Customer Retention Strategy (RFM Analysis)

The Business Requirement (User Story):

"The marketing team at an e-commerce retailer was struggling to identify high-value customers who had stopped purchasing. With data trapped in raw transaction logs (7,000+ rows), they lacked actionable visibility into their customer base, leading to lost revenue and increased churn."

The Solution:

I designed an end-to-end analytics workflow to segment customers based on purchasing behavior (Recency, Frequency, Monetary Value).

- Data Engineering (SQL): I utilized a Common Table Expression (CTE) to aggregate over 7,000 transaction records, calculating key RFM metrics (Recency, Frequency, Monetary) for each customer. To ensure dynamic scoring, I applied the NTILE(5) window function to rank customers into percentiles rather than using static thresholds. Finally, I implemented a logic-based CASE statement to synthesize these rankings into actionable business segments (e.g., 'Champions', 'At Risk'), creating a prioritized dataset ready for marketing execution.

- Analysis Logic: Applied RFM Segmentation to categorize users into 5 strategic groups, including "Champions" (High Spend, Recent) and "At Risk" (High Spend, Inactive).

- Visualization (Tableau): Built an interactive dashboard featuring:
  - Treemap: For a strategic overview of revenue distribution.
  - Cluster Analysis: A scatter plot to mathematically validate the segments.
  - Action List: A dynamic table for marketing managers to export "At-Risk" leads.

Key Insights and Results:

- Pareto Principle Validated: The top 20% of "Champions" are driving ~60% of total revenue.
- Churn Risk: Identified a specific segment of high-value customers who haven't purchased in >200 days.
- Strategic Opportunity: A targeted win-back campaign on the "At-Risk" list could recover an estimated 15% of potentially lost revenue.

Strategic Recommendations:

Based on the RFM segmentation analysis, I recommend the following immediate actions:

- Launch a "Win-Back" Campaign: The Marketing team should export the "At Risk" list (customers with high historical spend but >150 days inactivity) and send a targeted "We Miss You" email with a time-sensitive discount.
- Protect the VIPs: The "Champions" segment drives the majority of revenue. Implement an exclusive "Early Access" program for these users to maintain their high frequency and prevent churn.

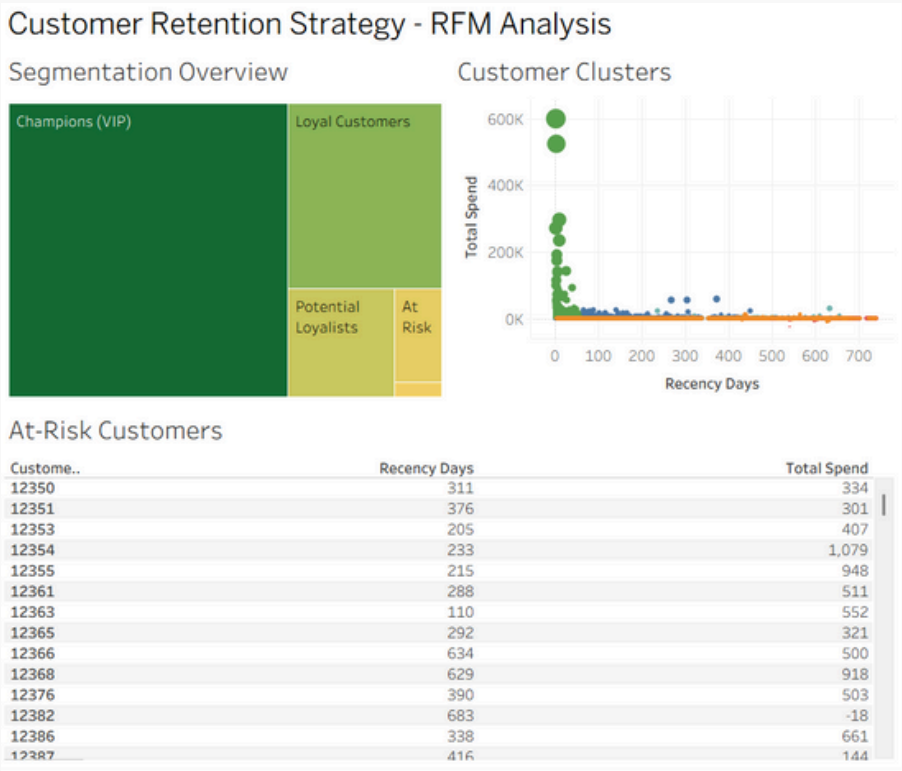
Live Project:

[Click here to view the interactive dashboard on Tableau Public]

```
...> WITH Customer_Metrics AS (
...>     SELECT
...>         "Customer ID" AS Customer_ID,
...>
...>         -- 1. Recency: Calculate days since 2011-12-10 using DuckDB date math
...>         -- We cast the string column to a DATE first
...>         date_diff('day', CAST(MAX(InvoiceDate) AS DATE), DATE '2011-12-10') AS Recency_Days,
...>
...>         -- 2. Frequency
...>         COUNT(DISTINCT Invoice) AS Frequency,
...>
...>         -- 3. Monetary
...>         SUM(Quantity * Price) AS Total_Spend
...>
...>     FROM 'online_retail_II.csv' -- Query the file directly
...>     WHERE "Customer ID" IS NOT NULL
...>     GROUP BY "Customer ID"
...> ),
...>
...> Ranked_Customers AS (
...>     SELECT
...>         *,
...>         -- Rank them 1-5 (5 is best)
...>         NTILE(5) OVER (ORDER BY Recency_Days DESC) AS R_Score,
...>         NTILE(5) OVER (ORDER BY Frequency ASC) AS F_Score,
...>         NTILE(5) OVER (ORDER BY Total_Spend ASC) AS M_Score
...>     FROM Customer_Metrics
...> )
...>
...> SELECT
...>     Customer_ID,
...>     Recency_Days,
...>     Frequency,
...>     Total_Spend,
...>     (R_Score + F_Score + M_Score) AS Total_Score,
...>     CASE
...>         WHEN (R_Score + F_Score + M_Score) >= 14 THEN 'Champions (VIP)'
...>         WHEN (R_Score + F_Score + M_Score) >= 11 THEN 'Loyal Customers'
...>         WHEN (R_Score + F_Score + M_Score) >= 8 THEN 'Potential Loyalists'
...>         WHEN (R_Score + F_Score + M_Score) >= 5 THEN 'At Risk'
...>         ELSE 'Lost / Hibernating'
...>     END AS Customer_Segment
...> FROM Ranked_Customers;
```

16007.0	48	7	3075.1399999999962	13	Loyal Customers
13464.0	59	7	3083.89	12	Loyal Customers
16584.0	81	7	3085.4699999999975	12	Loyal Customers
14312.0	47	11	3090.2900000000001	14	Champions (VIP)
12849.0	32	6	3090.5600000000004	13	Loyal Customers
15491.0	96	7	3090.8500000000017	12	Loyal Customers
13397.0	73	13	3094.11	13	Loyal Customers
12589.0	29	11	3094.1700000000001	14	Champions (VIP)
13772.0	34	9	3094.7200000000016	13	Loyal Customers
15105.0	19	9	3094.910000000002	13	Loyal Customers
12937.0	16	9	3097.6099999999997	14	Champions (VIP)
13650.0	17	16	3098.5000000000023	15	Champions (VIP)
16933.0	2	16	3099.6000000000013	15	Champions (VIP)
14897.0	8	9	3100.74	14	Champions (VIP)
13742.0	17	14	3102.4200000000001	15	Champions (VIP)
17870.0	3	14	3103.1999999999999	15	Champions (VIP)
14044.0	27	11	3106.1499999999965	14	Champions (VIP)
14978.0	73	12	3106.339999999999	13	Loyal Customers
17640.0	87	8	3107.5900000000006	12	Loyal Customers
13538.0	44	12	3112.7300000000002	14	Champions (VIP)
14700.0	33	5	3116.259999999999	12	Loyal Customers
13685.0	3	7	3119.4400000000014	14	Champions (VIP)
15199.0	60	10	3122.4	12	Loyal Customers
15095.0	104	11	3135.7400000000002	13	Loyal Customers
16918.0	50	19	3138.06	14	Champions (VIP)
17530.0	2	13	3140.5600000000004	15	Champions (VIP)
14306.0	52	7	3141.4999999999995	13	Loyal Customers
17354.0	51	4	3143.4600000000023	12	Loyal Customers
15632.0	16	10	3147.3100000000013	14	Champions (VIP)
12456.0	45	4	3153.889999999998	12	Loyal Customers
15353.0	255	8	3154.0499999999993	11	Loyal Customers
17353.0	130	7	3155.55	12	Loyal Customers
12976.0	212	8	3156.3799999999983	11	Loyal Customers
14963.0	10	13	3158.6300000000001	15	Champions (VIP)
14285.0	22	9	3158.64	13	Loyal Customers
16400.0	95	9	3159.4399999999999	12	Loyal Customers
12951.0	9	20	3164.42	15	Champions (VIP)
16041.0	31	3	3166.4	12	Loyal Customers
17679.0	53	11	3166.5600000000018	14	Champions (VIP)

SQL query performing data cleaning and RFM calculation.



Final Tableau Dashboard showing the "Champions" cluster and the "At-Risk" action list.