1a GI = 1 – ((1/2)^2 + (1/2)^2) = 0.5

1b GI will be 0 for customer ID since all values are unique or each element belongs to one class.

1c

By gender:

Gender Male = 10/20

Gender Female = 10/20

Gender Male and Class C0 = 6/10

Gender Male and Class C1 = 4/10

**Gini Index Male = 1 – ((6/10)^2 + (4/10)^2) = 0.48**

Gender Female and Class C0 = 4/10

Gender Female and Class C1 = 6/10

**Gini Index Female = 1 – ((4/10)^2 + (6/10)^2) = 0.48**

**Gini Index for gender = 10/20(0.48) + 10/20(0.48) = 0.48**

1d

By car type:

Family = 4/20

Sports = 8/20

Luxury = 8/20

Family and Class C0 = 1/4

Family and Class C1 = 3/4

**Gini Index Family = 1 – ((1/4)^2 + (3/4)^2) = 0.375**

Sports and Class C0 = 8/8

Sports and Class C1 = 0

**Gini Index Sports = 1 – ((8/8)^2 + 0 ) = 0**

Luxury and Class C0 = 1/8

Luxury and Class C1 = 7/8

**Gini Index Luxury = 1 – ((4/10)^2 + (6/10)^2) = 0.21875**

**Gini Index for car type = 4/20(0.375) + 8/20(0) + 8/20(0.21875) = 0.1625**

1e

By shirt size:

Small = 5/20

Medium = 7/20

Large = 4/20

Extra Large = 4/20

Small and Class C0 = 3/5

Small and Class C1 = 2/5

**Gini Index Small = 1 – ((3/5)^2 + (2/5)^2) = 0.48**

Medium and Class C0 = 3/7

Medium and Class C1 = 4/7

**Gini Index Medium = 1 – ((3/7)^2 + (4/7)^2) = 0.4898**

Large and Class C0 = 2/4

Large and Class C1 = 2/4

**Gini Index Large = 1 – ((2/4)^2 + (2/4)^2) = 0.5**

Extra Large and Class C0 = 2/4

Extra Large and Class C1 = 2/4

**Gini Index X Large = 1 – ((2/4)^2 + (2/4)^2) = 0.5**

**Gini Index for car type = 5/20(0.48) + 7/20(0.48) + 4/20(0.5) + 4/20(0.5) = 0.488**

1f We will choose the GI with the lowest value to get the best split. The lowest GI in our dataset is by car type.

2

Entropy = - 4/9 log(4/9) – 5/9 log (5/9) = 0.9911

Entropy for a1 = 4/9 [ - 3/4 log(3/4) – 1/4 log(1/4) ] + 5/9 [ - 1/5 log(1/5) – 4/5 log(4/5) ] = 0.7616

|  |  |  |
| --- | --- | --- |
| a1 | T | F |
| + | 3/4 | 1/5 |
| - | 1/4 | 4/5 |
| Total | 4/9 | 5/9 |

|  |  |  |
| --- | --- | --- |
| a2 | T | F |
| + | 2/5 | 2/4 |
| - | 3/5 | 2/4 |
| Total | 5/9 | 4/9 |