E-Commerce Consumer Behavior Analysis using SQL

1. Executive Summary

This project analyzes e-commerce consumer behavior using SQL to derive business insights across purchasing channels, seasonality, revenue concentration, customer satisfaction, return risk, loyalty programs, and marketing effectiveness. Advanced SQL techniques such as CTEs, window functions, CASE statements, and aggregations were used.

Key findings include balanced omnichannel usage, revenue concentration, high return rates not necessarily indicating poor customer satisfaction, loyalty programs improving satisfaction but not immediate spending, and moderate AD engagement driving the highest average spend.

2. Business Objective

To analyze consumer purchasing behavior and identify actionable insights that can help optimize revenue, customer experience, marketing efficiency, and operational performance using SQL-based analysis.

3. Dataset Overview

The dataset contains customer-level purchase records with demographic details, transaction values, product feedback, marketing engagement indicators, and logistics preferences. Each row represents a **single purchase event**, not repeated transactions per customer.

3. Purchase Channel Analysis
SQL Query and Output


```sql
1  SELECT "Purchase_Channel",
2  │  │  │  COUNT(*) AS total_orders
3  FROM "Ecommerce"
4  GROUP BY "Purchase_Channel"
5  ORDER BY total_orders DESC;
6
```

Results Chart Export ∨

Purchase_Channel	total_orders
Mixed	340
Online	334
In-Store	326

Observation

The distribution of orders across purchase channels is relatively balanced. The **Mixed channel** records the highest number of orders (340), followed closely by **Online** (334) and **In-store** (326). The small difference between channels indicates that customers are actively using **multiple touchpoints** rather than relying on a single channel.

Insights

- Customers exhibit strong **omnichannel behavior**, with similar order volumes across online, in-store, and mixed channels.
- The slightly higher usage of the **Mixed channel** indicates a preference for flexible purchase journeys across platforms.

4. Time of Purchase Analysis
SQL Query and Output

```sql
   1   SELECT
   2     TO_CHAR(
   3       TO_DATE("Time_of_Purchase", 'MM/DD/YYYY'),
   4       'Mon YYYY'
   5     ) AS purchase_month,
   6     COUNT(*) AS total_orders
   7   FROM "Ecommerce"
   8   GROUP BY purchase_month
   9   ORDER BY MIN(TO_DATE("Time_of_Purchase", 'MM/DD/YYYY'));
```

Results   Chart   Export ∨

purchase_mont	total_orders
Jan 2024	75
Feb 2024	68
Mar 2024	93
Apr 2024	100
May 2024	76
Jun 2024	89
Jul 2024	94
Aug 2024	98
Sep 2024	84
Oct 2024	79
Nov 2024	80
Dec 2024	64

Observation

Customer orders fluctuate across the year, with noticeable peaks during **April (100 orders)**, **July (94 orders)**, and **August (98 orders)**. Lower order volumes are observed in **February (68)** and **December (64)**, indicating uneven demand distribution across months.

Insights

- **Mid-year months (April–August)** experience higher purchase activity, suggesting seasonal demand or increased customer engagement during this period.
- **Early and year-end months** show comparatively lower order volumes, indicating opportunities for targeted promotions or campaigns to stabilize demand.

5. Category-Wise Revenue Contribution
SQL Query and Output

```sql
SELECT "Purchase_Category",
       ROUND(SUM("Purchase_Amount"), 2) AS total_revenue
FROM "Ecommerce"
GROUP BY "Purchase_Category"
ORDER BY total_revenue DESC;
```

Results | Chart | Export ∨

Purchase_Category	total_revenue
Jewelry & Accessories	15139.36
Sports & Outdoors	14610.51
Electronics	13842.41
Software & Apps	13601.41
Toys & Games	13536.46
Home Appliances	13191.82
Food & Beverages	12966.96
Packages)	12731.16
Health Care	12149.03

Gardening & Outdoors	11782.28
Mobile Accessories	11689.18
Groceries	11467.65
Animal Feed	11467.10
Baby Products	11172.52
Books	11122.69
Health Supplements	10908.31
Office Supplies	10835.01
Furniture	10282.29
Luxury Goods	10247.00
Travel & Leisure (Flights	9589.78
Hotels	8854.98
Clothing	8401.04
Beauty & Personal Care	7945.01
Arts & Crafts	7529.92

Observation

Revenue contribution varies significantly across product categories. **Jewelry & Accessories** generates the highest total revenue, followed by **Sports & Outdoors**, **Electronics**, and **Software & Apps**. In contrast, categories such as **Arts & Crafts**, **Beauty & Personal Care**, and **Clothing** contribute comparatively lower revenue.

Insights

- Revenue is concentrated in **premium and lifestyle-oriented categories**, indicating strong customer preference for higher-value discretionary purchases.
- Lower-performing categories present opportunities for **pricing optimization, bundling, or targeted promotions** to improve their revenue contribution.

6. Product Rating vs Purchase Category
SQL Query and Output

```
1  SELECT "Purchase_Category",
2         ROUND(AVG("Product_Rating"), 2) AS avg_rating
3  FROM "Ecommerce"
4  GROUP BY "Purchase_Category"
5  ORDER BY avg_rating DESC
6
```

Results   Chart   Export ∨

Purchase_Category	avg_rating
Travel & Leisure (Flights	3.58
Toys & Games	3.47
Baby Products	3.46
Mobile Accessories	3.32
Luxury Goods	3.32
Hotels	3.28
Beauty & Personal Care	3.26
Office Supplies	3.16
Software & Apps	3.16

Health Care	3.10
Health Supplements	3.07
Packages)	3.05
Animal Feed	3.00
Clothing	2.97
Furniture	2.95
Home Appliances	2.94
Electronics	2.89
Sports & Outdoors	2.88
Books	2.84
Jewelry & Accessories	2.76
Food & Beverages	2.70
Groceries	2.63
Arts & Crafts	2.62
Gardening & Outdoors	2.57

Observation

Average product ratings vary noticeably across purchase categories. **Travel & Leisure (Flights)** has the highest average rating (3.58), followed by **Toys & Games** and **Baby Products**. In contrast, categories such as **Gardening & Outdoors**, **Arts & Crafts**, **Groceries**, and **Food & Beverages** receive comparatively lower average ratings.

Insights
- **Service-oriented and experiential categories** tend to receive higher customer ratings than commodity-based categories.
- Lower-rated categories may require improvements in **product quality, expectation setting, or post-purchase support** to enhance customer satisfaction.

7. Return Rate Risk Analysis
SQL Query and Output

Ecommerce table preview

```sql
SELECT
    "Purchase_Category",
    ROUND(AVG("Frequency_of_Purchase"), 2) AS avg_purchase_frequency,
    ROUND(AVG("Return_Rate"), 2) AS avg_return_rate,
    COUNT(*) AS total_orders,
    CASE
        WHEN AVG("Return_Rate") >= 1.05 THEN 'High Risk'
        WHEN AVG("Return_Rate") BETWEEN 0.90 AND 1.04 THEN 'Improvement Needed'
        ELSE 'Healthy'
    END AS risk_segment
FROM "Ecommerce"
GROUP BY "Purchase_Category"
ORDER BY avg_return_rate DESC;
```

Results   Chart   Export ∨

Purchase_Category	avg_purchase_	avg_return_rate	total_orders	risk_segment
Travel & Leisure (Flights	7.08	1.32	38	High Risk
Gardening & Outdoors	6.34	1.14	44	High Risk
Office Supplies	6.74	1.13	38	High Risk
Jewelry & Accessories	6.12	1.10	50	High Risk
Hotels	6.06	1.03	36	Improvement Needed
Books	7.24	1.03	37	Improvement Needed
Baby Products	7.49	1.02	41	Improvement Needed
Groceries	6.95	1.02	41	Improvement Needed
Electronics	6.81	1.00	54	Improvement Needed
Animal Feed	7.11	0.98	44	Improvement Needed
Home Appliances	6.88	0.94	50	Improvement Needed
Software & Apps	5.81	0.93	43	Improvement Needed
Food & Beverages	7.36	0.93	44	Improvement Needed
Luxury Goods	7.38	0.92	37	Improvement Needed

Furniture	8.63	0.88	41	Healthy
Beauty & Personal Care	7.82	0.88	34	Healthy
Clothing	6.93	0.87	30	Healthy
Arts & Crafts	6.41	0.85	34	Healthy
Mobile Accessories	7.12	0.85	41	Healthy
Health Supplements	7.10	0.85	41	Healthy
Sports & Outdoors	6.37	0.84	51	Healthy
Toys & Games	7.40	0.83	47	Healthy
Packages)	7.30	0.81	43	Healthy
Health Care	6.56	0.73	41	Healthy

Observations
1. High Risk Categories (Avg Return Rate ≥ 1.05)
- **Travel & Leisure (Flights)**, **Gardening & Outdoors**, **Office Supplies**, and **Jewelry & Accessories** fall into the *High Risk* segment.
- These categories exhibit **elevated return rates (>1.10)** despite having **moderate to high purchase frequency (6–7+)**.
- Jewelry & Accessories and Office Supplies also have **high order volumes**, amplifying the financial impact of returns.

2. Improvement Needed Categories (Avg Return Rate 0.90–1.04)
- Categories such as **Hotels, Books, Baby Products, Groceries, Electronics, Food & Beverages, Luxury Goods**, and **Software & Apps** sit in the middle segment.
- These categories generally show **healthy purchase frequency (6–7.5)** but **return rates close to the risk threshold**.
- This segment represents the **largest group**, indicating broad but manageable operational leakage.

3. Healthy Categories (Avg Return Rate < 0.90)
- Categories including **Furniture, Beauty & Personal Care, Clothing, Mobile Accessories, Sports & Outdoors, Toys & Games, Packages, and Health Care** are classified as *Healthy*.
- These categories demonstrate a **strong balance of high purchase frequency and low return rates**.
- Furniture stands out with the **highest average purchase frequency (8.63)** and a **low return rate (0.88)**.

Insights
1. High Purchase Frequency Does Not Guarantee Low Returns

- Several *High Risk* and *Improvement Needed* categories (e.g., Travel, Jewelry, Office Supplies) have **frequent repeat purchases but poor post-purchase satisfaction**, as reflected by higher returns.
- This suggests **expectation mismatch, quality issues, or unclear product/service descriptions** rather than low demand.

2. Immediate Attention Areas
- **Travel & Leisure (Flights)** is the most critical category:
  - High purchase frequency
  - Highest return rate
  - Indicates systemic issues such as cancellations, policy dissatisfaction, or service variability.
- **Jewelry & Accessories** and **Office Supplies** should be prioritized due to **high volume × high return risk**, which directly impacts revenue and operational costs.

3. Optimization Opportunities in "Improvement Needed" Segment
- This segment offers the **highest ROI for intervention**.
- Small improvements in product information, pricing transparency, or customer education could shift these categories into the *Healthy* segment.

4. Best Practices from Healthy Categories
- Healthy categories provide a **benchmark for success**:
  - Clear value proposition
  - Lower expectation gaps
  - Better alignment between customer intent and product delivery
- Insights from **Furniture, Health Care, and Toys & Games** can inform process improvements in riskier categories.

8. Customer Satisfaction vs Returns
SQL Query and Output

```sql
SELECT "Return_Rate",
        ROUND(AVG("Customer_Satisfaction"), 2) AS avg_satisfaction
FROM "Ecommerce"
GROUP BY "Return_Rate";
```

Results   Chart   Export ∨

Return_Rate	avg_satisfaction
0	5.28
1	5.50
2	5.43

Observations

- Orders with **Return Rate = 0 (no returns)** have an average satisfaction score of **5.28**.
- Orders with **Return Rate = 1** show the **highest average satisfaction at 5.50**.
- Orders with **Return Rate = 2 (multiple returns)** have a slightly lower average satisfaction of **5.43**, but still **higher than no-return orders**.

Insights

1. **Returns do not necessarily indicate dissatisfaction**
   Contrary to common assumptions, customers who returned items (once or even multiple times) reported **equal or higher satisfaction** compared to those who did not return items.
2. **Effective return experience may boost satisfaction**
   The higher satisfaction for Return Rate = 1 suggests that a **smooth, customer-friendly return or replacement process** positively influences overall customer perception.
3. **Returns may reflect engagement, not churn risk**
   Customers who initiate returns are likely **active, engaged buyers** who continue interacting with the platform until their needs are met.
4. **Risk lies in friction, not the return itself**
   Since satisfaction remains high even with multiple returns, the real business risk is not the return behavior but **operational cost and logistics**, not customer loyalty.

9. Loyalty Program Impact (CTE)
SQL Query and Output

```sql
1  WITH loyalty_analysis AS (
2      SELECT "Customer_Loyalty_Program_Member",
3             ROUND(AVG("Purchase_Amount"),2) AS avg_spend,
4             ROUND(AVG("Customer_Satisfaction"), 2) AS avg_satisfaction
5      FROM "Ecommerce"
6      GROUP BY "Customer_Loyalty_Program_Member"
7  )
8  SELECT * FROM loyalty_analysis;
```

Results   Chart   Export ∨

Customer_Loyalty_Program_Member	avg_spend	avg_satisfaction
false	288.37	5.31
true	261.27	5.49

Observations
- **Non-loyalty members** have a higher average spend (**288.37**) compared to loyalty program members (**261.27**).
- **Loyalty program members** report higher average customer satisfaction (**5.49**) than non-members (**5.31**).

Insights
1. **Loyalty programs improve satisfaction, not immediate spend**
   Membership appears to enhance customer experience and perceived value, reflected in higher satisfaction, but does not directly translate into higher per-order spending.
2. **Non-members may represent high-value, transactional buyers**
   Higher average spend among non-members suggests the presence of occasional or premium buyers who purchase less frequently but spend more per transaction.

10. Marketing Effectiveness – Ad Engagement
SQL Query and Output

```sql
1  SELECT "Engagement_with_Ads",
2         ROUND(AVG("Purchase_Amount"), 2) AS avg_spend
3  FROM "Ecommerce"
4  GROUP BY "Engagement_with_Ads";
```

Results   Chart   Export ∨

Engagement_with_Ads	avg_spend
High	277.66
Medium	281.21
Low	268.26
None	272.58

Observation

- Customers with **Medium ad engagement** have the highest average spend (**281.21**).
- **High engagement** customers spend slightly less (**277.66**) than medium-engaged users.
- **Low or no engagement** groups show comparatively lower average spend (**268.26–272.58**).

Insights

1. **Moderate ad exposure is most effective for conversion value**
   Medium engagement likely reflects optimal ad frequency—enough to influence purchase without causing fatigue.
2. **High engagement does not guarantee higher spend**
   Highly engaged users may interact more with ads but could be more price-sensitive or comparison-driven, limiting spend per transaction.

Strategic Business Recommendations

1. **Improve Return Reduction Through Product & Logistics Optimization**
   Since higher return rates do not significantly lower customer satisfaction, efforts should focus on improving product information, sizing guidance, and delivery quality rather than tightening return policies.
2. **Adopt Category-Specific Strategies Based on Risk Segmentation**
   High-risk categories require quality and fulfilment improvements, while healthy

categories should receive increased marketing and inventory support to maximize revenue efficiency.

3. **Strengthen Omnichannel Experience Consistency**
Balanced usage across online, in-store, and mixed channels indicates the need for consistent pricing, promotions, and service levels across all customer touchpoints.

4. **Optimize Advertising Frequency to Avoid Saturation**
Medium ad engagement drives the highest average spend, suggesting that marketing effectiveness can be improved by reducing overexposure and focusing on optimal ad frequency.

5. **Redefine Loyalty Program Success Metrics**
Loyalty programs should be evaluated using long-term retention and lifetime value metrics rather than average order value, as members show higher satisfaction but not immediate spending increases.