PREDICT & CLASSIFY HEART ATTACK

98% ACCURACY (NO OVERFITTING)

DATA ANALYSIS

ANALYSIS PURPOSE

This dataset includes eight features that can be used to estimate the likelihood of a heart attack. In South Korea, my home country, heart attack-related deaths have been on the rise, while the average age of affected individuals has been decreasing. Therefore, identifying key risk factors and predicting heart attack risk may offer valuable insights for early intervention and prevention.

WHAT ARE KEY PREDICTORS?

RESEARCH QUESTIONS

- Q1. What are the key medical predictors of heart attack?
- Q2. Which medical features influence these key predictors?
- Q3. How does age affect the likelihood of having a heart attack?
- Q4. How does age influence the key medical predictors?
- Q5. How does gender affect the likelihood of having a heart attack?
- Q6. How does gender influence the key medical predictors?
- Q7. Can a predictive model be built to forecast the likelihood of having a heart attack based on the key medical predictors?
- Q8. Can patients be grouped based on medical features to identify the highest-risk group?

DATASET

#	Column	Non-Null Count	Dtype
0	age	1319 non-null	int64
1	gender	1319 non-null	int64
2	impulse	1319 non-null	int64
3	pressurehigh	1319 non-null	int64
4	pressurelow	1319 non-null	int64
5	glucose	1319 non-null	float64
6	kcm	1319 non-null	float64
7	troponin	1319 non-null	float64
8	heart attack	1319 non-null	object

The dataset has nine features including the target (heart attack). The size of the dataset is 1319.

Dummy columns of age groups were added for analysis.

ANSWER QUESTION 1

Q1. What are the key medical predictors of heart attack?

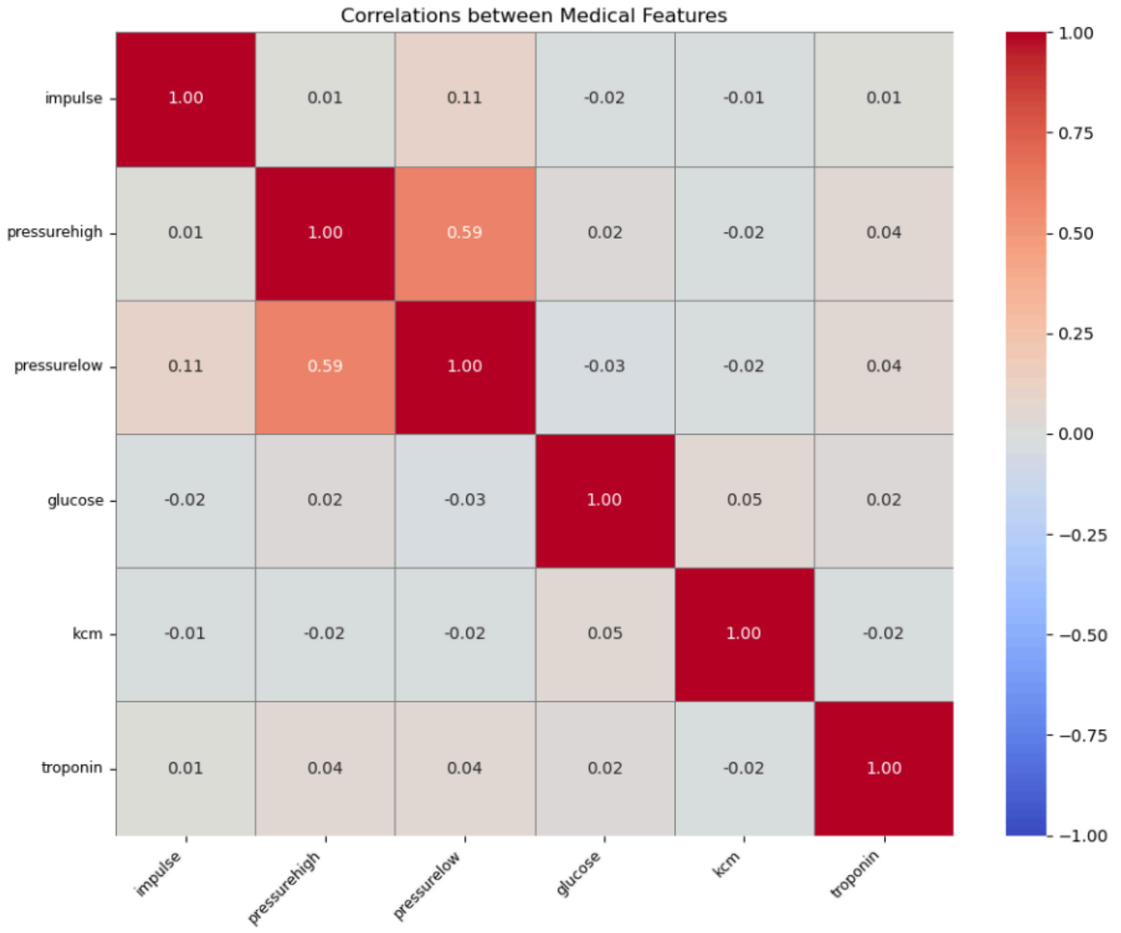
The key predictors of heart attack are kcm and troponin.

```
--------------------------------------------------
kcm vs heart attack
  U-statistic: 280469.5000
  p-value: 0.0000
  ✅ Statistically significant difference in the feature between heart attack and non heart attack classes.
--------------------------------------------------
troponin vs heart attack
  U-statistic: 368361.5000
  p-value: 0.0000
  ✅ Statistically significant difference in the feature between heart attack and non heart attack classes.
--------------------------------------------------
```

ANSWER QUESTION 2

Q2. Which medical features influence these key predictors?

Other medical features have weak statistical associations with the key predictors.



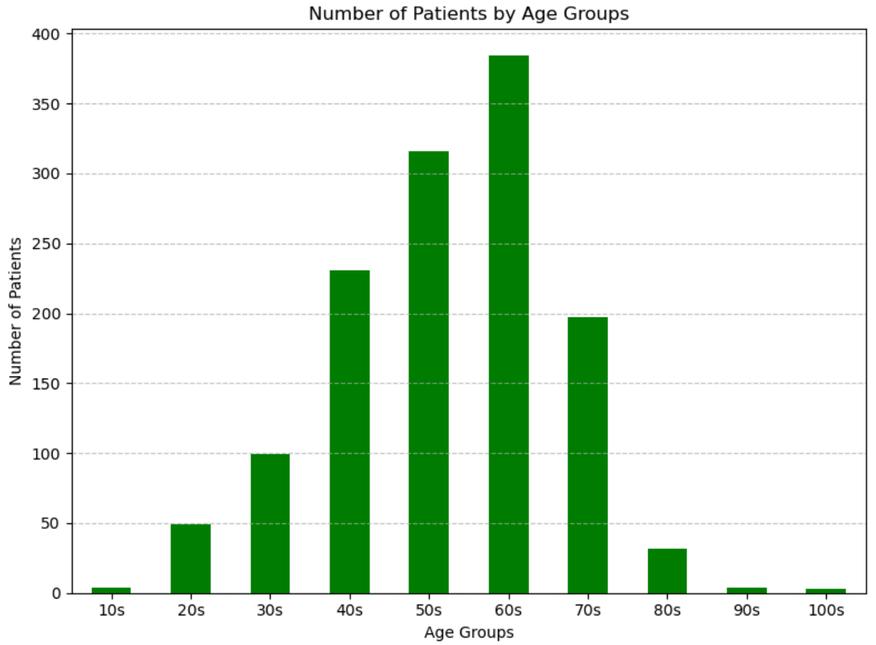
ANSWER QUESTION 3

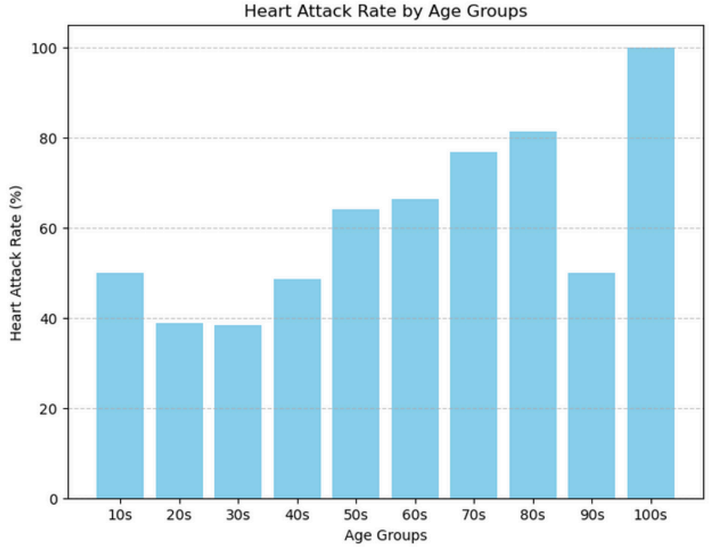
Q3. How does age affect the likelihood of having a heart attack?

The heart attack rate tends to increase with age. In the 100s age group, it reaches 100%, possibly due to the small sample size. While the rate is relatively low in younger age groups, it drastically increases from the 50s onward, indicating higher vulnerability among older populations. However, the result should be interpreted with caution.

```
Chi-squared statistic: 80.8613
P-value: 0.0000

⚠️ Some expected frequencies are less than 5. Chi-square test assumptions may be violated.
✅ There is a statistically significant relationship
```





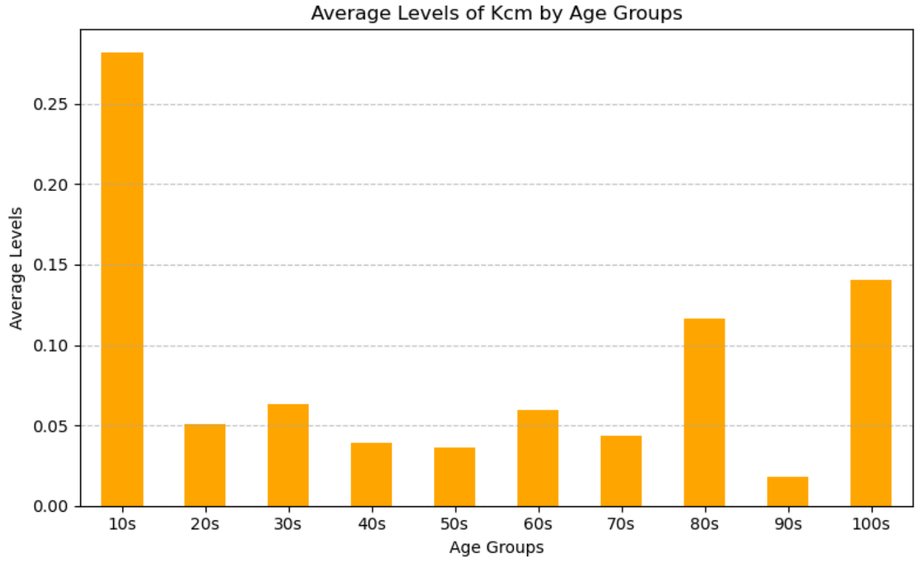
ANSWER QUESTION 4

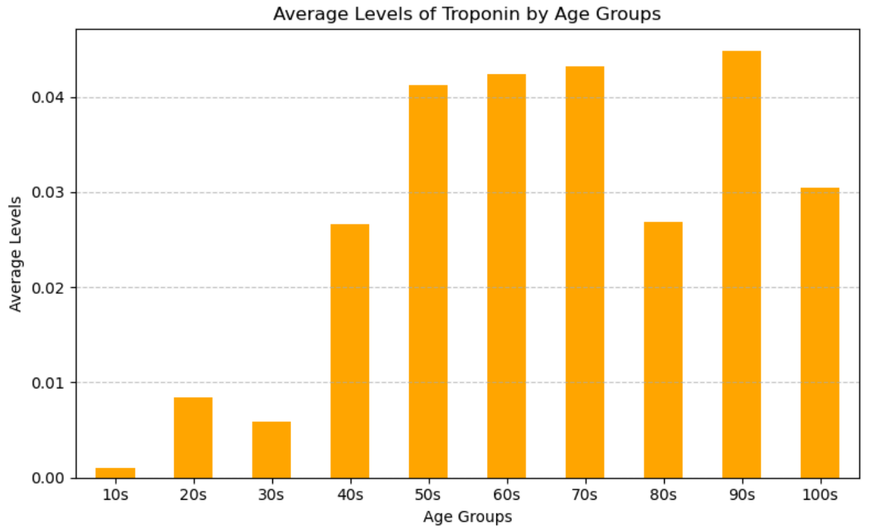
Q4. How does age influence the key medical predictors?

The distributions of both kcm and troponin levels differed significantly among age groups ($p < 0.05$). These results indicate a potential association between age and the levels of these key medical predictors.

```
Kruskal-Wallis Test for 'kcm' across age groups
--------------------------------------------------
p-value: 0.0347
✅ The differences are statistically significant (p < 0.05)

Kruskal-Wallis Test for 'troponin' across age groups
--------------------------------------------------
p-value: 0.0000
✅ The differences are statistically significant (p < 0.05)
```





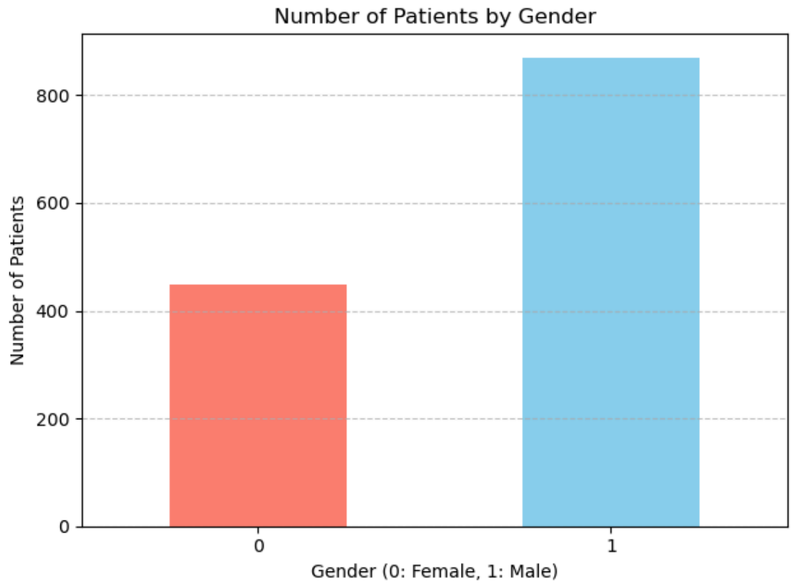
ANSWER QUESTION 5

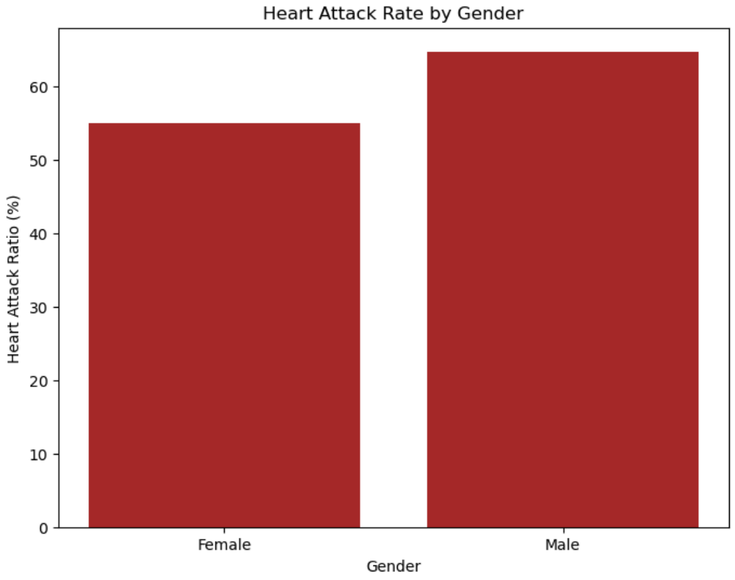
Q5. How does gender affect the likelihood of having a heart attack?

There is a statistically significant association between gender and heart attack occurrence (p-value < 0.05), indicating that gender affects the likeihood of having a heart attack.

```
Chi-squared statistic: 11.3563
p-value: 0.0008

✅ All expected frequencies are greater than or equal to 5. Chi-square test assumptions are satisfied.
✅ There is a statistically significant relationship
```





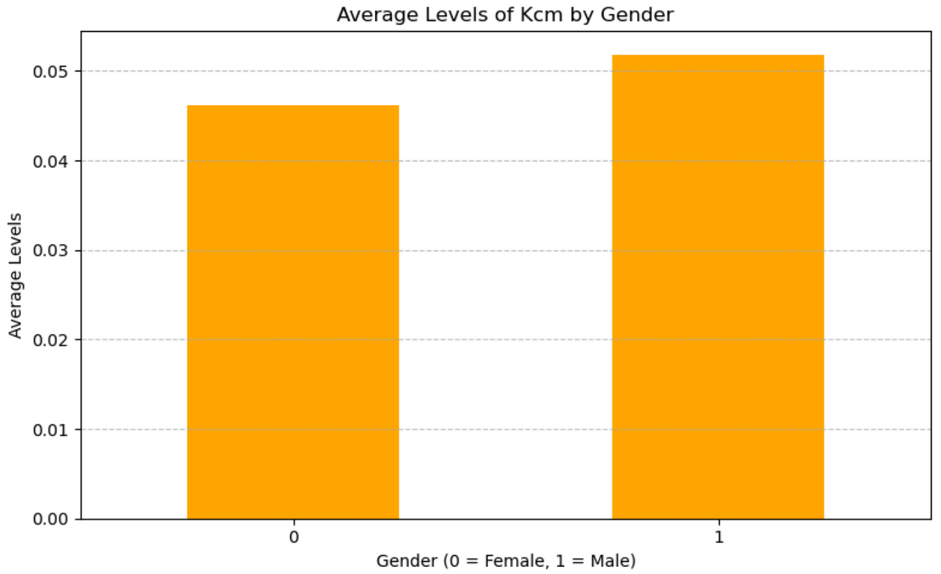
ANSWER QUESTION 6

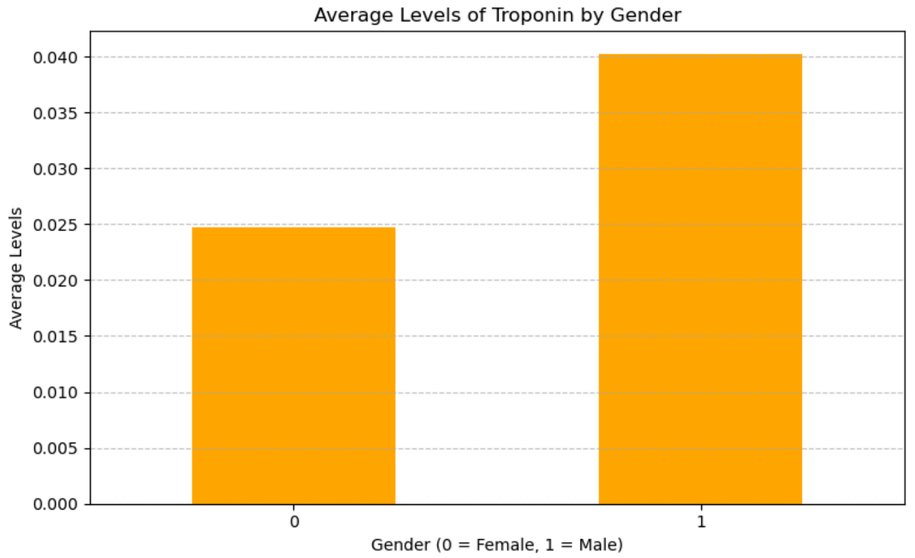
Q6. How does gender influence the key medical predictors?

Only troponin levels showed statistically significant differences between males and females ($p < 0.05$), indicating a potential association between gender and troponin. In contrast, the differences in kcm levels were not statistically significant ($p \geq 0.05$).

```
Mann-Whitney U Test for 'kcm' across gender
--------------------------------------------------
p-value: 0.3169
❌ The differences are not statistically significant (p ≥ 0.05)

Mann-Whitney U Test for 'troponin' across gender
--------------------------------------------------
p-value: 0.0000
✅ The differences are statistically significant (p < 0.05)
```



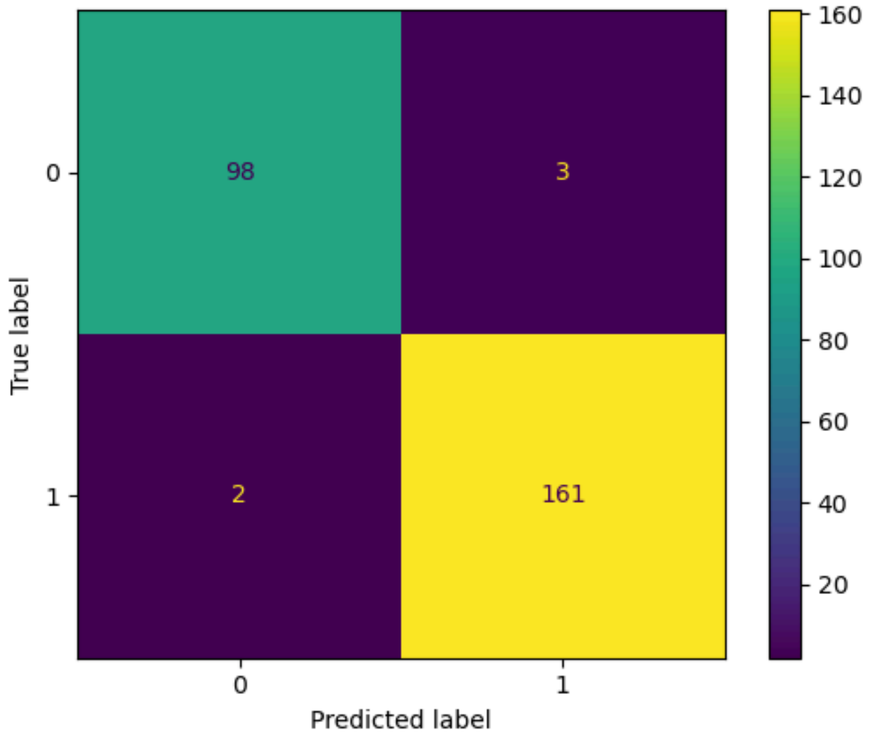


ANSWER QUESTION 7

DECISION TREE

Q7. Can a predictive model be built to forecast the likelihood of having a heart attack based on the key medical predictors?

A decision tree model was delveloped with four features (age, gender, kcm, and troponin) to classify or predict heart attack. The overall accuracy of the decision tree model is 98%. There are no signs of overfitting.



Classification Report:

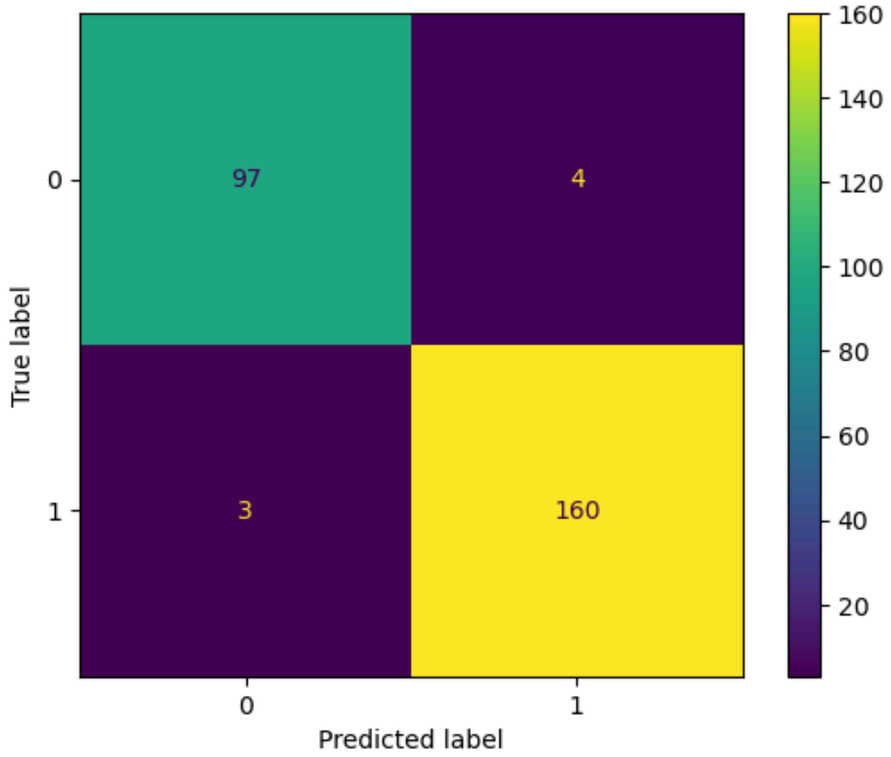
	precision	recall	f1-score	support
0	0.98	0.97	0.98	101
1	0.98	0.99	0.98	163
accuracy			0.98	264
macro avg	0.98	0.98	0.98	264
weighted avg	0.98	0.98	0.98	264

ANSWER QUESTION 7

RANDOM FOREST

Q7. Can a predictive model be built to forecast the likelihood of having a heart attack based on the key medical predictors?

A random forest model was delveloped with four features (age, gender, kcm, and troponin) to classify or predict heart attack. The overall accuracy of the decision tree model is 97%. There are no signs of overfitting.



Classification Report:

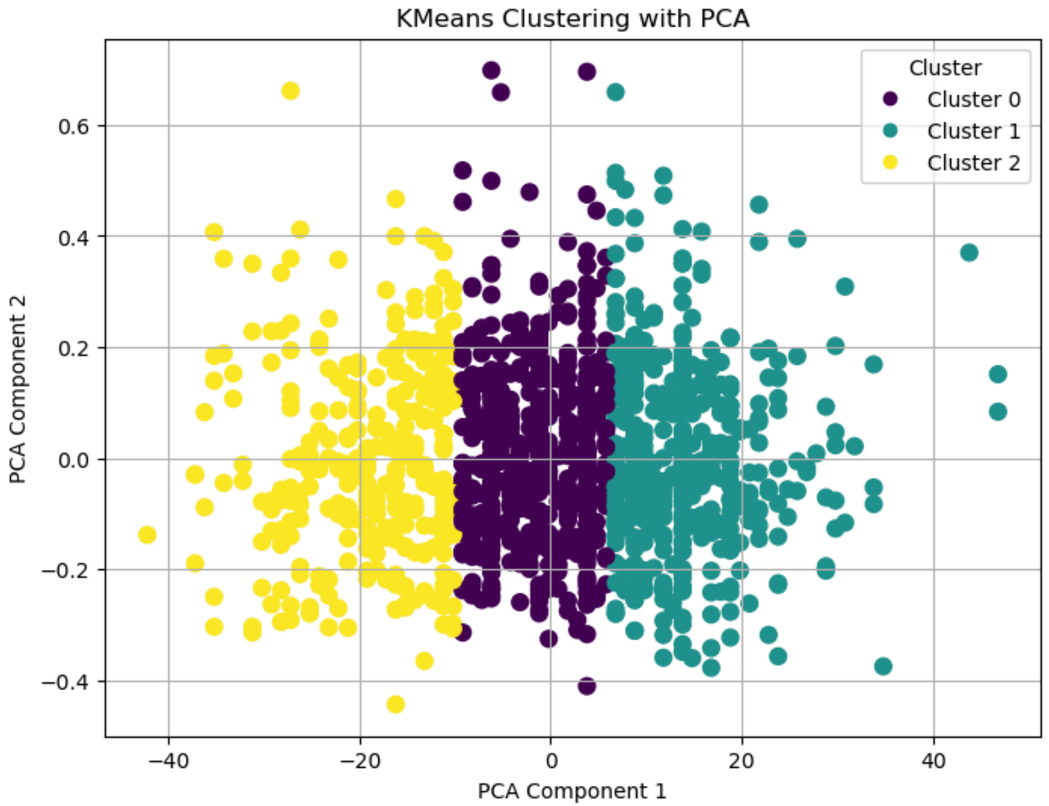
	precision	recall	f1-score	support
0	0.97	0.96	0.97	101
1	0.98	0.98	0.98	163
accuracy			0.97	264
macro avg	0.97	0.97	0.97	264
weighted avg	0.97	0.97	0.97	264

ANSWER QUESTION 8

Q8. Can patients be grouped based on medical features to identify the highest-risk group?

heart attack	0	1
cluster		
0	0.358595	0.641405
1	0.279318	0.720682
2	0.595469	0.404531

Patients in clusters 0 and 1 need focused care and monitoring to prevent or manage heart attack. The medical feature levels observed in cluster 2 can serve as a reference for identifying and managing high-risk patients.



*The Silhouette score of 0.518 indicates that the clusters are reasonably distinct, with some overlap. indicates that the clusters are reasonably distinct, with some overlap.