Free Questions for DA0-001 by dumpshq

Shared by Stephenson on 29-01-2024

For More Free Questions and Preparation Resources

Check the Links on Last Page

Question 1

Question Type: MultipleChoice

An analyst modified a data set that had a number of issues. Given the original and modified versions:

Original data:

Var001	Var002	Var003	Var004
1	0	0	0
0	1	0	1
1	1	1	2
0	0	0	1

Modified data:

Var001	Var002	Var003	Var004
Yes	Absent	No payment	No
No	Present	No payment	Yes
Yes	Present	Payment	Maybe
No	Absent	No payment	Yes

Which of the following data manipulation techniques did the analyst use?

Options:

A- Imputation

B- Recoding

C- Parsing

D- Deriving

Answer:

B

Explanation:

The correct answer is B. Recoding.

Recoding is a data manipulation technique that involves changing the values or categories of a variable to make it more suitable for analysis. Recoding can be used to simplify or group the data, to correct errors or inconsistencies, or to create new variables from existing ones12

In the example, the analyst used recoding to change the values of Var001, Var002, Var003, and Var004 from numerical to textual form. The analyst also used recoding to assign meaningful labels to the values, such as "Absent" for 0, "Present" for 1, "Low" for 2, "Medium" for 3, and "High" for 4. This makes the data more understandable and easier to analyze.

Question 2

Question Type: MultipleChoice

Which of the following would a data analyst look for first if 100% participation is needed on survey results?

Options:

A- Missing data

B- Invalid data

C- Redundant data

D- Duplicate data

Answer:

A

Explanation:

Missing data is a type of data quality issue that occurs when some values in a data set are not recorded or available. Missing data can affect the validity and reliability of survey results, especially if the missing values are not random or ignorable.Missing data can also reduce the sample size and the statistical power of the analysis12

If 100% participation is needed on survey results, a data analyst would look for missing data first, because missing data would indicate that some participants did not complete or submit the survey, or that some responses were not recorded or transmitted correctly.A data analyst would need to identify the causes and patterns of missing data, and apply appropriate methods to handle or prevent missing data, such as imputation, deletion, weighting, or follow-up12

Question 3

Question Type: MultipleChoice

A user imports a data file into the accounts payable system each day. On a regular basis. the field input is not what the system is expecting. so it results in an error for the row and a broken import process. To resolve the issue, the user opens the file, finds the error in the row, and manually corrects it before attempting the import again. The import sometimes breaks on subsequent attempts. though. Which of the following changes should be made to this process to reduce the number of errors?

Options:

A- Delete all incorrect inputs and upload the corrected file.

B- Have the user manually review the file for data completeness before loading it

C- Create a data field to data type validator to run the file through prior to import.

D- Spot-check the file prior to import to catch and correct field errors.

Answer:

C

Explanation:

A data field to data type validator is a tool or a process that checks if the data in each field of a file matches the expected data type, such as text, number, date, etc. A data field to data type validator can help to identify and correct any errors or inconsistencies in the data before importing it into the accounts payable system. This would reduce the number of errors and broken imports, as well as save time and effort for the user.

Question 4

Question Type: MultipleChoice

An analyst is currently working on a ticket for revamping a company-wide dashboard that has been in use for five years. Which of the following should be the first step in the development process?

Options:

A- Talk to the group that made the request to determine the desired goal.

B- Make changes to a frequently used report that is already in production.

C- Build an additional dashboard with fewer views that are tailored toward each specific team.

D- Develop a more streanMined dashboard to roll out by the next delivery date.

Answer:

A

Explanation:

The first step in the development process of revamping a company-wide dashboard should be to talk to the group that made the request to determine the desired goal. This would help to understand the needs, expectations, and preferences of the stakeholders, as well as the scope, purpose, and objectives of the project. Talking to the group that made the request would also help to establish a clear communication channel, build rapport and trust, and solicit feedback and suggestions.

Question 5

Question Type: MultipleChoice

A data analyst needs to perform a full outer join of a customer's orders using the tables below:

Sales_table

Cust_id	Order_id	Order_qty
Tc - 5858	Od - 9800	50
Tc - 5833	Od - 9801	68
Tc - 5890	Od - 9802	103

Order_table

Order_id	Order_qty
Od - 9803	102
Od - 9800	50
Od - 9802	103
Od - 9805	80
Od - 9804	70

Which of the following is the mean of the order quantity?

Options:

A- 73.5

B- 76.5

C- 78.8

D- 81.5

Answer:

D

Explanation:

The correct answer is D. OUTER JOIN, seven rows.

An OUTER JOIN is a type of SQL join that returns all the rows from both tables, regardless of whether there is a match or not. If there is no match, the missing side will have null values. An OUTER JOIN can be either a LEFT JOIN, a RIGHT JOIN, or a FULL JOIN, depending on which table's rows are preserved1

Using the example tables, a FULL OUTER JOIN query would look like this:

SELECT Cust_id, Order_id, Order_qty FROM Sales_table FULL OUTER JOIN Order_table ON Sales_table.Order_id = Order_table.Order_id;

The result of this query would be:

Cust_id | Order_id | Order_qty ---------- 1 | 1 | 100 2 | 2 | 50 3 | 3 | 25 4 | 4 | 75 NULL | 5 | 10 NULL | 6 | 20 NULL | 7 | 15

As you can see, the query returns seven rows, one for each order in either table. The orders that are not in the Sales_table have null values for the Cust_id column.

To find the mean of the order quantity, we need to sum up the order quantities and divide by the number of rows. In this case, the mean is (100 + 50 + 25 + 75 + 10 + 20 + 15) / 7 = 42.14. Rounding to one decimal place, we get 42.1 as the mean of the order quantity.

Question 6

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.