CS561 – Programming Assignment 2

Due Dates: Sec. A-12/8/2020 (Tue.) & Sec. B-12/10/2020 (Thu.)

***Objectives*:**

- You will continue with expressing "complex" OLAP queries in SQL. The reports below are similar in nature with the reports from the assignment #1; however, there are two main differences between the two: (1) the new reports will require aggregation "outside" the groups (in assignment #1, all of the aggregates were computed for the rows within the groups); (2) some of the aggregates in the new reports will be computed based on other aggregates of the same reports – they are known as "dependent aggregates".

***Description*:**

- Generate 4 separate reports based on the following queries (one report for query #1, one for query #2, one for query #3 and another for query #4):

  1. For each *customer*, *product*, *month* and *state* combination, compute (1) the customer's average sale of this product for the given month and state, (2) the customer's average sale for the given month and state, but for _all other products_ (3) the customer's average sale for the given product and state, but for _all other months_ and (4) the average sale of the product and the month but for _all other states_.

  2. For *customer*, *product* and *state*, show the average sales before and after each quarter (e.g., for Q2, show average sales of Q1 and Q3. For "before" Q1 and "after" Q4, display <NULL>. The "YEAR" attribute is not considered for this query – for example, both Q1 of 2007 and Q1 of 2008 are considered Q1 regardless of the year.

  3. For each product, find the median sales quantity (assume an odd number of sales for simplicity of presentation). (NOTE – "***median***" is defined as "*denoting or relating to a value or quantity lying at the midpoint of a frequency distribution of observed values or quantities, such that there is an equal probability of falling above or below it.*" E.g., Median value of the list {13, 23, 12, 16, 15, 9, 29} is 15.

     For example, given the following sales transactions for Bread, the median quant for Bread is 3.

     ```
     PRODUCT   QUANT

     =======  =====

     Bread       1
     Bread       1
     Bread       1
     Bread       2
     Bread       2
     Bread       3
     Bread       4
     Bread       5
     Bread       6
     Bread       7
     Bread       7
     ```

  4. For *customer* and *product*, find the month by which time, 75% of the sales quantities have been purchased. Again, for this query, the "YEAR" attribute is not considered. Another way to view this query is to pretend all 500 rows of sales data are from the same year.

  The following are sample report output (NOTE: the numbers shown below are not the actual aggregate values. You can write simple SQL queries to verify the actual aggregate values).

Report #1:

```
CUSTOMER PRODUCT MONTH STATE CUST_AVG OTHER_PROD_AVG OTHER_MONTH_AVG OTHER_STATE_AVG
======== ======= ===== ===== ======== ============== =============== ===============
Helen    Bread       1 NY         243           1493             199             268
Emily    Milk        3 NJ        1426            926             482             478

. . . .
```

Report #2:

```
CUSTOMER PRODUCT STATE Q1  BEFORE_AVG  AFTER_AVG
======== ======= ===== == ========== =========
Bloom    Bread      NJ  1     <NULL>       2434
Sam      Milk       CT  3        254        325

. . . .
```

Report #3:

```
PRODUCT  MEDIAN QUANT
=======  ============
Bread             422
Milk             1976

. . . .
```

Report #4:

```
CUSTOMER PRODUCT  75% PURCHASED BY MONTH
======== =======  ======================
Emily    Bread                         2
Bloom    Milk                          3

. . . .
```

Make sure that:

1. Character string data (e.g., customer name and product name) are left justified.

2. Numeric data (e.g., Maximum/minimum Sales Quantities) are right justified.

3. **Only standard SQL statements and aggregate function syntaxes** are used – if you're unsure, please ask the Tas.

Grading: **NOTE: A query with syntax errors will lose 50% of the points for the query**.

Submission: Submit **one file** on Canvas containing all of the 4 queries with your name and CWID on it. The file type must be "TXT".

Please include a "README" section in the same file if any special instructions are required.

I encourage you to discuss the "ideas" with your TAs as soon as possible (rather than your classmates, esp, if you have any specific questions), but the final queries must be your own work. If I determine that your queries are copies of someone else's, both you and that someone else will be disciplined (you will receive 0 for the entire assignment) and possibly receive additional penalties for the course.