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?	Is a column in a relational database table that's distinctive for each record.
How does this differ from a secondary key?	A secondary key is made on a field that you would like to be indexed for faster searches. A table can have more than one secondary key.
How are primary and foreign keys related?	The primary key is a unique identifier within its table, whereas a foreign key is a reference in one table to a primary key in another. Primary keys enforce uniqueness within their table, ensuring each record is identifiable. Foreign keys, however, are used to establish and navigate relationships between tables.
Provide a real-world example of a one-to-one relationship	Husband and Wife Passport and Person
Provide a real-world example of a one-to-many relationship	Mother and children Doctor and Patient Social Media Influencer and followers
Provide a real-world example of a many-to-many relationship	Students and Courses Authors and Books

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?	The non-relational database, or NoSQL database, stores data. However, unlike the relational database, there are no tables, rows, primary keys or foreign keys. Instead, the non-relational database uses a storage model optimized for specific requirements of the type of data being stored.
What type of data would benefit off the non-relational model? **Why?**	Semi-structured and unstructured content such as images, videos, documents. Non-relational databases are suitable for both operational and transactional data. They are more suitable for unstructured big data. Non-relational databases offer higher performance and availability. Flexible schema helps non-relational databases store more data of varied types that can be changed without major schema changes.

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

A self join is a regular join, but the table is joined with itself.

Result:

Number of Records: 88

CustomerName1	CustomerName2	City
Rancho grande	Cactus Comidas para llevar	Buenos Aires
Cactus Comidas para llevar	Océano Atlántico Ltda.	Buenos Aires
Rancho grande	Océano Atlántico Ltda.	Buenos Aires
Cactus Comidas para llevar	Rancho grande	Buenos Aires
Océano Atlántico Ltda.	Rancho grande	Buenos Aires

Right join

The RIGHT JOIN keyword returns all records from the right table (table2), and the matching records (if any) from the left table (table1).

Result:

Number of Records: 832

OrderID	CustomerName	OrderDate
10248	Wilman Kala	1996-07-04
10249	Tradição Hipermercados	1996-07-05
10250	Hanari Carnes	1996-07-08
10251	Victuailles en stock	1996-07-08
10252	Suprêmes délices	1996-07-09

Full join

FULL JOIN (or FULL OUTER JOIN) is a powerful technique used to combine records from two or more tables. Unlike an INNER JOIN, which only returns rows where there are matches in both tables, a FULL JOIN retrieves all rows from both tables, filling in NULL values where matches do not exist.

Inner join

The INNER JOIN keyword selects records that have matching values in both tables.

Result:

Number of Records: 830

OrderID	CustomerName	OrderDate
10248	Wilman Kala	1996-07-04
10249	Tradição Hipermercados	1996-07-05
10250	Hanari Carnes	1996-07-08
10251	Victuailles en stock	1996-07-08

Cross join

The CROSS JOIN keyword returns all records from both tables (table1 and table2).

Result:

Number of Records: 830

OrderID	CustomerName	OrderDate
10248	Wilman Kala	1996-07-04
10249	Tradição Hipermercados	1996-07-05
10250	Hanari Carnes	1996-07-08
10251	Victuailles en stock	1996-07-08

Left join

The LEFT JOIN keyword returns all records from the left table (table1), and the matching records (if any) from the right table (table2).

Result:

Number of Records: 830

OrderID	CustomerName	OrderDate
10248	Wilman Kala	1996-07-04
10249	Tradição Hipermercados	1996-07-05
10250	Hanari Carnes	1996-07-08
10251	Victuailles en stock	1996-07-08

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? Star Schema with Sales as the central table and customer and stock linked.*
    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

Conveniently Yours Store Database

Creation of a database schema to manage customer, stock and sales using a star schema that would require the following tables.
Staff – Dimension table that could be used to track staff interactions with customers.
Customer – Dimension table that would store customer details.
Sales – Main fact table that would track all sales of store items
Stock – Dimension table that would hold a list of all products sold within the store would include fields for suppler, cost and purchase date
Suppliers – Dimension table that would list all suppliers used by the store.
Loyalty – Dimension table that would be linked to the customer table to track points.
Table fields would include the following, primary key listed in bold.

*TblCustomer - CustomerName, CustomerAddress, **CustomerID**, LoyaltyID, EmailAddress*
*TblStock - **ProductID**, Product, SupplierID, SalePrice, Quantity, PurchaseDate*
*TblSupplier - **SupplierID**, ProductID, Price*
*TblStaff – **StaffID**, StaffName, StaffAddress, StaffContact, StaffRole*
*FCTSales – **SalesID**, CustomerID, ProductID, SoldQuantity, Cost, SalesDate, LoyaltyID*
TblLoyalty – LoyaltyID, CustomerID, Points

Access would be required by staff to maintain stock levels and confirm customer purchases or loyalty points, access could also be granted to the loyalty table for customers to confirm point totals.

As previous state a start schema could be utilised with Sales table central utilising 1-to-Many relationships from the Customer and Stock Table further 1-to-Many relationships would be required between the Customer and Loyalty tables, Stock and Supplier Tables.

A SQL script could be used to create the tables and create the customer and other required tables

```
CREATE TABLE customer (
  customer_id SMALLINT UNSIGNED NOT NULL AUTO_INCREMENT,
  first_name VARCHAR(45) NOT NULL,
  last_name VARCHAR(45) NOT NULL,
  address VARCHAR(45) NOT NULL,
  email VARCHAR(45) NOT NULL,
  last_update TIMESTAMP NOT NULL DEFAULT CURRENT_TIMESTAMP ON UPDATE
CURRENT_TIMESTAMP,
  PRIMARY KEY  (customer_id),
  KEY idx_loyalty_last_name (loyalty_ID)
  CONSTRAINT `fk_tblloyalty` FOREIGN KEY (loyalty_id) REFERENCES loyalty
(loyalty_id) ON DELETE RESTRICT ON UPDATE CASCADE
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4;
```

Depending on how the store currently stores inventory for example if it is currently stored within an excel sheet then it could be easily converted into information that can be INSERT INTO the following SQL script. Customers would need to be done as required by staff at point of sale if the customer is willing, Staff details could be manually entered by Management.

A GUI could also be created that would allow staff to manually create records for Customers and Stock entries.

```
SET AUTOCOMMIT=0;
INSERT INTO stock VALUES (1,'PEN','SBIC',’0.10’,’200’,'2025-01-15 04:34:33'),
(1,'BEANS','HEIZN',’0.30’,’50’,'2025-01-15 04:34:33'),
(1,'TOMATO SOUP','HEIZN',’0.50’,’50’,'2025-01-15 04:34:33'),
(Further stock items listed as required);
COMMIT;
```

Database maintenance could be achieved by implement data quality frameworks. Which could include some of the follow measures, Regular data audits which would ensure that the data is accurate. Automated validation checks again ensuring basic data accuracy. Training and education making sure staff know how to use the system and if needed correct any data errors. Feedback mechanisms that allow staff and customer to highlight and changes or issue that may arise with the data. Data source verification making sure that data input is correct and making sure that errors are minimised. Use data cleansing tools along with data audit this can make sure that all data is correct when required. Maintain documentation accurate and up to date documentation making sure that any updates made to the database is documented and updated training is rolled out when database is updated.

Data security and backups are key no matter the size of the organisation and can be seen as relatively unimportant as once the data is copied/backed up to whatever storage medium the business chooses to use. In this case the store needs to minimise the risk to its data, it is important that precautions are taken to ensure that data is part of a security strategy. Access rights to data backups should be restricted. Different backup locations should be used if possible. Physical access to data backups should also be restricted. Protected backup devices and make sure the network is secure.

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.

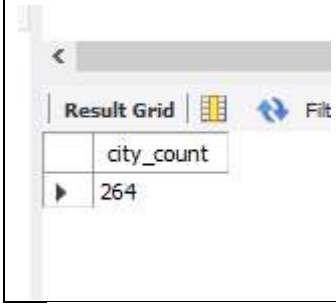
Setting up the database:

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.

```
SELECT COUNT(DISTINCT name) as city_count
FROM city
WHERE countrycode = 'USA';
```

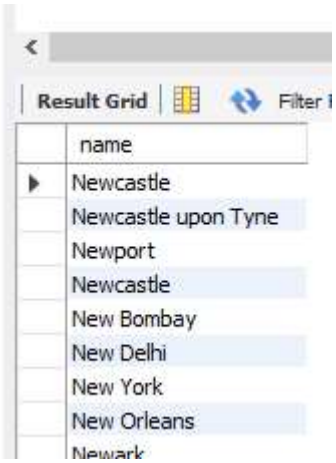


2. **Country with Highest Life Expectancy:** *Scenario:* As part of a global health initiative, you've been assigned to identify the country with the highest life expectancy. This information will be crucial for prioritising healthcare resources and interventions.

```
select Name, LifeExpectancy
From country
order by LifeExpectancy DESC
Limit 1;
```

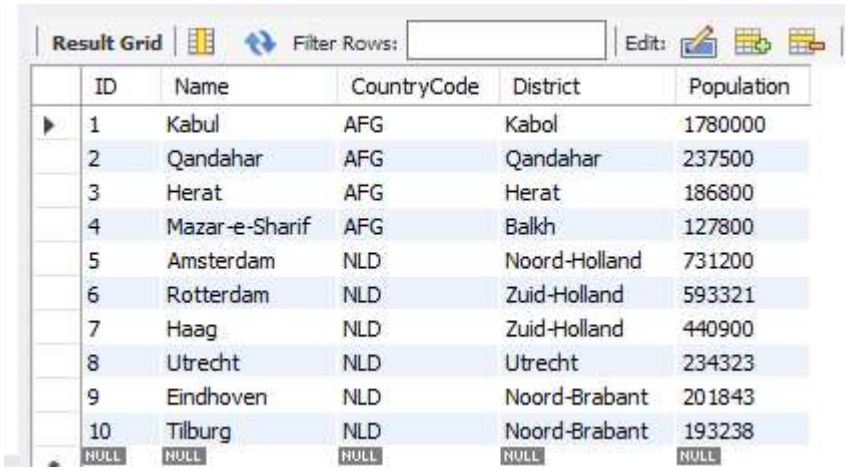
3. **"New Year Promotion: Featuring Cities with 'New :** *Scenario:* In anticipation of the upcoming New Year, your travel agency is gearing up for a special promotion featuring cities with names including the word 'New'. You're tasked with swiftly compiling a list of all cities from around the world. 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
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 * from city
Limit 10;
```



Result Grid | Filter Rows: | Edit:

ID	Name	CountryCode	District	Population
1	Kabul	AFG	Kabol	1780000
2	Qandahar	AFG	Qandahar	237500
3	Herat	AFG	Herat	186800
4	Mazar-e-Sharif	AFG	Balkh	127800
5	Amsterdam	NLD	Noord-Holland	731200
6	Rotterdam	NLD	Zuid-Holland	593321
7	Haag	NLD	Zuid-Holland	440900
8	Utrecht	NLD	Utrecht	234323
9	Eindhoven	NLD	Noord-Brabant	201843
10	Tilburg	NLD	Noord-Brabant	193238
NULL	NULL	NULL	NULL	NULL

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.

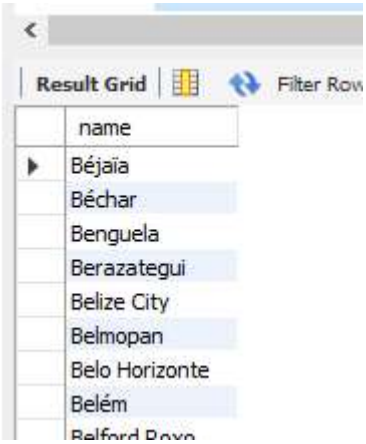
```
Select *
From city
Where Population > 2000000;
```

Result Grid | Filter Rows: | Edit:

ID	Name	CountryCode	District	Population
35	Alger	DZA	Alger	2168000
56	Luanda	AGO	Luanda	2022000
69	Buenos Aires	ARG	Distrito Federal	2982146
130	Sydney	AUS	New South Wales	3276207
131	Melbourne	AUS	Victoria	2865329
150	Dhaka	BGD	Dhaka	3612850
206	São Paulo	BRA	São Paulo	9968485
207	Rio de Janeiro	BRA	Rio de Janeiro	5598953
208	Salvador	BRA	Bahia	2302832

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
From city
Where Name like 'Be%';
```



name
Béjaïa
Béchar
Benguela
Berazategui
Belize City
Belmopan
Belo Horizonte
Belém

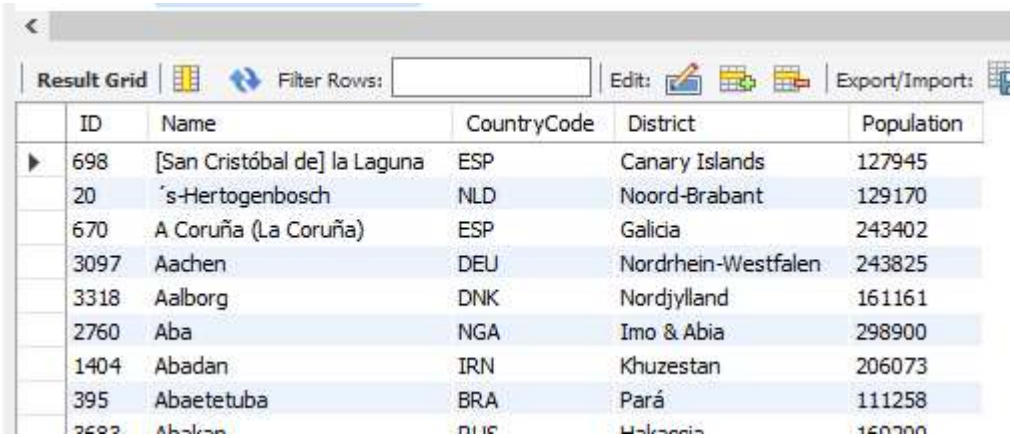
7. **Cities with Population Between 500,000-1,000,000:** *Scenario:* An urban planning committee needs to identify mid-sized cities suitable for infrastructure development projects. You're tasked with identifying cities with populations ranging between 500,000 and 1 million to inform their decision-making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

```
Select *
From city
Where Population between 500000 and 1000000;
```

ID	Name	CountryCode	District	Population
5	Amsterdam	NLD	Noord-Holland	731200
6	Rotterdam	NLD	Zuid-Holland	593321
36	Oran	DZA	Oran	609823
64	Dubai	ARE	Dubai	669181
72	Rosario	ARG	Santa Fé	907718
73	Lomas de Zamora	ARG	Buenos Aires	622013
74	Quilmes	ARG	Buenos Aires	559249
75	Almirante Brown	ARG	Buenos Aires	538918
76	La Plata	ARG	Buenos Aires	521936

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 *
From city
Order by Name Asc;
```



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

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 *
From city
Order by population desc;
```

Result Grid | Filter Rows: | Edit: | Exp

ID	Name	CountryCode	District	Population
1024	Mumbai (Bombay)	IND	Maharashtra	10500000
2331	Seoul	KOR	Seoul	9981619
206	São Paulo	BRA	São Paulo	9968485
1890	Shanghai	CHN	Shanghai	9696300
939	Jakarta	IDN	Jakarta Raya	9604900
2822	Karachi	PAK	Sindh	9269265
3357	Istanbul	TUR	Istanbul	8787958
2515	Ciudad de México	MEX	Distrito Federal	8591309

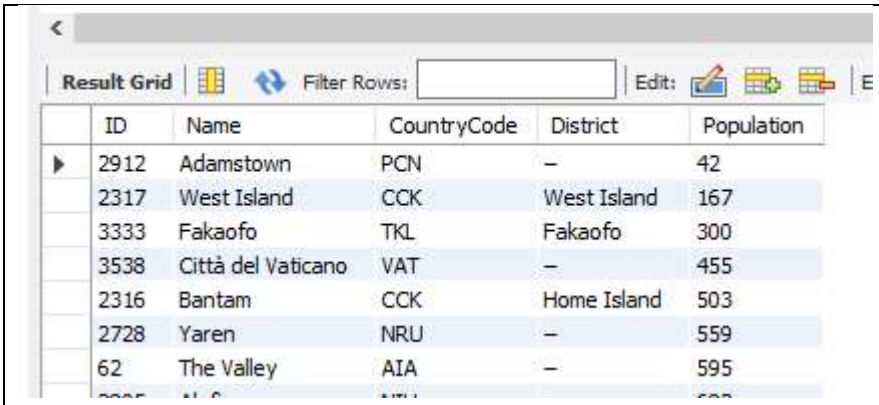
10. **City Name Frequency Analysis: Supporting Geography Education** *Scenario*: In a geography class, students are learning about the distribution of city names around the world. The teacher, in preparation for a lesson on city name frequencies, wants to provide students with a list of unique city names 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, COUNT(*) as city_count
FROM city
Group by name
order by name asc;
```

name	city_count
Antalya	1
Antananarivo	1
Antipolo	1
Antofagasta	1
Antsirabé	1
Antwerpen	1
Anyang	2

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 *
From city
Order by population ASC;
```



Result Grid | Filter Rows: | Edit:

ID	Name	CountryCode	District	Population
2912	Adamstown	PCN	–	42
2317	West Island	CCK	West Island	167
3333	Fakaofo	TKL	Fakaofo	300
3538	Città del Vaticano	VAT	–	455
2316	Bantam	CCK	Home Island	503
2728	Yaren	NRU	–	559
62	The Valley	AIA	–	595

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 *
From country
Order by population DESC;
```

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

Code	Name	Continent	Region	SurfaceArea	IndepYear	Population	LifeExpectancy	GNP	GNPOld
CHN	China	Asia	Eastern Asia	9572900.00	-1523	1277558000	71.4	982268.00	917719.00
IND	India	Asia	Southern and Central Asia	3287263.00	1947	1013662000	62.5	447114.00	430572.00
USA	United States	North America	North America	9363520.00	1776	278357000	77.1	8510700.00	8110900.00
IDN	Indonesia	Asia	Southeast Asia	1904569.00	1945	212107000	68.0	84982.00	215002.00
BRA	Brazil	South America	South America	8547403.00	1822	170115000	62.9	776739.00	804108.00
PAK	Pakistan	Asia	Southern and Central Asia	796095.00	1947	156483000	61.1	61289.00	58549.00

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.

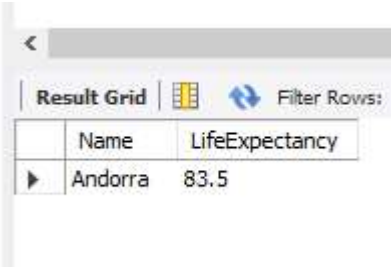
```
SELECT *
FROM city
WHERE countrycode = 'ESP';
```

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

ID	Name	CountryCode	District	Population
653	Madrid	ESP	Madrid	2879052
654	Barcelona	ESP	Katalonia	1503451
655	Valencia	ESP	Valencia	739412
656	Sevilla	ESP	Andalusia	701927
657	Zaragoza	ESP	Aragonia	603367
658	Málaga	ESP	Andalusia	530553
659	Bilbao	ESP	Baskimaa	357589
660	Las Palmas de Gran Canaria	ESP	Canary Islands	354757
661	Murcia	ESP	Murcia	353504
662	Palma de Mallorca	ESP	Balears	326993

14. **Country with Highest 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.

```
select Name, LifeExpectancy
From country
order by LifeExpectancy DESC
Limit 1;
```

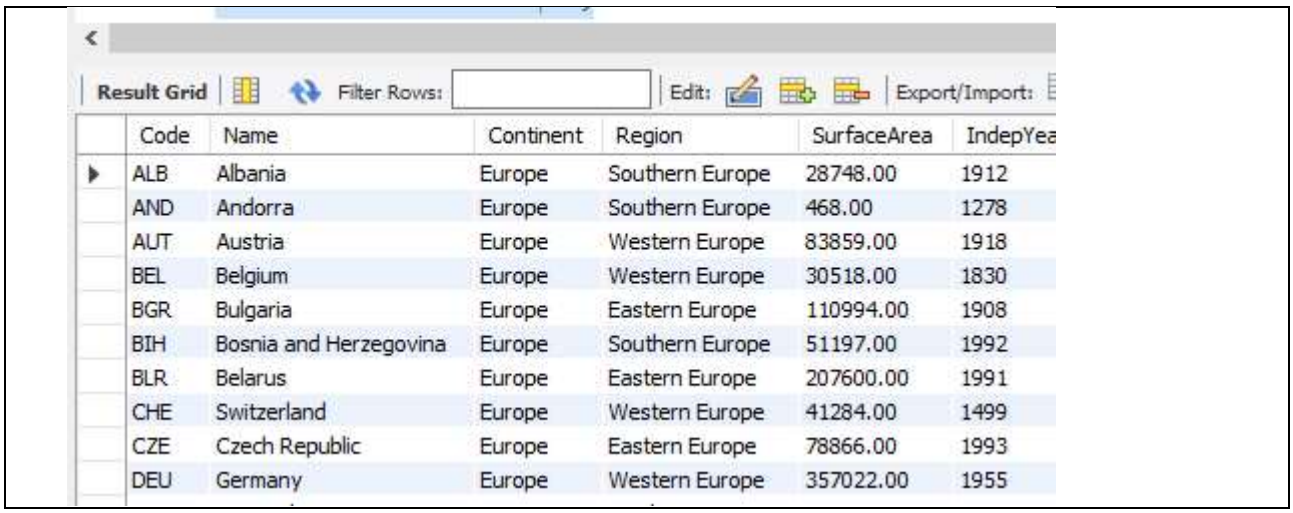


Result Grid | Filter Rows:

Name	LifeExpectancy
Andorra	83.5

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.

```
SELECT *
FROM country
WHERE Continent = 'Europe'
```

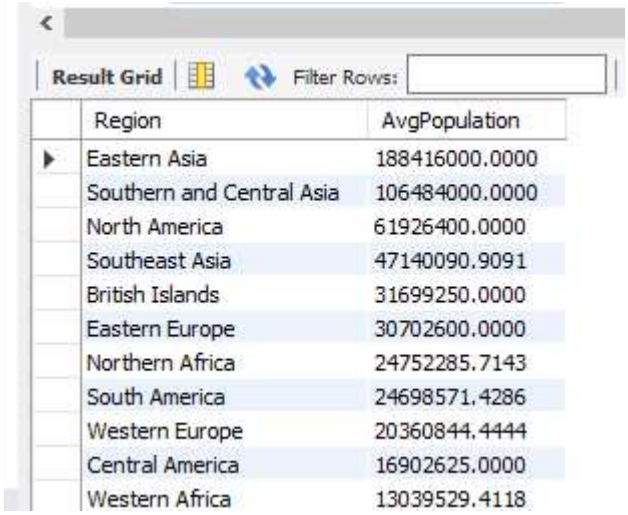


Result Grid | Filter Rows: | Edit: | Export/Import:

Code	Name	Continent	Region	SurfaceArea	IndepYea
ALB	Albania	Europe	Southern Europe	28748.00	1912
AND	Andorra	Europe	Southern Europe	468.00	1278
AUT	Austria	Europe	Western Europe	83859.00	1918
BEL	Belgium	Europe	Western Europe	30518.00	1830
BGR	Bulgaria	Europe	Eastern Europe	110994.00	1908
BIH	Bosnia and Herzegovina	Europe	Southern Europe	51197.00	1992
BLR	Belarus	Europe	Eastern Europe	207600.00	1991
CHE	Switzerland	Europe	Western Europe	41284.00	1499
CZE	Czech Republic	Europe	Eastern Europe	78866.00	1993
DEU	Germany	Europe	Western Europe	357022.00	1955

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 Region , AVG(Population) As AvgPopulation
From country
Group by Region
Order by AvgPopulation DESC;
```

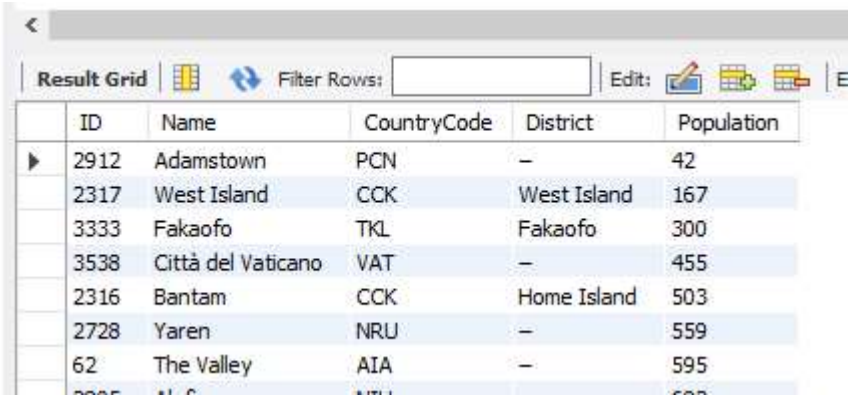


Result Grid | Filter Rows:

Region	AvgPopulation
Eastern Asia	188416000.0000
Southern and Central Asia	106484000.0000
North America	61926400.0000
Southeast Asia	47140090.9091
British Islands	31699250.0000
Eastern Europe	30702600.0000
Northern Africa	24752285.7143
South America	24698571.4286
Western Europe	20360844.4444
Central America	16902625.0000
Western Africa	13039529.4118

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 *
From city
Order by population ASC;
```



Result Grid | Filter Rows: | Edit:

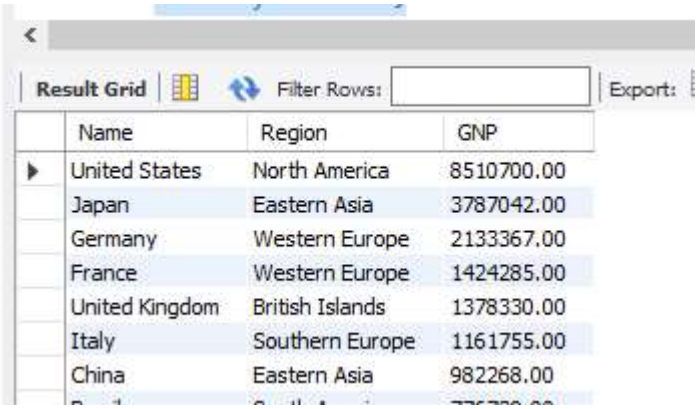
ID	Name	CountryCode	District	Population
2912	Adamstown	PCN	–	42
2317	West Island	CCK	West Island	167
3333	Fakaofo	TKL	Fakaofo	300
3538	Città del Vaticano	VAT	–	455
2316	Bantam	CCK	Home Island	503
2728	Yaren	NRU	–	559
62	The Valley	AIA	–	595

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.

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 Name, Region, GNP
From country
```

Where GNP > (Select AVG(GNP) from country)
Order by GNP Desc;



Result Grid | Filter Rows: | Export:

Name	Region	GNP
United States	North America	8510700.00
Japan	Eastern Asia	3787042.00
Germany	Western Europe	2133367.00
France	Western Europe	1424285.00
United Kingdom	British Islands	1378330.00
Italy	Southern Europe	1161755.00
China	Eastern Asia	982268.00

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.

Select *

From city

Order by population desc

Limit 10 offset 30;

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

ID	Name	CountryCode	District	Population
1896	Shenyang	CHN	Liaoning	4265200
1897	Kanton [Guangzhou]	CHN	Guangdong	4256300
3208	Singapore	SGP	–	4017733
3769	Ho Chi Minh City	VNM	Ho Chi Minh City	3980000
1027	Chennai (Madras)	IND	Tamil Nadu	3841396
2332	Pusan	KOR	Pusan	3804522
3794	Los Angeles	USA	California	3694820
150	Dhaka	BGD	Dhaka	3612850
3068	Berlin	DEU	Berliini	3386667
2710	Rangoon (Yangon)	MMR	Rangoon [Yangon]	3361700
NULL	NULL	NULL	NULL	NULL

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.