



AI-Powered E-commerce Decision System

Lakehouse + Machine Learning + Generative AI on Databricks

Presented by Saniya Noori
Databricks 14-Day AI Challenge

The Business Challenge: Why AI is Essential

Current Pain Points

E-commerce businesses face overwhelming data complexity that traditional systems cannot effectively address. Raw data arrives messy and duplicated across multiple sources, creating inconsistent customer views and operational blind spots.

Without visibility into high-value customers, businesses struggle to optimise marketing spend. Late deliveries erode customer trust, yet predicting delays remains challenging. Product reviews contain valuable feedback trapped in unstructured text, while executives lack time to sift through complex dashboards for critical decisions.

Why AI is the Answer

Customer behaviour follows non-linear patterns that simple rules cannot capture. Delivery delays depend on dozens of interconnected factors—from weather to warehouse capacity. Text data demands natural language processing to extract meaning at scale.

Leaders need concise, actionable summaries, not raw charts. AI bridges the gap between complex data and business decisions, transforming raw information into strategic advantage through automated insights and predictive intelligence.

Our Solution: End-to-End AI System on Databricks

A comprehensive framework that transforms raw e-commerce data into executive-ready decisions using Lakehouse architecture, machine learning, and generative AI—all powered by Databricks.

01

Data Engineering Layer

Built on Lakehouse architecture with Bronze, Silver, and Gold tiers for data quality and business readiness

02

Machine Learning Layer

Three predictive models for customer segmentation, delivery prediction, and sentiment analysis

03

Generative AI Layer

AI Summary Genie converts complex insights into executive-friendly recommendations





Data Engineering: The Bronze-to-Gold Pipeline

A systematic approach to data quality ensures reliable insights at every level. Each tier builds upon the previous, transforming raw data into business-ready intelligence.



Bronze Layer: Raw Ingestion

Raw data ingestion with enforced schema standards. Captures all source data exactly as received, maintaining historical fidelity for audit and reproducibility.



Silver Layer: Quality Assurance

Comprehensive data cleaning, deduplication, and standardisation. Resolves inconsistencies, handles missing values.



Gold Layer: Business Intelligence

Business metrics and ML-ready feature tables. Pre-calculated KPIs, engineered features, and aggregated views designed specifically for machine learning models.

Machine Learning: Three Models, One Purpose

Each model addresses a critical business question with interpretability and actionability in mind. These models work together to provide a complete view of e-commerce performance.



RFM Customer Segmentation

Technique: Clustering algorithm

Purpose: Identifies high-value customers

Output: Customer tiers by recency, frequency, monetary value

Enables targeted marketing campaigns and loyalty programmes based on actual purchasing behaviour rather than assumptions.



Delivery Delay Prediction

Technique: Classification model

Purpose: Flags at-risk orders

Output: Probability of late delivery

Proactive notifications and resource allocation prevent customer dissatisfaction before it occurs, improving NPS scores.



Review Sentiment Analysis

Technique: NLP pipeline

Purpose: Extracts feedback themes

Output: Sentiment scores and topic categories

Identifies product issues, service gaps, and improvement opportunities from thousands of reviews automatically.

Generative AI: The AI Summary Genie

Intelligent Insights at Executive Speed

Built using Mosaic AI SQL, the AI Summary Genie reads curated Gold tables and generates natural language summaries that translate complex data patterns into clear, actionable recommendations.

This layer bridges the gap between data scientists and business leaders. Rather than forcing executives to interpret charts and metrics, the system delivers concise narratives that highlight what matters most: revenue opportunities, operational risks, and customer experience improvements.

The genie learns from feedback, continuously improving summary quality and relevance. It identifies anomalies, trends, and correlations across all three ML models, presenting a unified story rather than isolated metrics.

Key Capabilities

- Automatic insight generation
- Executive-friendly language
- Actionable recommendations
- Continuous learning
- Multilingual support

Business Impact: From Data to Decisions

The system reveals critical insights that drive immediate business action across four key stakeholder groups.

Revenue Driven by Volume, Not Loyalty

Most revenue comes from one-time purchases rather than repeat customers. Marketing should focus on acquisition while developing retention programmes for identified high-value segments.

8% Late Delivery Rate

Delivery delays affect nearly one in twelve orders, primarily during peak seasons. Operations needs to invest in capacity planning and alternative logistics partners for high-risk routes.

Low Basket Size (1.14 Items)

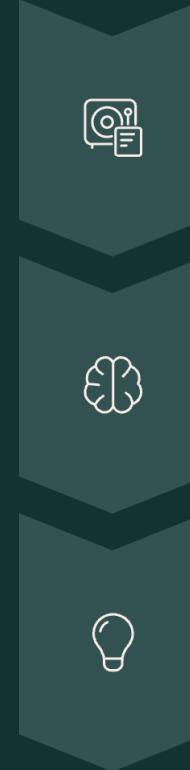
Average order contains just over one item, indicating missed cross-sell and upsell opportunities. Product page optimisation and recommendation engine improvements could significantly boost AOV.

Moderate Customer Sentiment

Overall sentiment is neutral-to-positive, but negative reviews highlight specific product quality and customer service issues. CX team should prioritise these pain points for immediate resolution.

Transforming Raw Data into Strategic Advantage

This end-to-end system demonstrates how Databricks enables organisations to move from data chaos to confident decision-making at scale.



Scalable Data Foundation

Lakehouse architecture handles terabytes of e-commerce data with consistent quality across Bronze, Silver, and Gold tiers

Explainable AI Models

Three interpretable models provide clear rationale for predictions, building trust and enabling business action

Executive-Ready Insights

Generative AI translates complex patterns into natural language recommendations that drive immediate business value

The result: A complete, production-ready AI system that delivers measurable business impact from day one—proving that advanced analytics and artificial intelligence are accessible to organisations of any size when built on the right platform.



THANK YOU