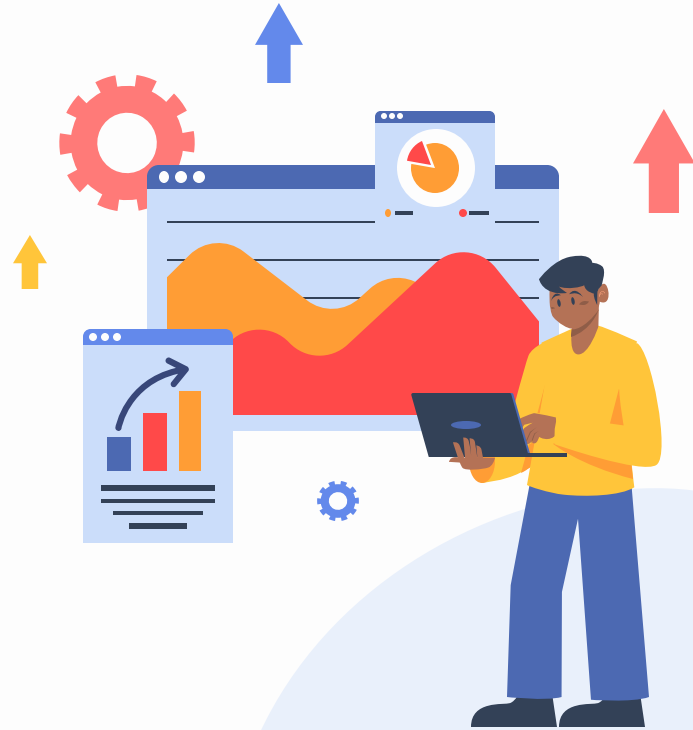


End-to-End Data Pipeline for Business Process Analysis

Kacper Talaga



Project Overview

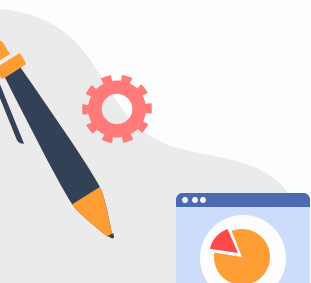


01 Developed an end-to-end data pipeline using Python & PostgreSQL

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 & standardized date/time formats
- Ensured schema consistency with database structure
- Performed SQL queries for initial insights



Invoice ID

Computer generated sales slip invoice identification number

1000
unique values

Valid	1000	100%
Mismatched	0	0%
Missing	0	0%
Unique	1000	
Most Common	750-67-8428	0%

Branch

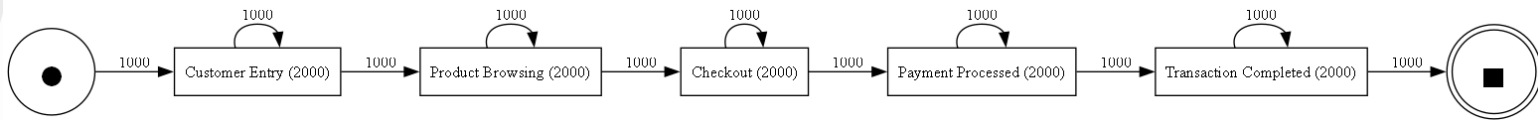
Branch of supercenter (3 branches are available identified by A, B and C).

A	34%	Valid	1000	100%
		Mismatched	0	0%
B	33%	Missing	0	0%
Other (328)	33%	Unique	3	
		Most Common	A	34%

City

Location of supercenters

Yangon	34%	Valid	1000	100%
		Mismatched	0	0%
Mandalay	33%	Missing	0	0%
Other (328)	33%	Unique	3	
		Most Common	Yangon	34%

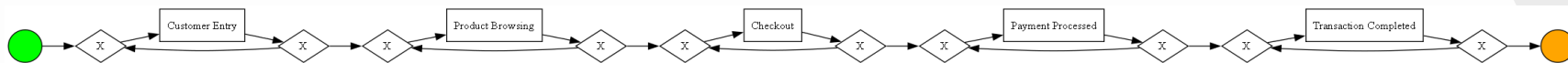


02

Process Mining



Process Mining



**Generated event logs
based on transactions**



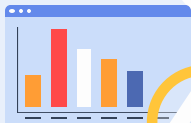
**Used PM4PY to analyze
customer transaction
flows**



**Identified bottlenecks
and inefficiencies**



**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**





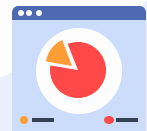
Insights & Visualizations

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

