**Role of SQL in Data Analysis**

**Prachi Bhandari**



**Introduction to SQL**

What is SQL? Why SQL is so important in various role like Data Analyst, Data Science, Business Analyst etc? These are general questions which came into mind of everyone. So here is the answer to all the questions. SQL stands for Structured Query Language. It is programming language used to manage, storing, manipulating, retrieving and modifying data in database. It is important because it a popular query language that is frequently used in all types of applications for few reasons

* It's semantically easy to understand and learn.
* Because it can be used to access large amounts of data directly where it's stored, analysts don't have to copy data into other applications.
* Compared to spreadsheet tools, data analysis done in SQL is easy to audit and replicate. For analysts, this means no more looking for the cell with the typo in the formula.

SQL is the language used to interact with relational database. Since most systems today capture the data using one or more database like MYSQL, Oracle, Redshift, SQL Server, MS Access, IBM DB2 and many others, you need to know SQL to extract data from these systems. The good thing about SQL syntax is you don’t need to learn a language specific to each database; you can learn SQL once and make small adjustments as you work in different database.

**Skills Data Analyst should have**

A Data Analyst is someone who uses their skills to collect, analyse, and report insights from company’s data. The major skills that Data Analyst should have:

1. Problem solving and critical thinking.
2. SQL
3. Statistics
4. Data Visualizations
5. Presentation Skills

**Role of SQL in Data Analysis**

SQL is the “meat and potatoes” of data analysis- it’s used for accessing, cleaning and analysing data that’s stored in database. Analyst can craft precise and tailored queries to extract relevant data subsets, facilitating the initial exploration and understanding of the dataset. SQL helps in data transformation, cleaning, preprocessing, standardise data. So, this ensure that data is suitable format and proper quality for in-depth analysis. Other than data transformation it also helps in bringing data from multiple sources which helps in analysis. As mentioned above one of the skills that data analyst should pursue is SQL, In SQL data analyst include modelling in specific format, and creating visualizations or view highlight information that can be used by decision-makers.

A data analyst needs to work on large volume of data within a short period of time and sometimes need to use same syntax for similar result so for this SQL Query are there. The SQL Query can fetch results from a few million rows within a minute. In addition to that, SQL Queries you write are repeatable and scalable. You can write a query once and then reuse it again and again. As the data increases in your database, you don’t need to change much in your query to accommodate similar results. In fact, even if new tables are added, you can use the previous query and make a few changes; it should still work just fine. Databases also generally support scalability for adding new data – e.g. easily increasing storage in the background or processing power – especially when they are cloud based.

**Top Tools in SQL used by Data Analyst.**

The key tools and features in SQL for data analysis:

**1.** **SELECT and FROM Statements**: The SELECT statement is the foundational tool for SQL for data analysis. It allows you to retrieve specific data from a database table, including specific columns, rows, or calculated values. The FROM statement specifies the location or table from which the data needs to be retrieved.

**2. WHERE Clause**: The WHERE clause is used for filtering data. It is used to extract only those record that fulfil a specified condition.

**3. GROUP BY Clause:** The GROUP BY clause is used for data aggregation. It groups rows that have the same values into summary rows, like “like find the number of customers in each country” and perform aggregate functions (e.g., SUM, AVG, COUNT) on those groups.

**4. JOIN Operations:** A JOIN Clause is used to combine rows from two or more tables, based on the related column between them. SQL supports different types of joins, including INNER JOIN, LEFT JOIN, RIGHT JOIN, and FULL JOIN, which enable you to combine data from multiple tables based on specified criteria.

**5. ORDER BY Clause:** The ORDER BY clause is used to sort the results-set in ascending or descending order based on one or more columns. It’s used for arranging data for analysis.

**6. Data Modification Statements:** SQL not only retrieves data but also allows you to modify data using statements like INSERT, UPDATE, and DELETE. This is important for data preparation and cleaning.

**7. Indexing:** SQL databases provide indexing mechanisms to improve query performance, making data retrieval faster, which is crucial for large datasets.

**8. Reporting Tools**: Many SQL-based database management systems offer reporting and visualisation tools that allow you to create charts, graphs, and reports based on SQL query results.

**Conclusion**

SQL functions are instrumental in efficiently managing vast amounts of data. By utilizing various functions, we enhance our ability to analyse data swiftly, facilitating informed decision-making within organizations. SQL’s pivotal role in data analysis cannot be overstated, as it empowers us to extract insights, discern patterns, and derive actionable intelligence from complex datasets.Top of Form