Week 1: SQL Basics

Introduction to Databases and SQL:

- What is a database?
- Introduction to Structured Query Language (SQL).
- Overview of popular SQL database management systems (e.g., MySQL, PostgreSQL).

Basic SQL Queries:

- SELECT statement for data retrieval.
- Filtering data using the WHERE clause.
- Sorting data using the ORDER BY clause.
- Limiting results with the LIMIT clause.

Week 1 covers the fundamentals of SQL, starting with an introduction to databases and SQL itself. Here's a simplified breakdown:

What is a database?

A database is like an organized collection of information stored on a computer. It's a structured way to store, manage, and retrieve data.

Introduction to SQL (Structured Query Language):

SQL is a language used to communicate with databases. It helps in performing various tasks like retrieving, updating, and managing data within a database.

Popular Database Management Systems:

Examples include *MySQL* and *PostgreSQL*, which are software used to create and manage databases.

Basic SQL Queries:

SELECT statement: Retrieves specific data from a database.

WHERE clause: Filters data based on specific conditions.

ORDER BY clause: Sorts the retrieved data in a specified order (like alphabetical or

numerical).

LIMIT clause: Sets a maximum number of results to be retrieved.

In essence, this week covers the foundational aspects of databases, SQL, and how to retrieve and manipulate data using basic SQL commands like SELECT, WHERE, ORDER BY, and LIMIT.

Data Modification Statements:

- INSERT, UPDATE, DELETE queries to add, update, and delete data.
- Best practices and considerations for data modification.

Data Cleaning and Data Manipulation:

- Identifying and handling missing or incorrect data.
- Using SQL functions for data cleansing (TRIM, LOWER, UPPER, etc.).
- Transforming data using string and date functions.

Here is a simplified breakdown of these concepts:

Data Modification Statements:

INSERT: Adds new data into a database.

UPDATE: Modifies existing data within a database.

DELETE: Removes data from a database.

These statements allow you to add, change, or remove information in a structured manner. Best practices involve being cautious while modifying data, ensuring accuracy, and considering backups before making significant changes.

Data Cleaning and Data Manipulation:

Identifying Missing or Incorrect Data: Recognizing and dealing with data that's either incomplete or inaccurate.

Using SQL Functions for Data Cleansing: Functions like *TRIM* (removes extra spaces), *LOWER* (converts text to lowercase), *UPPER* (converts text to uppercase), etc., help clean and standardize data.

Transforming Data: Utilizing functions specific to strings (text) or dates to change their format or structure within the database.

This part focuses on maintaining data accuracy by recognizing and handling missing or incorrect data, utilizing SQL functions to clean and transform data, and following best practices for modifying data within a database.

Here are some helpful resources to understand SQL basics, data modification statements, data cleaning, and manipulation:

W3Schools SQL Tutorial: Offers a beginner-friendly introduction to SQL with examples. https://www.w3schools.com/sql/

Khan Academy - Intro to SQL: Provides a simple introduction to databases and SQL. https://www.khanacademy.org/computing/computer-programming/sql

SQL Date Functions by SQLTutorial.org: Demonstrates date manipulation functions in SQL.

https://www.sqltutorial.org/sql-date-functions/

PostgreSQL Tutorial by PostgreSQL Tutorial Website: Provides examples and explanations for data modification in PostgreSQL. https://www.postgresqltutorial.com/

In the syntax:

SELECT is a list of one or more columns.

DISTINCT suppresses duplicates.

selects all columns

column selects the named column.

alias gives selected columns different headings.

FROM table specifies the table containing the columns.

Note: Throughout this module, the words keyword, clause, and statement are used.

A **keyword** refers to an individual SQL element. For example, **SELECT** and **FROM** are keywords.

A **clause** is a part of an SQL statement. For example. **SELECT empno**, **ename**, ... is a clause.

A **statement** is a combination of two or more clauses. For example. **SELECT** * **FROM** emp is a **SQL** statement.

Writing SQL Statements:

Using the following simple rules and guidelines, you can construct valid statements that are both easy to read and easy to edit:

- SQL statements are not case sensitive, unless indicated.
- > SQL statements can be entered on one or many lines.
- Keywords cannot be split across lines or abbreviated.
- Clauses are usually placed on separate lines for readability and ease of editing.
- Tabs and indents can be used to make code more readable.
- ➤ Keywords typically are entered in uppercase; all other words, such as table names and columns, are entered in lowercase.

Following are some resources which can help you get familiar with MySQL Workbench: https://www.youtube.com/watch?v=2bW3HuaAUcY&ab channel=365DataScience

https://youtu.be/x ez4IISGOE?feature=shared

Session 1:

Case Study: Telco Customer Management System

Background:

You've been appointed as a Database Analyst for a telecommunications company ("neon telco") that offers mobile, broadband, and streaming services. They're looking to revamp their Customer Management System to better track customer, their subscriptions, usage data, and customer service interactions.

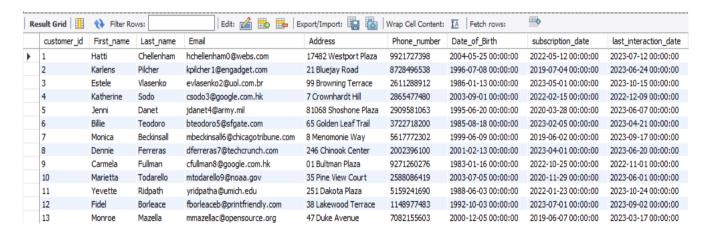
Basic SQL Queries:

Using SELECT

1. Exercise: Retrieve all columns and rows from the `customer` table.

Query: SELECT *

FROM customer;

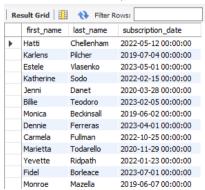


2. Exercise: List only the names and subscription dates of the customer.

Query:

SELECT first_name,last_name,subscription_date

FROM customer;



3. Exercise: Fetch all unique email addresses from the `customer` table and then count the number of unique email ids because it should be unique for all.

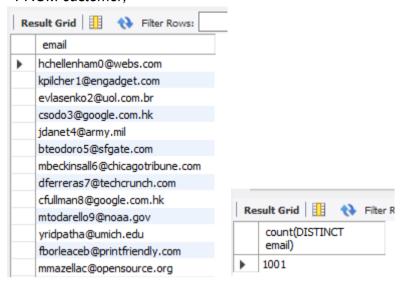
Query:

SELECT DISTINCT email

FROM customer;

SELECT count(DISTINCT email)

FROM customer;



4. Exercise: Display all columns from the `billing` table.

Query:

SELECT *

FROM billing;

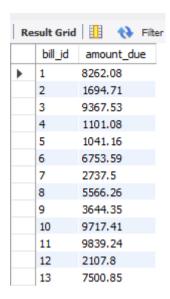
	_	· J/						
Re	sult Grid	│ Ⅲ 🙌 Fil	ter Rows:		Edit: 💪 📆 🖶 Export/Import: 📳 🐻 Wrap Cell Conte			
	bill_id	Customer_id	amount_due	due_date	payment_date	billing_cycle	discounts_applied	late_fee
•	1	100	8262.08	5/4/2023	12/22/2023	23-Nov	103.71	955.9
	2	101	1694.71	9/17/2023	8/17/2023	23-Jun	609.81	487.07
	3	102	9367.53	11/4/2023	9/25/2023	23-Jan	202.52	177.2
	4	103	1101.08	10/17/2023	11/1/2023	23-Jun	715.54	30.09
	5	104	1041.16	10/2/2023	3/13/2023	23-Jan	272.51	486.27
	6	105	6753.59	6/21/2023	1/13/2023	23-Apr	891.88	846.89
	7	106	2737.5	6/26/2023		23-Jun	15.51	605.7
	8	107	5566.26	1/28/2023	2/21/2023	23-Jun	299.44	70.45
	9	108	3644.35	6/5/2023	11/1/2023	23-Dec	307.98	682.41
	10	109	9717.41	9/4/2023		23-Feb	359.25	843.36
	11	110	9839.24	1/25/2023	1/24/2023	23-Apr	588.77	158.9
	12	111	2107.8	3/16/2023	5/20/2023	23-Dec	869.62	140.35
	13	112	7500.85	4/14/2023	11/29/2023	23-Mar	693.46	36.19

5. Exercise: Show only the bill ID and the amount due from the `billing` table.

Query:

SELECT bill_id, amount_due

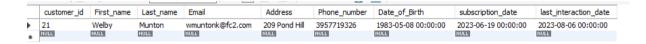
FROM billing;



Using WHERE

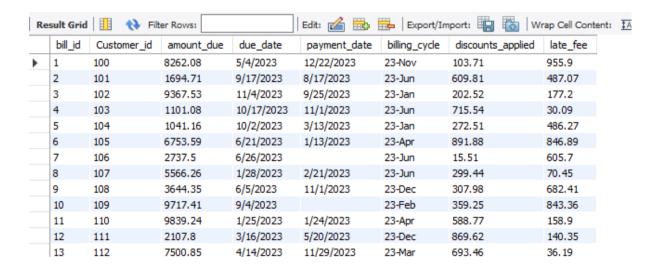
1. Exercise: Identify customer who live at "209 Pond Hill".

Query: SELECT * FROM customer WHERE address = '209 Pond Hill';



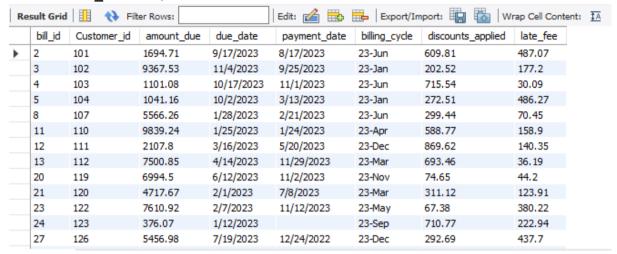
2. Exercise: Find bills in the 'billing' table with an amount due greater than 1000.

Query:
SELECT *
FROM billing
WHERE amount due > 1000;



3. Exercise: Find all the late fee less than 500

SELECT *
FROM billing
WHERE late fee <500;



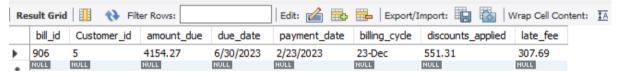
4. Exercise: Show bills that were generated for `customer id' 5 .

Query:

SELECT *

FROM billing

WHERE customer id = '5':



Using WHERE with (IN, OR, AND, NOT EQUAL TO, NOT IN)

1. Exercise: Identify customer who live at either '5 Northridge Road', '814 Kinsman Lane' Query:

SELECT *

FROM customer

WHERE address IN ('5 Northridge Road', '814 Kinsman Lane');



2. Exercise: using or and AND

SELECT *

FROM customer

WHERE address IN ('5 Northridge Road', '814 Kinsman Lane') AND phone_number LIKE '277%':

	customer_id	First_name	Last_name	Email	Address	Phone_number	Date_of_Birth	subscription_date	last_interaction_date
•	22	Tore	Vasishchev	tvasishchevl@shinystat.com	814 Kinsman Lane	2771639301	1989-01-25 00:00:00	2019-09-14 00:00:00	2023-03-19 00:00:00
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

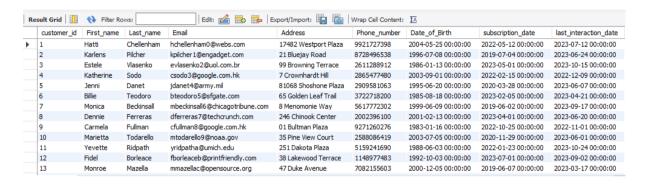
3. Exercise: Display customer whose phone number is NOT '123-456-7890'.

Query:

SELECT *

FROM customer

WHERE phone number <> '123-456-7890';



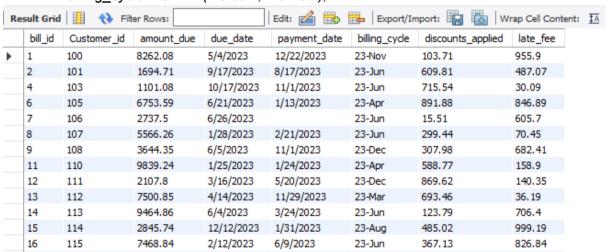
4. Exercise: List all bills except those with billing cycles in "January 2023" and "February 2023".

Query:

SELECT *

FROM billing

WHERE billing_cycle NOT IN ('23-Jan', '23-Feb');



Using ORDER BY

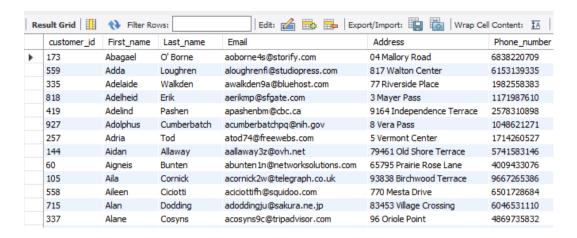
1. Exercise: Order customer by their names in ascending order.

Query:

SELECT *

FROM customer

ORDER BY first name ASC;

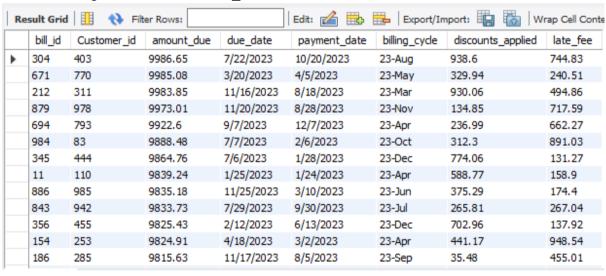


2. Exercise: Display bills from the `billing` table ordered by `amount_due` in descending order.

Query:

SELECT *

FROM billing ORDER BY amount_due DESC;

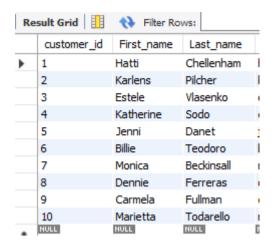


Using LIMIT

1. Exercise: Show only the first 10 customer.

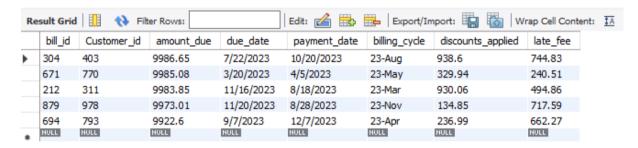
Query:

SELECT * FROM customer LIMIT 10:



2. Exercise: List the top 5 highest bills from the 'billing' table.

Query: SELECT * FROM billing ORDER BY amount_due DESC LIMIT 5;



3. Exercise: Retrieve the latest 3 bills based on the due date.

Query: SELECT * FROM billing ORDER BY due_date DESC LIMIT 3;

