



# DATABRICKS-CERTIFIED- PR OFESIONAL-DATA-SCIENTIST<sup>Q&As</sup>

Databricks Certified Professional Data Scientist Exam

**Pass Databricks DATABRICKS-CERTIFIED-  
PROFESSIONAL-DATA-SCIENTIST Exam with 100%  
Guarantee**

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

<https://www.passapply.com/databricks-certified-professional-data-scientist.html>

100% Passing Guarantee  
100% Money Back Assurance

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



VCE & PDF

PassApply.com

<https://www.passapply.com/databricks-certified-professional-data-scientist>.  
2024 Latest passapply DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST PDF and VCE dumps Download

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





### QUESTION 1

$y_1, y_2, y_3, \dots, y_{n-1}, y_n$

May have a trend component that is quadratic in nature. Which pattern of data will indicate that the trend in the time series data is quadratic in nature?

- A. Naive Bayesian classifier
- B. Decision tree
- C. Linear regression
- D. K-means clustering

Correct Answer: D

Explanation: kmeans uses an iterative algorithm that minimizes the sum of distances from each object to its cluster centroid, over all clusters. This algorithm moves objects between clusters until the sum cannot be decreased further. The result is a set of clusters that are as compact and well-separated as possible. You can control the details of the minimization using several optional input parameters to kmeans, including ones for the initial values of the cluster centroids, and for the maximum number of iterations. Clustering is primarily an exploratory technique to discover hidden structures of the data, possibly as a prelude to more focused analysis or decision processes. Some specific applications of k-means are image processing, medical and customer segmentation. Clustering is often used as a lead-in to classification. Once the clusters are identified, labels can be applied to each cluster to classify each group based on its characteristics. Marketing and sales groups use k-means to better identify customers who have similar behaviors and spending patterns.

---

### QUESTION 2

Classification and regression are examples of \_\_\_\_\_.

- A. supervised learning
- B. un-supervised learning
- C. Clustering
- D. Density estimation

Correct Answer: A

Explanation: In classification, our job is to predict what class an instance of data should fall into. Another task in machine learning is regression. Regression is the prediction of a numeric value. Most people have probably seen an example of regression with a best-fit line drawn through some data points to generalize the data points. Classification and regression are examples of supervised learning. This set of problems is known as supervised because we're telling the algorithm what to predict.

---

### QUESTION 3



Let's say you have two cases as below for the movie ratings

1.

You recommend to a user a movie with four stars and he really doesn't like it and he'd rate it two stars

2.

You recommend a movie with three stars but the user loves it (he'd rate it five stars). So which statement correctly applies?

- A. In both cases, the contribution to the RMSE is the same
- B. In both cases, the contribution to the RMSE is the different
- C. In both cases, the contribution to the RMSE, could varies
- D. None of the above

Correct Answer: A

---

#### QUESTION 4

You are creating a Classification process where input is the income, education and current debt of a customer, what could be the possible output of this process?

- A. Probability of the customer default on loan repayment
- B. Percentage of the customer loan repayment capability
- C. Percentage of the customer should be given loan or not
- D. The output might be a risk class, such as "good", "acceptable", "average", or "unacceptable".

Correct Answer: D

Explanation: Classification is the process of using several inputs to produce one or more outputs. For example the input might be the income, education and current debt of a customer The output might be a risk class, such as "good", "acceptable", "average", or "unacceptable". Contrast this to regression where the output is a number not a class.

---

#### QUESTION 5

Which of the following could be features?

- A. Words in the document
- B. Symptoms of a diseases
- C. Characteristics of an unidentified object
- D. Only 1 and 2
- E. All 1,2 and 3 are possible



Correct Answer: E

Explanation: Any dataset that can be turned into lists of features. A feature is simply something that is either present or absent for a given item. In the case of documents, the features are the words in the document but they could also be characteristics of an unidentified object symptoms of a disease, or anything else that can be said to be present or absent.

[DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST Practice Test](#)

[DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST Study Guide](#)

[DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST Exam Questions](#)