

BUILDING A MODERN ELT PIPELINE

Report - Assignment #1

Submitted by Mahnoor Imran (25280082)

This project focuses on building a modern ELT pipeline within the Financial Technology domain, specifically centered on digital payments. The thematic goal was to understand how transaction behavior, fraud patterns, financial market performance and fintech related news can be integrated into a unified data engineering workflow. Instead of working with a single dataset, the pipeline was designed to combine structured transactional data, semi-structured news data and time series stock market data. This approach reflects how actual financial systems operate, where insights often come from multiple interconnected data sources rather than one isolated table.

Through the transaction dataset, clear behavioral patterns emerged. Payment and cash-out transactions formed the majority of activity, highlighting the dominant use cases in digital financial platforms. Fraudulent transactions, although fewer in number, showed clustering around particular transaction types and amounts, suggesting that fraud is often systematic rather than random. The correlation analysis further confirmed strong relationships between account balance variables, which aligns with how digital transaction systems update balances deterministically. These findings reinforce the importance of structured feature engineering when preparing financial data for fraud detection models.

The stock market time series analysis of major digital payment companies demonstrated an overall upward trend across recent years, reflecting the rapid expansion and adoption of fintech services globally. This growth trend suggests strong investor confidence in digital payment platforms and provides a foundation for forecasting applications. Additionally, incorporating fintech related news articles introduces a qualitative dimension to the pipeline, enabling future sentiment analysis or topic modeling that could be linked with market performance or transaction trends.

Several challenges were encountered during the development of the pipeline. Managing API authentication securely required careful handling of environment variables rather than hard-coding keys. Data from different sources arrived in different formats (CSV, JSON), requiring consistent cleaning and standardization. Handling missing values in time series data, ensuring correct date formatting and maintaining modular code structure were also important considerations. Additionally, ensuring that the pipeline remained reproducible and organized into raw, processed and cleaned directories required disciplined project structuring.

Overall, this project demonstrates how an ELT pipeline can integrate heterogeneous financial datasets into a scalable architecture suitable for AI applications. The resulting engineered datasets provide a strong foundation for fraud detection systems, financial forecasting models and sentiment driven market analysis, illustrating the practical intersection of data engineering and artificial intelligence within the digital payments ecosystem.