

Analisi delle malattie cardiache

Progetto di Data Mining

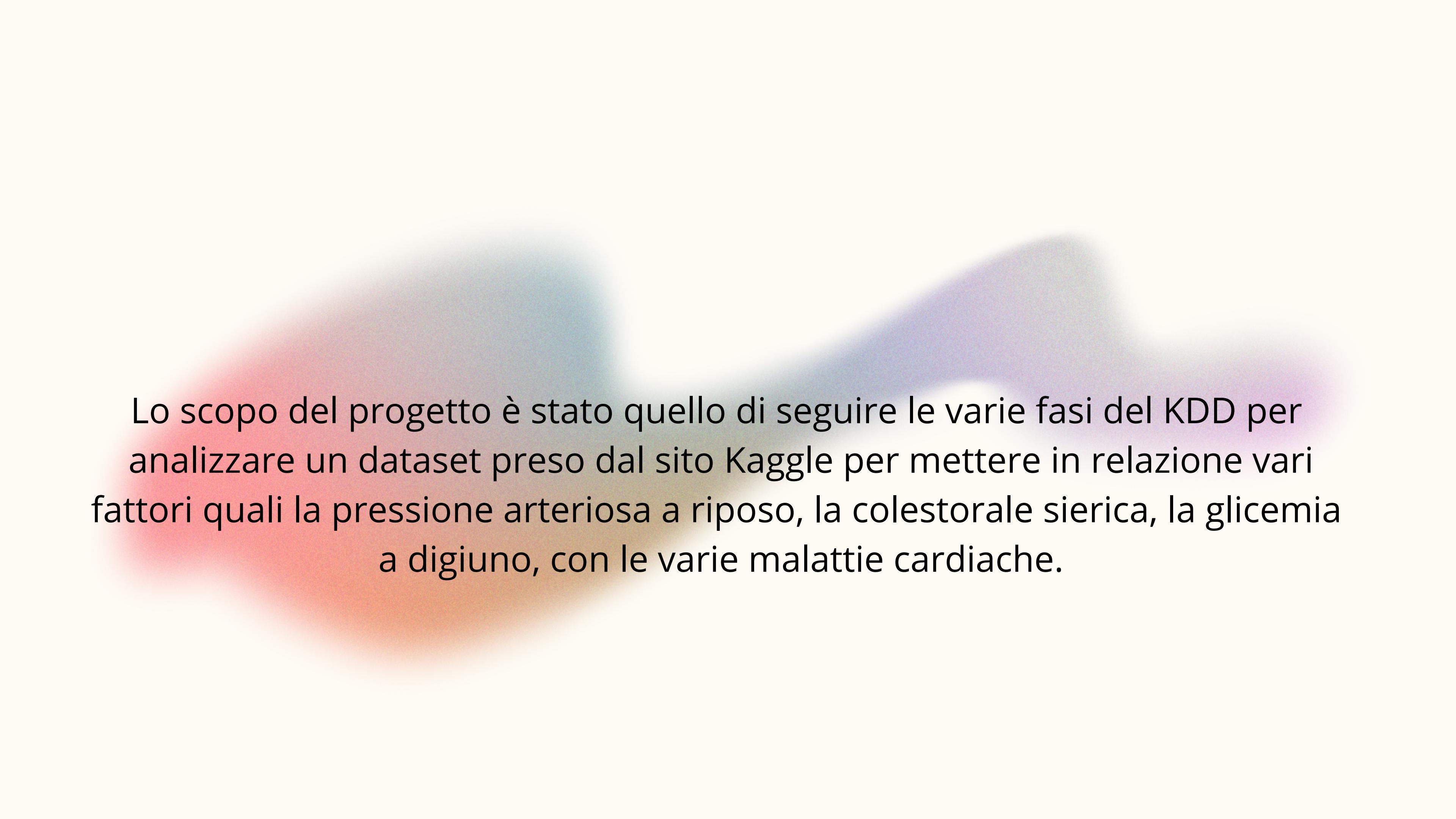
Federico Calò

Matricola 678191



Introduzione

Utilizzo della metodologia KDD per l'esplorazione
e l'analisi di un dataset relativo alle malattie
cardiache.



Lo scopo del progetto è stato quello di seguire le varie fasi del KDD per analizzare un dataset preso dal sito Kaggle per mettere in relazione vari fattori quali la pressione arteriosa a riposo, la colesterolemia sierica, la glicemia a digiuno, con le varie malattie cardiache.

Metodologia

Visualizzazione e cleaning del dataset

In una prima fase si è cercato di capire i dati presenti nel dataset ed eventualmente eliminare le anomalie.

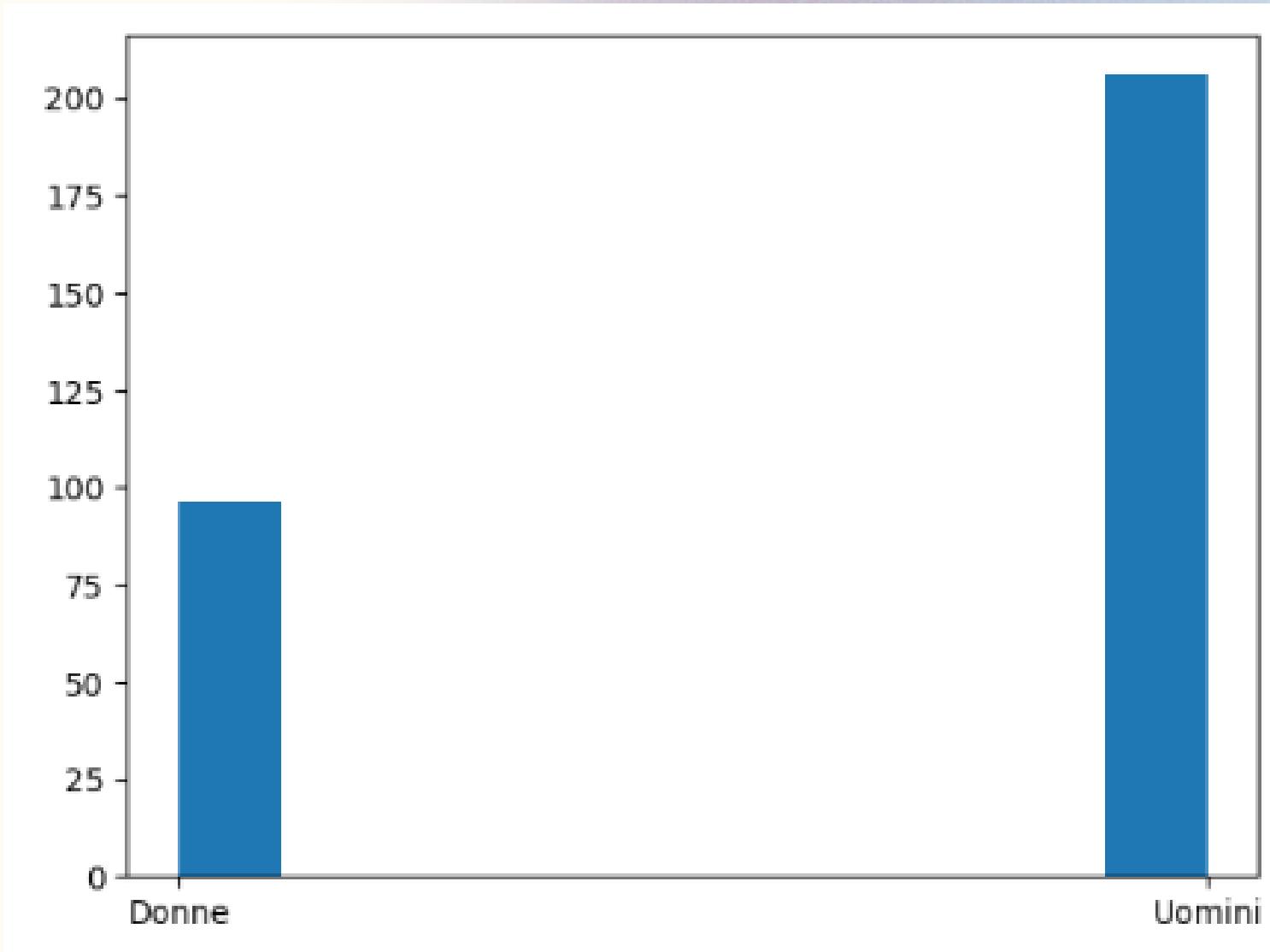
Matrice di correlazione

Successivamente si è cercato di comprendere le correlazioni presenti tra le varie colonne del dataset.

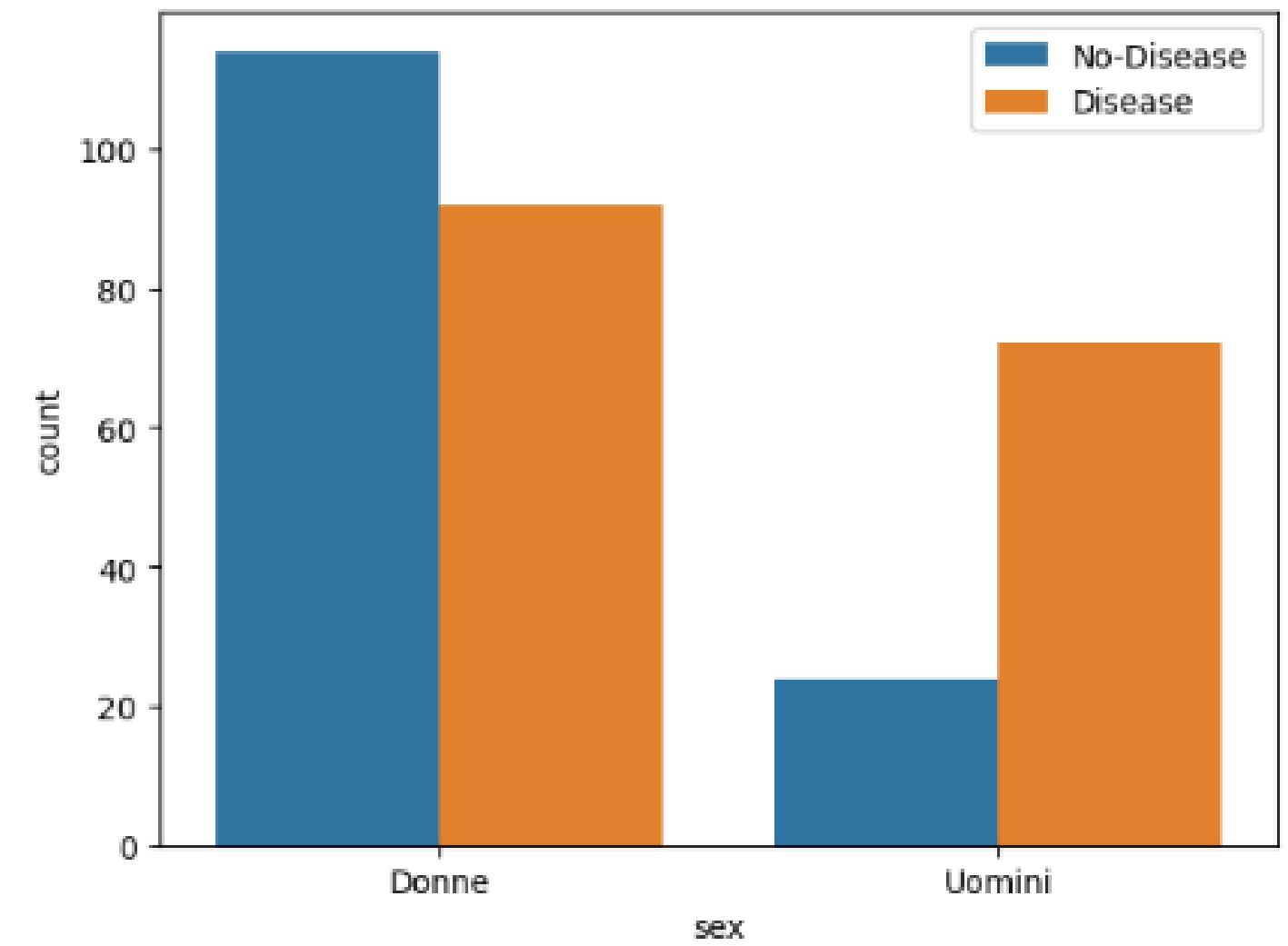
Manipolazione delle informazioni e visualizzazione risultati

Attraverso la manipolazione delle informazioni e la loro rappresentazione tramite opportuni grafici, sono state condotte alcune conclusioni

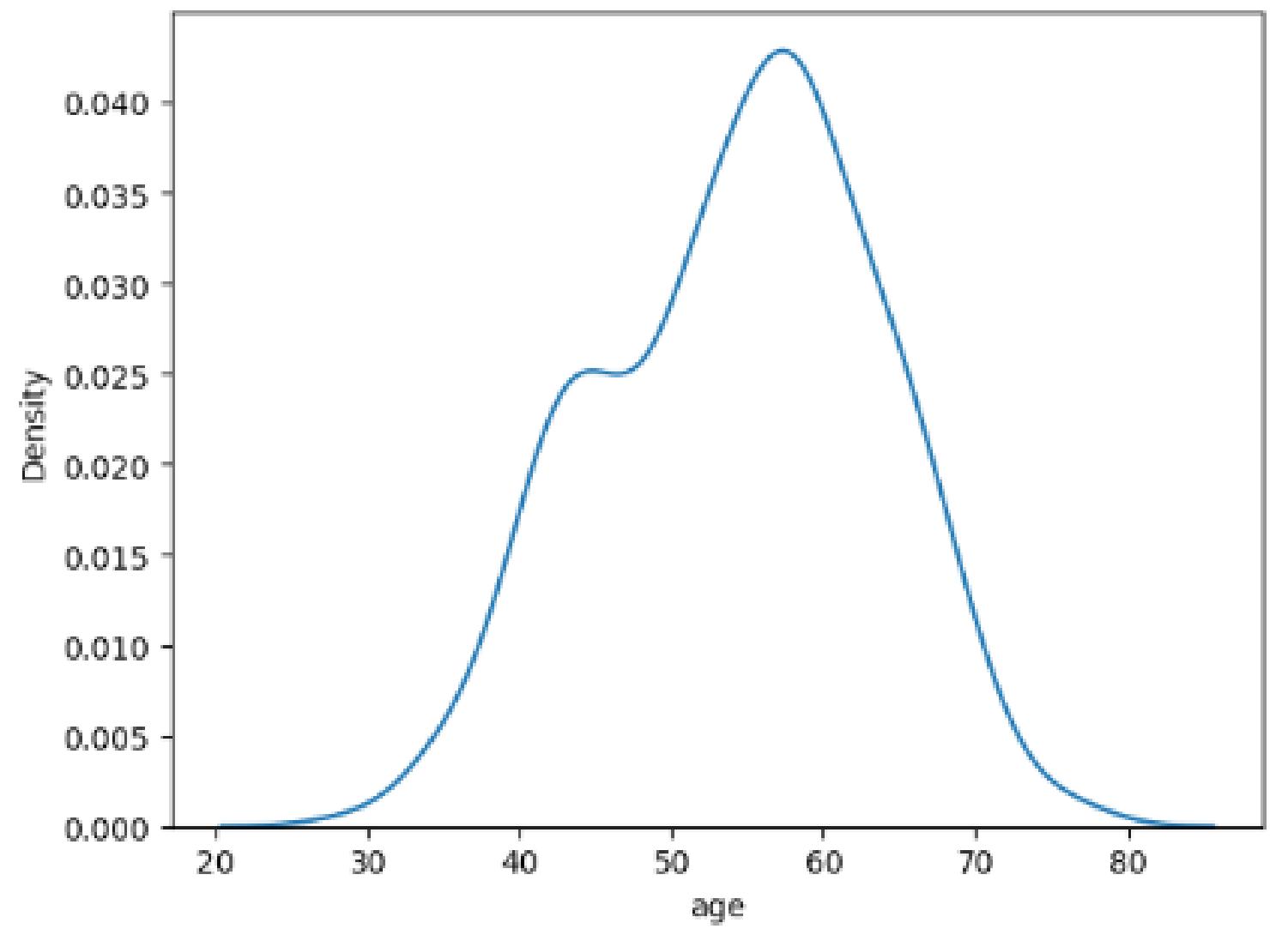
Informazioni sul dataset



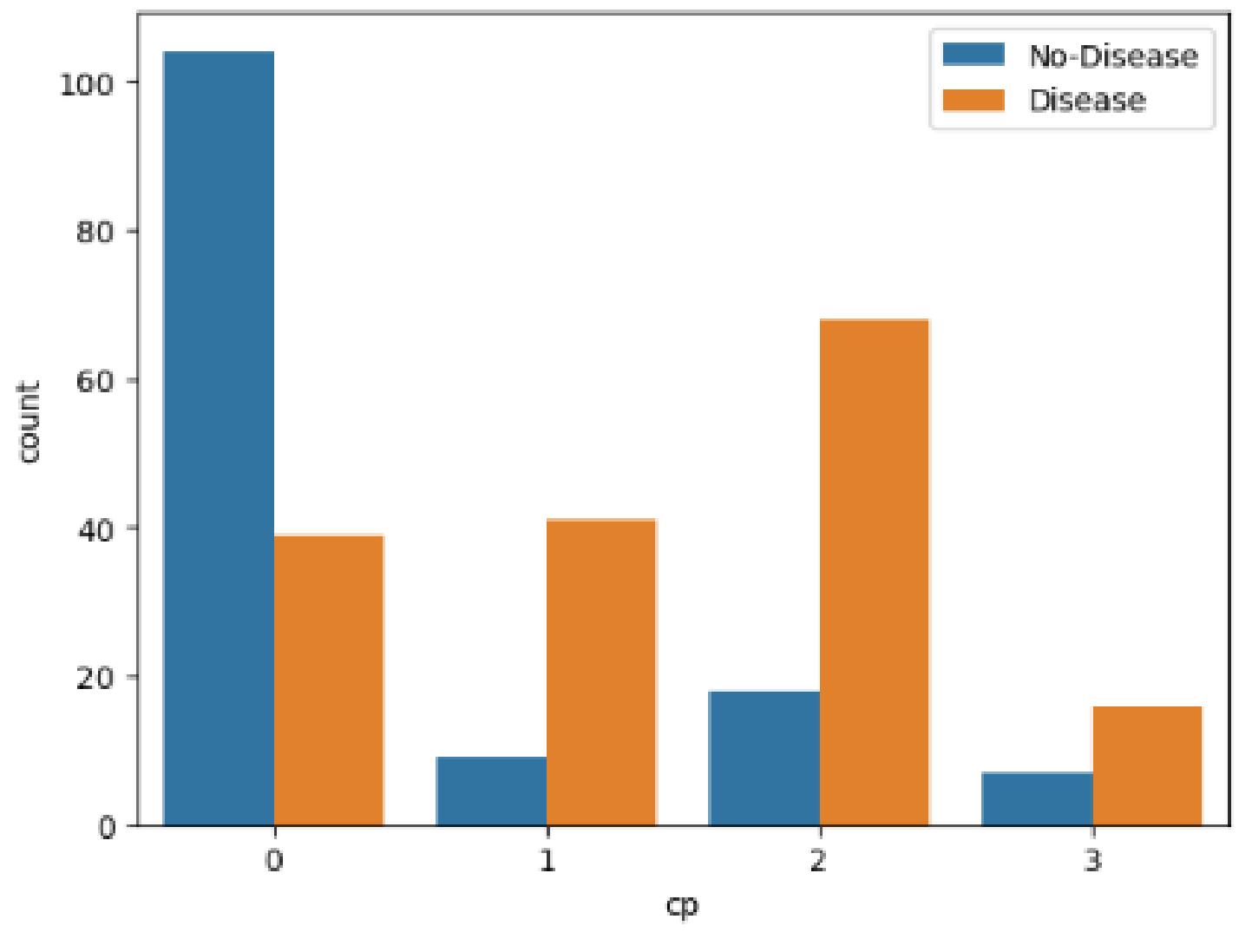
Su un campione di 302 record, 206 sono uomini e 96 sono donne



Mettendo in correlazione il numero di persone che hanno riscontrato una malattia cardiaca, si è riscontrato che negli uomini è stato diagnosticata una malattia con una frequenza nettamente maggiore rispetto al genere femminile.



In relazione all'età in cui è stata diagnosticata una problematica cardiaca, notiamo come i soggetti compresi tra i 50 e i 60 anni hanno un maggior rischi di contrarne una. Mentre tra i 40 e i 50 la probabilità è minore, ma comunque alta considerando i quantili del grafico.



0 - angina tipica

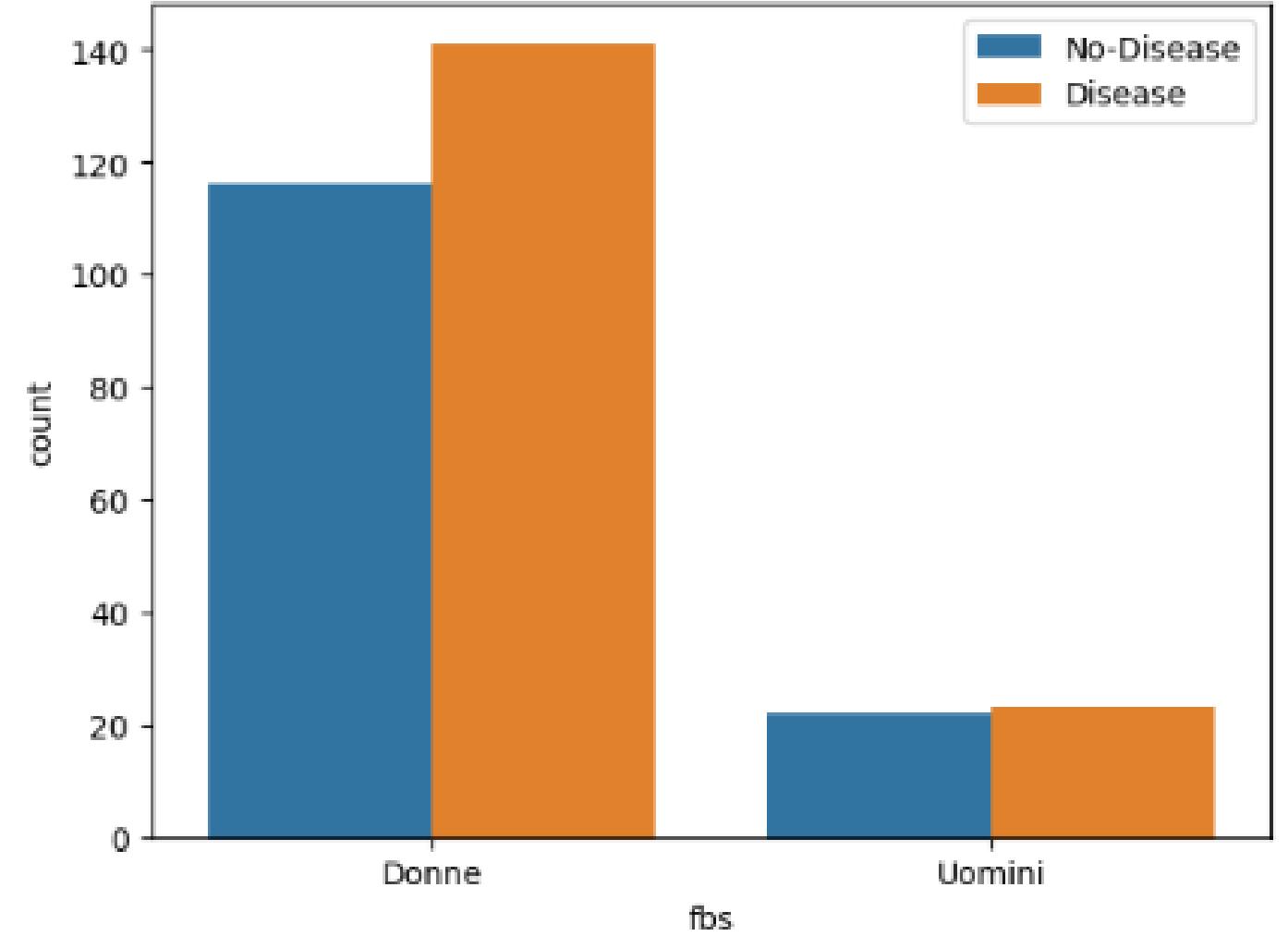
1- angina atipica

2 - nessun dolore di angina

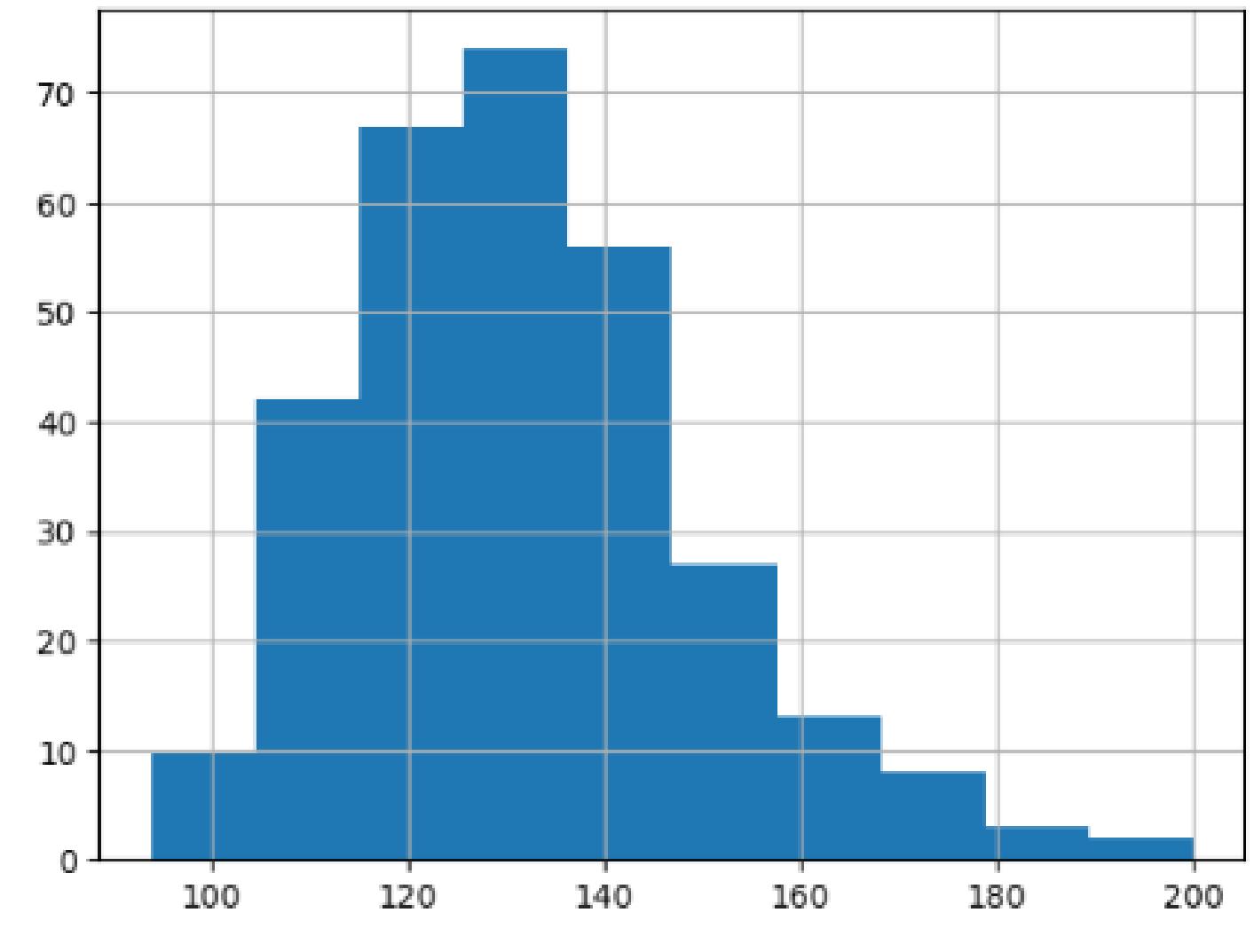
3 - asintomatico

Notiamo come la maggior parte dei casi diagnosticati, non avesse nessun tipo di dolore dovuto all'angina.

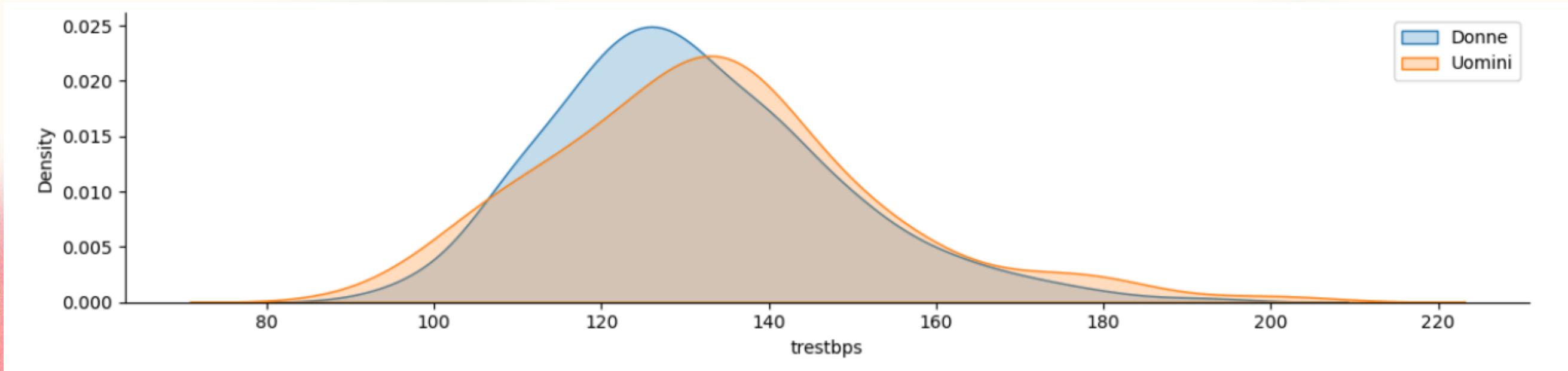
Inoltre notiamo che dolori tipici di angina non sono sempre legati a delle malattie cardiache.



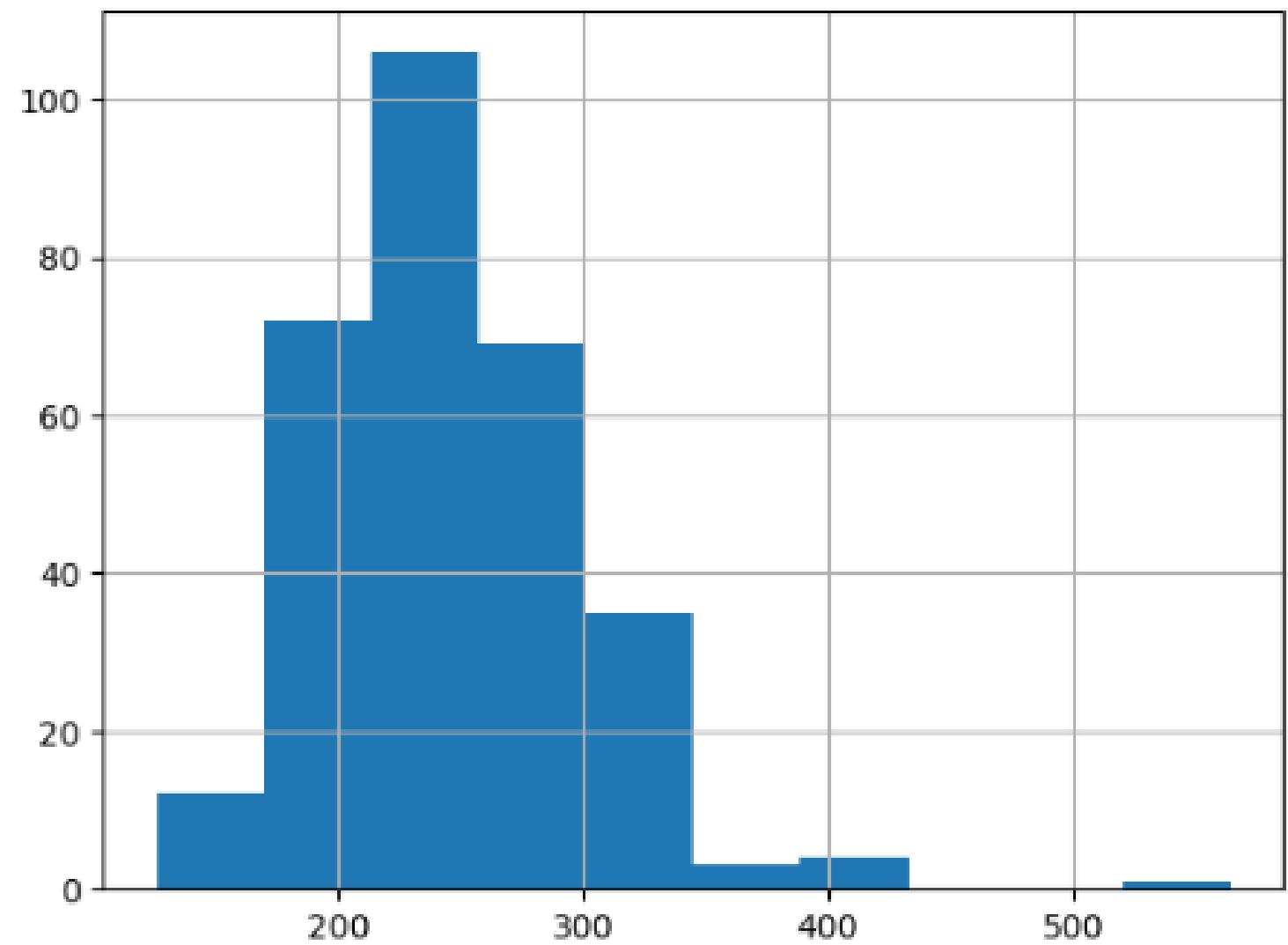
Considerando il valore della glicemia a digiuno in relazione al sesso, notiamo come questo è un forte sintomo di problematiche legate al cuore sia negli uomini, sia nelle donne.



Prendendo in considerazione la pressione arteriosa a riposo, si può evincere come i suoi valori che si aggirano tra i 120 mm Hg e i 130 mm Hg, siano degli indicatori di probabili attacchi cardiaci.



Nel grafico possiamo osservare che la pressione arteriosa per le donne si attesta intorno ai 130 mm Hg, mentre per gli uomini si aggira intorno ai 140 mm Hg.



In questo grafico vediamo come il colesterolo sierico influisce sulle malattie cardiache, soprattutto per un valore intorno ai 250 mg/dl.