Due date: Friday April 10, by 2pm
You must submit your assignment electronically and as a single file via the LMS page for this subject. Your solutions must include your workings. **Note that this assignment is worth 10% of your overall mark for this subject.** Advice for preparing an electronic version of your assignment is provided on **this LMS page (click here)** under the heading: 'Some ideas for submitting your assignments online'.

In submitting your work, you are consenting that it may be copied and transmitted by the University for the detection of plagiarism. Please start with the following statement of originality, which must be included near the top of your submitted assignment:

"This is my own work. I have not copied any of it from anyone else."

1. Police officers often set up sobriety checkpoints along busy roads. At these checkpoints, drivers are briefly questioned by officers, in order to ascertain whether or not an individual is driving while under the influence of alcohol. If the officer does not suspect a problem, drivers are released to go on their way. Otherwise, drivers are detained for the Breathalyzer test, that will determine whether or not the driver will be fined and/or arrested.

   Suppose that, based on police records, a trained officer makes the right decision 75% of the time, and will subsequently detain or not detain the driver based on their assessment (i.e. the officer will correctly detain someone driving under the influence 75% of the time, and correctly release a driver who is not intoxicated 75% of the time). For the following questions, consider a scenario in which police operate a sobriety checkpoint after 10 pm on a Friday night, a time at which national traffic safety experts suspect roughly 14% of drivers are driving under the influence of alcohol.

   (a) If you are stopped at the checkpoint and, being a law-abiding citizen, are not intoxicated, then what is the probability that you are detained for further testing?
   (b) What is the probability that any given driver will be detained?
   (c) What is the probability that a driver who is detained has actually been drinking?
   (d) What is the probability that a driver who was not detained had actually been drinking?

2. A travel company offers different bundles for flight, cruise, tour and insurance services. Each of the bundles has one of the labels (F, C, T, I, FF, CI). Consider the following set of six transactions at a selected store:

ID	Item Set
1	F, C, I
2	C, T, I, FF
3	F, I, FF, CI
4	F, C, T, CI
5	F, C, I, CI
6	F, T, I, FF

   When a customer buys a product, the travel company would like to be able to recommend another product that the customer would be most likely to buy as well. Suppose that the store manager therefore conducts market basket analysis using the above data and, in order to help the store cross-sell products and increase sales, is now considering selecting one of the following two association rules: $Rule_1 : F \Rightarrow I$ and $Rule_2 : T \Rightarrow I$.

   (a) What is the support and confidence of $Rule_1 : F \Rightarrow I$?
   (b) What is the support and confidence of $Rule_2 : T \Rightarrow I$?
   (c) On the basis of the given data, which of the above two association rules would be a better predictor of cross-sales? Explain your reasoning, referring to the lift and support values for the two association rules.

3. This question should be carried out in R. You must include your R code in your answers.
   Complete the following tasks, using the ***Groceries*** data set from the *arules* package:

   (a) Produce a figure to represent support for the first 9 items. *Hint: Use the command ItemFrequencyPlots.*
   (b) Produce a figure to represent the items with minimum support of 0.1.
   (c) Produce a grouped bubble matrix chart for items with minimum Support and Confidence 0.025. Comment on the association rules displayed in your chart that have, as their *antecedent*, {*tropical fruit, frankfurter*}.

4. Many recent statistical analyses of internet traffic suggest that users send messages whose size (measured in bytes) is a random variable T with finite mean and infinite variance, (i.e. $\mathrm{Var}(T) = \infty$). This infinite variance should not be a surprise when one considers the huge sizes of music and video files, as well as security updates for computer operating systems, that are transmitted across the internet.

   One widely used mathematical model for random variables with finite mean and infinite variance is the *Pareto distribution*, which may be defined as

$$F(T) = \begin{cases} 1 - \dfrac{1}{(1+t)^\alpha} & \text{where } (1 < \alpha < 2), t \geq 0 \\ 0 & \text{where } t < 0 \end{cases}.$$

   (a) Calculate $f(t)$, the probability density function of $F(T)$.

   (b) Considering the mean of T, $E(T) = \frac{1}{\alpha - 1}$, What unit of measurement should be used when discussing $E(T)$?

   (c) Using $f(t)$, compute the probability that a message will have size less than or equal to $E(T)$, when $\alpha = 1.5$.

5. This question should be carried out in R. You must include your R code in your answers.
   Hint: You can use the R function `integrate` *for many of the following questions.*

   Let X denote the October demand (measured in 1000's of units), for a particular product. Historical demand figures suggest that a suitable probability density function for X is

$$f(x) = \begin{cases} \dfrac{3}{\Gamma(2/3)} x \cdot e^{-x^3}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}.$$

   (a) Using R, create a plot of the probability density function $f(x)$. Your plot should have adequate labels and an appropriately chosen scale for the horizontal axis.

   (b) Using the R function `integrate`, show that $\int_0^\infty f(x)dx = 1$. In doing so, and by noting that $f(x) \geq 0$ for all x, you will therefore have shown that f is a valid probability density function.

   (c) Compute the mean of X:

      i. by using the R function `integrate`;

      ii. by hand

      Hint: You should obtain the same answer for both (i) and (ii).

   (d) Suppose that the company manufacturing the product in question will order 1200 units for the next October period. What is the probability that demand will exceed supply? In other words, what is $P(X > 1.2)$?

   (e) Unless the company sells at least 62 units of the item in question each October, it may be forced to file for bankruptcy. Compute the probability that the company exceeds this minimum amount of 62 units sold, while also having enough stock on hand to meet demand, having order 1200 units.

6. A clothing store sells three types of suits, priced at \$500, \$750 and \$1000 respectively, and two types of pairs of shoes, priced at \$250 and \$500 respectively. Let X denote suit price and let Y denote shoe price (for a pair). The joint probability mass function for (X, Y) is

$p(x, y)$	$y = 250$	$y = 500$
$x = 500$	0.25	0.20
$x = 750$	0.175	0.25
$x = 1000$	0.05	0.075

   (a) Calculate $P(X = 1000 \text{ and } Y = 500)$. What does this probability represent?

   (b) A customer buys a \$500 pair of shoes. What is the probability that she also bought an \$750 suit?

   (c) Calculate $F_{X,Y}(750, 250)$.

   (d) Derive the marginal probability mass function for the random variable Y.

   (e) The total cost of an outfit (suit and pair of shoes), is given by the random variable $T = X + Y$. Compute the probability function $f_T(t)$.

   (f) Compute $E(T)$.