Just IT

B2W group

Apprenticeships | Training | Recruitment

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 database table that uniquely identifies each row, ensuring each record is distinct and prevents duplicate data
How does this differ from a secondary key?	A primary key is a special column (or set of columns) that uniquely identifies every row in a table, with only one primary key per table, and it cannot contain NULL values. A secondary key (or alternate key) is any other field or combination of fields in the table that could also uniquely identify rows but wasn't chosen as the primary key
How are primary and foreign keys related?	Primary and foreign keys are related because a foreign key in one table points to a primary key in another table, establishing a link between the two and ensuring data integrity.
Provide a real-world example of a one-to-one relationship	A one-to-one relationship example is a school database, each student has only one student ID, and each student ID is assigned to only one person.
Provide a real-world example of a one-to-many relationship	A one-to-many relationship example is a blog can have many associated posts, but each Post is associated with only one Blog.
Provide a real-world example of a many-to-many relationship	A many-to many relationship example is one between students and classes. A student can register for many classes, and a class can include 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?	Relational databases store data in structured tables with predefined schemas, ensuring strong consistency and integrity (like in SQL databases), while non-relational (NoSQL) databases use more flexible, non-tabular formats like documents or graphs to store varied data types
What type of data would benefit off the non-relational model? **Why?**	Non-relational databases (NoSQL) are best suited for unstructured, semi-structured, or rapidly changing data that doesn't fit neatly into tables and rows. Social media posts, emails, sensor data, logs, documents (JSON, XML), photos, and videos are a few examples of unstructured or semi-structured data. Why: There is no set schema for these data types. Flexible, schema-less documents can be stored in NoSQL databases (such as Couchbase or MongoDB), making it simple to add new fields or modify the data structure without affecting already-existing entries.

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 a type of a JOIN query used to compare rows within the same table. An example of an SQL self-join is a query that retrieves employees and their managers. By joining the Employees table to itself by using their aliases for the employee and managers roles and a left join on the manager column.

Right join	The right join keyword returns all records from the right table (table2), and the matching records from the left table (table1). The result is 0 records from the left side, if there is no match.
Full join	Full join is a join operation that retrieves all records from both tables, including matching and non-matching records. If there are no matching records in one or both tables, NULL values are included for those columns.
Inner join	Inner joins combine records from two tables whenever there are matching values in a field common to both tables. You can use Inner join with the Departments and Employees tables to select all the employees in each department.
Cross join	Cross join is a join operation that produces the Cartesian product of two tables. Unlike other Join operators, it does not let you specify a join clause. You may, however, specify a WHERE clause in the Select statement.
Left join	Left Join is a type of Join operation that retrieves all records from the left table (table1) and the matching records from the right table (table2). If there are no matching records in the right table, NULL values are included for those columns.

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	1a) The types of data the database will need to store would be product/inventory data to track items in stock, item price and knowing when to re-order given the reorder level has been reached **2a)** In terms of structuring the database, I will break this part into three sections relative to the tables The first table will be centred around products sold by the retailer with fields such as Product_id – INT – Unique identifier for each product Name – VARCHAR – Product Name Category – VARCHAR – "Beverages", "Snacks", "Toiletries" etc. Brand – VARCHAR – Something that can help in terms of filtering Unit_Price – DECIMAL – Sales price per unit Stock_Quantity – INT – Current stock count Reorder_Level – INT – Threshold to trigger reorder Supplier_Id – INT – Links to Supplier Table The Second table will be about Customer Information relative to The loyalty program implemented by the retailer Customer_Id – INT – Unique Identifier Name – VARCHAR – Customer name Phone – VARCHAR – Useful for the lookup for loyalty program email – VARCHAR – optional as an alternative form of contact Join_Date – DATE – When loyalty membership started Loyalty_Points – INT – Current points balance The Third table will be a sales/transactions tailored where the fields contain the following Sales_Id – INT – Unique Transaction Identifier Customer_Id – INT – Nullable (supports guest sales) Sale_Date – DATETIME – Timestamp of transaction Total_Amount – DECIMAL – Total transaction value Points_Added – INT – Points accured from purchase

Points_Redeemed – INT – Points used in sale (If any)

The Fourth table will contain Information about Items on sale with the following fields

Sale_Item_Id – INT – Line Item Identifier
Sale_Id – INT - Links to sales table
Product_Id – INT – Product Sold
Quantity – INT – Number of units sold
Unit_Price_At_Sale – DECIMAL – Price at the time of transaction (may change later)
Subtotal – DECIMAL – Quantity x Unit_Price_At_Sale

b) The relationships between these tables disclosed above would be

- Sales to Customers which would be many to one as many sales can be made by one customer

- Sales to Sale Items where the relationship is one to many where one sale has multiple items included in the sale.

- Products and suppliers would be a many to one relationship where many products can be sent out by one supplier.

- Inventory transactions would be a many to one relationship since multiple stock movements can be related to a singular product.

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.

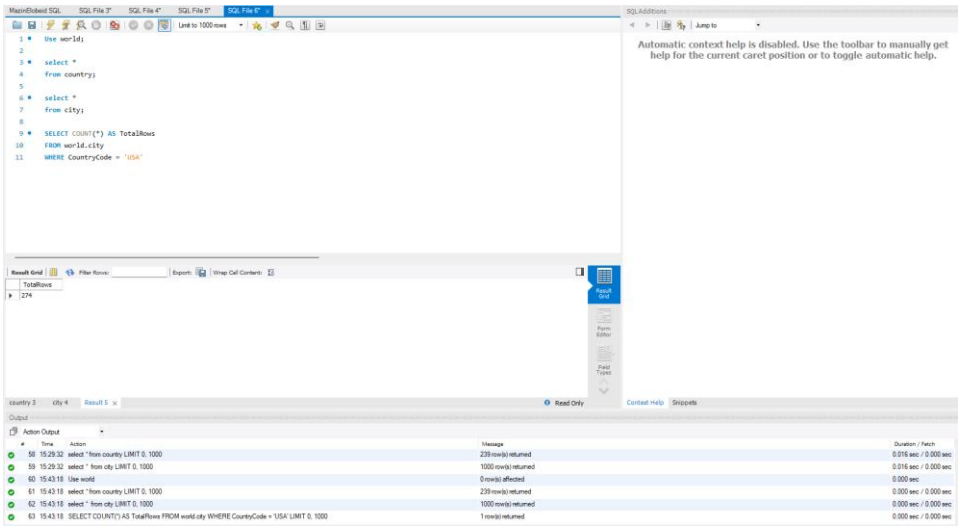
Setting up the database:

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

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

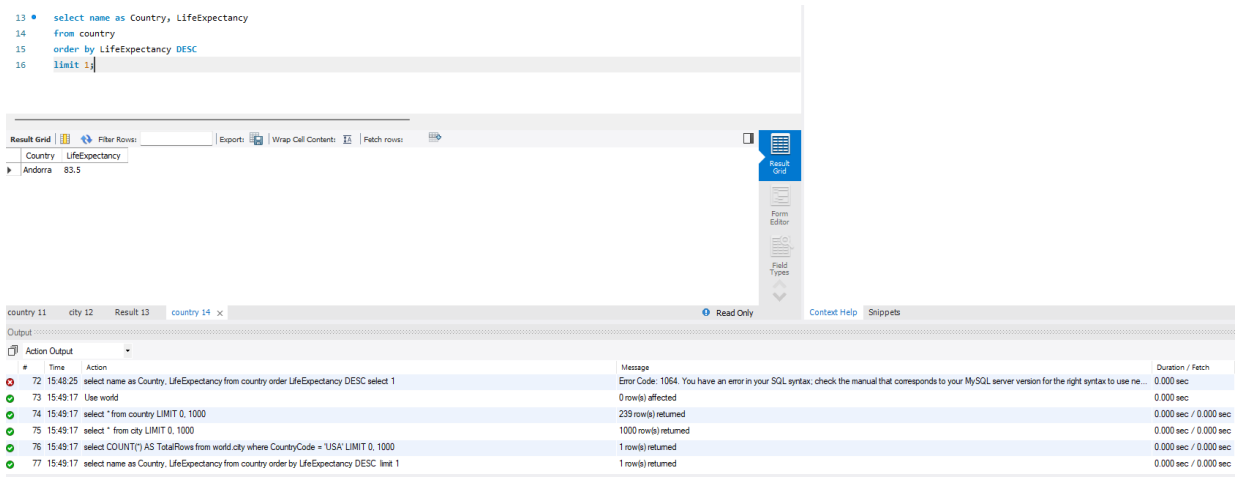
1. **Count Cities in USA:** *Scenario:* You've been tasked with conducting a demographic analysis of cities in the United States. Your first step is to determine the total number of cities within the country to provide a baseline for further analysis.

```
SELECT COUNT(*) AS TotalRows
FROM world.city
WHERE CountryCode = 'USA'
```



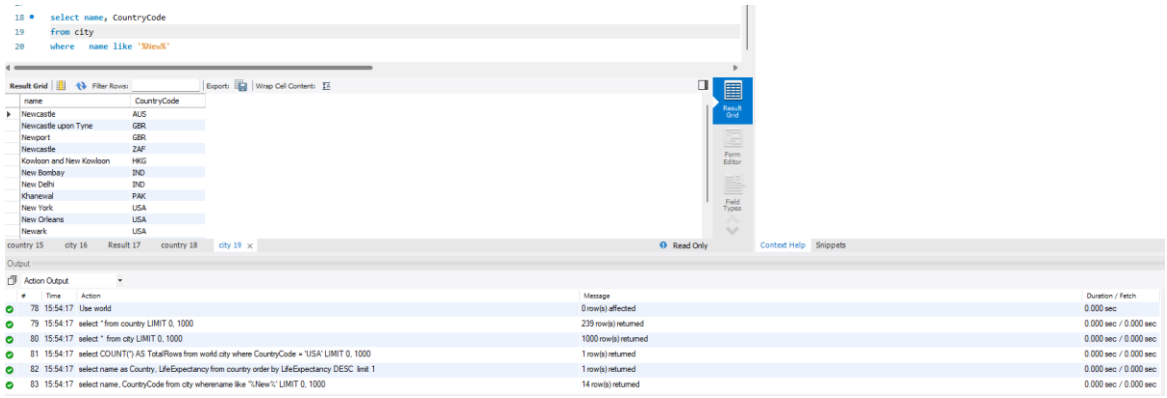
2. **Country with Highest Life Expectancy:** *Scenario:* As part of a global health initiative, you've been assigned to identify the country with the highest life expectancy. This information will be crucial for prioritising healthcare resources and interventions.

```
select name as Country, LifeExpectancy
from country
order by LifeExpectancy DESC
limit 1;
```



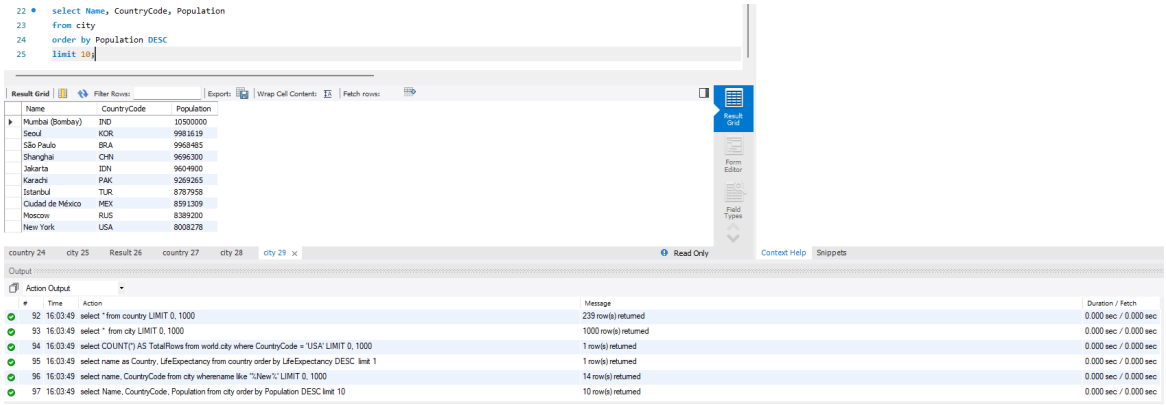
3. **"New Year Promotion: Featuring Cities with 'New :** *Scenario:* In anticipation of the upcoming New Year, your travel agency is gearing up for a special promotion featuring cities with names including the word 'New'. You're tasked with swiftly compiling a list of all cities from around the world.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

```
select name, CountryCode
from city
where  name like '%New%'
```



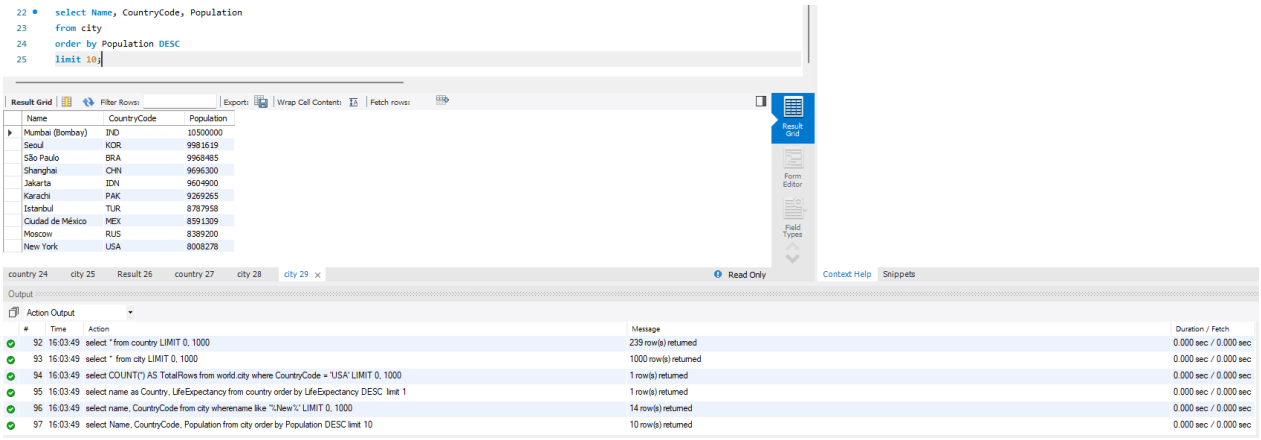
4. **Display Columns with Limit (First 10 Rows):** *Scenario:* You're tasked with providing a brief overview of the most populous cities in the world. To keep the report concise, you're instructed to list only the first 10 cities by population from the database.

```
select Name, CountryCode, Population
from city
order by Population DESC
limit 10;
```



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.

select name, CountryCode, Population
from city
where Population > 2000000
order by Population desc;



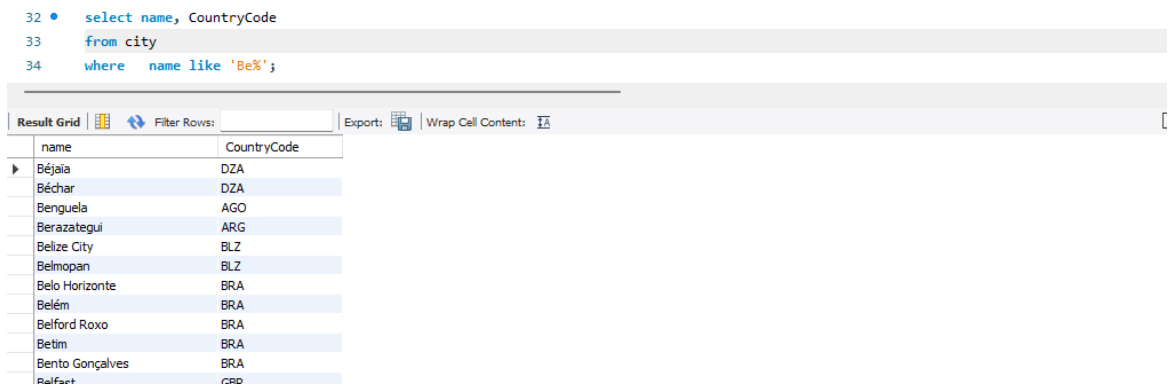
Result Grid | Filter Rows: | Export: | Wrap Cell Content:

name	CountryCode	Population
Mumbai (Bombay)	IND	10500000
Seoul	KOR	9981619
São Paulo	BRA	9968485
Shanghai	CHN	9696300
Jakarta	IDN	9604900
Karachi	PAK	9269265
Istanbul	TUR	8787958
Ciudad de México	MEX	8591309
Moscow	RUS	8389200
New York	USA	8008278
Tokyo	JPN	7980230
Peking	CHN	7472000
London	GBR	7285000
Delhi	IND	7206704
Cairo	EGY	6789479
Teheran	IRN	6758845
Lima	PER	6464693
Chongqing	CHN	6351600
Bangkok	THA	6320174
Santafé de Bogotá	COL	6260862
Rio de Janeiro	BRA	5598953
Tianjin	CHN	5286800
Kinshasa	COD	5064000
Lahore	PAK	5063499
Santiago de Chile	CHL	4703954
St Petersburg	RUS	4694000
Calcutta [Kolkata]	IND	4399819
Wuhan	CHN	4344600

country 30 | city 31 | Result 32 | country 33 | city 34 | city 35 | city 36

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.

```
select name, CountryCode
from city
where name like 'Be%';
```



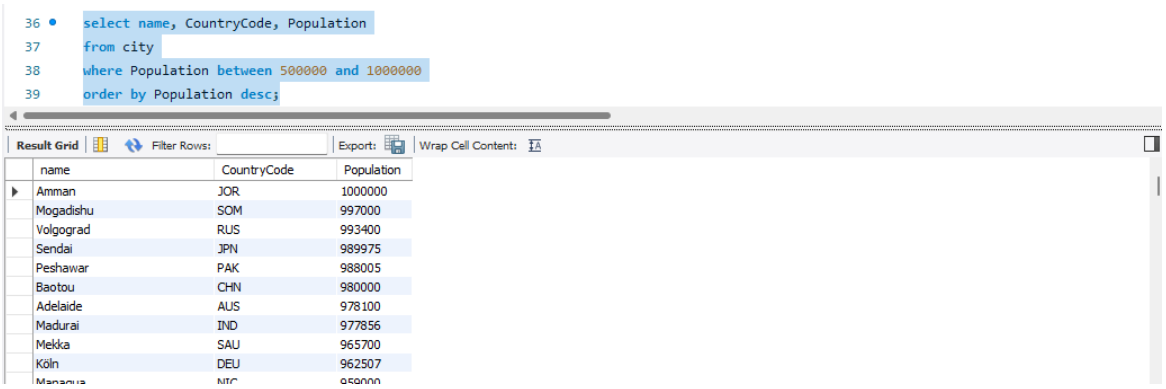
```
32 •  select name, CountryCode
33    from city
34    where   name like 'Be%';
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

name	CountryCode
Béjaïa	DZA
Béchar	DZA
Benguela	AGO
Berazategui	ARG
Belize City	BLZ
Belmopan	BLZ
Belo Horizonte	BRA
Belém	BRA
Belford Roxo	BRA
Betim	BRA
Bento Gonçalves	BRA
Belfast	GBR

7. **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.

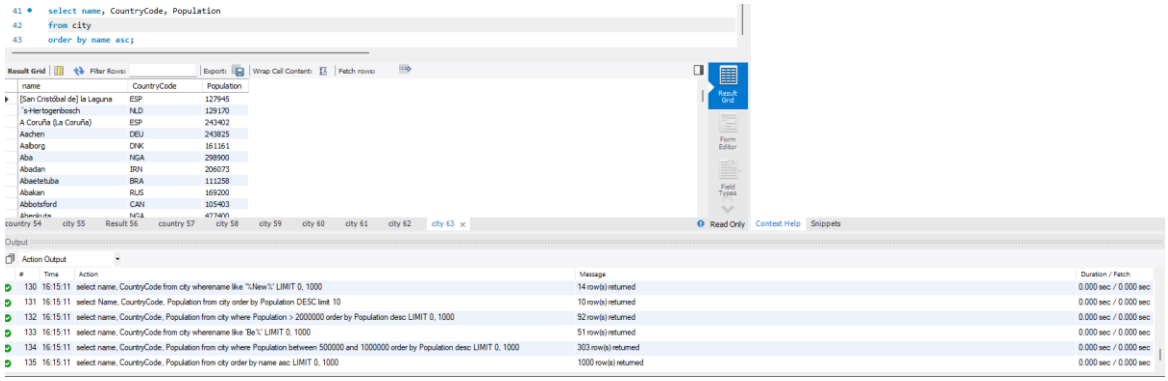
```
select name, CountryCode, Population
from city
where Population between 500000 and 1000000
order by Population desc;
```



name	CountryCode	Population
Amman	JOR	1000000
Mogadishu	SOM	997000
Volgograd	RUS	993400
Sendai	JPN	989975
Peshawar	PAK	988005
Baotou	CHN	980000
Adelaide	AUS	978100
Madurai	IND	977856
Mekka	SAU	965700
Köln	DEU	962507

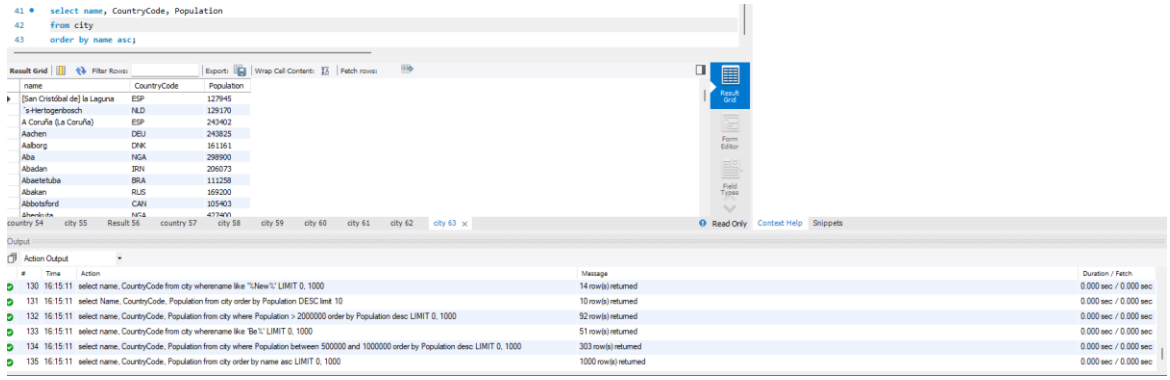
8. **Display Cities Sorted by Name in Ascending Order:** *Scenario:* A geography teacher is preparing a lesson on alphabetical order using city names. You're tasked with providing a sorted list of cities from the database in ascending order by name to support the lesson plan.

```
select name, CountryCode, Population
from city
order by name asc;
```



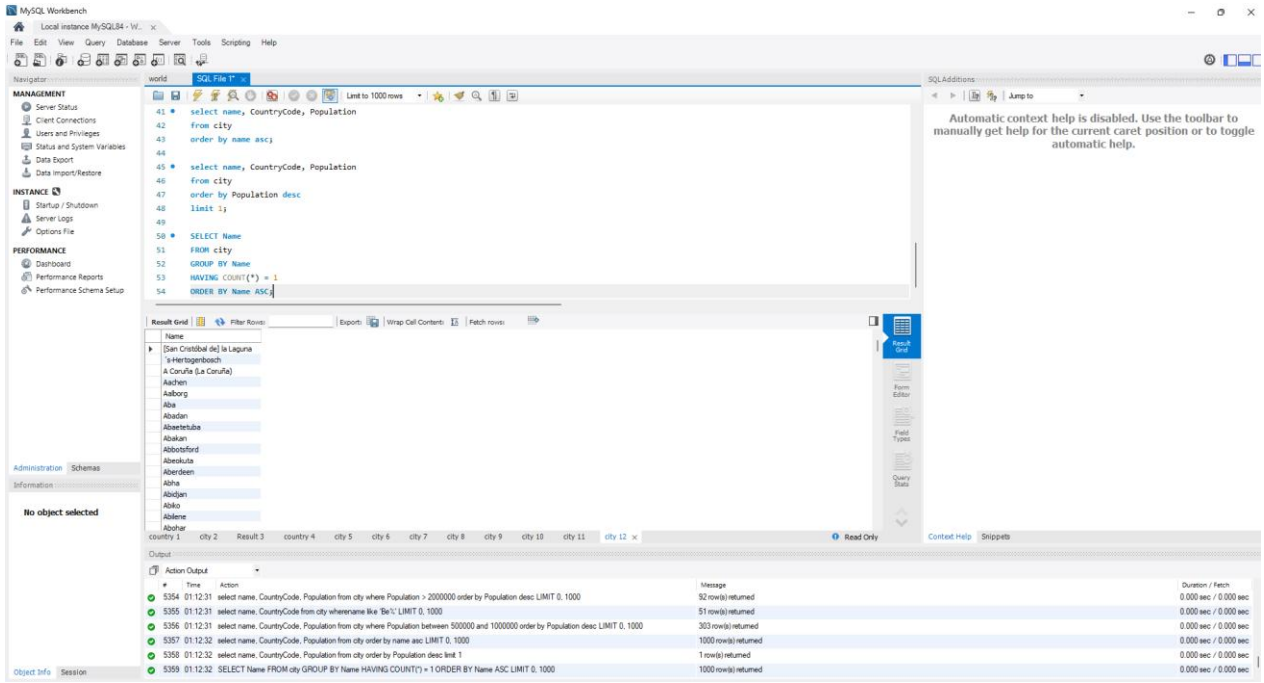
9. **Most Populated City:** *Scenario:* A real estate investment firm is interested in cities with significant population densities for potential development projects. You're tasked with identifying the most populated city from the database to guide their investment decisions and strategic planning.

```
select name, CountryCode, Population
from city
order by Population desc
limit 1;
```



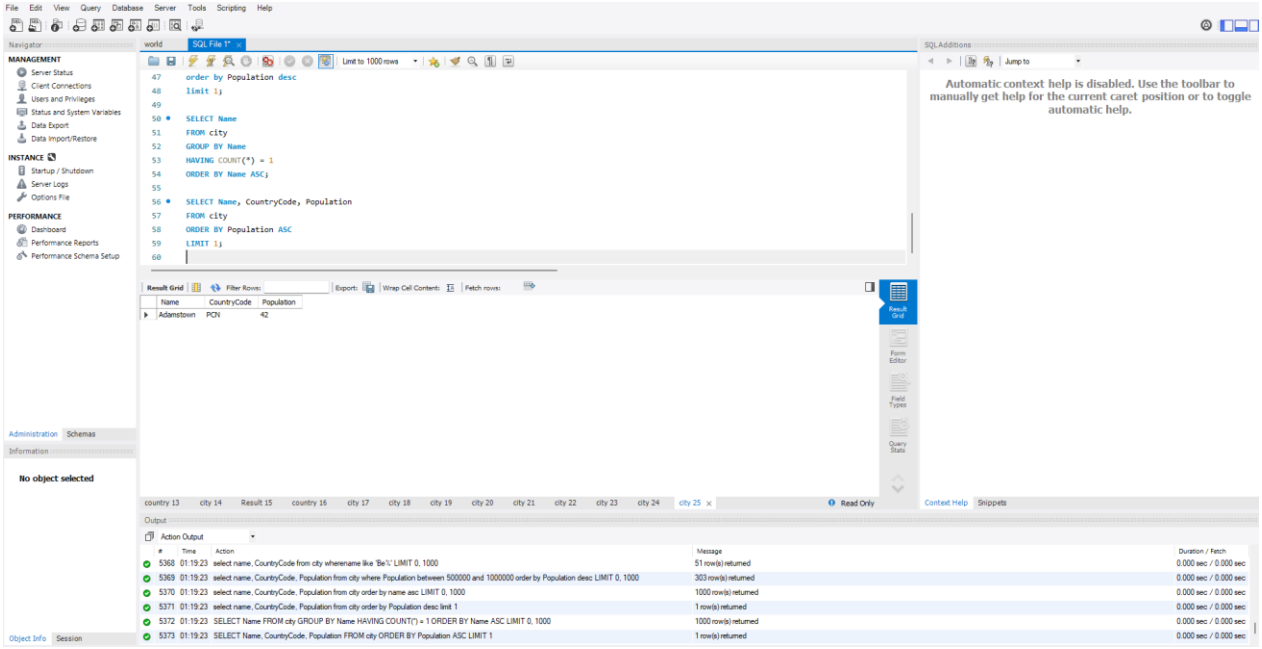
10. **City Name Frequency Analysis: Supporting Geography Education** *Scenario*: In a geography class, students are learning about the distribution of city names around the world. The teacher, in preparation for a lesson on city name frequencies, wants to provide students with a list of unique city names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

```
SELECT Name
FROM city
GROUP BY Name
HAVING COUNT(*) = 1
ORDER BY Name ASC;
```



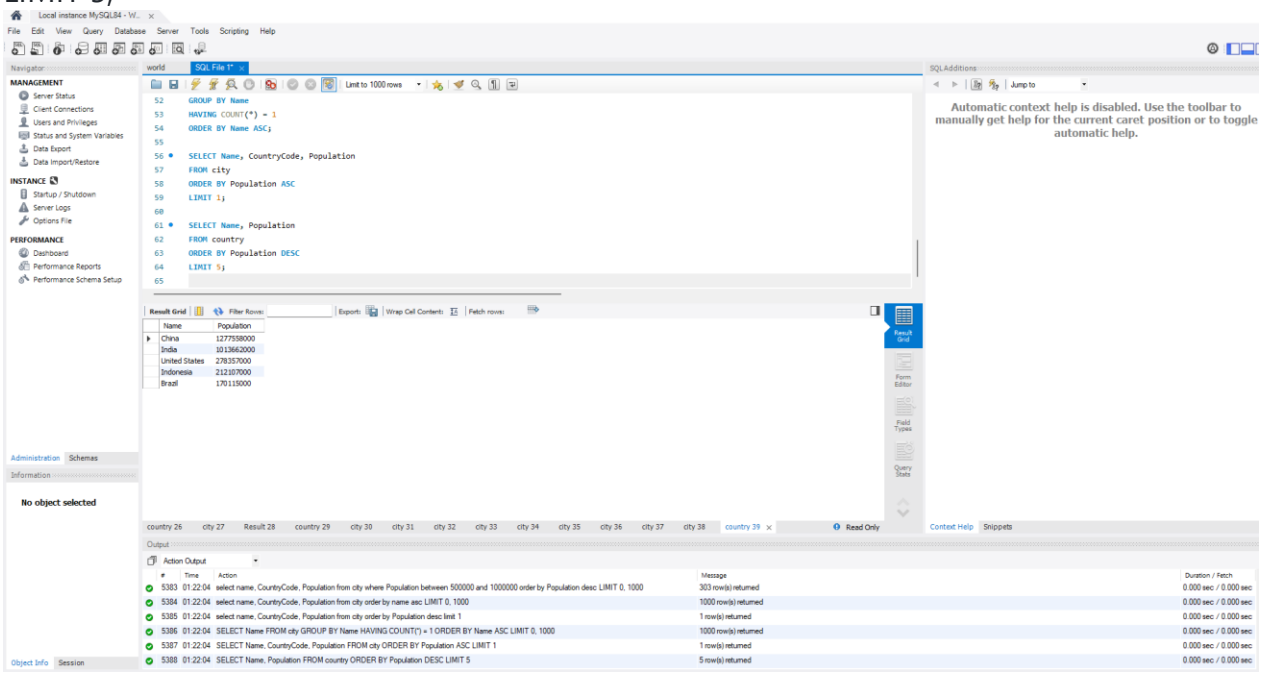
11. **City with the Lowest Population:** *Scenario:* A census bureau is conducting an analysis of urban population distribution. You're tasked with identifying the city with the lowest population from the database to provide a comprehensive overview of demographic trends.

```
SELECT Name, CountryCode, Population
FROM city
ORDER BY Population ASC
LIMIT 1;
```



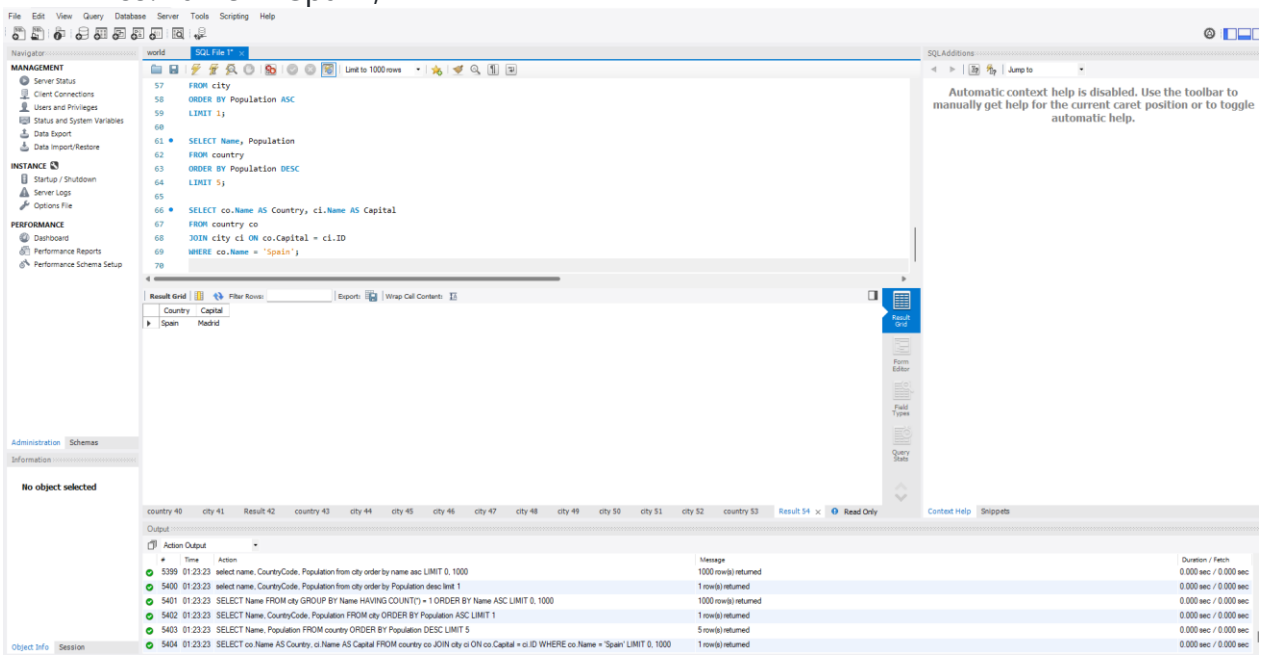
12. **Country with Largest Population:** *Scenario:* A global economic research institute requires data on countries with the largest populations for a comprehensive analysis. You're tasked with identifying the country with the highest population from the database to provide valuable insights into demographic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

```
SELECT Name, Population
FROM country
ORDER BY Population DESC
LIMIT 5;
```



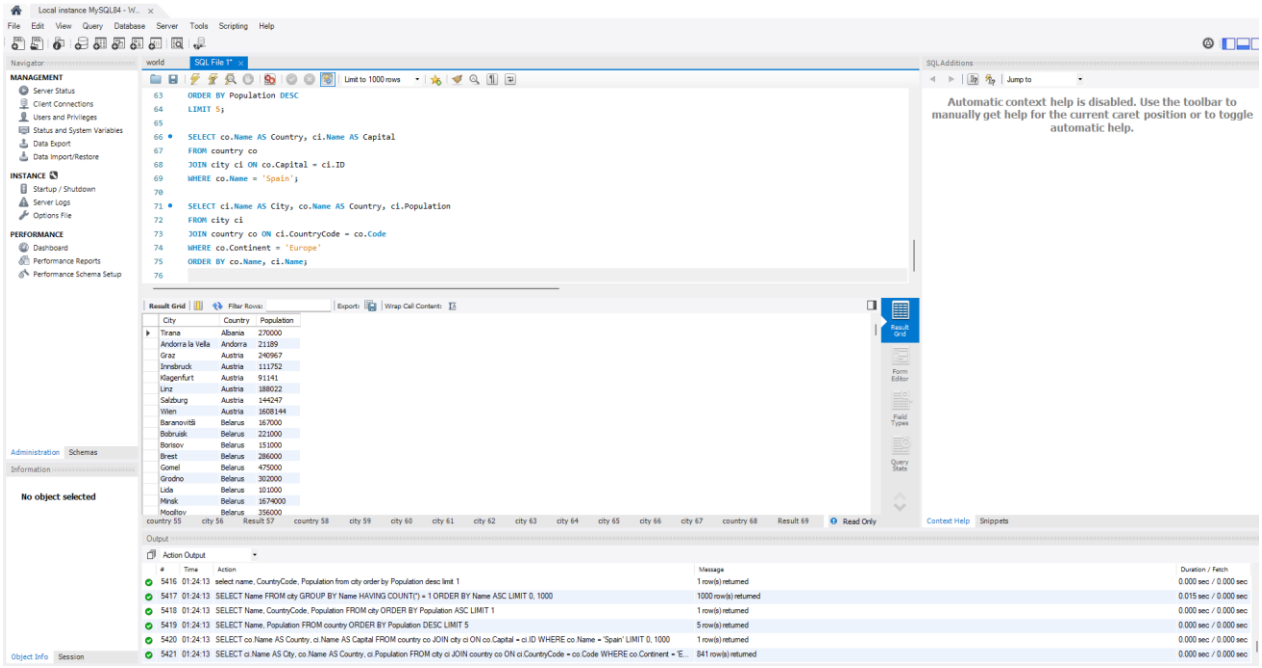
13. **Capital of Spain:** *Scenario:* A travel agency is organising tours across Europe and needs accurate information on capital cities. You're tasked with identifying the capital of Spain from the database to ensure itinerary accuracy and provide travellers with essential destination information.

```
SELECT co.Name AS Country, ci.Name AS Capital
FROM country co
JOIN city ci ON co.Capital = ci.ID
WHERE co.Name = 'Spain';
```



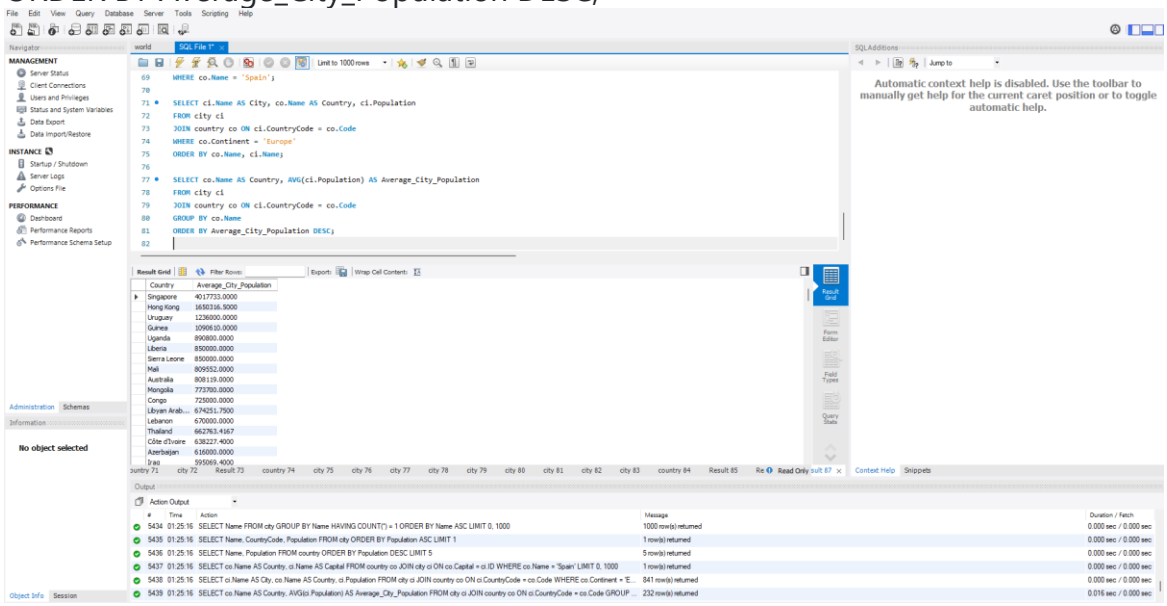
14. **Cities in Europe:** *Scenario:* A European cultural exchange program is seeking to connect students with cities across the continent. You're tasked with compiling a list of cities located in Europe from the database to facilitate program planning and student engagement.

```
SELECT ci.Name AS City, co.Name AS Country, ci.Population
FROM city ci
JOIN country co ON ci.CountryCode = co.Code
WHERE co.Continent = 'Europe'
ORDER BY co.Name, ci.Name;
```



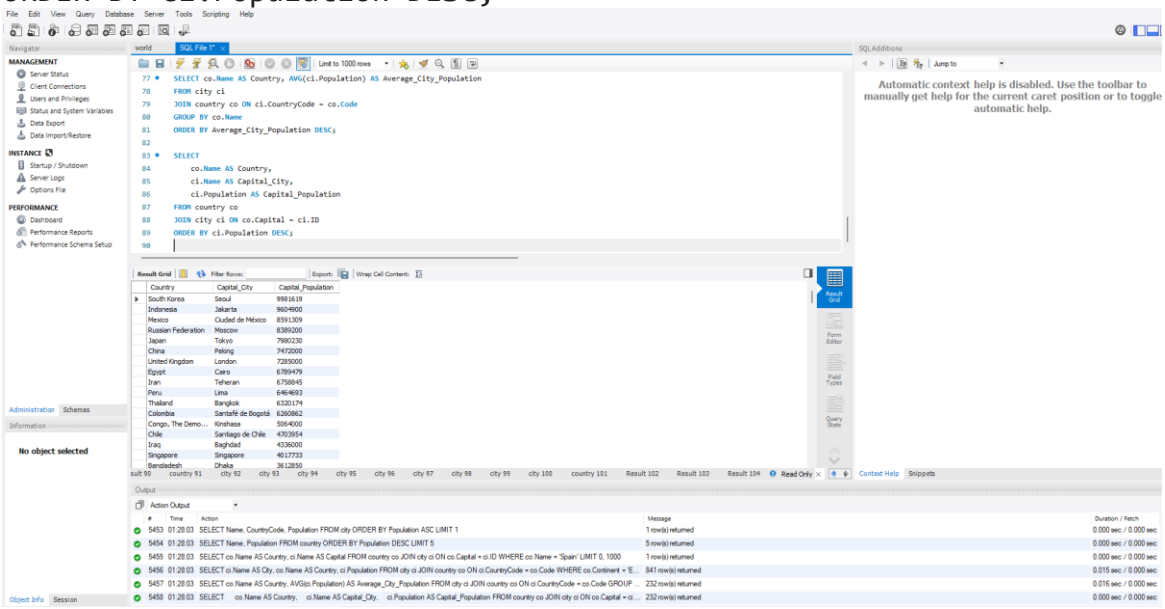
15. **Average Population by Country:** *Scenario:* A demographic research team is conducting a comparative analysis of population distributions across countries. You're tasked with calculating the average population for each country from the database to provide valuable insights into global population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

```
SELECT co.Name AS Country, AVG(ci.Population) AS Average_City_Population
FROM city ci
JOIN country co ON ci.CountryCode = co.Code
GROUP BY co.Name
ORDER BY Average_City_Population DESC;
```



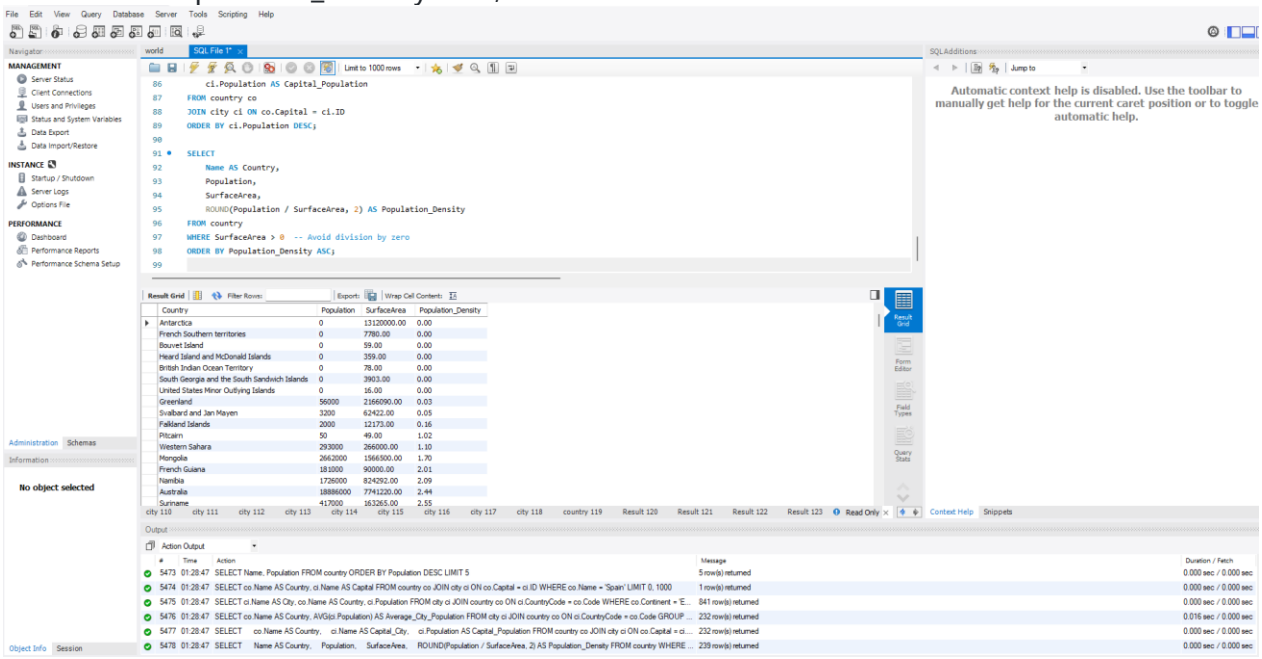
16. **Capital Cities Population Comparison:** *Scenario:* A statistical analysis firm is examining population distributions between capital cities worldwide. You're tasked with comparing the populations of capital cities from different countries to identify trends and patterns in urban demographics.

```
SELECT
    co.Name AS Country,
    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;
```



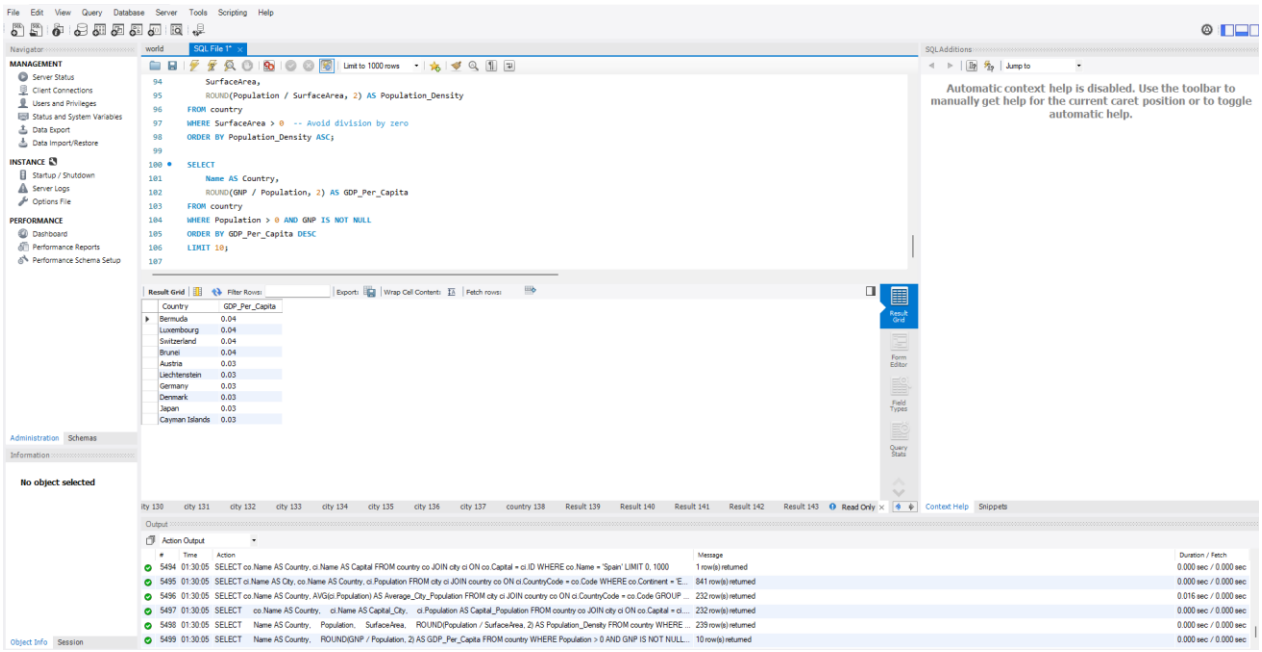
17. **Countries with Low Population Density:** *Scenario:* An agricultural research institute is studying countries with low population densities for potential agricultural development projects. You're tasked with identifying countries with sparse populations from the database to support the institute's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

```
SELECT
    Name AS Country,
    Population,
    SurfaceArea,
    ROUND(Population / SurfaceArea, 2) AS Population_Density
FROM country
WHERE SurfaceArea > 0  -- Avoid division by zero
ORDER BY Population_Density ASC;
```



18. **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
    Name AS Country,
    ROUND(GNP / Population, 2) AS GDP_Per_Capita
FROM country
WHERE Population > 0 AND GNP IS NOT NULL
ORDER BY GDP_Per_Capita DESC
LIMIT 10;
```



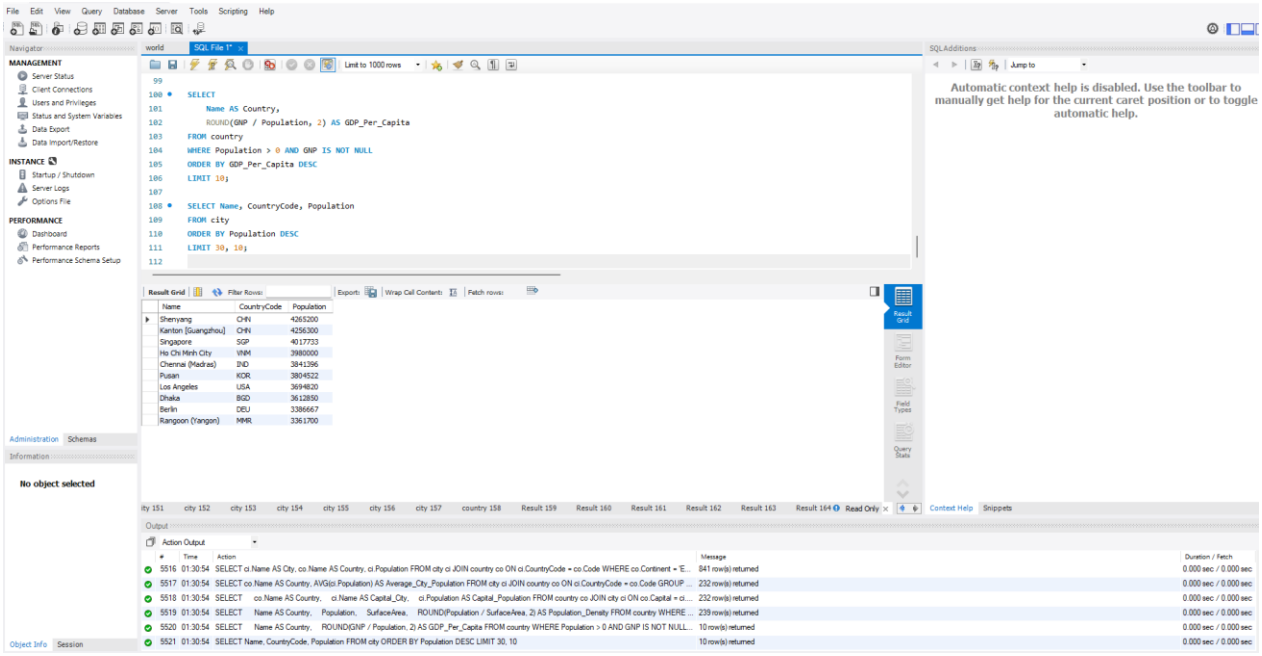
19. **Display Columns with Limit (Rows 31-40):** *Scenario:* A market research firm requires detailed information on cities beyond the top rankings for a comprehensive analysis. You're tasked with providing data on cities ranked between 31st and 40th by population to ensure a thorough understanding of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

```
SELECT Name, CountryCode, Population

FROM city

ORDER BY Population DESC

LIMIT 30, 10;
```



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.