

FinTech ELT Pipeline (Report)

Thematic Focus

The assignment area is the Financial Technology (FinTech) domain and in particular, the cryptocurrencies, the UPI transactions, and the stock prices of the key FinTech companies. The objective was to create a modular ELT system that gets the information out of various sources, cleans and processes it and stores it in an industry-friendly format and provides exploratory visualizations to interpret trends and patterns in the FinTech ecosystem.

Key Findings

1. Cryptocurrency Market

Bitcoin controls the market with a 63 percent of the total market capitalization, and Ethereum and Tether occupy the 11 and 8 percent respectively. Other leading coins such as XRP have less than 4 contribution. This demonstrates that the cryptocurrency industry is extremely concentrated, and a number of large coins affect the market dynamics.

2. UPI Transactions

The volume of UPI transactions is dominated by PhonePe (when compared to Google Pay and Paytm Payment Bank). Trends was calculated monthly and indicated that there is an increment in the volume of transactions, April and May have downward trend that are observable and indicate the seasonal impacts on digital payments.

3. FinTech Stock Prices

Visa and MasterCard are performing well in terms of strong positive relationships (0.99) meaning that they are performing equally. The trend of PayPal is negative, and it has a negative correlation with Visa and MasterCard, implying a different behavior of the market with these companies.

Challenges Encountered :

Data Extraction: Google Trends API was rate limited which meant that it had to be timed. Cryptocurrency API had missing values in roi and max supply which needed to be imputed. There were some missing days in stock price data where non-trading holidays had taken place and the data had to be treated gracefully.

Data Heterogeneity: The datasets consisted of structured data, semi-structured and time series data (stock prices). Preprocessing was important to align these disparate formats.

Storage Choices: It was necessary to store CSV and JSON since it was necessary to analyze these two formats. CSV enabled easy human readable summaries and fast visualizations whereas JSON stored the nested structures as presented by the API to be used later in advanced analysis.