**Assignment 01 - DQL, DDL, DML and Data Modeling**

**Due: by 11:59 PM on Sunday, March 10, 2024.**

**Overall Grade %**: 12

**Total Points**: 120

Part I is worth 60 points.

Part I of this assignment will help you become familiar with basic SQL syntax to query, create, and manipulate data in an SQL database. If you find yourself stuck after trying to resolve the issue on your own, start at [http://mysqltutorial.org](http://mysqltutorial.org/), a comprehensive resource that clearly explains SQL related concepts and functions.

Part II of the assignment will be for data modeling and will consist of creating the tables for a relational database. The required files, their formats as well as the actual tasks to be completed will be included in the file that contains the details of the Part II of the assignment (distributed separately).

**Submission Files**

To submit your answers to the tasks related to SQL, create a .SQL file in a text editor of your choice (Atom/Sublime/Notepad++/MySQL Workbench etc). The .SQL file will contain your SQL statements. Name the .SQL file in this format:

*assignment\_01\_part\_I\_<your\_last\_name>.SQL*

Example:

assignment\_01\_part\_I\_seal.SQL

Upload the .SQL file and your Excel files for the custom data requests to the Assignment\_01\_Part\_I link on Brightspace.

**Formatting**

At the top of the .SQL file, add ***your LMU.build MySQL database*** connection details using the syntax for multi-line comments (if you are using that host).

Example:

/\*

Host:

Username:

Password:

\*/

/\* You can create the Specialty Food database by executing the commands in the file SpecialtyFoodSchemaData.SQL (available from Brightspace). You have to create the database first though in your lmu build account and then you can either use phpMyAdmin to upload the SQL file, or can run it from your client.

/\*

SELECT …

FROM …

WHERE …;

Be sure the SQL conforms to the style guide and best practices as discussed in the class (<http://www.sqlstyle.guide>) **(3 points) \*/**

/\* The purpose of this assignment is to

* Make you practice your SQL skills and also
* To understand how sql can be used for answering business questions.

That is why you may not find a direct data request from many of the questions asked below. They are formed as business questions and you have to figure out what sort of information need to be collected to answer those business questions. \*/

/\* Set I. Specialty Food database. You can use the connection below or your own connections.

host: kcs.lmu.build

user: kcslmubu\_guest

pw: db4BSANcourses

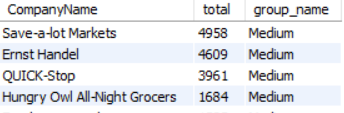
Schema: kcslmubu\_specialty\_food

The Specialty Food database tracks the entities and transactions for a global import export company.

**Create an ERD for the database. You just have to create the diagram. However, the diagram must clearly indicate all the of the PK, FK and the connections. (3 points)**

Write SQL statements to complete the following tasks. Please note that you can use your own lmu build server if you want to create VIEW (you cannot create CTE in lmu build but VIEW can replace CTE)to handle complex queries. \*/

1. The marketing department wants to run a direct mail marketing campaign to its American, Canadian, and Mexican customers. Write a query to gather the data needed for a mailing label. (1 points)
2. HR wants to celebrate hire date anniversaries for the sales representatives in the USA office. Develop a query that would give HR the information they need to coordinate hire date anniversary gifts. Sort the data as you see best fit. (1 points)
3. Customer service noticed an increase in shipping errors for orders handled by the employee, Janet Leverling. Return the OrderIDs handled by Janet so that the orders can be inspected for other errors. (1 points)
4. The warehouse packers want to label breakable products with a fragile sticker. Identify the products with glasses, jars, or bottles and are not discontinued (0 = not discontinued). (2 points)
5. Calculate the total order price per order and product before and after the discount. The products listed should only be for those where a discount was applied. Alias the before discount and after discount expressions. (2 points)
6. Supplier deliveries are confirmed via email and fax. Create a list of suppliers with a missing fax number to help the warehouse receiving team identify who to contact to fill in the missing information. (2 points)
7. Create a table each to identify the Low, Medium and High group companies based on their total order amount. Each table should contain the company name, the total order quantity and the group description that it belongs to. An example for Medium group (with partial results) is shown below. Note that you have to dynamically determine the group membership using the ranges in CustomerGroupThresholds table. (5 points)



**Custom Data Requests (10 points total – 3 points per request + 1 bonus)**

This is an exercise in asking good questions about data. Complex queries do not necessarily mean they produce higher value answers. Don't be afraid of simple solutions for important problems.

Connect to my lmu build database using the following credentials and select the database named **kcslmubu\_dw\_dillards\_small**. Create 3 data requests on this database. For each request, 1) write out a question to specify the required data and 2) give a business justification. How can a manager use the data for better decision making? 3) Write the SQL used to produce the final data set. 4) Export the final data set to a CSV file then convert it to an Excel spreadsheet with a .xlsx file extension. Collate all of the Excel files together in one file and name each worksheet according to the data request number, i.e. *data\_request\_1.xlsx, data\_request\_2.xslx*, etc.

Credentials to get into **kcslmubu\_dw\_dillards\_small**

**Server: kcs.lmu.build**

**User id: kcslmubu\_guest**

**Password: db4BSANcourses**

**Default Schema: kcslmubu\_dw\_dillards\_small**

Use the format below to document the custom data requests in your .SQL file.

/\*

# Data Request 1

Question

Business Justification

\*/

SQL statement(s)

You can use multiple queries to generate the data. You can modify the Excel spreadsheet to better communicate your findings and analysis. (highlight, bold, sort, graphs, etc.)

/\* Set II basket craft \*/

/\*

hostname: db.isba.co

username: analyst

password: go\_lions

database: basket\_craft

port: 3306

\*/

/\* quick look at the tables \*/

SELECT \*

FROM sessions

WHERE website\_session\_id = 1059;

SELECT \*

FROM pageviews

WHERE website\_session\_id = 1059;

SELECT \*

FROM orders

WHERE website\_session\_id = 1059;

SELECT \*

FROM order\_items

WHERE order\_id = 32;

SELECT \*

FROM products;

SELECT \*

FROM order\_item\_refunds

LIMIT 10;

# 8.0 (3 points)

/\*

INNER JOIN (JOIN)

Basket Craft wants to send a targeted email to customers who have placed an order.

Create a list of customer emails and the dates they placed an order.

\*/

=======================================================================

# 9.0(3 points)

/\*

Generate a report showing the number of orders placed by each customer.

This will help us identify our most active customers for targeted marketing efforts.

\*/

========================================================================

/\*

UTM Tracking Parameters (resource for understanding utm: <https://buffer.com/library/utm-guide/> )

\*/

SELECT DISTINCT utm\_source, utm\_campaign

FROM sessions

ORDER BY utm\_source;

=========================================================================

Traffic Source Analytics

# 10.0 (4 points)

/\*

Top Traffic Sources

From: Kara (CEO)

Subject: Site Traffic Breakdown

Date: April 12, 2021

We've been in business for nearly a month and have started to make sales. Could you assist me in understanding the primary sources of our website traffic up until yesterday (April 12, 2021)?

Business Question: ???

Expected Output: ???

\*/

SQL

/\*

Insight: ???

Recommendation: ???

\*/

========================================================================

# 11.0 (4 points)

/\*

Traffic Source Conversion Rates

From: Robert (Marketing Director)

Subject: Google Conversion Rate

Date: April 14, 2021

It appears that google nonbrand is the primary source of our website traffic, but we need to assess whether these sessions are resulting in actual sales. To do this, please calculate the Conversion Rate (CVR) from sessions to orders. Our target minimum CVR is 4% to ensure our advertising expenses are justified. If our CVR falls below this threshold, we will consider reducing our advertising bids. Conversely, if the CVR exceeds 4%, we can consider increasing our bids to generate more traffic and potential sales.

Business Question: ???

Expected Output: ???

\*/

SQL

/\*

Insight: ???

Recommendation: ???

Prediction: ???

\*/

=========================================================================

# 12.0 (4 points)

/\*

Traffic Source Bid Optimization

From: Robert (Marketing Director)

Subject: gsearch device-level performance

Date: May 11, 2021

While testing our website on my mobile device, I noticed that the user experience was suboptimal. Could you kindly retrieve the Conversion Rates (CVR) from sessions to orders categorized by device type? If we find that desktop performance surpasses mobile performance, we may consider increasing our bids for desktop to generate more traffic.

Business Question: ???

Expected Output: ???

\*/

SQL

/\*

Insight: ???

Recommendation: ???

Prediction: ???

\*/

=========================================================================

# 13.0 (2 points)

/\*

CASE to Categorize Data

Determine whether the total price of an order falls into the high, medium, or low category.

high: >= 100

medium: 50 - 99

low: < 50

\*/

SQL

=========================================================================

# 14.0 (5 points)

/\*

Top Landing Pages

From: Cheryl (Website Manager)

Subject: Top Entry Pages

Date: June 12, 2021

Can you pull a list of the most frequently visited entry pages on our website? I'd like to verify where our users are first landing when they visit the site.

Business Question:

???

Expected Output:

???

Concepts:

GROUP BY

Aggregate functions

CTE

\*/

SQL

/\*

Insight:

???

Recommendation:

???

Prediction:

???

\*/

=========================================================================

# 15.0 (5 points)

/\*

Bounce Rates

From: Cheryl (Website Manager)

Subject: Bounce Rate Analysis

Date: June 14, 2021

All our incoming traffic goes to the homepage. Let's assess the performance of this landing page. Could you please provide the following metrics related to the homepage:

Total Sessions

Bounced Sessions

Bounce Rate (Percentage of Sessions that Bounced)

STEP 1: find the first website\_pageview\_id for relevant sessions

STEP 2: identify the landing page URL of each session

STEP 3: count the # of pageviews for each session to identify bounces

If there's only one pageview, it means the user left the website after viewing just the landing page aka a bounce.

STEP 4: summarize total sessions and bounced sessions by landing page

Business Question:

???

Expected Output:

???

Concepts:

GROUP BY with HAVING

Multiple CTEs

\*/

SQL

/\*

Insight:

???

Recommendation:

???

Prediction:

???

\*/

=========================================================================