

Exploratory Data Analysis in SQL

Aggregate function: A function that computes a single summary value from multiple rows, such as COUNT, SUM, AVG, MIN, MAX, VARIANCE, or STDDEV

Cast (casting / ::): The operation of converting a value from one data type to another for the duration of a query, either using CAST(value AS type) or the shorthand value::type

Coalesce: A function that returns the first non-NULL value from a list of expressions for each row, commonly used to provide fallback or default values

Column: A named field in a table that defines the type of data stored for every row, often called a variable or attribute

Common table expression (WITH clause): A named, temporary result set defined at the start of a query that can be referenced like a table within that query to simplify complex SQL and improve readability

Data type: A specification for a column that determines the kind of values it can store (for example integer, numeric/decimal, text, date, timestamp, boolean, or interval) and how those values are represented and operated on

Database client: A program or interface used to connect to and run queries against a database, providing commands and tools to inspect tables, run SQL, and retrieve results

Document loader: A LangChain class designed to read and normalize data from various file formats (e.g., PDF, CSV, HTML) into document objects that can be processed for retrieval or indexing

Figure: The top-level container object in Matplotlib that holds all plot elements (Axes, titles, legends, etc.) and represents the entire image or page

extract / date_part: Functions that return a specific field (for example year, month, day, hour, minute, or week) from a date or timestamp value for analysis or grouping

Foreign key: A column (or columns) in one table whose values reference the primary key of another (or the same) table to define a formal relationship between rows

generate_series: A function that produces a sequence of numeric or timestamp values from a start to an end with a specified step, commonly used to build bins or fill missing time intervals

Group By: An SQL clause that groups rows sharing the same values of specified columns so aggregate functions can compute summaries per group

Interval: A data type that represents a duration of time (for example “3 days” or “1 hour 30 minutes”) and is produced by subtracting dates/timestamps or by casting literal interval expressions.

Lead and lag: Window functions that return the value from a following (lead) or preceding (lag) row within an ordered partition, commonly used to compute differences or changes over time

LIKE / ILIKE: Pattern-matching operators where LIKE matches case-sensitively using % (any sequence) and _ (single character) wildcards, while ILIKE performs the same match case-insensitively

NULL: A special marker that represents missing or unknown data in a database column and is not equal to any value, including itself

percentile_cont / percentile_disc: Ordered-set aggregate functions that compute percentiles where percentile_cont interpolates between values to return a continuous result and percentile_disc returns the nearest existing value from the data

Primary key: A column or set of columns designated to uniquely identify each row in a table and which must contain non-NULL, unique values

Row: A single record in a table representing one instance of the entity, with values for each column in that table

split_part: A string function that splits a text value on a specified delimiter and returns the nth part, useful for extracting structured pieces from concatenated fields

Table: A named collection of rows in a relational database that stores records with the same set of columns and represents one entity or relationship

Temporary table: A table created for the duration of a database session that stores intermediate query results or transformed data and is automatically dropped when the session ends

Trim (btrim / ltrim / rtrim): Functions that remove specified characters (by default spaces) from the start (ltrim), end (rtrim), or both ends (btrim/trim) of a string to clean whitespace or unwanted characters

Window function: A function that performs calculations across a set of table rows related to the current row (a window), allowing operations like running totals, ranks, and row offsets without collapsing rows