

# Retail Data Analysis with SQL & Python

## Libraries Used and Their Purpose

- **Pandas**  
For reading CSV files, cleaning data, handling missing values, and creating structured DataFrames.
- **MySQL Connector (mysql.connector)**  
To connect Python with the MySQL database and execute SQL queries for data insertion and analysis.
- **NumPy**  
Used for numerical computations such as calculating correlations between variables.
- **Matplotlib**  
A basic plotting library used to create bar charts and visualize data trends.
- **Seaborn**  
Built on top of Matplotlib, it provides enhanced statistical plots and is used for clear, aesthetically pleasing visualizations.

## CSV to MySQL Data Loader Script

### Description

This Python script automates the process of loading CSV data into a MySQL database. It connects to a local MySQL database named 'ecommerce' and processes several CSV files containing data such as customers, orders, sellers, products, payments, and more.

Each CSV file is read using the pandas library. The script dynamically detects the datatype of each column and creates a corresponding table in the MySQL database with appropriate column definitions. Column names are cleaned to remove spaces and special characters

The script replaces missing values (NaNs) with SQL-friendly NULL equivalents (None in Python) before inserting the data into the database. Each record from the CSV is inserted into the MySQL table one by one. All changes are committed to the database once each file is processed

This script enables a structured, scalable approach to loading large datasets into a relational database for further querying and analysis

### Purpose

To convert raw eCommerce dataset CSV files into structured relational tables within a MySQL database. This helps enable efficient querying, analysis, and integration with other tools for further data processing or dashboarding

# MySQL-Powered EDA and Visualization Script

## Description

This Python script performs SQL-based data analysis and visualizations using the data previously imported into the MySQL 'ecommerce' database. It connects to the database and runs a series of SQL queries to extract meaningful business insights

The analysis includes city and state-level customer distribution, total order counts, monthly order trends, product category-wise sales, percentage of installment-based payments, and sales performance over time. It also includes correlation analysis between product prices and order frequency, and examines top-selling sellers

For visual exploration, the script uses matplotlib and seaborn libraries to generate bar plots and trend charts, helping to communicate trends and patterns visually

Advanced SQL queries using window functions, CTEs, and aggregations are used to perform tasks like moving averages, year-over-year growth calculation, customer retention, and identifying top-paying customers per year

## Purpose

To perform exploratory data analysis (EDA) directly from a MySQL database and visualize key metrics such as customer behavior, sales distribution, growth trends, and seller performance. This helps derive actionable insights for business intelligence