CS442 – SQL Programming Assignment 1

Due Dates: 11/1/2022 (Tue)

Objectives

In this assignment, you will learn to _express_ "complex" OLAP queries in SQL. The key point of the exercise is to observe the complexity of expressing the type of such queries despite relatively simple ideas of the queries themselves. Your mission (in addition to writing the SQL queries) is to consider the reasons for the complexity of the _expression_ of these queries.

Description

Generate separate reports/output based on the following queries (one report for each of the queries):

1. For each _product_ compute the _maximum_ and _minimum_ sales quantities along with the _corresponding customer_ (who purchased the product)_, dates_ (i.e., dates of those maximum and minimum sales quantities) and the _state_ in which the sale transaction took place.

   For the same _product_, compute the _average_ sales quantity.

2. For each combination of _customer_ and _product_, output the _maximum sales quantities for NY_ and _minimum sales quantities for NJ and CT in 3 separate columns_. Like the first report, display the _corresponding dates_ (i.e., dates of those maximum and minimum sales quantities). Furthermore, for CT and NJ, include only the sales that occurred after 2000; for NY, include all sales.

3. For _each of the 12 months_ (regardless of the year), find the _most "popular"_ and _least "popular" states_ (those states with most and least total sales quantities) and the corresponding _total sales quantities (i.e., SUMs)_.

4. For each _customer_, find the "_most favorite_" _product_ (which the customer purchased the most), the "_least favorite_" _product_ (which the customer purchased the least), the "_most favorable state_" (where the most amounts of purchases were made) and the "_least favorable state_" (where the least amounts of purchases were made).

5. Show for each _customer_ and _product_ combination, the _total sales quantities for 4 quarters_, Q1, Q2, Q3 and Q4 (in four separate columns) – Q1 being the first 3 months of the year (Jan, Feb & Mar), Q2 the next 3 months (Apr, May & Jun), and so on – ignore the YEAR component of the dates (i.e., 3/11/2001 is considered the same date as 3/11/2002, etc.). Also compute the _average for the "whole" year_ (again ignoring the YEAR component, meaning simply compute AVG) along with the _total quantities_ (SUM) and the _counts_ (COUNT).

The following are sample output reports – quantities displayed are for illustration only (not the actual values.

Report #1:

```
PRODUCT   MAX_Q  MAX_CUST  MAX_DATE    ST  MIN_Q  MIN_CUST  MIN_DATE    ST  AVG_Q
========  =====  ========  ==========  ==  =====  ========  ==========  ==  =====
Pepsi      2893  Bloom     01/01/2006  NJ     12  Emily     09/25/2001  NY   1435
Banana      159  Dan       02/15/2002  NJ      1  Sam       03/23/2004  CT     56
Apple      3087  Helen     07/01/2005  NY      2  Sam       02/02/2001  NJ   1512
 . . . .
```

Report #2:

```
CUSTOMER   PRODUCT   NY_MAX   DATE         NJ_MIN   DATE         CT_MIN   DATE
========   =======   ======   ==========   ======   ==========   ======   ==========
Sam        Egg         1908   01/11/2000        2   07/24/2005        2   11/03/2008
Helen      Cookies      392   03/31/2002       42   09/14/2001       11   07/23/2002
Bloom      Butter      7045   09/22/2003       23   03/10/2004        8   09/11/2006
. . . .
```

Report #3:

```
MONTH   MOST_POPULAR_ST   MOST_POP_TOTAL_Q   LEAST_POPULAR_ST   LEAST_POP_TOTAL_Q
=====   ===============   ================   ================   =================
    1   NJ                          497214   CT                             55526
    2   NJ                         1874794   NY                             23126
    3   CT                          974531   NJ                             19958

. . . .
```

Report #4:

```
CUSTOMER   MOST_FAV_PROD   LEAST_FAV_PROD   MOST_FAV_ST   LEAST_FAV_ST
========   =============   ==============   ===========   ============
Claire             Apple          Jellies            NJ             NY
Emily               Eggs           Butter            NJ             CT
Chae                Fish              Ham            PA             NJ
. . . .
```

Report #5:

```
CUSTOMER   PRODUCT   Q1_TOT   Q2_TOT   Q3_TOT   Q4_TOT   AVERAGE   TOTAL   COUNT
========   =======   ======   ======   ======   ======   =======   =====   =====
Sam        Pepsi       1923     4241     2383     1325      2988   38848      13
Emily      Milk         239     9872      142     2435      2663   21307       8
Helen      Bread       2534      981     4239     1987      2781   25032       9

. . . .
```

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

Submission   Submit **one file** containing all of the 5 queries with your name and CWID on it on Canvas. The file type must be "TXT".

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

You can discuss the "ideas" with your classmates or your friends, 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.