Mentor training day

MINI HACKATHON - SQL

LEARNING OBJECTIVES

- Setting up a local MySQL server and MySQL workbench (feel free to use other SQL tools)
- Familiarising with a relevant database that learners will use.
- [Beginners] Getting started on simple SQL queries by following a list of well-crafted questions.
- [Intermediate Users] Using advanced SQL querying to identify a business problem for Adventureworks (arbitrary bike store).

Overall, this session aims to upskill mentors in SQL as a Data Manipulation Language (DML). This session will not be a lecture-style session. Rather, it will resemble a guided hackathon with some demonstrations at the start to kick things off.
If you have trouble getting SQL into your local machine, please refer to the pre-session set up below.

TASKS

Below is a list of questions designed to help you practice SQL querying. The questions become progressively more challenging as you go down. The first few questions are tailored for beginners. For intermediate users, feel free to skip these questions, and go straight to the challenging ones. Finally, there is an open-ended question at the end for advanced users.

1. **EASY** Select all of the data from the "department" table. How do you select just the "Name" from this table?

2. **EASY** What does the below query do? How would you rewrite this query so that it includes the number of customers that actually use those titles?

```
SELECT DISTINCT Title
FROM contact
```

3. **EASY-MEDIUM** Using the Entity Relationship Diagram (ERD) for inspiration, join the "customer" table to the "salesterritory" table via the appropriate columns ('keys'). (see below for instructions on getting the ERD)

4. **EASY-MEDIUM** Find some aggregate statistics (COUNT(), SUM(), AVG()...) from the "purchaseorderheader" table for each employeeId. Rename the columns appropriately.

5. **MEDIUM** Find some aggregate statistics (COUNT(), SUM(), AVG()...) from the "purchaseorderheader" table for each employeeID, but also include other employee information (employee information exists in the "employee" table). Rename the columns appropriately.

6. **MEDIUM** How would you rewrite the following query without the HAVING clause?

```
SELECT TerritoryID, COUNT(*)
FROM Customer
WHERE CustomerType <> 'S'
GROUP BY TerritoryID
HAVING COUNT(*) >= 1800
ORDER BY COUNT(*) DESC
```

7. **MEDIUM-HARD** By joining the contact table and the salesorderheader table, sum the SubTotal field in salesorderheader to make a Total field for all customers that have a first name that contains the letters 'll' and have no registered middle name, or, customers whose last name contains the letter 'v'.

   Then: filter the data so that the new Tot al field exceeds 1000, and round the Total field to the nearest 100. Ensure that the data is ordered.

   Finally, present the data with a dollar ($) sign (use the CAST()) and CONCAT() functions). Name this field "Total Spent".

8. **MEDIUM-HARD** Unfortunately, MySQL doesn't like the following query due to the LIMIT statement being embedded into a sub-query, albeit the fact that the SQL statement is quite intuitive. How could you rewrite the following query to obtain the desired result?

```
SELECT AddressLine1, City
FROM address a
JOIN customeraddress ca on ca.AddressID = a.AddressID
WHERE ca.customerID IN (
      SELECT customerid
      FROM salesorderheader
      GROUP BY customerid
      ORDER BY COUNT(*) DESC
      LIMIT 1
)
```

9. **HARD** Using the SalesOrderHeader table, show how much revenue (TotalDue field) was made and the number of orders made for each range of sales:

   0-99 dollars
   100-999 dollars
   1000-9999 dollars
   10000 dollars

10. As a data scientist/analyst, you have been tasked to identify **ONE** business problem for Adventureworks. You have access to the company's entire database which contains data on sales, employee, client, purchase, etc.

PRE-SESSION SET-UP

SETTING UP LOCAL MYSQL SERVER AND MYSQL WORKBENCH

For Mac Users, please update your MacOS to at least version 10.14 for Workbench to work. (This takes an hour, so please do it beforehand).

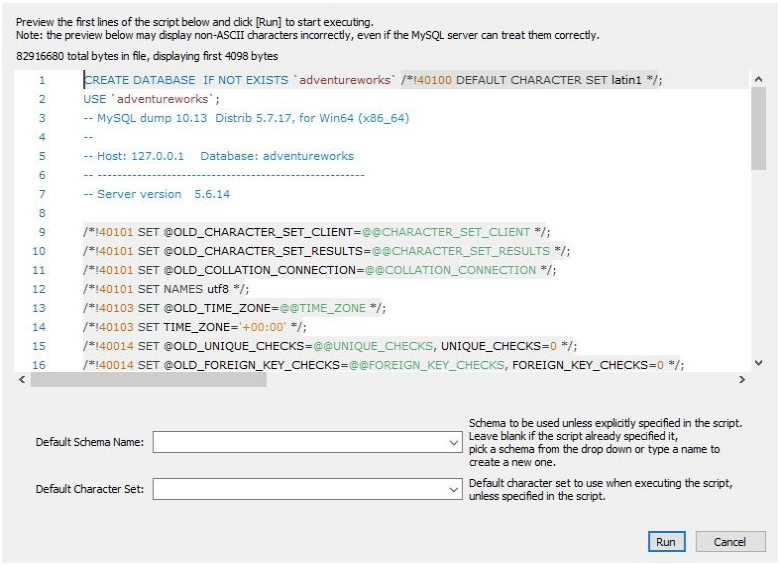
For SQL to work in your local machine, you will need to install two components:

1. MySQL Community Server - a popular Open Source RDBMS. This is the local MySQL server where your databases will sit.
2. MySQL Workbench - a graphical user interface to interact with MySQL servers.

SETTING UP ADVENTUREWORKS DATABASE

The database we will be using for this hackathon is a sample database from Microsoft SQL. This data provided is from a bike store known as Adventureworks. It contains data on sales, HR, client, etc. In order to import the database into your local workbench.

1. Download the MySQL version of the database from https://github.com/tapsey/AdventureWorksMYSQL/blob/master/AdventureWorks-MySQL-with-FKs.zip
2. Extract the zip file which will create `AdventureWorks-MySQL-with-FKs.sql`
3. Open MySQL Workbench and connect to your Local instance
4. Select `File -> Run SQL Script…` and select `AdventureWorks-MySQL-with-FKs.sql`



5. Select Run

If you are running on different SQL servers, here are other guides for installing the same database.

1. The Official Microsoft SQL AdventureWorks 2014 database
2. A PostgreSQL unofficial port

[OPTIONAL] SETTING UP ENTITY RELATIONSHIP DIAGRAM

If you want to create entity relationship diagram, the following steps will help you to quickly create it for an existing database.

1. Open your MySQL workbench and click on Database.
2. Click on Reverse Engineer.
3. Select your database connection and click on Next.
4. Click on Next.
5. Select the database schema for which you want to create the ER diagram and then click on Next.
6. I wanted to make ER diagram for the adventureworks schema, so I selected adventureworks.
7. After this it will retrieve objects from the selected database and check results, after the blue ticks in similar to the image below appear, click on Next.
8. Now there will be an option to select objects to reverse engineer i.e. the objects for which you want to create a diagram, by default all options are selected. After selecting objects click on Execute.
9. Click on Next.
10. Now click on Finish.
11. Once we click on the Finish, a new tab will open in the workbench which will contain the ER diagrams of the selected objects.