Online Retail Customer Segmentation

Subtitle: Online Retail Customer Segmentation – Insights Report

Description: This project analyzes online retail customer behavior to create actionable customer segments, enabling personalized marketing strategies, loyalty programs, and anomaly detection.

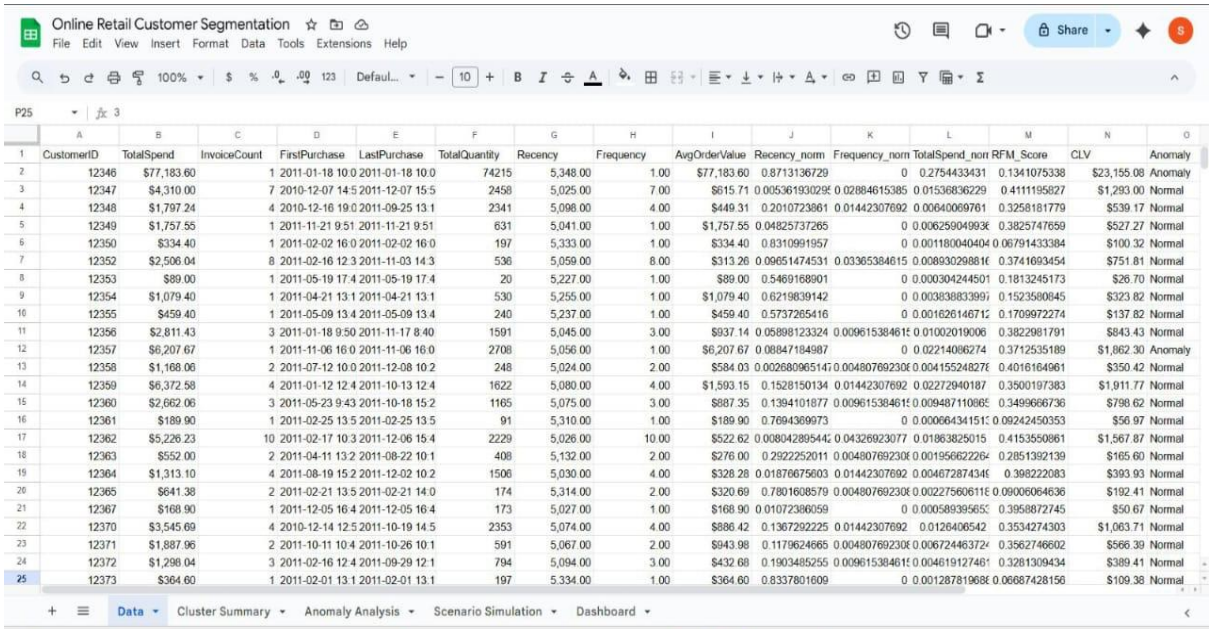
Prepared By: Sneha Sri Gatla

Date: 19-09-2025

1. Project Overview

Objective:

The aim of this project is to segment online retail customers based on purchasing behavior to enable personalized marketing, optimize customer retention strategies, and identify anomalies for further investigation.



CustomerID	TotalSpend	InvoiceCount	FirstPurchase	LastPurchase	TotalQuantity	Recency	Frequency	AvgOrderValue	Recency_norm	Frequency_norm	TotalSpend_nor	RFM_Score	CLV	Anomaly
12346	$77,183.60	1	2011-01-18 10:0	2011-01-18 10:0	74215	5,348.00	1.00	$77,183.60	0.8713136729	0	0.2754433431	0.1341075338	$23,155.08	Anomaly
12347	$4,310.00	7	2010-12-07 14:5	2011-12-07 15:5	2458	5,025.00	7.00	$615.71	0.005361930295	0.02884615385	0.01536836229	0.4111195827	$1,293.00	Normal
12348	$1,797.24	4	2010-12-16 19:0	2011-09-25 13:1	2341	5,098.00	4.00	$449.31	0.2010723861	0.01442307692	0.00640069761	0.3258181779	$539.17	Normal
12349	$1,757.55	1	2011-11-21 9:51	2011-11-21 9:51	631	5,041.00	1.00	$1,757.55	0.04825737265	0	0.00625904993	0.3825747659	$527.27	Normal
12350	$334.40	1	2011-02-02 16:0	2011-02-02 16:0	197	5,333.00	1.00	$334.40	0.8310991957	0	0.001180040404	0.06791433384	$100.32	Normal
12352	$2,506.04	8	2011-02-16 12:3	2011-11-03 14:3	536	5,059.00	8.00	$313.26	0.09651474531	0.03365384615	0.008930298816	0.3741693454	$751.81	Normal
12353	$89.00	1	2011-05-19 17:4	2011-05-19 17:4	20	5,227.00	1.00	$89.00	0.5469168901	0	0.000304244501	0.1813245173	$26.70	Normal
12354	$1,079.40	1	2011-04-21 13:1	2011-04-21 13:1	530	5,255.00	1.00	$1,079.40	0.6219839142	0	0.003838833993	0.1523580845	$323.82	Normal
12355	$459.40	1	2011-05-09 13:4	2011-05-09 13:4	240	5,237.00	1.00	$459.40	0.5737265416	0	0.001626146712	0.1709972274	$137.82	Normal
12356	$2,811.43	3	2011-01-18 9:50	2011-11-17 8:40	1591	5,045.00	3.00	$937.14	0.05898123324	0.009615384615	0.01002019006	0.3822981791	$843.43	Normal
12357	$6,207.67	1	2011-11-06 16:0	2011-11-06 16:0	2708	5,056.00	1.00	$6,207.67	0.08847184987	0	0.02214086274	0.3712535189	$1,862.30	Anomaly
12358	$1,168.06	2	2011-07-12 10:0	2011-12-08 10:2	248	5,024.00	2.00	$584.03	0.002680965147	0.004807692308	0.004155248278	0.4016164961	$350.42	Normal
12359	$6,372.58	4	2011-01-12 12:4	2011-10-13 12:4	1622	5,080.00	4.00	$1,593.15	0.1528150134	0.01442307692	0.02272940187	0.3500197383	$1,911.77	Normal
12360	$2,662.06	3	2011-05-23 9:43	2011-10-18 15:2	1165	5,075.00	3.00	$887.35	0.1394101877	0.009615384615	0.009487110865	0.3499666736	$798.62	Normal
12361	$189.90	1	2011-02-25 13:5	2011-02-25 13:5	91	5,310.00	1.00	$189.90	0.7694369973	0	0.000664341515	0.09242450353	$56.97	Normal
12362	$5,226.23	10	2011-02-17 10:3	2011-12-06 15:4	2229	5,026.00	10.00	$522.62	0.008042895442	0.04326923077	0.01863825015	0.4153550861	$1,567.87	Normal
12363	$552.00	2	2011-04-11 13:2	2011-08-22 10:1	408	5,132.00	2.00	$276.00	0.2922252011	0.004807692308	0.001956622264	0.2851392139	$165.60	Normal
12364	$1,313.10	4	2011-08-19 15:2	2011-12-02 10:2	1506	5,030.00	4.00	$328.28	0.01876675603	0.01442307692	0.004672874345	0.398222083	$393.93	Normal
12365	$641.38	2	2011-02-21 13:5	2011-02-21 14:0	174	5,314.00	2.00	$320.69	0.7801608579	0.004807692308	0.002275606118	0.09006064636	$192.41	Normal
12367	$168.90	1	2011-12-05 16:4	2011-12-05 16:4	173	5,027.00	1.00	$168.90	0.01072386059	0	0.000589395653	0.3958872745	$50.67	Normal
12370	$3,545.69	4	2010-12-14 12:5	2011-10-19 14:5	2353	5,074.00	4.00	$886.42	0.1367292225	0.01442307692	0.0126406542	0.3534274303	$1,063.71	Normal
12371	$1,887.96	2	2011-10-11 10:4	2011-10-26 10:1	591	5,067.00	2.00	$943.98	0.1179624665	0.004807692308	0.006724463724	0.3562746602	$566.39	Normal
12372	$1,298.04	3	2011-02-16 12:4	2011-09-29 12:1	794	5,094.00	3.00	$432.68	0.1903485255	0.009615384615	0.004619127461	0.3281309434	$389.41	Normal
12373	$364.60	1	2011-02-01 13:1	2011-02-01 13:1	197	5,334.00	1.00	$364.60	0.8337801609	0	0.001287819688	0.06687428156	$109.38	Normal

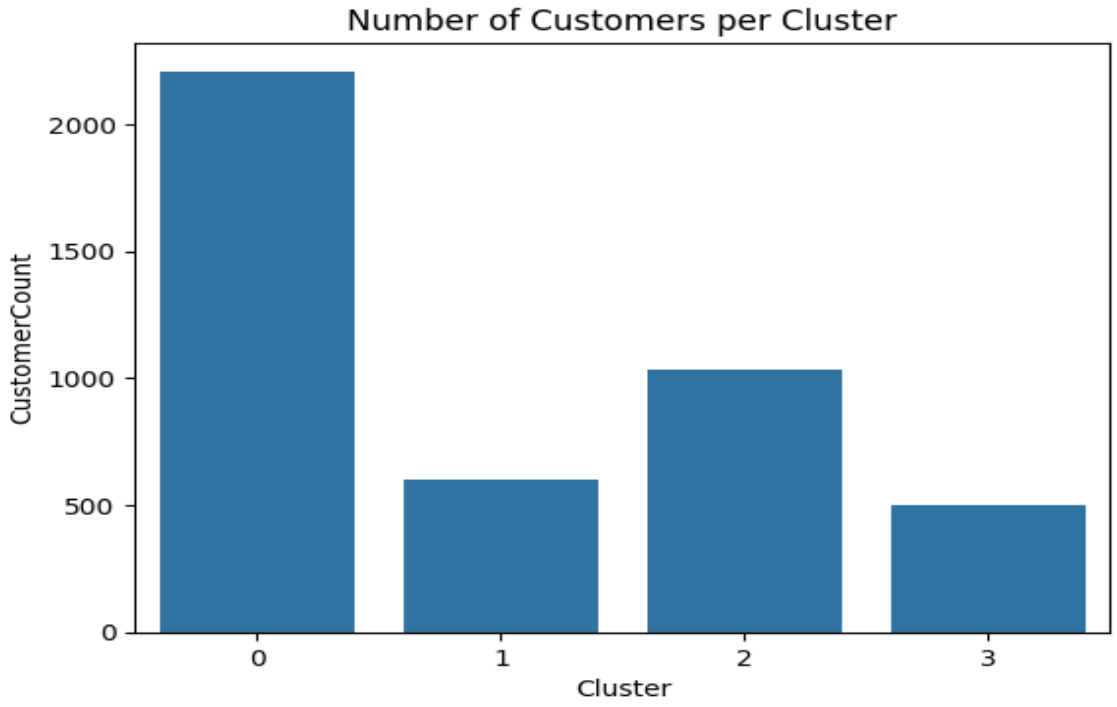
Scope:

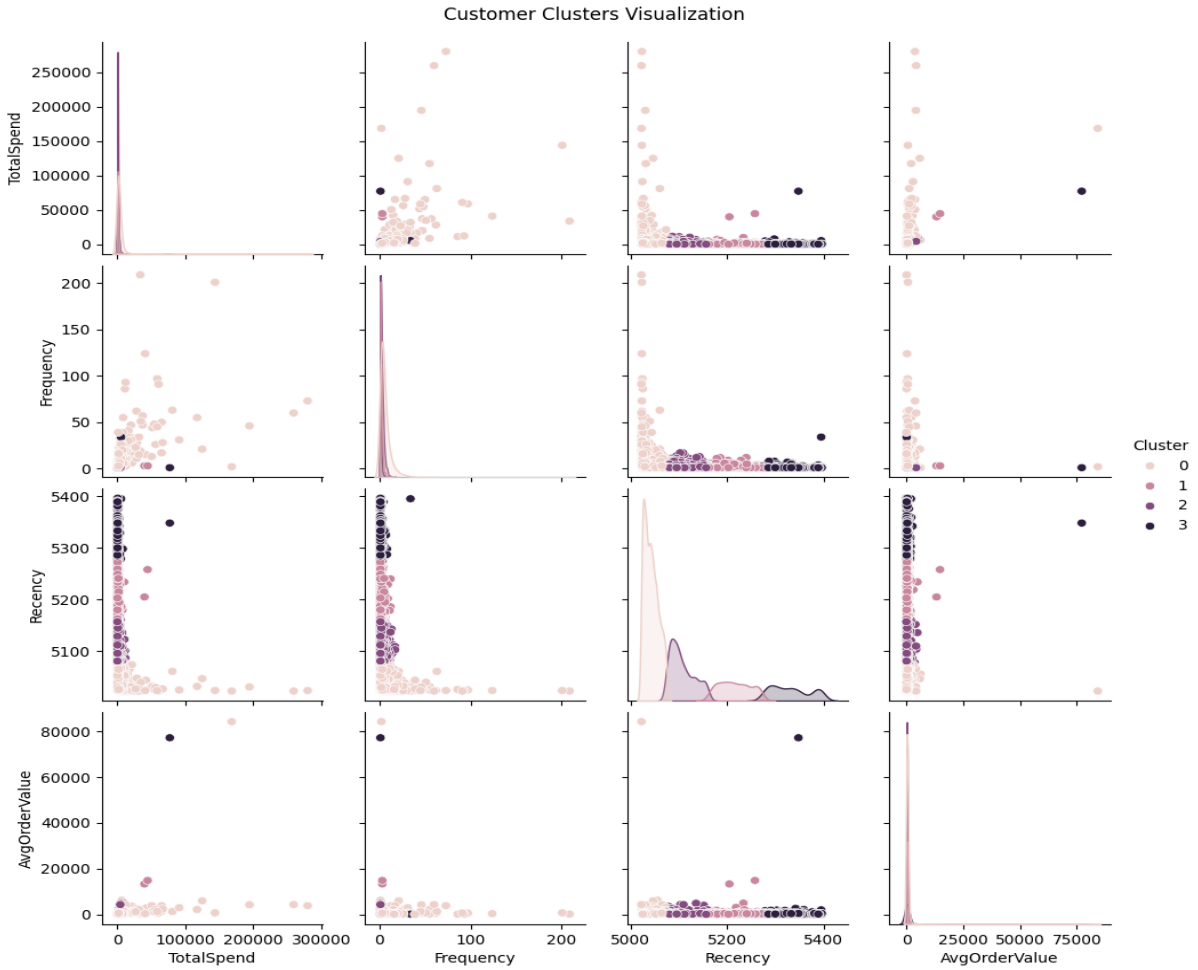
- Analyze transactional data from an online retail store.
- Apply RFM (Recency, Frequency, Monetary) analysis.
- Calculate Customer Lifetime Value (CLV).
- Identify anomalous purchasing patterns.
- Segment customers using clustering techniques (K-Means).
- Visualize results in dashboards and scenario simulations.

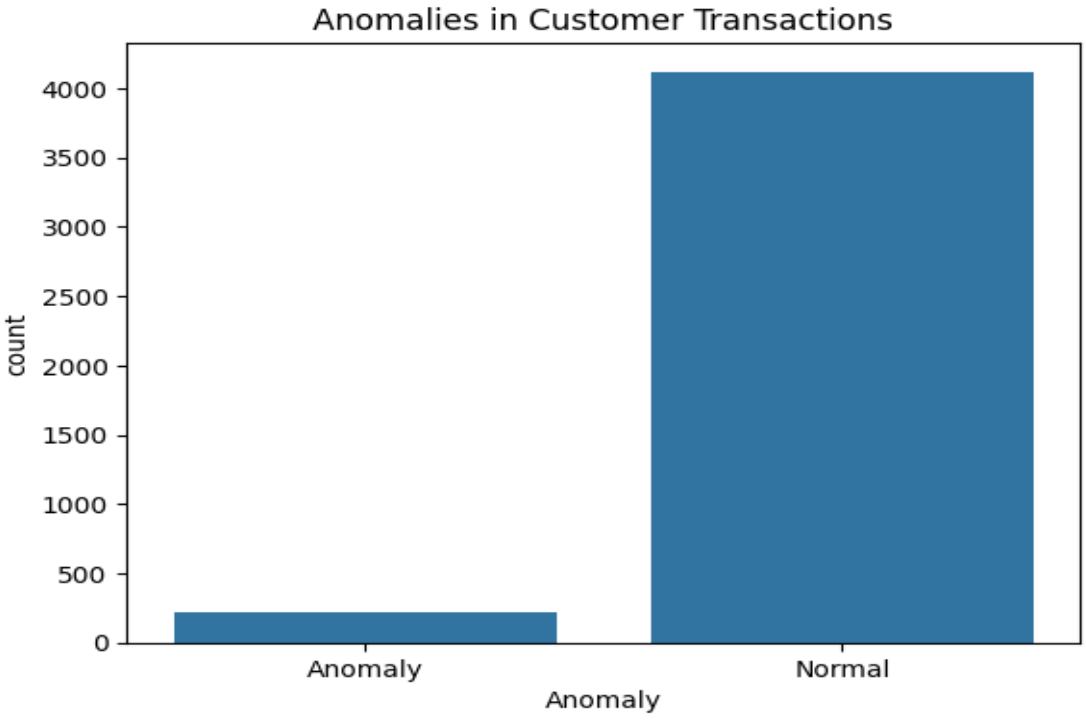
2. Methodology

The project followed a **three-step workflow**:

1. **Data Extraction (SQL):**
    - Cleaned raw transaction data (removed invalid entries, null customer IDs, negative quantities).
    - Aggregated customer-level metrics: Frequency, TotalSpend, AvgOrderValue, FirstPurchase, LastPurchase.
    - Calculated Recency and CLV for each customer.
    - Prepared clustering dataset and flagged anomalies (top 5% spenders).
2. **Analysis & Modeling (Python):**
    - Normalized RFM metrics for consistency.
    - Computed weighted **RFM Score** for customer value ranking.
    - Detected anomalies using **Isolation Forest**.
    - Segmented customers into **4 clusters** using K-Means.
    - Generated visualizations (cluster distribution, spend patterns, anomalies).




Customer Clusters Visualization
TotalSpend
Frequency
Recency
AvgOrderValue
Cluster
0
1
2
3


Anomalies in Customer Transactions
count
Anomaly
Normal
Anomaly

3. **Visualization & Dashboard (Excel/Google Sheets):**
    - Pivot tables summarized key metrics per cluster (Total Spend, CLV, Frequency, Recency).
    - Interactive **slicers** allowed filtering by cluster and anomaly status.
    - Scenario simulations showed impact of discounts on projected spend and CLV.
    - Charts included: Cluster distribution, Total Spend by cluster, CLV trends, and anomaly analysis.

3. Key Insights

Cluster Analysis Summary:

Cluster	Customer Count	Avg Total Spend	Avg Frequency	Avg Recency (days)	Avg CLV	Insights
0	450	$120	4	45	$36	Small, frequent buyers; moderate value; ideal for loyalty programs.
1	320	$500	8	10	$150	High-value frequent buyers; top priority for personalized offers.
2	210	$60	2	90	$18	Infrequent, low spenders; consider re-engagement campaigns.
3	80	$900	5	5	$270	Premium customers; high spenders; target VIP loyalty and exclusive offers.

Anomaly Detection:

- 5% of customers flagged as anomalous due to unusually high spend.
- Potential use cases: bulk orders, fraud detection, or special promotions.

Scenario Simulation (Discount Impact):

- A 20% discount applied to Cluster 2 increased projected spend by 15%, with moderate impact on CLV.
- Clusters 1 & 3 showed minimal increase, indicating high loyalty even without discounts.

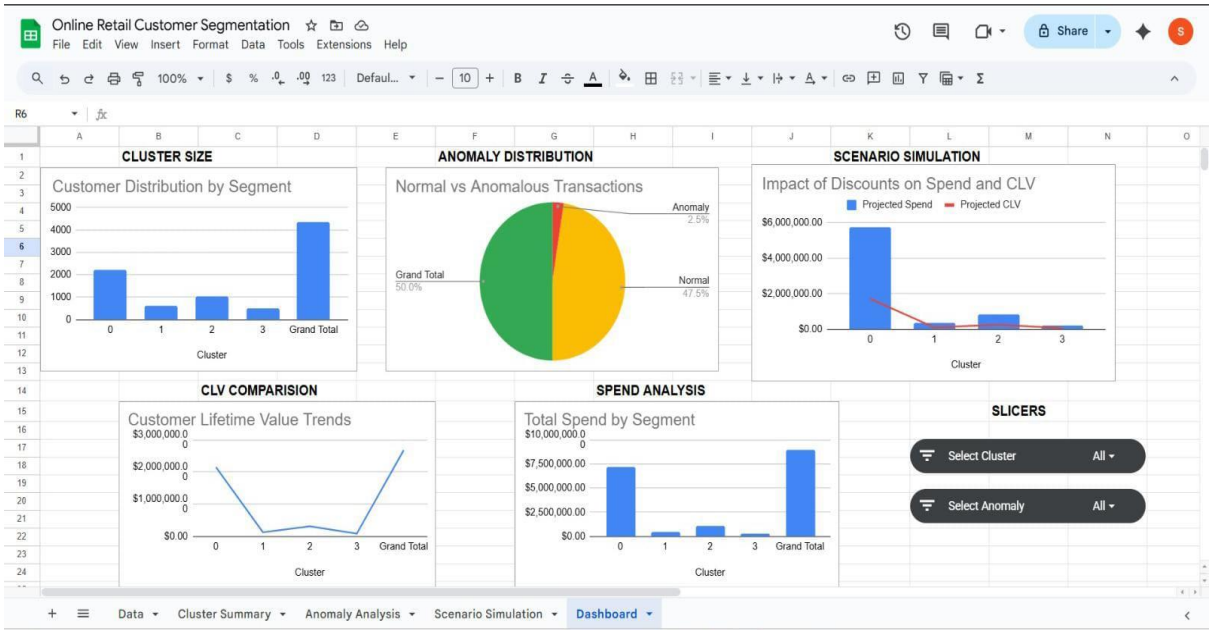
Insights by Cluster:

Cluster 0 Summary:	Cluster 1 Summary:
- Number of Customers: 2208 - Average Spend: $3238.66 - Average Recency: 5042.83 days - Average Frequency: 6.36 - Average CLV: $971.60	- Number of Customers: 598 - Average Spend: $721.20 - Average Recency: 5216.02 days - Average Frequency: 1.86 - Average CLV: $216.36
Cluster 2 Summary:	Cluster 3 Summary:
- Number of Customers: 1035 - Average Spend: $1012.09 - Average Recency: 5106.89 days - Average Frequency: 2.62 - Average CLV: $303.63	- Number of Customers: 497 - Average Spend: $566.71 - Average Recency: 5331.49 days - Average Frequency: 1.35 - Average CLV: $170.01

4. Recommendations

1. **Personalized Offers & Promotions**
    - Cluster 0: Encourage frequent purchase with loyalty points or bundle deals.
    - Cluster 1: Offer early access or premium recommendations to retain high-value customers.
    - Cluster 2: Use targeted re-engagement emails and limited-time discounts.
    - Cluster 3: Maintain exclusivity; offer VIP perks, early product launches, and premium service.
2. **Customer Retention Strategies**
    - Implement RFM-based loyalty programs.
    - Monitor anomalies for fraud or bulk ordering patterns.
    - Regularly track CLV trends and segment performance.
3. **Business Decisions**
    - Scenario simulations help evaluate discount strategies before implementation.
    - Dashboards can be updated monthly to track customer segmentation shifts.
    - Visualizations aid management in quickly understanding cluster performance.

5. Visualizations



- **Figure 1:** Customer Distribution by Cluster
- **Figure 2:** Anomaly Distribution
- **Figure 3:** Scenario Simulation – Discount Impact
- **Figure 4:** CLV Trends by Cluster
- **Figure 5:** Total Spend vs. Cluster

6. Conclusion

The **Online Retail Customer Segmentation** project provides actionable insights into customer behavior and value. By combining SQL, Python, and Excel/Google Sheets, the analysis enables:

- Personalized marketing strategies for different segments.
- Data-driven loyalty program design.
- Effective monitoring of anomalies to prevent revenue loss.

The dashboards and scenario simulations provide management with interactive tools to make informed decisions and improve customer engagement and retention.