

CloudWalk Data Agent

Natural Language Interface for Transaction Analytics

Requirements



- Build a natural language interface for querying transaction data.



- Calculate and display key business KPIs (e.g., Total Payment Volume, Average Ticket).



- Enable multi-dimensional data segmentation by entity, product, and payment method.



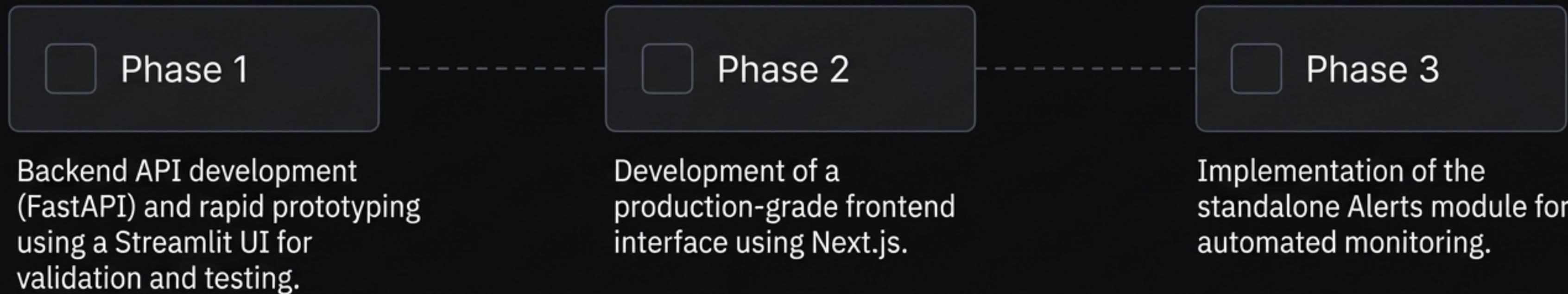
- Generate automated visualizations (charts) based on query results.



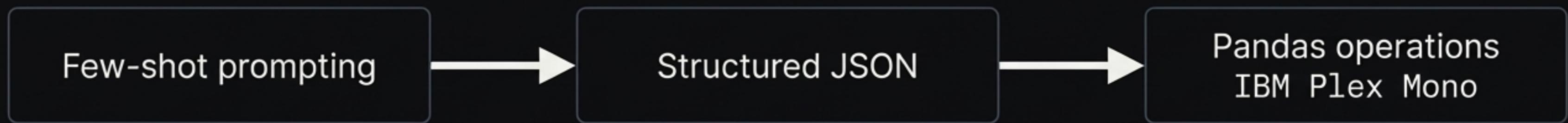
- Implement a system for daily alerts with anomaly detection.

Methodology

Development Phases

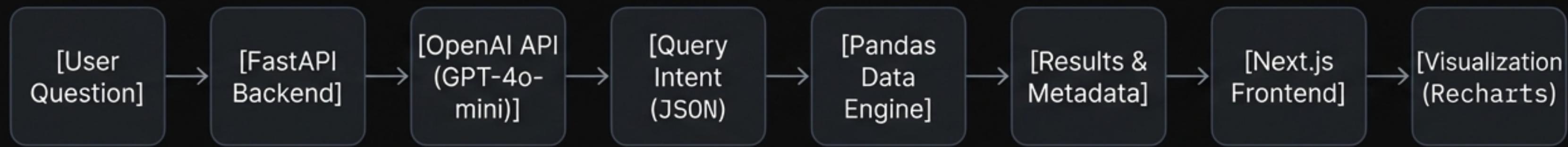


Core Technical Approach



- This approach avoids the complexity of larger frameworks like LangChain/LlamaIndex while ensuring reliable, structured output from the LLM.

Architecture



Backend (FastAPI, Python)

Handles user requests, communicates with the OpenAI API, and executes data operations.

Uses Pandas for in-memory data processing from a CSV source.

Frontend (Next.js, React)

Provides the user interface for inputting queries and renders the returned data and associated charts.

Data Layer

In-memory data processing via Pandas, loaded from a local CSV file. This decision prioritizes simplicity and performance for this application's scope, avoiding database setup overhead.

Features Delivered

- **Natural Language Query Interface:** Users can ask questions in plain English to analyze data.
- **KPI Calculations:** On-demand calculation of key metrics including TPV, Average Ticket, Transaction Counts, and Merchant Counts.
- **Multi-dimensional Segmentation:** Data can be sliced by entity, product, payment method, price tier, and anticipation method.
- **Automatic Chart Generation:** The system intelligently selects and renders appropriate bar or line charts to visualize results.
- **Query Transparency:** Each result is accompanied by a ‘Reasoning Summary’ explaining the LLM's interpretation and the exact query intent (JSON) it generated.
- **Daily Alerts:** The dashboard includes automated alerts with D-1, D-7, and D-30 comparisons.

Example Queries

Which product has the highest TPV?

How do weekdays influence TPV?

I want the TPV from the last 3 days of the pos product

Which segment has the highest Average Ticket?

What is the most used anticipation method by individuals?

Compare price tier performance by TPV

Query Input

Which product has the lowest TPV?

Results

Reasoning Summary

PRODUCT

METRIC VALUE

bank_slip

R\$ 42.039.460,51

How do weekdays influence TPV?

Results

Reasoning Summary

TPV

40V

30V

20V

10V

0

Mon

Tue

Wed

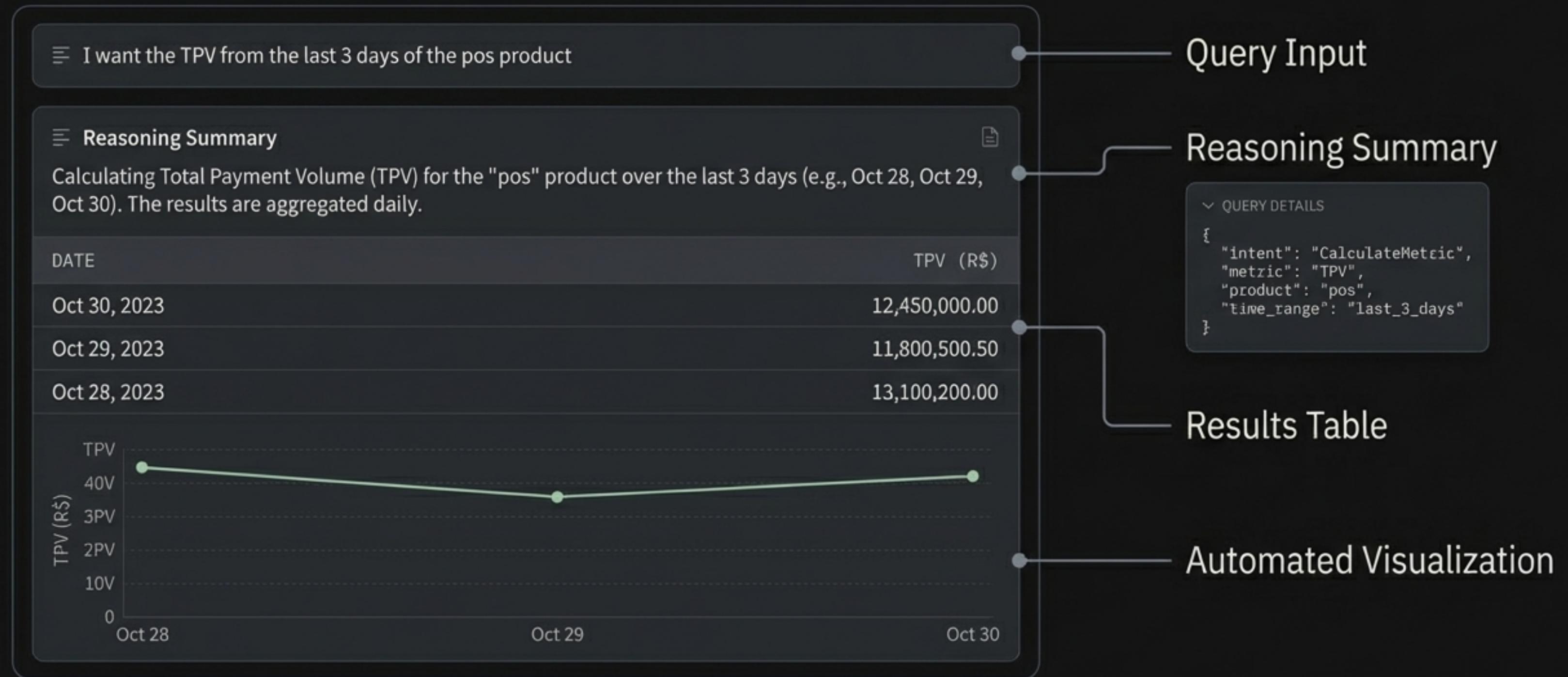
Thu

Fri

Sat

Sun

User Interface



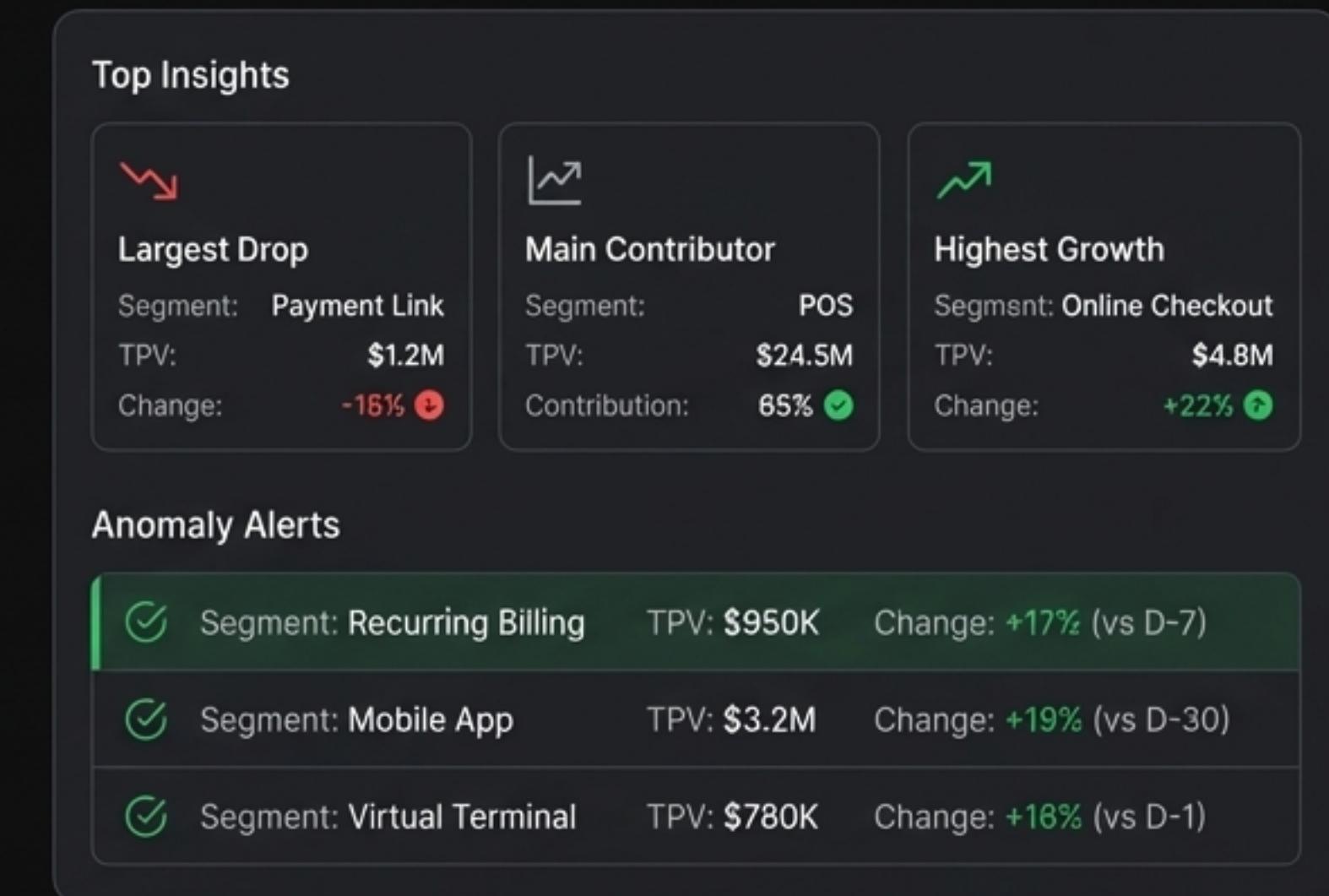
Alerts Module

Automated Monitoring & Anomaly Detection

- Provides a daily summary of key metrics with percentage variations against D-1, D-7, and D-30 baselines.
- Uses a Z-score statistical method (threshold: $\pm 2\sigma$ or $\pm 15\%$) to automatically detect significant anomalies in **TPV** and **Average Ticket** across key business segments.

Top Insights

- Automatically surfaces the most important daily changes, categorized as:
 - Largest Drop:** The segment with the most significant negative change.
 - Main Contributor:** The segment with the highest absolute TPV.
 - Highest Growth:** The segment with the most significant positive change.

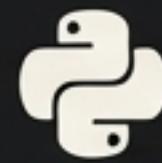


Technology Stack

Backend



Framework
FastAPI



Language
Python 3.12+



Data Operations
Pandas



AI/LLM
OpenAI API (GPT-4o-mini)



Validation
Pydantic

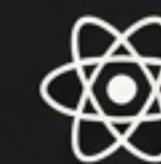
Frontend



Framework
Next.js 15 (App Router)



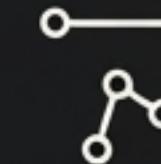
Language
TypeScript



UI Library
React 19



Styling
Tailwind CSS



Visualizations
Recharts

Delivered

- ✓ Natural language query interface
- ✓ Business KPI calculations
- ✓ Multi-dimensional segmentation
- ✓ Automated visualizations
- ✓ Daily alerts with anomaly detection
- ✓ Production-ready implementation