ShopSense

ITCS 6190 – Cloud Computing for Data Analysis
Team Members:
    Deepashree Srinivasa
    Vasudha Maganti
    Sai Tarun Vedagiri
    Manish Yendeti
    Sasi Vadana Atluri
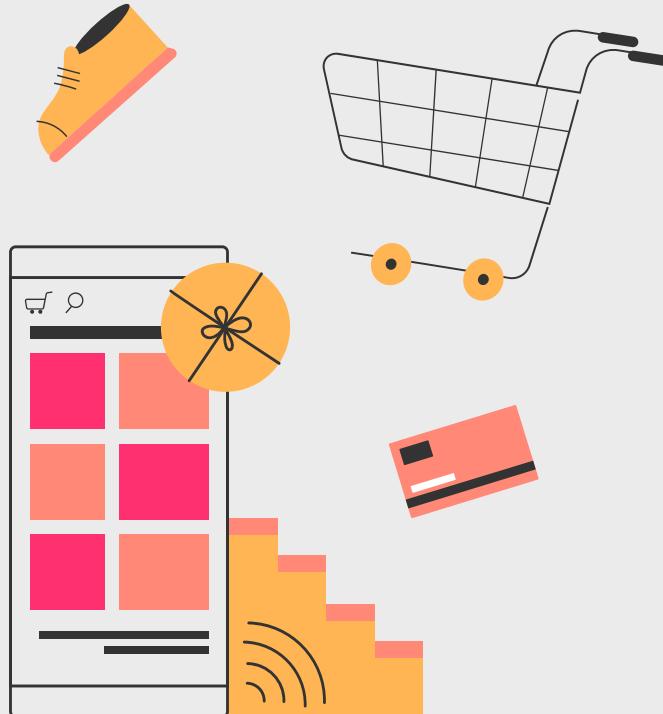
Group 8

About the Product

ShopSense transforms customer data into observable patterns using a unified Apache Spark pipeline that combines robust ETL with real-time streaming. All insights are delivered through an interactive Gradio web interface, bridging the gap between complex big data infrastructure and real-time business decision-making.
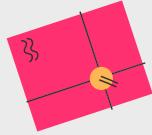
Table of contents

01 About

02 Key Insights

03 The "Brain"
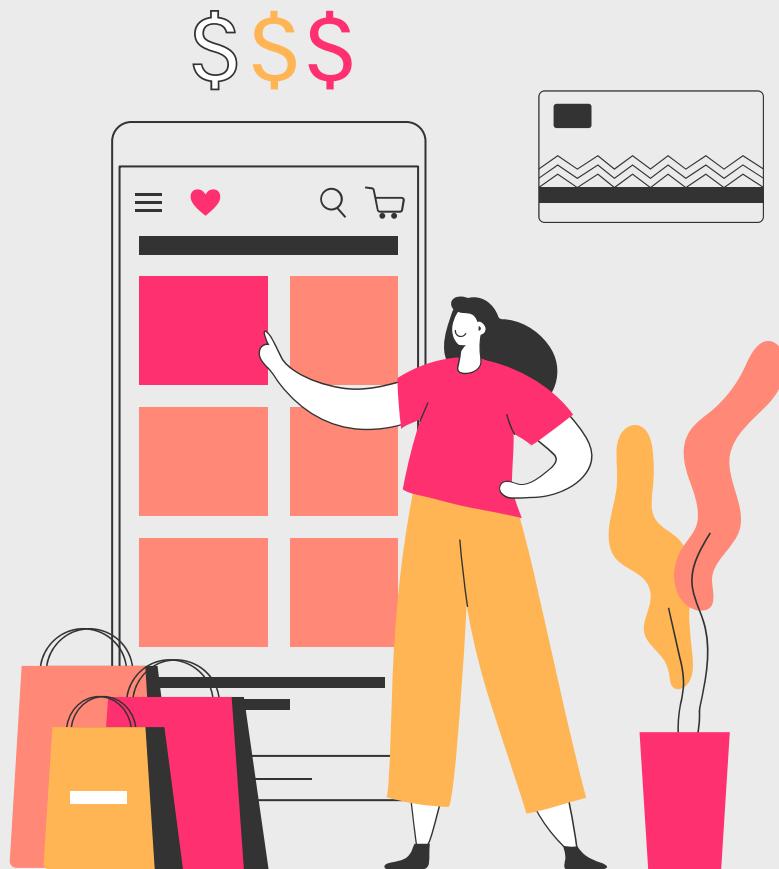
04 Conclusion

Problem Statement

Retailers today are "data-rich but insight-poor." They possess massive volumes of raw transaction logs but suffer from a Data-Action Gap:

- Latency
- Generic Marketing
- Dirty Data

Why Is This Product Unique?

It Watches "Live" & "Past" Simultaneously: Unlike standard tools that only look at historical reports, ShopSense monitors transactions the moment they happen while still analyzing past trends.

It Cleans Its Own Data: The system automatically detects errors, removes duplicates, and filters out useless information before you even see it.

Three AI "Brains" in One: Instead of doing just one thing, it runs three different predictive engines at once: it guesses *what* a customer wants to buy, *if* they will use a coupon, and *which* seasonal items they need.

No Coding Required for Users: It wraps all this complex math into a simple, easy-to-use web dashboard where store managers can get answers just by clicking buttons.

Data Engineering & Quality

Tech Stack:
- Apache Spark Structured API.

Automated Quality Gates:
- Null value detection & duplicate removal.
- Schema enforcement (Integer vs Float precision).
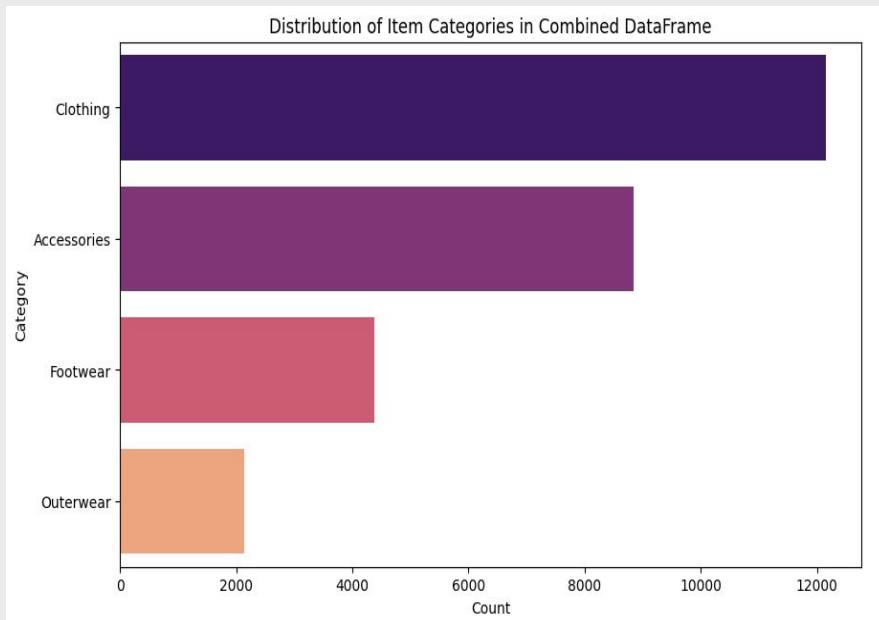
Smart Feature Reduction:
- Automatically drops redundant features to optimize ML model performance.
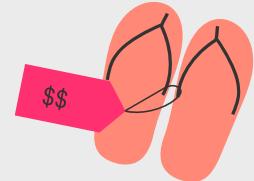
Storage:
- Optimized Parquet format for downstream speed.

Data Engineering & Quality

Distribution of Item Categories in Combined DataFrame

Purchase Distribution by Gender

Deep Dive Analytics

Complex Querying and Key Business Insights:
- Market Basket Analysis: Identifying product affinity (Lift Scores & Confidence).
- Customer Categories: Segmenting users into 'VIP', 'Loyal', and 'At Risk'.
- Seasonal Trends: Pivot-based analysis of category performance across seasons.
- Shipping Optimization: Correlating shipping types with customer satisfaction ratings.

Complex insights

Location	unique_customers	total_transactions	transactions_per_customer	total_revenue	avg_transaction	avg_customer_age	avg_rating	subscription_rate	male_pct	female_pct	avg_loyalty	top_category	revenue_market_share
Idaho	93	691	7.43	40838	59.1	42.8	3.76	19.54	63.24	36.76	25.4	Clothing	2.48
Nevada	87	665	7.64	40054	60.23	42.8	3.77	34.74	69.77	30.23	26.0	Clothing	2.43
Illinois	92	634	6.89	38556	60.81	44.0	3.76	23.50	67.35	32.65	25.3	Clothing	2.34
California	95	650	6.84	38402	59.08	42.7	3.76	27.38	67.23	32.77	25.3	Clothing	2.33
New York	87	622	7.15	38162	61.35	44.1	3.74	24.44	70.58	29.42	24.5	Clothing	2.32
Maryland	86	629	7.31	37476	59.58	46.4	3.79	20.83	68.04	31.96	26.5	Clothing	2.28
Montana	96	628	6.54	37373	59.51	44.9	3.75	27.23	60.19	39.81	25.8	Clothing	2.27
Nebraska	87	611	7.02	36764	60.17	42.5	3.74	28.81	67.59	32.41	25.0	Clothing	2.23
Georgia	79	598	7.57	36364	60.81	43.4	3.73	27.93	60.70	39.30	24.1	Clothing	2.21
Louisiana	84	599	7.13	35981	60.07	48.6	3.73	31.22	67.78	32.22	25.5	Clothing	2.19
Minnesota	88	616	7.00	35920	58.31	46.6	3.73	28.08	68.99	31.01	24.9	Clothing	2.18
Alabama	89	603	6.78	35857	59.46	43.6	3.73	26.37	69.65	30.35	25.5	Clothing	2.18
Tennessee	77	580	7.53	35688	61.53	44.7	3.79	17.93	69.83	30.17	24.2	Clothing	2.17
Vermont	85	589	6.93	35521	60.31	42.5	3.78	25.81	74.36	25.64	25.1	Clothing	2.16
Delaware	86	609	7.08	35210	57.82	43.1	3.79	34.48	74.38	25.62	25.2	Clothing	2.14
Arkansas	79	594	7.52	35196	59.25	43.6	3.73	24.24	69.02	30.98	25.6	Clothing	2.14
North Carolina	78	579	7.42	35054	60.54	46.5	3.74	28.67	74.27	25.73	24.6	Clothing	2.13
Virginia	77	581	7.55	34787	59.87	41.1	3.73	29.43	66.95	33.05	25.0	Clothing	2.11
Kentucky	79	571	7.23	34538	60.49	43.4	3.77	35.55	61.47	38.53	25.6	Clothing	2.10
New Mexico	81	570	7.04	34455	60.45	43.6	3.72	28.95	74.56	25.44	26.5	Clothing	2.09

Complex insights

customer_segment	customer_count	avg_lifetime_value	avg_transactions	avg_transaction_size	avg_age	avg_loyalty_score	avg_rating	subscriber_rate	segment_total_value	pct_of_total_revenue
VIP Champions	6172	102.49	1.3	85.95	44.0	25.3	3.76	26.46	632559	38.42
High Spenders	251	84.26	1.0	84.26	44.8	25.4	3.75	25.90	21149	1.28
Frequent Buyers	251	68.67	2.0	34.34	46.0	24.9	3.77	25.90	17237	1.05
Regular Customers	9486	62.18	1.0	62.18	44.0	31.5	3.74	27.80	589876	35.82
Loyal Customers	4656	51.67	1.0	51.67	44.0	6.4	3.76	27.28	240568	14.61
At Risk	4879	29.76	1.0	29.76	44.4	31.5	3.75	27.96	145221	8.82

age_group	Category	purchase_count	total_revenue	avg_purchase	avg_rating	percentage_of_age_group	popularity_rank
18-24 (Gen Z)	Clothing	1522	91526	60.14	3.76	44.70	1
18-24 (Gen Z)	Accessories	1077	65460	60.78	3.78	31.63	2
18-24 (Gen Z)	Footwear	514	31034	60.38	3.74	15.10	3
25-34 (Millennials)	Clothing	2354	142226	60.42	3.75	43.91	1
25-34 (Millennials)	Accessories	1729	102490	59.28	3.74	32.25	2
25-34 (Millennials)	Footwear	871	52173	59.9	3.76	16.25	3
35-44 (Gen X)	Clothing	2305	138032	59.88	3.76	45.01	1
35-44 (Gen X)	Accessories	1632	97773	59.91	3.74	31.87	2
35-44 (Gen X)	Footwear	800	46586	58.23	3.77	15.62	3
45-54 (Gen X)	Clothing	2315	138259	59.72	3.73	43.39	1
45-54 (Gen X)	Accessories	1714	103167	60.19	3.76	32.13	2
45-54 (Gen X)	Footwear	890	52647	59.15	3.8	16.68	3
55-64 (Boomers)	Clothing	2300	138267	60.12	3.74	43.80	1
55-64 (Boomers)	Accessories	1698	102184	60.18	3.78	32.34	2
55-64 (Boomers)	Footwear	831	48579	58.46	3.75	15.83	3
65+ (Seniors)	Clothing	1366	79468	58.18	3.77	44.64	1
65+ (Seniors)	Accessories	999	60766	60.83	3.72	32.65	2
65+ (Seniors)	Footwear	479	27990	58.43	3.72	15.65	3

The "Brain": Here are three important models
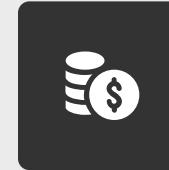
Recommendation

Recommend products users have not bought yet but are likely to buy.

Seasonal prediction

Predict specific items a user wants based on the Season.

Promo Code Model

Predict if a user will actually use a coupon code (saves marketing spend).

Recommendation
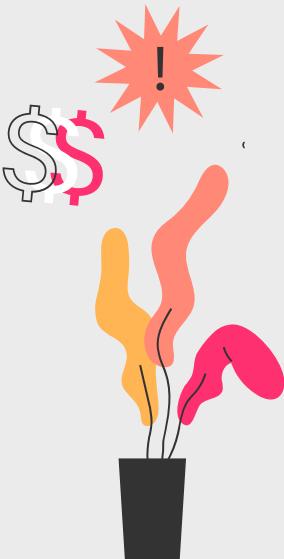
ALS (Alternating Least Squares) – Collaborative Filtering

Seasonal
Prediction

Random Forest Classifier (Multi-Class Classification)

Promo Code Model

Random Forest Classifier (Binary Classification)

Interactive User Experience

🛒 **ShopSense**

🛍️ Seasonal Prediction　　⭐ Recommendations　　🎟️ Promo Propensity

Customer ID

108

Season

Winter ▾

Predict

Result

🎯 Predictions for User 108 in Winter:
----
1. Sneakers　　(6.8%)
2. Pants　　(4.5%)
3. Sunglasses　　(4.3%)

Use via API 🚀 · Built with Gradio 🧡 · Settings ⚙️

Interactive User Experience

Interactive User Experience

Conclusion

ShopSense turns complex customer data into real-time, useful insights. Using Apache Spark and intelligent models for recommendations, seasonal trends, and promotion targeting, it helps businesses understand customers better, plan smarter, and market more effectively. With strong data quality checks and optimized storage, ShopSense delivers fast, accurate, and scalable results that increase revenue and improve the overall customer experience.

Real Business Impact

Key Business Impacts

- Personalized Recommendations → Boosts customer engagement & sales

- Seasonal Prediction → Improves demand planning & inventory control

- Promo Code Model → Increases campaign success & marketing ROI

📈 **Boost Sales**
Personalized recommendations that convert

💰 **Cut Costs**
Eliminate wasteful promo spending

⚡ **Move Faster**
Real-time data-driven decisions

Get Started Today

Free 30-Day Trial

No credit card required. Full access to all three ML models.

Step 1: Demo

Schedule a 20-minute personalized demo with our team.
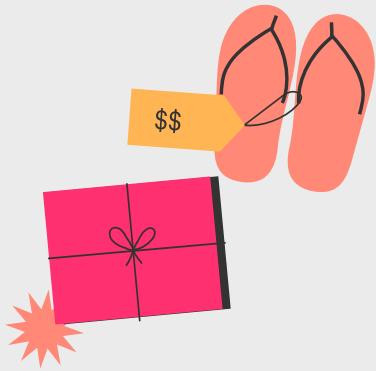
Step 2: Setup

Upload your customer data. We handle data prep and model training.

Step 3: Launch

Start making data-driven decisions via dashboard.

First 50 customers: Lifetime 30% discount + priority support

Thank you!