3.2.8 Home Assignment (for Submission)

Tasks:

1. Using the DVD rental dataset, provide the SQL code and the output for this query criterion

   A. Using the DVD rental dataset, provide the SQL code and the output for this query criterion. a Provide an Excel sheet report of the total sum of payments made by each customer in the payments table. Present it in ascending order. There should be four columns: customer_id`, first_name`, last_name`, [SEP] and sum.

   SQL QUERY ANSWER:

```
SELECT customer.customer_id, customer.first_name, customer.last_name,
SUM(payment.amount) AS total_payment
FROM customer
INNER JOIN payment
ON payment.customer_id = customer.customer_id
GROUP BY customer.customer_id, customer.first_name, customer.last_name
ORDER BY total_payment ASC
```

   OUTPUT:
   CSV File Link: Total Sum of Payments

   B. Provide the SQL syntax and output of customer IDs having an average payment amount of less than 3.

   SQL QUERY ANSWER:

```
SELECT customer.customer_id, AVG(payment.amount) AS average_payment
FROM customer
INNER JOIN payment
ON payment.customer_id = customer.customer_id
GROUP BY customer.customer_id
HAVING AVG(payment.amount) < 3
ORDER BY average_payment ASC
```

   OUTPUT:
   CSV File Link: List of Payments Less Than 3

Data Output | Messages | Notifications

	customer_id [PK] integer	average_payment numeric
1	288	2.7794736842105263
2	557	2.899090909090909
3	248	2.913076923076923
4	252	2.94
5	586	2.99
6	283	2.99

2. Why would these SQL syntaxes fail?

   A. SELECT first_name, last_name, district FROM customer INNER JOIN address ON address_id = address_id.

   ANSWER:
   The "address id" column name is used in the ON clause of the query, but the table from which it should be retrieved is not specified, hence this SQL syntax will fail. Due to the fact that they are connected on that column, both tables (customer and address) include a column called "address id." So, SQL is unable to determine which table should be used for this column in the ON clause.

   The "address id" column's table name or alias must be specified in the ON clause to correct this. The amended syntax is as follows:

   District: SELECT first name, last name
   from customer
   Address for INNER JOIN
   customer.address id = address.address id

   With this approach, the "address id" column to utilize for the join condition is specified, and the query should succeed.

B. SELECT customer_id, SUM(amount3 FROM payment GROUP BY customer_id HAVING amount > 100

   ANSWER:
   The column "amount" that is mentioned in the HAVING clause should actually be SUM(amount), since it is not included in the SELECT statement.

3. Answer the following business scenario questions. Provide the SQL code and the output.

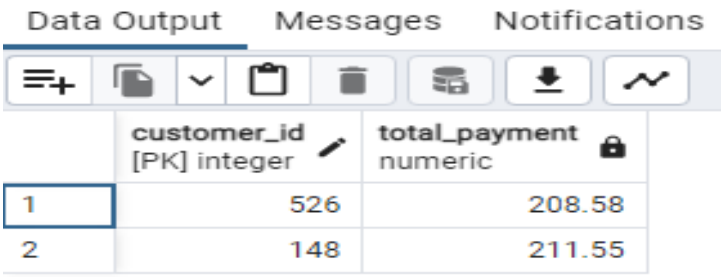
A. One of the criteria for a customer to become part of the customer loyalty program is that they need to have an accumulated payment made to the DVD rental store of at least $200 or more. As a data analyst, you need to provide a list of customer IDs with a total accumulated payment of at least $200 or more.

SQL QUERY ANSWER:

```
SELECT customer.customer_id, SUM(payment.amount) AS total_payment
FROM customer
INNER JOIN payment
ON payment.customer_id = customer.customer_id
GROUP BY customer.customer_id
HAVING SUM(payment.amount) >= 200
ORDER BY total_payment ASC
```

OUTPUT:

CSV File Link: List of customer with 200 Dollars or more payment



Data Output | Messages | Notifications

	customer_id [PK] integer	total_payment numeric
1	526	208.58
2	148	211.55