CERT24.COM

100% Success Guarantee of IT Exams

-The fastest and guaranteed way to certy now!

http://cert24.com/

Exam Number/Code:70-773

Exam Name: Analyzing Big Data with Microsoft R

Version: Demo

Q1
You are running a parallel function that uses the following R code segment. (Line numbers are included for reference only.)

```
01  cp <- 0.01 xval <- 0 maxdepth <- 5
02
03  (form, data = "segmentationDataBig",maxDepth = maxdepth, cp = cp, xVal = xval, blocksPerRead = 250
```
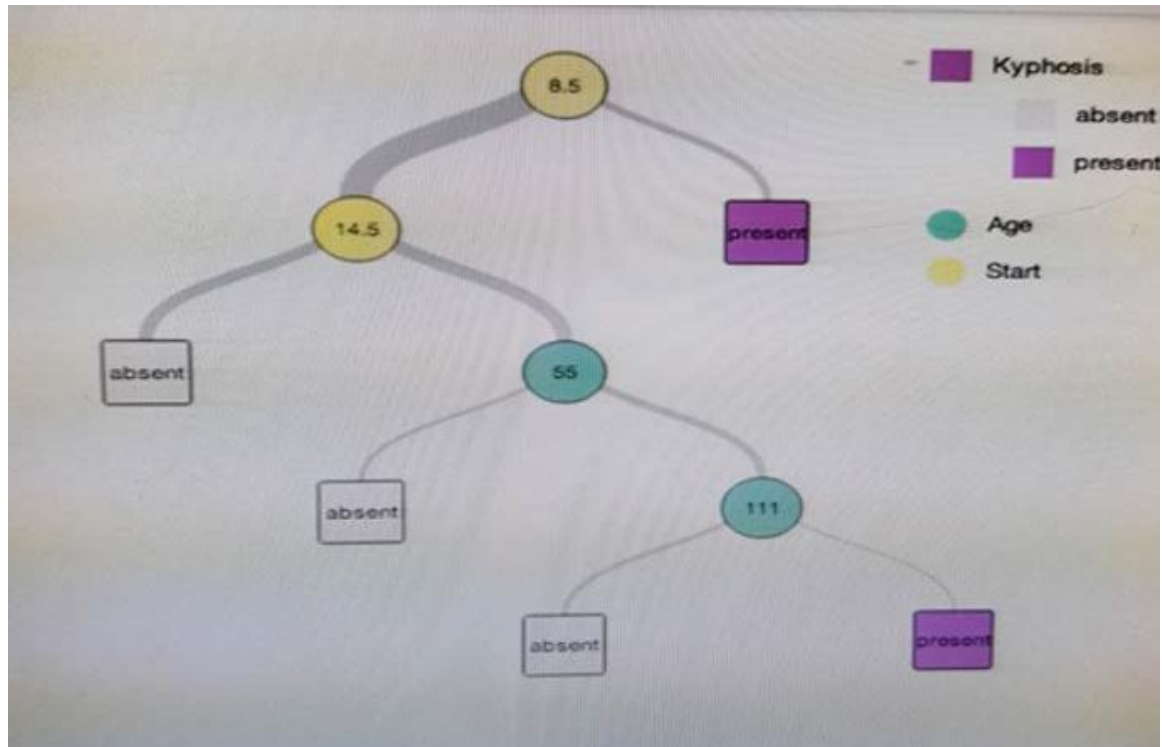
You need to complete the R code. The solution must support chunking.
Which function should insert at line 02?

A. rxBTrees
B. rxExec
C. rxDForest
D. rxDTree

Answer: C

Q2
You perform an analysis that produces the decision tree shown in the exhibit.

How many leaf nodes are there on the tree?

A. 2
B. 3
C. 5
D. 7

Answer: B

Q3
You need to generate a residual based on two columns. The solution must build a trend indicator.
Which function should you use?

A. rxPredict
B. rxLogit
C. Summary
D. rxLinMod
E. rxTweedie
F. stepAic
G. rxTransform
H. rxDataStep

Answer: C

Q4
You have following regression forest.


```
rxDForest(formula = stack.loss ~ Air.Flow + Water.Temp + Acid.Conc., data = stackloss, maxDepth = 3, nTree = 200, mTry
= 2)

Type of decision forest: anova
Number of trees: 200
No. of variables tried at each split: 2
Mean of squared residuals: 44.54992 % Var explained: 65
```

Which variable contributes the most to the dependent variable?

A. stack.loss
B. Water.Temp
C. Air.Flow
D. Acid.Conc

Answer: A

Q5

HOTSPOT

You are developing a Microsoft R Open solution that will leverage the computing power of the database server for some of your datasets.

You are performing feature engineering and data preparation for the datasets.

The following is a sample of the dataset.

```
rxGetInfo(df)
head(df)

  age incwage perwt  wkswork1   state
1  50    9000    30    48      Indiana
2  41   35000    20    48      Indiana
3  55   40400    21    49      Indiana
4  56   45000    30    52      Indiana
5  46   17200    60    51      Indiana
6  49   35000    21    52      Indiana
```

You need to sort the data from the dataset sample and to remove duplicates by using wkswork1. Which R code segment should you use? to answer, select the appropriate options in the answer area.

Note: Each correct selection is worth one point.

```
rxSort(inData = [sampleInData], outFile = [sampleOutData], sortByVars = c("incwage", "perwt", "age", "wkswork1"),
                                           ▼  ,                                    ▼  = "wkswork1")
  removeDupKeys = FALSE                        dupFreqVar
  removeDupKeys = TRUE                          varsToDrop
  rxMerge = FALSE                               VarsToKeep
  rxMerge = TRUE
```

Answer:

```
rxSort(inData = [sampleInData], outFile = [sampleOutData], sortByVars = c("incwage", "perwt", "age", "wkswork1"),
                                           ▼  ,                                    ▼  = "wkswork1")
  removeDupKeys = FALSE                        dupFreqVar
 [removeDupKeys = TRUE]                         varsToDrop
  rxMerge = FALSE                               VarsToKeep
  rxMerge = TRUE
```

Q6

You have a data source that is larger than memory. You need to visualize the distribution of the values for a variable in the data source.

What should you use?

A. the Describe package
B. the rxHistogram function
C. the rxSummary function
D. the rxQuantile function
E. the rxCube function

F. the summary function

G. the rxCrossTabs function

H. the ggplot2 package

Answer: B