Data Technician

Name:

Course Date:

Table of contents

Day 1: Task 1

Please research and complete the below questions relating to key concepts of databases.

What is a primary key?	A primary key is a column (or set of columns) in a table that uniquely identifies each record.
How does this differ from a secondary key?	A secondary key is a field or combination of fields used for searching or sorting, it is not required to be unique, it is not used to enforce entity integrity.
How are primary and foreign keys related?	A foreign key in one table is a field that references the primary key of another table. They establish relationships between tables and enforce referential integrity, ensuring that a referenced record actually exists.
Provide a real-world example of a one-to-one relationship	Person ↔ Passport One person has exactly one passport and only each passport belongs to exactly one person.
Provide a real-world example of a one-to-many relationship	School ↔ Pupil A School can have many Pupils however each Pupil can only have one School.
Provide a real-world example of a many-to-many relationship	Students ↔ Courses One student can enroll in many Courses and each course can have many Students

Day 1: Task 2

Please research and complete the below questions relating to key concepts of databases.

What is the difference between a relational and non-relational database?	A **relational database (SQL)** stores data in tables made of rows and columns. It uses a fixed schema, meaning the structure of the data must be defined in advance. Relational databases support relationships through primary and foreign keys, which help maintain strong data integrity. A **non-relational database (NoSQL)** stores data in flexible formats such as documents (JSON-like), key-value pairs, graphs, or wide-column structures. It uses a flexible or schema-less design, making it well-suited for handling large, fast-changing, complex, or unstructured data.
What type of data would benefit off the non-relational model? **Why?**	The type of data that would benefit from the non-relational model is big, large and unstructured data or semi-structured data. non-relational databases are ideal for handling data which changes often. It is easier to store larger volumes.

Day 3: Task 1

Please research the below 'JOIN' types, explain what they are and provide an example of the types of data it would be used on.

Self-join	Self-join is when a table is joined to itself. It is used when rows in the same table can relate to each other. Example of use: A self join can match each employee to their manager.
Right join	Right join allows all rows from the right table to join with matching rows from the left table, however non-matching rows from the left table return as NULL. Example of use: Show all customers even those with no orders (NULL).
Full join	full join allows all rows from both tables with matching rows to be combined, however non-matching rows will be filled with NULL Example of use: Comparing two list such as Old_Product_List and New_Product_List
Inner join	Inner join returns only rows that match on both tables. Example of use: Students who are enrolled.
Cross join	A cross join pairs every row in the left table with every row in the right table using every possible combination. Example of use: To generate all possible product variations
Left join	Left join allows all rows from the left table to join with matching rows from the right table, however non-matching rows from the right table return as NULL. Example of use: A self join can match each employee to their manager.

Day 4: Task 1: Written

In your groups, discuss and complete the below activity. You can either nominate one writer or split the elements between you. Everyone however must have the completed work below:

Imagine you have been hired by a small retail business that wants to streamline its operations by creating a new database system. This database will be used to manage inventory, sales, and customer information. The business is a small corner shop that sells a range of groceries and domestic products. It might help to picture your local convenience store and think of what they sell. They also have a loyalty program, which you will need to consider when deciding what tables to create.

Write a 500-word essay explaining the steps you would take to set up and create this database. Your essay should cover the following points:

1. ***Understanding the Business Requirements**:*
    a. *What kind of data will the database need to store?*
    b. *Who will be the users of the database, and what will they need to accomplish?*
2. ***Designing the Database Schema**:*
    a. *How would you structure the database tables to efficiently store inventory, sales, and customer information?*
    b. *What relationships between tables are necessary (e.g., how sales relate to inventory and customers)?*
3. ***Implementing the Database**:*
    a. *What SQL commands would you use to create the database and its tables?*
    b. *Provide examples of SQL statements for creating tables and defining relationships between them.*
4. ***Populating the Database**:*
    a. *How would you input initial data into the database? Give examples of SQL INSERT statements.*
5. ***Maintaining the Database**:*
    a. *What measures would you take to ensure the database remains accurate and up to date?*
    b. *How would you handle backups and data security?*

Your essay should include specific examples of SQL commands and explain why each step is necessary for creating a functional and efficient database for the retail business.

Please write your 500-word essay here	1. Building a database and storing key information regarding the store's products, customers, sales and trends.

This can include:

Inventory data such as **product names, categories, costs, stock levels** and **supplier information**.
Example: Biscuits (20 units in stock, £1 each), Butter (30 units in stock, 60p each).
Sales - this includes what was sold, when it was sold, the quantity purchased, and the method of payment.
Customer data such as contact information, names, and loyalty points.
Loyalty data from loyalty programs including the points gained and spent on each transaction.

Users of the database:
Shop staff: They will track sales, manage loyalty points, update transactions and inventory.
Manager: They will track customer loyalty activity and trends, view sales reports, update product descriptions, and check stock levels.
IT: They would manage the database, back up data and guarantee data security.

2.
Structuring the database as a collection of tables, each of them corresponding to distinct types of data. The tables needed for this shop:

Products table example:

Product ID	Product Name	Category	Price	Stock Level	Supplier
1	Butter	Dairy	£1.00	100	Farm 4 u
2	Biscuits	baked goods	0.60p	200	Crumbs

Customers table example:

Customer ID	Full Name	D.O.B	Email	Loyalty Points	Join date
1	John Spark	21.04.1992	J.Sparks@emaple.com	100	02.11.2025
2	Sven Marks	03.03.2002	S.Marks@example.com	200	04.11.2025

Sales table example:

Sale ID	Product ID	Customer ID	Date	Quantity
1	2	2	04.12.2025	2
2	1	2	03.12.2025	3

Connections between the tables by linking the primary key in each table to the corresponding foreign key in the other table:
Products and Sales: allows to update stock, each sale will make reference to a particular product and calculate revenue.
Sales and Customers: A customer's transactions should link to their record if they are a member of the loyalty program and trends.

3.

Building the database using tools such as MySQL or Microsoft Access

Customer Table	Suppliers Table	Inventory Table
CREATE TABLE `Customer` (`Loyalty_Card_Number` <type>, `First_Name` <type>, `Loyalty_Points` <type>, `Email` <type>, `Contact_Number` <type>, `Last_Name` <type>, `DOB` <type>, `Date_of_join` <type>, PRIMARY KEY (`Loyalty_Card_Number`), KEY `Key` (`First_Name`, `Loyalty_Points`, `Email`, `Contact_Number`, `Last_Name`, `DOB`, `Date_of_join`));	CREATE TABLE `Suppliers` (`Supplier_ID` <type>, `Supplier_Name` <type>, `Email` <type>, `Contact_Number` <type>, `Supplier_Address` <type>, `Supplier_Postcode` <type>, PRIMARY KEY (`Supplier_ID`), KEY `Key` (`Supplier_Name`, `Email`, `Contact_Number`, `Supplier_Address`, `Supplier_Postcode`));	CREATE TABLE `Inventory` (`Product_ID` <type>, `Supplier_ID` <type>, `Product_Name` <type>, `Product_Category` <type>, `Selling_Price` <type>, `Stock_Quantity` <type>, `Stock_Count` <type>, PRIMARY KEY (`Product_ID`), FOREIGN KEY (`Supplier_ID`) REFERENCES `Suppliers`(`Supplier_ID`), KEY `Key` (`Product_Name`, `Product_Category`, `Selling_Price`, `Stock_Quantity`, `Stock_Count`));

4.

To add new information, an SQL INSERT command allows you to place new rows of data into each table. or a spreadsheet to manually enter data, such as Excel.

An example of this would be:

Product Table	Customers Table
INSERT INTO Products (Name, Price, StockQty) VALUES ('Orange juice', 1.20, 50)	INSERT INTO Customers (FullName, Email, LoyaltyPoints) VALUES ('Ali Smith', 'A.Smith@example.com', 10);

Or importing data from pre-existing files like Excel. Spreadsheets can be uploaded into the system.

5.

To maintain the database is accurate and up-to-date:

Update stock levels to keep the inventory accurate on a regular basis following each transaction.

Check for errors, look for mistakes such as duplicate clients or inaccurate prices, and correct them.

Update product details when new products are introduced or when pricing is changed.

Clean data Remove unnecessary information, such as discontinued products.

Permissions only allow personnel with the necessary training to change information like stock or prices, this minimises the risk of errors.

Backups and Security Make regular backups of the database and store backups in a safe place.

Using passwords so only authorised users can access the data and protect sensitive info and keep software up to date.

These steps help keep the information reliable for both staff and managers.

Day 4: Task 2: SQL Practical

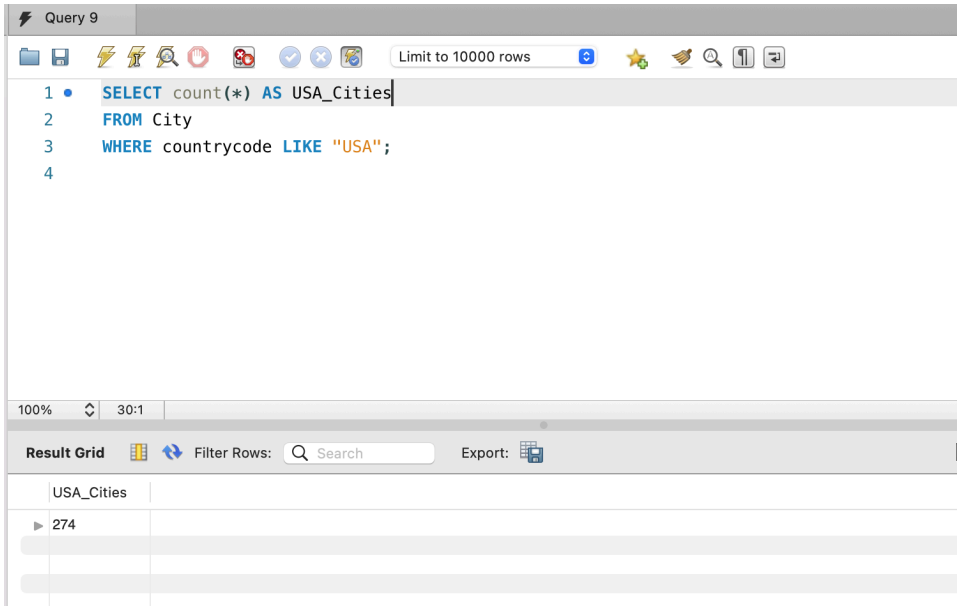
In your groups, work together to answer the below questions. It may be of benefit if one of you shares your screen with the group and as a team answer / take screen shots from there.

Setting up the database:

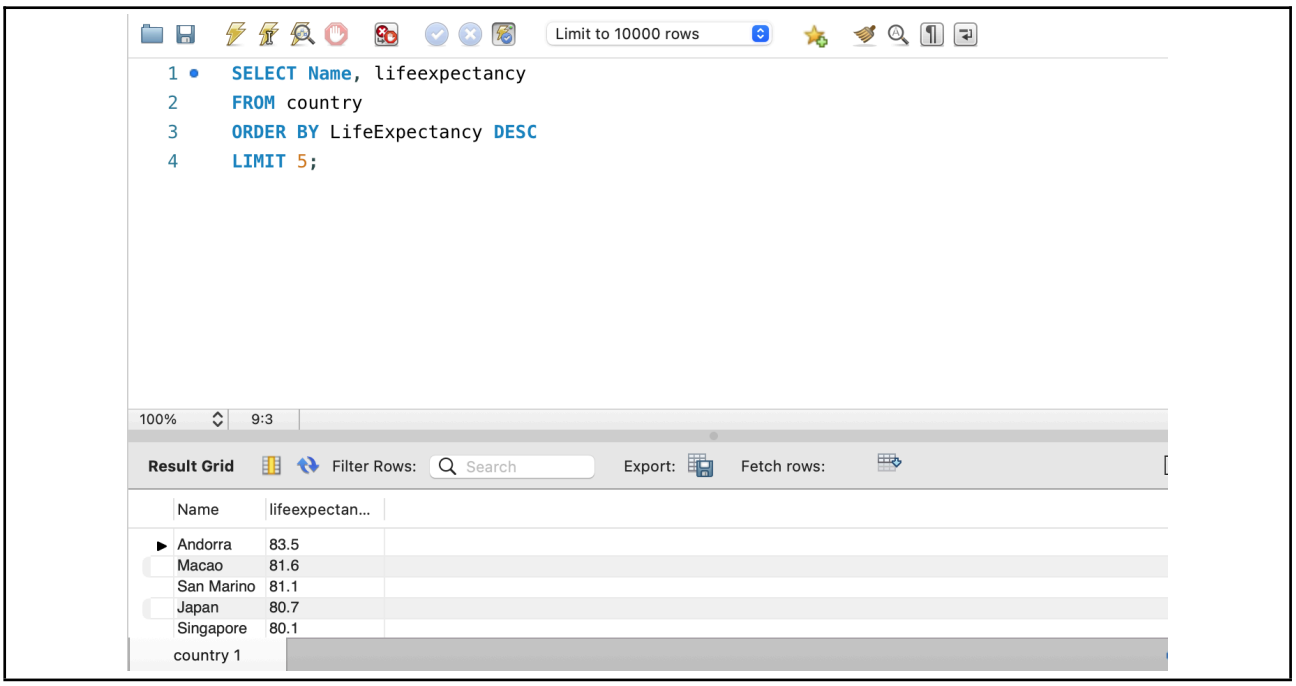
1. **Download world_db(1)**
2. **Follow each step to create your database**

For each question I would like to see both the syntax used and the output.

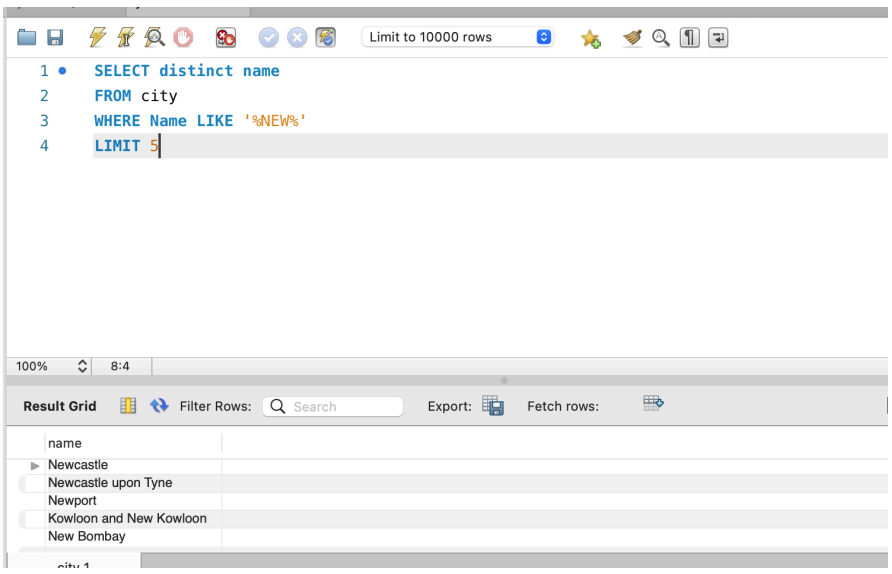
1. **Count Cities in USA:** *Scenario:* You've been tasked with conducting a demographic analysis of cities in the United States. Your first step is to determine the total number of cities within the country to provide a baseline for further analysis.



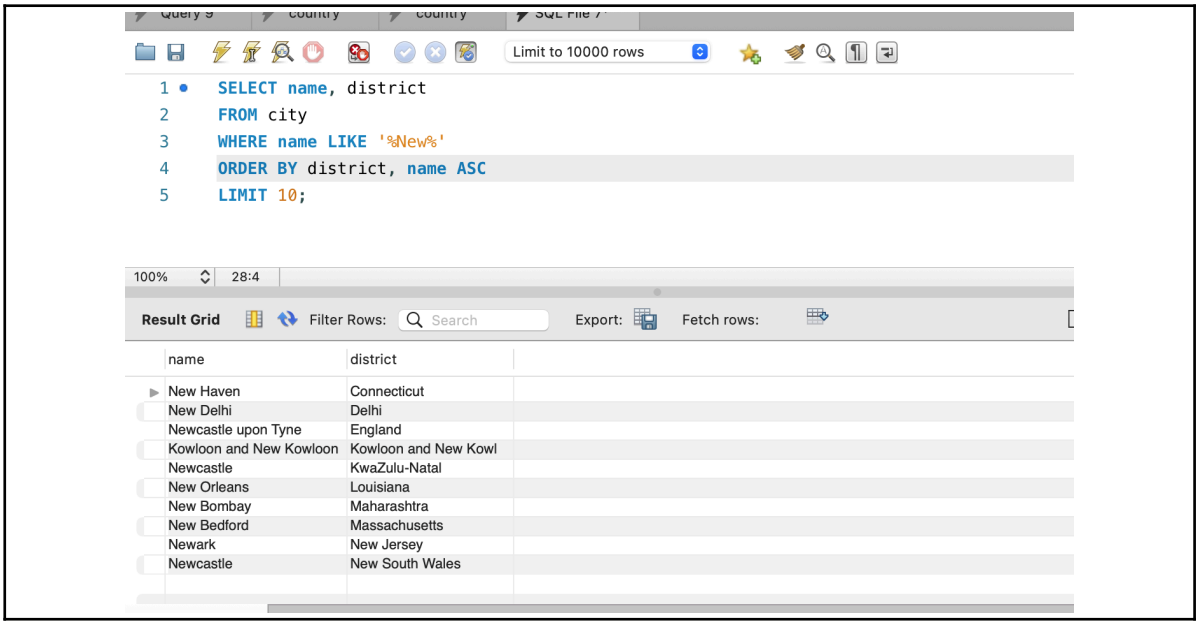
2. **Country with Highest Life Expectancy:** *Scenario:* As part of a global health initiative, you've been assigned to identify the country with the highest life expectancy. This information will be crucial for prioritising healthcare resources and interventions.



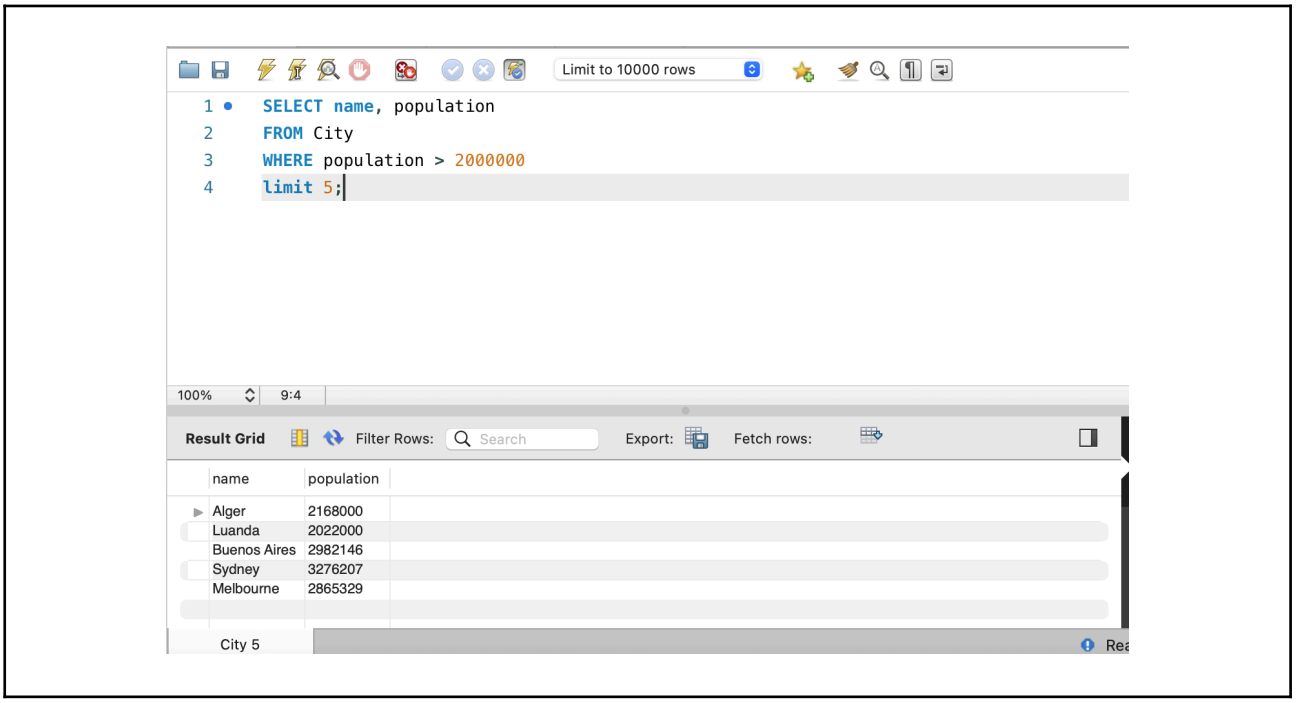
3. **"New Year Promotion: Featuring Cities with 'New :** *Scenario:* In anticipation of the upcoming New Year, your travel agency is gearing up for a special promotion featuring cities with names including the word 'New'. You're tasked with swiftly compiling a list of all cities from around the world. This curated selection will be essential in creating promotional materials and enticing travellers with exciting destinations to kick off the New Year in style.



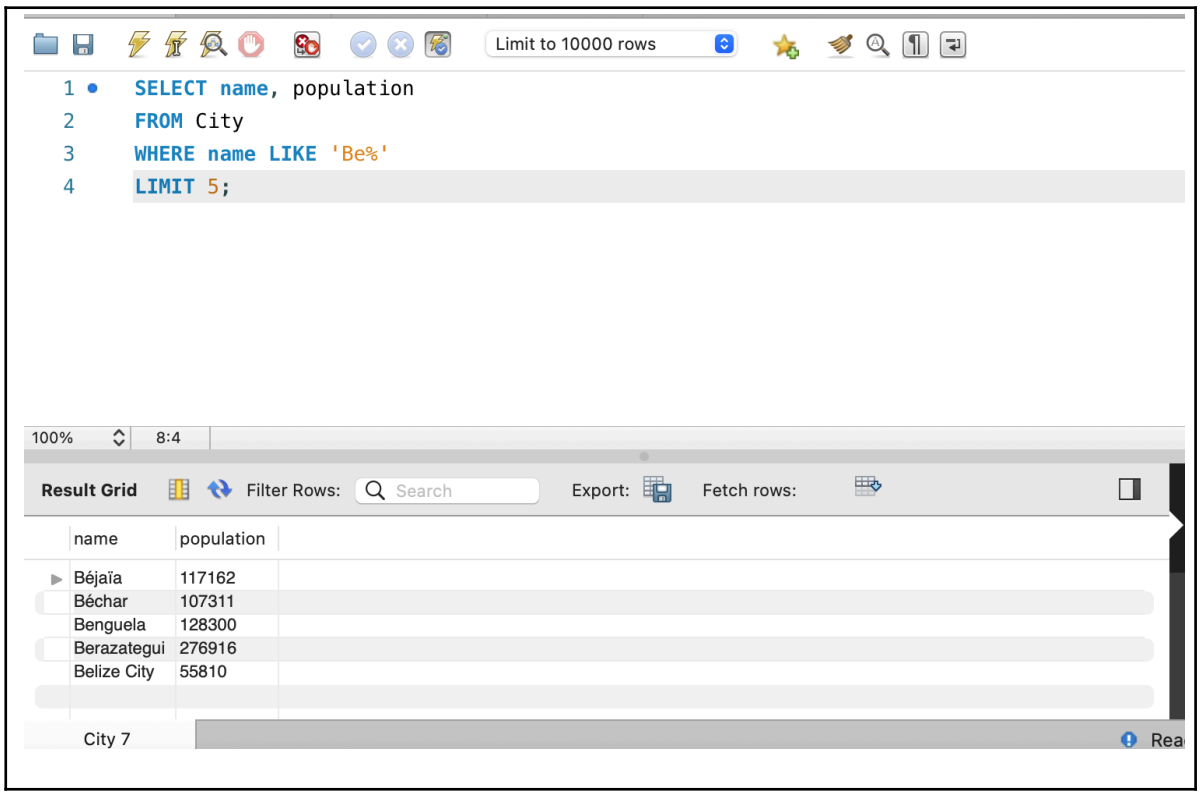
4. **Display Columns with Limit (First 10 Rows):** *Scenario:* You're tasked with providing a brief overview of the most populous cities in the world. To keep the report concise, you're instructed to list only the first 10 cities by population from the database.



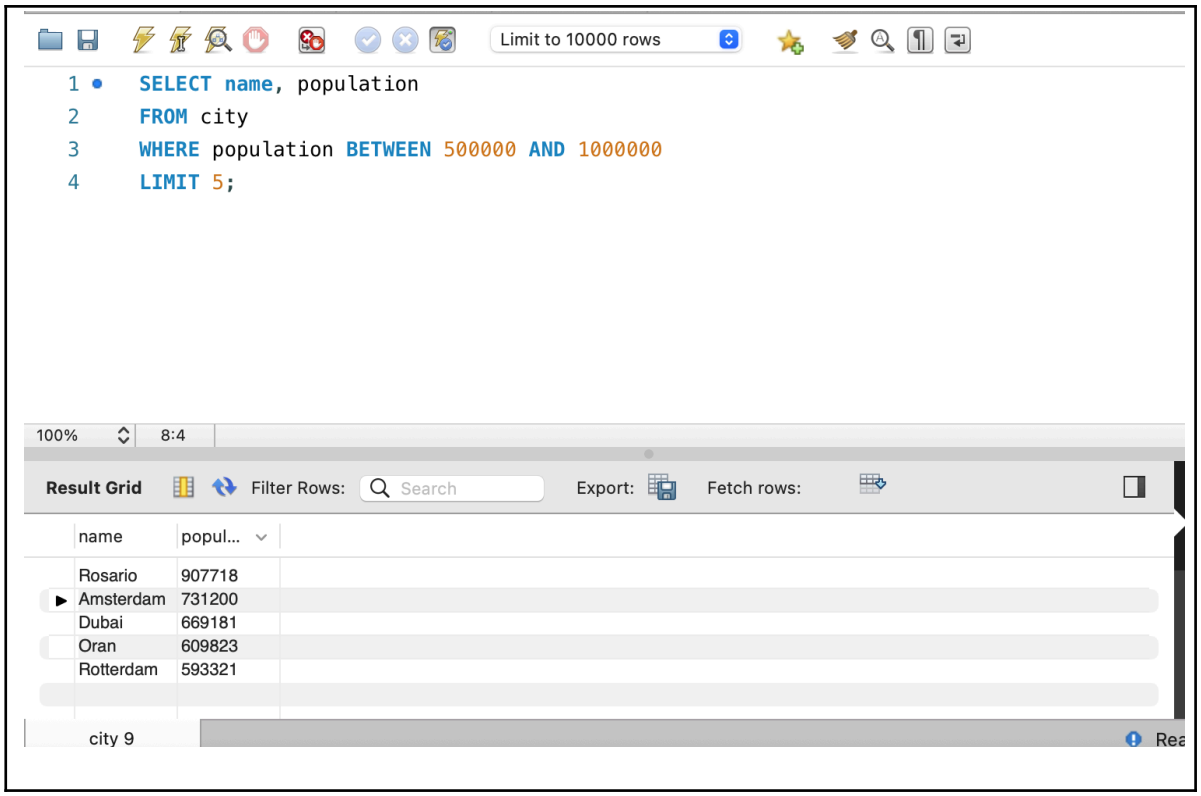
5. **Cities with Population Larger than 2,000,000:** *Scenario:* A real estate developer is interested in cities with substantial population sizes for potential investment opportunities. You're tasked with identifying cities from the database with populations exceeding 2 million to focus their research efforts.



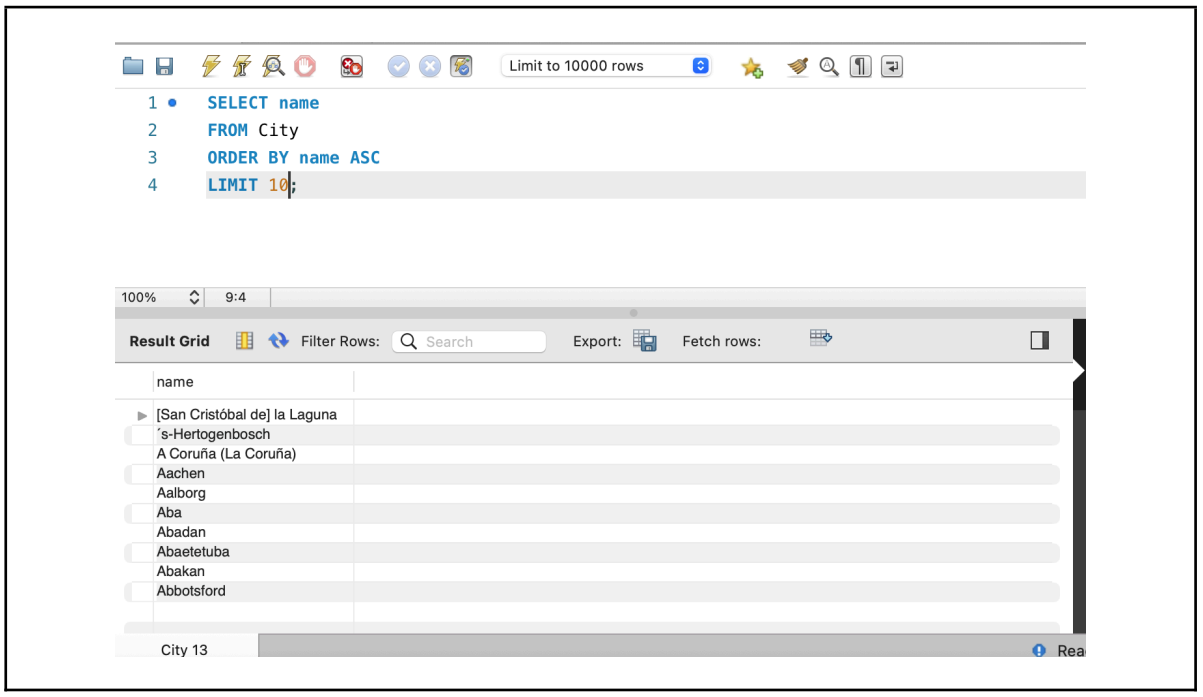
6. **Cities Beginning with 'Be' Prefix:** *Scenario:* A travel blogger is planning a series of articles featuring cities with unique names. You're tasked with compiling a list of cities from the database that start with the prefix 'Be' to assist in the blogger's content creation process.



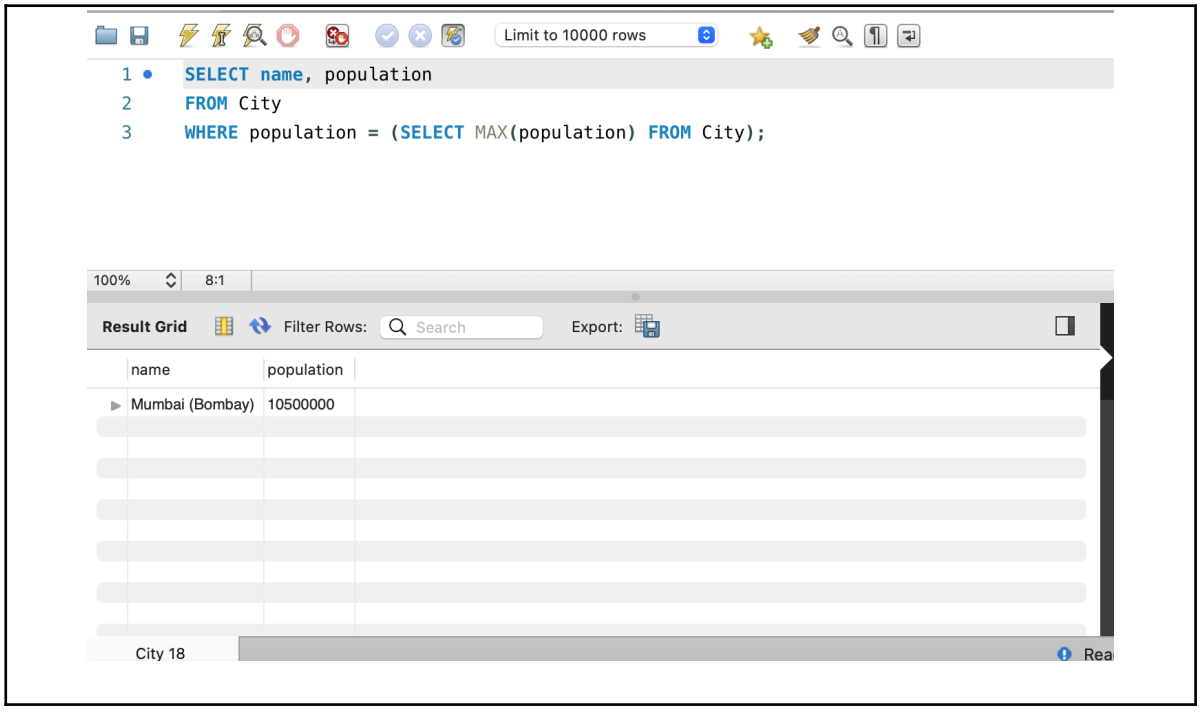
Cities with Population Between 500,000-1,000,000: *Scenario:* An urban planning committee needs to identify mid-sized cities suitable for infrastructure development projects. You're tasked with identifying cities with populations ranging between 500,000 and 1 million to inform their decision-making process.



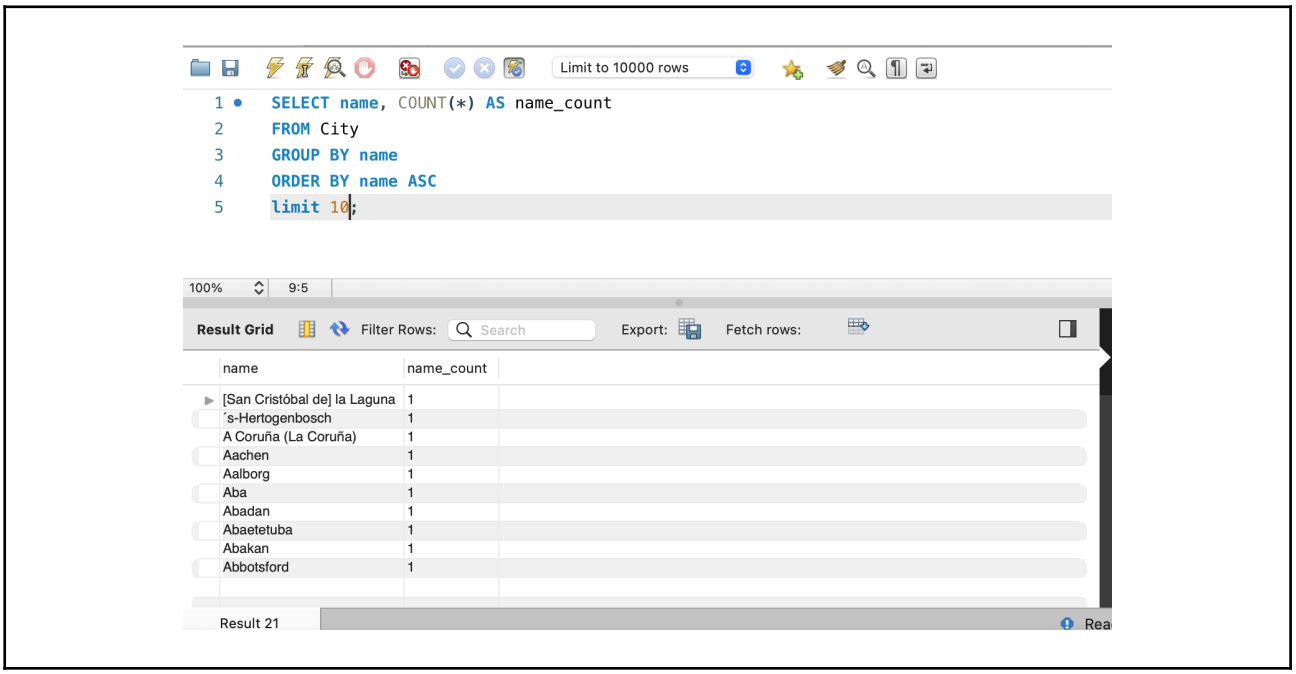
8. **Display Cities Sorted by Name in Ascending Order:** *Scenario:* A geography teacher is preparing a lesson on alphabetical order using city names. You're tasked with providing a sorted list of cities from the database in ascending order by name to support the lesson plan.



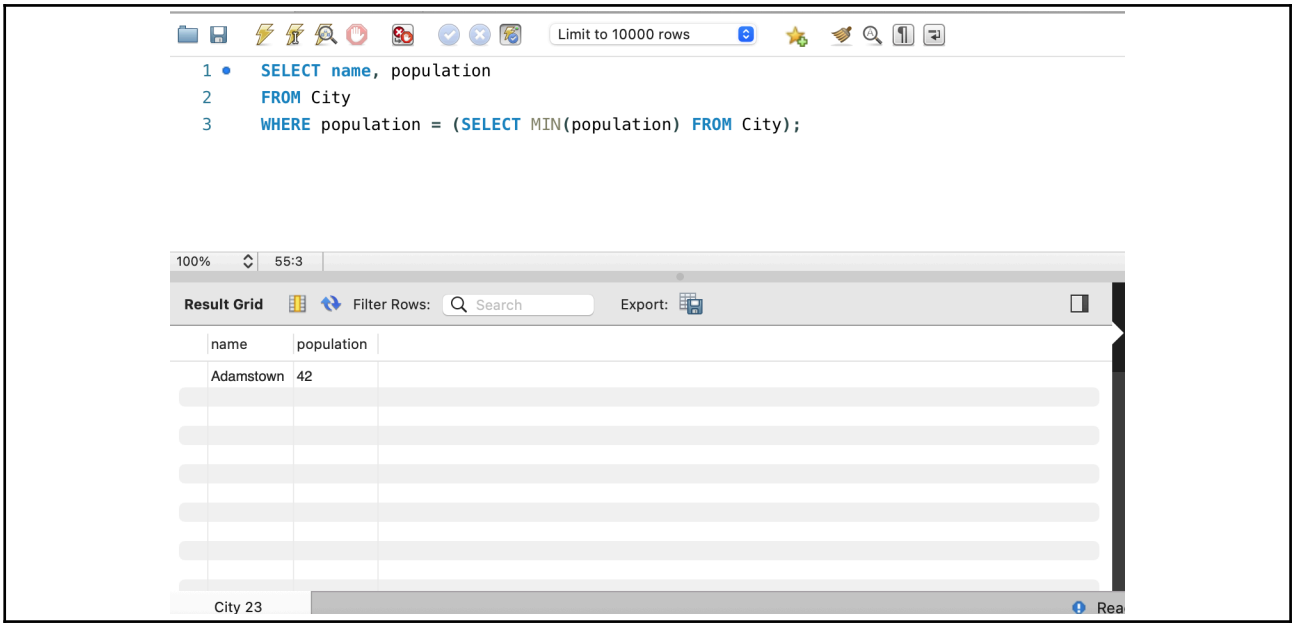
9. **Most Populated City:** *Scenario:* A real estate investment firm is interested in cities with significant population densities for potential development projects. You're tasked with identifying the most populated city from the database to guide their investment decisions and strategic planning.



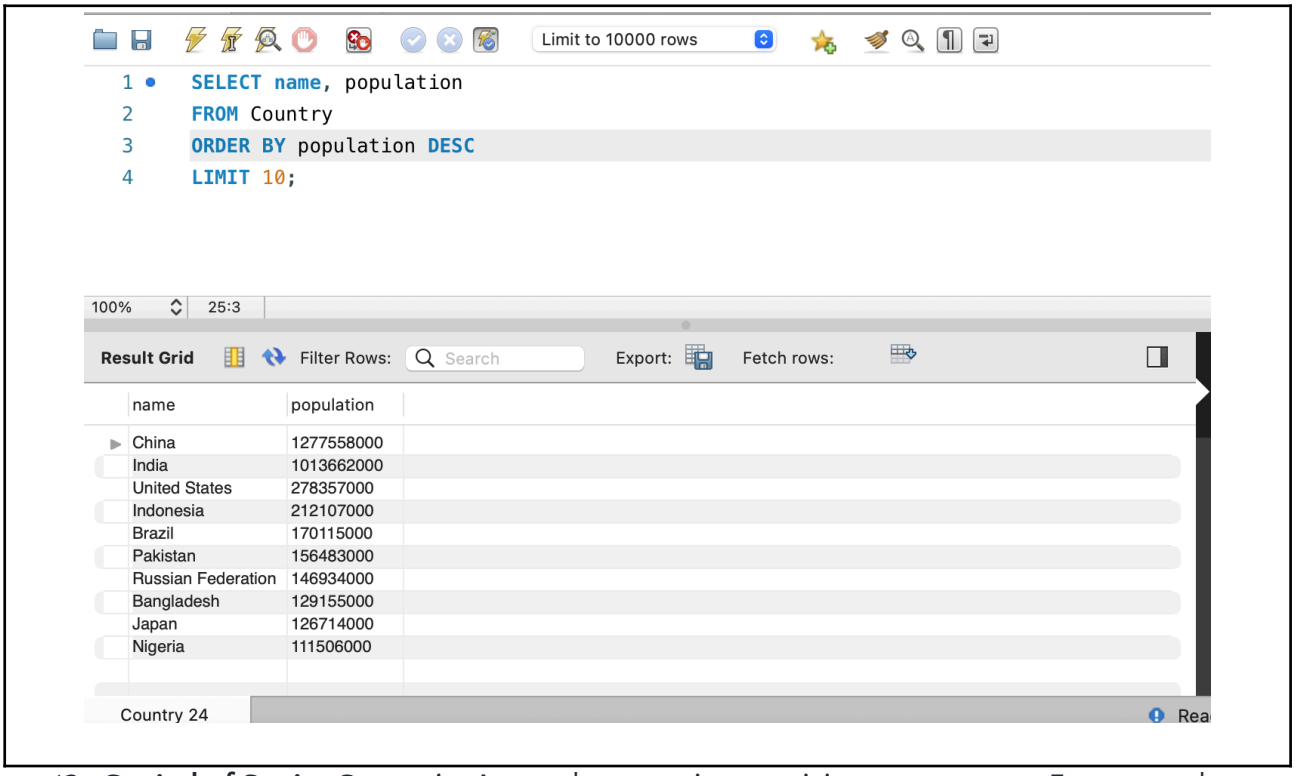
10. **City Name Frequency Analysis: Supporting Geography Education** *Scenario*: In a geography class, students are learning about the distribution of city names around the world. The teacher, in preparation for a lesson on city name frequencies, wants to provide students with a list of unique city names sorted alphabetically, along with their respective counts of occurrences in the database. You're tasked with this sorted list to support the geography teacher.



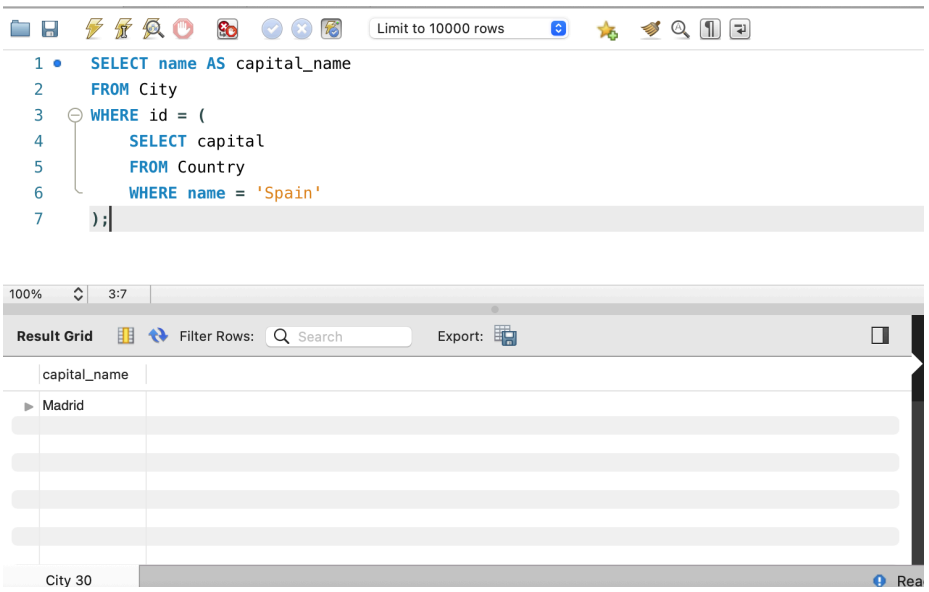
11. **City with the Lowest Population:** *Scenario:* A census bureau is conducting an analysis of urban population distribution. You're tasked with identifying the city with the lowest population from the database to provide a comprehensive overview of demographic trends.



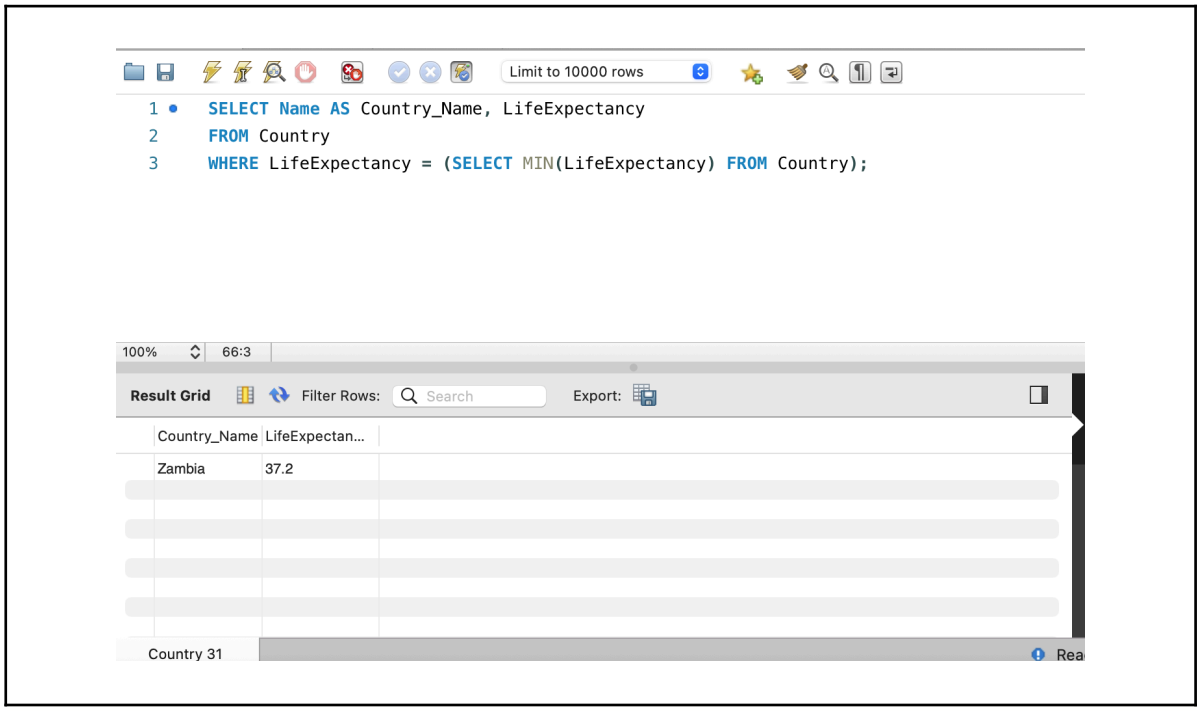
12. **Country with Largest Population:** *Scenario:* A global economic research institute requires data on countries with the largest populations for a comprehensive analysis. You're tasked with identifying the country with the highest population from the database to provide valuable insights into demographic trends.



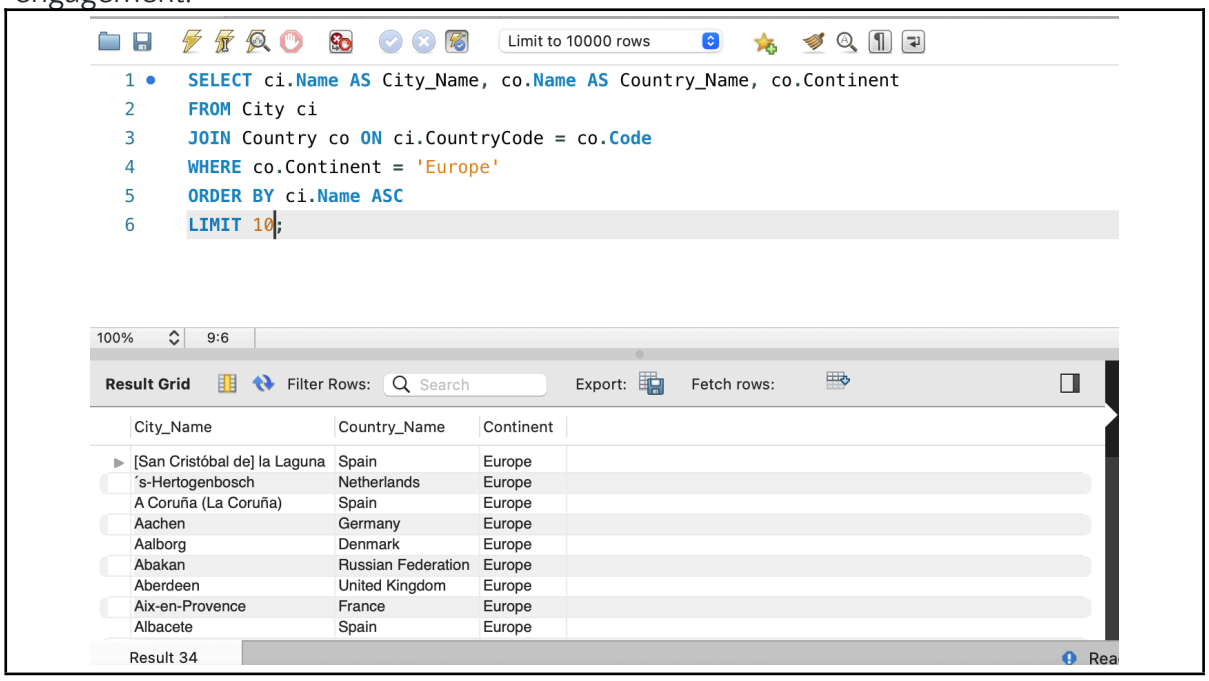
13. **Capital of Spain:** *Scenario:* A travel agency is organising tours across Europe and needs accurate information on capital cities. You're tasked with identifying the capital of Spain from the database to ensure itinerary accuracy and provide travellers with essential destination information.



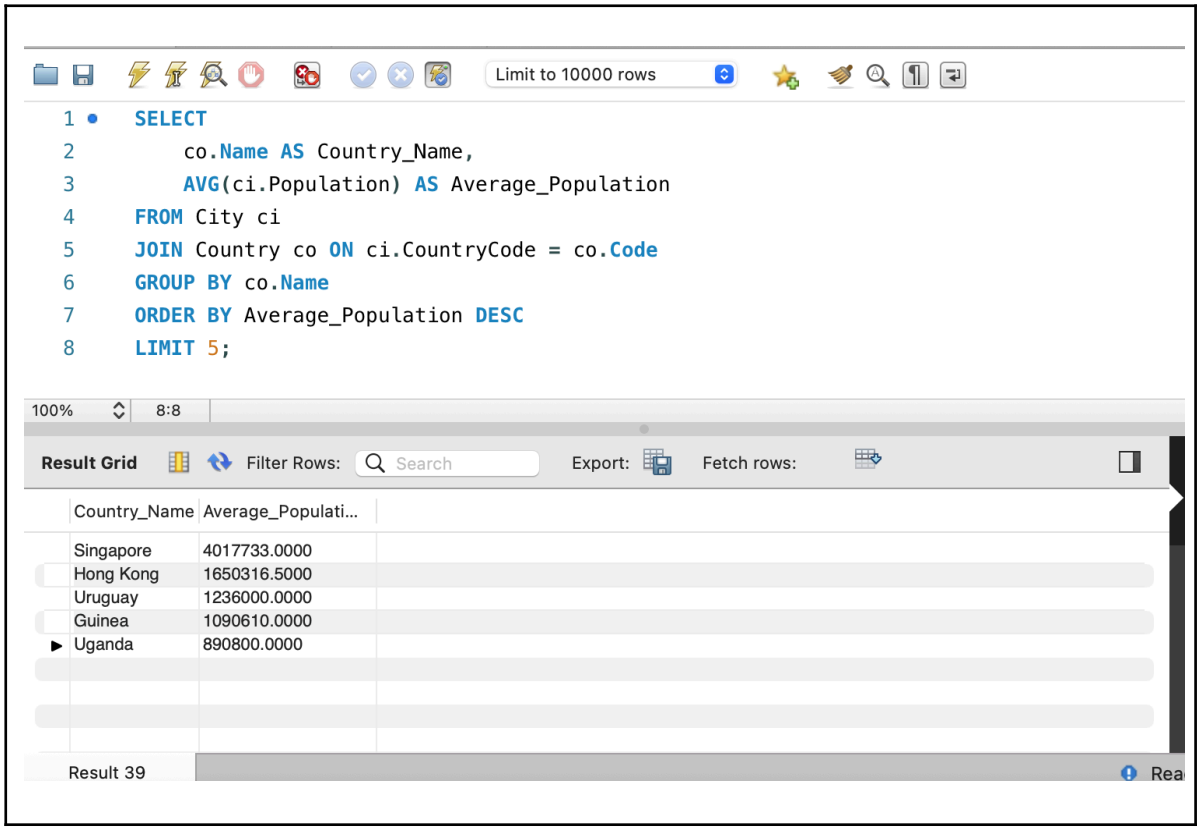
14. **Country with Shortest Life Expectancy:** *Scenario:* A healthcare foundation is conducting research on global health indicators. You're tasked with identifying the country with the highest life expectancy from the database to inform their efforts in improving healthcare systems and policies.



15. **Cities in Europe:** *Scenario:* A European cultural exchange program is seeking to connect students with cities across the continent. You're tasked with compiling a list of cities located in Europe from the database to facilitate program planning and student engagement.



16. **Average Population by Country:** *Scenario:* A demographic research team is conducting a comparative analysis of population distributions across countries. You're tasked with calculating the average population for each country from the database to provide valuable insights into global population trends.

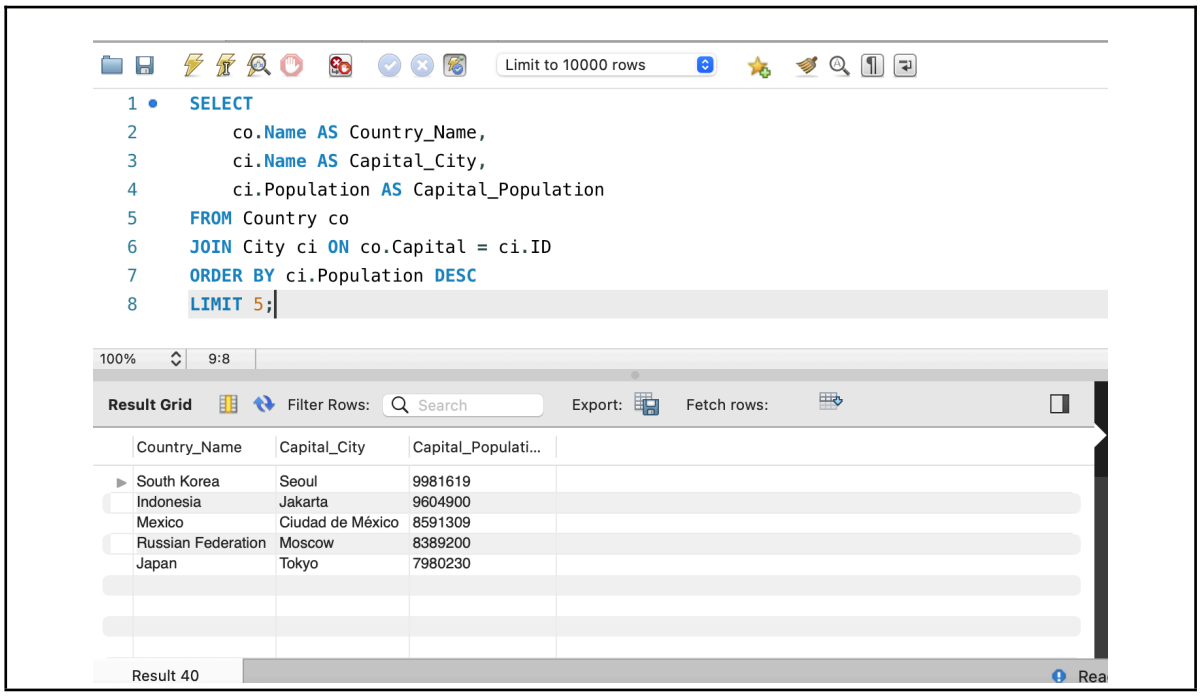


```
SELECT
    co.Name AS Country_Name,
    AVG(ci.Population) AS Average_Population
FROM City ci
JOIN Country co ON ci.CountryCode = co.Code
GROUP BY co.Name
ORDER BY Average_Population DESC
LIMIT 5;
```

Country_Name	Average_Populati...
Singapore	4017733.0000
Hong Kong	1650316.5000
Uruguay	1236000.0000
Guinea	1090610.0000
Uganda	890800.0000

Result 39

17. **Capital Cities Population Comparison:** *Scenario:* A statistical analysis firm is examining population distributions between capital cities worldwide. You're tasked with comparing the populations of capital cities from different countries to identify trends and patterns in urban demographics.



```
SELECT
    co.Name AS Country_Name,
    ci.Name AS Capital_City,
    ci.Population AS Capital_Population
FROM Country co
JOIN City ci ON co.Capital = ci.ID
ORDER BY ci.Population DESC
LIMIT 5;
```

Country_Name	Capital_City	Capital_Populati...
South Korea	Seoul	9981619
Indonesia	Jakarta	9604900
Mexico	Ciudad de México	8591309
Russian Federation	Moscow	8389200
Japan	Tokyo	7980230

Result 40

18. **Countries with Low Population Density:** *Scenario:* An agricultural research institute is studying countries with low population densities for potential agricultural development projects. You're tasked with identifying countries with sparse populations from the database to support the institute's research efforts.

```
SELECT
    Name AS Country_Name,
    Population,
    SurfaceArea,
    (Population / SurfaceArea) AS Population_Density
FROM Country
WHERE (Population / SurfaceArea) < 50
ORDER BY Population_Density ASC
LIMIT 5;
```

Result Grid

Country_Name	Population	SurfaceArea	Population_Dens...
British Indian Ocean Territory	0	78.00	0.0000
Bouvet Island	0	59.00	0.0000
Heard Island and McDonald Islands	0	359.00	0.0000
Antarctica	0	13120000.00	0.0000
French Southern territories	0	7780.00	0.0000

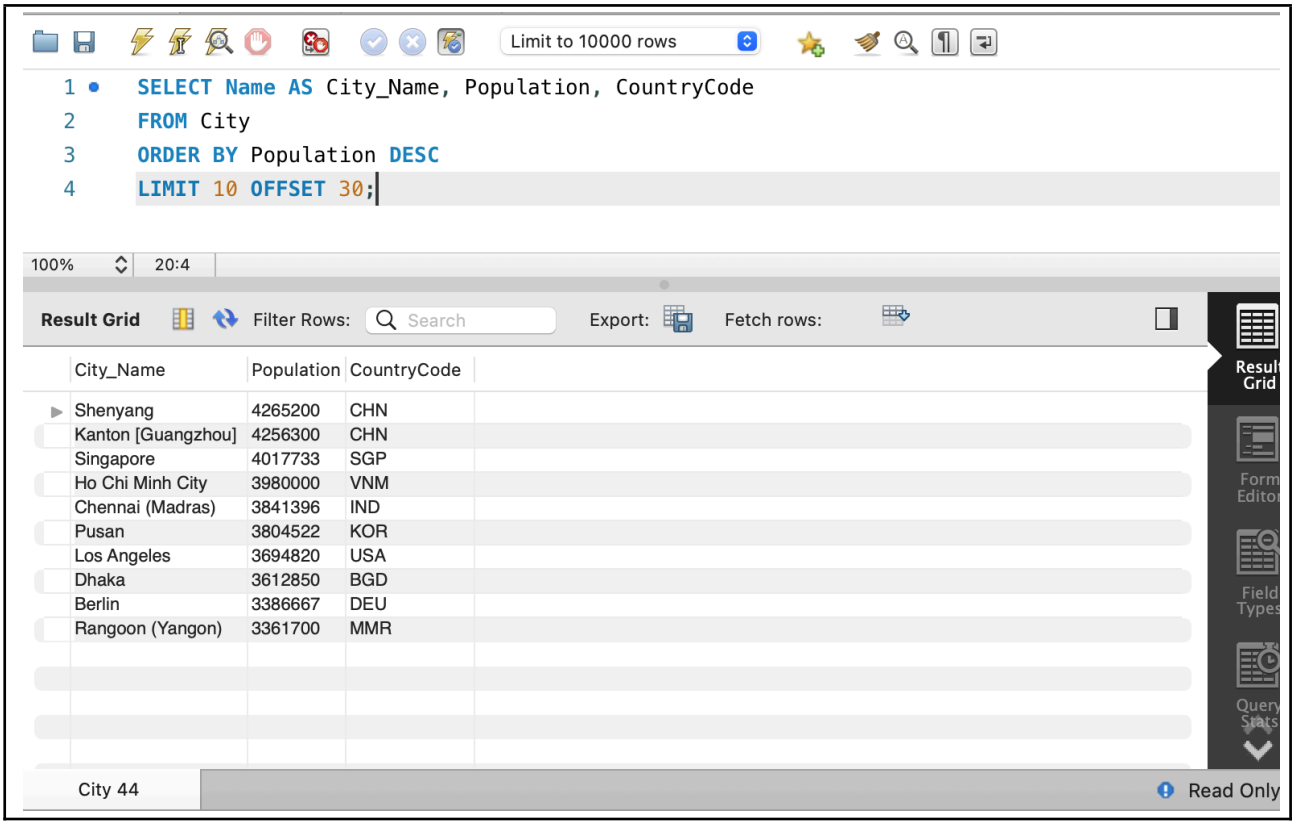
19. **Cities with High GDP per Capita:** *Scenario:* An economic consulting firm is analysing cities with high GDP per capita for investment opportunities. You're tasked with identifying cities with above-average GDP per capita from the database to assist the firm in identifying potential investment destinations.

```
SELECT
    ci.Name AS City_Name,
    co.Name AS Country_Name,
    (co.GNP / ci.Population) AS GDP_Per_Capita
FROM City ci
JOIN Country co ON ci.CountryCode = co.Code
WHERE (co.GNP / ci.Population) > (
    SELECT AVG(co2.GNP / ci2.Population)
    FROM City ci2
    JOIN Country co2 ON ci2.CountryCode = co2.Code
)
ORDER BY GDP_Per_Capita DESC
LIMIT 5;
```

Result Grid

City_Name	Country_Name	GDP_Per_Capita
Charleston	United States	95.558200
Carson	United States	95.530312
Odessa	United States	95.312063
Elgin	United States	95.189469

20. **Display Columns with Limit (Rows 31-40):** *Scenario:* A market research firm requires detailed information on cities beyond the top rankings for a comprehensive analysis. You're tasked with providing data on cities ranked between 31st and 40th by population to ensure a thorough understanding of urban demographics.



Course Notes

It is recommended to take notes from the course, use the space below to do so, or use the revision guide shared with the class:

We have included a range of additional links to further resources and information that you may find useful, these can be found within your revision guide.

END OF WORKBOOK

Please check through your work thoroughly before submitting and update the table of contents if required.

Please send your completed work booklet to your trainer.