Data Engineer Intern Assessment Answers

# Explanation of the Query

# Query Explanation

The provided SQL query uses a Common Table Expression (CTE) to generate a ranked list of currency exchange rates. It then determines the effective and expiration dates for each exchange rate based on the dates of the subsequent rates.

# Steps Involved

1. Common Table Expression (CTE) Definition:  
- The CTE (`WITH cte AS (...)`) selects columns such as `currencytype`, `fromcurrency`, `tocurrency`, `date`, and `currencyrate` from the `curr` table.  
- It uses `ROW\_NUMBER()` to rank the records within each group of `(currencytype, fromcurrency, tocurrency)` based on the `date`, in descending order (latest first).

2. Main Query Logic:  
- Selects data from the CTE with aliases for columns (`ExchangeRateType`, `FromCurrency`, `ToCurrency`, `EffectiveDate`, `ExchangeRate`).  
- Uses a `LEFT JOIN` on the CTE itself to get the next row in the sequence (`rownum = t.rownum + 1`), matching currency type and combination.  
- Sets the `ExpirationDate` using a `CASE` statement: if the next date is `NULL`, defaults to '2999-01-01'; otherwise, uses the next date.

3. Filtering:  
- The `WHERE` clause restricts results to rows where `currencytype = 'MAR'.

# Concepts Involved

- CTE simplifies complex queries.  
- Window Functions (`ROW\_NUMBER()`) rank rows within partitions.  
- Self-Joins get the next row in sequence for expiration dates.  
- Conditional Logic (`CASE` statement) handles missing values.

# Purpose

The query generates a timeline of currency exchange rates, providing valid time ranges for each rate.

# 2. SQL Query to Find the Average Salary Excluding Highest and Lowest Salaries

Here’s how you can write an SQL query to find the average salary from an employee table without including the highest and lowest salaries:

SELECT AVG(salary)   
 FROM employee   
 WHERE salary NOT IN (SELECT MAX(salary) FROM employee)   
 AND salary NOT IN (SELECT MIN(salary) FROM employee);

This query calculates the average salary by filtering out the highest and lowest salaries from the data using subqueries.

# 3. Python Script for Reading a File with Error Handling

The following Python script opens a file named 'data.csv' and reads its contents, including error handling to manage cases where the file may not exist or may be unreadable:

try:  
 with open('data.csv', 'r') as file:  
 content = file.read()  
 print(content)  
 except FileNotFoundError:  
 print("The file 'data.csv' does not exist.")  
 except IOError:  
 print("The file could not be read.")

In this script:

- FileNotFoundError handles cases where the file is not found.  
- IOError manages other issues related to file reading, such as permission problems.

# 4. Explanation of Environmental Variables and Virtual Environments in Python

- Environmental Variables: These are variables that are defined outside of the Python program and affect the way running processes behave. In Python, they can be accessed using the os module.  
  
- Setting Up a Virtual Environment:  
 - You can create a virtual environment using the command: python -m venv myenv  
 - Activate it using: myenv\Scripts\activate (Windows) or source myenv/bin/activate (Linux/Mac).  
  
- Why Create a Virtual Environment?:  
 - To isolate dependencies for different projects, avoiding conflicts between packages and maintaining a consistent development environment.

# 5. Parsing and Loading a JSON Response into a Database

If tasked with parsing a JSON response and loading it into a database, it is important to consider whether to use a single table or multiple tables:  
- Multiple Tables Approach: If the JSON has nested structures or related entities (like users and their orders), it’s better to normalize the data across multiple tables.  
- Single Table Approach: If the JSON is flat with simple key-value pairs, it can be stored in a single table.  
  
For example, if the JSON response contains user information with a list of addresses, the tables could be structured as:  
- users (user\_id as the primary key, name, email)  
- addresses (address\_id as the primary key, user\_id as the foreign key, address, city)