EMC

Exam E20-007

Data Science Associate Exam

Version: 7.0

[Total Questions: 198]

Question No : 1

You have been assigned to perform a study of the daily revenue effect of a pricing model of online transactions. When is the analytics lifecycle considered completed?

A.
When written documentation has been produced and the code has been handed off to the DBA/operations.
B.
When a model has been completely developed and the results have shown statistically acceptable results.
C.
When the results of the model have been presented to both the internal analytics team and the business owner of the project.
D.
When a model has been completely developed based on both a sample of the data and the entire set of data available.

Answer: A

Question No : 2

Which word or phrase completes the statement; “A theater actor is to ‘artistic and expressive’ as a data scientist is to ____________.”?

A.
Communicative and collaborative
B.
Introverted and technical
C.
Logical and steadfast
D.
Independent and intelligent

Answer: A

Question No : 3

You are provided four different datasets. Initial analysis on these datasets show that they have identical mean, variance and correlation values. What should your next step in the analysis be?

A. Visualize the data to further explore the characteristics of each data set
B. Select one of the four datasets and begin planning and building a model
C. Combine the data from all four of the datasets and begin planning and bulding a model
D. Recalculate the descriptive statistics since they are unlikely to be identical for each dataset

Answer: A

Question No : 4

What is a core deliverable at the end of the analytic project?

A. An implemented database design
B. A whitepaper describing the project and the implementation
C. A presentation for project sponsors
D. The training materials

Answer: C

Question No : 5

In which phase of the analytic lifecycle would you expect to spend most of the project time?

A. Discovery
B. Data preparation
C. Communicate Results
D. Operationalize

Answer: B

Question No : 6

What is LOESS used for?

A. It fits a smoothed curve to scatterplot data, to give a general sense of the data's behavior.
B. It is a significance test for the correlation between two variables.
C. It plots a continuous variable versus a discrete variable, to compare distributions across classes.
D. It is run after a one-way ANOVA, to determine which population has the highest mean value.

Answer: A

Question No : 7

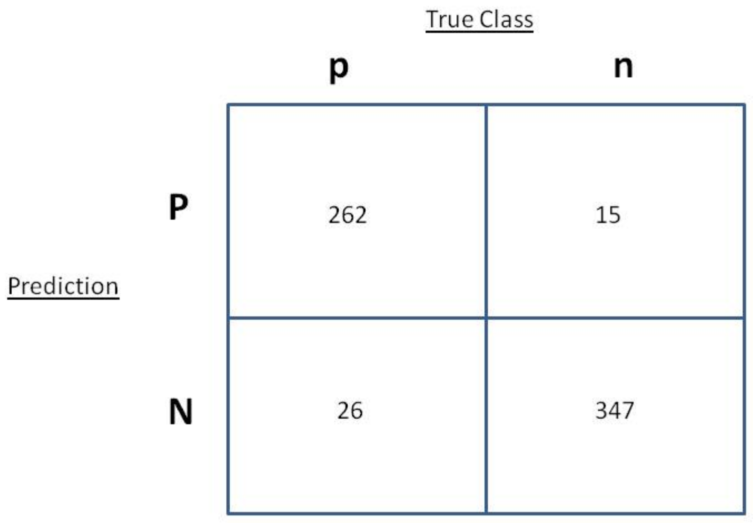
What is required in a presentation for business analysts?

A. Budgetary considerations and requests
B. Operational process changes
C. Detailed statistical explanation of the applicable modeling theory
D. The presentation author's credentials

Answer: B

Question No : 8

Refer to the exhibit.



True Class

		p	n
Prediction	P	262	15
	N	26	347

You have scored your Naive bayesian classifier model on a hold out test data for cross validation and determined the way the samples scored and tabluated them as shown in the exhibit.

What are the Precision and Recall rate of the model?

A. Precision = 262/277
Recall = 262/288
B. Precision =262/288
Recall = 262/277
C. Precision = 277/262
Recall = 288/262
D. Precision = 288/262
Recall = 277/262

Answer: A

Question No : 9

Which word or phrase completes the statement; “A data scientist would consider a RDBMS is to a table as R is to a ___________.”?

A.
Data frame
B.
List
C.
Matrix
D.
Array

Answer: A

Question No : 10

What is an appropriate data visualization to use in a presentation to a project sponsor?

A.
Bar Chart
B.
Pie Chart
C.
Box and Whisker Plot
D.
Density Plot

Answer: A

Question No : 11

Which word or phrase completes the statement? Business Intelligence is to monitoring trends as Data Science is to _________ trends.

A. Predicting
B. Discarding
C. Driving
D. Optimizing

Answer: A

Question No : 12

If your intention is to show trends over time, which chart type is the most appropriate way to depict the data?

A. Line chart
B. Bar chart
C. Stacked bar chart
D. Histogram

Answer: A

Question No : 13

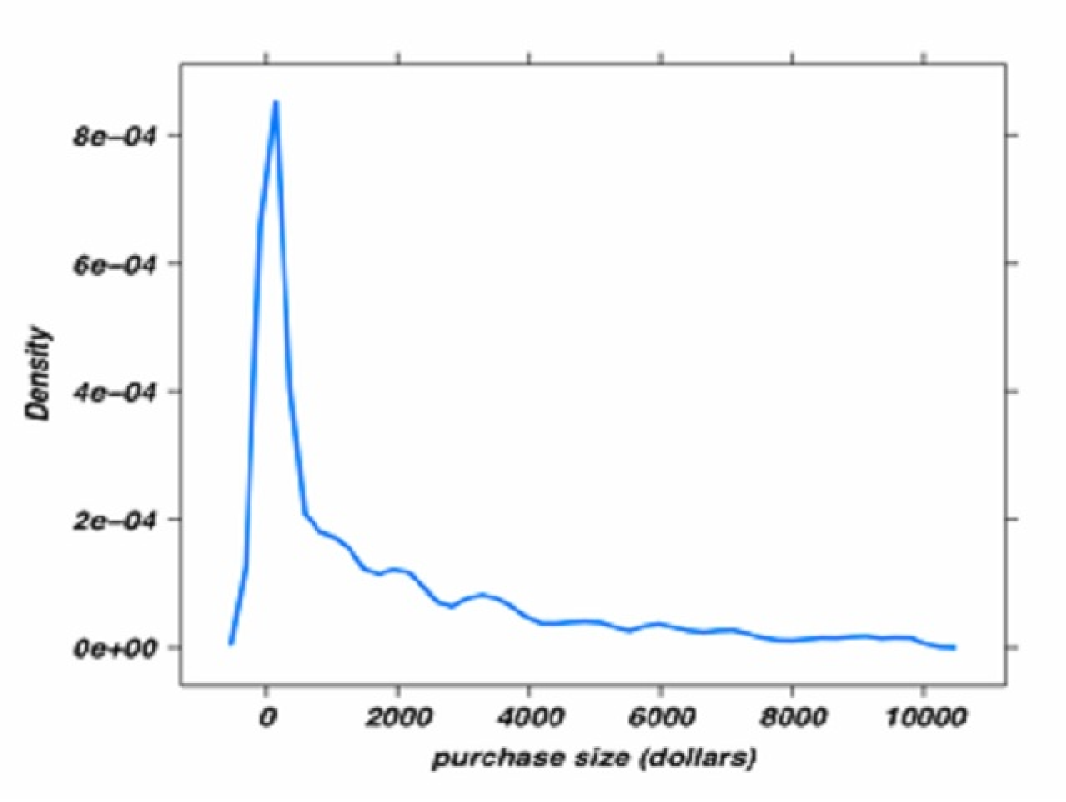
Which process in text analysis can be used to reduce dimensionality?

A. Stemming
B. Parsing
C. Digitizing
D. Sorting

Answer: A

Question No : 14

Refer to the exhibit.



You have created a density plot of purchase amounts from a retail website as shown. What should you do next?

A. Recreate the plot using the barplot() function
B. Use the rug() function to add elements to the plot
C. Recreate the density plot using a log normal distribution of the purchase amount data
D. Reduce the sample size of the purchase amount data used to create the plot

Answer: C

Question No : 15

Which key role for a successful analytic project can consult and advise the project team on the value of end results and how these will be used on a day-to-day basis?

A. Business User
B. Project Manager
C. Data Scientist
D. Business Intelligence Analyst

Answer: A

Question No : 16

To ensure a successful analytic project, which key role can provide business domain expertise with a deep understanding of the data and key performance indicators?

A.
Business Intelligence Analyst
B.
Project Manager
C.
Project Sponsor
D.
Business User

Answer: A

Question No : 17

You are performing a market basket analysis using the Apriori algorithm. Which measure is a ratio describing the how many more times two items are present together than would be expected if those two items are statistically independent?

A. Lift
B. Leverage
C. Support
D. Confidence

Answer: A

Question No : 18

Which word or phrase completes the statement? Data-ink ratio is to data visualization as ___________ .

A. Confusion matrix is to classifier

B. Data scientist is to big data
C. Seasonality is to ARIMA
D. K-means is to Naive Bayes

Answer: A

Question No : 19

You have been assigned to run a logistic regression model for each of 100 countries, and all the data is currently stored in a PostgreSQL database. Which tool/library would you use to produce these models with the least effort?

A. MADlib
B. Mahout
C. RStudio
D. HBase

Answer: A

Question No : 20

Your colleague, who is new to Hadoop, approaches you with a question. They want to know how best to access their data. This colleague has previously worked extensively with SQL and databases.

Which query interface would you recommend?

A. Hive
B. Pig
C. Howl
D. HBase

Answer: A

Question No : 21

Which functionality do regular expressions provide?