E-commerce Sales Analysis

Python & MySQL Data Pipeline

Project Overview

Complete Workflow

Raw CSV files → structured database → business insights

Real-World Simulation

Data collection, cleaning, storage, and analysis for business decisions

Project Objectives

Data Pipeline

Automated CSV import to MySQL

Customer Analysis

Locations and purchasing behavior

Revenue Trends

Product category performance

Payment Patterns

Installment usage analysis

Retention

Repeat purchase behavior

Seller Performance

Revenue-based evaluation

Multi-Table Dataset

01

Customers

Customer information and locations

02

Orders

Order details and timelines

03

Products

Product catalog and categories

04

Sales

Transaction records

05

Payments

Payment methods and installments

06

Delivery

Shipping and delivery status

Methodology

Data Loading

CSV import with Python & Pandas

Data Cleaning

Missing values, standardization, verification

Database Creation

Dynamic table generation in MySQL

Data Insertion

Parameterized queries for security

Analysis

SQL joins, aggregations, ranking

Business Questions

Geographic Analysis

Unique customer locations

Payment Behavior

Installment percentage

Temporal Trends

Monthly order patterns

Seller Metrics

Top revenue performers

Category Performance

Highest revenue products

Growth Analysis

Year-over-year trends

Key Insights

Geographic Distribution

Wide customer reach across multiple locations

Revenue Concentration

Uneven contribution across product categories

Seasonal Patterns

Monthly order fluctuations observed

Seller Performance

Revenue concentrated among top sellers

Payment Usage

Significant installment payment adoption

Customer Retention

Evidence of repeat purchase behavior

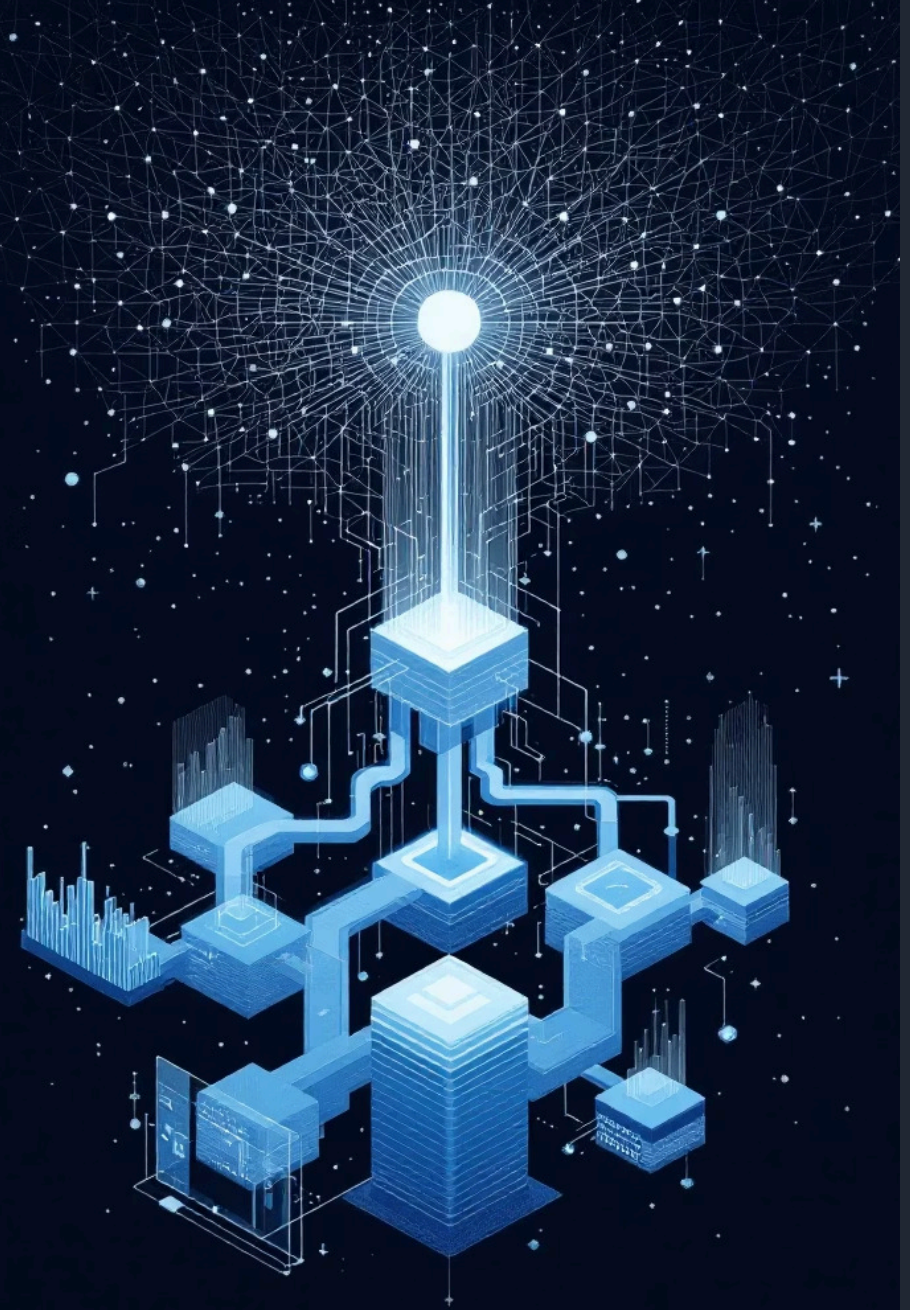
Challenges & Learning

Challenges Faced

- **Missing Values**
  Data quality issues
- **Type Mapping**
  Pandas to MySQL conversion
- **Large Datasets**
  Performance optimization
- **Complex Queries**
  Multi-table SQL operations
- **Relationships**
  No foreign key constraints

Learning Outcomes

- **Automated Pipelines**
  End-to-end data workflows
- **Data Preprocessing**
  Cleaning and validation
- **Database Design**
  Relational structure principles
- **SQL Analytics**
  Advanced querying techniques
- **Business Translation**
  Problem to solution mapping



Conclusion

Raw transactional data transformed into structured insights using Python and SQL

Organized Workflows

Systematic approach to data processing

Analytical Thinking

Business problem translation and solution

Technical Skills

Python, Pandas, MySQL, and SQL mastery

Future Improvements

Interactive Dashboard

Real-time visualization interface

Cloud Deployment

Scalable infrastructure

Query Optimization

Performance enhancement

Automated Reporting

Scheduled insights delivery

Visualization Tools

Power BI or Tableau integration