

100% Money Back
Guarantee

Vendor:Microsoft

Exam Code:70-773

Exam Name:Analyzing Big Data with Microsoft R

Version:Demo

QUESTION 1

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 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

Correct Answer: B

QUESTION 2

You are planning the compute contexts for your environment.

You need to execute rx-function calls in parallel.

What are three possible compute contexts that you can use to achieve this goal? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

- A. local parallel
- B. Spark
- C. local sequential
- D. Map Reduce
- E. SQL

Correct Answer: ABD

QUESTION 3

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

Correct Answer: D

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 estimate a model where the outcome variable is continuous, is in the range of $[0, \infty]$, 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

Correct Answer: F

QUESTION 5

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 evaluate the significance of coefficients 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

Correct Answer: D

References: <https://docs.microsoft.com/en-us/machine-learning-server/r/how-to-revoscaler-linear-model>

QUESTION 6

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, while

others might not have a correct solution.

After you answer a question in this sections, 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 rxSetComputeContext with the local parameter before performing operations that save results.

Does this meet the goal?

A. Yes

B. No

Correct Answer: B

QUESTION 7

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 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.

Hot Area:

Answer Area

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

, = "wkswork1")

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

Correct Answer:

Answer Area

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

, = "wkswork1")

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

QUESTION 8

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. featurizeText()
- B. categoricalHash()
- C. concat()
- D. selectFeatures()
- E. categorical()

Correct Answer: A

References: <https://docs.microsoft.com/en-us/machine-learning-server/python-reference/microsoftml/featurize-text>

QUESTION 9

You are running a large logistic regression for 1,000 feature variables by using the LogisticRegression() function in the MicrosoftML package. All of the predictor variables are numeric.

Currently, you specify the input variables separately by using the following formula.

Outcome ~ Feature000 + Feature001 + Feature002 + ... + Feature999

You discover that it takes 20 minutes to estimate each model.

You need to reduce the amount of time required to estimate each model without losing any information in the predictors.

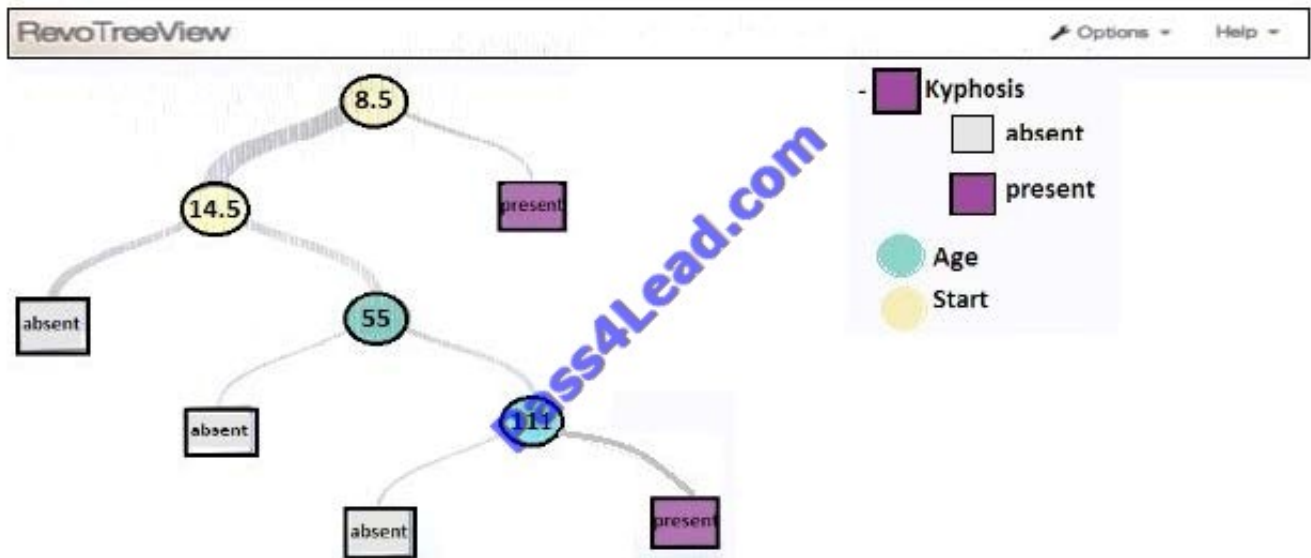
What should you do?

- A. Use `stepControl()` to perform stepwise regression to limit the number of variables that contribute to the model.
- B. Use `selectFeatures()` to select the features that provide the most information about the outcome variable.
- C. Use `princomp()` on the correlation matrix of Features, and then use only the first 100 principle components to reduce the number of input variables.
- D. Use `concat()` to create a single array variable named Features, and then specify a new formula named Outcome ~ Features.

Correct Answer: D

QUESTION 10

You perform an analysis that produces the decision tree shown in the exhibit. (Click the Exhibit button.)



How many leaf nodes are there on the tree?

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

Correct Answer: C

References: <https://docs.microsoft.com/en-us/machine-learning-server/r/how-to-revoscaler-decision-tree>

QUESTION 11

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, while

others might not have a correct solution.

After you answer a question in this sections, 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

Correct Answer: A

References: <https://docs.microsoft.com/en-us/sql/advanced-analytics/r/extended-events-for-sql-server-r-services>

QUESTION 12

You need to set the compute context for three different target environments.

Which statement should you use for each environment? To answer, drag the appropriate statements to the correct execution contexts. Each statement may be used once, more than once, or not at all. You may need to drag the split bar

between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Select and Place:

Statements	Answer Area
<input type="text" value="RxHadoopMR()"/>	Parallelized execution across the cores of the edge node server, except for rxExec calls, which are executed serially: <input type="text" value="Statement"/>
<input type="text" value="rxSetComputeContext('local')"/>	
<input type="text" value="rxSetComputeContext('localpar')"/>	Parallelized execution across the cores of the edge node server: <input type="text" value="Statement"/>
<input type="text" value="RxSpark()"/>	Parallelized distributed execution via Map Reduce across the nodes of the cluster: <input type="text" value="Statement"/>

Correct Answer:

Statements

RxSpark()

Answer Area

Parallelized execution across the cores of the edge node server, except for rxExec calls, which are executed serially:

Parallelized execution across the cores of the edge node server:

Parallelized distributed execution via Map Reduce across the nodes of the cluster:

To Read the [Whole Q&As](#), please purchase the [Complete Version](#) from [Our website](#).

Try our product !

100% Guaranteed Success

100% Money Back Guarantee

365 Days Free Update

Instant Download After Purchase

24x7 Customer Support

Average **99.9%** Success Rate

More than **800,000** Satisfied Customers Worldwide

Multi-Platform capabilities - **Windows, Mac, Android, iPhone, iPod, iPad, Kindle**

Need Help

Please provide as much detail as possible so we can best assist you.

To update a previously submitted ticket:



 <p>One Year Free Update Free update is available within One Year after your purchase. After One Year, you will get 50% discounts for updating. And we are proud to boast a 24/7 efficient Customer Support system via Email.</p>	 <p>Money Back Guarantee To ensure that you are spending on quality products, we provide 100% money back guarantee for 30 days from the date of purchase.</p>	 <p>Security & Privacy We respect customer privacy. We use McAfee's security service to provide you with utmost security for your personal information & peace of mind.</p>
---	---	--

Any charges made through this site will appear as Global Simulators Limited.

All trademarks are the property of their respective owners.