Free Questions for DA0-001

Shared by Welch on 06-06-2022

For More Free Questions and Preparation Resources

Check the Links on Last Page

Question 1

Question Type: MultipleChoice

A financial analyst is creating a daily billing report for a company. One night, the company's data warehouse did not update the data, which caused the data to be reported incorrectly the next day. Which of the following documentation elements should the analyst add to catch this error?

Options:

A- Version number

B- Data refresh

C- Frequently asked questions tab

D- Summary

Answer:

B

Explanation:

A data refresh is a documentation element that indicates when the data was last updated or refreshed from the source. A data refresh can help the analyst to catch the error of the data warehouse not updating the data, as it will show a discrepancy between the expected and actual date of the data update.A data refresh can also help the users of the report to verify the timeliness and accuracy of the data, and to avoid making decisions based on outdated or incorrect data

Question 2

Question Type: MultipleChoice

A customer's telephone number is in the format 123-456-7890. Which of the following data types is used for the phone number?

Options:

A- Boolean

B- Date

C- Text
D- Number

Answer:

C

Question 3

Question Type: MultipleChoice

Which of the following is an example of a flat file?

Options:

A- CSV file
B- PDF file
C- JSON file
D- JPEG file

Answer:

A

Explanation:

A CSV file is a type of flat file that stores data as plain text in a table-like structure with rows and columns. Each row represents a single record, while columns represent fields or attributes of the dat

a. A CSV file uses commas or other delimiters to separate the values in each row.A CSV file can be easily imported or exported by various applications and programs12

Question 4

Question Type: MultipleChoice

An analyst develops an IT document and needs to describe the technical terms used in the document. Which of the following is where the analyst should include descriptions of the technical terms?

Options:

A- Glossary

B- System diagram

C- User requirements

D- Index

Answer:

A

Question 5

Question Type: MultipleChoice

Which of the following tools would be best to use to calculate the interquartile range, median, mean, and standard deviation of a column in a table that has 5.000.000 rows?

Options:

A- Microsoft Excel

B- R

C- Snowflake

D- SQL

Answer:

B

Question 6

Question Type: MultipleChoice

A data analyst needs to collect a similar proportion of data from every state. Which of the following sampling methods would be the most appropriate?

Options:

A- Systematic sampling

B- Convenience sampling

C- Stratified sampling

D- Random sampling

Answer:

C

Explanation:

The best sampling method for the data analyst's need is C. Stratified sampling.

Stratified sampling is a type of probability sampling that involves dividing the population into homogeneous groups or strata based on some characteristic, such as state, and then randomly selecting a proportional number of individuals from each stratum. Stratified sampling ensures that every group is adequately represented in the sample, and reduces the sampling error and variability12

Systematic sampling is not correct, because it involves selecting every nth individual from the population, starting from a random point. Systematic sampling does not guarantee that every state will have a similar proportion of data in the sample, and may introduce bias or error if there is a hidden pattern or order in the population12

Convenience sampling is not correct, because it involves selecting individuals who are easily accessible or available to the researcher. Convenience sampling is a type of non-probability sampling that does not involve random selection, and may result in a biased or unrepresentative sample12

Random sampling is not correct, because it involves selecting individuals from the population at random, without any grouping or stratification. Random sampling may not produce a sample that has a similar proportion of data from every state, especially if the population is large or heterogeneous. Random sampling may also have a higher sampling error and variability than stratified sampling12

Question 7

Question Type: MultipleChoice

Which of the following is the most likely reason for a data analyst to optimize a query using parameterization?

Options:

A- To return a subset of records

B- To insert a temporary table

C- To prevent SQL injections

D- To increase the query speed

Answer:

C

Question 8

Question Type: MultipleChoice

A company's human resources department has asked a data analyst to categorize the income of all employees into five salary bands:

Employee_ID	Salary	Salary_band
003	$130,000	
014	$120,000	
004	$110,000	
013	$90,000	
002	$140,000	
012	$122,000	
016	$132,000	
006	$70,000	
017	$53,000	
009	$111,000	
019	$107,000	
008	$111,000	
018	$50,000	

Which of the following types of functions would be the most appropriate to use?

Options:

A- Statistical
B- Aggregate
C- Logical
D- Mathematical

Answer:

C

Explanation:

Short explanation: Logical functions are the most appropriate to use for categorizing data into bands, because they allow the data analyst to apply conditional statements and criteria to the data values. For example, the IF function can be used to assign a band name based on whether a value meets a certain condition or not.Other logical functions that can be useful for categorizing data are AND, OR, NOT, and IFERROR12

Question 9

Question Type: MultipleChoice

A data analyst has received a data set that contains actual and projected sales for the fourth quarter of 2019. Which of the following statistical methods should the analyst use to find the measure of dispersion?

Options:

A- Mean
B- Variance
C- Correlation
D- Confidence interval

Answer:

B

Question 10

Question Type: MultipleChoice

Given the following data table:

CandidateID	Status	Date	HireDate
01	Hired	05-23-87	05-23-87
02	Hired	11-30-96	11-30-96
03	Hired	13-05-99	13-05-99

Which of the following are appropriate reasons to undertake data cleansing? (Select two).

Options:

A- Non-parametric data

B- Missing data

C- Duplicate data

D- Invalid data

E- Redundant data

F- Normalized data

Answer:

D, E

Question 11

Question Type: MultipleChoice

Which of the following best describes the law of large numbers?

Options:

A- As a sample size decreases, its standard deviation gets closer to the average of the whole population.

B- As a sample size grows, its mean gets closer to the average of the whole population

C- As a sample size decreases, its mean gets closer to the average of the whole population.

D- When a sample size doubles. the sample is indicative of the whole population.

Answer:

B

Explanation:

The best answer is B. As a sample size grows, its mean gets closer to the average of the whole population.

The law of large numbers, in probability and statistics, states that as a sample size grows, its mean gets closer to the average of the whole population. This is due to the sample being more representative of the population as it increases in size. The law of large numbers guarantees stable long-term results for the averages of some random events1

A) As a sample size decreases, its standard deviation gets closer to the average of the whole population is not correct, because it confuses the concepts of standard deviation and mean. Standard deviation is a measure of how much the values in a data set vary from the mean, not how close the mean is to the population average. Also, as a sample size decreases, its standard deviation tends to increase, not decrease, because the sample becomes less representative of the population.

C) As a sample size decreases, its mean gets closer to the average of the whole population is not correct, because it contradicts the law of large numbers. As a sample size decreases, its mean tends to deviate from the average of the whole population, because the sample becomes less representative of the population.

D) When a sample size doubles, the sample is indicative of the whole population is not correct, because it does not specify how close the sample mean is to the population average. Doubling the sample size does not necessarily make the sample indicative of the whole population, unless the sample size is large enough to begin with. The law of large numbers does not state a specific number or proportion of samples that are indicative of the whole population, but rather describes how the sample mean approaches the population average as the sample size increases indefinitely.

Question 12

Question Type: MultipleChoice

A customer survey reveals 90% positive feedback. Which of the following statistical methods would be best to utilize to determine the reliability of a data set and predict how a larger sample of customers over the same time period might respond?

Options:

A- Calculate a high variance on survey responses.

B- Calculate the maximum range of the survey responses.

C- Calculate a low standard deviation on survey responses.

D- Remove any data more than 4 standard deviation from the mean.

Answer:

C

To Get Premium Files for DA0-001 Visit
https://www.p2pexams.com/products/da0-001

For More Free Questions Visit
https://www.p2pexams.com/comptia/pdf/da0-001