



DS-200^{Q&As}

Data Science Essentials

Pass Cloudera DS-200 Exam with 100% Guarantee

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

<https://www.pass4itsure.com/ds-200.html>

100% Passing Guarantee
100% Money Back Assurance

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

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



**QUESTION 1**

Consider the following sample from a distribution that contains a continuous X and label Y that is either A or B:

X	Y
1	A
2	A
3	A
4	B
5	A
6	B
7	A
8	B
9	B
10	B

Which is the best cut point for X if you want to discretize these values into two buckets in a way that minimizes the sum of chi-square values?

- A. X 8
- B. X 6
- C. X 5
- D. X 4
- E. X 2

Correct Answer: D

QUESTION 2

A company has 20 software engineers working to fix on a project. Over the past week, the team has fixed 100 bugs. Although the average number of bugs. Although the average number of bugs fixed per engineer id five. None of the engineer fixed exactly five bugs last week.

You want to understand how productive each engineer is at fixing bugs. What is the best way to visualize the distribution of bug fixes per engineer?

- A. A bar chart of engineers vs. number of bugs fixed
- B. A scatter plot of engineers vs. number of bugs fixed
- C. A normal distribution of the mean and standard deviation of bug fixes per engineer
- D. A histogram that groups



engineers to together based on the number of bugs they fixed

Correct Answer: A

QUESTION 3

How can the naiveté of the naive Bayes classifier be advantageous?

- A. It does not require you to make strong assumptions about the data because it is a non- parametric
- B. It significantly reduces the size of the parameter space, thus reducing the risk of over fitting
- C. It allows you to reduce bias with no tradeoff in variance
- D. It guarantees convergence of the estimator

Correct Answer: A

QUESTION 4

Given the following sample of numbers from a distribution: 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89

What are two benefits of using the five-number summary of sample percentiles to summarize a data set?

- A. You can calculate unbiased estimators for the parameters of the distribution
- B. It's robust to outliers
- C. It's well-defined for any probability distribution
- D. You can calculate it quickly using a relational database like MySQL, even when we have a large sample

Correct Answer: D

QUESTION 5

Which best describes the primary function of Flume?

- A. Flume is a platform for analyzing large data sets that consists of a high-level language for expressing data analysis programs, coupled with an infrastructure consisting of sources and sinks for importing and evaluating large data sets
- B. Flume acts as a Hadoop filesystem for log files
- C. Flume Imports data from SQL/relational database into your Hadoop cluster



D. Flume provides a query languages for Hadoop similar to SQL

E. Flume is a distributed server for collecting and moving large amount of data into HDFS as it\\'s produced from streaming data flows

Correct Answer: D

[DS-200 PDF Dumps](#)

[DS-200 VCE Dumps](#)

[DS-200 Study Guide](#)