Assessment 1

Health Care Industry

Data :
https://github.com/Priya-tops/Tops-Assessment-Data/blob/main/Data%20Science/stock_market_dataset.csv

Blood pressure consists of:
Systolic (ap_hi) – pressure when the heart beats (should be higher than diastolic)
Diastolic (ap_lo) – pressure when the heart rests

Explore the Data by Answering the Following Questions and Summarize Your Insights
Use Python libraries such as Pandas, NumPy, Matplotlib, and Seaborn to analyze the data.
After answering each question, write a brief summary of your findings or observations.

1. What is the shape of the dataset?
2. Are there any missing values in the dataset?
3. What are the unique values for categorical features like gender, cholesterol, gluc?
4. What's the average age of patients (in years)?
5. What is the distribution of the target variable (cardio)?
6. What is the average age (in years) of patients?
7. What is the distribution of BMI? (Create new BMI feature)
8. Are there outliers in height or weight?
9. Are there implausible blood pressure values (e.g., ap_hi < ap_lo or too high)?
10. How many rows have incorrect blood pressure values?

Column	Condition	Reason
ap_hi	between 80 and 250	realistic systolic
ap_lo	between 50 and 200	realistic diastolic
ap_hi >= ap_lo	systolic must be ≥ diastolic	logical

11. Remove rows with invalid blood pressure, height, or weight?
     Height between 120cm to 220cm
     Weight between 40 to 200kg

12. What is the distribution of cholesterol and glucose levels after cleaning?
13. How many smokers have cardiovascular disease?
14. Does alcohol intake correlate with higher cardio risk?
15. What's the correlation between features?
16. Compare mean BMI for cardio vs. non-cardio
17. Plot age distribution for those with and without cardio disease
18. Boxplot of systolic blood pressure by cardio status

19. What is the distribution of cholesterol levels?
20. What percentage of patients have above-normal glucose levels?

Insightful Analysis Questions
These questions aim to uncover key relationships and gain deeper knowledge about cardiovascular disease. To answer them, you'll need to analyze the data and write down any patterns, trends, or observations you notice during the process.

1. What percentage of the dataset has cardiovascular disease?
2. Is there a link between cholesterol and heart disease?
3. Does age impact heart disease prevalence?
4. Is BMI higher in those with heart disease?
5. Does physical activity reduce heart disease risk?
6. Do smokers have more heart disease?
7. Is systolic pressure significantly higher in those with heart disease?