MSiA-413 Introduction to Databases and Information Retrieval

Homework 3: Basic SQL Queries

Name 1: Ayush Agarwal

NetID 1: scg1143

Name 2: Lowan(Sydney) Li

NetID 2: chr0390

Instructions

You should submit this homework assignment via Canvas. Acceptable formats are word files, text files, and pdf files. Paper submissions are not allowed and they will receive an automatic zero.

As explained during lecture and in the syllabus, assignments are done in groups. The groups have been created and assigned. Each group needs to submit only one assignment (i.e., there is no need for both partners to submit individually the same homework assignment).

Each group can submit solutions multiple times (for example, you may discover an error in your earlier submission and choose to submit a new solution set). We will grade only the last submission and ignore earlier ones.

Make sure you submit your solutions before the deadline. The policies governing academic integrity, tardiness and penalties are detailed in the syllabus.

Homework Instructions

For this assignment, you will use the program "DB Browser for SQLite" (available at http://sqlitebrowser.org/). This is the same software we have worked with in class. I posted several sample database files on Canvas in the Lecture Slides page. These database files can be opened with the DB Browser for SQLite. The database files we will use for this homework are:

- SalesOrders.sqlite
- SchoolScheduling.sqlite

Using these datasets, please answer each of the questions that follow with one query only. The query can have subqueries, if needed. For every question, we expect to see both your SQL code and the resulting data. Copy and paste both the SQL code and the results into a document and submit it following the submission instructions.

You may find it helpful to use the "Basic SQL Cheat Sheet" posted on Canvas.

Each one of the questions below is worth **10 points**.

EXAMPLE:

Question: Which bikes cost more than $1000?

Answer:

```
SELECT ProductName, RetailPrice
FROM Products
WHERE CategoryID = (SELECT CategoryID
                    FROM Categories
                    WHERE CategoryDescription = "Bikes")
 AND RetailPrice > 1000;
```

Output:

```
"Trek 9000 Mountain Bike"       "1200"
"Eagle FS-3 Mountain Bike"      "1800"
"GT RTS-2 Mountain Bike" "1650"
```

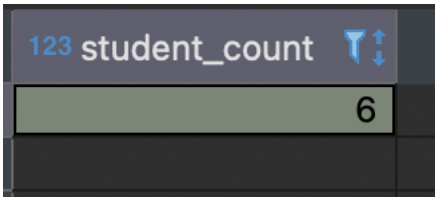
SchoolScheduling.sqlite

1) How many students are majoring in English or Mathematics?

```
SELECT COUNT(DISTINCT(s.StudentID)) as student_count
FROM
Students s
inner join
Majors m
on
s.StudMajor = m.MajorID
where
 m.Major = 'English' or m.Major = 'Mathematics';
```

Response:

There are 6 students who are majoring in English or Mathematics

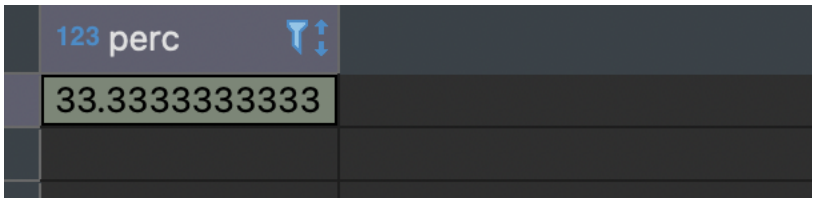


2) What is the percentage of students with majors in English or Mathematics?

```
SELECT 1.0 * COUNT(*) / (SELECT COUNT(*) FROM Students s2) * 100 as perc
FROM
Students s
where
s.StudMajor in
(
    SELECT m.MajorID
    FROM Majors m
    where m.Major = 'English' or m.Major = 'Mathematics'
)
```

Response:

% of students majoring in English or Mathematics is 33.33 %

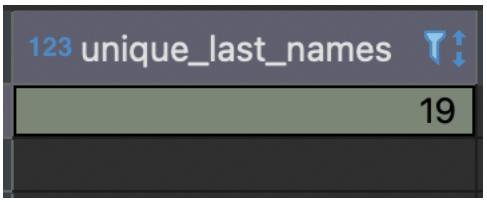


3) How many unique last names does the staff have?

SELECT COUNT(DISTINCT(s.StfLastname)) as unique_last_names
FROM Staff s

Response

There are 19 unique last names

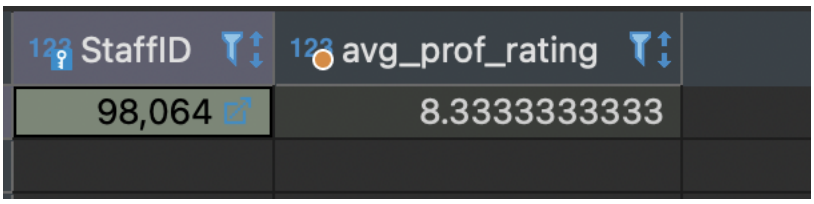


4) Each staff member has a proficiency rating for a number of subjects. For each staff member we can calculate its average proficiency rating (average across all subjects). What is the minimum value of the average proficiency rating of the staff?

SELECT fs.StaffID, AVG(fs.ProficiencyRating) as avg_prof_rating
FROM
Faculty_Subjects fs
group by
fs.StaffID
order by
avg_prof_rating
ASC
limit 1

Response

Minimum Average Proficiency is 8.333 for staff ID 98064

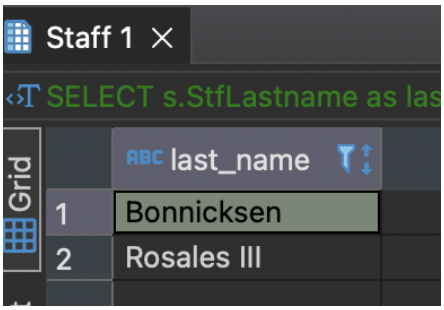


5) In the Staff table, which last names have a length longer than 9 characters?

```
SELECT s.StfLastname
FROM
Staff s
where
length(s.StfLastname) > 9;
```

Response:

There are 2 last names which have length larger than 9 characters



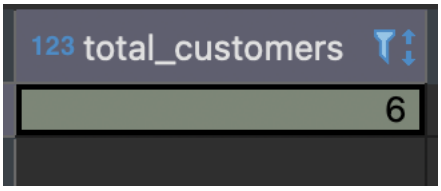
SalesOrders.sqlite

6) How many customers live in TX?

```
SELECT COUNT(*) as total_customers
FROM Customers c
where
c.CustState = 'TX';
```

Response:

Total Customers living in State TX is 6

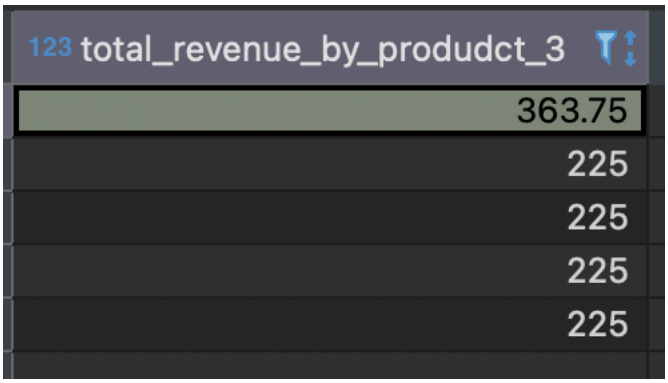


7) What are the top 5 highest revenue amounts that product number 3 has ever generated in a sale?

```
SELECT od.QuotedPrice * od.QuantityOrdered as total_revenue_by_produdct_3
FROM Order_Details od
where
od.ProductNumber = 3
order by
total_revenue_by_produdct_3
DESC
limit 5
```

Response:

Top 5 Revenues generated by Product are 363.75, 225, 225, 225, 225

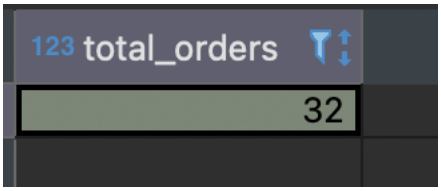


8) How many orders has a customer named Angel Kennedy placed so far?

```
SELECT COUNT(DISTINCT(o.OrderNumber)) as total_orders
FROM Orders o
where
o.CustomerID in
(
        SELECT c.CustomerID
        FROM
        Customers c
        where
        c.CustFirstName = 'Angel' and c.CustLastName = 'Kennedy'
)
```

Response:

Angel Kennedy has placed 32 orders so far



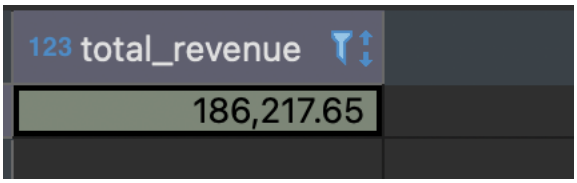
9) What is the total revenue that a customer named Angel Kennedy has brought through product sales?

```
SELECT SUM(o.OrderTotal) as total_revenue
FROM Orders o
```

```
where
o.CustomerID in
(
	SELECT c.CustomerID
	FROM
	Customers c
	where
	c.CustFirstName = 'Angel' and c.CustLastName = 'Kennedy'
)
```

Response:

Angel Kennedy has brought total revenue of worth 186,217.65 USD so far



10) In which state do most customers live? Report both the state name and the number of customers living in that state.

```
SELECT c.CustState , COUNT(*) as total_customers
FROM
Customers c
group by
c.CustState
order by
count(*) DESC
limit 1
```

Response

Most Customers live in WA and the count is 11