Digital Assignment 2

1. Consider the dataset from the below link

   https://drive.google.com/file/d/1faHS-mVx2UP1D32AfeYAv0DekQ0gUknw/view?usp=sharing

   a. Find out how many data is missing in each attribute
   b. For all the missing values in 'car' attribute, fill the missing value with the mode.
   c. For the 'BuildingArea' attribute, fill the missing value with the linear interpolation and quadratic interpolation.
   d. Fill the 'yearbuilt' attribute with forward fill approach
   e. Remove all the rows which doesn't have a councilarea data.

2. Consider the below data set

   dataset = [['Milk', 'Onion', 'Nutmeg', 'Kidney Beans', 'Eggs', 'Yogurt'],

   ['Dill', 'Onion', 'Nutmeg', 'Kidney Beans', 'Eggs', 'Yogurt'],

   ['Milk', 'Apple', 'Kidney Beans', 'Eggs'],

   ['Milk', 'Unicorn', 'Corn', 'Kidney Beans', 'Yogurt'],

   ['Corn', 'Onion', 'Onion', 'Kidney Beans', 'Ice cream', 'Eggs']]

   a. Find the frequent itemset from the dataset with support count = 4 and support count = 2.
   b. Generate the association rules for support count 2 using the two methods 'confidence' and 'lift'.

3. Suppose if the dataset has a Nan value

   dataset = [['Milk', 'Onion', Nan, 'Kidney Beans', 'Eggs', 'Yogurt'],

   ['Dill', 'Onion', 'Nutmeg', 'Kidney Beans', 'Eggs', 'Yogurt'],

   ['Milk', 'Apple', 'Kidney Beans', 'Eggs'],

   ['Milk', 'Unicorn', 'Corn', Nan, 'Yogurt'],

   ['Corn', 'Onion', Nan, 'Kidney Beans', 'Ice cream', 'Eggs']]

   How does the frequent itemset change for the support count of 2