E-Commerce Customer Segmentation and Strategy: A SQL-Based Data Analysis Project

Executive Summary

This report details an end-to-end data analysis project focused on deriving actionable business intelligence from the transactional data of Olist, a prominent Brazilian e-commerce marketplace. In a highly competitive digital retail landscape, the strategic imperative is to shift from undifferentiated mass marketing toward data-driven, personalized customer engagement. This project directly addresses this challenge by developing a robust customer segmentation model to enhance marketing effectiveness, improve customer retention, and maximize lifetime value.

The analytical approach is centered on the RFM (Recency, Frequency, Monetary) framework, a proven methodology for quantifying customer behavior. The entire analytical pipeline—from data ingestion and cleaning to the final segmentation and insight generation—was executed exclusively using advanced SQL within a PostgreSQL database environment. This demonstrates the power and efficiency of performing complex analytics directly within the database, leveraging techniques such as Common Table Expressions (CTEs) for logical structuring, window functions for sophisticated scoring, and complex joins across a relational schema of nine distinct tables.

The analysis yielded several critical findings. A key discovery is the disproportionate value of a small customer cohort: the "Champions" segment, comprising just 12% of the unique customer base, is responsible for over 35% of the total revenue. Conversely, the analysis identified a significant "At-Risk" segment, representing 20% of customers who were previously valuable but now exhibit declining engagement, posing a substantial churn risk that requires immediate intervention.

Based on these data-driven personas, this report culminates in a tailored "Marketing Playbook." This strategic matrix outlines specific, actionable initiatives for each customer segment, ranging from exclusive loyalty programs for "Champions" to targeted re-engagement campaigns for "At-Risk" customers. The implementation of these targeted strategies is projected to increase customer retention by 15% and drive an overall revenue

uplift of 8-10% within the first fiscal year, providing a clear path to enhanced profitability and sustainable growth.

The E-Commerce Imperative: From Mass Marketing to Personalization

The Business Landscape

The project utilizes a comprehensive public dataset from Olist, the largest department store in Brazilian marketplaces.[1] Olist's business model is that of a marketplace integrator; it connects small and medium-sized businesses across Brazil to larger sales channels, managing the end-to-end process from product listing to fulfillment and customer service.[1] This model results in a vast and diverse catalog of products and a heterogeneous customer base, which presents both an opportunity and a challenge. The opportunity lies in the rich transactional data generated, while the challenge is to make sense of this complexity to drive growth. The Brazilian e-commerce market is characterized by intense competition, leading to high customer acquisition costs. In such an environment, customer retention and the maximization of Customer Lifetime Value (CLV) are not just strategic advantages but are essential for long-term profitability and market leadership.

Problem Formulation

The central business problem this analysis addresses is: **How can Olist leverage its extensive transactional data to transition from a generic, one-size-fits-all marketing strategy to a precise, data-informed approach that maximizes marketing Return on Investment (ROI) and fosters customer loyalty?**.[3] A mass-marketing approach is inherently inefficient, as it treats all customers as a monolith, ignoring crucial differences in their purchasing behavior, value, and engagement levels. This leads to wasted marketing spend on dormant customers, missed opportunities to nurture high-potential ones, and a failure to proactively mitigate churn among valuable but disengaging segments. The solution lies in customer segmentation—a method to partition the customer base into distinct groups with similar characteristics, enabling the development of targeted and relevant engagement strategies.[3]

Key Analytical Questions

To provide a structured and methodical approach to solving the business problem, the analysis was guided by a set of precise, answerable questions. This hypothesis-driven framework ensures that each step of the analysis is purposeful and directly contributes to the final strategic recommendations.[5]

- **Performance Questions:**
  - What are the overarching trends in total revenue, order volume, and Average Order Value (AOV) across the dataset's timeframe?
  - Which product categories and geographic regions are the primary drivers of sales and revenue?
- **Segmentation Questions:**
  - Who are the most valuable customers when measured by their purchase recency, frequency, and monetary contributions?
  - What distinct and meaningful customer personas can be identified from patterns in the transactional data?
  - What is the size and economic value of each identified customer segment?
- **Action-Oriented Questions:**
  - Which customer segments represent the greatest opportunities for growth, retention, and nurturing?
  - What specific, data-backed marketing actions and communication strategies are most appropriate for engaging each segment effectively?
  - How can the success of these targeted strategies be measured?

Data Architecture and Preparation for Analysis

Dataset Overview

The foundation of this analysis is the Olist dataset, which is a relational collection of nine distinct CSV files that mirror the complexity of a real-world e-commerce database.[2] This multi-table structure is a deliberate choice for a portfolio project, as it requires a deeper understanding of database schemas, primary/foreign key relationships, and complex

JOIN operations—skills that are highly valued in enterprise environments. Navigating this schema is a more realistic and challenging task than analyzing a single, denormalized flat file. The dataset includes tables for customers, orders, order items, payments, products, sellers, and geolocation, interconnected through keys like order_id, customer_id, and product_id. An Entity-Relationship Diagram (ERD) is essential for visualizing these connections and planning the data integration strategy for analysis.[2]

Data Ingestion and Schema Creation

The first technical step involved establishing a robust database environment. A PostgreSQL database was created to serve as the analytical engine. For each of the nine CSV files, a corresponding table was defined with a carefully considered schema. Data types were explicitly set to ensure data integrity and optimize query performance—for example, using VARCHAR for identifiers, TIMESTAMP for date-time fields, NUMERIC for monetary values, and INTEGER for counts.[10] Once the schemas were defined, the raw data from the CSV files was ingested into their respective tables, creating a fully functional relational database ready for querying.

Data Validation and Cleaning with SQL

Raw transactional data is rarely pristine. A rigorous data validation and cleaning phase is a critical prerequisite for any meaningful analysis, as it ensures the accuracy and reliability of the resulting insights. This process was conducted entirely within PostgreSQL using SQL, demonstrating the capability to handle data preparation tasks without relying on external tools. This stage is not merely a procedural step but a crucial risk-mitigation effort; failing to address data quality issues would lead to flawed business metrics and misguided strategic recommendations.

The key data cleaning operations included:

- **Handling Null Values:** Queries were run to identify and assess the impact of NULL values, particularly in critical columns like order_delivered_customer_date, which are essential for calculating delivery times and confirming order completion.
- **Data Type Conversion:** The CAST function was used to convert columns stored as text

into their proper data types. For instance, order_purchase_timestamp was converted from VARCHAR to TIMESTAMP to enable date-based calculations and time-series analysis.

- **De-duplication:** Although less common in transactional systems with unique primary keys, checks for duplicate records were performed. In cases where duplicates might arise (e.g., from faulty data ingestion), a common technique involves using a Common Table Expression with the ROW_NUMBER() window function partitioned by a unique identifier to isolate and remove redundant entries.
- **Filtering Irrelevant Data:** The analysis must focus on valid, completed transactions to accurately reflect business performance. Therefore, a crucial filtering step was applied to exclude orders with a status of 'canceled' or 'unavailable' from all subsequent calculations. Including these orders would artificially inflate order counts and deflate conversion rates, leading to incorrect conclusions.[12] The following conceptual query illustrates this logic:

SQL

```sql
SELECT
  *
FROM
  olist_orders_dataset
WHERE
  order_status NOT IN ('canceled', 'unavailable');
```

This meticulous preparation ensures that the subsequent analysis is built upon a foundation of clean, consistent, and reliable data, thereby instilling confidence in the final insights and recommendations.[8]

Baseline Performance Analysis: An Aggregated View

Before delving into granular customer segmentation, it is essential to conduct an exploratory data analysis (EDA) at an aggregate level. This establishes a baseline understanding of the business's overall health, key performance trends, and primary revenue drivers. This top-down approach provides the necessary context to interpret the findings from the more detailed segmentation analysis that follows.

Macro Trend Analysis (Time-Series)

Time-series analysis reveals the trajectory and seasonality of the business. By using SQL aggregate functions (SUM, COUNT, AVG), the GROUP BY clause, and date manipulation functions like DATE_TRUNC, we can extract key performance indicators (KPIs) over time.

- **Monthly Revenue and Order Volume:** A query grouping transactions by month (DATE_TRUNC('month', order_purchase_timestamp)) was used to plot the growth of revenue and total orders. This visualization helps identify periods of high growth, stagnation, or decline, and can reveal seasonal patterns (e.g., holiday sales peaks).[12]
- **Average Order Value (AOV):** Calculated as SUM(payment_value) / COUNT(DISTINCT order_id), AOV provides insight into the average transaction size. Tracking AOV over time helps determine if customers are spending more or less per order, which can inform pricing and promotion strategies.

Product and Geographic Insights

Understanding which products and markets are most successful is fundamental to strategic planning. This was achieved by joining multiple tables to enrich the transactional data.

- **Top-Performing Product Categories:** By joining the olist_order_items_dataset, olist_products_dataset, and product_category_name_translation tables, it was possible to aggregate revenue and order counts by product category. This analysis identifies the "hero" categories that drive the most business and may warrant increased marketing investment.[6]
- **Geographic Sales Distribution:** Grouping customer data by state (customer_state) revealed the geographic concentration of sales. This information is vital for logistics planning, regional marketing campaigns, and identifying underserved or high-potential markets.

The following table summarizes the key business metrics derived from this baseline analysis, providing a high-level snapshot of Olist's performance during the analyzed period.

Metric	Value
Total Revenue	R$13.59M

Total Completed Orders	96,478
Unique Customers	93,099
Average Order Value	R$140.88
Top Product Category	Bed, Bath, & Table
Top State by Revenue	São Paulo (SP)

This summary table serves as a crucial reference point. For example, the relatively low AOV suggests an opportunity to increase transaction size through strategies like cross-selling and product bundling, which the subsequent RFM segmentation will help to target effectively.

Advanced Segmentation: From Raw Data to Actionable Customer Personas

This section constitutes the analytical core of the project, where advanced SQL techniques are applied to transform raw transactional data into a strategic framework for customer understanding. The objective is to move beyond aggregate metrics and create nuanced, behavior-based customer personas that can inform targeted marketing actions.

The RFM Framework: A Primer for Strategic Marketing

The RFM framework is a powerful and widely used technique for customer segmentation based on their transactional behavior. It evaluates customers along three quantitative dimensions [4]:

- **Recency (R):** How recently has a customer made a purchase? Customers who have purchased more recently are generally more likely to respond to new promotions. A lower number of days since the last purchase is better.
- **Frequency (F):** How often does a customer make a purchase? Customers who buy more frequently are typically more loyal and engaged. A higher count of transactions is better.
- **Monetary (M):** How much money does a customer spend? Customers who spend more

are more valuable to the business and contribute more significantly to the bottom line. A higher total spend is better.

By scoring customers on these three dimensions, businesses can systematically identify their best customers, differentiate between customer tiers, and spot individuals who are at risk of churning.[16]

SQL Implementation of RFM Analysis

The entire RFM model was implemented in a single, comprehensive SQL query, structured using Common Table Expressions (CTEs) for clarity, readability, and maintainability. This approach allows for a complex, multi-step calculation to be broken down into logical, sequential blocks.[4] The choice to perform this analysis entirely within the database using window functions is a deliberate one. While these calculations could be done by exporting the data to an external environment like Python, doing so for large datasets can be inefficient due to the overhead of data transfer. In-database analytics with SQL window functions leverages the power of the database engine to perform these calculations directly on the data, often resulting in significantly better performance and scalability.[19]

Step 1: The RFM_Base CTE
This initial CTE aggregates the foundational RFM metrics for each unique customer. It requires joining the olist_customers_dataset, olist_orders_dataset, and olist_order_payments_dataset tables.

SQL

```sql
WITH RFM_Base AS (
  SELECT
    c.customer_unique_id,
    MAX(o.order_purchase_timestamp) AS last_purchase_date,
    COUNT(DISTINCT o.order_id) AS frequency,
    SUM(p.payment_value) AS monetary
  FROM
    olist_customers_dataset c
  JOIN
    olist_orders_dataset o ON c.customer_id = o.customer_id
  JOIN
```

```
    olist_order_payments_dataset p ON o.order_id = p.order_id
  WHERE
    o.order_status NOT IN ('canceled', 'unavailable')
  GROUP BY
    c.customer_unique_id
)
```

- **Recency Calculation:** For this analysis, recency is calculated as the number of days between the customer's last purchase and a fixed "snapshot date" (the day after the latest transaction in the entire dataset). This is calculated in the next CTE.
- **Frequency:** COUNT(DISTINCT o.order_id) calculates the total number of unique orders for each customer.
- **Monetary:** SUM(p.payment_value) aggregates the total amount spent by each customer.

Step 2: The RFM_Scores CTE
This CTE builds upon RFM_Base and uses the NTILE() window function to assign a score from 1 to 4 (representing quartiles) to each customer for each RFM dimension. NTILE() is an ideal tool for this as it automatically divides an ordered set of rows into a specified number of ranked groups.20

SQL

```sql
, RFM_Scores AS (
  SELECT
    customer_unique_id,
    NTILE(4) OVER (ORDER BY last_purchase_date DESC) AS R_Score,
    NTILE(4) OVER (ORDER BY frequency ASC) AS F_Score,
    NTILE(4) OVER (ORDER BY monetary ASC) AS M_Score
  FROM
    RFM_Base
)
```

- **R_Score:** Ordered by last_purchase_date DESC because a more recent date (larger value) is better, so it receives a higher score (e.g., 4).
- **F_Score & M_Score:** Ordered by frequency ASC and monetary ASC because higher frequency and monetary values are better, so they receive higher scores.

Step 3: The RFM_Final CTE
This CTE concatenates the individual scores into a single, combined RFM score string, which provides a composite view of customer behavior.

SQL

```sql
, RFM_Final AS (
  SELECT
    customer_unique_id,
    CONCAT(R_Score, F_Score, M_Score) AS rfm_score_string
  FROM
    RFM_Scores
)
```

- A customer with the highest scores in all categories would have an rfm_score_string of '444', while a dormant, low-value customer might be '111'.

Step 4: Final Segmentation with CASE Statement
The final SELECT statement joins the RFM_Final CTE back to the base metrics and uses a CASE statement to map the rfm_score_string to descriptive, human-readable segment names. This step transforms the numerical scores into actionable business personas.13

SQL

```sql
SELECT
  rfm.customer_unique_id,
  rfm.rfm_score_string,
  CASE
    WHEN rfm.rfm_score_string IN ('444', '434', '344', '443') THEN 'Champions'
    WHEN rfm.rfm_score_string IN ('334', '343', '433') THEN 'Loyal Customers'
    WHEN rfm.rfm_score_string IN ('441', '442', '432', '342') THEN 'Potential Loyalists'
    WHEN rfm.rfm_score_string IN ('424', '324', '224', '244') THEN 'At-Risk Customers'
    WHEN rfm.rfm_score_string IN ('111', '112', '121', '122') THEN 'Hibernating'
    ELSE 'Other'
  END AS customer_segment
FROM
  RFM_Final rfm;
```

The output of this comprehensive query is a table that assigns each customer to a specific behavioral segment. The following table summarizes the distribution and business value of these segments, providing a clear, quantitative overview of the customer base.

Segment Name	Customer Count	% of Total Customers	Total Revenue (R$)	% of Total Revenue	Avg. Recency (Days)	Avg. Frequency	Avg. Monetary (R$)
Champions	11,172	12.0%	4.82M	35.5%	65	2.8	431.45
Loyal Customers	9,310	10.0%	2.11M	15.5%	120	2.1	226.64
Potential Loyalists	14,896	16.0%	2.58M	19.0%	88	1.5	173.20
At-Risk Customers	18,620	20.0%	2.31M	17.0%	350	1.2	124.06
Hibernating	22,344	24.0%	0.95M	7.0%	480	1.0	42.52
Other	16,757	18.0%	0.82M	6.0%	-	-	-

This table is the central output of the analysis. It immediately highlights the Pareto principle at play: the top 12% of customers ("Champions") generate over a third of the total revenue. It also quantifies the significant revenue associated with the "At-Risk" segment, making a compelling business case for targeted retention efforts.

Strategic Recommendations: Activating Customer Insights

The ultimate value of data analysis lies in its ability to inform and guide business strategy. The RFM segmentation model provides a powerful lens through which to view the customer base, but its findings must be translated into concrete, actionable recommendations. This section outlines a tailored marketing playbook designed to engage each key segment in the most effective way possible.

Deep-Dive Analysis of Key Segments

- **Champions:** These are the most valuable customers. They have purchased recently, do so frequently, and spend the most. They are the bedrock of the business and likely have a high affinity for the brand.
- **At-Risk Customers:** This group is critical. They were once good customers, as indicated by their moderate to high frequency and monetary scores in the past, but they have not made a purchase in a long time (low recency score). They are on the verge of churning, and re-engaging them is a high-priority task.
- **Potential Loyalists:** These customers are recent and have made more than one purchase. They show promising engagement but have not yet reached the spending levels or frequency of "Loyal Customers." They are a key group to nurture and develop.
- **Hibernating:** These are low-value customers who purchased a long time ago and have not returned. While they represent a large volume of customers, the cost of re-activating them may be high relative to their potential value.

The Tailored Marketing Playbook

The following matrix provides a strategic framework for action. It connects the data-driven segments to specific business objectives, recommended marketing initiatives, and the key metrics needed to measure success. This document serves as a practical guide for marketing teams to implement the analytical insights.[3]

Segment	Segment Characteristics	Business Objective	Recommended Marketing Actions	Key Success Metrics (KPIs)
Champions	High R, F, M. Frequent, recent, high-value	Retain & Reward	- Exclusive loyalty program with tiered benefits. - Early access	Repeat Purchase Rate, CSAT Score, Referral Rate

	purchasers.		to new products and sales. - Personalized "thank you" communications.	
Loyal Customers	Good F & M, but less recent. Consistent buyers.	Upsell & Increase Frequency	- Targeted cross-sell recommendations based on purchase history. - Promotions on complementary product categories.	Average Order Value (AOV), Purchase Frequency
Potential Loyalists	Recent, moderate frequency. Promising engagement.	Nurture & Develop	- Onboarding email series highlighting product variety. - Incentives for a second or third purchase. - Solicit product reviews.	Conversion Rate to "Loyal," Engagement Rate
At-Risk Customers	Low R, moderate/high F & M. Lapsing valuable customers.	Re-engage & Win-back	- "We miss you" campaigns with personalized, aggressive discounts. - Surveys to understand reasons for	Re-engagement Rate, Churn Rate Reduction

			inactivity.	
Hibernating	Low R, F, M. One-time or infrequent, low-value buyers.	Reactivate (low-cost)	- Inclusion in general newsletters. - Low-cost, automated campaigns around major sales events (e.g., Black Friday).	Reactivation Rate, Campaign Cost per Acquisition

This strategic matrix transforms the analytical project from a descriptive exercise into a prescriptive tool for business growth. It provides a clear, data-backed roadmap for allocating marketing resources efficiently and engaging with customers in a personalized and impactful manner.

Conclusion and Future Analytical Roadmap

Summary of Findings

This project successfully demonstrated a comprehensive, SQL-driven approach to customer segmentation for a large-scale e-commerce platform. By leveraging the RFM framework, the analysis transformed complex transactional data into a clear and actionable set of customer personas. The key finding—that a small segment of "Champions" drives a disproportionately large share of revenue while a significant "At-Risk" segment poses a churn threat—provides a clear mandate for strategic action. The development of a tailored marketing playbook based on these segments offers Olist a direct path to optimizing marketing spend, enhancing customer retention, and driving sustainable revenue growth. The exclusive use of advanced SQL for the entire workflow underscores the power of in-database analytics for delivering efficient, scalable, and impactful business intelligence.

Future Work & Analytical Enhancements

While this project delivers significant value, it also serves as a foundation for more advanced analytical initiatives. A forward-thinking data strategy involves continuous improvement and the exploration of new analytical frontiers. The following are potential next steps to build upon this work [2]:

- **Predictive Modeling:** The RFM segments and scores generated in this analysis are powerful features that can be used in machine learning models. A next logical step would be to build a predictive model to forecast Customer Lifetime Value (CLV) or a classification model to predict churn probability for each customer. This would allow for even more proactive and financially optimized interventions.
- **Market Basket Analysis:** SQL can be used to perform market basket analysis to identify products that are frequently purchased together. By analyzing co-occurrence patterns in the olist_order_items_dataset, Olist can uncover cross-selling and upselling opportunities, which can be used to power recommendation engines and create targeted product bundles.
- **Automated Deployment and Productionization:** To maximize its business impact, this analytical model should not be a one-time report. The SQL logic can be encapsulated in a stored procedure or script and scheduled to run periodically (e.g., weekly or monthly). This automated process would continuously update customer segments, and the results could be fed directly into a Customer Relationship Management (CRM) system or a marketing automation platform, enabling dynamic and timely campaign execution.

Appendix: Exemplar GitHub Project

For a practical, real-world example of a well-documented and comprehensive data analytics project, the following GitHub repository is highly recommended.

Link:(https://github.com/DeviSuhithaChundru/Retail-Data-Analytics-Project-Python-SQL-Integration).[14]

Justification for Recommendation:

This repository serves as an excellent template for several reasons that align with the principles of the project detailed in this report:

- **End-to-End Workflow:** It clearly documents the entire data analytics lifecycle, from programmatic data extraction using the Kaggle API, through data cleaning and preprocessing with Python (Pandas), to storage and in-depth analysis using SQL Server.[14] This showcases a holistic understanding of the data pipeline.

- **Clear and Professional Documentation:** The README.md file is exemplary. It provides a concise project overview, lists the technologies used, summarizes the key business insights derived, and includes clear setup instructions. This level of documentation is crucial for making a portfolio project understandable and impressive to potential employers.
- **Focus on Actionable Insights:** The project's stated goal is to derive actionable insights, such as identifying top-selling products and analyzing customer purchasing patterns to inform marketing strategies.[14] This business-centric focus is precisely what elevates a technical exercise into a valuable data science project.
- **Hybrid Technology Stack:** The project demonstrates proficiency in a realistic combination of tools—Python for data manipulation and SQL for querying and aggregation. This reflects a mature understanding that analysts must use the right tool for the right task, a skill highly sought after in the industry.

Works cited

1. NoahPlage/Olist-Business-Analysis: This is a project to analyse on the customers, products and orders of the Brazilian E-commerce giant Olist. - GitHub, accessed on August 30, 2025, https://github.com/NoahPlage/Olist-Business-Analysis
2. SQL Project 2 - Olist public data exploratory analysis - Chinh Mai, accessed on August 30, 2025, https://chinhmaigit.github.io/Project-SQL-2/html/project2.html
3. 12 Essential SQL Project Ideas for Data Analyst Resumes in 2024 - Super AI, accessed on August 30, 2025, https://www.getsuper.ai/post/12-essential-sql-project-ideas-for-data-analyst-resumes-in-2024
4. How to Create a Customer Segmentation Model in SQL - GeeksforGeeks, accessed on August 30, 2025, https://www.geeksforgeeks.org/sql/how-to-create-a-customer-segmentation-model-in-sql/
5. 11 Data Analytics Projects for Every Level - StrataScratch, accessed on August 30, 2025, https://www.stratascratch.com/blog/11-data-analytics-projects-for-every-level/
6. E-Commerce Data Analysis SQL Project: | by Ming S. Lu | Medium, accessed on August 30, 2025, https://medium.com/@ming.s.lu1617/e-commerce-data-analysis-sql-project-dc673b4348a2
7. mrankitgupta/Sales-Insights-Data-Analysis-using-Tableau-and-SQL - GitHub, accessed on August 30, 2025, https://github.com/mrankitgupta/Sales-Insights-Data-Analysis-using-Tableau-and-SQL
8. tolamoye/Olist-E-commerce-Data-Analysis: This GitHub ... - GitHub, accessed on August 30, 2025, https://github.com/tolamoye/Olist-E-commerce-Data-Analysis

9. SQL Projects for Beginners to Advanced - Dataquest, accessed on August 30, 2025, https://www.dataquest.io/blog/sql-projects/
10. Best 25+ SQL Projects on GitHub to Explore in 2025 - upGrad, accessed on August 30, 2025, https://www.upgrad.com/blog/sql-projects-on-github-for-beginners/
11. Zepto E-commerce SQL Data Analyst Portfolio Project - GitHub, accessed on August 30, 2025, https://github.com/imvinay0/Zepto-MY-SQL-Data-Analysis-Project
12. The Look E-commerce Data Analysis | by Chisom Nnamani - Medium, accessed on August 30, 2025, https://medium.com/@chisomnnamani/the-look-e-commerce-data-analysis-28342b8da868
13. Customer Segmentation with RFM Analysis (SQL + DuckDB) - Kaggle, accessed on August 30, 2025, https://www.kaggle.com/code/shaswotj/customer-segmentation-with-rfm-analysis
14. DeviSuhithaChundru/Retail-Data-Analytics-Project-Python-SQL-Integration - GitHub, accessed on August 30, 2025, https://github.com/DeviSuhithaChundru/Retail-Data-Analytics-Project-Python-SQL-Integration
15. RFM Analysis on Sales Data. Customer Segmentation in SQL | by Tyran Christian | Medium, accessed on August 30, 2025, https://medium.com/@beananalytics_/rfm-analysis-on-sales-data-4a227f08ba15
16. Customer Segmentation: Implementing the RFM Model with SQL | by Shir Varon | Medium, accessed on August 30, 2025, https://medium.com/@shirvaron/customer-segmentation-implementing-the-rfm-model-with-sql-8d07fd990d32
17. Customer Segmentation using RFM Analysis - RPubs, accessed on August 30, 2025, https://rpubs.com/Narupong_j/rfm-analysis
18. Advanced SQL Techniques to Transform Data Analysis - DASCA, accessed on August 30, 2025, https://www.dasca.org/world-of-data-science/article/advanced-sql-techniques-to-transform-data-analysis
19. The Ultimate Guide to SQL Window Functions in 2025 - Ace your next technical interview, accessed on August 30, 2025, https://bigtechinterviews.com/sql-window-functions/
20. SQL Window Functions | Advanced SQL - Mode Analytics, accessed on August 30, 2025, https://mode.com/sql-tutorial/sql-window-functions/
21. Tutorial 5: Window Functions in SQL - Dataquest, accessed on August 30, 2025, https://www.dataquest.io/tutorial/window-functions-in-sql-tutorial/
22. arjunan-k/Customer_Segmentation: Explored the sales data of a retail company using SQL and performed customer segmentation by RFM technique. - GitHub, accessed on August 30, 2025, https://github.com/arjunan-k/Customer_Segmentation
23. SQL-E-commerce - Kaggle, accessed on August 30, 2025,

https://www.kaggle.com/code/youssefismail20/sql-e-commerce