

# CS561 – Programming Assignment 1

Due Date: 7/22/2022 (Fri.)

## Objectives

- In this assignment, you will express “complex” OLAP queries in SQL. The key point of the exercise is to observe a large gap between the complexity of expressing the type of such queries and that of evaluating them (such as writing Java programs to produce the same results). Your mission (in addition to writing the SQL queries) is to consider the reasons for the gap (between the *expression* and *evaluation* of such queries) and how to narrow it.

## Description

- Generate 5 separate reports based on the following queries (one report for query #1, one for query #2, one for query #3, one for query #4, and another for query #5):
  - For each *customer*, compute the minimum and maximum sales quantities along with the corresponding products (purchased), dates (i.e., dates of those minimum and maximum sales quantities) and the states 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 average sales quantity.
  - For each of the 12 months (regardless of the year), find the most “popular” and least “popular” products (those products with most and least total sales quantities) and the corresponding total sales quantities (i.e., SUMs).
  - For each *product*, find the “most favorable” month (when most amount of the product was sold) and the “least favorable” month (when the least amount of the product was sold).
  - Show for each *customer* and *product* combination, the average 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).
  - For each combination of *customer* and *product*, output the maximum sales quantities for NJ, NY and CT in 3 separate columns. Like the first report, display the corresponding dates (i.e., dates of those corresponding maximum sales quantities). Furthermore, show the output only if maximum for NY is greater than NJ or CT.

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.**

### Report #1:

CUSTOMER	MIN_Q	MIN_PROD	MIN_DATE	ST	MAX_Q	MAX_PROD	MAX_DATE	ST	AVG_Q
=====	=====	=====	=====	==	=====	=====	=====	==	=====
Bloom	12	Pepsi	01/01/2006	NJ	2893	Apple	09/25/2001	NY	1435
Sam	1	Milk	02/15/2002	NJ	259	Banana	03/23/2004	CT	56
Emily	2	Bread	07/01/2005	NY	3087	Milk	02/02/2001	NJ	1512
. . . .									

**Report #2:**

MONTH	MOST_POPULAR_PROD	MOST_POP_TOTAL_Q	LEAST_POPULAR_PROD	LEAST_POP_TOTAL_Q
=====	=====	=====	=====	=====
1	Eggs	497214	Pepsi	55526
2	Milk	1874794	Banana	23126
3	Pepsi	974531	Milk	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	01/11/2001	2405	07/24/2005	1932	11/03/2008
Helen	Cookies	392	03/31/2002	1042	09/14/2000	811	07/23/2002
Bloom	Butter	1045	09/22/2003	2023	03/10/2004	2988	09/11/2006

. . . . .

**Grading**

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

**Submission**

Submit **one file** containing all the 5 queries on Canvas – Please include your name and CWID in the file. The file type must be “TXT”.

Please include a “README” file if any special instructions are required.

I encourage you to discuss the “ideas” with your TAs (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.