# CS442 – Programming Assignment 1

## Due Date: 4/6/2023 (Thu.)

***Objectives***

In this assignment, you will <u>express</u> "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 *customer*, compute the <u>minimum</u> and <u>maximum</u> sales quantities along with the <u>corresponding products</u> (purchased), <u>dates</u> (i.e., dates of those minimum and maximum sales quantities) and the <u>states</u> in which the sale transactions took place. If there are >1 occurrences of the min or max, display all.

   For the same *customer*, compute the <u>average</u> sales quantity.

2. For *each of the 12 months* (regardless of the year), find the <u>most "productive"</u> and <u>least "productive" days</u> (those days with most and least total sales quantities) and the corresponding <u>total sales quantities (i.e., SUMs)</u>.

3. For each *product*, find the "<u>most favorable</u>" <u>month</u> (when most amount of the product was sold) and the "<u>least favorable</u>" <u>month</u> (when the least amount of the product was sold).

4. Show for each *customer* and *product* combination, the <u>average sales quantities for 4 quarters</u>, 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., 3/11/2016 is considered the same date as 3/11/2020, etc.). Also compute the <u>average for the "whole" year</u> (again ignoring the YEAR component, meaning simply compute AVG) along with the <u>total quantities</u> (SUM) and the <u>counts</u> (COUNT).

5. For each combination of *customer* and *product*, output the <u>maximum sales quantities for NJ, NY and CT in 3 separate columns</u>. Like the first report, display the <u>corresponding dates</u> (i.e., dates of those corresponding maximum sales quantities). Furthermore, show the output <u>only if maximum for NY is greater than NJ or CT</u>.

The following is a sample output – quantities displayed are for illustration only (not the actual values). **For dates (e.g., MAX_DATE, MIN_DATE), you can display 'month', 'day' and 'year' as 3 separate columns – i.e., you don't need to concatenate them into MM/DD/YYYY format.**

<u>Report #1:</u>

```
CUSTOMER  MIN_Q  MIN_PROD  MIN_DATE    ST  MAX_Q  MAX_PROD  MAX_DATE    ST  AVG_Q
========  =====  ========  ==========  ==  =====  ========  ==========  ==  =====
Bloom        12  Pepsi     01/01/2016  NJ   2893  Apple     09/25/2017  NY   1435
Sam           1  Milk      02/15/2018  NJ    259  Banana    03/23/2020  CT     56
Emily         2  Bread     07/01/2017  NY   3087  Milk      02/02/2029  NJ   1512

. . . .
```

Report #2:

```
MONTH  MOST_PROFIT_DAY  MOST_PROFIT_TOTAL_Q  LEAST_PROFIT_DAY  LEAST_PROFIT_TOTAL_Q
=====  ===============  ===================  ================  ====================
    1               12               497214                31                 55526
    2               23              1874794                15                 23126
    3                4               974531                 2                 19958

. . . .
```

Report #3:

```
PRODUCT  MOST_FAV_MO  LEAST_FAV_MO
=======  ===========  ============
Egg                4            12
Apple              1            11
Banana             3             2
. . . .
```

Report #4:

```
CUSTOMER  PRODUCT  Q1_AVG  Q2_AVG  Q3_AVG  Q4_AVG  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

. . . .
```

Report #5:

```
CUSTOMER  PRODUCT  NJ_MAX  DATE        NY_MAX  DATE        CT_MAX  DATE
========  =======  ======  ==========  ======  ==========  ======  ==========
Sam       Egg        7908  01/11/2016    2405  07/24/2017    1932  11/03/2020
Helen     Cookies     392  03/31/2018    1042  09/14/2020     811  07/23/2019
Bloom     Butter     1045  09/22/2020    2023  03/10/2016    2988  09/11/2018
```

**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.