PROBLEM SET 3, MRIDUL HARISH, CED18I034

Question 1 - Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70.

(a) Use min-max normalization to transform the values of age to the range [0:1].

(b) Use z-score normalization to transform the values of age.

(c) Use normalization by decimal scaling to transform the values of age such that the transformed value is less than 1.

In [ ]:
```
import pandas as pd
import numpy as np
import random
import matplotlib.pyplot as plt
```

In [ ]:
```
Age = np.array([13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 3
```

(i) Min-Max

In [ ]:
```
Age_min = min(Age)
Age_max = max(Age)
Age_min_max = [(i - Age_min)/(Age_max - Age_min) for i in Age]
```

In [ ]:
```
Age_min
```

Out[ ]:
```
13
```

In [ ]:
```
Age_max
```

Out[ ]:
```
70
```

In [ ]:
```
Age_min_max
```

Out[ ]:
```
[0.0,
 0.03508771929824561,
 0.05263157894736842,
 0.05263157894736842,
 0.10526315789473684,
 0.12280701754385964,
 0.12280701754385964,
 0.14035087719298245,
 0.15789473684210525,
 0.15789473684210525,
 0.21052631578947367,
 0.21052631578947367,
 0.21052631578947367,
```

```
0.21052631578947367,
0.2982456140350877,
0.3508771929824561,
0.3508771929824561,
0.38596491228070173,
0.38596491228070173,
0.38596491228070173,
0.38596491228070173,
0.40350877192982454,
0.47368421052631576,
0.5614035087719298,
0.5789473684210527,
0.6842105263157895,
```

In [ ]:
```
Age_mean = np.mean(Age)
Age_std = np.std(Age)
```

(ii)Z score

In [ ]:
```
Age_mean
```

Out[ ]:
```
29.962962962962962
```

In [ ]:
```
Age_std
```

Out[ ]:
```
12.700193878606099
```

In [ ]:
```
Age_Zscore = [(i - Age_mean)/Age_std for i in Age]
```

In [ ]:
```
Age_Zscore
```

Out[ ]:
```
[-1.3356459850221374,
 -1.1781680741243308,
 -1.0994291186754275,
 -1.0994291186754275,
 -0.8632122523287176,
 -0.7844732968798143,
 -0.7844732968798143,
 -0.705734341430911,
 -0.6269953859820077,
 -0.6269953859820077,
 -0.39077851963529775,
 -0.39077851963529775,
 -0.39077851963529775,
 -0.39077851963529775,
 0.0029162576092187234,
 0.23913312395592862,
 0.23913312395592862,
 0.3966110348537352,
 0.3966110348537352,
 0.3966110348537352,
 0.3966110348537352,
 0.4753499903026385,
 0.7903058120982517,
 1.1840005893427683,
 1.2627395447916716,
```

```
1.7351732774850914,
```

We can see that here that the dataset has at max 2 digit numbers, we can easily realize that j = 2

In [ ]:
```
Age_decimalscale = [i/100 for i in Age]
```

In [ ]:
```
Age_decimalscale
```

Out[ ]:
```
[0.13,
 0.15,
 0.16,
 0.16,
 0.19,
 0.2,
 0.2,
 0.21,
 0.22,
 0.22,
 0.25,
 0.25,
 0.25,
 0.25,
 0.3,
 0.33,
 0.33,
 0.35,
 0.35,
 0.35,
 0.35,
 0.36,
 0.4,
 0.45,
 0.46,
 0.52,
 0.7]
```

In [ ]:
```
Output = pd.DataFrame({"Age":Age, "Min-Max":Age_min_max, "Z-score":Age_Zscore,
```

In [ ]:
```
display(Output)
```

	Age	Min-Max	Z-score	Decimal Scale
0	13	0.000000	-1.335646	0.13
1	15	0.035088	-1.178168	0.15
2	16	0.052632	-1.099429	0.16
3	16	0.052632	-1.099429	0.16
4	19	0.105263	-0.863212	0.19
5	20	0.122807	-0.784473	0.20
6	20	0.122807	-0.784473	0.20
7	21	0.140351	-0.705734	0.21
8	22	0.157895	-0.626995	0.22

	Age	Min-Max	Z-score	Decimal Scale
9	22	0.157895	-0.626995	0.22
10	25	0.210526	-0.390779	0.25
11	25	0.210526	-0.390779	0.25
12	25	0.210526	-0.390779	0.25
13	25	0.210526	-0.390779	0.25
14	30	0.298246	0.002916	0.30
15	33	0.350877	0.239133	0.33
16	33	0.350877	0.239133	0.33
17	35	0.385965	0.396611	0.35
18	35	0.385965	0.396611	0.35
19	35	0.385965	0.396611	0.35
20	35	0.385965	0.396611	0.35
21	36	0.403509	0.475350	0.36
22	40	0.473684	0.790306	0.40
23	45	0.561404	1.184001	0.45
24	46	0.578947	1.262740	0.46
25	52	0.684211	1.735173	0.52