Customer Analysis: Cohort, Segmentation, and Retention — PostgreSQL

1. Introduction

In today's data-driven landscape, understanding customer behavior is crucial for business growth and customer retention. This project utilizes PostgreSQL to perform a comprehensive analysis of customer data, focusing on segmentation, cohort-based revenue, and retention trends. By exploring how customer value evolves over time and identifying key patterns in customer behavior, the project provides insights that can help optimize marketing efforts, improve customer engagement, and enhance overall business strategy.

2. Objectives

The objectives of this report are as follows:

- **Segment Customers by Value:** Categorize customers into high, mid, and low value groups based on total LTV to understand contribution levels and prioritize retention strategies.
- **Analyze Revenue by Cohort:** Examine average customer revenue across different cohort years to detect trends, growth patterns, and anomalies over time.
- **Assess Retention and Churn:** Measure customer retention versus churn rates within each cohort to evaluate long-term engagement and identify areas for improvement.
- **Generate Visual Insights:** Create visualizations such as pie charts, bar graphs, and stacked columns to effectively communicate data patterns and trends.
- **Leverage SQL for Business Intelligence:** Demonstrate how PostgreSQL can be used to derive actionable insights from raw customer data through advanced querying and analytical techniques.

3. Tools Used

The following tools were used to format, extract, & analyze data and to visualize the insights:

- PostgreSQL and DBeaver (Data querying and extraction)
- ChatGPT (Data Visualization and Insights)
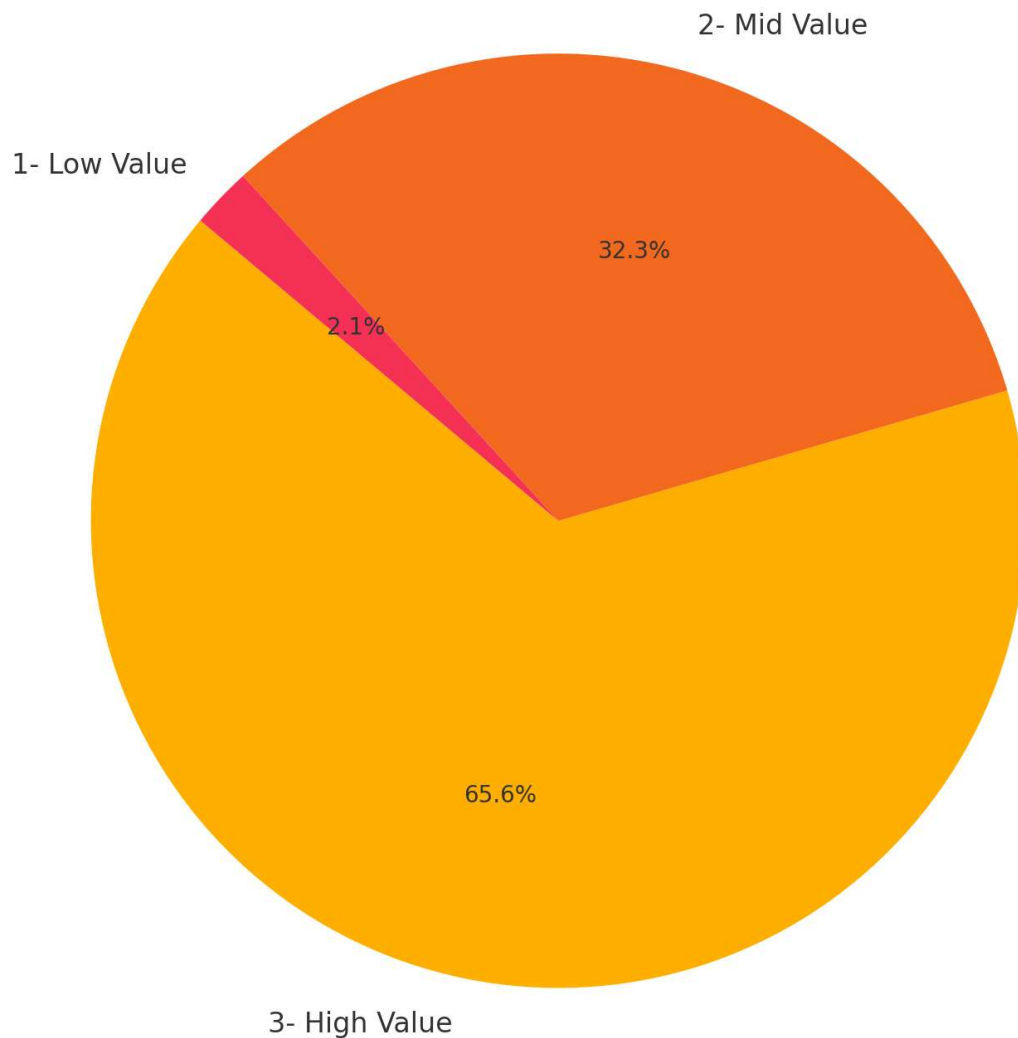
4. Customer Segmentation Analysis

Figure 1: Customer Segmentation by Total Lifetime Value

The pie chart represents customer segmentation based on total lifetime value (LTV). It divides customers into three groups: Low Value (Red), Mid Value (Orange), and High Value (Yellow) of the total customer base. Each segment is visually differentiated by color and percentage, showing the proportional distribution of customers across different value categories.

- High Value customers contribute the majority of revenue (65.6%), despite being equal in number to Low Value customers—highlighting their critical importance to the business.

- Mid Value customers contribute around 32.3% of total LTV and represent a key segment for potential growth.

- Low Value customers contribute only 2.1% of revenue, showing limited financial impact despite their size.

- Strategic focus should be on retaining High Value customers and nurturing Mid Value customers to boost overall profitability.

SQL Query:

```sql
WITH customer_ltv AS (
    SELECT
        customerkey,
        cleaned_name,
        sum(total_net_revenue) AS total_ltv
    FROM
        cohort_analysis
    GROUP BY
        customerkey,
        cleaned_name
),
customer_segments AS (
    SELECT
        percentile_cont(0.25) WITHIN GROUP (
        ORDER BY
            total_ltv
        ) AS ltv_25th_percentile,
        percentile_cont(0.75) WITHIN GROUP (
        ORDER BY
            total_ltv
        ) AS ltv_75th_percentile
    FROM
        customer_ltv
),
segment_values AS (
    SELECT
        c.*,
        CASE
            WHEN c.total_ltv < cs.ltv_25th_percentile THEN '1- Low Value'
            WHEN c.total_ltv <= cs.ltv_75th_percentile THEN '2- Mid Value'
            ELSE '3- High Value'
        END AS customer_segment
    FROM
        customer_ltv c,
        customer_segments cs
)
SELECT
    customer_segment,
    sum(total_ltv) AS total_ltv,
    count(customerkey) AS customer_count,
    sum(total_ltv) / count(customerkey) AS avg_ltv
FROM
    segment_values
GROUP BY
    customer_segment
ORDER BY
    customer_segment DESC;
```

Table:

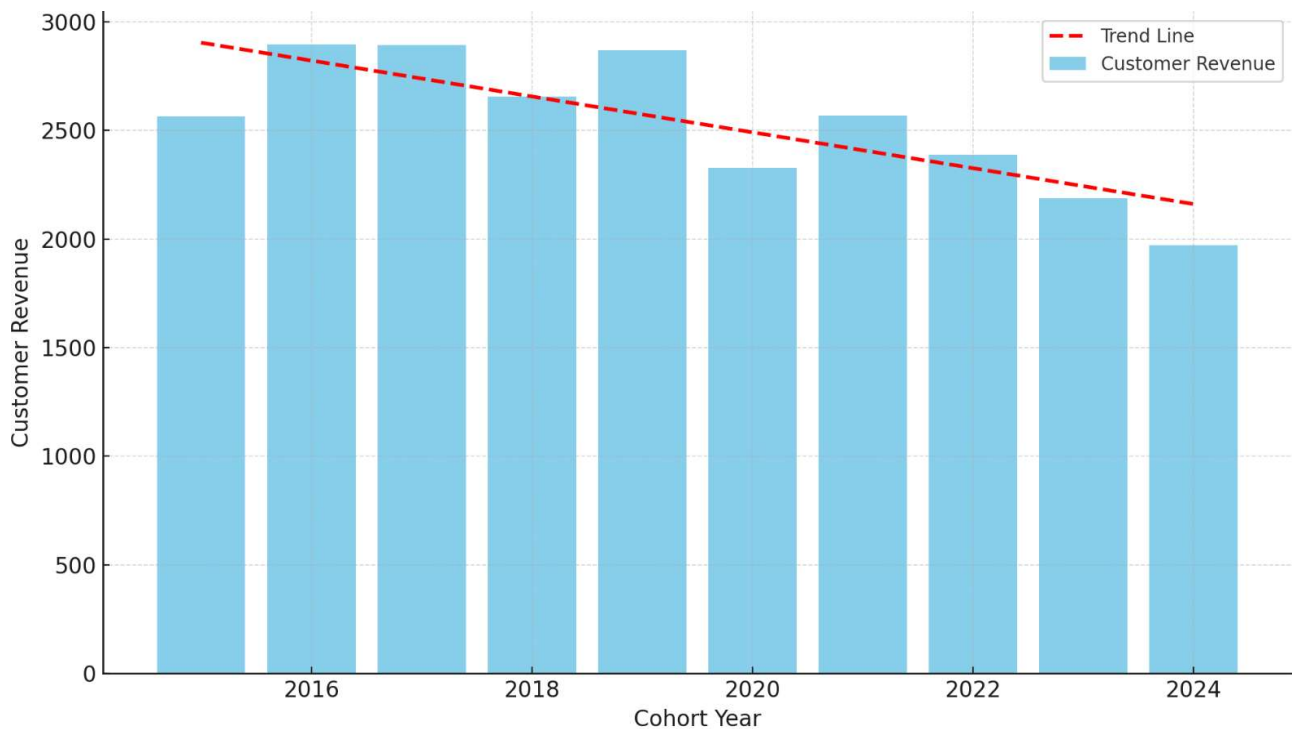|	customer_segment	total_ltv	customer_count	avg_ltv
1	3- High Value	135,429,277.26549858	12,372	10,946.4336619381
2	2- Mid Value	66,636,451.787238546	24,743	2,693.1435875698
3	1- Low Value	4,341,809.527328128	12,372	350.9383711064

5. Cohort Analysis

Figure 2: Customer Revenue by Cohort Year

This bar chart presents customer revenue across different cohort years from 2015 to 2024. The y-axis represents customer revenue, ranging from 0 to 3000, while the x-axis represents cohort years, with each bar corresponding to a specific year. Additionally, a trend line is overlaid to show revenue changes over time.

- Peak revenue per customer occurred in 2016 and 2017, indicating strong early engagement or high-value acquisitions during those years.

- A gradual decline is visible from 2018 onward, with revenue dropping steadily each year—falling to its lowest in 2024.

- This downward trend suggests a need to reassess recent acquisition or retention strategies, as newer cohorts are contributing less revenue on average.

- Retention efforts, onboarding experiences, or customer targeting for recent years may need improvement to boost future cohort performance.

SQL Query:

```sql
--create view cohort_analysis

CREATE OR REPLACE VIEW public.cohort_analysis
AS WITH customer_revenue AS (
    SELECT
        s.customerkey,
        s.orderdate,
        sum(s.quantity * s.netprice * s.exchangerate) AS total_net_revenue,
        count(s.orderkey) AS num_orders,
        max(c.countryfull) AS countryfull,
        max(c.age) AS age,
        max(c.givenname) AS givenname,
        max(c.surname) AS surname
    FROM
        sales s
    JOIN customer c ON
        c.customerkey = s.customerkey
    GROUP BY
        s.customerkey,
        s.orderdate
)
 SELECT
    customerkey,
    orderdate,
    total_net_revenue,
    num_orders,
    countryfull,
    age,
    concat(TRIM(BOTH FROM givenname), ' ', TRIM(BOTH FROM surname)) AS
cleaned_name,
    min(orderdate) OVER (
        PARTITION BY customerkey
    ) AS first_purchase_date,
    EXTRACT(YEAR FROM min(orderdate) OVER (PARTITION BY customerkey)) AS
cohort_year
FROM
    customer_revenue cr;


-------------------
-- Cohort Analysis

SELECT
    cohort_year,
    SUM(total_net_revenue) AS total_revenue,
    COUNT(DISTINCT customerkey) AS total_customers,
    SUM(total_net_revenue) / COUNT(DISTINCT customerkey) AS customer_revenue
FROM cohort_analysis
WHERE orderdate = first_purchase_date
GROUP BY
    cohort_year;
```

Table:

| cohort_year	total_revenue	total_customers	customer_revenue
2,015	7,245,612.977386287	2,825	2,564.8187530571
2,016	9,839,134.340623533	3,397	2,896.418704923
2,017	11,771,496.313936565	4,068	2,893.6814931014
2,018	19,773,770.555444762	7,446	2,655.6232279673
2,019	22,245,058.216363013	7,755	2,868.4794605239
2,020	7,058,614.523282488	3,031	2,328.8071670348
2,021	11,974,082.363847343	4,663	2,567.8924220132
2,022	21,507,554.55017317	9,010	2,387.0759767118
2,023	12,890,580.83980028	5,890	2,188.5536230561
2,024	2,764,779.660859805	1,402	1,972.025435706

6. Customer Retention Analysis
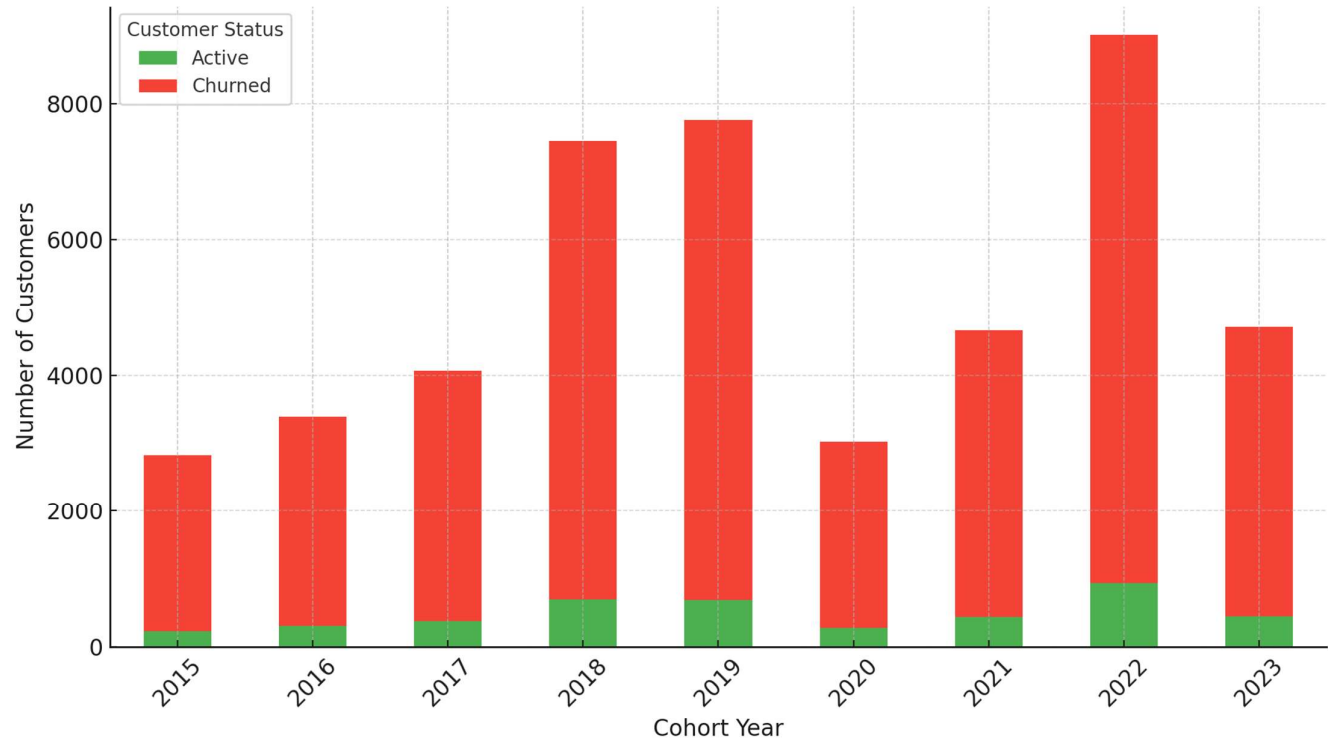
Figure 3: Total Customer Retention by Cohort Year

This bar chart displays the number of customers from 2015 to 2023 and their status as either active (green) or churned (red). The y-axis represents the number of customers, while the x-axis represents cohort years,

- Churn rates are consistently high across all cohorts, with 90–92% of customers becoming inactive within each year group.
- Retention is low, averaging only 8–10% active customers per cohort year, indicating limited long-term engagement.
- While the total number of customers acquired has increased over time (e.g., from 2,825 in 2015 to 9,010 in 2022), the proportion of retained customers has not improved, signaling a persistent retention challenge.
- The 2022 and 2023 cohorts have the highest total customers, yet their retention rates are still around 10%, showing that volume does not compensate for weak retention.
- Businesses should invest in post-acquisition strategies like onboarding, loyalty programs, or personalized communication to improve customer longevity and reduce churn.

SQL Query:

```sql
WITH customer_last_purchase AS
(
    SELECT
            customerkey,
            cleaned_name,
            orderdate,
            ROW_NUMBER() OVER (
                    PARTITION BY customerkey
            ORDER BY
                    orderdate DESC
            ) AS number,
            first_purchase_date,
            cohort_year
    FROM
            cohort_analysis
),
churned_customers AS (
    SELECT
            customerkey,
            cleaned_name,
            first_purchase_date,
            orderdate AS last_purchase_date,
            CASE
                    WHEN orderdate < (SELECT max(orderdate) FROM sales) - INTERVAL
'6 months' THEN 'Churned'
                    ELSE 'Active'
            END AS customer_status,
            cohort_year
    FROM
            customer_last_purchase
    WHERE
            number = 1
            AND first_purchase_date < (SELECT max(orderdate) FROM sales) -
INTERVAL '6 months'
)
SELECT
    cohort_year,
    customer_status,
    count(DISTINCT customerkey) AS total_customers,
```

```
     sum(count(DISTINCT customerkey)) OVER(PARTITION BY cohort_year) AS
total_customers,
     (round(count(DISTINCT customerkey) / sum(count(DISTINCT customerkey))
over(PARTITION BY cohort_year), 2)) * 100  AS percentage
FROM churned_customers
GROUP BY cohort_year, customer_status
```

Table:

	cohort_year	customer_status	total_customers	total_customers	percentage
1	2,015	Active	237	2,825	8
2	2,015	Churned	2,588	2,825	92
3	2,016	Active	311	3,397	9
4	2,016	Churned	3,086	3,397	91
5	2,017	Active	385	4,068	9
6	2,017	Churned	3,683	4,068	91
7	2,018	Active	704	7,446	9
8	2,018	Churned	6,742	7,446	91
9	2,019	Active	687	7,755	9
10	2,019	Churned	7,068	7,755	91
11	2,020	Active	283	3,031	9
12	2,020	Churned	2,748	3,031	91
13	2,021	Active	442	4,663	9
14	2,021	Churned	4,221	4,663	91
15	2,022	Active	937	9,010	10
16	2,022	Churned	8,073	9,010	90
17	2,023	Active	455	4,718	10
18	2,023	Churned	4,263	4,718	90

7. Findings

- High Value customers generate the majority of revenue (~64%), even though they constitute the same number as Low Value customers—underscoring their strategic importance.
- Mid Value customers account for ~31% of total revenue, representing a valuable group for upselling and targeted growth strategies.
- Low Value customers contribute less than 3% of revenue despite their large numbers, suggesting limited ROI on broad marketing to this segment.
- Revenue per customer peaked in 2016 and 2017, followed by a gradual decline through 2024, pointing to a possible drop in customer quality or engagement over time.
- Retention rates remain consistently low, with only 8–10% of customers remaining active across all cohorts. This highlights a significant churn issue.
- Despite an increase in total customer acquisition in recent years, the percentage of retained customers has not improved, indicating a gap in post-acquisition engagement strategies.

8. Recommendations

- **Prioritize Retention of High Value Customers**: Implement loyalty programs, personalized offers, and VIP services to preserve this highly profitable segment.
- **Upsell to Mid Value Customers**: Use targeted campaigns and personalized engagement to move these customers toward the High Value tier.
- **Reassess Acquisition Strategy for Low Value Customers**: Analyze acquisition channels and behaviors to determine if investments in this segment are justified.
- **Revamp Onboarding and Engagement Processes**: Improve the initial customer experience, use email journeys, and encourage product adoption to reduce churn.
- **Implement Data-Driven Retention Campaigns**: Use behavioral segmentation and cohort analysis to time and personalize outreach based on churn risk.
- **Track Cohort Performance Over Time**: Establish regular cohort revenue tracking to quickly identify shifts in customer quality or behavior.

9. Conclusion

This analysis, powered by PostgreSQL, provides key insights into the value, revenue trends, and retention behavior of customers across various cohorts. It highlights a heavy reliance on a small group of High Value customers for overall profitability, a concerning downward trend in revenue from recent cohorts, and consistently low retention rates.

To ensure long-term sustainability, businesses must adopt a more strategic approach—focusing on retaining their most valuable customers, nurturing mid-tier ones, and addressing churn through better onboarding and engagement. Leveraging data analytics to inform these strategies can significantly enhance customer lifetime value and drive sustained business growth.