Microsoft

Exam Questions 70-773

Analyzing Big Data with Microsoft R (beta)

NEW QUESTION 1
You perform an analysis that produces the decision tree shown in the exhibit.
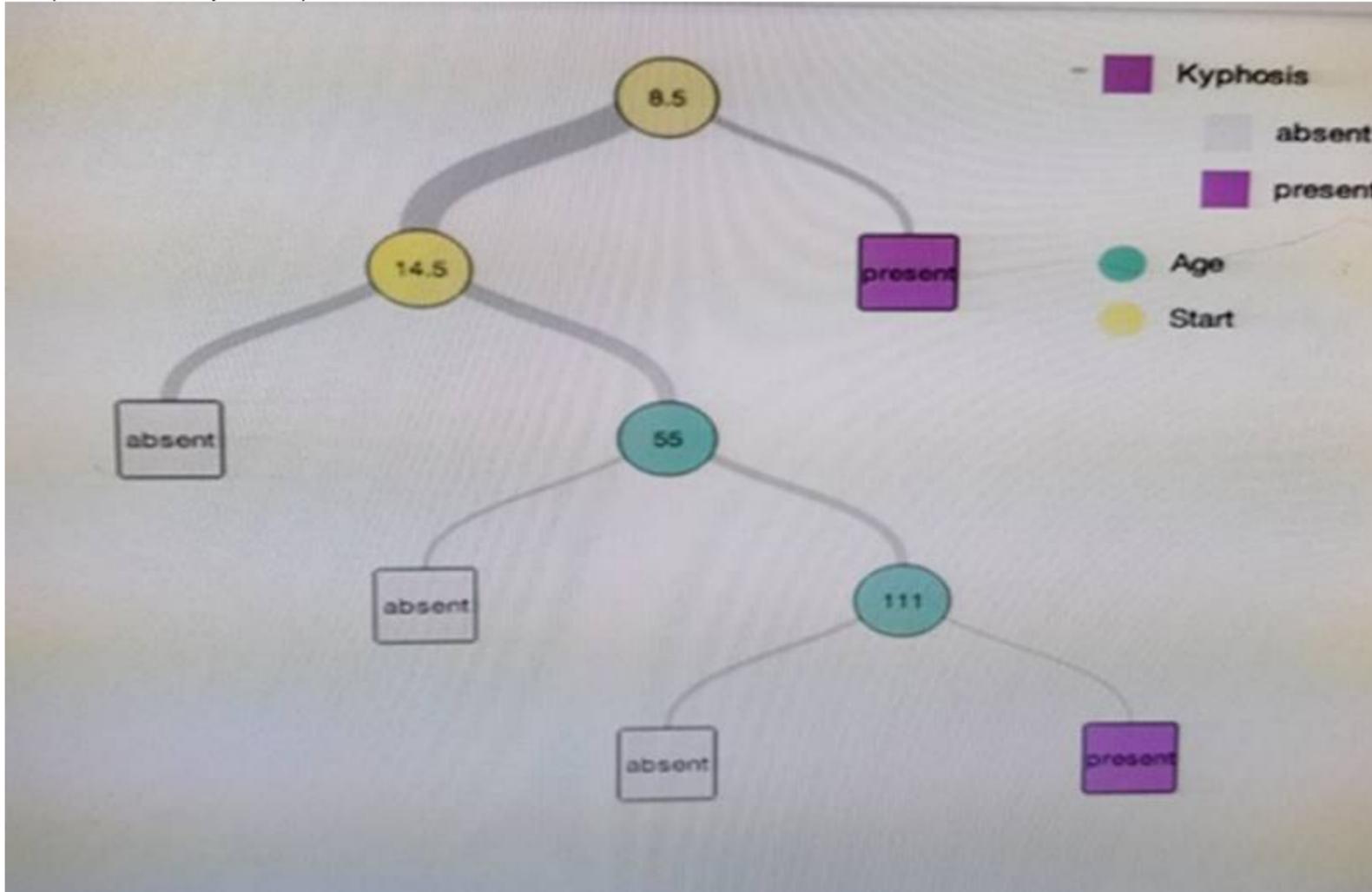
How many leaf nodes are there on the tree?

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

Answer: B

NEW QUESTION 2
Note: This question Is part of a series of questions that use the same or similar answer choice. An answer choice may be correct for more than one question in the series. Each question is independent of the other questions in this series.
Information and details provided In a question apply only to that question.
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

NEW QUESTION 3
HOTSPOT
Note: This question is part of a series of questions that use the same scenario. For your convenience, the scenario is repeated in each question. Each question presents a different goal and answer choices, but the text of the scenario is exactly the same in each question in this series.
Start of repeated scenario
You are developing a Microsoft R Open solution that will leverage the computing power 0f 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

End of repeated scenario
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:

Explanation:
```
rxSort(inData = [sampleInData], outFile = [sampleOutData], sortByVars = c("incwage", "perwt", "age", "wkswork1"),
```
```
▼    ,    ▼    =  "wkswork1")

removeDupKeys = FALSE          dupFreqVar
removeDupKeys = TRUE           varsToDrop
rxMerge = FALSE               VarsToKeep
rxMerge = TRUE
```

NEW QUESTION 4
Note: This question is part of a series of questions that use the same or similar answer choices. An answer choice may be correct for more than one question in
the series. Each question is independent of the other questions in this series. Information and details provided in a question apply only to that question.
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
D. the rxQuantile function
E. the rxCube function
F. the summary function
G. the rxCrossTabs function
H. the ggplot2 package

Answer: C

NEW QUESTION 5
You need to build a model that looks at the probability of an outcome. You must regulate between L1 and L2.
Which classification method should you use?

A. Two-Class Neural Network
B. Two-Class Support Vector Machine
C. Two-Class Decision Forest
D. Two-Class Logistic Regression

Answer: A

NEW QUESTION 6
Note: This question Is part of a series of questions that use the same or similar answer choice. An answer choice may be correct for more than one question in the
series. Each question is independent of the other questions in this series.
Information and details provided In a question apply only to that question.
You need to evaluate the significance of coefficient that are produced by using a model that was estimated already.
Which function should you use?

A. rxPredict
B. rxLogit
C. Summary

D. rxLinMod
E. rxTweedie
F. stepAic
G. rxTransform
H. rxDataStep

Answer: D

Explanation: https://docs.microsoft.com/en-us/r-server/r/how-to-revoscaler-linear-model

NEW QUESTION 7
Note: This question is part of a series of Questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, whale others might not have a correct solution-After you answer a question in this section, you will NOT be able to return to it- As a result, these questions will not appear in the review screen.
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 dplyrXdf with the persist verb.
Does this meet the goal?

A. Yes
B. No

Answer: A

NEW QUESTION 8
You have cloud and on-premises resources that include Microsoft SQL Server and a big data environment in Apache Hadoop.
You have 50 billion fact records.
You need to build time series models to execute forecasting reports on the fact records. What should you use?

A. RxSpark on the Hadoop cluster
B. RxHadoopMR on the Hadoop cluster
C. RxLocalseq on the SQL Server database
D. RxLocalParallel on the SQL Server database

Answer: A

NEW QUESTION 9
You have one class support vector machines (SVMs).
You have a large dataset, but you do not have enough training time to fully test the model. What is an alternative method to validate the model?

A. Use Principal Components Analysis (PCA) Based Anomaly detection
B. Replace the SVMs with two class SVMs.
C. Perform feature selection.
D. Use outlier detection.

Answer: A

NEW QUESTION 10
Note: This question is part of a series of Questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, whale others might not have a correct solution-After you answer a question in this section, you will NOT be able to return to it- As a result, these questions will not appear in the review screen.
You have a Microsoft SQL Server instance that has R Services (In-Database) installed. You need to monitor the R jobs that are sent to SQL Server.
Solution: You create an events trace configuration file and place the file in the same directory as the BXLServer process.
Does this meet the goal?

A. Yes
B. No

Answer: B

NEW QUESTION 10
HOTSPOT
Note: This question is part of a series of questions that use the same scenario. For your convenience, the scenario is repeated in each question. Each question presents a different goal and answer choices, but the text of the scenario is exactly the same in each question in this series.
Start of repeated scenario
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	52	Indiana
4	56	45000	30	52	Indiana
5	46	17200	60	52	Indiana
6	49	35000	21	52	Indiana

End of repeated scenario

You plan to score some data to create data features to address empty rows. You have the following R code.

```
xdfPath <- file.path(rxGetOption("[sampleInData], "), "inputfile.xdf")
xdfLagged <- [sampleOutDatainclud ingFeatures](fileext = ".xdf")
rxSort(inData = xdfPath,
       outFile = xdfLagged,
       sortByVars = "Date")
rxDataStep(inData = xdfLagged,
           outFile = xdfLagged,
           transformObjects = list(
               varToLag = "Open",
               newName = "previousOpen"),
           transformFunc = lagVar,
           append = "cols",
           overwrite = TRUE)
rxDataStep(xdfLagged,
           varsToKeep = c("Date", "Open", "previousOpen"),
           numRows = 10)
```

You need to transform the data and overwrite the current dataset.

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.

```
rxCube
rxDataStep          (inData = [sampleInData], outFile=[sampleOutDatainclud ingFeatures],
rxExec
transform

      transformFunc = computeNonLagFeatures,
                                              overwrite=FALSE
                                              overwrite=TRUE
                                              varsToDrop=All
                                              varsToDrop=NONE
```

Answer:

Explanation:

```
rxCube
rxDataStep          (inData = [sampleInData], outFile=[sampleOutDatainclud ingFeatures],
rxExec
transform

      transformFunc = computeNonLagFeatures,
                                              overwrite=FALSE
                                              overwrite=TRUE
                                              varsToDrop=All
                                              varsToDrop=NONE
```
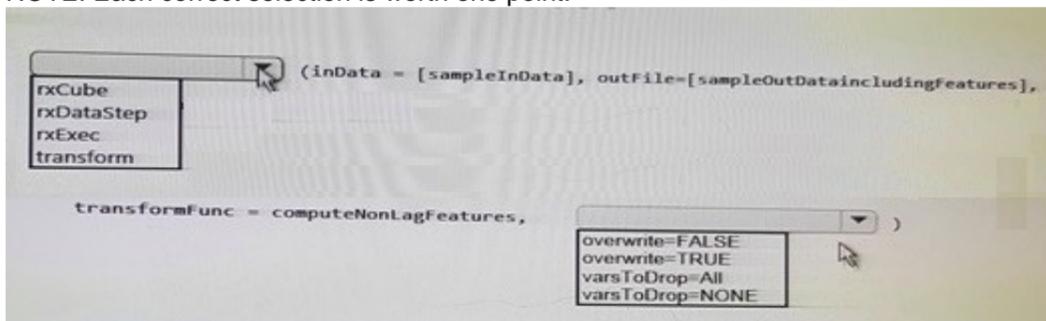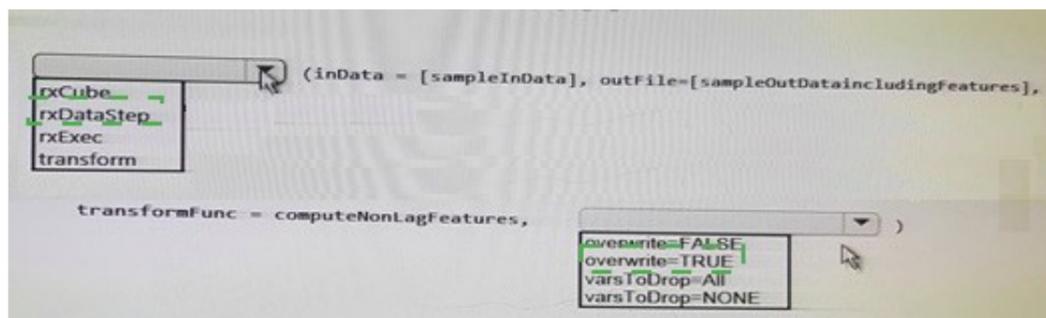
NEW QUESTION 13

You have a dataset that has a character variable. You need to create a bag of counts of n-grams. Which function should you use?

A. featurizeText0
B. categoricalHash0
C. concat0
D. selcctFeatures0
E. categorical0

Answer: A

Explanation: featurizeText: Produces a bag of counts of sequences of consecutive words, called n-grams, from a given corpus of text. It offers language detection, tokenization, stopwords removing, text normalization and feature generation.

NEW QUESTION 18
Note: This question is part of a series of questions that use the same or similar answer choices. An answer choice may be correct for more than one question in the series. Each question is independent of the other questions in this series. Information and details provided in a question apply only to that question.
You need to get all of the deciles for a variable in a data frame. 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: F

NEW QUESTION 22
You need to run a larger data tree model by using rsDForest. The model must use cross validation.
Which rxDForest option should you use?

A. maxSurrogate
B. maxNumBins
C. maxDepth
D. maxCompete
E. xVal

Answer: E

Explanation: https://docs.microsoft.com/en-us/r-server/r/how-to-revoscaler-decision-tree

NEW QUESTION 26
Note: This Question is part of a series of Questions that use the same or similar answer choices. An answer choice may be correct than one question in the series. Each question is independent of the other questions in this series. Information and details provided in a question apply only to that question.
You have a dataset that contains the physical characteristics of people.
You need to visualize a relationship between height and weight for a subset of observations in the dataset.
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: E

NEW QUESTION 29
Note: This question Is part of a series of questions that use the same or similar answer choice. An answer choice may be correct for more than one question in the series. Each question is independent of the other questions in this series.
Information and details provided In a question apply only to that question. You build a model that uses xyz regression.
You need to estimate a model that predicts a binary variable.
Which function should you use?

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

Answer: B

Explanation: https://docs.microsoft.com/en-us/r-server/r/how-to-revoscaler-logistic- regression

NEW QUESTION 34
Note: This question Is part of a series of questions that use the same or similar answer choice. An answer choice may be correct for more than one question in the series. Each question is independent of the other questions in this series.

Information and details provided In a question apply only to that question.
You need to estimate a model where the outcome variable is continuous, is in the range of [0,inf], and has a substantial mass at an exact value of 0.
Which function should you use?

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

Answer: H

NEW QUESTION 38
You plan to read data from an Oracle database table and to store the data in the file system for later processing by dplyrXdf, The size of the data is larger than the memory on the server to used for modelling.
You need to ensure that the data can be processed by dplyrXdf in the least amount of time possible.
How should you transfer the data from the Oracle database?

A. Use the RODBC library, connect to the Oracle database server by using odbcConnec
B. and then use rxDataStep to export the data to a comma-separated values (CSV) file.
C. Define a data source to the Oracle database server by using RxOdbcData, and then use rxImport to save the data to an XDF file.
D. Use the RODBC library, connect to the Oracle database server by using odbcConnec
E. and then use rxSplit to save the data to multiple comma-separated values (CSV) files.

Answer: C

NEW QUESTION 39
Note: This question is part of a series of Questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, whale others might not have a correct solution-After you answer a question in this section, you will NOT be able to return to it- As a result, these questions will not appear in the review screen.
You have a Microsoft SQL Server instance that has R Services (In-Database) installed. You need to monitor the R jobs that are sent to SQL Server.
Solution: You register an Extended Events package.
Does this meet the goal?

A. Yes
B. No

Answer: A

NEW QUESTION 40
......

Thank You for Trying Our Product

We offer two products:

1st - We have Practice Tests Software with Actual Exam Questions

2nd - Questons and Answers in PDF Format

70-773 Practice Exam Features:

* 70-773 Questions and Answers Updated Frequently

* 70-773 Practice Questions Verified by Expert Senior Certified Staff

* 70-773 Most Realistic Questions that Guarantee you a Pass on Your FirstTry

* 70-773 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year

100% Actual & Verified — Instant Download, Please Click
Order The 70-773 Practice Test Here