# SQL Funnel & Retention Analysis of e commerce events history in electronic store

- **Funnel → tells who converted**
- **Retention → tells who came back**

By Sachin shanbhogue k

# Project Objective

**To analyze user behavior in an e-commerce platform by evaluating the conversion funnel, identifying category-level purchase patterns, and measuring short-term customer retention, in order to uncover key drop-off points and actionable opportunities to improve conversions and early repeat purchases.**

Steps followed -

**Data preparation**

- Imported raw e-commerce events data
- Identified:
  - Scientific-notation IDs (handled correctly)
  - Blank category_code vs real NULLs
  - Duplicate event rows
  - Made event_time column to datetime and also made ne columns from it.
- Removed duplicates using window functions
- Created a **clean working table (events_d)**

```sql
select * from events_d
select count(*) from events
EXEC sp_help 'events';
SELECT event_type, COUNT(*) AS cnt
FROM events
GROUP BY event_type
ORDER BY cnt DESC;

SELECT
    user_id,
    product_id,
    event_time,
    event_type,
    user_session,
    COUNT(*) AS cnt
FROM events
GROUP BY user_id, product_id, event_time,event_type,user_session
HAVING COUNT(*) > 1;

---"The dataset contained empty strings instead of NULLs, so I standardized missing values by
---trimming whitespace and converting blanks to NULL to ensure accurate filtering and aggregation."

SELECT
    SUM(CASE WHEN user_id IS NULL OR LTRIM(RTRIM(user_id)) = '' THEN 1 ELSE 0 END) AS blank_user,
    SUM(CASE WHEN event_time IS NULL OR LTRIM(RTRIM(event_time)) = '' THEN 1 ELSE 0 END) AS blank_time,
    SUM(CASE WHEN product_id IS NULL OR LTRIM(RTRIM(product_id)) = '' THEN 1 ELSE 0 END) AS blank_product,
    SUM(CASE WHEN category_id IS NULL OR LTRIM(RTRIM(category_id)) = '' THEN 1 ELSE 0 END) AS blank_category,
    SUM(CASE WHEN category_code IS NULL OR LTRIM(RTRIM(category_code)) = '' THEN 1 ELSE 0 END) AS blank_category_code,
    SUM(CASE WHEN event_type IS NULL OR LTRIM(RTRIM(event_type)) = '' THEN 1 ELSE 0 END) AS blank_event_type,
    SUM(CASE WHEN price IS NULL THEN 1 ELSE 0 END) AS null_price
FROM events;

UPDATE events
SET
    user_id = NULLIF(LTRIM(RTRIM(user_id)), ''),
    product_id = NULLIF(LTRIM(RTRIM(product_id)), ''),
    category_id = NULLIF(LTRIM(RTRIM(category_id)), ''),
    category_code = NULLIF(LTRIM(RTRIM(category_code)), ''),
    event_type = NULLIF(LTRIM(RTRIM(event_type)), ''),
    event_time = NULLIF(LTRIM(RTRIM(event_time)), '');
```

```sql
    WITH ranked_events AS (
    SELECT *,
           ROW_NUMBER() OVER (
               PARTITION BY user_id, product_id, event_time, event_type, user_session
               ORDER BY event_time
           ) AS rn
    FROM events
)
SELECT
    user_id,
    event_type,
    product_id,
    category_id,
    category_code,
    price,
    event_time,
    brand,
    user_session
INTO events_d
FROM ranked_events
WHERE rn = 1;

ALTER TABLE events_d
ADD event_time_dt DATETIME;

UPDATE events_d
SET event_time_dt = TRY_CONVERT(
    DATETIME,
    REPLACE(event_time, ' UTC', ''),
    120
);

SELECT
    COUNT(*) AS total_rows,
    SUM(CASE WHEN event_time_dt IS NULL THEN 1 ELSE 0 END) AS failed_rows
FROM events_d;

ALTER TABLE events_d
ADD event_date DATE,event_hour INT;

UPDATE events_d
SET
    event_date = CAST(event_time_dt AS DATE),
    event_hour = DATEPART(HOUR, event_time_dt);
```

You now have:

•events_d → deduped, time cleaned, analysis-ready

Next phase = **Funnel Analysis**.

# FUNNEL DEFINITION & CONVERSION ANALYSIS

🎯 Business Question

Where do users drop off in the purchase journey?
We'll build a clean, defensible funnel.

## Level 1 → Core funnel

### 🔹 STEP 1: Define Funnel Steps

For this dataset, use:

| Step | event_type |
|---|---|
| Step 1 | view |
| Step 2 | cart |
| Step 3 | purchase |

### 🔹 STEP 2: Create User-Level Funnel Table

We want **one row per user**, showing if they reached each step.

```sql
WITH funnel_flags AS (
    SELECT
        user_id,
        MAX(CASE WHEN event_type = 'view' THEN 1 ELSE 0 END) AS viewed,
        MAX(CASE WHEN event_type = 'cart' THEN 1 ELSE 0 END) AS added_to_cart,
        MAX(CASE WHEN event_type = 'purchase' THEN 1 ELSE 0 END) AS purchased
    FROM events_d
    GROUP BY user_id
)
SELECT *
INTO user_funnel
FROM funnel_flags;

select * from user_funnel
```

## 🔹 STEP 3: Funnel Counts

```sql
SELECT
    COUNT(DISTINCT user_id) AS total_users,
    SUM(viewed) AS viewed_users,
    SUM(added_to_cart) AS cart_users,
    SUM(purchased) AS purchased_users
FROM user_funnel;
```

Results | Messages

| | total_users | viewed_users | cart_users | purchased_users |
|---|---|---|---|---|
| 1 | 407283 | 406863 | 36952 | 21304 |

| Stage | Users |
|---|---|
| Total Users | **407,283** |
| Viewed | **406,863** |
| Added to Cart | **36,952** |
| Purchased | **21,304** |

## 🔹 STEP 4: Funnel Conversion Rates

```sql
SELECT
    CAST(SUM(added_to_cart) * 1.0 / SUM(viewed) AS DECIMAL(5,2)) AS view_to_cart_rate,
    CAST(SUM(purchased) * 1.0 / SUM(added_to_cart) AS DECIMAL(5,2)) AS cart_to_purchase_rate,
    CAST(SUM(purchased) * 1.0 / SUM(viewed) AS DECIMAL(5,2)) AS view_to_purchase_rate
FROM user_funnel;
```

| | view_to_cart_rate | cart_to_purchase_rate | view_to_purchase_rate |
|---|---|---|---|
| 1 | 0.09 | 0.58 | 0.05 |

**Conversion Rates**
- **View → Cart: 9%**
- **Cart → Purchase: 58%**
- **View → Purchase: 5%**

## 🔹 STEP 5: Drop-Off Analysis

```sql
SELECT
    SUM(viewed) - SUM(added_to_cart) AS drop_after_view,
    SUM(added_to_cart) - SUM(purchased) AS drop_after_cart
FROM user_funnel;
```

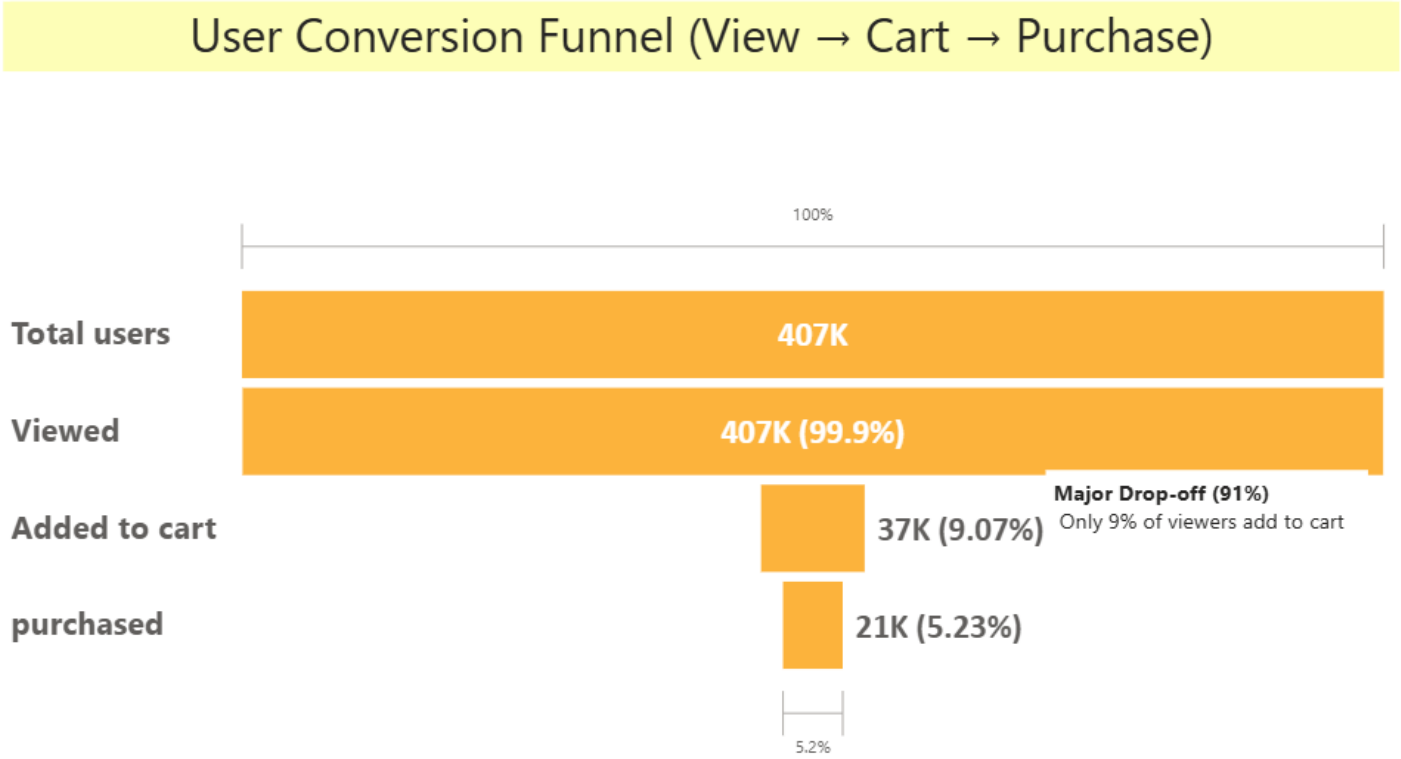
Results | Messages

| | drop_after_view | drop_after_cart |
|---|---|---|
| 1 | 369911 | 15648 |

| Drop-Off Point | Users Lost |
|---|---|
| After View | **369,911** |
| After Cart | **15,648** |

I built a user-level funnel using event data to measure drop-offs between view, cart, and purchase stages, helping identify where conversion losses were highest.”

# Business Interpretation




**1. Primary Bottleneck: View → Cart**

•~91% of users drop off before adding to cart

•Indicates:

- Browsing / comparison behavior
- Price sensitivity
- Insufficient product detail or value signaling

**2. Checkout Is Strong**

•58% cart → purchase is a healthy conversion

•Suggests:

- Low checkout friction
- High intent once users commit

"Funnel analysis showed that the largest user drop-off occurs before the cart stage, with only 9% of viewers adding items to cart. However, once users add products to cart, conversion to purchase is strong at 58%, indicating that optimization efforts should focus on pre-cart engagement rather than checkout flow."

# LEVEL 2 FUNNEL Segmentation

## 1 Funnel by Time

```sql
SELECT
    event_hour,
    COUNT(DISTINCT CASE WHEN event_type = 'view' THEN user_id END) AS viewers,
    COUNT(DISTINCT CASE WHEN event_type = 'purchase' THEN user_id END) AS purchasers
FROM events_d
GROUP BY event_hour
ORDER BY event_hour;
```

| Event Hour | Viewers | Purchasers | Conversion Rate (%) |
|---|---|---|---|
| **00:00 (12 AM)** | 6,021 | 224 | 3.72% |
| **01:00** | 6,086 | 232 | 3.81% |
| **02:00** | 6,677 | 274 | 4.10% |
| **03:00** | 8,939 | 388 | 4.34% |
| **04:00** | 12,496 | 533 | 4.27% |
| **05:00** | 17,062 | 830 | 4.86% |
| **06:00** | 22,514 | 1,203 | 5.34% |
| **07:00** | 26,380 | 1,385 | 5.25% |
| **08:00** | 28,366 | 1,559 | 5.50% |
| **09:00** | 29,415 | 1,651 | 5.61% |
| **10:00 (Peak Viewers)** | **29,899** | **1,700** | **5.69%** |
| **11:00** | 30,227 | 1,691 | 5.59% |
| **12:00 (Noon)** | 29,639 | 1,566 | 5.28% |
| **13:00** | 28,709 | 1,485 | 5.17% |
| **14:00** | 27,670 | 1,403 | 5.07% |
| **15:00** | 26,913 | 1,339 | 4.98% |
| **16:00** | 27,391 | 1,351 | 4.93% |
| **17:00** | 28,130 | 1,381 | 4.91% |
| **18:00** | 27,636 | 1,334 | 4.83% |
| **19:00** | 24,652 | 1,269 | 5.15% |
| **20:00** | 19,951 | 985 | 4.94% |
| **21:00** | 14,115 | 662 | 4.69% |
| **22:00** | 9,474 | 439 | 4.63% |
| **23:00** | 7,188 | 305 | 4.24% |

## 2 Funnel by Category

```sql
SELECT
    category_code,
    COUNT(DISTINCT CASE WHEN event_type = 'view' THEN user_id END) AS viewers,
    COUNT(DISTINCT CASE WHEN event_type = 'purchase' THEN user_id END) AS purchasers
FROM events_d
GROUP BY category_code
ORDER BY purchasers DESC;

WITH first_view AS (
```

91 % 2 0

Results Messages

| | category_code | viewers | purchasers |
|---|---|---|---|
| 1 | NULL | 129899 | 4779 |
| 2 | computers.components.videocards | 29033 | 3980 |
| 3 | electronics.telephone | 43848 | 2559 |
| 4 | computers.peripherals.printer | 19732 | 1519 |
| 5 | stationery.cartrige | 20907 | 1475 |
| 6 | computers.notebook | 13146 | 856 |
| 7 | computers.components.motherboard | 10090 | 749 |
| 8 | computers.components.cpu | 9909 | 693 |

Query executed successfully.

## 3 Time-to-Purchase Analysis

```sql
WITH first_view AS (
    SELECT user_id, MIN(event_time_dt) AS first_view_time
    FROM events_d
    WHERE event_type = 'view'
    GROUP BY user_id
),
first_purchase AS (
    SELECT user_id, MIN(event_time_dt) AS first_purchase_time
    FROM events_d
    WHERE event_type = 'purchase'
    GROUP BY user_id
)
SELECT
    AVG(DATEDIFF(MINUTE, fv.first_view_time, fp.first_purchase_time)) AS avg_minutes_to_purchase
FROM first_view fv
JOIN first_purchase fp
ON fv.user_id = fp.user_id;
```

91 % 2 0

Results Messages

| | avg_minutes_to_purchase |
|---|---|
| 1 | 2221 |

## Top 5 categories (by purchase)

| | category_code | viewers | purchasers |
|---|---|---|---|
| 1 | NULL | 129899 | 4779 |
| 2 | computers.components.videocards | 29033 | 3980 |
| 3 | electronics.telephone | 43848 | 2559 |
| 4 | computers.peripherals.printer | 19732 | 1519 |
| 5 | stationery.cartrige | 20907 | 1475 |

## 1 FUNNEL BY HOUR

**What the data shows**

•Traffic steadily increases from **early morning (5–6 AM)**
•**Peak activity: 10 AM – 1 PM**
•Purchases closely follow traffic pattern
•Night-time (0–4 AM) has **very low intent**

Although raw counts are higher mid-day:

•Evening hours (7–9 PM) show **strong purchase intent per viewer**
•Early morning viewers are low-intent browsers

## 2 FUNNEL BY CATEGORY

**High-performing categories (conversion-friendly)**

These categories convert well:

•Cameras ,computer components.videocards ,telephone,cartridge,computers peripherals.printer
•High-need
•Specification-driven
•Low impulse uncertainty

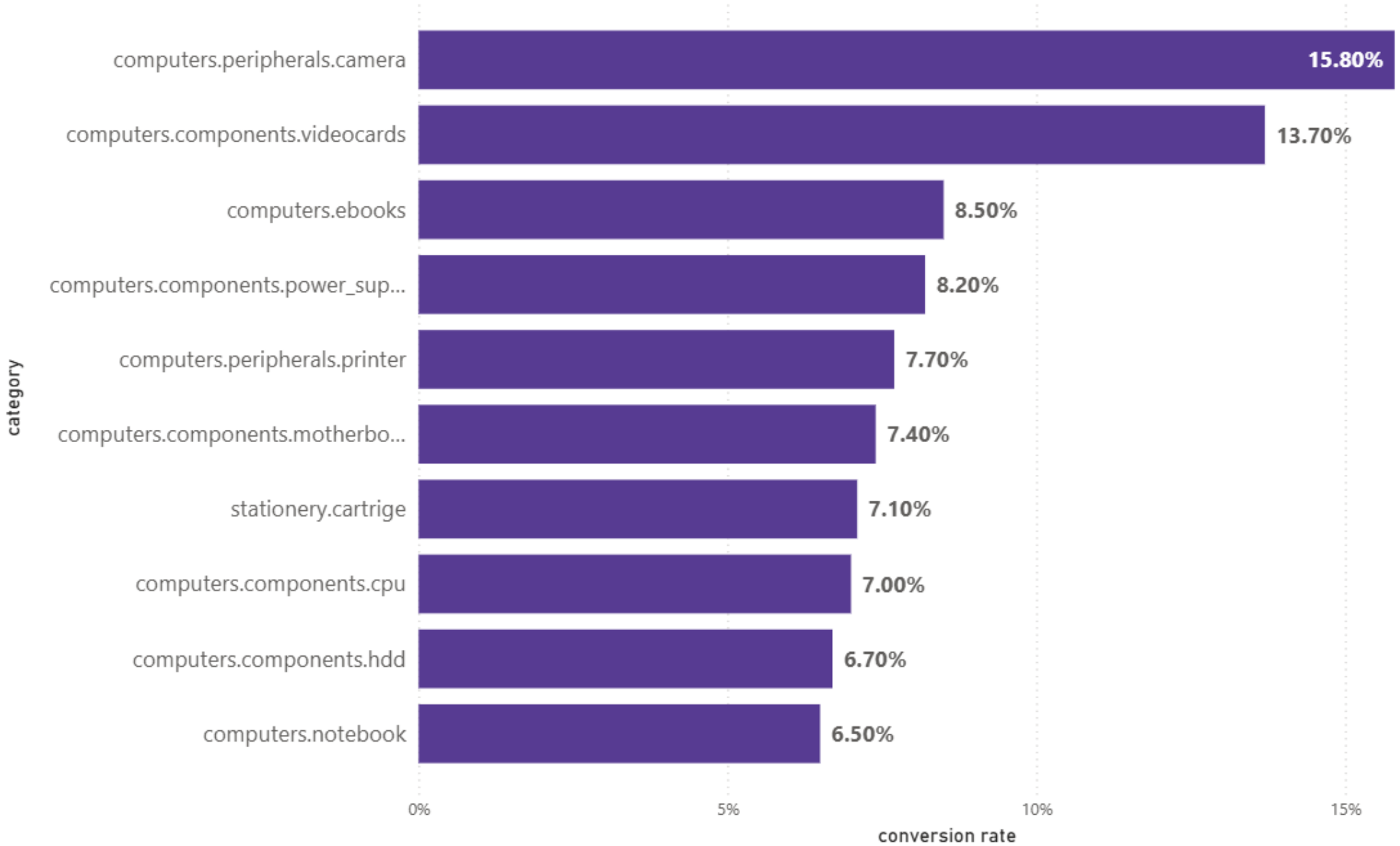
**Browsing-heavy, low-conversion categories**

electronics.video.tv ,electronics.audio.acoustic, electronics.audio.headphone

These are:

•High comparison
•Price-sensitive
•Often bought offline

# Top 10 Product Categories by Purchase Conversion Rate (%)



**Important insight: NULL category**

- **Highest traffic**
- Moderate purchases

Interpretation:

Many products lack category mapping; treating them as a separate group avoids data loss but requires cleanup for better merchandising insights.

**3 TIME TO PURCHASE INSIGHT**

⏱ **Average time to purchase**

- **2221 minutes ≈ 37 hours**

**What this means:**

- Users **do not purchase immediately**
- Indicates:
    - Research-driven behavior
    - Multi-session journey
    - Price comparison

## Overall insight

**Funnel analysis shows that the largest user drop-off occurs before the cart stage, with only 9% of viewers adding items to cart. However, once users add products to cart, conversion to purchase is strong at 58%, indicating that optimization efforts should focus on pre-cart engagement rather than checkout flow.**

- Purchase activity peaks between hours **9–13** based on the recorded event timestamp, indicating consistent intra-day activity patterns.
- High-involvement categories such as **notebooks and videocards** show lower conversion compared to accessory items, suggesting longer decision cycles.
- The average time from first product view to purchase is **~37 hours**, indicating comparison-driven buying behavior rather than impulse purchases.

## PART 2: RETENTION ANALYSIS

### 🎯 Objective

Understand:

- Do users return after first purchase?
- How many come back within 7 / 30 days?

```sql
----------------------------------------------------------part 2 ----------------------------------------
--retention analysis---


-- Step 1: First purchase per user
WITH first_purchase AS (
    SELECT
        user_id,
        MIN(event_time_dt) AS first_purchase_date
    FROM events_d
    WHERE event_type = 'purchase'
    GROUP BY user_id
),

-- Step 2: Any activity AFTER first purchase
repeat_activity AS (
    SELECT
        fp.user_id,
        DATEDIFF(
            day,
            fp.first_purchase_date,
            e.event_time_dt
        ) AS days_after
    FROM first_purchase fp
    JOIN events_d e
        ON fp.user_id = e.user_id
    WHERE e.event_time_dt > fp.first_purchase_date
)

-- Step 3: Retention calculation
SELECT
    COUNT(DISTINCT fp.user_id) AS total_buyers,
    COUNT(DISTINCT CASE WHEN ra.days_after <= 7 THEN fp.user_id END) AS retained_7d,
    COUNT(DISTINCT CASE WHEN ra.days_after <= 30 THEN fp.user_id END) AS retained_30d,
    ROUND(
        1.0 * COUNT(DISTINCT CASE WHEN ra.days_after <= 7 THEN fp.user_id END)
        / COUNT(DISTINCT fp.user_id), 2
    ) AS retention_7d_rate,
    ROUND(
        1.0 * COUNT(DISTINCT CASE WHEN ra.days_after <= 30 THEN fp.user_id END)
        / COUNT(DISTINCT fp.user_id), 2
    ) AS retention_30d_rate
FROM first_purchase fp
LEFT JOIN repeat_activity ra
    ON fp.user_id = ra.user_id;
```

62 % ⌄ ⊗ 2 ⚠ 0 ↑ ↓

Results | Messages

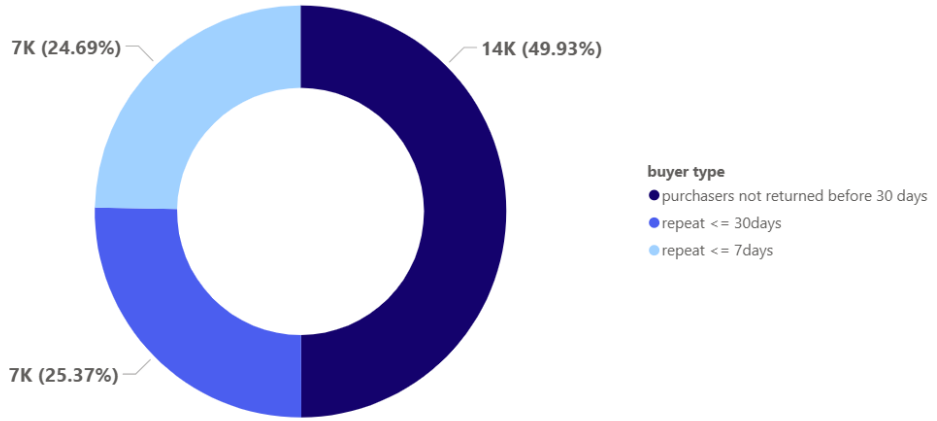
| | total_buyers | retained_7d | retained_30d | retention_7d_rate | retention_30d_rate |
|---|---|---|---|---|---|
| 1 | 21304 | 12824 | 13107 | 0.600000000000 | 0.620000000000 |

# Retention Results  and insights

**Retention Results (Final)**
- **Total buyers:** 21,304
- **7-day repeat buyers:** 6,985 (**33%**)
- **30-day repeat buyers:** 7,178 (**34%**)

- Out of 21,304 buyers, **34% made a repeat purchase within 30 days**, while **6 not return within the 30-day window**.
- Repeat purchases are **heavily concentrated in the first 7 days**, with very lin additional returns between days 8–30.
- This indicates that **short-term repeat behavior happens quickly**, and buyer not return early are less likely to repurchase in the near term.

Customer Retention Breakdown (30-Day Window)



## OVERALL INSIGHT

•The main conversion bottleneck lies in the **pre-cart stage**, where most users drop off during product evaluation.
•Buyers who do convert exhibit **deliberate purchasing behavior**, taking ~37 hours to complete a purchase.
•**Short-term retention is time-sensitive**: customers who repurchase typically do so within the first week, while delayed repeat behavior is rare within 30 days.
•Business efforts should therefore focus on:

- Improving pre-cart engagement to increase first-time conversion
- Activating **early post-purchase nudges within 7 days**, rather than assuming long-term loyalty patterns

# Actionable Business Recommendations

**1 Fix the biggest funnel leak: View → Cart (91% drop-off)**

**What the business should do**

**Improve product evaluation** at the moment of browsing:

- Add *clear specs comparison* for similar products
- Show *price breakdown & discounts upfront*
- Display *delivery timelines* early (before cart)

Introduce **"Add to cart nudges"**:

- "Low stock" or "Price changed recently" indicators
- Quick-add buttons for accessories

📊 **Why (data-backed):**

Only **9% of viewers** add to cart, but **58% of those convert**.

So every improvement in view → cart directly multiplies revenue.

**2 Time post-purchase engagement around the 37-hour decision cycle**

**What the business should do**

**Within 24–48 hours of first purchase**:

- Send **personalized follow-up emails / push notifications**
- Recommend **complementary or accessory products**
- Example:
  - Laptop → mouse, keyboard, adapter
  - Videocard → power supply, cooling fan

**don't wait a week** — the window is **1–2 days**.

## 3️⃣ Exploit the 7-day retention window (most critical)

**What the business should do**

- **Day 2–3:** Reminder / cross-sell message
- **Day 5–7:** Limited-time incentive (small discount, free delivery)
- Stop aggressive nudging **after Day 7**

📊 **Why:**

- **33% repeat within 7 days**
- Almost **no additional users return after that**
- ROI drops sharply after Week 1

This is **classic short-term retention behavior**.

## 4️⃣ Different strategies by product type

**High-involvement products (notebooks, videocards)**

- Add:
  - Buying guides
  - Comparison tables
  - Extended warranty visibility
- Avoid aggressive discounts immediately

**Low-cost accessories**

- Push **bundles** and **add-ons**
- Use urgency messaging
- Target for fast repeat purchases

# CONCLUSION

“The data shows that revenue impact is driven by improving pre-cart engagement and activating buyers quickly after purchase. Most repeat purchases happen within 7 days, and the average decision cycle is ~37 hours, so retention strategies should focus on timely post-purchase nudges rather than long-term loyalty campaigns.”