Customer Cohort Analysis Using SQL

Customer Retention and CLV Analysis from Transaction Data

Objective: Understand customer behavior over time via cohort analysis.

Key Questions:

How well are we retaining customers?

What is the lifetime value of different customer cohorts(monthly)?

Dataset: 512,909 records from RETAIL table

Data Exploration Highlights

- SELECT COUNT(*) FROM RETAIL; Total Records: **512,909**
- SELECT COUNT(*) FROM RETAIL

  WHERE QUANTITY <= 0; Suspicious Records (Quantity ≤ 0): **10,126**
- SELECT COUNT(*) FROM RETAIL

  WHERE INVOICENO LIKE 'C%'; Canceled Orders (InvoiceNo starts with "C"): **8,836**
- SELECT COUNT(*) FROM RETAIL

  WHERE CUSTOMERID = ' '; Blank CustomerID: **128,676**
- SELECT COUNT(CUSTOMERID) FROM RETAIL

  WHERE CUSTOMERID IS NOT NULL AND CUSTOMERID != ' '

  AND INVOICENO NOT LIKE 'C%' AND QUANTITY > 0

  AND UNITPRICE > 0; Valid Records Used for Cohort: **375,736**

Cohort Analysis Logic

- **CTE1**: Clean and format data
- **CTE2**: Get first transaction date per customer
- **CTE3**: Assign each purchase to a “Cohort Month”
- **Final Step**: Pivot table of customer counts by cohort index (Month_0 to Month_12)

Cohort Analysis [Customer Retention]

```
WITH CTE1 AS
(SELECT
    CUSTOMERID,
    date(str_to_date(INVOICEDATE, '%m/%d/%Y %H:%i')) AS FORMATTED_DATE,
    ROUND(QUANTITY*UNITPRICE, 2) AS SALE_VALUE
FROM RETAIL
WHERE
    CUSTOMERID IS NOT NULL
    AND CUSTOMERID != ''
    AND INVOICENO NOT LIKE 'C%'
    AND QUANTITY > 0
    AND UNITPRICE > 0),
CTE2 AS
(SELECT
    CUSTOMERID,
    FORMATTED_DATE AS PURCHASE_DATE,
    MIN(FORMATTED_DATE) OVER (PARTITION BY CUSTOMERID) AS FIRST_TRANSACTION_DATE
FROM CTE1),
```

Cohort Analysis [Customer Retention]

```
CTE3 AS
(SELECT
    CUSTOMERID,
    FIRST_TRANSACTION_DATE,
    PURCHASE_DATE,
    CONCAT(
        'Month_',
        ROUND(DATEDIFF(PURCHASE_DATE, FIRST_TRANSACTION_DATE)/30, 0)
        ) AS COHORT_MONTH,
    DATE_FORMAT(PURCHASE_DATE, '%Y-%m-01') as PURCHASE_MONTH,
    DATE_FORMAT(FIRST_TRANSACTION_DATE, '%Y-%m-01') AS FIRST_TRANSACTION_MONTH
FROM CTE2)
```

Cohort Analysis [Customer Retention]

```sql
SELECT
    FIRST_TRANSACTION_MONTH AS COHORT,
    COUNT(DISTINCT CASE WHEN COHORT_MONTH = 'Month_0' THEN CUSTOMERID ELSE NULL END) AS "MONTH_0",
    COUNT(DISTINCT CASE WHEN COHORT_MONTH = 'Month_1' THEN CUSTOMERID ELSE NULL END) AS "MONTH_1",
    COUNT(DISTINCT CASE WHEN COHORT_MONTH = 'Month_2' THEN CUSTOMERID ELSE NULL END) AS "MONTH_2",
    COUNT(DISTINCT CASE WHEN COHORT_MONTH = 'Month_3' THEN CUSTOMERID ELSE NULL END) AS "MONTH_3",
    COUNT(DISTINCT CASE WHEN COHORT_MONTH = 'Month_4' THEN CUSTOMERID ELSE NULL END) AS "MONTH_4",
    COUNT(DISTINCT CASE WHEN COHORT_MONTH = 'Month_5' THEN CUSTOMERID ELSE NULL END) AS "MONTH_5",
    COUNT(DISTINCT CASE WHEN COHORT_MONTH = 'Month_6' THEN CUSTOMERID ELSE NULL END) AS "MONTH_6",
    COUNT(DISTINCT CASE WHEN COHORT_MONTH = 'Month_7' THEN CUSTOMERID ELSE NULL END) AS "MONTH_7",
    COUNT(DISTINCT CASE WHEN COHORT_MONTH = 'Month_8' THEN CUSTOMERID ELSE NULL END) AS "MONTH_8",
    COUNT(DISTINCT CASE WHEN COHORT_MONTH = 'Month_9' THEN CUSTOMERID ELSE NULL END) AS "MONTH_9",
    COUNT(DISTINCT CASE WHEN COHORT_MONTH = 'Month_10' THEN CUSTOMERID ELSE NULL END) AS "MONTH_10",
    COUNT(DISTINCT CASE WHEN COHORT_MONTH = 'Month_11' THEN CUSTOMERID ELSE NULL END) AS "MONTH_11",
    COUNT(DISTINCT CASE WHEN COHORT_MONTH = 'Month_12' THEN CUSTOMERID ELSE NULL END) AS "MONTH_12"
FROM CTE3
GROUP BY FIRST_TRANSACTION_MONTH
ORDER BY FIRST_TRANSACTION_MONTH;
```

Customer Retention Analysis

COHORT	MONTH_0	MONTH_1	MONTH_2	MONTH_3	MONTH_4	MONTH_5	MONTH_6	MONTH_7	MONTH_8	MONTH_9	MONTH_10	MONTH_11	MONTH_12
01/12/2010	885	266	304	310	350	327	323	304	308	322	333	409	217
01/01/2011	417	99	103	107	127	110	104	98	124	132	143	22	
01/02/2011	380	67	85	98	99	88	100	102	102	97	21		
01/03/2011	452	73	97	102	98	76	103	112	111	28			
01/04/2011	300	65	65	56	61	58	61	75	17				
01/05/2011	284	50	47	46	57	63	71	20					
01/06/2011	242	43	36	55	61	73	16						
01/07/2011	188	39	31	46	41	10							
01/08/2011	169	34	43	40	4								
01/09/2011	299	73	80	3									
01/10/2011	358	75	14										
01/11/2011	306	7											

Customer Retention Rate (%)

COHORT	MONTH_0	MONTH_1	MONTH_2	MONTH_3	MONTH_4	MONTH_5	MONTH_6	MONTH_7	MONTH_8	MONTH_9	MONTH_10	MONTH_11	MONTH_12
01/12/2010	885	30.1%	34.4%	35.0%	39.5%	36.9%	36.5%	34.4%	34.8%	36.4%	37.6%	46.2%	24.5%
01/01/2011	417	23.7%	24.7%	25.7%	30.5%	26.4%	24.9%	23.5%	29.7%	31.7%	34.3%	5.3%	
01/02/2011	380	17.6%	22.4%	25.8%	26.1%	23.2%	26.3%	26.8%	26.8%	25.5%	5.5%		
01/03/2011	452	16.2%	21.5%	22.6%	21.7%	16.8%	22.8%	24.8%	24.6%	6.2%			
01/04/2011	300	21.7%	21.7%	18.7%	20.3%	19.3%	20.3%	25.0%	5.7%				
01/05/2011	284	17.6%	16.5%	16.2%	20.1%	22.2%	25.0%	7.0%					
01/06/2011	242	17.8%	14.9%	22.7%	25.2%	30.2%	6.6%						
01/07/2011	188	20.7%	16.5%	24.5%	21.8%	5.3%							
01/08/2011	169	20.1%	25.4%	23.7%	2.4%								
01/09/2011	299	24.4%	26.8%	1.0%									
01/10/2011	358	20.9%	3.9%										
01/11/2011	306	2.3%											

Customer Churn Rate (%)

COHORT	MONTH_0	MONTH_1	MONTH_2	MONTH_3	MONTH_4	MONTH_5	MONTH_6	MONTH_7	MONTH_8	MONTH_9	MONTH_10	MONTH_11	MONTH_12
01/12/2010	885	69.9%	65.6%	65.0%	60.5%	63.1%	63.5%	65.6%	65.2%	63.6%	62.4%	53.8%	75.5%
01/01/2011	417	76.3%	75.3%	74.3%	69.5%	73.6%	75.1%	76.5%	70.3%	68.3%	65.7%	94.7%	
01/02/2011	380	82.4%	77.6%	74.2%	73.9%	76.8%	73.7%	73.2%	73.2%	74.5%	94.5%		
01/03/2011	452	83.8%	78.5%	77.4%	78.3%	83.2%	77.2%	75.2%	75.4%	93.8%			
01/04/2011	300	78.3%	78.3%	81.3%	79.7%	80.7%	79.7%	75.0%	94.3%				
01/05/2011	284	82.4%	83.5%	83.8%	79.9%	77.8%	75.0%	93.0%					
01/06/2011	242	82.2%	85.1%	77.3%	74.8%	69.8%	93.4%						
01/07/2011	188	79.3%	83.5%	75.5%	78.2%	94.7%							
01/08/2011	169	79.9%	74.6%	76.3%	97.6%								
01/09/2011	299	75.6%	73.2%	99.0%									
01/10/2011	358	79.1%	96.1%										
01/11/2011	306	97.7%											

Cohort Analysis [Customer Lifetime Value (CLV)]

```
WITH CTE1 AS
(SELECT
    CUSTOMERID,
    str_to_date(INVOICEDATE, '%m/%d/%Y %H:%i') AS FORMATTED_DATE,
    ROUND(QUANTITY*UNITPRICE, 2) AS SALE_VALUE
FROM RETAIL
WHERE
    CUSTOMERID IS NOT NULL
    AND CUSTOMERID != ''
    AND INVOICENO NOT LIKE 'C%'
    AND QUANTITY > 0
    AND UNITPRICE > 0),
CTE2 AS
(SELECT
    CUSTOMERID,
    FORMATTED_DATE AS PURCHASE_DATE,
    MIN(FORMATTED_DATE) OVER (PARTITION BY CUSTOMERID) AS FIRST_TRANSACTION_DATE,
    SALE_VALUE
FROM CTE1),
```

Cohort Analysis [Customer Lifetime Value (CLV)]

```
CTE3 AS
(SELECT
    CUSTOMERID,
    FIRST_TRANSACTION_DATE,
    PURCHASE_DATE,
    SALE_VALUE,
    CONCAT(
        'Month_',
        ROUND(DATEDIFF(PURCHASE_DATE, FIRST_TRANSACTION_DATE)/30, 0)
        ) AS COHORT_MONTH,
    DATE_FORMAT(PURCHASE_DATE, '%Y-%m-01') as PURCHASE_MONTH,
    DATE_FORMAT(FIRST_TRANSACTION_DATE, '%Y-%m-01') AS FIRST_TRANSACTION_MONTH
FROM CTE2)
```

Cohort Analysis [Customer Lifetime Value (CLV)]

```sql
SELECT
    FIRST_TRANSACTION_MONTH AS COHORT,
    ROUND(SUM(CASE WHEN COHORT_MONTH = 'Month_0' THEN SALE_VALUE ELSE 0 END),0) AS "MONTH_0",
    ROUND(SUM(CASE WHEN COHORT_MONTH = 'Month_1' THEN SALE_VALUE ELSE 0 END),0) AS "MONTH_1",
    ROUND(SUM(CASE WHEN COHORT_MONTH = 'Month_2' THEN SALE_VALUE ELSE 0 END),0) AS "MONTH_2",
    ROUND(SUM(CASE WHEN COHORT_MONTH = 'Month_3' THEN SALE_VALUE ELSE 0 END),0) AS "MONTH_3",
    ROUND(SUM(CASE WHEN COHORT_MONTH = 'Month_4' THEN SALE_VALUE ELSE 0 END),0) AS "MONTH_4",
    ROUND(SUM(CASE WHEN COHORT_MONTH = 'Month_5' THEN SALE_VALUE ELSE 0 END),0) AS "MONTH_5",
    ROUND(SUM(CASE WHEN COHORT_MONTH = 'Month_6' THEN SALE_VALUE ELSE 0 END),0) AS "MONTH_6",
    ROUND(SUM(CASE WHEN COHORT_MONTH = 'Month_7' THEN SALE_VALUE ELSE 0 END),0) AS "MONTH_7",
    ROUND(SUM(CASE WHEN COHORT_MONTH = 'Month_8' THEN SALE_VALUE ELSE 0 END),0) AS "MONTH_8",
    ROUND(SUM(CASE WHEN COHORT_MONTH = 'Month_9' THEN SALE_VALUE ELSE 0 END),0) AS "MONTH_9",
    ROUND(SUM(CASE WHEN COHORT_MONTH = 'Month_10' THEN SALE_VALUE ELSE 0 END),0) AS "MONTH_10",
    ROUND(SUM(CASE WHEN COHORT_MONTH = 'Month_11' THEN SALE_VALUE ELSE 0 END),0) AS "MONTH_11",
    ROUND(SUM(CASE WHEN COHORT_MONTH = 'Month_12' THEN SALE_VALUE ELSE 0 END),0) AS "MONTH_12"
FROM CTE3
GROUP BY FIRST_TRANSACTION_MONTH
ORDER BY FIRST_TRANSACTION_MONTH;
```

Customer Lifetime Value (CLV)

COHORT	MONTH_0	MONTH_1	MONTH_2	MONTH_3	MONTH_4	MONTH_5	MONTH_6	MONTH_7	MONTH_8	MONTH_9	MONTH_10	MONTH_11	MONTH_12
01/12/2010	544,267	236,708	244,378	254,988	262,905	288,674	308,568	264,147	374,082	353,317	506,318	470,828	181,757
01/01/2011	290,806	60,139	68,033	69,921	83,228	75,809	66,837	75,889	65,560	114,835	106,163	8,555	
01/02/2011	157,777	27,216	45,545	43,186	40,264	35,625	44,544	63,433	53,324	56,229	10,519		
01/03/2011	199,305	33,276	48,180	48,576	45,584	42,788	59,189	74,495	65,637	9,516			
01/04/2011	120,641	27,251	25,026	22,184	27,615	27,714	27,891	33,163	6,237				
01/05/2011	123,791	17,313	19,041	19,917	22,888	34,302	25,942	9,489					
01/06/2011	135,529	15,307	12,768	26,622	30,303	32,119	6,951						
01/07/2011	72,987	15,250	13,488	18,104	15,613	1,665							
01/08/2011	78,668	26,008	35,655	39,590	676								
01/09/2011	154,368	29,994	31,141	825									
01/10/2011	165,753	39,496	3,362										
01/11/2011	126,904	2,343											

Customer Average Spend

COHORT	MONTH_0	MONTH_1	MONTH_2	MONTH_3	MONTH_4	MONTH_5	MONTH_6	MONTH_7	MONTH_8	MONTH_9	MONTH_10	MONTH_11	MONTH_12
01/12/2010	615	890	804	823	751	883	955	869	1,215	1,097	1,520	1,151	838
01/01/2011	697	607	661	653	655	689	643	774	529	870	742	389	
01/02/2011	415	406	536	441	407	405	445	622	523	580	501		
01/03/2011	441	456	497	476	465	563	575	665	591	340			
01/04/2011	402	419	385	396	453	478	457	442	367				
01/05/2011	436	346	405	433	402	544	365	474					
01/06/2011	560	356	355	484	497	440	434						
01/07/2011	388	391	435	394	381	167							
01/08/2011	465	765	829	990	169								
01/09/2011	516	411	389	275									
01/10/2011	463	527	240										
01/11/2011	415	335											

Insights & Takeaways

- Summarization of patterns:
  - Retention drops sharply after Month 1
  - Some cohorts (e.g., Feb or Mar) show better retention/lifetime value

- Suggestion:
  - Re-engagement campaigns after first month
  - Personalized retention strategy by cohort