CUET-UG Information Practices Sample Paper-1

Duration: 1 Hour **Maximum Marks: 250**

Instructions

- This paper contains a total of 50 Multiple Choice Questions.
- Each correct answer carries **+5 marks**.
- Each incorrect answer carries **-1 mark**.
- No negative marking for unattempted questions.

Q1. Consider the SQL command: `SELECT ROUND(15.193, 1), TRUNCATE(15.193, 1);`. What will be the output?

(A) 15.2, 15.1

(B) 15.1, 15.1

(C) 15.2, 15.2

(D) 15.1, 15.2

Q2. Which SQL function is used to return the position of the first occurrence of a substring within a string, and what does it return if the substring is not found?

(A) `SEARCH()`; returns NULL

(B) `LOCATE()`; returns 0

(C) `INSTR()`; returns 0

(D) `SUBSTR()`; returns -1

Q3. A table 'Employee' has a column 'DOB' (Date of Birth). A user wants to display the name of the day (e.g., Monday, Tuesday) on which each employee was born. Which function is most appropriate?

(A) `DAYNAME(DOB)`

(B) `DAYOFWEEK(DOB)`

(C) `DATE(DOB)`

(D) `MONTHNAME(DOB)`

Q4. What will be the result of the following SQL query?

```
SELECT LTRIM(" CUET 2026 "), RTRIM(" CUET 2026 ");
```

(A) ”CUET 2026”, ”CUET 2026”

(B) ”CUET 2026 ”, ” CUET 2026”

(C) ” CUET 2026”, ”CUET 2026 ”

(D) ”CUET2026”, ”CUET2026”

Q5. Match the SQL Functions in List I with their descriptions in List II:

List I	Function	List II	Description
(i)	MOD(x, y)	(1)	Returns the remainder of x divided by y
(ii)	LENGTH(s)	(2)	Returns the number of characters in a string
(iii)	MID(s, n, m)	(3)	Returns a substring starting from n of length m
(iv)	NOW()	(4)	Returns the current system date and time

(A) i-1, ii-2, iii-3, iv-4

(B) i-2, ii-1, iii-3, iv-4

(C) i-3, ii-4, iii-1, iv-2

(D) i-4, ii-3, iii-2, iv-1

Q6. Identify the output of the following SQL statement:

```
SELECT POWER(INSTR("PANDAS", "A"), 3);
```

(A) 8

(B) 27

(C) 1

(D) 64

Q7. Which of the following expressions will return the current year from the system date in MySQL?

(A) `SELECT YEAR(CURDATE());`

(B) `SELECT EXTRACT(YEAR FROM NOW());`

(C) `SELECT YEAR(SYSDATE());`

(D) All of the above

Q8. Match the Mathematical Results in List I with the SQL Expressions in List II:

List I	Result	List II	SQL Expression
(i)	100	(1)	`ROUND(123.45, -2)`
(ii)	123	(2)	`TRUNCATE(123.99, 0)`
(iii)	-1	(3)	`SIGN(-25)`
(iv)	5	(4)	`LENGTH("Informatics") - 6`

(A) i-1, ii-2, iii-3, iv-4

(B) i-2, ii-1, iii-4, iv-3

(C) i-1, ii-3, iii-2, iv-4

(D) i-3, ii-2, iii-1, iv-4

Q9. A table 'Employee' has attributes (EmpID, Name, Salary, DeptID). If a query is executed as $\sigma_{Salary>50000}(Employee)$, which Relational Algebra operation is being performed?

(A) Projection

(B) Selection

(C) Join

(D) Intersection

Q10. In a Relational Database Management System, a candidate key that is not chosen as the primary key is specifically known as a/an:

(A) Foreign Key

(B) Secondary Key

(C) Alternate Key

(D) Composite Key

(E) Super Key

Q11. Which property of a database transaction ensures that either the entire transaction occurs or nothing occurs at all?

(A) Consistency

(B) Isolation

(C) Durability

(D) Atomicity

Q12. Consider a relation R(A, B, C). The operation $\Pi_{A,C}(R)$ will result in a relation containing:

(A) Only rows where A and C are not null

(B) Only columns A and C from the relation R

(C) All columns where A and C are keys

(D) Only rows that are common to A and C

Q13. Which network topology requires a central controller or a hub to connect all the nodes in the network?

(A) Mesh Topology

(B) Bus Topology

(C) Star Topology

(D) Tree Topology

Q14. What is the size of an IPv4 address and an IPv6 address respectively?

(A) 32 bits, 64 bits

(B) 32 bits, 128 bits

(C) 64 bits, 128 bits

(D) 48 bits, 128 bits

Q15. Which hardware device operates at the Data Link Layer of the OSI model and uses MAC addresses to forward data to specific ports?

(A) Hub

(B) Repeater

(C) Switch

(D) Gateway

Q16. Given below are Two statements, Choose Correct One:
Statement 1: A Pandas Series can store only homogeneous data types.
Statement 2: A Pandas DataFrame is a 2D labeled data structure with columns of potentially different types.

(A) Both Statement 1 and Statement 2 are correct

(B) Both Statement 1 and Statement 2 are incorrect

(C) Statement 1 is correct but Statement 2 is incorrect

(D) Statement 1 is incorrect but Statement 2 is correct

Q17. What will be the output of the following code?

```
import pandas as pd
s = pd.Series([10, 20, 30], index=['a', 'b', 'c'])
print(s['a':'b'])
```

(A) 10

(B) 10, 20

(C) 10, 20, 30

(D) Error

Q18. Which attribute of a Pandas DataFrame is used to check the dimensions (rows, columns) of the object?

(A) .size

(B) .shape

(C) .ndim

(D) .index

Q19. Given below are Two statements, Choose Correct One:
Statement 1: The `loc` method in Pandas is label-based slicing.
Statement 2: The `iloc` method includes the start index but excludes the stop index.

(A) Both Statement 1 and Statement 2 are correct

(B) Both Statement 1 and Statement 2 are incorrect

(C) Statement 1 is correct but Statement 2 is incorrect

(D) Statement 1 is incorrect but Statement 2 is correct

Q20. Which method is used to export a DataFrame object named `df` to a comma-separated file named 'data.csv'?

(A) `df.save_csv('data.csv')`

(B) `df.to_csv('data.csv')`

(C) `df.write_csv('data.csv')`

(D) `df.csv_export('data.csv')`

Q21. If a DataFrame `df` has 5 rows and 3 columns, what will `df.size` return?

(A) 5

(B) 3

(C) 15

(D) 8

Q22. To display the last 3 rows of a DataFrame `df`, which command is correct?

(A) `df.head(-3)`

(B) `df.tail(3)`

(C) `df.last(3)`

(D) `df.bottom(3)`

Q23. Given below are Two statements, Choose Correct One:
Statement 1: Missing data in Pandas is represented by NaN (Not a Number).
Statement 2: The `dropna()` method is used to fill missing values with a specific constant.

(A) Both Statement 1 and Statement 2 are correct

(B) Both Statement 1 and Statement 2 are incorrect

(C) Statement 1 is correct but Statement 2 is incorrect

(D) Statement 1 is incorrect but Statement 2 is correct

Q24. Match the SQL Functions in List I with their categories in List II:

List I	Function	List II	Category
(i)	COUNT()	(1)	Scalar Function
(ii)	MAX()	(2)	Aggregate Function
(iii)	ROUND()	(3)	String Function
(iv)	LENGTH()	(4)	Numeric Function

(A) i-2, ii-2, iii-4, iv-3

(B) i-1, ii-1, iii-3, iv-4

(C) i-2, ii-1, iii-4, iv-3

(D) i-4, ii-3, iii-2, iv-1

Q25. Which SQL clause is used to filter the results produced by the GROUP BY clause?

(A) WHERE

(B) HAVING

(C) ORDER BY

(D) LIMIT

Q26. In SQL, which join returns all records when there is a match in either left or right table records?

(A) INNER JOIN

(B) LEFT JOIN

(C) RIGHT JOIN

(D) FULL OUTER JOIN

Q27. The default sort order for the ORDER BY clause in SQL is:

(A) DESC

(B) ASC

(C) RANDOM

(D) NO ORDER

Q28. Match the SQL Join types in List I with their descriptions in List II:

List I	Join Type	List II	Description
(i)	Equi Join	(1)	Cartesian Product
(ii)	Natural Join	(2)	Comparison using '='
(iii)	Cross Join	(3)	Join on same column names without duplicates

(A) i-2, ii-3, iii-1

(B) i-1, ii-2, iii-3

(C) i-3, ii-1, iii-2

(D) i-2, ii-1, iii-3

Q29. Which function is used in Matplotlib to add a title to the plot?

(A) `plt.set_title()`

(B) `plt.title()`

(C) `plt.head()`

(D) `plt.label_title()`

Q30. In `plt.bar(x, y)`, what does the 'x' parameter represent?

(A) Frequency of the data

(B) Height of the bars

(C) Categories or positions on the x-axis

(D) Color of the bars

Q31. Which of the following is used to display the legend in a Matplotlib chart?

(A) `plt.show_legend()`

(B) `plt.legend()`

(C) `plt.display(legend=True)`

(D) `plt.plot(legend=True)`

Q32. To plot a histogram in Matplotlib, which function is used?

(A) `plt.histogram()`

(B) `plt.hist()`

(C) `plt.bar_hist()`

(D) `plt.plot_hist()`

Q33. What is the primary purpose of 'E-waste' management?

(A) To increase the sale of electronic goods

(B) To safely dispose or recycle old electronic devices

(C) To track the location of discarded computers

(D) To create more hardware components

Q34. Which term refers to the trail of data you leave behind while using the internet?

(A) Digital Cookie

(B) Digital Footprint

(C) Cyber Trail

(D) Web Shadow

Q35. Stealing someone else's intellectual work and representing it as your own is called:

(A) Phishing

(B) Spamming

(C) Plagiarism

(D) Hacking

Q36. Which of the following falls under 'Intellectual Property Rights' (IPR)?

(A) Patents

(B) Copyrights

(C) Trademarks

(D) All of the above

Q37. A DataFrame `df` has a column 'Sales'. Which command shows the average sales?

(A) `df['Sales'].average()`

(B) `df['Sales'].mean()`

(C) `df['Sales'].sum() / 10`

(D) `df.mean('Sales')`

Q38. In Matplotlib, what does the 'bins' parameter in a histogram define?

(A) The color of the bars

(B) The number of intervals for the data distribution

(C) The width of each individual bar

(D) The labels on the y-axis

Q39. Given below are Two statements, Choose Correct One:
Statement 1: The `rename()` function in Pandas allows changing row indexes and column labels.
Statement 2: By default, `rename()` modifies the original DataFrame without needing `inplace=True`.

(A) Both Statement 1 and Statement 2 are correct

(B) Both Statement 1 and Statement 2 are incorrect

(C) Statement 1 is correct but Statement 2 is incorrect

(D) Statement 1 is incorrect but Statement 2 is correct

Q40. Which command will correctly drop a column named 'Address' from a DataFrame `df` permanently?

(A) `df.drop('Address', axis=0)`

(B) `df.drop('Address', axis=1, inplace=True)`

(C) `df.remove('Address')`

(D) `df.drop(column='Address')`

Q41. Case Study: A school uses a DataFrame `df_stud` to store (RollNo, Name, Marks). If the school wants to see the top 5 performers, which command is most appropriate?

(A) `df_stud.sort_values('Marks').head(5)`

(B) `df_stud.sort_values('Marks', ascending=False).head(5)`

(C) `df_stud.max(5)`

(D) `df_stud.head(5).sort()`

Q42. Which SQL function is used to return the current date and time?

(A) `DATE()`

(B) `GETDATE()`

(C) `NOW()`

(D) `TIME()`

Q43. In SQL, if you want to find the total number of unique departments in an 'Employees' table, the query would be:

(A) `SELECT COUNT(DeptID) FROM Employees;`

(B) `SELECT DISTINCT(COUNT(DeptID)) FROM Employees;`

(C) `SELECT COUNT(DISTINCT DeptID) FROM Employees;`

(D) `SELECT COUNT(*) FROM Employees GROUP BY DeptID;`

Q44. Which of the following is a legitimate way to create a Pandas DataFrame?

(A) From a Dictionary of Lists

(B) From a List of Dictionaries

(C) From a 2D NumPy array

(D) All of the above

Q45. Given below are Two statements, Choose Correct One:
Statement 1: `df.describe()` provides statistical summary for only numeric columns by default.
Statement 2: `df.info()` is used to get a concise summary of the DataFrame including non-null counts.

(A) Both Statement 1 and Statement 2 are correct

(B) Both Statement 1 and Statement 2 are incorrect

(C) Statement 1 is correct but Statement 2 is incorrect

(D) Statement 1 is incorrect but Statement 2 is correct

Q46. In Matplotlib, which parameter in `plt.plot()` is used to change the line style to dashed?

(A) `marker='--'`

(B) `linestyle='--'`

(C) `format='dashed'`

(D) `line='dashed'`

Q47. Which protocol is primarily used for transferring files between a client and a server on a computer network?

(A) HTTP

(B) SMTP

(C) FTP

(D) IP

Q48. Using someone else's Wi-Fi connection without their permission is an example of:

(A) Ethical Hacking

(B) Phishing

(C) Unauthorized Access

(D) Intellectual Property Theft

Q49. Which attribute of a Pandas Series returns the number of bytes used by the data?

(A) `s.nbytes`

(B) `s.size`

(C) `s.memory`

(D) `s.bytes`

Q50. Which SQL keyword is used with aggregate functions to group the result-set by one or more columns?

(A) ORDER BY

(B) GROUP BY

(C) SORT BY

(D) CLUSTER BY

Detailed Solutions

Q1.

Solution

Concept:

In SQL, both `ROUND()` and `TRUNCATE()` are mathematical functions used to handle decimal numbers, but they operate with different logic. `ROUND(number, decimals)` reduces the number of decimal places by looking at the digit following the specified precision; if that digit is 5 or greater, it rounds the last kept digit up. Conversely, `TRUNCATE(number, decimals)` simply "cuts off" the number at the specified decimal place without any consideration for the value of the succeeding digits. Understanding this distinction is vital for financial and scientific applications where precision and rounding rules can significantly impact the final result.

Solution:

(a) The first part of the command is `ROUND(15.193, 1)`. We are rounding to one decimal place. The digit in the second decimal place is '9'. Since 9 is greater than or equal to 5, the first decimal digit '1' is incremented by 1, resulting in **15.2**.

(b) The second part of the command is `TRUNCATE(15.193, 1)`. Here, we are truncating to one decimal place. The function ignores the digits after the first decimal (9 and 3) completely. It does not matter that '9' is large; the function simply chops the number to yield **15.1**.

(c) Therefore, the combined output of the query is `15.2, 15.1`.

(d) Comparing this with the options, Option (A) is the correct match.

Final Answer: The output is 15.2, 15.1.

Answer: (A)

Q2.

Solution

Concept:

The `INSTR()` function (short for In-String) is a scalar string function in SQL used to search for the presence of a substring within a larger string. It is a zero-indexed or one-indexed search depending on the SQL dialect, but in MySQL, it is 1-indexed. A critical aspect of searching functions is their "failure state"—the value they return when the search term does not exist within the target string. While some programming languages return -1 for a failed search (like Java or Python), SQL standard functions typically use 0 to indicate that the substring was not found, as SQL string positions start at 1.

Solution:

(a) The question asks for the function that finds the position of a substring. The standard function for this in the IP syllabus is `INSTR(string, substring)`.

(b) Let's evaluate the options:

- **Option A:** `SEARCH()` is not a standard MySQL string function.
- **Option B:** `LOCATE()` is a valid function, but the question asks for a specific pairing. `LOCATE` and `INSTR` are similar, but `INSTR` is more commonly emphasized in the curriculum.
- **Option C:** `INSTR()` correctly identifies the position. If the substring is absent, it returns **0**. This is the standard behavior in MySQL.
- **Option D:** `SUBSTR()` is used to *extract* a substring, not find its position.

(c) Since `INSTR()` returns 0 when no match is found, Option (C) is the most accurate technical description.

Final Answer: The function is `INSTR()`; returns 0.

Answer: (C)

Q3.

Solution

Concept:

Date and time functions in SQL are designed to extract specific components from a `DATE` or `DATETIME` value. For reporting purposes, it is often necessary to convert a numerical date into a human-readable format, such as the name of the month or the name of the weekday. MySQL provides dedicated functions for these conversions. Understanding the difference between a function that returns a number (like `DAYOFWEEK`, which returns 1-7) and a function that returns a name (like `DAYNAME`, which returns 'Monday', 'Tuesday', etc.) is essential for meeting the specific display requirements of a user interface or report.

Solution:

(a) The user wants to display the "name of the day" (e.g., Monday, Tuesday).

(b) Let's analyze the available functions:

- `DAYNAME(date)`: This function takes a date and returns the full name of the day of the week. This matches the user's requirement exactly.
- `DAYOFWEEK(date)`: This function returns the index of the day (1 for Sunday, 2 for Monday, etc.). It returns a number, not a name.
- `DATE(date)`: This simply extracts the date part from a datetime expression.
- `MONTHNAME(date)`: This returns the name of the month (e.g., 'January'), which is not what was requested.

(c) Therefore, `DAYNAME(DOB)` is the most appropriate and direct function to use.

Final Answer: The function is `DAYNAME(DOB)`.

Answer: (A)

Q4.

Solution

Concept:

Trimming functions are essential for data cleaning and preprocessing in SQL. They are used to remove unwanted characters—most commonly spaces—from the beginning or the end of a string. This is crucial because extra spaces can cause errors in string comparisons, joins, and data validation. The `LTRIM()` function targets leading spaces (left side), while the `RTRIM()` function targets trailing spaces (right side). There is also a `TRIM()` function that removes spaces from both sides simultaneously. It is important to note that these functions do not affect spaces located in the middle of the string (between words).

Solution:

(a) The query is `SELECT LTRIM(" CUET 2026 "), RTRIM(" CUET 2026 ").`

(b) **LTRIM Analysis:** This function removes the three leading spaces at the start. However, it leaves the trailing spaces at the end untouched. So, the result is `"CUET 2026 "`.

(c) **RTRIM Analysis:** This function removes the three trailing spaces at the end. However, it leaves the leading spaces at the start untouched. So, the result is `" CUET 2026"`.

(d) Comparing these results to the options:

- Option A suggests both sides are trimmed in both functions, which is incorrect.
- Option B suggests LTRIM removes trailing and RTRIM removes leading, which is the opposite of reality.
- Option C matches our analysis: LTRIM removes left spaces (result: `"CUET 2026 "`) and RTRIM removes right spaces (result: `" CUET 2026"`).

Final Answer: The result is ”CUET 2026 ”, ” CUET 2026”.

Answer: (B)

Q5.

Solution

Concept:

In Structured Query Language (SQL), functions are specialized tools used to manipulate and transform data within a query. These functions are categorized based on the type of data they process: Mathematical functions (like MOD), String functions (like LENGTH and MID), and Date/Time functions (like NOW). Understanding these functions is fundamental for data processing tasks, such as calculating remainders, validating input lengths, extracting specific data segments, or timestamping records. Each category follows specific syntax rules and return types that are essential for accurate database management and reporting.

Solution:

(a) **(i) MOD(x, y):** This is a mathematical (arithmetic) function. It stands for the modulo operation, which calculates the remainder after dividing the first number (x) by the second number (y). For example, `SELECT MOD(10, 3)` would return 1. Therefore, (i) matches with (1).

(b) **(ii) LENGTH(s):** This is a string function used to measure the size of a text value. It counts the total number of characters, including spaces and special symbols, present within the string s. For instance, `LENGTH('CUET')` returns 4. Therefore, (ii) matches with (2).

(c) **(iii) MID(s, n, m):** This is a string extraction function, often synonymous with `SUBSTR` or `SUBSTRING`. It extracts a part of the string s, starting at the specified position n, for a length of m characters. For example, `MID('Informatics', 1, 4)` returns 'Info'. Therefore, (iii) matches with (3).

(d) **(iv) NOW():** This is a temporal function. It retrieves the current date and time from the server's operating system where the database is running. It is commonly used for creating timestamps in tables. Therefore, (iv) matches with (4).

(e) Combining these logic steps, the correct sequence is i-1, ii-2, iii-3, iv-4.

Final Answer: The matching is i-1, ii-2, iii-3, iv-4.

Answer: (A)

Q6.

Solution

Concept:

Nested functions in SQL involve using the output of one function as the input for another. To solve such expressions, one must follow the order of operations, starting from the innermost parentheses and working outwards. In this specific case, the query combines a string function, `INSTR()`, with a mathematical function, `POWER()`. The `INSTR()` function identifies the position of a substring, and the `POWER(base, exponent)` function calculates the value of the base raised to the power of the exponent. This type of question tests a student's ability to track data transformations across different functional domains.

Solution:

(a) **Inner Function:** Analyze `INSTR("PANDAS", "A")`. The function searches for the first occurrence of the character "A" in the string "PANDAS".

(b) In the string "PANDAS": P is at 1, A is at 2, N is at 3, D is at 4, A is at 5, S is at 6.

(c) The first "A" is found at position **2**.

(d) **Outer Function:** Now substitute the result into the outer function: `POWER(2, 3)`.

(e) This calculation is 2^3, which is $2 \times 2 \times 2 = 8$.

(f) Looking at the options, the value 8 matches Option (A).

Final Answer: The output is 8.

Answer: (A)

Q7.

Solution

Concept:

MySQL is flexible in how it allows users to access system information. To retrieve the current year, one can use a combination of extraction functions and system constants. The `YEAR()` function is a specific date function that extracts the year part (YYYY) from a given date. The source date can be provided by various system functions such as `CURDATE()` (returns current date), `NOW()` (returns date and time), or `SYSDATE()` (returns date and time at the moment of execution). Additionally, the standard SQL `EXTRACT()` function provides a more verbose but highly readable way to isolate specific components like year, month, or day.

Solution:

(a) **Option A:** `YEAR(CURDATE())`. `CURDATE()` provides the current date (e.g., '2026-05-07'), and `YEAR()` extracts '2026'. This is valid.

(b) **Option B:** `EXTRACT(YEAR FROM NOW())`. This is the standard SQL syntax for retrieving the year part from the current timestamp. This is also valid.

(c) **Option C:** `YEAR(SYSDATE())`. `SYSDATE()` is similar to `NOW()` but reflects the actual time the function executes. `YEAR()` will correctly extract the year from it. This is valid.

(d) Since all three methods are syntactically correct and will yield the same result (the current year), the correct answer is "All of the above."

Final Answer: The correct answer is All of the above.

Answer: (D)

Q8.

Solution

Concept:

SQL mathematical and string functions allow for complex data manipulation by performing arithmetic operations, rounding values to specific decimal or positional places, determining the sign of a number, and calculating string lengths. The `ROUND()` function with a negative precision parameter is used to round numbers to the nearest tens, hundreds, or thousands. The `TRUNCATE()` function is distinct from rounding as it simply discards the digits beyond the specified position. The `SIGN()` function is a mathematical utility that identifies whether a value is positive, negative, or zero. Finally, `LENGTH()` is a string function that returns the total count of characters, which can be combined with standard arithmetic to derive specific numerical results.

Solution:

(a) **(i) Result 100:** Looking at SQL Expression (1) `ROUND(123.45, -2)`. Here, the precision is -2, which tells SQL to round to the nearest hundredth place (left of the decimal). Since 123.45 is closer to 100 than to 200, the result is 100. Thus, (i) matches (1).

(b) **(ii) Result 123:** Looking at SQL Expression (2) `TRUNCATE(123.99, 0)`. The truncate function with a 0 precision simply removes all digits after the decimal point without any rounding logic. Even though the decimal is .99, it is discarded, leaving 123. Thus, (ii) matches (2).

(c) **(iii) Result -1:** Looking at SQL Expression (3) `SIGN(-25)`. The `SIGN()` function returns 1 for positive numbers, 0 for zero, and -1 for negative numbers. Since -25 is negative, the function returns -1. Thus, (iii) matches (3).

(d) **(iv) Result 5:** Looking at SQL Expression (4) `LENGTH("Informatics") - 6`. The word "Informatics" contains 11 characters. Performing the arithmetic $11 - 6$ yields a final result of 5. Thus, (iv) matches (4).

(e) Consequently, the correct matching sequence is i-1, ii-2, iii-3, iv-4, which corresponds to option (A).

Final Answer: The matching is i-1, ii-2, iii-3, iv-4.

Answer: (A)

Q9.

Solution

Concept:

In Relational Algebra, operations are used to manipulate relations. The Selection operation (σ) is a unary operation that returns a subset of tuples (rows) from a relation that satisfy a specific condition or predicate.

Solution:

(a) The question provides the expression $\sigma_{Salary>50000}(Employee)$.

(b) The symbol σ (sigma) specifically denotes the Selection operation in relational algebra.

(c) The condition ”Salary > 50000” acts as a filter to pick specific rows from the ’Employee’ table.

(d) Projection (Π) is used for selecting columns, not rows.

(e) Join and Intersection are binary operations involving two tables, whereas this is a unary operation.

(f) Therefore, the operation being performed is Selection.

Final Answer: The operation is Selection.

Answer: (B)

Q10.

Solution

Concept:

Relational database design involves various types of keys to uniquely identify records. While a Primary Key is the main unique identifier, other unique keys within the same table serve as backups or alternatives.

Solution:

(a) A Candidate Key is any minimal set of attributes that can uniquely identify a tuple.

(b) A table can have multiple Candidate Keys.

(c) One of these is selected by the Database Administrator to be the Primary Key.

(d) All other Candidate Keys that were not selected to be the Primary Key are called Alternate Keys.

(e) A Foreign Key is used to link two tables, and a Secondary Key is usually used for indexing purposes, not necessarily for unique identification.

Final Answer: The key is the Alternate Key.

Answer: (C)

Q11.

Solution

Concept:

The ACID properties (Atomicity, Consistency, Isolation, Durability) are a set of properties that guarantee database transactions are processed reliably.

Solution:

(a) The question asks for the property that ensures an ”all or nothing” execution.

(b) Atomicity ensures that if any part of the transaction fails, the entire transaction is aborted, and the database is left unchanged.

(c) Consistency ensures the database moves from one valid state to another.

(d) Isolation ensures that concurrent transactions do not interfere with each other.

(e) Durability ensures that once a transaction is committed, it remains so even in the event of a system failure.

Final Answer: The property is Atomicity.

Answer: (D)

Q12.

Solution

Concept:

Projection (Π) is a unary operation in relational algebra used to select specific attributes (columns) from a relation while discarding the rest.

Solution:

(a) The expression provided is $\Pi_{A,C}(R)$.

(b) The symbol Π (pi) denotes the Projection operation.

(c) The subscripts A and C represent the specific attributes to be extracted from the relation R.

(d) Unlike Selection, which filters rows, Projection filters columns vertically.

(e) The resulting relation will contain only the data from columns A and C for all existing rows in R.

Final Answer: The operation results in only columns A and C.

Answer: (B)

Q13.

Solution

Concept:

Network topology defines the physical or logical layout of a network. Different topologies have different methods of connecting nodes (computers/devices).

Solution:

(a) In a Star Topology, all nodes are individually connected to a central connection point, such as a hub, switch, or router.

(b) In a Bus Topology, all nodes share a single communication line (backbone).

(c) In a Mesh Topology, nodes are interconnected with many redundant interconnections.

(d) In a Tree Topology, nodes are arranged in a hierarchy, but it is essentially a collection of star networks connected to a bus.

(e) The requirement for a central controller or hub is the defining characteristic of the Star Topology.

Final Answer: The topology is Star Topology.

Answer: (C)

Q14.

Solution

Concept:

IP (Internet Protocol) addresses are unique numerical identifiers assigned to every device connected to a computer network. There are two primary versions of IP addresses currently in use: IPv4 and IPv6, each with different bit-lengths to accommodate the growing number of devices globally.

Solution:

(a) IPv4 (Internet Protocol version 4) uses a 32-bit address scheme, which allows for approximately 4.3 billion unique addresses. It is typically represented in dotted-decimal notation (e.g., 192.168.1.1).

(b) Because the world ran out of IPv4 addresses, IPv6 was developed.

(c) IPv6 (Internet Protocol version 6) uses a 128-bit address scheme, providing a vastly larger address space (3.4×10^{38} addresses). It is represented in hexadecimal notation.

(d) Therefore, the sizes are 32 bits and 128 bits respectively.

Final Answer: The sizes are 32 bits and 128 bits.

Answer: (B)

Q15.

Solution

Concept:

Networking devices operate at different layers of the OSI (Open Systems Interconnection) model. Understanding where a device sits helps determine how it handles data packets or frames.

Solution:

(a) The question specifies the ”Data Link Layer” and the use of ”MAC addresses.”

(b) A Hub is a physical layer device that broadcasts data to all ports; it does not look at addresses.

(c) A Repeater also works at the physical layer to regenerate signals.

(d) A Switch is a Data Link Layer device (Layer 2). it maintains a MAC address table to intelligently forward frames only to the specific port where the destination device is connected.

(e) A Gateway usually operates at the higher layers (Network or Application) to connect different networks.

Final Answer: The device is a Switch.

Answer: (C)

Q16.

Solution

Concept:

Pandas is a powerful Python library for data manipulation. It provides two main data structures: Series (1D) and DataFrame (2D). Understanding the homogeneity and dimensionality of these structures is fundamental.

Solution:

(a) Statement 1 claims a Series can store only homogeneous data. This is correct because a Series is built on top of NumPy arrays, meaning all elements in a single Series typically share the same data type (dtype).

(b) Statement 2 claims a DataFrame is a 2D labeled structure with potentially different types. This is also correct; a DataFrame is essentially a container for Series objects, where each column can have a different data type (e.g., one column of integers, one of strings).

(c) Since both statements accurately describe the characteristics of Pandas structures, option A is correct.

Final Answer: Both Statement 1 and Statement 2 are correct.

Answer: (A)

Q17.

Solution

Concept:

Slicing in a Pandas Series allows for retrieving a subset of data. However, the behavior of slicing depends on whether you use positional (integer) indices or labeled indices.

Solution:

(a) The Series `s` is created with data [10, 20, 30] and labels [’a’, ’b’, ’c’].

(b) The command `s['a':'b']` uses label-based slicing.

(c) In Pandas, label-based slicing is inclusive of the end point. This is a critical distinction from Python’s standard integer slicing.

(d) Therefore, it will include both the entry for ’a’ (10) and the entry for ’b’ (20).

(e) The output will show index ’a’ with 10 and index ’b’ with 20.

Final Answer: The output is 10, 20.

Answer: (B)

Q18.

Solution

Concept:

Pandas DataFrames have several attributes that provide metadata about the object. Knowing which attribute returns which specific piece of information (size, shape, or dimension) is essential for data inspection.

Solution:

(a) `.size` returns the total number of elements in the DataFrame (rows multiplied by columns).

(b) `.shape` returns a tuple representing the dimensionality of the DataFrame in the format (number of rows, number of columns).

(c) `.ndim` returns the number of axes or dimensions (which is always 2 for a DataFrame).

(d) `.index` returns the row labels of the DataFrame.

(e) To check the specific dimensions (rows and columns), `.shape` is the correct attribute.

Final Answer: The attribute is .shape.

Answer: (B)

Q19.

Solution

Concept:

Accessing data in Pandas can be done through two primary methods: `loc` and `iloc`. These methods have different rules regarding how they interpret the indices provided to them.

Solution:

(a) Statement 1 states that `loc` is label-based slicing. This is correct. `loc` uses the explicit row and column labels assigned to the DataFrame.

(b) Statement 2 states that `iloc` includes the start index but excludes the stop index. This is also correct. `iloc` is integer-position based and follows the standard Python slicing convention (start to stop-1).

(c) In contrast, `loc` is inclusive of both the start and the stop labels.

(d) Since both descriptions of the indexing methods are accurate, the statements are both correct.

Final Answer: Both Statement 1 and Statement 2 are correct.

Answer: (A)

Q20.

Solution

Concept:

Pandas provides various functions to save data to external files. For CSV (Comma Separated Values) files, specific methods exist to convert the in-memory DataFrame into a persistent file format.

Solution:

(a) The question asks for the method to export a DataFrame to a file named 'data.csv'.

(b) The standard Pandas naming convention for export functions follows the pattern `to_-<format>()`.

(c) Therefore, `to_csv()` is the built-in function used for this purpose.

(d) Other options like `save_csv()`, `write_csv()`, or `csv_export()` do not exist in the standard Pandas library.

(e) The correct syntax is `df.to_csv('filename.csv')`.

Final Answer: The method is `df.to_csv()`.

Answer: (B)

Q21.

Solution

Concept:

The `size` attribute of a DataFrame is different from the `shape` attribute. While `shape` gives the dimensions, `size` gives the total volume of data points stored in the object.

Solution:

(a) The DataFrame `df` is described as having 5 rows and 3 columns.

(b) The `size` attribute in Pandas returns an integer representing the total number of elements in the object.

(c) The calculation for `size` is: Number of Rows $\times$ Number of Columns.

(d) In this case, $5 \times 3 = 15$.

(e) Thus, `df.size` will return 15.

Final Answer: The value returned is 15.

Answer: (C)

Q22.

Solution

Concept:

To quickly inspect the beginning or the end of a dataset, Pandas provides the `head()` and `tail()` methods. These are essential for verifying data loading or processing results.

Solution:

(a) The question asks to display the "last 3 rows" of a DataFrame.

(b) The `head(n)` function returns the first n rows.

(c) The `tail(n)` function returns the last n rows.

(d) Passing the integer 3 into the `tail()` function (`df.tail(3)`) will retrieve exactly the last 3 records.

(e) `df.head(-3)` would actually return all rows *except* the last 3, which is not what was requested.

Final Answer: The command is `df.tail(3)`.

Answer: (B)

Q23.

Solution

Concept:

Data cleaning often involves handling missing values. Pandas uses specific markers for missing data and provides different methods to either remove or replace those missing values.

Solution:

(a) Statement 1: In Pandas, missing data is indeed represented by NaN (Not a Number), which is a floating-point value defined by NumPy. This statement is correct.

(b) Statement 2: The `dropna()` method is used to *remove* rows or columns that contain missing values.

(c) To *fill* missing values with a specific constant, the method `fillna()` is used, not `dropna()`.

(d) Therefore, Statement 1 is correct, but Statement 2 is incorrect.

Final Answer: Statement 1 is correct but Statement 2 is incorrect.

Answer: (C)

Q24.

Solution

Concept:

SQL functions are broadly divided into two main types based on the number of rows they process: **Aggregate Functions** and **Scalar Functions**. Aggregate functions (also known as Multiple-Row functions) take a set of values from multiple rows and return a single summarized value. They are commonly used with the `GROUP BY` clause. Scalar functions (Single-Row functions) operate on a single value from a single row and return one result for every row processed. Scalar functions are further categorized based on the data type they handle, such as **Numeric** (mathematical operations), **String** (text manipulation), and **Date/Time** functions.

Solution:

(a) **(i) COUNT():** This function is used to find the total number of rows or non-null values in a column. Since it looks at a group of rows to return one total count, it is an **Aggregate Function**. Matches with (2).

(b) **(ii) MAX():** This function identifies the highest value within a specified column across a set of rows. Like `COUNT()`, it processes multiple records to return one result, making it an **Aggregate Function**. Matches with (2).

(c) **(iii) ROUND():** This function is used to round a numeric value to a specified number of decimal places. It operates on individual numbers in each row and performs a mathematical calculation, so it is a **Numeric (Scalar) Function**. Matches with (4).

(d) **(iv) LENGTH():** This function calculates the number of characters in a string. It works on a single text input per row and returns the character count, qualifying it as a **String (Scalar) Function**. Matches with (3).

(e) The combined mapping is (i)-2, (ii)-2, (iii)-4, and (iv)-3. Note that in this specific question format, both (i) and (ii) share the same category from List II.

Final Answer: The correct match is (i)-2, (ii)-2, (iii)-4, (iv)-3.

Answer: (A)

Q25.

Solution

Concept:

In SQL, the order of execution for clauses is crucial. While the `WHERE` clause filters rows before grouping occurs, there is a specific clause designed to filter the groups themselves after they have been created.

Solution:

(a) The question asks for the clause that filters results "produced by the GROUP BY clause."

(b) The `WHERE` clause cannot be used with aggregate functions (like SUM or AVG) directly because it acts on individual rows.

(c) The `HAVING` clause was specifically introduced to allow for search conditions on groups or aggregates.

(d) `ORDER BY` is used for sorting the final result, and `LIMIT` is used to restrict the number of rows displayed.

(e) Therefore, `HAVING` is the correct clause for filtering grouped data.

Final Answer: The clause is HAVING.

Answer: (B)

Q26.

Solution

Concept:

SQL Joins are used to combine rows from two or more tables based on a related column. "Outer" joins are used when you want to include rows even if there is no match in the joining table.

Solution:

(a) An `INNER JOIN` only returns rows where there is a match in both tables.

(b) A `LEFT JOIN` returns all rows from the left table and matched rows from the right.

(c) A `RIGHT JOIN` returns all rows from the right table and matched rows from the left.

(d) A `FULL OUTER JOIN` combines the behavior of both Left and Right joins. It returns all records when there is a match in either the left OR the right table.

(e) If a row has no match, the result set will contain NULL values for every column of the table that lacks a matching row.

Final Answer: The join is FULL OUTER JOIN.

Answer: (D)

Q27.

Solution

Concept:

Sorting is a fundamental operation in database management. The `ORDER BY` clause allows users to arrange the output in a specific sequence based on one or more columns.

Solution:

(a) When the `ORDER BY` clause is used without specifying a direction, SQL must apply a default behavior.

(b) `ASC` stands for Ascending (smallest to largest, or A to Z).

(c) `DESC` stands for Descending (largest to smallest, or Z to A).

(d) In SQL, if you write `SELECT * FROM Students ORDER BY Name;`, the database will sort the names in ascending order by default.

Final Answer: The default sort order is ASC.

Answer: (B)

Q28.

Solution

Concept:

In relational databases, joins are operations used to combine rows from two or more tables based on a related column between them. Understanding the logical differences between various join types is fundamental for data retrieval. An **Equi Join** is the most common form of join, utilizing the equality operator to find matches. A **Natural Join** is a specialized version of an Equi Join that automates the process based on common column names. A **Cross Join** represents the most basic form of table combination, where every possible pairing of rows is created regardless of shared data.

Solution:

(a) **(i) Equi Join:** An Equi Join is a specific type of join that uses the equality operator (=) to join tables. In the SQL `WHERE` or `ON` clause, it specifies that values in one column must equal values in another column for a row to be included in the result set. Therefore, (i) matches with **(2)**.

(b) **(ii) Natural Join:** A Natural Join is an Equi Join that occurs automatically on all columns with the same name in both tables. A defining characteristic of a Natural Join in standard SQL is that it suppresses redundant columns, meaning that the common joining column appears only once in the output. Therefore, (ii) matches with **(3)**.

(c) **(iii) Cross Join:** A Cross Join produces the **Cartesian Product** of two tables. This means that every row from the first table is combined with every single row from the second table. If Table A has 5 rows and Table B has 10 rows, the Cross Join result will have 50 rows. No joining condition is required. Therefore, (iii) matches with **(1)**.

(d) Aligning these results: (i)-2, (ii)-3, and (iii)-1. This configuration corresponds to option (A).

Final Answer: The correct matching is i-2, ii-3, iii-1.

Answer: (A)

Q29.

Solution

Concept:

Matplotlib is the most widely used plotting library in Python for creating static, interactive, and animated visualizations. To make a chart informative, it requires descriptive metadata such as axis labels, legends, and a main title. The library provides a specific sub-module, `pyplot`, which contains a collection of functions that make Matplotlib work like MATLAB, allowing users to modify figures easily.

Solution:

(a) When creating a plot, the primary way to identify the content of the entire visualization is by adding a title at the top.

(b) The function `plt.title()` is the standard method used to define this string. It accepts a string argument and offers parameters to customize the font size, color, and alignment.

(c) It is important to distinguish this from other similar-sounding methods:

- `plt.set_title()` is typically used when working with the Object-Oriented interface (on an *Axes* object), not the functional `pyplot` interface.
- `plt.xlabel()` and `plt.ylabel()` are used for the horizontal and vertical axes respectively.
- There are no functions named `plt.head()` or `plt.label_title()` in the Matplotlib library for this purpose.

(d) Therefore, in the context of the `pyplot` module (usually imported as `plt`), `plt.title('Your Title')` is the correct syntax.

Final Answer: The function used is `plt.title()`.

Answer: (B)

Q30.

Solution

Concept:

A bar chart (or bar graph) is a category-based visualization where the length or height of the bars is proportional to the values they represent. In Matplotlib, the `bar()` function is used to generate vertical bar plots. This function requires at least two primary positional arguments to define the structure of the graph on the Cartesian plane.

Solution:

(a) The syntax for a basic bar plot is `plt.bar(x, height)`.

(b) The first parameter, **'x'**, represents the coordinates of the bars on the horizontal axis. In the context of Informatics Practices, these are usually the "Categories," "Labels," or "Names" of the data points being compared (e.g., Names of Cities, Years, or Names of Students).

(c) The second parameter, **'y'** (often referred to as *height* in documentation), represents the numerical values associated with those categories. This determines how tall the bar will be.

(d) It is a common mistake to confuse the two; however, the 'x' parameter always dictates the placement and identity of the bars along the horizontal baseline, while the numerical magnitude is handled by the height parameter.

(e) Therefore, 'x' represents the categories or positions on the x-axis.

Final Answer: The 'x' parameter represents categories or positions on the x-axis.

Answer: (C)

Q31.

Solution

Concept:

A legend is an area describing the elements of the graph. In Matplotlib, when you plot multiple datasets on the same axes (for example, two different lines in a line chart), you need a way to distinguish them. This is done by assigning a ”label” to each plot and then calling a specific function to render the box that explains these labels.

Solution:

(a) To display a legend, a two-step process is usually followed. First, the `label` argument must be provided inside the plotting function (e.g., `plt.plot(x, y, label='Growth')`).

(b) Once labels are assigned, they are not automatically visible on the screen. The user must explicitly call the `plt.legend()` function.

(c) This function searches for all active labels in the current figure and compiles them into a legend box.

(d) You can further customize the legend using the `loc` parameter (to change its position like 'upper left' or 'best') and the `shadow` or `fontsize` parameters.

(e) Options like `show_legend()` or `display()` are incorrect as they do not exist in the Matplotlib namespace for this specific task.

Final Answer: The function used is `plt.legend()`.

Answer: (B)

Q32.

Solution

Concept:

Histograms are specialized charts used to represent the distribution of a continuous numerical variable. Unlike bar charts, which compare discrete categories, histograms group data into ”bins” (ranges of values) and show the frequency of data points falling within each range. This is a critical tool in statistics for understanding the spread, skewness, and outliers in a dataset.

Solution:

(a) Matplotlib provides a direct and efficient function for this called `plt.hist()`.

(b) When you pass a list or a Pandas Series to `plt.hist(data)`, the function automatically calculates the range of the data, divides it into a default number of bins (usually 10), and counts how many values fall into each bin.

(c) It then renders these counts as vertical bars that are typically touching each other (signifying the continuous nature of the underlying data).

(d) Other functions like `plt.histogram()` might exist in the NumPy library (to return numerical values of bins), but for the purpose of *plotting* the visual chart in Matplotlib, `plt.hist()` is the correct command.

(e) Options such as `bar_hist()` or `plot_hist()` are invented names and do not exist in the library.

Final Answer: The function used is `plt.hist()`.

Answer: (B)

Q33.

Solution

Concept:

E-waste (Electronic Waste) refers to discarded electrical or electronic devices. Used electronics which are destined for refurbishment, reuse, resale, salvage recycling through material recovery, or disposal are also considered e-waste. With the rapid advancement of technology, the volume of e-waste is increasing globally, posing significant environmental and health risks due to toxic components like lead, mercury, and cadmium.

Solution:

(a) The primary goal of E-waste management is to mitigate the negative impact of discarded technology on the environment.

(b) This involves two main strategies: Safe Disposal and Recycling.

(c) **Recycling:** Many components of electronic devices (like gold, copper, and specialized plastics) are valuable and can be recovered to manufacture new products, reducing the need for mining.

(d) **Safe Disposal:** Electronic items contain hazardous substances that can leach into soil and groundwater if left in standard landfills. Proper management ensures these toxins are neutralized or contained.

(e) While tracking devices or creating hardware are technical activities, they do not define the ”management” aspect of waste. The core objective is environmental protection through systematic recycling and disposal.

Final Answer: The primary purpose is to safely dispose or recycle old electronic devices.

Answer: (B)

Q34.

Solution

Concept:

The concept of a Digital Footprint is central to modern digital citizenship and cyber ethics. Every time an individual interacts with the internet—whether it is through social media, online shopping, or simply browsing a website—they leave behind a trail of data. This trail is not temporary; it is often permanent and can be used to track a user's behavior, preferences, and location. There are two types of digital footprints: active and passive. An active footprint is created when a user intentionally shares information (like posting a photo), while a passive footprint is created without the user's direct knowledge (like a website logging their IP address).

Solution:

(a) The question asks for the specific term that refers to the "trail of data" left behind by internet usage.

(b) A "Digital Cookie" is a small piece of data stored on the user's computer by a web browser, which is a *component* of the trail but not the name of the trail itself.

(c) "Digital Footprint" is the comprehensive term used in Informatics Practices and Cybersecurity to describe the entire history and record of a person's online activity.

(d) "Cyber Trail" and "Web Shadow" are descriptive terms occasionally used in literature, but "Digital Footprint" is the official technical term recognized in the CUET syllabus.

(e) Understanding one's digital footprint is crucial because it contributes to one's online reputation, which can be seen by future employers, colleges, and digital platforms.

Final Answer: The term is Digital Footprint.

Answer: (B)

Q35.

Solution

Concept:

Plagiarism is a significant ethical violation in the academic and professional world. It involves the act of taking someone else's original ideas, words, or creative work and presenting them as if they were one's own creation, without providing proper credit or citation to the original author. In the digital age, where information is easily accessible via "copy and paste," plagiarism has become a widespread issue. It is not limited to text; it can include images, music, code, and even the structural logic of a project.

Solution:

(a) The core of the question lies in "stealing intellectual work" and "representing it as your own."

(b) **Phishing** is a form of cybercrime where attackers pose as legitimate institutions to lure individuals into providing sensitive data (like passwords).

(c) **Spamming** is the act of sending unsolicited, bulk messages over the internet.

(d) **Hacking** refers to the unauthorized access or manipulation of a computer system or network.

(e) **Plagiarism** specifically addresses the theft of intellectual content and the lack of attribution. It is considered an infringement of intellectual property rights and an act of academic dishonesty.

Final Answer: The act is called Plagiarism.

Answer: (C)

Q36.

Solution

Concept:

Intellectual Property Rights (IPR) are legal rights that provide creators protection for their original inventions, designs, and artistic works. The primary goal of IPR is to encourage innovation by allowing the creators to benefit financially and socially from their work for a specific period. Without these protections, creators would have little incentive to invest time and resources into developing new technologies or art forms. IPR is a broad umbrella that covers several distinct legal categories, each designed for a different type of intellectual ”property.”

Solution:

(a) **Patents:** These protect new inventions or functional processes, preventing others from making, using, or selling the invention without permission.

(b) **Copyrights:** These protect original works of authorship, such as books, music, software code, and paintings. They give the creator the sole right to reproduce and distribute the work.

(c) **Trademarks:** These protect symbols, names, and slogans used to identify goods or services and distinguish them from those of competitors (e.g., the Nike ”Swoosh”).

(d) Since Patents, Copyrights, and Trademarks are all fundamental pillars of the IPR framework, they all fall under this category.

Final Answer: All of the above fall under IPR.

Answer: (D)

Q37.

Solution

Concept:

In Pandas, descriptive statistics allow users to summarize the central tendency, dispersion, and shape of a dataset's distribution. Calculating the "average" is one of the most common requirements in data analysis. Pandas provides highly optimized methods to perform these calculations on specific columns of a DataFrame. It is important to remember that these methods ignore missing values (NaN) by default, which ensures that the statistical calculation is not skewed by null entries.

Solution:

(a) The user wants to find the average value of the 'Sales' column in a DataFrame `df`.

(b) In the Pandas library, the function for the arithmetic mean (average) is `.mean()`.

(c) The syntax for this operation is `df['ColumnName'].mean()`.

(d) While "average" is the mathematical concept, there is no `.average()` method in the Pandas Series object; using it would result in an `AttributeError`.

(e) Option C (`sum() / 10`) is only correct if the DataFrame has exactly 10 non-null entries, which is not stated. Option D is syntactically incorrect for standard Pandas usage.

(f) Therefore, `df['Sales'].mean()` is the correct and most robust command.

Final Answer: The command is `df['Sales'].mean()`.

Answer: (B)

Q38.

Solution

Concept:

A histogram is a graphical representation that organizes a group of data points into user-specified ranges. It is a tool for visualizing the frequency distribution of a dataset. One of the most important parameters when generating a histogram in Matplotlib is the ”bins” parameter. ”Binning” is the process of taking a set of continuous data and dividing it into discrete intervals. The visual appearance and the interpretability of the histogram depend heavily on the size and number of these bins.

Solution:

(a) When we use `plt.hist(data, bins=n)`, the parameter `bins` tells Matplotlib how many equal-width intervals the data should be divided into.

(b) For example, if our data ranges from 0 to 100 and we set `bins=10`, Matplotlib will create 10 bars, each representing a range of 10 units (0-10, 10-20, etc.).

(c) If the number of bins is too small, the data distribution is over-summarized and patterns may be lost. If the number is too large, the histogram becomes ”noisy” and hard to read.

(d) The width of the bars is automatically calculated based on the data range divided by the number of bins.

(e) Thus, the bins define the number of intervals for the data distribution.

Final Answer: The bins define the number of intervals for the data distribution.

Answer: (B)

Q39.

Solution

Concept:

The `rename()` function in the Pandas library is a powerful tool for modifying the metadata of a DataFrame. In data analysis, it is frequently necessary to change column names or row indices to make them more descriptive or to align them with other datasets. This function allows for mapping old names to new names using a dictionary-like structure. A key feature of many Pandas functions, including `rename()`, is the `inplace` parameter, which determines whether the changes are applied to the existing object or if a new object is returned with the changes.

Solution:

(a) Statement 1: This is correct. The `rename()` function can accept an `index` argument and a `columns` argument, allowing for the simultaneous renaming of both row labels and column headers. For example, `df.rename(columns={"old_col": "new_col"}, index={0: "first_row"})` is a valid operation.

(b) Statement 2: This is incorrect. In Pandas, most data transformation functions (including `rename`, `drop`, and `fillna`) default to `inplace=False`. This means they return a "copy" of the DataFrame with the changes, leaving the original DataFrame untouched. To modify the original object without creating a new one, the user must explicitly set `inplace=True`.

(c) Because Statement 1 accurately describes the function's capability and Statement 2 fails to correctly identify the default behavior of the `inplace` parameter, option C is the only logical choice.

Final Answer: Statement 1 is correct but Statement 2 is incorrect.

Answer: (C)

Q40.

Solution

Concept:

Removing unnecessary data is a fundamental part of data cleaning. In Pandas, the `drop()` method is used to delete rows or columns from a DataFrame. To use this method effectively, one must understand the `axis` parameter. In the Pandas coordinate system, `axis=0` refers to rows (the vertical axis), and `axis=1` refers to columns (the horizontal axis). Furthermore, to make a change "permanent" (meaning the original variable is updated), the user must use the `inplace=True` parameter; otherwise, the function only returns a view of the data without the dropped column.

Solution:

(a) The task is to drop a column named 'Address' permanently.

(b) Option A (`axis=0`) is incorrect because it would attempt to find a row with the label 'Address', which would result in a `KeyError`.

(c) Option B specifies `axis=1`, which correctly targets the columns. It also includes `inplace=True`, which ensures the change is permanent and not just a temporary display.

(d) Option C (`remove`) is not a valid method for a Pandas DataFrame.

(e) Option D (`column='Address'`) is syntactically incomplete as it lacks both the `axis` specification (unless using the `columns=` keyword argument) and the `inplace` parameter required for permanence.

(f) Thus, Option B is the only complete and correct command for the specific requirement.

Final Answer: The command is `df.drop('Address', axis=1, inplace=True)`.

Answer: (B)

Q41.

Solution

Concept:

Case study problems in Informatics Practices often require combining multiple Pandas operations to solve a real-world scenario. To find the "top performers" in a dataset, one must first arrange the data based on a specific numerical criterion (like Marks) and then select a specific number of records from the top of that sorted list. This involves the use of `sort_values()` for ordering and `head()` for retrieval. Understanding the `ascending` parameter is crucial here, as it defaults to `True` (lowest to highest), which would show the worst performers first.

Solution:

(a) The objective is to identify the "top 5 performers" based on 'Marks'.

(b) To get the highest marks at the top of the list, we must sort the DataFrame in descending order. This is achieved by setting `ascending=False` in the `sort_values()` method.

(c) Option A sorts by marks in ascending order (default), so `head(5)` would return the students with the 5 lowest marks.

(d) Option B correctly sorts the 'Marks' column from highest to lowest (`ascending=False`) and then selects the first 5 rows using `head(5)`. This perfectly identifies the top performers.

(e) Option C (`max(5)`) is not a valid way to retrieve 5 rows; `max()` typically returns the single highest value.

(f) Option D is syntactically incorrect as `head()` is usually applied to the sorted result, not before it, and the sort logic is missing parameters.

Final Answer: The command is `df_stud.sort_values('Marks', ascending=False).head(5)`.

Answer: (B)

Q42.

Solution

Concept:

SQL provides a variety of built-in functions to handle temporal data (Date and Time). These functions are essential for logging events, calculating intervals, or filtering records based on when they were created or updated. Different database systems (like MySQL, PostgreSQL, or SQL Server) may have slightly different names for these functions, but the standard curriculum usually focuses on the most common implementations. In MySQL, which is the primary dialect taught in Informatics Practices, there is a specific function that retrieves both the date and the exact time of the system.

Solution:

(a) The question asks for a function that returns both current "date and time."

(b) `DATE()` is a function that extracts the date part from a date-time expression, but it does not return the current system time on its own.

(c) `GETDATE()` is the standard function used in T-SQL (Microsoft SQL Server), but it is not the primary function taught in the standard IP syllabus.

(d) `NOW()` is the standard MySQL function that returns the current date and time in the format 'YYYY-MM-DD HH:MM:SS'. It is the most common answer in competitive exams like CUET for this specific requirement.

(e) `TIME()` is used to extract the time portion from a given date-time expression.

(f) Therefore, `NOW()` is the correct function for retrieving the current timestamp.

Final Answer: The function is `NOW()`.

Answer: (C)

Q43.

Solution

Concept:

Data distinctness and counting are common requirements in SQL reporting. The `COUNT()` function is an aggregate function that returns the number of rows that match a specified criterion. By default, `COUNT(column_name)` counts all non-null entries in that column, including duplicates. To find the number of ”unique” or ”distinct” entries, the `DISTINCT` keyword must be used. The placement of this keyword is critical: it must be placed *inside* the parentheses of the `COUNT` function so that the database first identifies the unique values and then counts them.

Solution:

(a) The requirement is to find the total number of ”unique departments.”

(b) Option A: `COUNT(DeptID)` counts every row where DeptID is not null. If there are 10 employees in 2 departments, this would return 10.

(c) Option B: `DISTINCT(COUNT(DeptID))` is logically flawed because `COUNT` returns a single number (e.g., 10), and the ”distinct” version of a single number is just that number itself.

(d) Option C: `COUNT(DISTINCT DeptID)` correctly instructs the database to look at the ’DeptID’ column, find the set of unique values (e.g., ’Sales’, ’IT’), and then count how many items are in that unique set. This returns 2.

(e) Option D: This would return a list of counts (one for each department), rather than a single total count of the departments themselves.

Final Answer: The query is `SELECT COUNT(DISTINCT DeptID) FROM Employees;`.

Answer: (C)

Q44.

Solution

Concept:

The versatility of the Pandas library stems from its ability to ingest data from a wide variety of sources and formats. Creating a DataFrame is often the first step in any data analysis pipeline. Since a DataFrame is essentially a two-dimensional labeled data structure, it can be visualized as a table or a spreadsheet. Pandas provides the `pd.DataFrame()` constructor, which is highly flexible and can convert various Python data structures into this tabular format. Understanding these input types is crucial for data scientists who need to transition data from raw Python objects or numerical libraries like NumPy into the structured environment of Pandas.

Solution:

(a) **From a Dictionary of Lists:** This is perhaps the most common method. The keys of the dictionary become the column headers, and the lists associated with those keys become the column data. For example, `{'Name': ['Alice', 'Bob'], 'Age': [25, 30]}`.

(b) **From a List of Dictionaries:** In this format, each dictionary in the list represents a single row (a record). The keys in each dictionary represent the column names. This is very common when dealing with JSON data from APIs.

(c) **From a 2D NumPy array:** Since Pandas is built on top of NumPy, it can directly take a two-dimensional array of numbers and wrap it in a DataFrame structure. You can optionally provide column and index labels; otherwise, they default to integers.

(d) Because all three methods are standard, built-in ways to initialize a DataFrame, the correct answer is ”All of the above.”

Final Answer: The correct answer is All of the above.

Answer: (D)

Q45.

Solution

Concept:

Data exploration is the process of summarizing the main characteristics of a dataset, often with visual methods, but primarily using descriptive statistics and structural summaries. Pandas provides two essential methods for this: `describe()` and `info()`. While they may seem similar to a beginner, they serve very different purposes. One focuses on the "what" (the statistical distribution of the data), while the other focuses on the "how" (the technical structure and memory usage of the DataFrame). Distinguishing between these two is vital for effective data auditing and preparation.

Solution:

(a) **Statement 1:** This is correct. The `df.describe()` method generates descriptive statistics that summarize the central tendency, dispersion, and shape of a dataset's distribution. By default, it analyzes only numeric columns (int, float) and provides the count, mean, standard deviation, min, max, and quartiles. To include categorical data, one would need to use `include='all'`.

(b) **Statement 2:** This is also correct. The `df.info()` method prints information about a DataFrame including the index dtype and columns, non-null values, and memory usage. It is the go-to tool for checking if there are missing values (by comparing the non-null count to the total number of entries).

(c) Since both statements accurately reflect the primary use cases and default behaviors of these two diagnostic tools in the Pandas library, they are both considered correct.

Final Answer: Both Statement 1 and Statement 2 are correct.

Answer: (A)

Q46.

Solution

Concept:

Visualizing data with line plots is a core task in Informatics Practices. In Matplotlib, the `plt.plot()` function is used to create these charts. Beyond simply drawing lines between data points, users often need to customize the aesthetic properties of the line to make the graph more readable or to distinguish between multiple data series. Line styles are one of the most common customizations. Matplotlib provides a specific keyword argument that accepts either a descriptive string (like 'dashed' or 'dotted') or a shorthand symbol (like '–' or ':').

Solution:

(a) To change the appearance of the line itself, we use the `linestyle` (or shorthand `ls`) parameter.

(b) The value `'--'` is the specific shorthand for a dashed line. Other common values include `'-'` for a solid line, `':'` for a dotted line, and `'-.'` for a dash-dot line.

(c) **Marker:** This parameter (Option A) is used to highlight individual data points (e.g., circles 'o' or stars '*'), not the line connecting them.

(d) **Format:** While Matplotlib has a "format string" (e.g., `'r--'` for a red dashed line), there is no specific keyword argument named `format` that takes 'dashed' as a value.

(e) **Line:** There is no keyword argument simply named `line` in the `plt.plot()` function.

(f) Therefore, the `linestyle='--'` parameter is the correct way to achieve a dashed line.

Final Answer: The parameter is linestyle='–'.

Answer: (B)

Q47.

Solution

Concept:

In computer networking, protocols are sets of rules that govern how data is exchanged between devices. Different protocols are designed for different specific tasks. For instance, some are meant for browsing the web, others for sending emails, and others for the underlying routing of data packets. Understanding the specific function of each protocol is a key component of the Computer Networks unit in the IP syllabus. File transfer is a specific requirement that involves maintaining a stable connection to move potentially large amounts of data without corruption.

Solution:

(a) **HTTP (Hypertext Transfer Protocol):** Primarily used for transferring web pages and related content from a server to a web browser. While it can ”download” files, it is not its primary specialized purpose.

(b) **SMTP (Simple Mail Transfer Protocol):** Specifically used for sending email messages between servers.

(c) **FTP (File Transfer Protocol):** This is the standard network protocol designed specifically for transferring files between a client and a server on a computer network. It allows for uploading, downloading, and managing files on a remote server.

(d) **IP (Internet Protocol):** This is a fundamental network layer protocol responsible for addressing and routing packets across network boundaries; it does not handle high-level tasks like ”file transfer” directly.

(e) Thus, FTP is the protocol whose primary definition matches the requirement.

Final Answer: The protocol is FTP.

Answer: (C)

Q48.

Solution

Concept:

Cyber ethics and security deal with the moral and legal aspects of internet and network usage. Unauthorized access is a broad category of cybercrime that involves entering a computer system, network, or data storage without the permission of the owner. In many jurisdictions, this is a punishable offense. One of the most common real-world examples of unauthorized access is "piggybacking" or "mooching" off a private wireless network. Even if the network is not password-protected, using it without the owner's consent is considered unethical and, in many cases, illegal.

Solution:

(a) **Ethical Hacking:** This involves a person being *hired* or *permitted* to find vulnerabilities in a system to improve security. Since there is no permission here, it cannot be ethical hacking.

(b) **Phishing:** This is a technique used to trick users into revealing sensitive information like passwords. Using someone's Wi-Fi doesn't necessarily involve trickery to gain info.

(c) **Unauthorized Access:** This term perfectly describes the situation. By connecting to a private Wi-Fi network without permission, the user is accessing a network resource they are not entitled to use.

(d) **Intellectual Property Theft:** This refers to stealing ideas, patents, or creative works. While you might use the Wi-Fi to steal IP, the act of connecting to the Wi-Fi itself is not IP theft.

Final Answer: This is an example of Unauthorized Access.

Answer: (C)

Q49.

Solution

Concept:

When working with data structures in Pandas, such as Series or DataFrames, it is often necessary to understand the memory footprint of the objects being processed. This is particularly important when dealing with large-scale datasets that might approach the limits of the system's available RAM. Pandas provides several attributes to inspect the size and memory usage of data. While the `.size` attribute tells you the number of elements, it does not account for the varying bit-depths or memory requirements of different data types (e.g., an integer versus a complex object). To get the actual technical memory consumption, a more specific attribute is required.

Solution:

(a) The question asks for the attribute that returns the "number of bytes" used by the data in a Pandas Series.

(b) **s.size:** This attribute returns the total number of elements in the Series (e.g., 5). It does not represent bytes.

(c) **s.nbytes:** This is the correct attribute. It returns the total number of bytes consumed by the elements in the underlying array. For example, if a Series has 5 integers (each 8 bytes), `nbytes` would return 40.

(d) **s.memory:** This is not a standard attribute. However, there is a method called `memory_usage()`, which provides more detailed info including index memory, but the property specifically for the data bytes is `nbytes`.

(e) **s.bytes:** This is not a valid attribute in the Pandas Series object.

(f) Therefore, `s.nbytes` is the standard way to retrieve the memory size of the data values in bytes.

Final Answer: The attribute is `s.nbytes`.

Answer: (A)

Q50.

Solution

Concept:

Data aggregation in SQL is the process of collecting data from multiple rows and grouping it based on one or more criteria to perform calculations. For example, a company might want to find the total sales per region or the average salary per department. To achieve this, SQL uses a specific clause that works in tandem with aggregate functions like `SUM()`, `AVG()`, `COUNT()`, and `MAX()`. This clause essentially partitions the table into "buckets" based on the unique values in the specified column, and then the aggregate function is applied to each bucket independently.

Solution:

(a) The question focuses on the keyword used to "group the result-set by one or more columns."

(b) **ORDER BY:** This clause is used exclusively for sorting the output in ascending or descending order; it does not perform any grouping logic for aggregates.

(c) **GROUP BY:** This is the correct SQL clause. It is used in a `SELECT` statement to collect data across multiple records and group the results by one or more columns. It is almost always used with aggregate functions.

(d) **SORT BY:** While used in some distributed SQL environments (like Hive), it is not standard SQL and is typically used for local sorting within a reducer.

(e) **CLUSTER BY:** This is also specific to Big Data SQL environments (like Hive/Spark) and relates to how data is distributed across nodes, not standard grouping for aggregates.

(f) Thus, the `GROUP BY` clause is the standard requirement for partitioning data into sets for aggregate calculation.

Final Answer: The keyword is GROUP BY.

Answer: (B)

Answer Key

Q	Ans	Q	Ans	Q	Ans	Q	Ans	Q	Ans
1	A	2	C	3	A	4	B	5	A
6	A	7	D	8	A	9	B	10	C
11	D	12	B	13	C	14	B	15	C
16	A	17	B	18	B	19	A	20	B
21	C	22	B	23	C	24	A	25	B
26	D	27	B	28	A	29	B	30	C
31	B	32	B	33	B	34	B	35	C
36	D	37	B	38	B	39	C	40	B
41	B	42	C	43	C	44	D	45	A
46	B	47	C	48	C	49	A	50	B