1. This classic dataset “Diamonds”contains the prices and other attributes of almost 54,000 diamonds. It's a great dataset for beginners learning to work with data analysis and visualization.

price price in US dollars (\$326--\$18,823)

carat weight of the diamond (0.2--5.01)

cut quality of the cut (Fair, Good, Very Good, Premium, Ideal)

color diamond colour, from J (worst) to D (best)

clarity a measurement of how clear the diamond is (I1 (worst), SI2, SI1, VS2, VS1, VVS2, VVS1, IF (best))

x length in mm (0--10.74)

y width in mm (0--58.9)

z depth in mm (0--31.8)

depth total depth percentage = z / mean(x, y) = 2 \* z / (x + y) (43--79)

table width of top of diamond relative to widest point (43--95)

1. What is the probability that a random diamond you selected is IDEAL cut?
2. What is the probability that a random diamond you selected has VS1 type clarity?
3. What is the probability that a random diamond you pick up is PREMIUM and color E?
4. What is the probability that a random diamond you selected is IDEAL or has VS1 type clarity?
5. This dataset “heart.csv” contains 76 attributes, but all published experiments refer to using a subset of 14 of them. In particular, the Cleveland database is the only one that has been used by ML researchers to this date. The "goal" field refers to the presence of heart disease in the patient. It is integer valued from 0 (no presence) to 4.  
     
   Attribute Information:  
   1. age  
   2. sex  
   3. chest pain type (4 values)  
   4. resting blood pressure  
   5. serum cholestoral in mg/dl  
   6. fasting blood sugar > 120 mg/dl  
   7. resting electrocardiographic results (values 0,1,2)  
   8. maximum heart rate achieved  
   9. exercise induced angina  
   10. oldpeak = ST depression induced by exercise relative to rest  
   11. the slope of the peak exercise ST segment  
   12. number of major vessels (0-3) colored by flourosopy  
   13. thal: 3 = normal; 6 = fixed defect; 7 = reversable defect

The names and social security numbers of the patients were recently removed from the database,   
replaced with dummy values. One file has been "processed", that one containing the Cleveland   
database. All four unprocessed files also exist in this directory.  
To see Test Costs (donated by Peter Turney), please see the folder "Costs"  
Acknowledgements  
Creators:  
1. Hungarian Institute of Cardiology. Budapest: Andras Janosi, M.D.  
2. University Hospital, Zurich, Switzerland: William Steinbrunn, M.D.  
3. University Hospital, Basel, Switzerland: Matthias Pfisterer, M.D.  
4. V.A. Medical Center, Long Beach and Cleveland Clinic Foundation: Robert Detrano, M.D., Ph.D.  
Donor: David W. Aha (aha '@' ics.uci.edu) (714) 856-8779  
Inspiration Experiments with the Cleveland database have concentrated on simply attempting to distinguish   
presence (values 1,2,3,4) from absence (value 0).  
See if you can find any other trends in heart data to predict certain cardiovascular events or find any clear indications of heart health.

i. Find the probability of a patient having fasting sugar level > 120 mg/l given that the   
patient is a male.  
ii. Find the probability of a patient having fasting sugar level > 120 mg/l given that the   
patient is a female.  
iii. Find the probability of a patient having serum level 200 – 239 (moderately elevated)   
given that the patient is male?  
iv. Find the probability of a patient having serum level 200 – 239 (moderately elevated)   
given that the patient is female?

1. Use the Naïve Bayes Classifier model to predict a customer who is **self-employed, has average credit history, low income and No collateral**. Would the bank approve a loan from his/her customer?

The datasets named loan\_approval.csv has following variables:

Employment Status: Salaried, Self-Employed, Unemployed

Credit History: Good, Average, Poor

Income Level: Low, Medium, High

Collateral: Yes, No

Target variable

Loan Approved: Yes, No

Solve the problem step-by-step and explain as you go. You can follow the problem that we went over the class as reference. Link to the example: <https://www.geeksforgeeks.org/machine-learning/naive-bayes-classifiers/>