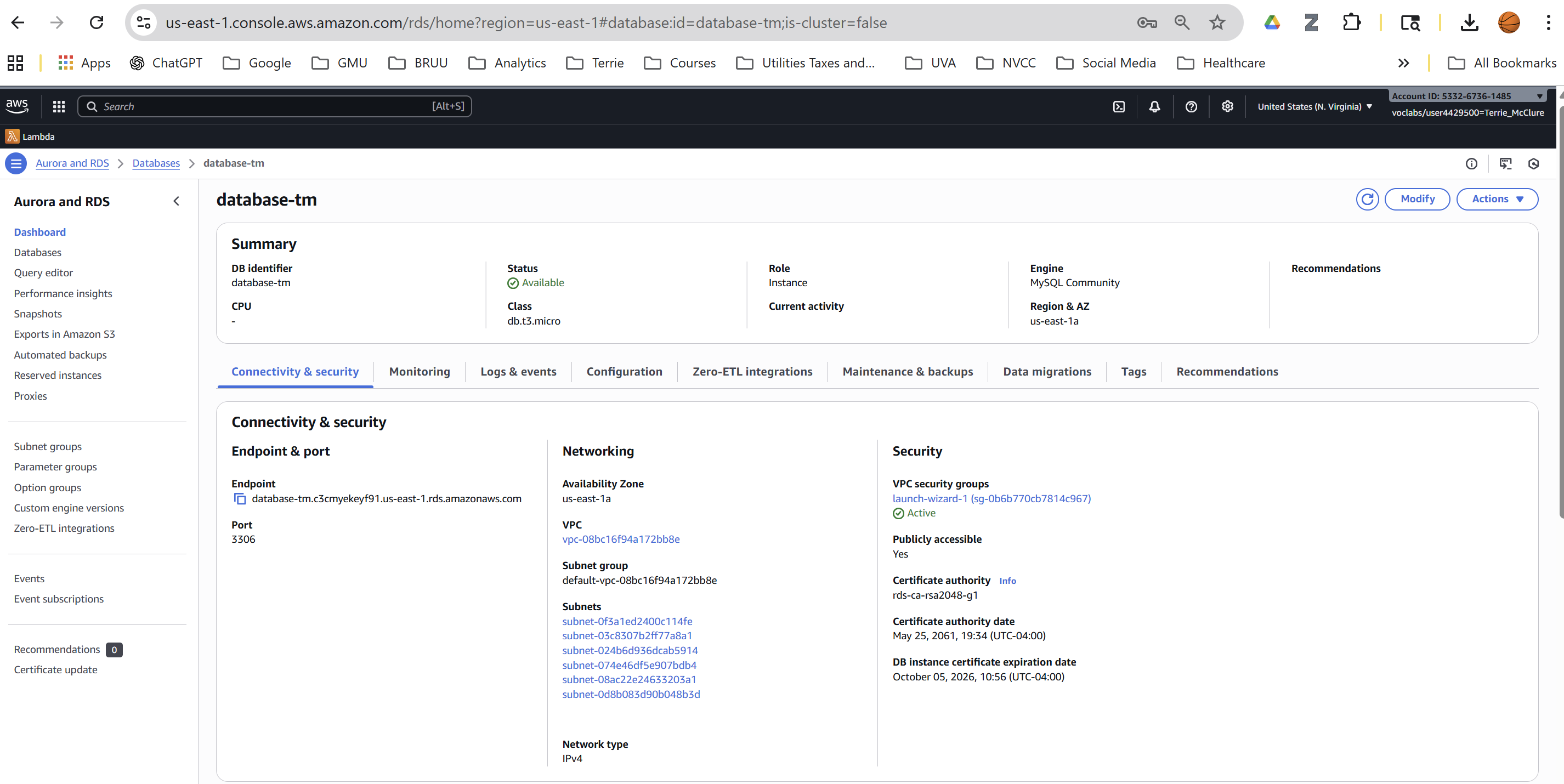
M8: Final project – SQL queries

Name: Terrie McClure

Email: tmcclur@gmu.edu

GNumber: G00239649

Course section: Analytics Big Data to Info, AIT 580, Section 011



mysql Ver 9.4.0 for Win64 on x86\_64 (MySQL Community Server - GPL)

C:\cygwin64\home\Terri\M6>mysql -h database-tm.c3cmyekeyf91.us-east-1.rds.amazonaws.com -P 3306 -u admin -p

Enter password: \*\*\*\*\*\*\*\*\*\*\*\*\*\*

Welcome to the MySQL monitor. Commands end with ; or \g.

Your MySQL connection id is 97

Server version: 8.4.6 Source distribution

Copyright (c) 2000, 2025, Oracle and/or its affiliates.

Oracle is a registered trademark of Oracle Corporation and/or its

affiliates. Other names may be trademarks of their respective

owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

* Create SQL database schema and table

mysql> create database sol\_cs\_db;

Query OK, 1 row affected (1.074 sec)

mysql> use sol\_cs\_db

Database changed

mysql> create table sol\_cs (

-> school\_year text,

-> division\_number integer,

-> division\_name text,

-> school\_number integer,

-> school\_name text,

-> subject text,

-> race text,

-> gender text,

-> test\_level text,

-> test\_source text,

-> pass\_advanced\_count integer,

-> pass\_proficient\_count integer,

-> fail\_count integer,

-> pass\_count integer,

-> total\_count integer,

-> pass\_advanced\_rate decimal(10,3),

-> pass\_proficient\_rate decimal(10,3),

-> fail\_rate decimal(10,3),

-> pass\_rate decimal(10,3),

-> year integer,

-> stem\_flag bool,

-> cs\_enrollment\_count integer,

-> school\_level text

-> );

Query OK, 0 rows affected (1.433 sec)

mysql> show tables

-> ;

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

| Tables\_in\_sol\_cs\_db |

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

| sol\_cs |

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

1 row in set (0.931 sec)

mysql> describe sol\_cs;

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

| Field | Type | Null | Key | Default | Extra |

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

| school\_year | text | YES | | NULL | |

| division\_number | int | YES | | NULL | |

| division\_name | text | YES | | NULL | |

| school\_number | int | YES | | NULL | |

| school\_name | text | YES | | NULL | |

| subject | text | YES | | NULL | |

| race | text | YES | | NULL | |

| gender | text | YES | | NULL | |

| test\_level | text | YES | | NULL | |

| test\_source | text | YES | | NULL | |

| pass\_advanced\_count | int | YES | | NULL | |

| pass\_proficient\_count | int | YES | | NULL | |

| fail\_count | int | YES | | NULL | |

| pass\_count | int | YES | | NULL | |

| total\_count | int | YES | | NULL | |

| pass\_advanced\_rate | decimal(10,3) | YES | | NULL | |

| pass\_proficient\_rate | decimal(10,3) | YES | | NULL | |

| fail\_rate | decimal(10,3) | YES | | NULL | |

| pass\_rate | decimal(10,3) | YES | | NULL | |

| year | int | YES | | NULL | |

| stem\_flag | tinyint(1) | YES | | NULL | |

| cs\_enrollment\_count | int | YES | | NULL | |

| school\_level | text | YES | | NULL | |

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

23 rows in set (0.926 sec)

* Load dataset

mysql> LOAD DATA LOCAL INFILE 'SOL\_CS.csv' INTO TABLE sol\_cs FIELDS TERMINATED BY ',' ENCLOSED BY '"' LINES TERMINATED BY '\r\n' IGNORE 1 LINES;

Query OK, 18493 rows affected (39.897 sec)

Records: 18493 Deleted: 0 Skipped: 0 Warnings: 0

mysql> select \* from sol\_cs limit 1

-> ;

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

| school\_year | division\_number | division\_name | school\_number | school\_name | subject | race | gender | test\_level | test\_source | pass\_advanced\_count | pass\_proficient\_count | fail\_count | pass\_count | total\_count | pass\_advanced\_rate | pass\_proficient\_rate | fail\_rate | pass\_rate | year | stem\_flag | cs\_enrollment\_count | school\_level |

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

| 2020-2021 | 1 | Accomack County | 540 | Arcadia High | English:Reading | Black, not of Hispanic origin | F | EOC | SOL | 0 | 15 | 10 | 15 | 25 | 0.000 | 60.000 | 40.000 | 60.000 | 2021 | 0 | 13 | High School |

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

1 row in set (1.974 sec)

* + Identify and select a column variable for each NOIR data type

|  |  |  |
| --- | --- | --- |
| Column Name | Computational Data Type | NOIR Data Type |
| pass\_rate  pass\_advanced\_rate | Float | **Ratio** |
| year | Integer | **Interval** |
| school\_level | String | **Ordinal** |
| stem\_flag | Boolean | **Nominal** |

## Summary statistics using Interval column Year

*group by statistics (mean pass rates by year)*

mysql> select year, round(avg(pass\_rate),1) as pass\_rate from sol\_cs group by year;

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

| year | pass\_rate |

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

| 2021 | 70.2 |

| 2022 | 77.5 |

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

2 rows in set (2.340 sec)

The above summary shows that school pass rates for End of Course SOL tests increased from 2021 to 2022. In 2021 the average (mean) pass rate was 70.2%. In 2022 the average pass rate was 77.5%.

## Summary statistics using Ordinal column School Level.

*group by statistics (mean pass rates by school level):*

mysql> SELECT

-> school\_level,

-> round(AVG(pass\_rate),1) AS mean\_pass\_rate,

-> round(AVG(pass\_advanced\_rate),1) AS mean\_pass\_advanced\_rate -- ,

-> FROM sol\_cs

-> GROUP BY school\_level

-> ORDER BY FIELD(school\_level, 'Elementary School','Middle School','High School');

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

| school\_level | mean\_pass\_rate | mean\_pass\_advanced\_rate |

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

| Elementary School | 96.8 | 14.8 |

| Middle School | 86.0 | 15.3 |

| High School | 71.8 | 12.4 |

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

3 rows in set (2.026 sec)

mysql>

The above summary shows that elementary schools have the highest pass rates, followed by middle schools, then high schools – with a mean pass rate of 96.8% - for elementary students taking End of Course SOL tests. End of Course SOL tests are usually taken in high school. An elementary age student would typically take an End of Course Algebra test only if they are already an advanced student in Algebra.

## Summary statistics using Nominal column Stem Flag

*number of schools in table sol\_cs*

mysql> select count(distinct division\_name, school\_name) as schools from sol\_cs;

+---------+

| schools |

+---------+

| 658 |

+---------+

1 row in set (2.799 sec)

*number of tests given*

mysql> select count(distinct division\_name, school\_name, subject) as tests from sol\_cs;

+-------+

| tests |

+-------+

| 1990 |

+-------+

1 row in set (2.253 sec)

*number of EOC SOL tests given by subject and STEM/Not STEM*

mysql> select subject, stem\_flag, count(distinct division\_name, school\_name, subject) as tests from sol\_cs group by subject, stem\_flag;

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

| subject | stem\_flag | tests |

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

| English:Reading | 0 | 324 |

| English:Writing | 0 | 254 |

| History and Social Science | 0 | 353 |

| Mathematics | 1 | 655 |

| Science | 1 | 404 |

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

5 rows in set (1.349 sec)

mysql>

The above summary shows that the number of End of Course SOL tests are close to evenly split between STEM and Non-STEM subjects.

## Summary statistics using Ratio columns Pass Rate and Advanced Pass Rate.

mysql> select min(pass\_rate) as 'min pass rate', max(pass\_rate) as 'max pass rate', round(avg(pass\_rate),1) as 'mean pass rate' from sol\_cs;

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

| min pass rate | max pass rate | mean pass rate |

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

| 0.000 | 100.000 | 74.2 |

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

1 row in set (0.990 sec)

mysql> select min(pass\_advanced\_rate) as 'min pass advanced rate', max(pass\_advanced\_rate) as 'max pass advanced rate', round(avg(pass\_advanced\_rate),1) as 'mean pass advanced rate' from sol\_cs;

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

| min pass advanced rate | max pass advanced rate | mean pass advanced rate |

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

| 0.000 | 100.000 | 12.9 |

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

1 row in set (6.913 sec)

The above summaries show that both pass\_rate and pass\_advanced\_rate range from 0% to 100%. The average pass\_rate is 74.2%, while the average pass\_advanced\_rate is 12.9%.