Internship Program on Data Science Using R-Programming

R – Hackathon

Date : 10/8/2021 **Time: 90 Mins**

1. The attached car data have 400 records, with attributes - Age, Gender, Work Exp, Salary, Distance, License, Transport.

   Sample data for the first few records -

Age	Gender	Work Exp	Salary	Distance	license	Transport
20	Male	2	8.5	7	0	Public Transport
25	Male	3	10.5	7.1	0	Public Transport
33	Male	13	36.6	7.1	1	Public Transport
23	Female	4	8.4	7.1	0	Public Transport
21	Male	3	9.5	7.1	0	2Wheeler
30	Male	8	14.6	7.1	0	Public Transport
28	Female	5	14.6	7.2	0	Public Transport
23	Female	1	7.5	7.2	0	Public Transport
23	Male	3	11.7	7.2	0	2Wheeler

   Build a model and specify, what would be your prediction regarding their choice of transport for the following Employee. (Attach the R code, model output and predicted output of the following record)

Age	Gender	Work Exp	Salary	Distance	license
30	Male	5	20	7	1

2. With the car data, find the following - Attach the R code with Output
   a. Number of records and columns
   b. Structure of dataset
   c. Summary of the dataset
   d. Data type of attribute “Transport”
   e. Convert attribute “Transport” to categorical data type
3. Using car Data, generate the following output - Attach the R code and output
   a. List the records with Transport = “car”
   b. List the datasets with Age, Gender, Salary, Distance, Transport
   c. Add a new attribute and display the output
   d. Display the mean, standard deviation and variance of the Salary attribute
4. Using Car Dataset, apply the following Visualization Techniques to the appropriate attributes – Attach the R code and output
   a. Histogram
   b. Bar plot
   c. Box plot