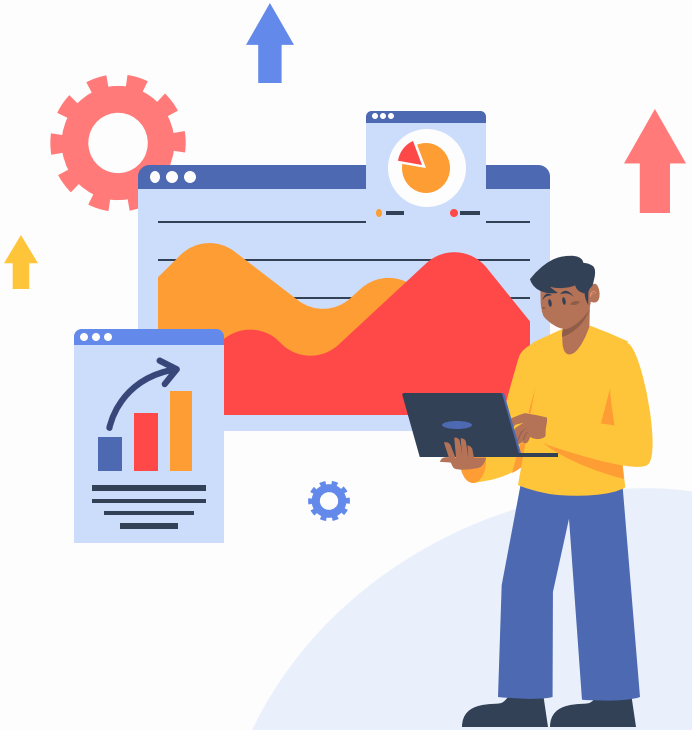
End-to-End Data Pipeline for Business Process Analysis

Kacper Talaga



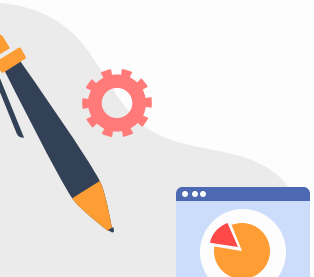
Project Overview

01 Developed an end-to-end data pipeline using Python & PostgreSQL

02 Integrated supermarket sales data for process mining & analysis

03 Extracted key business insights and visualized patterns

04 Simulated ERP data formatting for integration



Dataset Overview

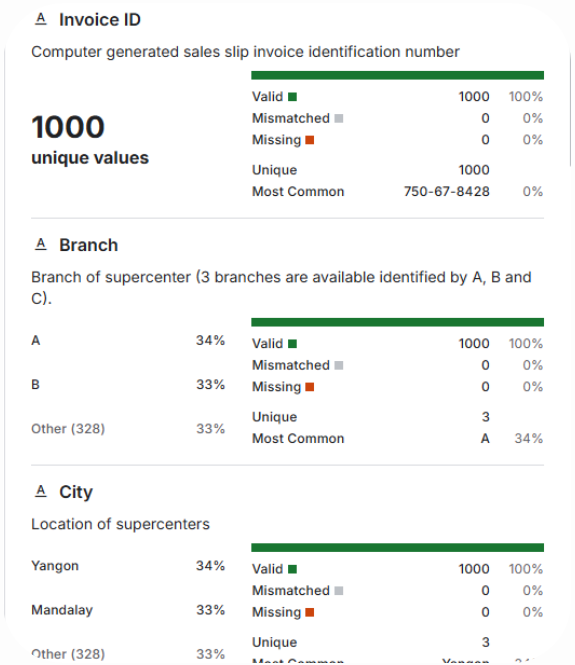
Dataset: Supermarket Sales from Kaggle
Total Recods: 1,000+ transactions
Key Attributes: Invoice ID, Branch, Date, Payment, Revenue
Used for transaction analysis & process mining

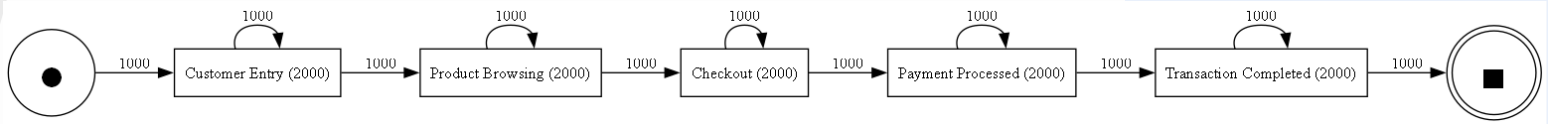
01

ETL Pipeline

ETL Pipeline

- Extracted data from CSV & loaded into PostgreSQL
- Cleaned & standarized date/time formats
- Ensured schema consistency with database structure
- Performed SQL queries for initial insights





02

Process Mining

Process Mining

Generated event logs based on transactions

Identified bottlenecks and inefficiencies

Used PM4PY to analyze customer transaction flows

Visualized customer journey with process flow diagrams

03

ERP Integration & Reporting

ERP Integration & Reporting

Transformed data into ERP-compatible format(SAP structure)

Generated reports on sales trends & customer segments

Automated data exports for business intelligence tools

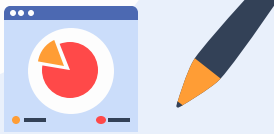
Insights & Visualizations

Analyzed customer behavior(peak hours, preferred payment methods

Created revenue distribution & sales trends using Matplotlib & Seaborn

Found correlations between customer type and purchase patterns

Identified process efficiency gaps



Supermarket Sales Summary Dashboard