**Customer Segmentation and CLV Analysis: Annual Sales Insights**

## **Executive Summary**

This report presents a business intelligence project based on the Online Retail dataset, analyzing customer behavior and financial performance to support data-driven decision-making in e-commerce. The main goal was to **evaluate annual sales** by identifying valuable customer segments and tracking revenue and retention trends, helping define **strategic focus areas for the next financial year.**

## **Project Goals & Background**

**Problem**E-commerce businesses often lack clarity on which customers generate the most value over time. This leads to ineffective marketing, unfocused retention strategies, and potential revenue loss.

**Goal**To assess annual sales performance by segmenting customers, identifying retention patterns, estimating customer lifetime value, and visualizing insights in Power BI — helping determine where to focus marketing and sales efforts next year.

The project aimed to:

* Segment customers based on purchasing behavior
* Identify retention trends using cohort analysis
* Measure revenue efficiency (ARPU)
* Estimate customer lifetime value
* Develop a Power BI dashboard with filters by country, segment, and timeframe

The dataset contains over 500,000 transactions from a UK-based online retailer (2010–2011), including customer IDs, countries, timestamps, quantities, and prices.

## **Approach & Tools**

**Technologies Used:**

* SQL – for data cleaning and feature engineering
* Power BI – for interactive dashboard development

**Steps:**

* **Data Cleaning**: Removed canceled orders, missing customer IDs, and invalid values
* **RFM Segmentation**: Quantile-based ranking by Recency, Frequency, and Monetary value
* **Cohort Analysis**: Grouped customers by first purchase month
* **ARPU & CLV Estimation**: Monthly ARPU calculated; CLV projected using historical revenue
* **Dashboard Creation**: Built a dynamic Power BI dashboard with filters

## **Key Insights & Results**

### **1. RFM Analysis**

To better understand customer behavior and value, an RFM (Recency, Frequency, Monetary) analysis was conducted. This method scores each customer based on:

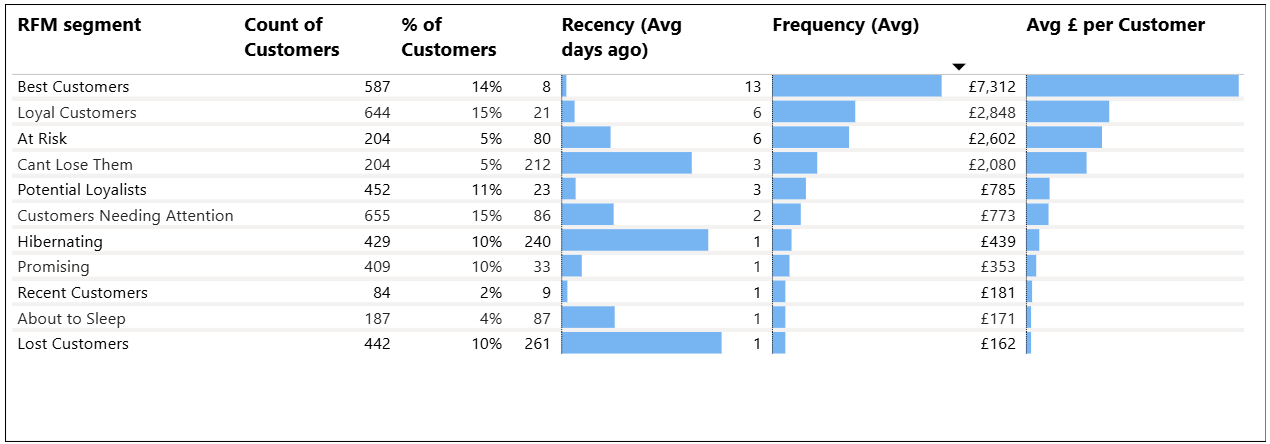
* **Recency** – how recently a customer made a purchase
* **Frequency** – how often the customer made purchases
* **Monetary** – how much the customer spent in total

Customers are assigned scores from 1 to 4 in each category and grouped into meaningful segments. These segments help identify which customers are most valuable, which ones are at risk of churning, and where to focus marketing efforts.

Table below summarizes the main RFM segments used in the analysis:

| **Segment Name** | **Recency Score** | **Frequency/Monetary Score** | **Segment Description** |
| --- | --- | --- | --- |
| **Best Customers** | 4 | 4 | Most recent, most frequent, and highest spending customers. |
| **Loyal Customers** | 4 or 3 | 4 or 3 | Repeat buyers with strong loyalty. |
| **Potential Loyalists** | 4 or 3 | 3 or 2 | New or fairly active customers with potential for loyalty. |
| **Recent Customers** | 4 | 1 | Recently made a purchase, but not yet frequent or high-value. |
| **Promising** | 3 | 2 or 1 | Recent but low-value customers; might become more active |
| **At Risk** | 2 | 4 | Previously valuable customers, but haven’t purchased in a while. |
| **About to Sleep** | 2 | 1 | Low-value customers who haven’t purchased recently. |
| **Need Attention** | 2 | 3 or 2 | Average customers needing engagement to stay active. |
| **Can’t Lose Them** | 1 | 4 or 3 | High-value customers who haven’t purchased in a long time. |
| **Hibernating** | 1 | 2 | Low-to-medium value and haven’t purchased in a long time. |
| **Lost Customers** | 1 | 1 | Inactive and low-value customers, unlikely to return. |

Below this segmentation overview, we analyze the actual customer counts, recency, frequency, and average monetary value for each segment to identify key trends and business insights.

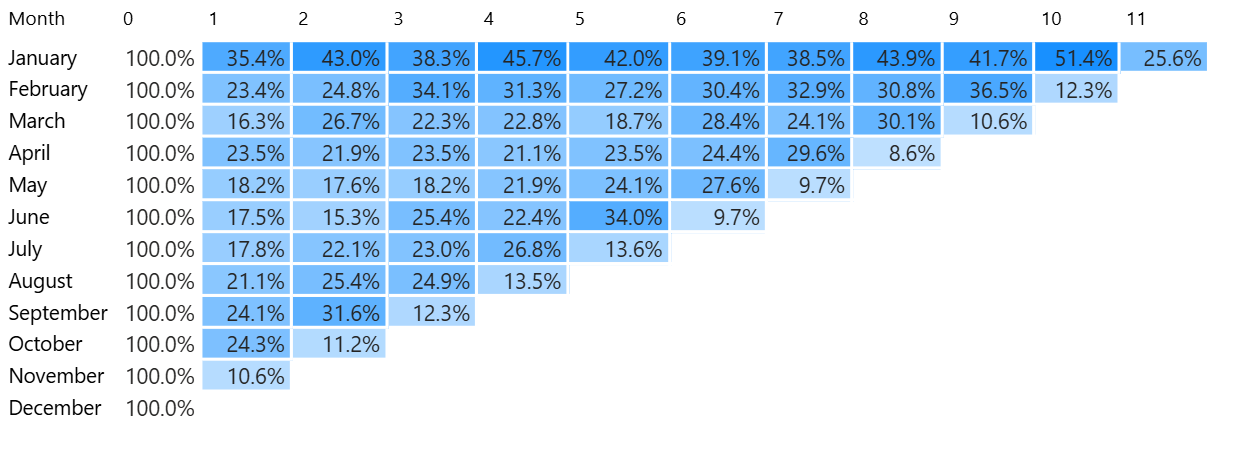


### **Key Insights:**

* **Best and Loyal Customers** (29% of all customers) generate the largest share of revenue thanks to frequent purchases and high average spend. These are the most valuable and engaged clients.
* A combined **10% of customers (At Risk and Can’t Lose Them)** were previously strong spenders but have stopped buying. Without action, this could lead to noticeable revenue loss.
* **20% of customers (Hibernating and Lost)** are inactive and bring very low value. Re-engagement efforts here are unlikely to deliver strong returns.
* **21% of the base (Potential Loyalists and Promising)** show early signs of loyalty. With the right follow-up, they could become long-term, high-value customers.

### **2. Cohort Analysis**

Monthly retention was tracked to evaluate how long customers remained active

after their first purchase.

### **Key Insights:**

### **Initial retention is low (~18-24%)**, but many customers return **after several months**, with retention peaking between months 4 and 10.

### This pattern aligns with **wholesale customer behavior**, who buy less often but in larger quantities and on a predictable schedule.

### The **January cohort** is particularly loyal, with over **50% retention at month 10**, indicating strong, long-term relationships.

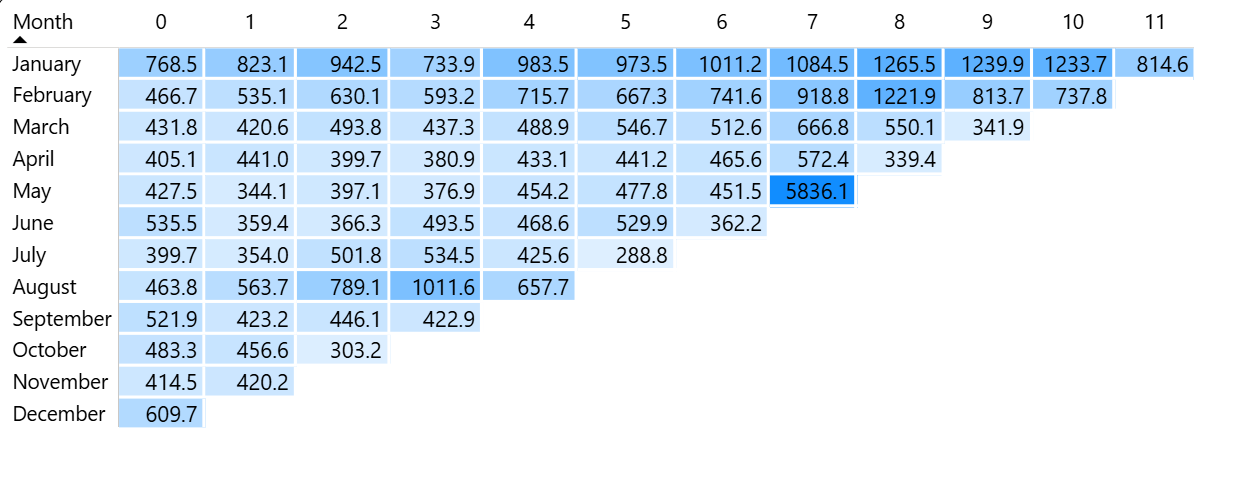
### Understanding these cycles allows the business to **tailor marketing and sales efforts by customer type**, increasing the efficiency of re-engagement campaigns and inventory planning.

### 

### **3. Average Revenue Per User**

ARPU was analyzed on a monthly basis to evaluate revenue efficiency per customer.

### **Key Insights:**

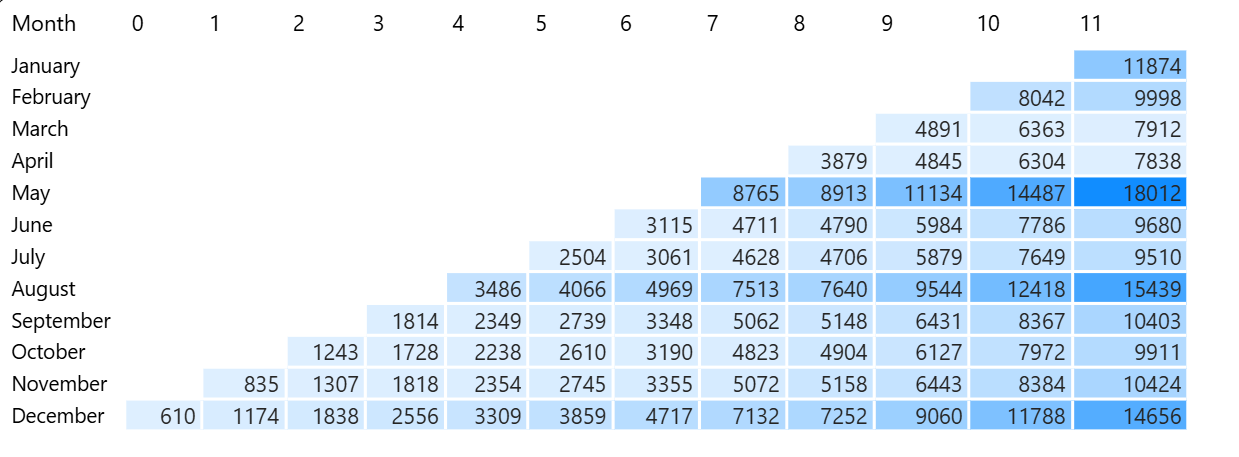
* **January cohort demonstrates strong revenue growth over time,** starting near £770 per customer and increasing steadily to over £1,200 by month 8–10. This signals solid repeat purchasing behavior typical for wholesale clients.  
  
* **February cohort follows a similar pattern,** with ARPU rising from £467 initially to peaks above £1,200, confirming value in nurturing customers beyond their first purchase.
* Most cohorts start with an ARPU around £400–£600, reflecting initial order size. **Subsequent months show moderate growth or stability.**
* The **May cohort’s month 7 spike to £5,836 is a notable outlier,** likely driven by a large wholesale order or special event. This warrants further investigation to understand drivers and potential replication.
* Overall, revenue patterns reinforce that **long-term wholesale customers deliver increasing value over time.**

### 

### **4. Customer Lifetime Value**

CLV was projected using predictive cumulative ARPU over a one-year horizon. The estimation helps prioritize marketing efforts and forecast future revenue.

### **Key Insights:**

* The annual CLV was calculated based on cumulative monthly averages. Using this predictive model, the estimated **average annual value per user is approximately 11.3K.**

## **Conclusion: Strategic Recommendations for the Next Fiscal Year**

Based on the customer segmentation and financial performance analysis, the following key strategic directions are recommended:

1. **Invest in Retaining Top Customers**Best and Loyal segments bring the most revenue. Keep them engaged with loyalty programs, targeted emails, and exclusive deals.
2. **Re-engage Lost High-Value Clients**At Risk and Can’t Lose Them customers were previously strong spenders. Use personalized campaigns to win them back.
3. **Support Growing Segments with Potential**Potential Loyalists and Promising customers could become loyal buyers. Keep in touch with them and offer small deals to build the relationship.
4. **Adapt to Wholesale Buying Cycles**Cohort trends show that wholesale customers buy less often but in larger volumes. Plan campaigns and stock accordingly.
5. **Prioritize Data-Driven Marketing**Use RFM and CLV insights to focus marketing budgets on the segments that drive long-term profit, not just short-term sales.

**Appendices**

# **Query for Sales Dashboard**

WITH revenue\_by\_country AS (

SELECT

Country,

SUM(UnitPrice \* Quantity) AS total\_revenue

FROM

`retail-465908.online\_retail.retail`

WHERE

InvoiceDate BETWEEN '2010-12-01' AND '2011-12-01'

AND Quantity > 0

AND UnitPrice > 0

AND CustomerID IS NOT NULL

GROUP BY

Country

),

top\_5\_countries AS (

SELECT

Country

FROM

revenue\_by\_country

ORDER BY

total\_revenue DESC

LIMIT 5

)

SELECT

\*,

SUM(UnitPrice \* Quantity) AS total\_revenue,

CASE

WHEN r.Country IN (SELECT Country FROM top\_5\_countries) THEN r.Country

ELSE 'Other countries'

END AS Region\_Group

FROM

`retail-465908.online\_retail.retail` r

WHERE

r.InvoiceDate BETWEEN '2010-12-01' AND '2011-12-01'

AND r.Quantity > 0

AND r.UnitPrice > 0

AND r.CustomerID IS NOT NULL

GROUP BY

ALL

# **Query for RFM**

---calculated monetary,frequency and recenecy and cleaned negative price data

WITH rfm AS (

SELECT

CustomerID,

SUM(Quantity \* UnitPrice) AS monetary,

COUNT(DISTINCT InvoiceNo) AS frequency,

DATE\_DIFF(DATE '2011-12-01', DATE(MAX(InvoiceDate)), DAY) AS recency

FROM

`retail-465908.online\_retail.retail`

WHERE

InvoiceDate BETWEEN '2010-12-01' AND '2011-12-01'

AND UnitPrice > 0

AND Quantity > 0

AND CustomerID IS NOT NULL

GROUP BY

CustomerID

),

--- set quantiles

quant AS (

SELECT

APPROX\_QUANTILES(frequency, 100) AS f\_q,

APPROX\_QUANTILES(recency, 100) AS r\_q,

APPROX\_QUANTILES(monetary, 100) AS m\_q

FROM

rfm

),

--- divided to 4 equal parts

percent AS (

SELECT

m\_q[OFFSET(25)] AS m25, m\_q[OFFSET(50)] AS m50, m\_q[OFFSET(75)] AS m75,

f\_q[OFFSET(25)] AS f25, f\_q[OFFSET(50)] AS f50, f\_q[OFFSET(75)] AS f75,

r\_q[OFFSET(25)] AS r25, r\_q[OFFSET(50)] AS r50, r\_q[OFFSET(75)] AS r75

FROM

quant

),

--- cross joined percents and set rfm scores

scored AS (

SELECT

rfm.\*,

CASE

WHEN monetary <= m25 THEN 1

WHEN monetary <= m50 THEN 2

WHEN monetary <= m75 THEN 3

ELSE 4

END AS m\_score,

CASE

WHEN frequency <= f25 THEN 1

WHEN frequency <= f50 THEN 2

WHEN frequency <= f75 THEN 3

ELSE 4

END AS f\_score,

CASE

WHEN recency <= r25 THEN 4

WHEN recency <= r50 THEN 3

WHEN recency <= r75 THEN 2

ELSE 1

END AS r\_score

FROM

rfm

CROSS JOIN

percent

),

fm\_score\_cte AS (

SELECT

\*,

CAST(ROUND((f\_score + m\_score) / 2) AS INT64) AS fm\_score

FROM

scored

)

--- set customer segements by their rfm scores

SELECT

CustomerID,

recency,

frequency,

monetary,

r\_score,

f\_score,

m\_score,

fm\_score,

CASE

WHEN r\_score = 4 AND fm\_score = 4 THEN 'Best Customers'

WHEN (r\_score = 4 AND fm\_score = 3) OR (r\_score = 3 AND fm\_score = 4) THEN 'Loyal Customers'

WHEN (r\_score = 4 AND fm\_score = 2) OR (r\_score = 3 AND fm\_score = 3) THEN 'Potential Loyalists'

WHEN r\_score = 4 AND fm\_score = 1 THEN 'Recent Customers'

WHEN (r\_score = 3 AND fm\_score = 1) OR (r\_score = 3 AND fm\_score = 2) THEN 'Promising'

WHEN r\_score = 2 AND fm\_score = 4 THEN 'At Risk'

WHEN r\_score = 2 AND fm\_score = 1 THEN 'About to Sleep'

WHEN (r\_score = 2 AND fm\_score = 3) OR (r\_score = 2 AND fm\_score = 2) THEN 'Customers Needing Attention'

WHEN (r\_score = 1 AND fm\_score = 4) OR (r\_score = 1 AND fm\_score = 3) THEN 'Cant Lose Them'

WHEN r\_score = 1 AND fm\_score = 2 THEN 'Hibernating'

WHEN r\_score = 1 AND fm\_score = 1 THEN 'Lost Customers'

END AS rfm\_segment

FROM

fm\_score\_cte

ORDER BY

fm\_score DESC, r\_score DESC

# **Query for Cohort Analysis**

-- filtered valid transactions in 2011

WITH filtered\_data AS (

SELECT

\*

FROM

`retail-465908.online\_retail.retail`

WHERE

InvoiceDate BETWEEN '2011-01-01' AND '2011-12-31'

AND Quantity > 0

AND UnitPrice > 0

AND CustomerID IS NOT NULL

),

-- got each customer's first purchase month

first\_purchase AS (

SELECT

CustomerID,

DATE\_TRUNC(DATE(MIN(InvoiceDate)), MONTH) AS cohort\_month

FROM

filtered\_data

GROUP BY

CustomerID

),

-- counted customers and revenue by cohort and months since first purchase

actual\_data AS (

SELECT

fp.cohort\_month,

DATE\_DIFF(DATE\_TRUNC(DATE(p.InvoiceDate), MONTH), fp.cohort\_month, MONTH) AS cohort\_index,

COUNT(DISTINCT p.CustomerID) AS customer\_count,

SUM(p.Quantity \* p.UnitPrice) AS revenue

FROM

filtered\_data p

JOIN

first\_purchase fp ON p.CustomerID = fp.CustomerID

WHERE

DATE\_DIFF(DATE\_TRUNC(DATE(p.InvoiceDate), MONTH), fp.cohort\_month, MONTH) BETWEEN 0 AND 11

GROUP BY

fp.cohort\_month,

cohort\_index

),

-- got the size of each cohort

cohort\_sizes AS (

SELECT

cohort\_month,

COUNT(DISTINCT CustomerID) AS cohort\_size

FROM

first\_purchase

GROUP BY

cohort\_month

)

SELECT

ad.cohort\_month,

ad.cohort\_index,

ad.customer\_count,

ad.revenue,

SAFE\_DIVIDE(ad.revenue, NULLIF(ad.customer\_count, 0)) AS average\_revenue\_per\_customer,

cs.cohort\_size,

SAFE\_DIVIDE(ad.customer\_count, cs.cohort\_size) \* 100 AS retention\_percent

FROM

actual\_data ad

JOIN

cohort\_sizes cs ON ad.cohort\_month = cs.cohort\_month

ORDER BY

ad.cohort\_month,

ad.cohort\_index;