Feature Engineering -Assignment 1

Answer (a)

|  |  |  |
| --- | --- | --- |
|  | Continuous or Discrete | Scale of Measurement |
| Amount of sugar in a Coke bottle | Continuous | Real Data (gm)/Ratio |
| Birth place of BITS university students | Discrete | Category (string) / Nominal |
| The size of Easter eggs represented as Small, Medium, Large, Extra Large, Jumbo | Discrete | Ordinal (cms) |
| Pan card number | Discrete | Category (string) / Nominal |
| Dates | Discrete | Interval |
| Number of defective shirts in packs of 100 manufactured by color plus | Continuous | Ratio |

Answer (b)

No, A noise is introduced in some readings due to incorrect transmission, signal distortion. A noise has lot of variance and differ for each sample. It simply cannot be constant like in this case.

Answer (c)

|  |  |  |
| --- | --- | --- |
| Student ID | Height(cm) | Weight(kg) |
| S1 | 169 | 60 |
| S2 | 171 | 70 |
| S3 | 150 | 80 |
| S4 | 180 | 75 |
| S5 | 150 | 60 |

We will use min-max scaling to contain the values of these features. Min-max scaling can be achieved using the formula,

X^ = [X - min(X) ] / ( max(X) - min (X) )

min(Height) = 150, min(Weight) = 60

max(Height) = 180 , max(Weight) = 80

X^ s4 = 75 - 60 / 80 - 60 = 15 / 20 = .75

Answer (d) See attached PY file.