



E20-007^{Q&As}

Data Science and Big Data Analytics

Pass EMC E20-007 Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

<https://www.passapply.com/e20-007.html>

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by EMC
Official Exam Center

-  **Instant Download** After Purchase
-  **100% Money Back** Guarantee
-  **365 Days** Free Update
-  **800,000+** Satisfied Customers





QUESTION 1

Refer to exhibit.

Independent Variable	Coefficient	P-Value
A	0.45	0.0000
E	3.67	0.0000
C	1.23	0.0000

$$R^2 = 0.10$$

You are asked to write a report on how specific variables impact your client's sales using a data set provided to you by the client. The data includes 15 variables that the client views as directly related to sales, and you are restricted to these variables only.

After a preliminary analysis of the data, the following findings were made:

1.

Multicollinearity is not an issue among the variables

2.

Only three variables--A, B, and C--have significant correlation with sales

You build a linear regression model on the dependent variable of sales with the independent variables of A, B, and C. The results of the regression are seen in the exhibit.

You cannot request additional data. What is a way that you could try to increase the R^2 of the model without artificially inflating it?

- A. Create clusters based on the data and use them as model inputs
- B. Force all 15 variables into the model as independent variables
- C. Create interaction variables based only on variables A, B, and C
- D. Break variables A, B, and C into their own univariate models

Correct Answer: A

QUESTION 2

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

- A. Bar chart



- B. Pie chart
- C. Box and Whisker plot
- D. Density plot

Correct Answer: A

QUESTION 3

In MADlib what does MAD stand for?

- A. Magnetic, Agile, Deep
- B. Machine Learning, Algorithms for Databases
- C. Mathematical Algorithms for Databases
- D. Modular, Accurate, Dependable

Correct Answer: A

QUESTION 4

How does Pig's use of a schema differ from that of a traditional RDBMS?

- A. Pig's schema is optional
- B. Pig's schema requires that the data is physically present when the schema is defined
- C. Pig's schema is required for ETL
- D. Pig's schema supports a single data type

Correct Answer: A

QUESTION 5

What is the reason for using LOESS?

- A. Fits a smoothed curve to scatterplot data; providing a general idea of the data's behavior
- B. Significance test for the correlation between two variables
- C. Plots a continuous variable versus a discrete variable; comparing distributions across classes
- D. Runs after a one-way ANOVA; determining which population has the highest mean value

Correct Answer: A



VCE & PDF

PassApply.com

<https://www.passapply.com/e20-007.html>

2024 Latest passapply E20-007 PDF and VCE dumps Download

[Latest E20-007 Dumps](#)

[E20-007 VCE Dumps](#)

[E20-007 Practice Test](#)