Just IT

B2W group

Apprenticeships | Training | Recruitment

Data Technician

Name: Almas Mansuri

Course Date: 12/05/2025

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 field (or a set of fields) that uniquely identifies each record in a table.
How does this differ from a secondary key?	A secondary key is alternative way to retrieve or sort data.
How are primary and foreign keys related?	Primary key and foreign key connect two tables by establishing the relationship between them.
Provide a real-world example of a one-to-one relationship	A Person has one passport
Provide a real-world example of a one-to-many relationship	One student can enrol in to many classes
Provide a real-world example of a many-to-many relationship	An employee can work on many projects and a project can have multiple employees.

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, organized tables with rows and columns using SQL to manage relationship between data. Non-relational database which are also known as NoSQL, store data as objects (key-Value pairs), as documents, graphs.
What type of data would benefit off the non-relational model? **Why?**	Unstructured or semi structures data • No SQL databases don't enforce a strict schema that's why the data can be stored in Various formats • They can adapt changing data needs and structures. Big data Application • No SQL databases are designed to handle large volumes of data and are often used for data analytics. Scalability and Performance • Non-relational databases are optimized for fast read and write operations, making them suitable for real-time applications and big data analytics where speed and efficiency are important.

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 regular join. In this join the table is joined with itself.

Right join	Right-join returns all the data from the right table and matching data from the left table.
	`Select Orders.item, Orders.amount,Shippings.status`
	`from Orders`
	`Right Join Shippings`
	`on Orders.order_id=Shippings.shipping.id;`

Full join	The full join returns common data between two tables and all the data from left and right table.

```
Select Customers.first_name,Customers.last_name,Customers_country,Orders.item
Orders.amount
from Customers
Full Join Orders
on Customers.customer_id=Orders.customer_id;
```

Inner join

Inner-join returns data that have matching values in both tables.

```
Select first_name,last_name,item,amount
from Customers
Inner Join Orders on Customers.customer_id= Orders.customer_id;
```

Output

first_name	last_name	item	amount
John	Reinhardt	Keyboard	400
John	Reinhardt	Mouse	300
David	Robinson	Monitor	12000
John	Doe	Keyboard	400
Robert	Luna	Mousepad	250

Cross join

The cross join returns all the data from both tables.

```
Select *
from Customers C
cross join Orders O
on C.customer_id = O.customer_id;
```

Output

stomer_id	first_name	last_name	age	country	order_id	item	amount
	John	Doe	31	USA	4	Keyboard	400
	Robert	Luna	22	USA	5	Mousepad	250
	David	Robinson	22	UK	3	Monitor	12000
	John	Reinhardt	25	UK	1	Keyboard	400
	John	Reinhardt	25	UK	2	Mouse	300

Left join

Left-join returns all the data from the left table and matching data from the right table.

```
Select *
from Customers C
left join Orders O
on C.customer_id = O.customer_id;
```

Output

John	25
Betty	28

customer_id	first_name	last_name	age	country	order_id	item
1	John	Doe	31	USA	4	Keybo
2	Robert	Luna	22	USA	5	Mouse
3	David	Robinson	22	UK	3	Monito
4	John	Reinhardt	25	UK	1	Keybo

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 a. To help streamline operations for small retail business I will create simple but effective database. This structure will useful to track inventory, customer information and overall sales performance.

The database will have:

1. Products Data-This section will have Product id, Names, categories (groceries, other domestic products), price, stock (product available or not).
2. Customers- This section will have personal detail of customers such as Names, contact information.
3. Sales-This section will have information of product sold, transaction, date & time, customer.
4. Sales Items- This section will have product Id, quantity.
5. Supplier- This section will have supplier information such as supplier name, supplier Id, contact.

1 b. The database will be access by:

1. Store owner/Manager
2. Sales Team
3. Marketing Team

2 a. I will create the database in SQL with clearly defined normalized tables as below

1. Products –Product table stores all the product related information
    a. Product Id (integer)
    b. Product Name (varchar)
    c. Product Category (varchar)
    d. Price (decimal)
    e. Stock (varchar)
2. Customers- Customer table contains customer records
    a. Customer Id (integer)
    b. Customer Name (Varchar)
    c. Customer Phone (integer)
    d. Customer Email (varchar)
3. Sales – Sales table contains transaction related information
    a. Sales Id (integer)
    b. Customer Id (integer)
    c. Date (dateTime)
    d. Amount
4. Sales Items Table contains information about product sold
    a. Sales Items Id(varchar)
    b. Sales Id (integer)
    c. Product Id (integer)
    d. Quantity (integer)
5. Supplier-This table contain information about supplier information
    a. Supplier Id (integer)
    b. Supplier Name (varchar)
    c. Supplier contact (varchar)

2 b. Primary key and Foreign Keys are the first necessary things to connect the tables to each other.

The following relationships between tables are necessary:

1. Products to Customers: one to Many
   One product is purchased by many customers
2. Sales to Customers: Many to one
   Each sale is linked to one customer but a customer can have multiple sales
3. Sales to sales Item: One to many
   Each sale can include multiple items but each sale item belongs to one sale.

3a. Below SQL commands will be used to create database

Create Database Corner Shop;(This command creates database in SQL workbench)

Use Database Corner Shop; (This command use the database in SQL. Before running any queries make sure to use the database)

3a. Below command will be used to create table in particular Database

```
Create Table Products (
Product Id Int Primary Key Auto_Increment,
Product Name varchar(100) Not Null,
Product Category varchar(150),
Price decimal (10,2) No Null,
Stock varchar(50) Not Null
    );
```

4a. Below command will be used to insert data in to product table

```
Insert into Products
(product Id, Product Name, Product Category, Price, Stock)
Values
(1, 'Milk', 'Dairy',2.15,'In stock'),
(2, 'Chicken','Poultry',4,''In stock');
```

5a. I will take following measures to make sure database remains accurate and up to date:

1. Use Constraints and Relationships:

- Use of constraints like Not null, Unique where necessary
- Use of Primary keys, foreign keys to maintain data integrity

2. Data Backups:

- Do regular backups of the database and allowing recovery if any mistake occurs.

3. Access Control:

- Limited access to sensitive information.
- Reduce any human errors and unauthorized changes

5b. Below steps I will take to handle backups:

- Schedule daily automatic backups of the database (full backup).
- Store backups both locally (e.g., external hard drive) and in the cloud (e.g., Google Drive, Dropbox)

Day 4: Task 2: SQL Practical

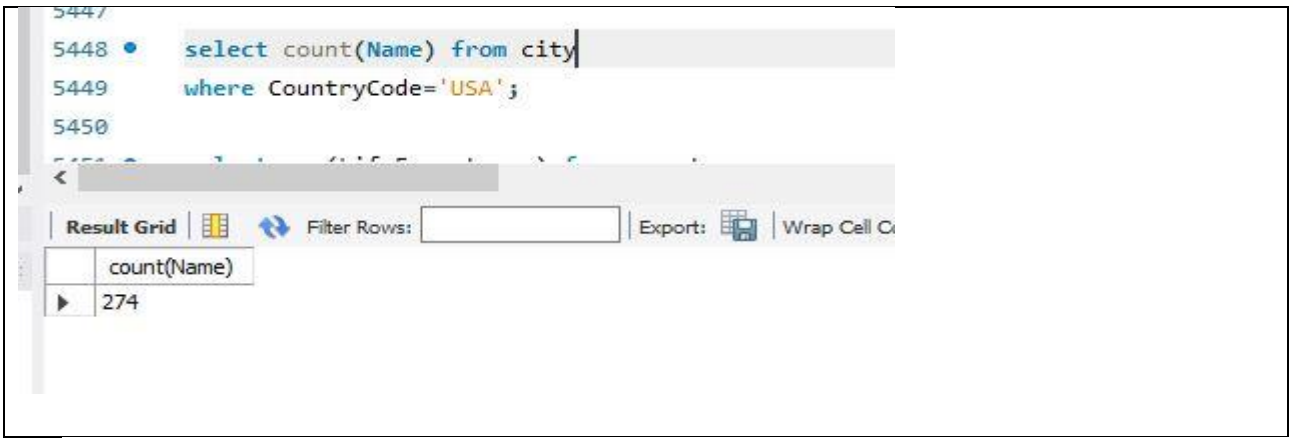
In your groups, work together to answer the below questions. It may be of benefit if one of you shares your screen with the group and as a team answer / take screen shots from there.

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.

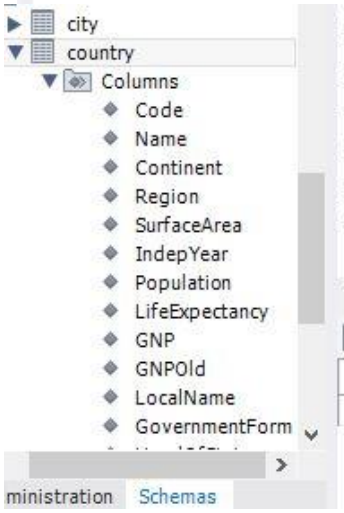


2. **Country with Highest Life Expectancy:** *Scenario:* As part of a global health initiative, you've been assigned to identify the country with the highest life expectancy. This information will be crucial for prioritising healthcare resources and interventions.

```
5453    SELECT Name, LifeExpectancy
5454    FROM country
5455    WHERE LifeExpectancy = (
5456        SELECT MAX(LifeExpectancy)
5457        FROM country
5458    );
5459
5460
```

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

Name	LifeExpectancy
Andorra	83.5




```
5444    select Name, Continent,Population from country;
5445    select count(Name) from country;
5446    select distinct count(Name) from country;
5447
5448    select max(LifeExpectancy) from country;
5449
```

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

max(LifeExpectancy)
83.5

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.

```
5448 •  select max(LifeExpectancy) from country;
5449 •  select * from city
5450    where Name like '%New%';
5451
```

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content:

ID	Name	CountryCode	District	Population
137	Newcastle	AUS	New South Wales	270324
482	Newcastle upon Tyne	GBR	England	189150
502	Newport	GBR	Wales	139000
734	Newcastle	ZAF	KwaZulu-Natal	222993
936	Kowloon and New Kowloon	HKG	Kowloon and New Kowl	1987996
1106	New Bombay	IND	Maharashtra	307297
1109	New Delhi	IND	Delhi	301297
2857	Khanewal	PAK	Punjab	133000
3793	New York	USA	New York	8008278

city 9 | Apply

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.

```
5452 •  select Name, Population from city
5453    order by Population DESC
5454    limit 10;
5455
5456
```

Result Grid | Filter Rows: | Export:

Name	Population
Mumbai (Bombay)	10500000
Seoul	9981619
São Paulo	9968485
Shanghai	9696300
Jakarta	9604900
Karachi	9269265
Istanbul	8787958
Ciudad de México	8591309
Moscow	8389200
New York	8008278

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.

```
5477 •  select Name,Population from city
5478    where Population > 2000000;
```

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

Name	Population
Alger	2168000
Luanda	2022000
Buenos Aires	2982146
Sydney	3276207
Melbourne	2865329
Dhaka	3612850
São Paulo	9968485
Rio de Janeiro	5598953
Salvador	2302832
Belo Horizonte	2139125
Fortaleza	2097757
London	7285000

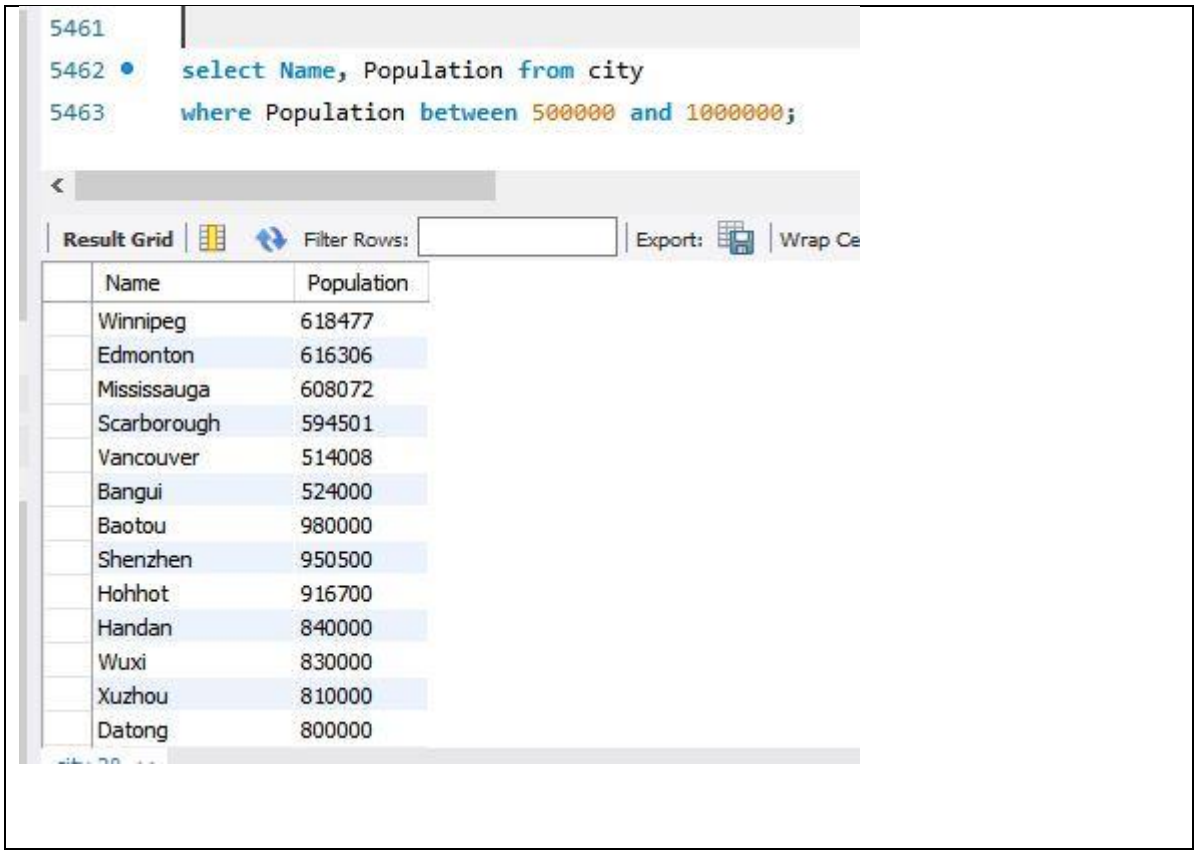
6. **Cities Beginning with 'Be' Prefix:** *Scenario:* A travel blogger is planning a series of articles featuring cities with unique names. You're tasked with compiling a list of cities from the database that start with the prefix 'Be' to assist in the blogger's content creation process.

```
5459 •  select distinct Name from city
5460    where Name like 'Be%';
5461
```

Result Grid | Filter Rows: | Export:

Name
Béjaïa
Béchar
Benguela
Berazategui
Belize City
Belmopan
Belo Horizonte
Belém
Belford Roxo
Betim
Bento Gonçal...
Belfast
Benoni

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.

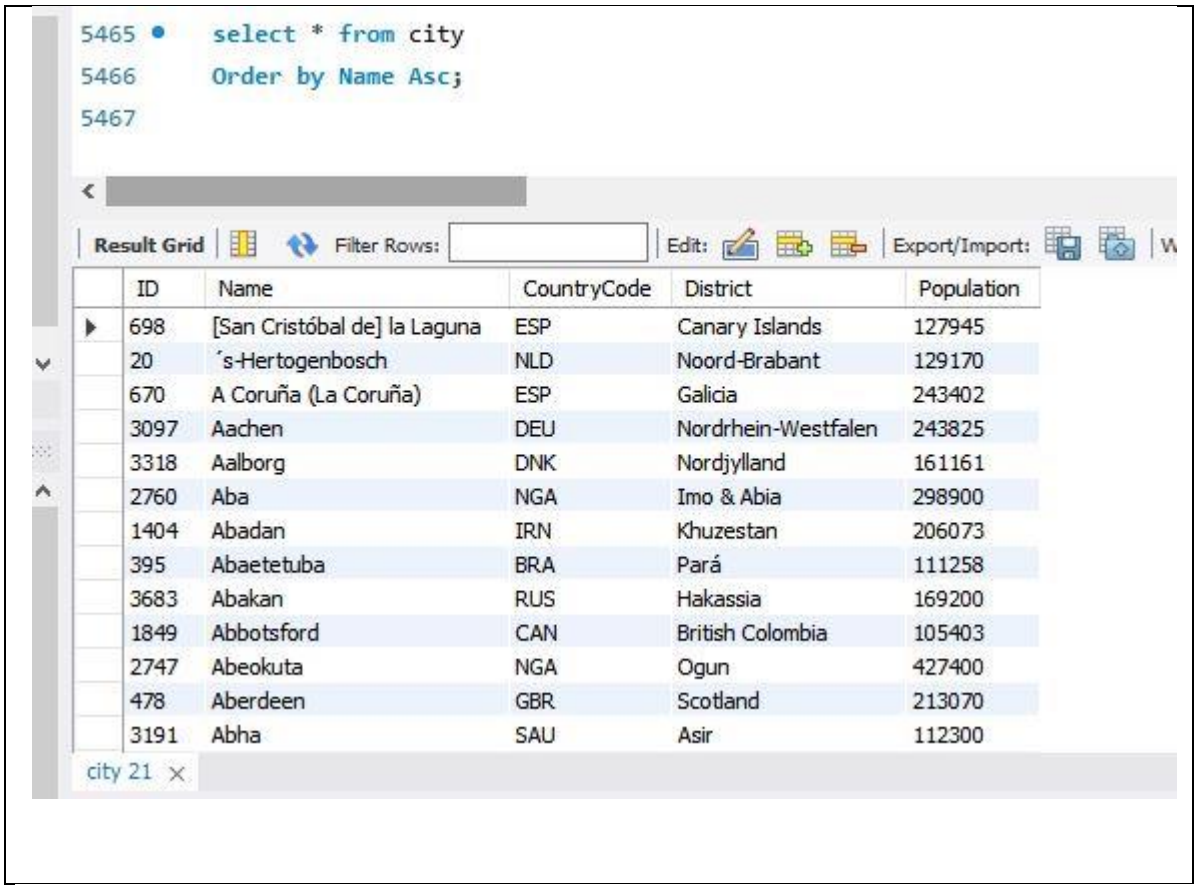


```
5461
5462 •  select Name, Population from city
5463    where Population between 500000 and 1000000;
```

Result Grid | Filter Rows: | Export: | Wrap Ce

Name	Population
Winnipeg	618477
Edmonton	616306
Mississauga	608072
Scarborough	594501
Vancouver	514008
Bangui	524000
Baotou	980000
Shenzhen	950500
Hohhot	916700
Handan	840000
Wuxi	830000
Xuzhou	810000
Datong	800000

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.



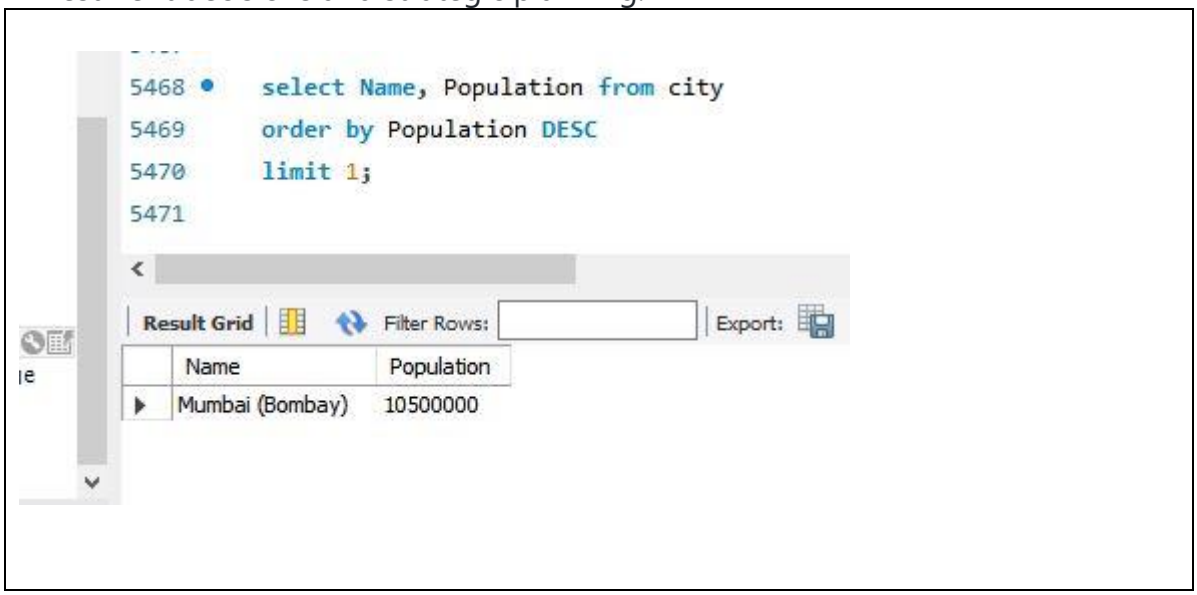
```
5465 •  select * from city
5466    Order by Name Asc;
5467
```

Result Grid | Filter Rows: | Edit: | Export/Import:

ID	Name	CountryCode	District	Population
698	[San Cristóbal de] la Laguna	ESP	Canary Islands	127945
20	´s-Hertogenbosch	NLD	Noord-Brabant	129170
670	A Coruña (La Coruña)	ESP	Galicia	243402
3097	Aachen	DEU	Nordrhein-Westfalen	243825
3318	Aalborg	DNK	Nordjylland	161161
2760	Aba	NGA	Imo & Abia	298900
1404	Abadan	IRN	Khuzestan	206073
395	Abaetetuba	BRA	Pará	111258
3683	Abakan	RUS	Hakassia	169200
1849	Abbotsford	CAN	British Colombia	105403
2747	Abeokuta	NGA	Ogun	427400
478	Aberdeen	GBR	Scotland	213070
3191	Abha	SAU	Asir	112300

city 21

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.

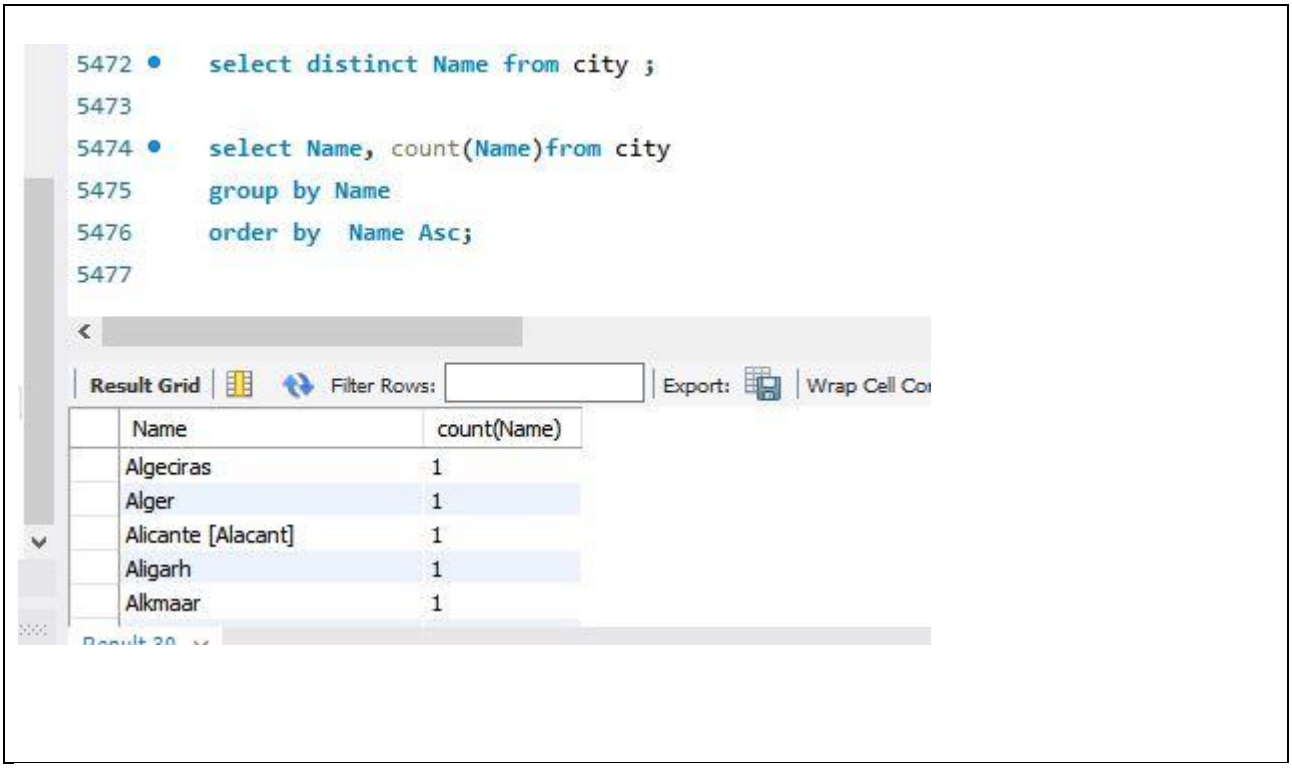


```
5468 •  select Name, Population from city
5469    order by Population DESC
5470    limit 1;
5471
```

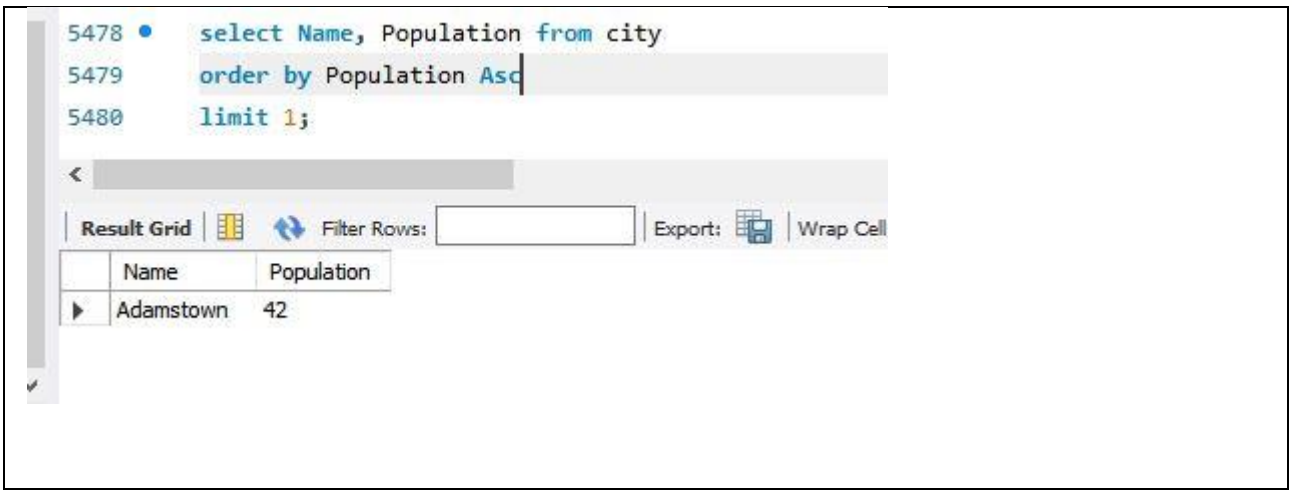
Result Grid | Filter Rows: | Export:

Name	Population
Mumbai (Bombay)	10500000

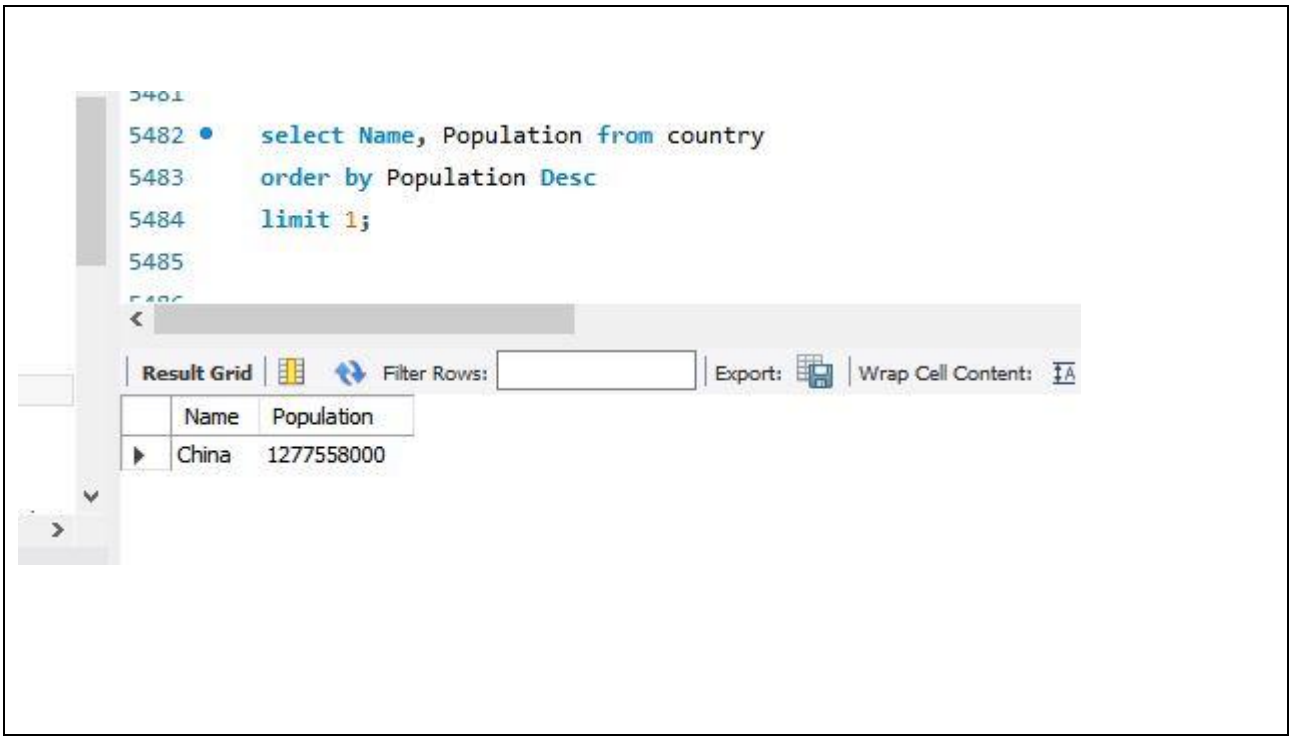
10. **City Name Frequency Analysis: Supporting Geography Education** *Scenario*: In a geography class, students are learning about the distribution of city names around the world. The teacher, in preparation for a lesson on city name frequencies, wants to provide students with a list of unique city names 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. **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.

```
5499 •  select city.Name As cityName, country.Name As countryName, country.Continent
5500    from country
5501    inner join city
5502    on country.capital = city.ID
5503    where Continent='Europe'
5504    ;
```

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

cityName	countryName	Continent
Sarajevo	Bosnia and Herzegovina	Europe
Minsk	Belarus	Europe
Bern	Switzerland	Europe
Praha	Czech Republic	Europe
Berlin	Germany	Europe

Result 51

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.

```
556 •  SELECT country.Name AS CountryName, AVG(city.Population) AS AvgPopulation
557    FROM city
558    JOIN country
559    ON city.CountryCode = country.Code
```

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

CountryName	AvgPopulation
Aruba	29034.0000
Afghanistan	583025.0000
Angola	512320.0000
Anguilla	778.0000
Albania	270000.0000
Andorra	21189.0000
Netherlands Antilles	2345.0000

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.

```
5508 •  select country.Name,city.Name,city.Population
5509    from country
5510    left join city
5511    on country.capital= city.ID
5512    order by city.Population Desc;
5513
```

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

Name	Name	Population
Russian Federation	Moscow	8389200
Japan	Tokyo	7980230
China	Peking	7472000
United Kingdom	London	7285000
Egypt	Cairo	6789479

Result 49

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.

```
5517 •  select Name, Population,SurfaceArea ,
5518    (Population/SurfaceArea)As PopulationDensity from country
5519    where(Population/SurfaceArea)<50
5520    order by PopulationDensity Asc;
5521
5522
5523
```

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

Name	Population	SurfaceArea	PopulationDensity
Estonia	1439200	45227.00	31.8217
Eritrea	3850000	117600.00	32.7381
South Africa	40377000	1221037.00	33.0678
Guinea-Bissau	1213000	36125.00	33.5779
Yemen	18112000	527968.00	34.3051
Afghanistan	22720000	652090.00	34.8418
Colombia	42321000	1138914.00	37.1591

Result 58

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.

```
5480    SELECT ci.Name AS CityName, ci.CountryCode, co.Name AS CountryName,
5481           ROUND(co.GNP / co.Population, 2) AS GDPPerCapita
5482    FROM city ci
5483    JOIN country co ON ci.CountryCode = co.Code
5484    WHERE (co.GNP / co.Population) > (
5485        SELECT AVG(GNP /Population)
5486        FROM country
5487         WHERE GNP IS NOT NULL AND Population > 0
5488    );
```

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

```
5522  WITH RankedCities AS (
5523      SELECT *,
5524             ROW_NUMBER() OVER (ORDER BY population DESC) AS rownum
5525      FROM city
5526  )
5527  SELECT *
5528  FROM RankedCities
5529  WHERE rownum between 31 and 40;
5530
```

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

ID	Name	CountryCode	District	Population	rownum
1896	Shenyang	CHN	Liaoning	4265200	31
1897	Kanton [Guangzhou]	CHN	Guangdong	4256300	32
3208	Singapore	SGP	–	4017733	33
3769	Ho Chi Minh City	VNM	Ho Chi Minh City	3980000	34
1027	Chennai (Madras)	IND	Tamil Nadu	3841396	35
2332	Pusan	KOR	Pusan	3804522	36
3794	Los Angeles	USA	California	3694820	37

Result 64

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.