

Update Statement

① T/F UPDATE statement is a DML Statement.
True

② UPDATE statement is used to _____ alter or modify the data.

③ The syntax for UPDATE Statement is _____.

UPDATE tablename SET column_a = 'new-a', Column_b = 'new-b',
... column_n = 'new-n' Where predicate;
Optional for specificity

④ The syntax to change first name and last name to Lakshmi Katta where AUTHOR-ID = A2 is _____.

UPDATE AUTHOR set LASTNAME = 'KATTA', FIRSTNAME = 'LAKSHMI'
Where AUTHOR-ID = 'A2';

DELETE Statement

① T/F DELETE statement is DML Statement.
True

② The syntax for DELETE statement by a column_name is _____.

DELETE FROM tablename
Where column_name in ('property_a', 'property_b',, 'property_n');

③ The result of the code

DELETE FROM AUTHOR Where AUTHOR-ID in ('A2', 'A3');
is _____.

Deletes a row where AUTHOR-ID is either 'A2' or 'A3'.

④ T/F If you do not specify a where clause, all the rows will be deleted.

True

⑤ T/F You do not use DELETE to delete columns.

True

⑥ The general syntax for deleting columns is _____.

ALTER TABLE tablename DROP COLUMN column_name;

⑦ T/F MySQL, PostgreSQL have the same syntax for dropping multiple columns but SQL Server and Oracle database each have their own syntaxes.

True.

Relational Model

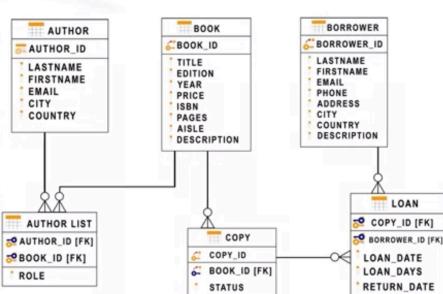
① The relational model is the most used data model because it allows for _____ data independence

② The three types of independence preserved are _____

1. logical independence
2. physical data independence
3. physical storage independence

③ An entity relationship (ER) data model is _____ an alternative to relational data model

④ This is an example of an ER (entity relationship diagram):



⑤ ERD is used to represent _____ entities called table and their relationships.

⑥ T/F The ER model is used as a tool to design relational databases.
True

⑦ In an ER model, an entity is an _____ object

⑧ T/F Entities in an ER model exists independently of any entities in the database.

⑨ The building blocks of an ER diagram are _____.

(2)

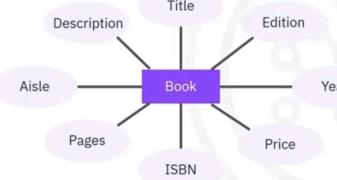
1. entities
2. attributes

ER Diagram

① An Entity can be _____.

Noun: person, place, or thing

(2)



In this diagram, the entity is _____

Book

(2)

In this diagram, the attributes are _____

Title, Edition, etc.

(3)

Attributes are _____

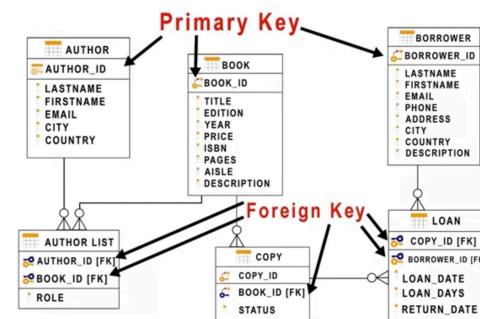
data elements that characterize the entity

Entity-Relationship Model

① Entity becomes a _____ in the database.
Table

② Attributes become the _____ in a table.
Columns

Primary keys and foreign keys



① T/F Each table is assigned a primary key.
True.

② The primary key uniquely identifies each _____ of a table.
tuple or row

③ The primary key prevents _____.

duplication of data

④ Foreign keys are _____.

primary keys defined in other tables

⑤ The foreign key helps create a _____ link between tables

DDL vs DML

① You can interchange row with _____ tuple

② You can interchange column with _____ attribute

③ DDL stands for _____ data definition language

④ DML stands for _____ data manipulation language

⑤ DDL statements are used to _____ database objects (such as tables).

1. define
2. change
3. drop

⑥ Common DDL statement types are _____

1. CREATE
2. ALTER
3. TRUNCATE
4. DROP

Common DDL statements

① CREATE is used for _____

1. creating tables
2. defining its columns

② ALTER is used for _____

1. Adding / dropping columns
2. modifying their datatypes

③ TRUNCATE is used for _____

deleting data in a table but not the table itself

④ DROP is used for _____

deleting tables

DML statements

- ① DML statements are used to _____
read and modify data
- ② Read and modify data with _____.
CRUD operations
- ③ CRUD stands for _____
Create, Read, Update & Delete rows

Common DML statements

- ① Common DML statements are _____

1. INSERT
2. SELECT
3. UPDATE
4. DELETE

- ② INSERT is used for _____

Inserting a row or several rows of data into a table.

- ③ SELECT is used for selecting data in a _____ rows

rows

- ④ UPDATE is used for editing _____

rows

- ⑤ DELETE is used for deleting data in a _____ rows

CREATE TABLE Statement

- ① The syntax for CREATE TABLE is _____.

`CREATE TABLE tablename (`

`column_1 datatype optional parameters,
 column_2 datatype,
 ...
 column_n datatype
)`

- ② A sample code:

```
CREATE TABLE provinces(  
    id CHAR(2) PRIMARY KEY NOT NULL;  
    name VARCHAR(24) NOT NULL;  
    population BIGINT  
)
```

- ③ The CHAR(2) tells us the datatype is character string for id and has a _____ fixed length of 2.

- ④ The varchar is a character string of a _____ variable length

- ⑤ VARCHAR(24) means the variable character length can be _____ up to 24.

ALTER TABLE Statement

- ① The syntax for ALTER TABLE is _____.

```
ALTER TABLE tablename  
ADD COLUMN column_1 datatype,  
...  
ADD COLUMN column_n  
)
```

- ② T/F The ALTER TABLE does not allow you to perform multiple operations (e.g. adding and modifying columns) in a single statement.

True.

- ③ The actions you can perform with ALTER TABLE _____.

1. Add or remove columns
2. Modify the data types of columns
3. Add or remove keys
4. Add or remove constraints

- ④ The syntax for modifying the datatype is _____.

```
ALTER TABLE tablename  
MODIFY column-name datatype;
```

- ⑤ T/F The CHAR datatype does not include '(', '+', '*' .

True

- ⑥ To delete a column, the syntax is _____.

```
ALTER TABLE tablename  
DROP COLUMN column-name;
```

- ⑦ The DROP COLUMN is known as a _____ clause

`SELECT column_name(s) FROM tablename`

- ⑧ The code for deleting a table is _____.

```
DROP TABLE tablename;
```

- ⑨ The code for deleting the data and not the table itself is _____.

```
TRUNCATE TABLE tablename
```

IMMEDIATE;

- ⑩ If IMMEDIATE is used to process the statement immediately and the action cannot be undone.

`SELECT column_name(s) FROM tablename
ORDER BY column-name DESC;`

Retrieving Rows of Data: String Patterns

- ① If you don't know which value to specify in the predicate, (e.g. not knowing the exact value to search for), we can do _____.

String patterns

- ② Suppose we know the first name of the author begins with 'R'. The code for using string pattern is _____.

```
SELECT * FROM tablename  
WHERE column LIKE 'R%';
```

- ③ The '%' serves as a _____ placeholder for other characters.

- ④ T/F You can put '%' before, middle, and after characters.

True

Range

- ⑤ For numbers, you can use the operators _____.

1. ' $>=$ ' ... ' $<=$ '
2. 'BETWEEN' ... 'AND'

- ⑥ The syntax to use range are _____.

```
1. SELECT column_name(s) FROM tablename  
WHERE column_name BETWEEN X AND Y;  
(X,Y ∈ R.)
```

2. `SELECT column_name(s) FROM tablename
WHERE column_name >= X AND column_name <= Y;
(X,Y ∈ R).`

- ⑦ T/F Most prefer the first approach because it is easier and quicker to write.

Time

A Set of Values

- ① If you cannot group numbers into a range, you can use _____ in operator

- ② The syntax for in operator is _____.

`SELECT column_name(s) FROM tablename
WHERE column_name IN (value_1, value_2, ..., value_n);`

- ③ The in operator takes a list of expressions to _____ compare against

Sorting Results Set: ORDER BY Clause

- ① The ORDER BY clause is used in a query to sort the result set by _____ a specified column

- ② The syntax for ORDER BY is _____.

`SELECT column_name(s) FROM tablename
ORDER BY column-name;`

- ③ By default, the result set is sorted in _____ order.

ascending

- ④ The syntax for sorting the result set in descending order is _____.

`SELECT column_name(s) FROM tablename
ORDER BY column-name DESC;`

Column Sequence Number

- ① Another way to specify sort column is to indicate the _____ column sequence number

- ② Suppose the second column in the table was named pages. The code to sort the result set using column sequence number for pages is _____.

`SELECT column_name(s) FROM tablename
ORDER BY 2;`

Grouping Result Sets: Eliminating Duplicates, DISTINCT Clause

- ① The DISTINCT clause removes duplicates within a _____ column

- ② The syntax for using the DISTINCT clause is _____.

`SELECT DISTINCT (column-name)
FROM tablename;`

③ To count the number of unique values of a column, use _____.

COUNT function & GROUP BY clause

④ The syntax for GROUP BY clause is _____.

SELECT column_name, COUNT(column_name)
FROM tablename GROUP BY column_name;

⑤ The GROUP BY clause groups a result into _____ subsets that has matching values for one or more columns.

⑥ Example of the code and output using GROUP BY:

| db2 => select country, count(country) from Author GROUP BY country | |
|---|---|
| Country | 2 |
| AU | 1 |
| BR | 1 |
| CA | 3 |
| CN | 6 |
| IN | 6 |
| RO | 3 |
| 6 record(s) selected. | |

⑦ T/F First the 'country' is grouped. Then they are counted.

True

⑧ The number '2' is displayed as a column name because the column name is _____ not directly available in the table

⑨ The second column in the result set was calculated by the _____ count function

⑩ We can assign a column name to the result set with this code _____.

SELECT column_name, COUNT(column_name)
AS Column-name_1 FROM tablename
GROUP BY column_name;

⑪ AS is a _____ keyword

Condition on GROUP BY Clause: HAVING Clause

⑫ T/F The WHERE clause works with the entire result set, but the HAVING clause only works with GROUP BY clause.

True

⑬ The syntax for using HAVING clause is _____.

SELECT column_name, COUNT(column_name)
AS Column-name_1 FROM tablename
GROUP BY column_name HAVING predicate;

Built-In Database Functions

⑭ Built-in functions allows you to perform operations on data right within _____ the database itself

⑮ T/F Using database functions reduces the amount of data that needs to be retrieved from the database.

True

⑯ T/F Database functions can significantly reduce network traffic and use of bandwidth.

True

⑰ T/F Database functions can speed up data processing because rather than first retrieving the data into your app and then executing functions on retrieved data.

True

⑱ T/F You can create your own functions called user-defined functions in the database.

True

⑲ The syntax for using the average function is _____.

SELECT AVG(column_name) FROM tablename;

⑳ T/F You can perform mathematical operations between columns and apply aggregate functions on them.

True

㉑ Suppose we have the COST, QUANTITY, and ANIMAL column. We can calculate the average cost per dog by doing _____.

SELECT AVG(COST / QUANTITY) FROM tablename
WHERE ANIMAL = 'Dog';

Aggregate or Column Functions

㉒ Some examples of aggregate or column functions are _____.

1. SUM() 3. MAX()
2. MIN() 4. AVG()

㉓ The sum function adds up all the values in a _____ column.

㉔ Suppose the tablename was PETRESCUE. The syntax to use SUM() is _____.

SELECT SUM(column_name) FROM tablename;

㉕ T/F When you use an aggregate function, the column in the result set is given a number by default.

True

Column Aliases

㉖ T/F It is possible to explicitly name the resulting column.

True

㉗ Suppose we want to call the output column SUM_OF_COST. The code for this is _____.

SELECT SUM(COST) AS SUM_OF_COST
FROM PETRESCUE;

㉘ The SUM_OF_COST becomes a _____ in the result set.

the column_name (instead of the default number)

MAX & MIN

㉙ The MAX and min returns the max and min value of the _____ respectively.

column

㉚ The syntax for MAX function is _____.

SELECT MAX(column_name) FROM PETRESCUE;

㉛ T/F Aggregate functions can be applied on a subset of data instead of an entire column.

True

㉜ Suppose we want to get the minimum value of a column with a condition within a table. The syntax is _____.

SELECT MIN(column_name) FROM tablename

WHERE <Condition>;

Average

㉝ T/F You can calculate the mean of a column using the average function.

True

㉞ The syntax for using the average function is _____.

SELECT AVG(column_name) FROM tablename;

㉟ T/F You can perform mathematical operations between columns and apply aggregate functions on them.

True

㉟ Suppose we have the COST, QUANTITY, and ANIMAL column. We can calculate the average cost per dog by doing _____.

SELECT AVG(COST / QUANTITY) FROM tablename
WHERE ANIMAL = 'Dog';

Scalar and String Functions

㉟ Scalar functions perform operations on _____ every input value.

㉟ Some examples of scalar and string functions are _____.

1. ROUND() 3. UCASE
2. LENGTH() 4. LCASE

㉟ The following syntax

SELECT ROUND(column_name)

FROM tablename;
will round every value of the column to the nearest _____ integer

㉟ The string functions apply to _____.

1. Char
2. Varchar

The Syntax

SELECT LENGTH(ANIMAL)

FROM tablename;
will retrieve the length of _____ each char/varchar in ANIMAL

UCASE, LCASE

㉟ The syntax to capitalize every value in a column is _____.

SELECT UCASE(column_name)

FROM tablename;

㉟ T/F You can use the UCASE/LCASE function in a WHERE clause.

True

㉟ Suppose you want to find all cat items. You are unsure that the ANIMAL column stores this value as. A solution is to _____.

㉟ SELECT * FROM tablename
WHERE LCASE(column_name) = 'cat';

㉟ T/F You can have one function operate on the output of another function.

True

㉟ Suppose you want the unique values from a column regardless of their cases. One solution is _____.

㉟ SELECT DISTINCT UCASE(column_name)

FROM tablename;

③ An example code for a derived table is _____.

```
SELECT * FROM
  (SELECT column_1, column_2, column_3
   FROM tablename) AS derived_table_name;
```

④ Derived tables can be useful for _____.

1. Working with multiple tables
2. Doing joins

Multiple Tables

Given these tables:

| EMPLOYEES: | | | | | | | | | | |
|------------|--------|--------|--------|------------|-----|------------------------|--------|--------|------------|--------|
| EMP_ID | F_NAME | L_NAME | SSN | B_DATE | SEX | ADDRESS | JOB_ID | SALARY | MANAGER_ID | DEP_ID |
| E1001 | John | Thomas | 123456 | 1976-01-09 | M | 5631 Rice, OakPark,IL | 100 | 100000 | 30001 | 2 |
| E1002 | Alice | James | 123457 | 1972-07-31 | F | 980 Berry Ln, Elgin,IL | 200 | 80000 | 30002 | 5 |
| E1003 | Steve | Wells | 123458 | 1980-08-10 | M | 291 Springs, Gary,IL | 300 | 50000 | 30002 | 5 |

| DEPARTMENTS: | | | |
|--------------|----------------------|------------|--------|
| DEPT_ID_DEP | DEP_NAME | MANAGER_ID | LOC_ID |
| 5 | Software Development | 30002 | L0002 |
| 7 | Design Team | 30003 | L0003 |

① Suppose we want to retrieve only the employee records from the Employees for which a Department ID exists. We can use a _____.

1. Multiple tables
2. IN operator
3. sub-query

② The code retrieve employees from the Employees for which a department ID exists is _____.

```
SELECT * FROM employees
WHERE DEP_ID IN
  (SELECT DEPT_ID_DEP FROM departments);
```

③ Suppose we want to retrieve only the list of employees from a specific location. Assume the employees table does not have location info, but the departments table has a column called Location-ID. A sample syntax would be _____.

```
SELECT * FROM tablename1 WHERE
  column_1 IN
    ( SELECT column_1_table2 FROM tablename2
      WHERE predicate);
```

④ A sample code and the results for using two tables to refine a search using sub-queries is _____.

Query:
`select * from employees where DEP_ID IN
 (select DEPT_ID_DEP from departments where LOC_ID = 'L0002');`

Result:

| EMP_ID | F_NAME | L_NAME | SSN | B_DATE | SEX | ADDRESS | JOB_ID | SALARY | MANAGER_ID | DEP_ID |
|--------|---------|--------|--------|-----------|-----|-------------------------------|--------|--------|------------|--------|
| E1002 | Alice | James | 123457 | 7/31/1977 | F | 980 Berry Ln, Elgin,IL | 200 | 80000 | 30002 | 5 |
| E1003 | Steve | Wells | 123458 | 8/10/1998 | M | 291 Springs, Gary,IL | 300 | 50000 | 30002 | 5 |
| E1004 | Santosh | Kumar | 123459 | 7/20/1998 | M | 511 Aurora Av, Aurora,IL | 400 | 60000 | 30004 | 5 |
| E1010 | Ann | Jacob | 123415 | 3/30/1998 | F | 111 Britany Springs,Elgin,220 | 70000 | 30004 | 5 | |

⑤ Suppose we want to retrieve the department ID and department name for employees who earn more than \$70k. We can get _____.

1. A sub-query on the Employees table to satisfy salary criteria
2. A matching department info by feeding 1. as input to an outer query on Departments table

⑥ A sample code for retrieving department ID and department name for employees who earned more than \$70k is _____.

```
SELECT DEPT_ID_DEP, DEP_NAME
FROM departments
WHERE DEPT_ID_DEP IN
  (SELECT DEP_ID FROM employees
   WHERE SALARY > 70000);
```

Access Multiple Tables with Implicit Joins

① T/F An Implicit Join is done by specifying 2 tables in the FROM clause. True

② The syntax for implicit join is _____.

```
SELECT * FROM
  tablename1, tablename2;
```

③ T/F The result of an implicit join is a full join. True

④ A full join is when _____ is joined with _____.

- every row in the first table
- every row in the second table

Operands to limit Implicit Join

① Suppose we want the result set with matching department ID. A sample code is _____.

```
SELECT * FROM employees, departments
WHERE employees.DEP_ID = departments.DEP_ID_DEP;
```

Qualify Column Name

② To fully qualify the column name, we prefix the name of the column with the name of the _____ table.

③ T/F It is possible that different tables can have same column names that are exactly the same. True

④ For table names that are long, you can use shorter _____ alias

⑤ An example code for shorter alias is _____.

```
SELECT * FROM employee E, departments D
WHERE E.DEP_ID = D.DEP_ID_DEP;
```

⑥ We can access multiple table name in these ways _____.

1. SELECT EMP_ID, DEP_NAME
 FROM employees E, departments D
 WHERE E.DEP_ID = D.DEP_ID_DEP;
2. SELECT E.EMP_ID, D.DEP_NAME
 FROM employees E, departments D
 WHERE E.DEP_ID = D.DEP_ID_DEP;

Connecting Databases using Python

⑦ T/F Python supports relational database systems. True

⑧ T/F Writing Python code to access database is made easier by the presence of the Python database API. True.

⑨ The Python Database API can be interchanged with _____.

⑩ Some examples of notebook interfaces include _____.

1. Mathematica notebook
2. Maple worksheet
3. Matlab notebook
4. iPython Jupyter
5. R Markdown
6. Apache Zeppelin
7. Apache Spark Notebook
8. Databricks iCloud

⑪ Some of the features of the Jupyter notebook include _____.

1. Supports 40+ languages including Python, R, Julia, and Scala.
2. Jupyter notebook can be shared via email, Dropbox, GitHub, and Jupyter notebook viewer
3. Jupyter notebooks can produce images, videos, LaTeX, and customized types.
4. Leverage big data tools such as Apache Spark from Python, R, Scala and explore data using pandas, scikit-learn, ggplot2, and TensorFlow.

Accessing databases using Python

⑫ The Python code connects to the database using _____ API calls.

SQL API

⑬ API stands for _____.

⑭ Application Programming Interface

⑮ API is a _____ set of functions

⑯ You call API so that you can get _____ access to some type of service

⑰ SQL API consists of _____ library function calls as an API for DBMS

⑱ An application program calls functions in API to pass SQL statements to DBMS

APIs used in SQL-based DBMS

⑲ The applications or database and their respective SQL API are _____.

⑳ MySQL : _____ MySQL C API / Python

㉑ PostgreSQL : _____ psycopg2

㉒ IBM DB2 : _____ ibm-db

㉓ SQL Server : _____ dblib API

㉔ Database access for Microsoft Windows OS : _____ ODBC

㉕ Oracle : _____ OCI

㉖ Java : _____ JDBC

㉗ MongoDB : _____ PyMongo

Why Code with DB-API?

- ① DB-API is _____.
Python's standard API for accessing relational databases
② DB-API allows a _____ single program that work with multiple databases

- ③ DB-API allows for _____ (5)

1. Easy implementation and understanding
2. similarity between Python modules used to access databases
3. Consistency
4. portability across databases
5. broad reach of database connectivity from Python

Python DB API : 2 Objects
Connection & Query Objects

- ① Connection objects are used to _____ (2)

1. Connect to a database
2. manage your transactions.

- ② Cursor objects are used to _____ (3)

1. run queries
2. Scroll/Scan through result set.
3. Retrieve results

- ③ The DB-API includes a _____.

Connect constructor.

- ④ The connect constructor is used for creating a connection to the _____ database.

- ⑤ The connect constructor returns a _____.

Connection object

- ⑥ A connection object is then used by the _____ various connection methods

Connection Methods

- ① The connection methods used are _____ (4)

1. cursor method
2. commit method
3. rollback method
4. close method

- ② The cursor method, :cursor(),

returns a _____ new cursor object using the connection

- ③ The commit method, .commit(), is used to commit _____ any pending transaction to the database.

- ④ The rollback method, .rollback(), is used to roll back to _____ the start of any pending transaction

- ⑤ The close method, .close(), is used to close _____ a database connection.

Cursor Methods

- ① The cursor methods used are _____.

- | | |
|--------------------|----------------|
| 1. Callproc method | .callproc() |
| 2. execute method | .execute() |
| 3. executemany " | .executemany() |
| 4. fetchone " | .fetchone() |
| 5. fetchmany " | .fetchmany() |
| 6. fetchall " | .fetchall() |
| 7. nextset " | .nextset() |
| 8. arraysize " | .arraysize() |
| 9. close " | .close() |

- ② An object is an instance of a _____ class

- ③ A method is a _____ associated with a class or object.

function

- ④ T/F All methods are objects, but not all objects are methods.

True

- ⑤ The cursor methods are used to _____

Manage the content of a fetch operation

- ⑥ Cursors created from the same connection are _____.

linked.

- ⑦ T/F Any changes done to the database by a cursor are immediately visible by the other cursor.

True

- ⑧ A database cursor is a _____.

Control structure.

- ⑨ A database cursor enables _____ traversing over records in a database

- ⑩ A database cursor behaves like a _____.
1. file name
 2. file handle in a programming language

- ⑪ A program opens a file to access its contents. It opens a cursor to gain _____ access to the query results

- ⑫ A program closes a file to end its access. It closes a cursor to end access to the _____ query result

- ⑬ A file handle keeps track of the program's current position. A cursor keeps track of the program's _____ within the query results.

Current position

Syntax for using DB-API in Python

- ① You import the database module by using the _____ - Connect API from that module

- ⑭ The syntax is _____.

from dbmodule import connect

- ⑮ You open up the connection to the database by _____.

1. Using the connect constructor
2. pass in parameters
3. The connect constructor returns a connection object

- ③ The full cycle of writing code using DB-API is _____ (5, in order)

1. Importing database module
2. Create Connection object
3. Create Cursor object
4. run queries
5. free resources

- ④ The syntax for importing then creating connection object is _____.

from dbmodule import connect

connection = connect(

 database="databaseName",
 user="username",
 password="password")

- ⑤ You create the cursor object on the _____ object.

connection

- ⑥ The cursor object is used to _____

(2, in order)

1. run queries
2. fetch results

- ⑦ The syntax for running and fetching queries is _____.

cursor.execute(

 SELECT * FROM tablename
 results = cursor.fetchall())

- ⑧ After the system is done running the queries, apply the _____

Close method

- ⑨ The close method is used to _____

free all resources

- ⑩ The syntax to avoid unused cursor and connection from taking up resources is _____.

cursor.close()
connection.close()

Analyzing Data with Python

- ① Suppose we want to create a database table on an SQL server.

T/F We can use SQLite3.

True

- ② SQLite3 is an _____.

in process Python library

- ③ SQLite3 implements a _____.

1. self-contained

2. severless

3. zero configuration

transactional SQL database engine

Load CSV to SQLite3 with Panda

- ① The syntax for loading CSV to SQLite3 is _____.

import pandas as pd

import sqlite3

data = pd.read_csv('file-path.csv')

conn = sqlite3.connect('databaseName.db')

e.g. conn = sqlite3.connect('McDonalds.db')

data.to_sql('tablename', conn)

e.g. data.to_sql('MCDONALDS_NUTRITION', conn)

Using Pandas to Retrieve Data from Database Tables

① T/F To load a table into a DataFrame, you use the function `pd.read_sql()`. True

- ② The syntax for loading a table into a DataFrame is _____.
- ```
df = pd.read_sql(
 "SELECT * FROM tablename",
 conn)
print(df)
```

Learn About Data using Pandas:

### Using Categorical Scatter Plots

① T/F `df.head()` returns the first few rows of the DataFrame.

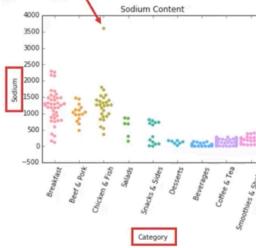
② The `df.describe(include='all')` allows you to see statistics about each \_\_\_\_\_ of the table, column

| Category | Name         | Beverage | Sodium     | Calories   | From Fat   | Total Fat  | Total Fat (% Daily Value) | Saturated Fat | Trans Fat  | Carbohydrates | Cholesterol | Protein    | Fiber      | Sugars     | Calories From Sugars | Vitamin D  | Calcium    | Iron       |
|----------|--------------|----------|------------|------------|------------|------------|---------------------------|---------------|------------|---------------|-------------|------------|------------|------------|----------------------|------------|------------|------------|
| Count    | 260          | 0.000000 | 260.000000 | 260.000000 | 260.000000 | 260.000000 | 260.000000                | 260.000000    | 260.000000 | 260.000000    | 260.000000  | 260.000000 | 260.000000 | 260.000000 | 260.000000           | 260.000000 | 260.000000 | 260.000000 |
| unique   | 9            | 0.000000 | 9.000000   | 9.000000   | 9.000000   | 9.000000   | 9.000000                  | 9.000000      | 9.000000   | 9.000000      | 9.000000    | 9.000000   | 9.000000   | 9.000000   | 9.000000             | 9.000000   | 9.000000   | 9.000000   |
| top      | Coffee & Tea | 0.000000 | 140.000000 | 140.000000 | 140.000000 | 140.000000 | 140.000000                | 140.000000    | 140.000000 | 140.000000    | 140.000000  | 140.000000 | 140.000000 | 140.000000 | 140.000000           | 140.000000 | 140.000000 | 140.000000 |
| freq     | 10           | 0.000000 | 140.000000 | 140.000000 | 140.000000 | 140.000000 | 140.000000                | 140.000000    | 140.000000 | 140.000000    | 140.000000  | 140.000000 | 140.000000 | 140.000000 | 140.000000           | 140.000000 | 140.000000 | 140.000000 |
| mean     | Nan          | Nan      | 365.203217 | 127.095162 | 14.165388  | 21.815588  | 0.0707                    | 0.065398      | 0.020544   | 47.348154     | 1.0         | Nan        | Nan        | Nan        | Nan                  | Nan        | Nan        | Nan        |
| std      | Nan          | Nan      | 320.000000 | 100.000000 | 100.000000 | 100.000000 | 0.0707                    | 0.065398      | 0.020544   | 47.348154     | 1.0         | Nan        | Nan        | Nan        | Nan                  | Nan        | Nan        | Nan        |
| 25%      | Nan          | Nan      | 230.000000 | 50.000000  | 5.000000   | 2.000000   | 0.050000                  | 0.020000      | 0.005000   | 3.000000      | 0.0         | Nan        | Nan        | Nan        | Nan                  | Nan        | Nan        | Nan        |
| 50%      | Nan          | Nan      | 210.000000 | 20.000000  | 2.000000   | 2.000000   | 0.050000                  | 0.020000      | 0.005000   | 3.000000      | 0.0         | Nan        | Nan        | Nan        | Nan                  | Nan        | Nan        | Nan        |
| 75%      | Nan          | Nan      | 200.000000 | 20.000000  | 2.000000   | 2.000000   | 0.050000                  | 0.020000      | 0.005000   | 3.000000      | 0.0         | Nan        | Nan        | Nan        | Nan                  | Nan        | Nan        | Nan        |
| max      | Nan          | Nan      | 600.000000 | 200.000000 | 20.000000  | 48.000000  | 0.080000                  | 0.030000      | 0.010000   | 40.000000     | 0.0         | Nan        | Nan        | Nan        | Nan                  | Nan        | Nan        | Nan        |

③ A categorical scatterplot shows \_\_\_\_\_ for different items by category.

the values

④ An example of a categorical scatterplot is \_\_\_\_\_.



⑤ A code for creating categorical scatterplots is \_\_\_\_\_.

```
1 import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns

2 plot = sns.swarmplot(x="Category",
 y="Sodium", data=df)
plt.setp(plot.get_xticklabels(), rotation=90)
plt.title("Sodium Content")
plt.show()
```

### Basic Data Analysis

① To get the count, mean, std, min, quartile in the 25%, 50%, 75%, and max, we use the function \_\_\_\_\_.

`describe()`

② An example of using `describe` is \_\_\_\_\_.

```
In [17]: df['Sodium'].describe()
Out[17]: count 260.000000
mean 495.750000
std 577.026323
min 0.000000
25% 107.500000
50% 190.000000
75% 865.000000
max 3600.000000
Name: Sodium, dtype: float64
```

③ T/F The results shown from applying the described() function is known as a summary of statistics. True

### Finding the Row of the Max Value

① To find the row of the max value of a column, we can use the function \_\_\_\_\_.

② A code for finding the row with max value of a column is \_\_\_\_\_.

```
df['column_name'].idxmax()
```

③ T/F A way to find the item with the maximum value is through the `at` function.

True

④ A code to find an associated column of the maximum value is \_\_\_\_\_.

```
row_max_value = df['column_name'].idxmax()
```

```
df.at[row_max_value, 'column_name']
```

⑤ An example code of the routine above is \_\_\_\_\_.

```
1 df['Sodium'].idxmax()
out 82
2 df.at[82, 'Item']
out 'Chicken McNuggets (40 pieces)'
```

### Data Exploration

① For initial data exploration, \_\_\_\_\_ is very useful.

Visualization

② Visualizations are useful for \_\_\_\_\_ understanding.

1. relationships
2. patterns, and
3. outliers

in the data.

### Making Scatter plots (with Seaborn)

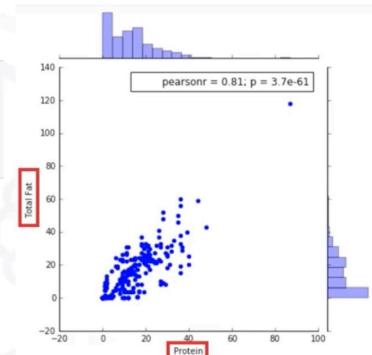
① To make a scatter plot using seaborn, we can use the \_\_\_\_\_ function.

`joinplot`

② A code for scatterplot is \_\_\_\_\_.

```
1 import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns

2 plot = sns.joinplot(x="Protein", y="Total Fat", data=df)
plot.show()
```



③ A correlation is \_\_\_\_\_.

a measure

Correlation measures the association between two variables.

④ T/F Correlation has a value between -1 and 1. True

⑤ On the top and to the right are \_\_\_\_\_ histograms

⑥ The top represents \_\_\_\_\_ protein

⑦ The right represents \_\_\_\_\_ Total fat

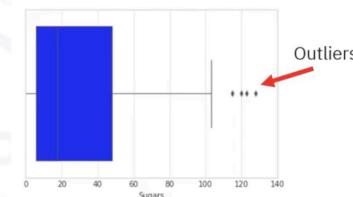
### Box Plots

① To indicate the distribution of one or more variable, we can use \_\_\_\_\_ box plots.

② A sample code for boxplot is \_\_\_\_\_.

```
1: import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns

2: plot = sns.set_style("whitegrid")
ax = sns.boxplot(x=df["Sugars"])
plot.show()
```



### Seaborn and Matplotlib

① Seaborn is \_\_\_\_\_.

a python library

② Seaborn is used for \_\_\_\_\_ data visualization.

③ Seaborn is built on top of \_\_\_\_\_.

④ Matplotlib is a \_\_\_\_\_.

python visualization library