook

Topic: AI in Healthcare & Exploratory Data Analysis (EDA)

1. Lab Objectives

By the end of this lab, you will be able to:
- Load and inspect a healthcare dataset
- Clean missing and duplicate data
- Perform exploratory data analysis using Python
- Visualise healthcare data using charts
- Interpret insights for healthcare decision-making

2. Dataset Description

You will use a Heart Disease dataset containing patient information such as:
- Age, Sex, Blood Pressure, Cholesterol
- ECG results and heart rate
- Target variable: Presence of heart disease

3. Task 1 – Load and Inspect Data

Python Code:

```
import pandas as pd
df = pd.read_csv("messy_heart_disease.csv")
df.head()
df.info()
```

4. Task 2 – Check and Handle Missing Values

```
df.isnull().sum()
df.fillna(df.median(numeric_only=True), inplace=True)
```

5. Task 3 – Remove Duplicate Records

```
df = df.drop_duplicates()
```

6. Task 4 – Standardise Numerical Features

```
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
num_cols = df.select_dtypes(include=['int64','float64']).columns
df[num_cols] = scaler.fit_transform(df[num_cols])
```

7. Task 5 – Exploratory Data Analysis

Histogram of Age:

```
import matplotlib.pyplot as plt
plt.hist(df['age'], bins=20)
plt.title("Age Distribution")
plt.show()
```

Bar Chart of Heart Disease Outcome:

```
df['target'].value_counts().plot(kind='bar')
plt.title("Heart Disease Outcome")
plt.show()
```

8. Interpretation Questions

1. What patterns do you observe in patient age distribution?
2. Is heart disease more common in certain groups?
3. How did cleaning the data change your results?
4. Why is standardisation important in healthcare analytics?

9. Extension Task (Advanced)

- Perform correlation analysis using df.corr()
- Visualise correlations using a heatmap
- Identify the strongest risk factors for heart disease

10. Submission Instructions

Submit:

- Screenshots of your charts
- Cleaned dataset file

- Short report (300 words) discussing findings