END-TO-END PYTHON + SQL PROJECT - WALMART DATASET

Problem Statement -

This is an end-to-end data analysis project that uses Python and SQL to solve some queries and generate insights from the Walmart dataset available on Kaggle. Python is basically used for data cleaning and processing, while PostgreSQL to solve some basic to advanced queries.

About the dataset

Introduction to Walmart dataset - 10k rows

Let's look at the attributes of this dataset.

Column name	Description
invoice_id	Invoice id
branch	Name of the branch
city	City
category	Type of category product belongs to
unit_price	Price per unit
quantity	Quantity bought
date	Date of purchase
time	At what time sell took place
payment_method	Type of payment
rating	Rating received
profit_margin	Profit

Python - We download the dataset from Kaggle and use VS code as a platform to run the code.

1. **Data cleaning -** We use pandas to perform data cleaning like removing duplicates, removing null values and changing the data types.
2. The cleaned dataset is then used for connecting to a RDBMS, PostgreSQL.

PostgreSQL -

- Perform EDA and get to know your data better
- We use PostgreSQL to solve some basic to advanced SQL queries that will help us dig insights about the sales at Walmart.