

Title : 7 to 10 - sql challenge - spark sql

7. Which item was purchased just before the customer became a member?

Code :

```
val bfMemberPurchaseDF = salesMembersTableDF.join(menuDataFrame,
salesDataFrame("product_id") === menuDataFrame("product_id")&&
salesDataFrame("order_date") < membersDataFrame("join_date"))
.groupBy(salesDataFrame("customer_id"),menuDataFrame("product_name"))
.agg(first("product_name").alias("before_purchase"))
.orderBy(salesDataFrame("customer_id"))
bfMemberPurchaseDF.show()
```

OutPut:

```
+-----+-----+-----+
|customer_id|product_name|before_purchase|
+-----+-----+-----+
|      B|    curry|    curry|
+-----+-----+-----+
```

-- 8. What is the total items and amount spent for each member before they became a member?

Code :

```
val bfCustomeSpendDF = salesMembersTableDF.join(menuDataFrame,
salesDataFrame("product_id") === menuDataFrame("product_id")
&&(salesDataFrame("order_date") < membersDataFrame("join_date")))
.groupBy(salesDataFrame("customer_id"))
.agg(sum(menuDataFrame("price")))
bfCustomeSpendDF.show()
```

Output:

```
+-----+
|customer_id|sum(price)|
+-----+
```

+-----+-----+

| B| 30|

+-----+-----+

-- 9. If each \$1 spent equates to 10 points and sushi has a 2x points multiplier - how many points would each customer have?

Code: PyCham Spark code

```
saleDataFrame.createOrReplaceTempView("saleTable")
memberDataFrame.createOrReplaceTempView("memberTable")
menuDataFrame.createOrReplaceTempView("menuTable")

purchasePointsDF = spark.sql(
    """WITH Points AS (
        SELECT s.customer_id,
            SUM(
                CASE
                    WHEN m.product_name = 'sushi' THEN 20 * m.price
                    ELSE 10 * m.price
                END
            ) AS no_of_points
        FROM saleTable s
        INNER JOIN menuTable m ON s.product_id = m.product_id
        GROUP BY s.customer_id
    )
    SELECT customer_id, COALESCE(no_of_points, 0) AS total_points
    FROM
        Points
    """
    .strip()
)
purchasePointsDF.show()
```

OutPut:

+-----+-----+

|customer_id|total_points|

+-----+-----+

| B| 540|

| C| 360|

| A| 660|

+-----+-----+

-- 10. In the first week after a customer joins the program (including their join date) they earn 2x points on all items, not just sushi - how many points do customer A and B have at the end of January?

Code:

```
# how many points do customer A and B have at the end of January?
janMonthPointsDF = spark.sql(
    """SELECT s.customer_id,
           SUM(
               CASE
                   WHEN s.order_date >= mb.join_date AND s.order_date <
DATE_ADD(mb.join_date, 7) THEN 20 * m.price  -- First week after joining
                   ELSE 10 * m.price
               END
           ) AS total_points
    FROM saleTable s
    INNER JOIN menuTable m ON s.product_id = m.product_id
    INNER JOIN memberTable mb ON s.customer_id = mb.customer_id
    WHERE s.order_date <= '2021-01-31'  -- End of January
    GROUP BY s.customer_id
    """
)
janMonthPointsDF.show()
```

OutPut:

```
+-----+-----+
|customer_id|total_points|
+-----+-----+
|      B|      540|
+-----+-----+
```