

Microsoft 70-773 Exam

Volume: 39 Questions

Question: 1

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
```

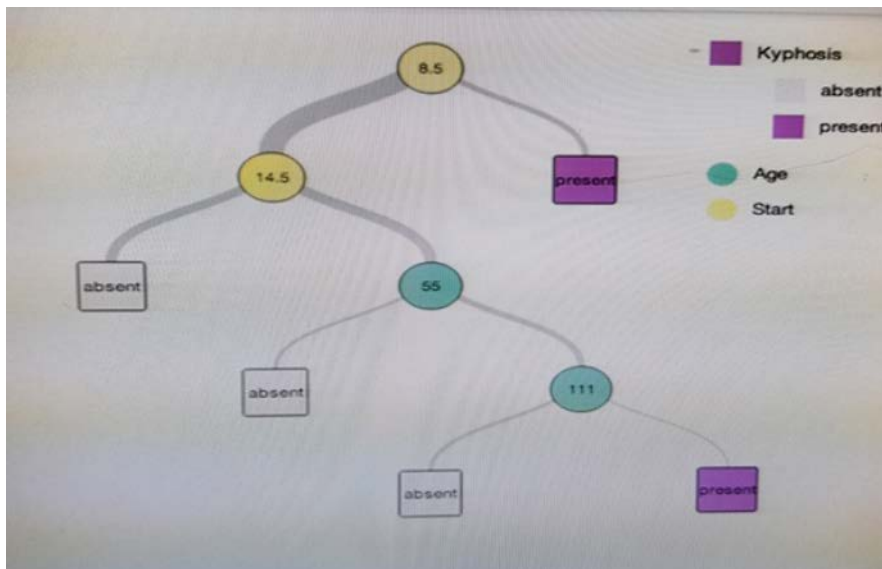
You need to complete the R code. The solution must support chunking. Which function should insert at line 02?

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

Answer: C

Question: 2

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

Microsoft 70-773 Exam

C. 5

D. 7

Answer: B

Question: 3

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

Question: 4

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

Microsoft 70-773 Exam

C. Air.Flow

D. Acid.Conc

Answer: A

Question: 5 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.

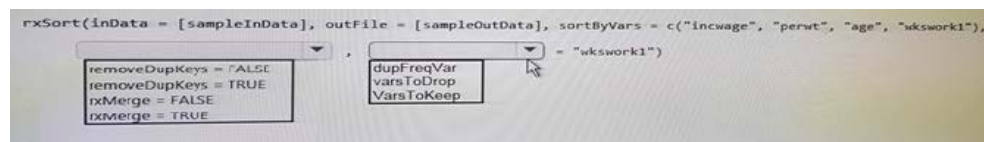
The following is a sample of the dataset.

```
rxGetInfo(df)
head(df)
```

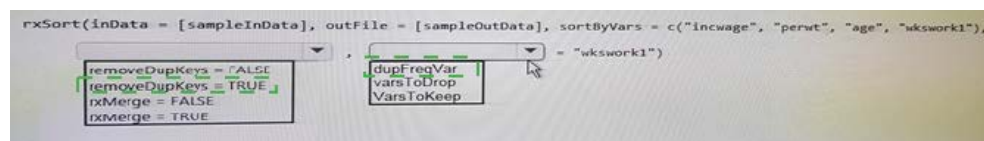
	age	incwage	perwt	wkswork1	state
1	50	9000	30	48	Indiana
2	41	35000	20	48	Indiana
3	55	40400	21	49	Indiana
4	56	45000	30	52	Indiana
5	46	17200	60	51	Indiana
6	49	35000	21	52	Indiana

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.



Answer:



Question: 6

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

Microsoft 70-773 Exam

- 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

Question: 7

You have an Apache Hadoop Hive data warehouse. RevoScaleR is not installed. You need to sort the data according to the variables in the dataset. What should you do?

- A. Connect to the database by using an ODBC connection, and then use the rxSort function.
- B. Create a table in the ORC file format.
- C. Connect to the database by using an ODBC connection, and then use the rxDataStep function.
- D. Execute a Hive query that sorts the data, and then reads the results.

Answer: D

Question: 8

You need to calculate a measure of central tendency and variability for the variables in a dataset that is grouped by using another categorical variable. What should you use?

- A. the Describe package
- B. the rxHistogram function
- C. the rxSummary function

Microsoft 70-773 Exam

- D. the rxQuantile function
- E. the rxCube function
- F. the summary function
- G. the rxCrossTabs function
- H. the ggplot2 package

Answer: C

Question: 9

You have a dataset that has multiple blocks and only numeric variables.

You are computing in a local compute context.

You plan to lag a variable named x to create a new variable named x_lagged by using a transform function. You will create a new element in the output of the function.

You need to minimize the number of missing values.

Which three actions should you perform? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. Assign a value to the first value of x_lagged in the current block.
- B. Use rxSet to store the last value of x_lagged in the current block.
- C. Use rxSet to store the last value of x in the current block.
- D. Use rxGet to retrieve the first value of x in the next block to be processed.
- E. Use rxGet to retrieve a value stored in processing of the prior block.

Answer: A,C,D

Question: 10

You use dplyrXdf. and you discover that after you exit the session, the output files that were created were deleted. You need to prevent the files from being deleted.

Solution:

You use rxSetComputeContext with the local parameter before performing operations that save results.

Does this meet the goal?