Data Mining in R
Assignment 3

a) Short answer questions

Q.1 - What is the main difference between unsupervised and supervised tasks? **[2 points]**

Q.2 - What is the relationship between clustering and outlier detection? **[2 points]**

Q.3 - What is the difference between precision and classification accuracy? **[2 points]**

Q.4 - Imagine that someone told you that a model achieved recall of 100% on the problem addressed in this lesson. Can you say something about the quality of this model based on this information? **[2 points]**

Q.5 - What is the reason why a problem with an imbalanced class distribution creates problems to evaluation metrics like classification accuracy? **[2 points]**

Q.6 - What is the main difference between the holdout method and cross validation? **[2 points]**

b) Assignments [18 points]

i) Obtain a bar plot that shows the diversity of the product portfolio of the 5 salesmen that sell a larger total quantity. The diversity of the product portfolio of a salesman is the proportion of all products that the salesmen sold.