



DATABRICKS-CERTIFIED- PR OFESIONAL-DATA-SCIENTIST^{Q&As}

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.geekcert.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

GeekCert.com

<https://www.geekcert.com/databricks-certified-professional-data-scientist.ht>
2024 Latest geekcert 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

In which lifecycle stage are appropriate analytical techniques determined?

- A. Model planning
- B. Model building
- C. Data preparation
- D. Discovery

Correct Answer: A

Explanation: In Phase 3, the data science team identifies candidate models to apply to the data for clustering, classifying, or finding relationships in the data depending on the goal of the project. It is during this phase that the team refers to the hypotheses developed in Phase 1, when they first became acquainted with the data and understanding the business problems or domain area. These hypotheses help the team frame the analytics to execute in Phase 4 and select the right methods to achieve its objectives. Some of the activities to consider in this phase include the following: Assess the structure of the datasets. The structure of the datasets is one factor that dictates the tools and analytical techniques for the next phase. Depending on whether the team plans to analyze textual data or transactional data, for example, different tools and approaches are required. Ensure that the analytical techniques enable the team to meet the business objectives and accept or reject the working hypotheses. Determine if the situation warrants a single model or a series of techniques as part of a larger analytic workflow. A few example models include association rules and logistic regression. Other tools, such as Alpine Miner, enable users to set up a series of steps and analyses and can serve as a front-end user interface (UI) for manipulating Big Data sources in PostgreSQL.

QUESTION 2

In which of the scenarios can you use the linear regression model?

- A. Predicting Home Price based on the location and house area
- B. Predicting demand of the goods and services based on the weather
- C. Predicting tumor size reduction based on input as number of radiation treatment
- D. Predicting sales of the text book based on the number of students in state

Correct Answer: ABCD

Explanation: : You can use the linear regression model for predicting the continuous output variable based on the input variables. In all the cases mentioned in the question options, you can see that output can be predicted based on the input variable. Option-A: Input: Location, House Area and Output: House Price Option-B : Input: Weather condition, Output: Demand for the goods and services Option-C : Input: Number of Radiation Session Output: Tumor Size Reduction Option-D : Input: Number of students and Output: Sale quantity of text book

QUESTION 3

A data scientist is asked to implement an article recommendation feature for an on-line magazine.



The magazine does not want to use client tracking technologies such as cookies or reading history. Therefore, only the style and subject matter of the current article is available for making recommendations. All of the magazine's articles are stored in a database in a format suitable for analytics.

Which method should the data scientist try first?

- A. K Means Clustering
- B. Naive Bayesian
- C. Logistic Regression
- D. Association Rules

Correct Answer: A

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 4

In which of the scenario you can use the regression to predict the values?

- A. Samsung can use it for mobile sales forecast
- B. Mobile companies can use it to forecast manufacturing defects
- C. Probability of the celebrity divorce
- D. Only 1 and 2
- E. All 1 ,2 and 3

Correct Answer: E

Explanation: Regression is a tool which Companies may use this for things such as sales forecasts or forecasting manufacturing defects. Another creative example is predicting the probability of celebrity divorce.

QUESTION 5

- A. 2.4
- B. 24 0
- C. .24



D. .48

E. 4.8

Correct Answer: C

Explanation: Given no additional information, the MLE for the probability of an item in the test set is exactly its frequency in the training set. The method of maximum likelihood corresponds to many well-known estimation methods in statistics.

For example, one may be interested in the heights of adult female penguins, but be unable to measure the height of every single penguin in a population due to cost or time constraints. Assuming that the heights are normally (Gaussian)

distributed with some unknown mean and variance, the mean and variance can be estimated with MLE while only knowing the heights of some sample of the overall population. MLE would accomplish this by taking the mean and variance as

parameters and finding particular parametric values that make the observed results the most probable (given the model).

In general, for a fixed set of data and underlying statistical model the method of maximum likelihood selects the set of values of the model parameters that maximizes the likelihood function. Intuitively, this maximizes the "agreement" of the

selected model with the observed data, and for discrete random variables it indeed maximizes the probability of the observed data under the resulting distribution. Maximum-likelihood estimation gives a unified approach to estimation, which is

well-defined in the case of the normal distribution and many other problems. However in some complicated problems, difficulties do occur: in such problems, maximum-likelihood estimators are unsuitable or do not exist.

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

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

[DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST Braindumps](#)