DWDM LAB – 21AUG2018

1. Find the car with the best mpg
2. Find the car with the worst horse power
3. Tabulate mpg for different number of gears
4. Find the 5 number summary for displacement
5. Draw a boxplot of mpg for different cylinders
6. Draw a boxplot for weights for each cyinder type
7. Compare the mpg for different transmission types (Automatic or Manual)

DWDM LAB – 28AUG2018

USING WEKA

1. Preprocessing: Use numeric to nominal filter
2. For question 2 change property "car" = 2 and set index to cook books.

USING R and TRAIN.CSV

3. Find the mean of train fare per passenger class and visualize using box plot.
4. Tabulate the age of passengers for each passenger class.
5. Is there a correlation between survival and gender?
6. Is there a correlation between survival and Passenger class?
7. Preprocess the data and find all associations rules for survived = "Yes" using the Apriori algorithm. Note down the support, confidence and lift of the rules.

****

DWDM LAB EXERCISES - 11 SEP 2018

****

1. Display the sales of each product and for each salesman as a % of the entire sales.
2. Display the sales of each product for each suppler
3. Display the sale of each product for each customer.
4. Visualize the sales of products as a pie chart
5. Visualize the sales of products category wise in a column chart. Identify the top 5 selling products and the salesman who sell them. Identify the bottom 5 selling products in the company
6. Summarize the total sales for year-wise.
7. Display the sale of each product for each country it is shipped to as a % of the entire sales.
8. Identify the top 5 selling product categories and the countries to which they were shipped.
9. Identify the top 5 selling products and the customers who buy them.

****

DWDM LAB EXERCISES - 14 SEP 2018

****

1. What is the total sales for each sales person
2. Summarize the % sales for products yearwise.
3. What is the total sales for each company in categories of products?
4. Find the top 10 selling items in the company.
5. Find the total sales of Boston crab meat customer wise
6. Visualize the sale of products as a pie chart.
7. Identify the bottom 5 selling products in the company and the salesman who sold it.
8. Display the sale of each product and the country it is shipped to as a % of the entire sales.
9. Identify the top 5 selling products and the customer who buys them.
10. Find the customers region wise sale of items in each category.

****

DWDM LAB EXERCISES - 21 SEP 2018

****

1. FOR THE BANK.xls dataset perform preprocessing and find atleast 5 interesting rules with pep on the RHS of the rule.

2. Use the supermarket.arff dataset and perform the following tasks:
   (i) Split the dataset into 2 datasets: one containing items and another containing departments.
   (ii) For the item dataset find the most frequent itemsets ranked by lift.
   (iii) Find the top five selling items in the dataset
   (iv) For the topmost selling item, find association rules with the item on the RHS of the rule.
   (v) Find the top 5 association rules for the department dataset ranked as per the lift.
   (vi) Find top 5 association rules with total in the RHS using the item dataset.
   (vii)

****

DWDM LAB EXERCISES – 25SEP2018

****

1. Using the Weather Nominal dataset and J48 algorithm, write down the classifier rules where the class label is play. Compare the accuracy of the classifier with the confusion matrix for two testing strategies: cross-validation and 66% split.

2. Diabetes.arff: Using this dataset perform required preprocessing and generate the top 5 association rules using the apriori algorithm with minimum lift of 1.

3. Using the same dataset, create classifiers using Naïve Bayesian and the Bayesian Belief Network. Compare the classifier accuracy using a Confusion Matrix.