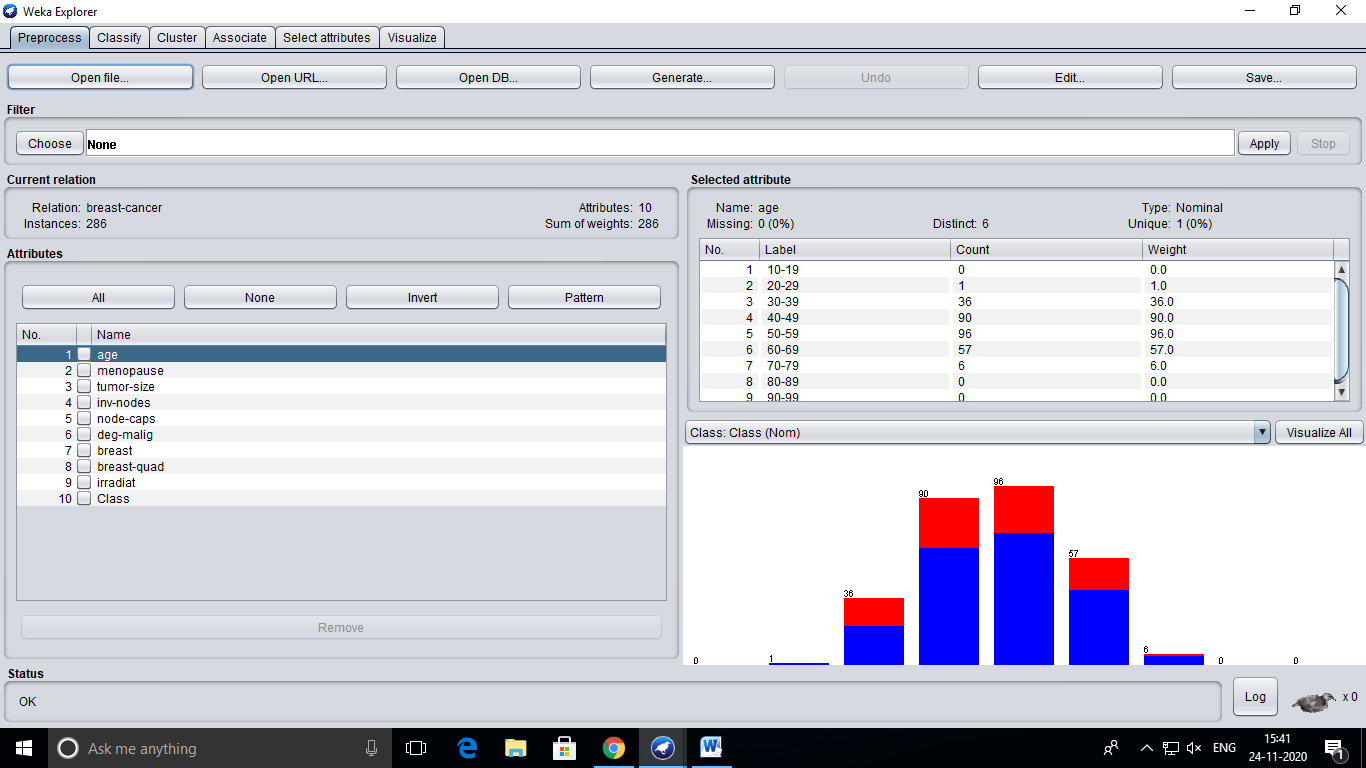
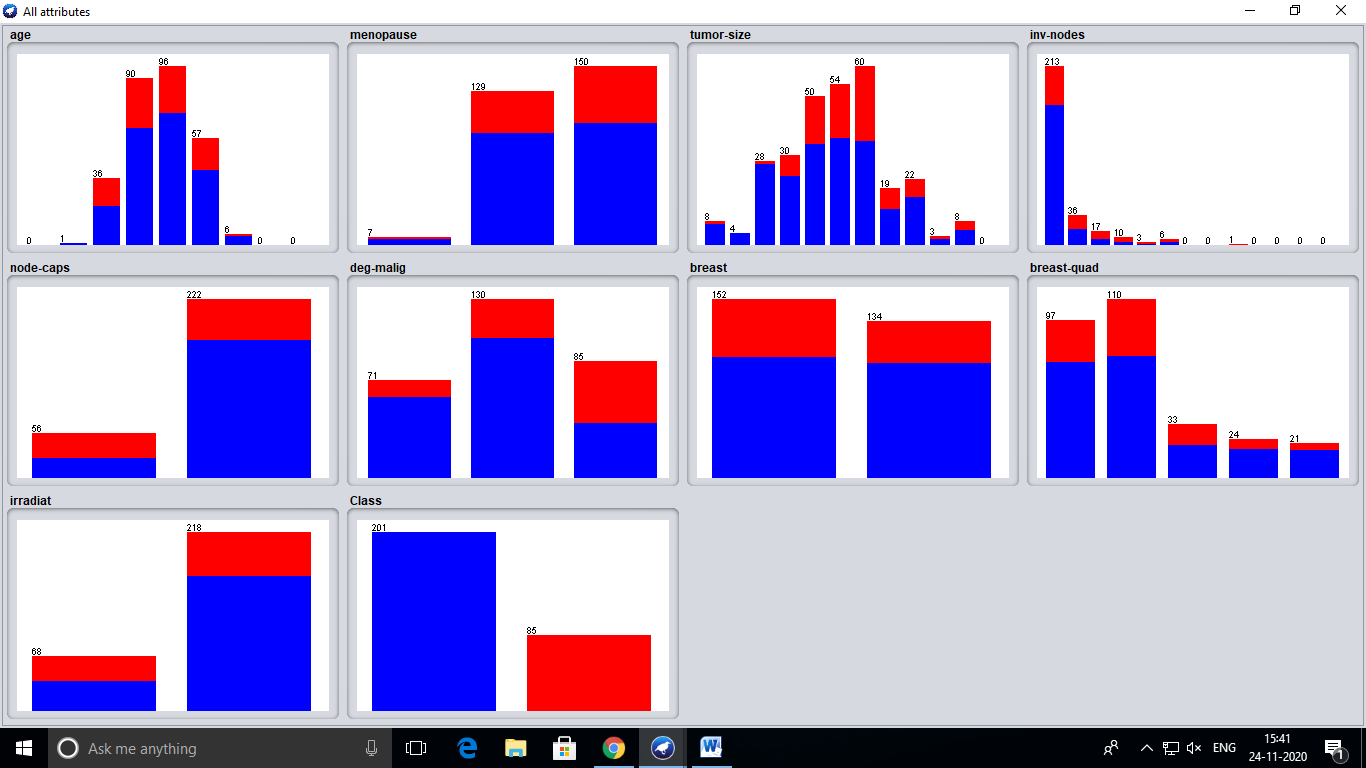
1. **Write a program to implement k-means clustering algorithm**





=== Run information ===

Scheme: weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10

Relation: breast-cancer

Instances: 286

Attributes: 10

age

menopause

tumor-size

inv-nodes

node-caps

deg-malig

breast

breast-quad

irradiat

Class

Test mode: evaluate on training data

=== Clustering model (full training set) ===

kMeans

======

Number of iterations: 3

Within cluster sum of squared errors: 1177.0

Initial starting points (random):

Cluster 0: 50-59,premeno,10-14,0-2,no,2,right,left\_up,no,no-recurrence-events

Cluster 1: 40-49,premeno,15-19,0-2,yes,3,right,left\_up,no,recurrence-events

Missing values globally replaced with mean/mode

Final cluster centroids:

Cluster#

Attribute Full Data 0 1

(286.0) (225.0) (61.0)

======================================================================================

age 50-59 50-59 40-49

menopause premeno premeno premeno

tumor-size 30-34 25-29 30-34

inv-nodes 0-2 0-2 0-2

node-caps no no yes

deg-malig 2 2 3

breast left left left

breast-quad left\_low left\_low left\_low

irradiat no no no

Class no-recurrence-events no-recurrence-events recurrence-events

Time taken to build model (full training data) : 0.01 seconds

=== Model and evaluation on training set ===

Clustered Instances

0 225 ( 79%)

1 61 ( 21%)