

Politecnico di Torino

Database systems

Laboratory 2

Google Data Studio - The MySQL connector

Prof. Silvia Chiusano
Prof. Alessandro Fiori
Andrea Avignone

2021-2022

1 Summary and objectives

As you remember from the previous laboratory activity, Google Data Studio is highly flexible to different data sources. If data samples and Google Sheets are probably the most intuitive, the other alternatives represent more convenient solutions to interact with real-life cases for data visualization. At the end of this laboratory, you should acquire some useful experience with:

- Leveraging the skills of the Data Studio environment learned in the last lab to have them applied to a new scenario.
- Investigating the structure of a complete database with multiple tables to perform the main SQL queries.
- Connecting your report with the **MySQL**¹ Relational Database Management System (RDBMS) by exploiting the provided Google Connector.
- Observing the preliminary usage of databases administration tools like **phpMyAdmin**² to have a graphical interface for managing your database.

2 General overview

Before facing the exercises, it is important to have a clear idea of the general overview we are dealing with. Conceptually, the basic aspects are exactly the same of the ones described in *Laboratory 1*. If you observe Figure 1, our objective is to use the proper connector to link our data report with the information stored in the selected database. In this case, we will remotely connect Google Data Studio with the associated MySQL instance to import our data.

In general, the system users have the possibility to interact with the database according to their privileges. The native approach for MySQL is to use the specific commands through the available MySQL shell. However, this method might be less intuitive and tricky. For this reason, graphical interfaces have been proposed to support users in creating, editing and visualizing databases. One of the most popular ones is certainly the open source tool *phpMyAdmin*, written in PHP. It was designed for performing the database-related maintenance operations.

¹<https://www.mysql.com>

²<https://www.phpmyadmin.net>

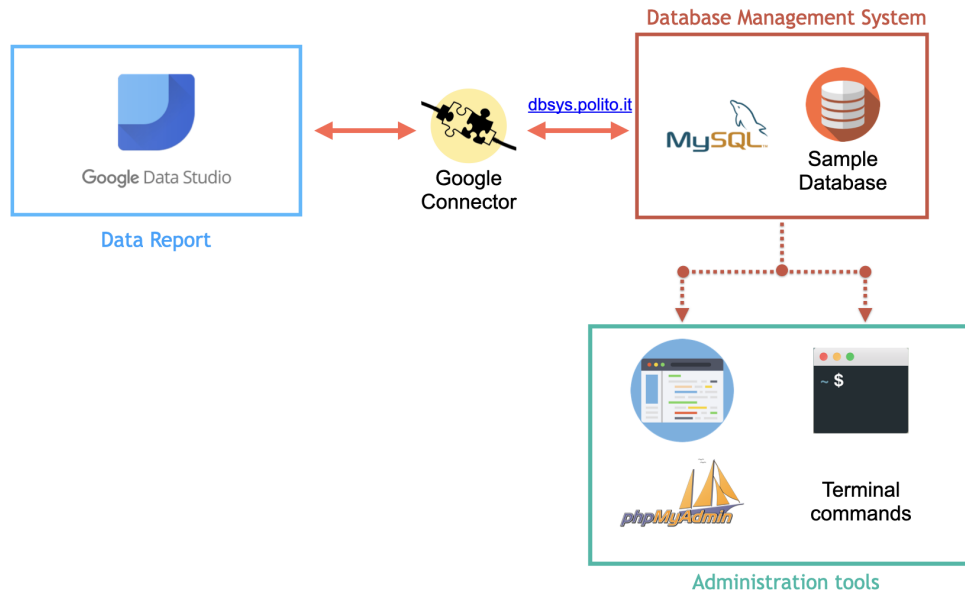


Figure 1: General overview - schema

3 Addressing the problem

Data are a powerful source of general details and crucial information for companies. Imagine that a retailer of scale models of classic cars asked you to generate a useful report highlighting the major insights about the corresponding sales, customers and staff.

In order to carry out this activity, they have made their database available for further analysis through MySQL connection (for more information, you can visit <https://www.mysqltutorial.org/mysql-sample-database.aspx>). In Figure 2, you can observe the schema reporting all the tables to better understand the given scenario. Therefore, your job consists in preparing a comprehensive dashboard exploiting your Data Studio and SQL skills. So, let's start!

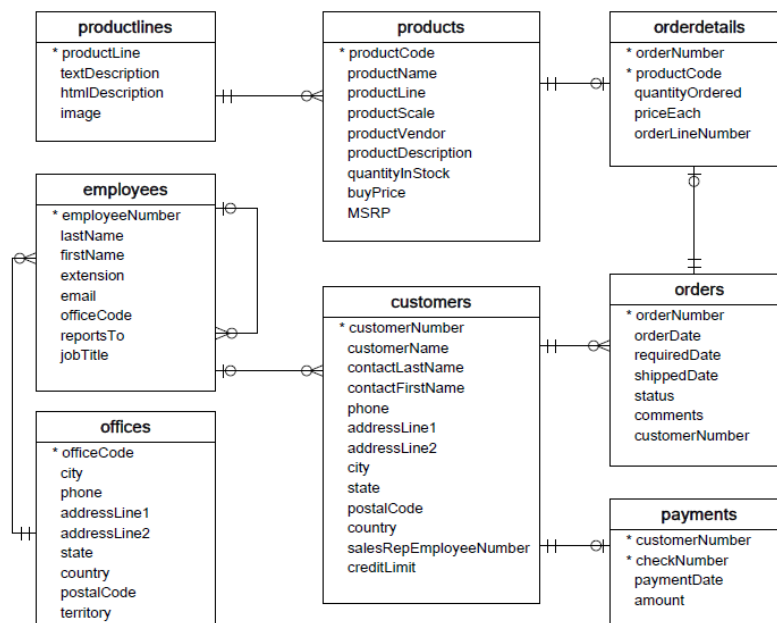


Figure 2: Sample Database

TIP. When you need to interact with a new database, it is suggested to spend some time analyzing the general schema to get familiar with the provided information and relationships.

4 Create the report

As always, the preliminary stage consists in connecting to <https://datastudio.google.com/> and performing the login if necessary. Once you have access to your personal page, create a new **Blank report** to begin.

4.1 Connection and authentication

The first automatic step is obviously to **Add data to report**. Among the *Google Connectors*, select **MySQL**. You will then be asked for database authentication (use *BASIC*), as shown in Figure 3. Use the following information to complete the form and press **Authenticate** at the bottom to connect. If the authentication is successful, a new section will appear showing the tables available for the selected database.

- Host Name: dbsys.polito.it
- Port: 3306 (default)
- Database: lab2
- Username: user
- Password: dbstudents22

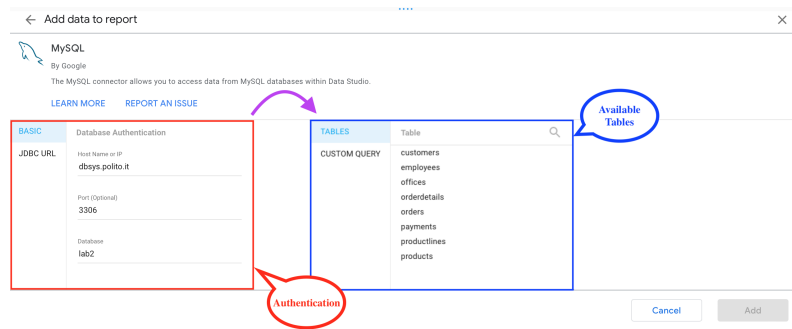


Figure 3: Authentication menu for MySQL connector

4.2 Tables selection

At this point, you can individually choose which table should be imported as data source from the corresponding list. This is helpful if you desire to investigate in your report any specific table using the **Add** button. However, the whole table is often unnecessary and inefficient to load. The smartest solution is then to exploit your SQL knowledge by defining the proper queries to simplify your work. To do that, click on **Custom Query**.

4.3 Warm up - Load your first table

First of all, you should get familiar with retrieving data and the related SQL commands. By using the custom query option you left pending in Section 4.2, prepare your report to meet the following company requirements.

Question 1. It is important to have a general overview of the products on sale. Define the proper query to import the corresponding data from the *products* table. In particular, we are interested in *productCode*, *productName*, *quantityInStock*, *buyPrice* and *MSRP* only.

TIP. You can easily verify the correctness of your query by taking a look to the right panel of the created table. Are the dimensions the expected ones?

4.4 Inspecting the data source

As we learned last time, a common step is to explore the data source. In this case especially, it can be extremely useful. Select **Resource** from the top menu and **Manage added data sources**. Press the **Edit** action for your data source to obtain something similar to Figure 4.

Before any modification, you can press **Edit Connection** on the left. This allows you to change the previously defined query according to your needs. Clearly, most of the tasks can be reproduced directly in Data Studio, but it is interesting to observe how changing the query affects the report.

Then, you can select **Reconnect** and observe the pop-up message of Google Data Studio to recap the new configuration. If everything is correct, press **Apply** to conclude.

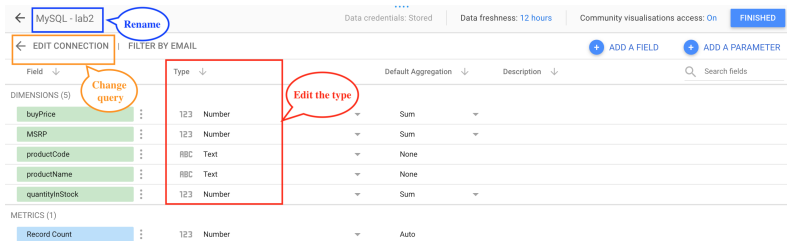


Figure 4: Data source - edit menu

Question 2a. Frequently, the final dashboard is expected to be shared with different people. Using the database column names as default may be indeed unclear. Update your SQL query to automatically rename all the selected dimensions other than *MSRP*.

Question 2b. Rename the data source with a more appropriate name. Finally, change the type of *buyPrice* and *MSRP* to be more consistent in your visual representation (i.e. from number to US Dollar). Next, save your changes with **Finished** and return to the report with the **Close** button.

4.5 Company products in a nutshell

Since the products data have been correctly imported, we can present the associated information in our dashboard. Prepare the report with title, header and layout you prefer.

Question 3. Create/update a table reporting the previous information about the available products. Use the right-side panel to select all the filtered dimensions (productCode, productName, quantityInStock, buyPrice, MSRP). The company wants the list about the models ordered by quantity in stock (showing at least the top 5 elements).

4.6 Payments and profit

Moving on to the other tables, we should include additional data sources. There are two methods: you can select **Add data** from the top menu as usual or you can **Duplicate** existing data sources from the **Manage added data sources** menu and change the desired query. The latter one avoids to fill in the form each time, but it is strongly not recommended due to performance issue.

Question 4a. Try to import the *payments* table as separate data source using the custom query. How should you write the command to import all the columns? Observe the dimensions if they need some additional change.

Question 4b. Create a specific box to summarize general performance indicators. Place a Scorecard reporting the total amount of received payments in US dollars (i.e. sum) supported by a *bar chart* showing the top 3 dates with higher amount (use the style section to choose the bars number). Can you spot which period of the year is the most beneficial?

TIP. Remember to rename the new data source, otherwise they are all assigned to the same default name!

4.7 Where are the customers from?

Things get more interesting if you look at the customer table. In particular, it would be appealing for the company to have a geographical overview of the origin of orders. Understanding the customers' specifications is always a key aspect to evaluate the performance and prepare specific strategies.

Question 5a. Insert a new data source to access customers information. Define efficiently the SQL query by importing only the necessary columns to provide the geographical insights about the orders origin. Then, try to use a **Bubble Map** for representation. Remember to choose the **Location** field and to set the useful metric for the bubbles size.

TIP. Once a new table has been imported, Data Studio automatically makes available the *Record Count* metric.

Question 5b. Keep updated the summary box designed for *Question 4b*. Prepare another Scorecard to report the total number of tracked customers. Use texts and shapes to explain your data!

4.8 Focus on US customers

Retrieving data is typically a matter of filters to highlight and deal with the essential information only. Until now we have limited ourselves to the basics of the *SELECT* command, but it is significantly useful to include some specific clauses in our queries with *WHERE*.

Question 6. To learn more about a particular type of clients, the company would expect to extract - based on your report - a list of clients from the U.S. (i.e. *USA*) with a credit limit greater than \$100k. After defining the custom query to filter data accordingly, insert a table reporting the customer name and the credit limit expressed in *USD*. Sort the results by credit limit (descending order).

4.9 Investigating order details

Once observed the customers, it is time to analyze the performed orders more in detail. In particular, this is helpful to gain confidence with some advanced constructs. In fact, while querying the database, you can apply specific mathematical functions to retrieve additional information.

Question 7a. Look again at the columns available for the *orderdetails* table. Can you define the SQL query to return the order numbers and the total amount of each order?

TIP. Remember to group your orders according to the order number. Moreover, renaming columns (especially if you apply mathematical functions) is always a good idea.

Question 7b. Complete the summary box with the information you have just retrieved. Create two Scorecards to express respectively the maximum and average amount of the given orders. In Google Data Studio, the faster approach is to use customization menu available for the considered metric (Figure 5). It is accessible from the pencil icon next to the metric name.

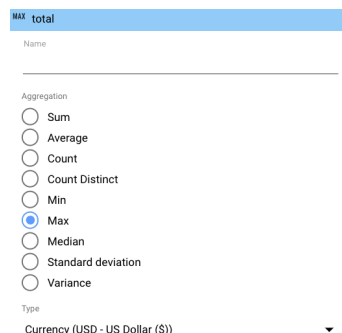


Figure 5: Metric menu

4.10 Staff and Offices

The last step consists in providing an intuitive overview about the staff members and the available offices. These additional insights make the report more complete and ensure that we have studied most of the database. At the end, you can design a report similar to the one shown in Figure 6.

Question 8a. Investigate the composition of the company staff. Use the *employees* table to select the job titles. Include in your report a **Doughnut chart** to understand the proportions. You should notice a particularly crowded department compared to the others. In this case, visualizing the absolute value of each slice is more useful to have an idea about the number of employees.

Question 8b. Given its multinational nature, the company is careful to emphasise the presence of offices around the world. Support the above analysis by delving into the information stored in the *offices* table. Define the proper query and prepare a *Bar chart* to present the number of offices per each country in descending order.

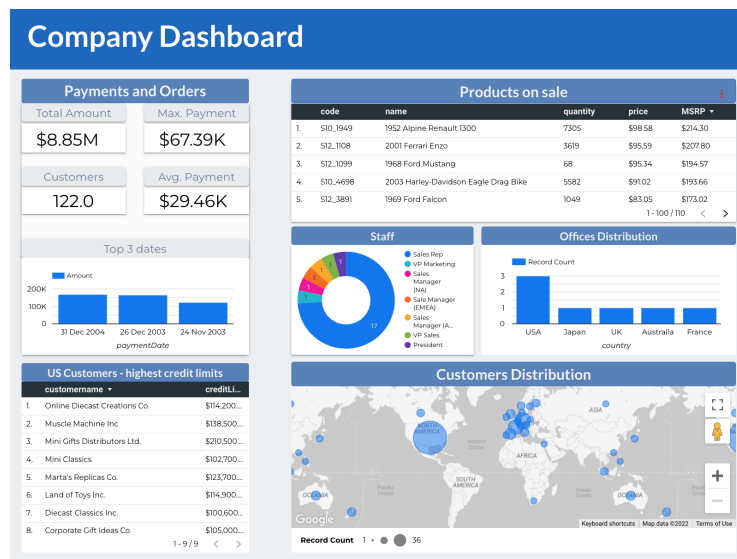


Figure 6: Final company dashboard

5 Customize and share the report

We have successfully imported all the necessary information to complete our dashboard according to the fixed requirements. Now, leave space to creativity. Customize your report layout to present a valuable, clear and informative work.

Question 9. Complete the report by focusing on the stylistic customization of each single block. Feel free to design it the exact way you find most convenient and observe the final result in **View** mode. Finally, test the **Share** button with any of your colleagues to perform the invitation to collaborate. As you can notice in Figure 7, you need to specify if other members can view or edit the project.

Sharing with others

Share as Andrea Avignone

[Add people](#) [Manage access](#)

Enter names or email addresses...

Can view

☒ Notify people

[Cancel](#) [Send](#)

Figure 7: Share project menu

6 PhpMyAdmin - brief introduction

As we have already mentioned, *phpMyAdmin* is an administration tool to manage your databases. It offers an intuitive GUI to create and visualize databases with several functions, including SQL queries, tables configuration and users privileges.

For your final project, your group will receive the credential to connect using your browser. The login page is shown in Figure 8 in which you have also the possibility to choose the language.

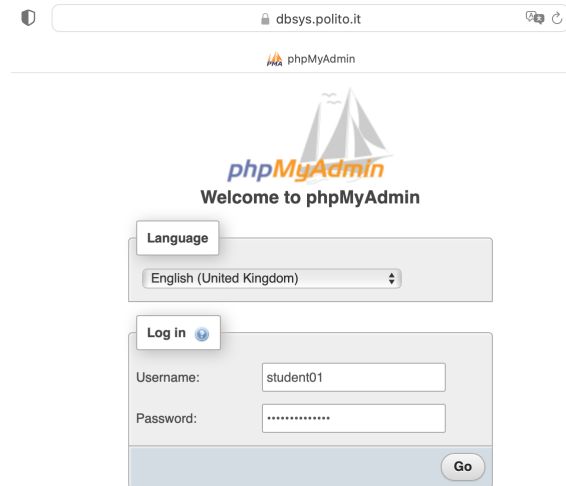


Figure 8: PhpMyAdmin - login page

Once logged in, the homepage is extremely user-friendly. As you can observe in Figure 9, on the left you can explore the databases you have access to. This is helpful to explore the database structure and the information stored for each table. In fact, you can click on each single column to learn more.

On the top instead, there is the toolbar you can use for all the actions you need. For example, from **Databases** you can access the existing ones or create a new database according to the desired specifications (if you have the privileges). By selecting a specific database instead, you can explore the structure, create new tables or search for specific fields.

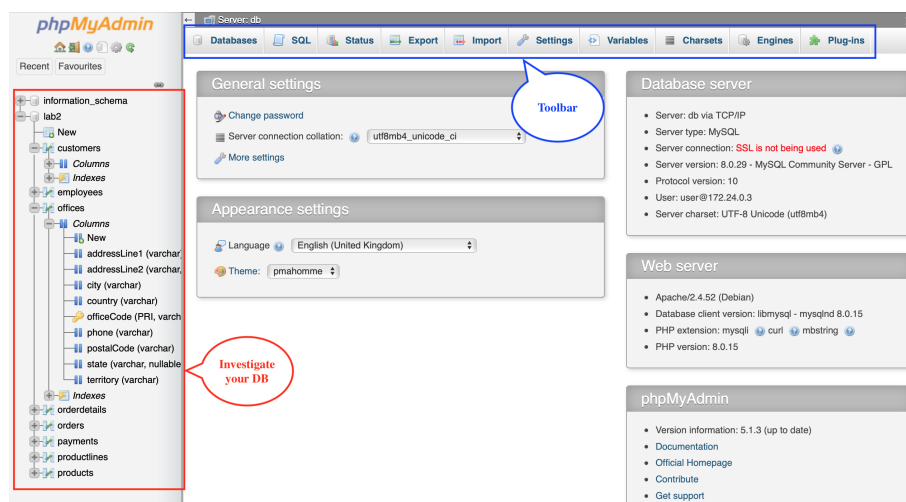
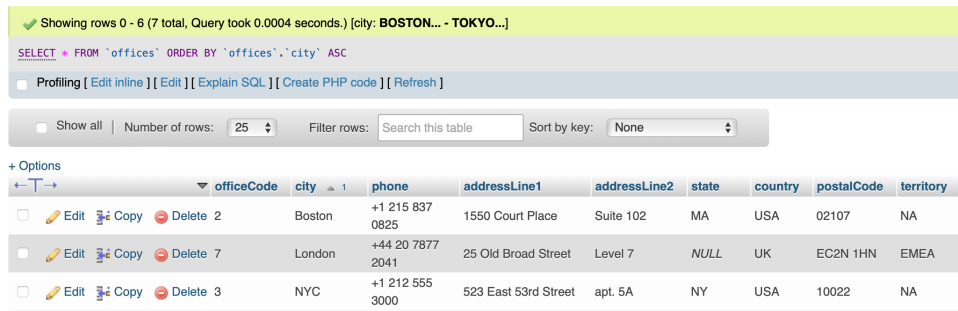


Figure 9: PhpMyAdmin - homepage

It is interesting that when you browse the available tables, *phpMyAdmin* automatically generates the corresponding query. In Figure 10 the *city* column was clicked to sort by city and the query on the top was then obtained.

If the previous feature is important to get familiar with SQL queries while using a graphical interface, to take a step further you should refer to Figure 11. Clearly, *phpMyAdmin* supports the possibility to integrate SQL scripts to execute different queries. This is useful to be completely free of facing your data and creating new tables if you need them.



Showing rows 0 - 6 (7 total, Query took 0.0004 seconds.) [city: BOSTON... - TOKYO...]

SELECT * FROM `offices` ORDER BY `offices`.`city` ASC

Profiling [Edit inline] [Edit] [Explain SQL] [Create PHP code] [Refresh]

Show all | Number of rows: 25 | Filter rows: Search this table | Sort by key: None

+ Options

	officeCode	city	phone	addressLine1	addressLine2	state	country	postalCode	territory
<input type="checkbox"/> Edit Copy Delete	2	Boston	+1 215 837 0825	1550 Court Place	Suite 102	MA	USA	02107	NA
<input type="checkbox"/> Edit Copy Delete	7	London	+44 20 7877 2041	25 Old Broad Street	Level 7	NULL	UK	EC2N 1HN	EMEA
<input type="checkbox"/> Edit Copy Delete	3	NYC	+1 212 555 3000	523 East 53rd Street	apt. 5A	NY	USA	10022	NA

Figure 10: PhpMyAdmin - browsing tables

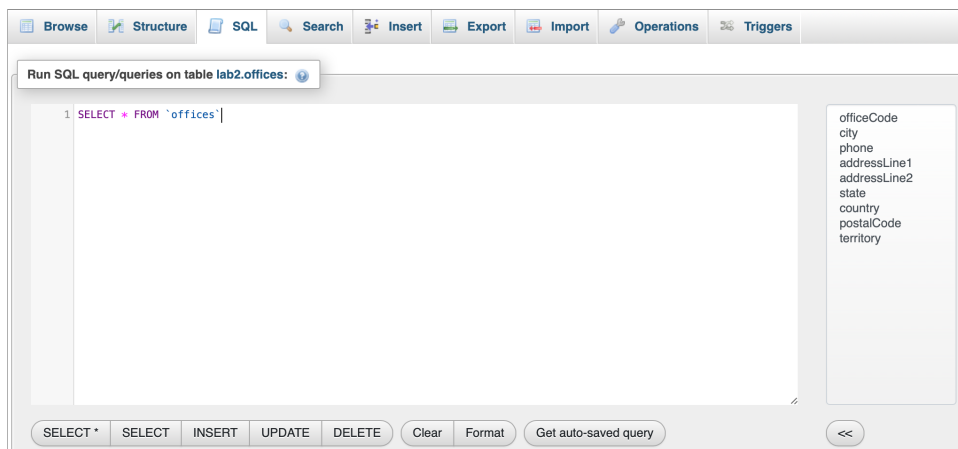


Figure 11: PhpMyAdmin - SQL scripts

However, *phpMyAdmin* offers several features and operations that can be extremely useful. Besides those already mentioned, we can briefly summarize the main possibilities:

- **Create/remove/rename** databases, tables and fields, with support to SQL queries.
- **Status.** Retrieve information about the traffic (received and sent) and connections concerning the MySQL server.
- **Export.** Export databases from the server according to the desired format. You can choose among different file types, including: SQL, JSON, CSV, PHP array. This is useful to migrate databases, associated with the import function.
- **Import.** Upload your databases into the server based on a specific file (format examples: SQL, CSV, XML).
- **Settings.** Customize user preferences related to the current session.
- **Documentation.** PhpMyAdmin is certainly a user-friendly tool and it is easy to gain confidence with. In addition, it offers the detailed documentation to learn more about installation, configuration and general features.
- **Designer.** For each database, users can access this graphical function to create, edit or understand the database schema (similarly to the one previously shown in Figure 2)