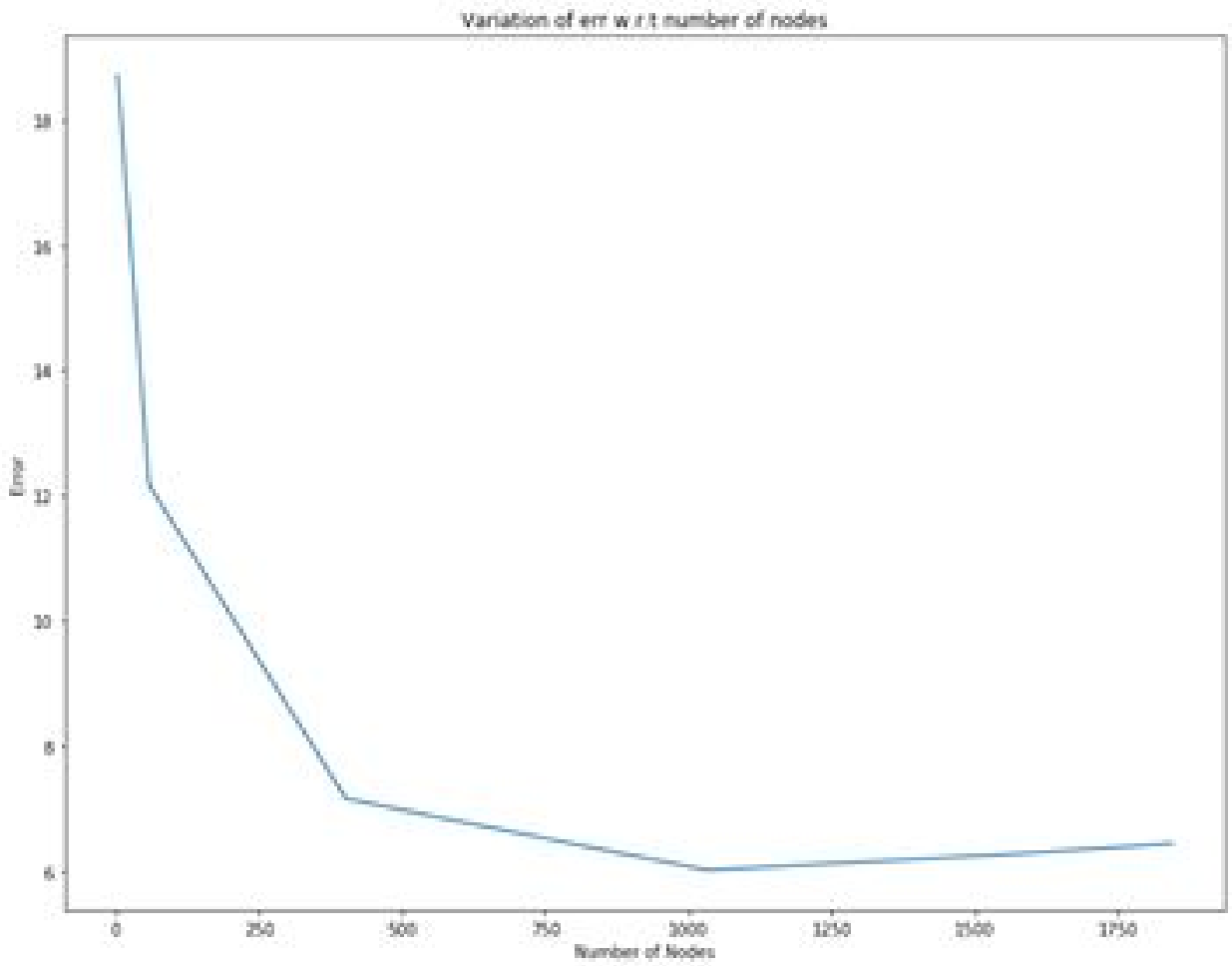


Question 5:

1. The graph is of U shape, the error first decreases then it increases after some point.
2. The graph is shown above

Question 6:

There are several methods used by various decision trees. We can simply ignore the missing values or treating the missing values as another category are not real handling missing values.

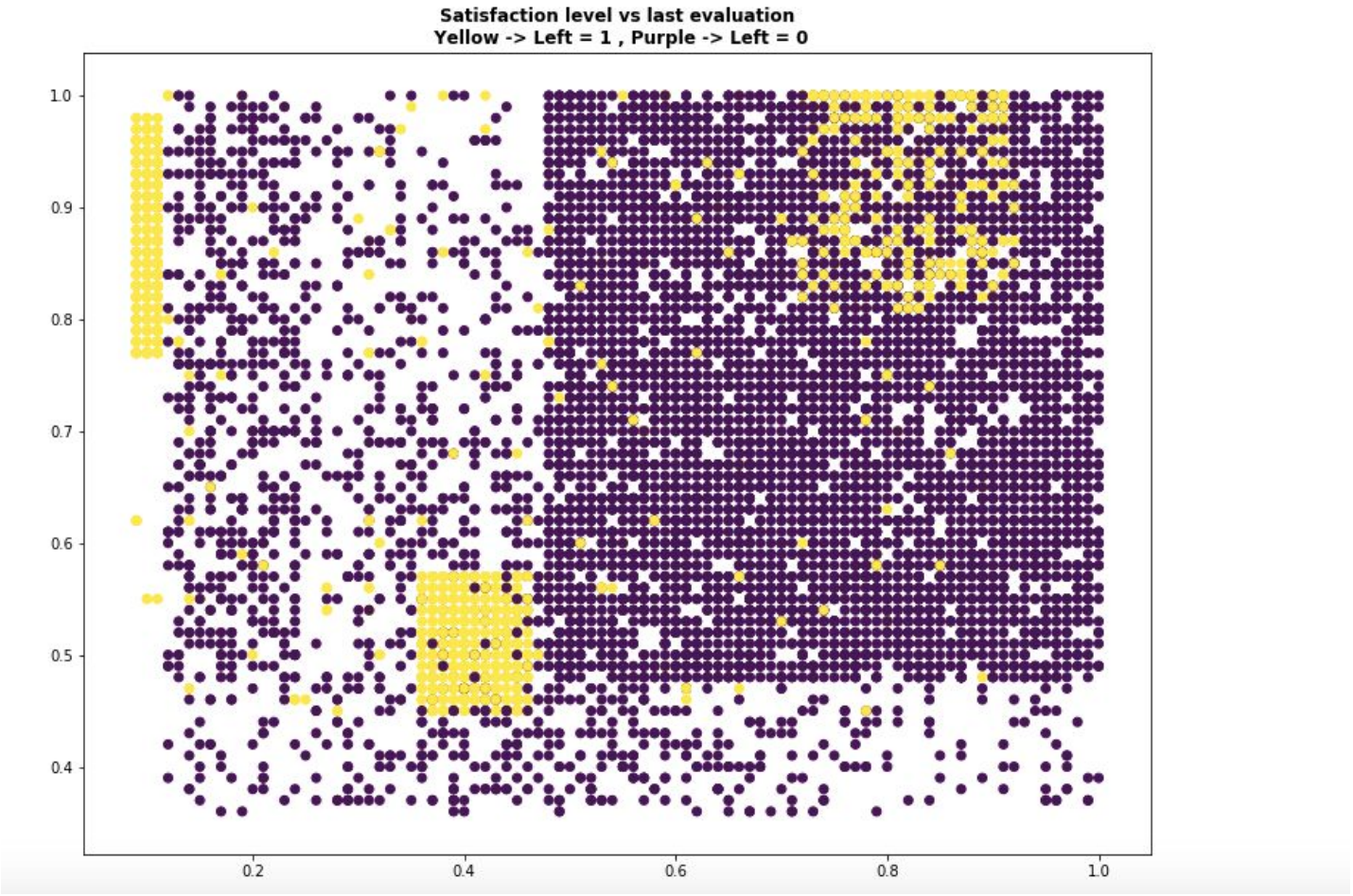
The real handling approaches to missing data does not use data point with missing values in the evaluation of a split. However, when child nodes are created and trained, those instances are distributed somehow.

The following approaches to distribute the missing value instances to child nodes:

1. All goes to the node which already has the biggest number of instances (CART, is not the primary rule) distribute to all children, but with diminished weights, proportional with the number of instances from each child node (C45 and others)

2. Distribute randomly to only one single child node, eventually according with a categorical distribution (I have seen that in various implementations of C45 and CART for a faster running time)
3. Build, sort and use surrogates to distribute instances to a child node, where surrogates are input features which resembles best how the test feature send data instances to left or right child node (CART, if that fails, the majority rule is used)

Question 4:



1. This is the best comparison graph where we could easily visualise decision boundaries.
2. The graph is with respect to satisfaction level and last evaluation