Homework 1

Name: Damilola David Popoola
ID: 249157974

Task 1(a)

The QI's are: Age, Education, Marital Status and Race

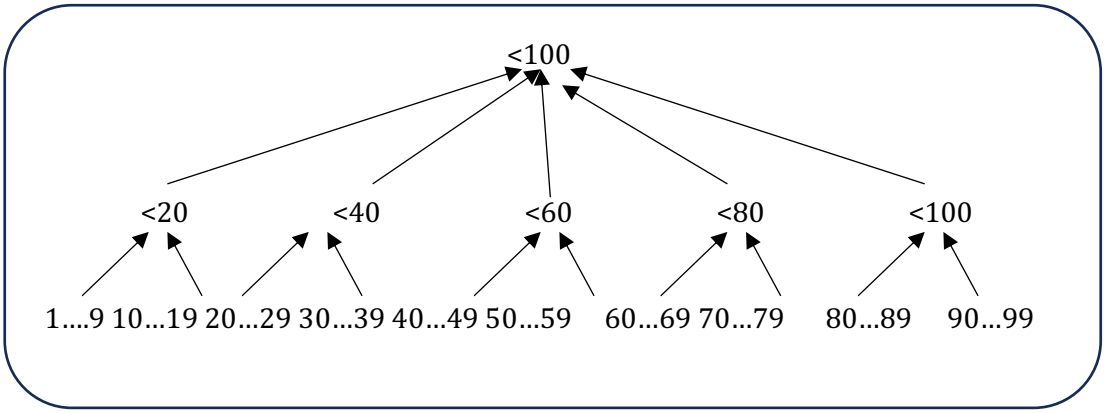

Figure 1: Age Hierarchy

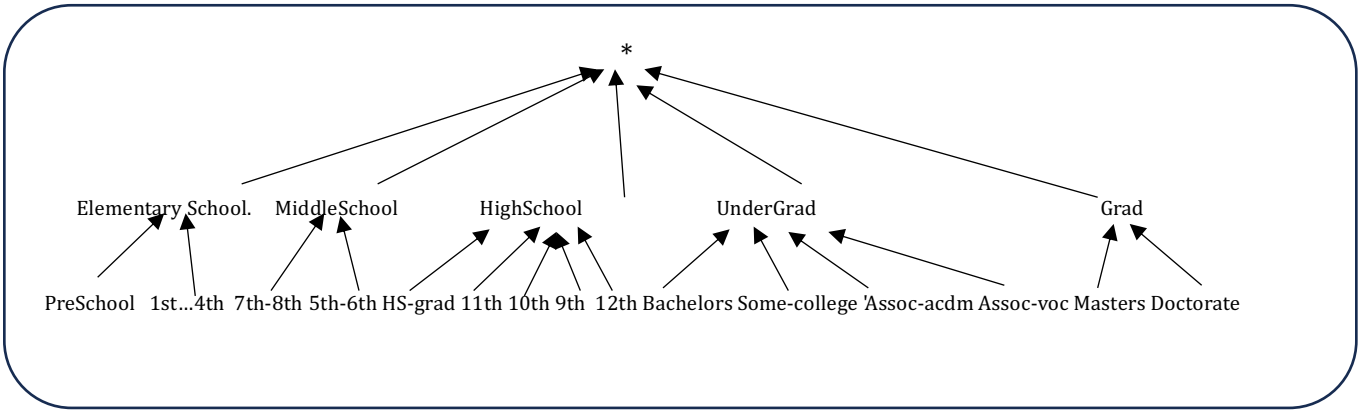

Figure 2: Education Hierarchy

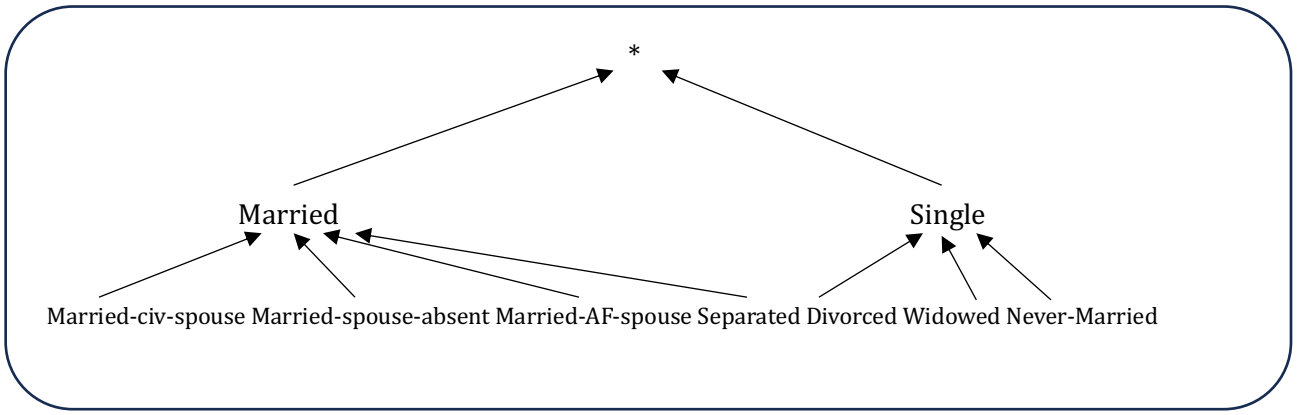

Figure 3: Marital Status Hierarchy

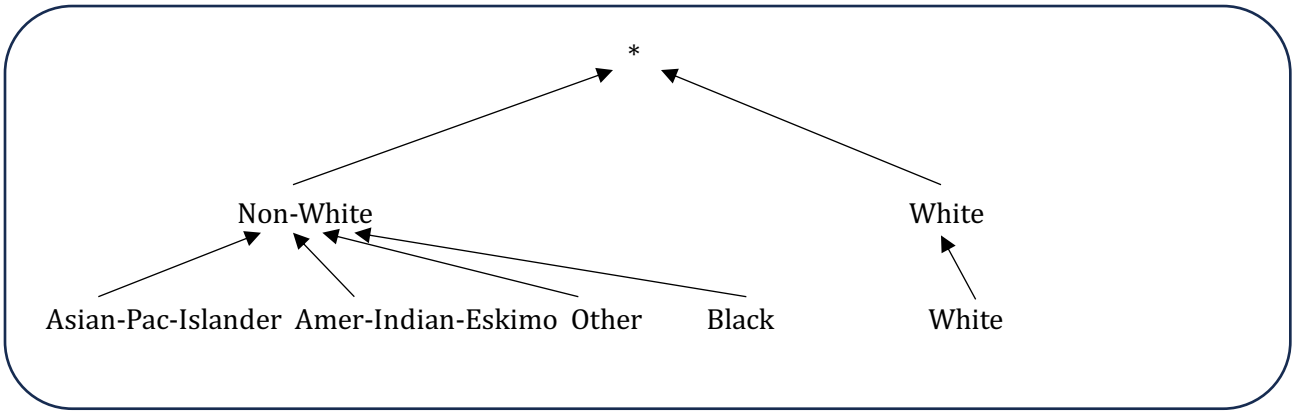

Figure 4: Race Hierarchy

Task 1(C) Computing Precision and Distortion

Calculation Precision and Distortion

In [686]:
```
def compute_distortion(max_hierarchy_level, generalization_level, attributes):
    return sum(generalization_level / level for level in hierarchy_levels)/attributes

max_hierarchy_level = [2, 2, 2, 2] #Hierarchy used for each attribute is 2
generalization_level = 1 #Generalization used for each attribute is 1
attributes = 4    #Total number of attributes is 4

# Compute the Distortion
total_distortion = compute_distortion(max_hierarchy_level, generalization_level, attributes)

def compute_precision(max_hierarchy_level, generalization_level):
    precisions = [(1 - (generalization_level / level)) for level in max_hierarchy_level]
    return sum(precisions) / len(precisions)

# Compute the Precision
precisions = compute_precision(max_hierarchy_level, generalization_level)

print(f"Distortion is: {total_distortion:.2f} and  Precision: {precisions:.2f}")
```

```
Distortion is: 0.50 and  Precision: 0.50
```

Figure 5: Precision and Distortion for the k-Anonymity

Task 2(c)

In [1240]:
```
# Define parameters for c-diversity and other configurations
k_value = 5

# Use the modified 'entropy_c_diversity' function and generalize_age function
c_values = [0.5, 1, 2]
sensitive_attribute = 'Occupation'

# Iterate over different values of 'c'
for c_value in c_values:
    result_entropy_c_diversity = entropy_c_diversity(df, k_value, c_value, sensitive_attribute)
    precision_sensitive = 1.0  # Assuming perfect precision


    # Print the distortion and precision values for each 'c' value
    print(f"Results for c={c_value}:")
    print(f"Precision for Sensitive attribute: {precision_sensitive}")

    # Write the output dataset to a file (you can customize the file path)
    result_entropy_c_diversity.to_csv(f'hw1-2-output_entropy_c_diversity_c{c_value}.csv', index=False)
```

```
Results for c=0.5:
Precision for Sensitive attribute: 1.0
Results for c=1:
Precision for Sensitive attribute: 1.0
Results for c=2:
Precision for Sensitive attribute: 1.0
```

Figure 5: Precision for the L-Diversity