E-Commerce Website Usage

By Faury A. Abreu
09/02/2023

Objective:

The Service Optimization and The Marketing team of an E-commerce company wants to understand the most important metrics for the website visitors using the past 6-months data collected from the User Behavior Tracking (UBT) reports. The insights found in this analysis will lead to data driven decisions to adapt the services and marketing to customer preference and consequently increase the number of subscriptions.

Visualizations:

Tableau Dashboard

I highly recommend looking to the dashboard during your journey through this report.

Data information:

- The dataset contains 6-month data (Sep 2020 – Feb 2021) from a large electronics online store.
- The data in Events table is driven by (UBT) reports from the official website.
- Each row in the table represents an event. All events are related to products and users. Each event is like many-to-many relation between products and users.
- Source: Data Source Data collected by Open CDP from open source UBT data.

Table Structure:

Property	Description
event_time	Time when event happened at (in UTC).
event_type	Only one kind of event: purchase.
product_id	ID of a product
category_id	Product's category ID

category_code	Product's category taxonomy (code name) if it was possible to make it. Usually present for meaningful categories and skipped for different kinds of accessories.
brand	Down cased string of brand name. Can be missed.
price	Float price of a product. Present.
user_id	Permanent user ID.
user_session	Temporary user's session ID. Same for each user's session. Is changed every time user come back to online store from a long pause.

- **Event types:**

  - **view** - a user viewed a product.
  - **cart** - a user added a product to shopping cart.
  - **remove_from_cart** - a user removed a product from shopping cart.
  - **purchase** - a user purchased a product.
  - **Time_on_website**: time spend by user on the website.

Business Tasks:

1) Find the most popular device category based on sales.
2) Distribution of all sales based on different event type.
3) Distribution of all event_type over daytime.
4) Top 10 brands by sales.
5) Distribution of website activity over daytime.

Data Preparation/Formatting:

Since it is a relatively large table, I will use wizard, then input the data using a LOAD based query. Note that the file should be in an exact location so MySQL can detect it without error 1290. The indicated location is at the “Upload” file within the MySQL program file.

```
LOAD DATA INFILE "/datapath"
INTO TABLE events
FIELDS TERMINATED BY ','
ENCLOSED BY '"' LINES TERMINATED BY '\n'
IGNORE 1 ROWS;
```

Also, in order to make the exploratory analysis process clearer and more concise, it is recommended to split the datetime field into year, month, day, and time. To do this I modified the table by converting the data to the correct datatypes then I altered the table to create new variables and input the data.

```sql
/*split event_time*/
alter table events rename column event_time to event_date;

alter table events add column event_time_zone varchar(4) after event_date;
alter table events add column event_year int after event_date;
alter table events add column event_month int after event_year;
alter table events add column event_day int after event_month;
alter table events add column event_time time after event_day;

-- INPUT DATA INTO NEW COLUMNS FROM EVENT_DATE:
SET SQL_SAFE_UPDATES = 0;
update events set event_time_zone = right(event_date,3);
update events set event_year = left(event_date,4);
update events set event_month = mid(event_date,6,2);
update events set event_day = mid(event_date,9,2);
update events set event_time = mid(event_date,12,8);
SET SQL_SAFE_UPDATES = 1;
```

Now that all necessary changes have been executed, its time for the analysis:

Exploratory Analysis:

For the exploratory analysis I decided to start by exploring the factors that affected the sales by filtering the data and applying specific conditions:

1) Sales by Month:

Query

```sql
/*sales by month*/
SELECT
    event_year,
    monthname(event_date) as month,
    COUNT(event_type) AS net_sales,
    ROUND(SUM(price), 3) AS total_sales
FROM
    events
WHERE
    event_type = 'purchase'
GROUP BY event_year,month
ORDER BY total_sales desc;
```

Output

event_year	month	net_sales	total_sales
2021	January	8315	1488410.57
2021	February	7643	1376815.54
2020	December	6531	813339.87
2020	November	7574	787886.24
2020	October	6276	562590.39
2020	September	1007	96353.01

Observations:

- The month with most sales was January 2021, with $ 1,488,410.57 in sales.
- The month with less sales was September 2020 with $ 96,353.01 in sales.
- Sales show a growth trend, except for February 2021, with a drop in sales of $ 111,595.03.

2) Time of the day with most views on the website:

Query	Output

```sql
/*time most views*/
SELECT
    HOUR(event_time) AS time_hour_UTC, COUNT(event_type) AS views
FROM
    events
WHERE
    event_type = 'view'
        OR event_type = 'cart'
GROUP BY time_hour
ORDER BY views DESC;
```

time_hour	views
11	51103
10	50503
12	50228
9	50077
17	49678
13	48652
18	48643
16	47983
8	47725
14	47177
15	46745
7	44968
19	43710
6	37913
20	35462
5	28595
21	25035
4	20202
22	16495
3	14694
23	11861
2	10772
0	9933
1	9629

Observations:

- The website reports the most activity at 11:00 AM UTC.
- Least activity is reported at 1:00 AM UTC.

3) Time of the day with most purchases:

Query	Output

```sql
/* daytime with the most sales*/

SELECT
    HOUR(event_time) AS time_hour_UTC,
    COUNT(event_type) AS purchases
FROM
    events
WHERE
    event_type = 'purchase'
GROUP BY time_hour_UTC
ORDER BY purchases DESC;
```

time_hour_UTC	purchases
10	2535
11	2496
9	2474
12	2282
8	2215
13	2171
14	2160
17	2085
16	2019
7	2000
15	1957
18	1950
19	1934
6	1691
20	1533
5	1165
21	1041
4	788
22	701
3	575
23	482
2	403
0	361
1	328

Observations:

- Time of the day with most purchases is during 10:00 AM UTC.
- Least sales reported during 1:00 AM
- In general terms, from 8:00 AM to 2:00PM is the period with most sales.

4) Top 10 brands by sales

Query

```sql
/*top 10 brands by sales*/

SELECT
    brand, count(brand) as times_sold, round(SUM(price),3) AS total_Sales
FROM
    events
WHERE
    event_type = 'purchase'
GROUP BY brand
ORDER BY times_sold desc
limit 10;
```

Output

brand	times_sold	total_Sales
	9501	594729.99
msi	1788	643492.34
gigabyte	1741	556183.04
asus	1261	330147.11
palit	1077	484210.58
canon	1045	137964.79
amd	1020	191987.4
sapphire	917	306192.94
samsung	758	54624.5
sirius	742	17103.55

Observations:

- MSI was the brand with most sales with $643,492.34.
- Even though there is a considerable amount of sales with no specified brand $594,729.99, the Gigabyte and Palit were the second and third most sold brands with $556,183.04 and $484,210.58 respectively.

- The brand that sold the most times in terms of product was MSI (1788 times sold), followed by Gigabyte (1741 times sold) and Asus (1261 times sold)

5) Demand for items. Top 10 Bestselling products:

Query

```
/*bestselling products*/

SELECT
    category_code,
    COUNT(category_code) AS times_purchased,
    ROUND(SUM(price), 3) AS total_sales
FROM
    events
WHERE
    event_type = 'purchase'
        AND category_code != ''
GROUP BY category_code
ORDER BY times_purchased DESC
LIMIT 10;
```

Output

category_code	times_purchased	total_sales
computers.components.videocards	6888	2604764.8
electronics.telephone	4119	150915.69
stationery.cartrige	2739	103595.99
computers.peripherals.printer	2557	364566.8
computers.notebook	1297	63604.79
computers.components.motherboard	1266	102871.16
computers.components.cpu	1213	219168.98
electronics.audio.acoustic	848	84724.8
electronics.tablet	819	53347.43
auto.accessories.player	685	81558.66

Observations:

- Computer products are predominating the top.
- Being the three bestselling products: Video cards $ 2,604,764.8, printers $ 364,566.8, and CPU $ 219,168.98
- Most purchased products are Video cards (6888 purchases), telephones (4119), and cartage (2739 purchases)

6) Frequency of purchase by customers

Query	Output

```sql
/*Client Retention and Frequency */
CREATE TABLE purchase_frequency  SELECT user_id, COUNT(user_id) AS times_purchased
FROM events
WHERE event_type = 'purchase' and event_type  != ''
GROUP BY user_id
ORDER BY times_purchased DESC;

ALTER TABLE purchase_frequency ADD COLUMN ranges VARCHAR(10);
SET SQL_SAFE_UPDATES = 0; -- SET  safe mode off to make changes without specify primary
UPDATE purchase_frequency SET ranges=('1') WHERE times_purchased = 1;
UPDATE purchase_frequency SET ranges=('2-10') WHERE times_purchased BETWEEN 2 AND 10;
UPDATE purchase_frequency SET ranges=('10-20') WHERE times_purchased BETWEEN 10 AND 20;
UPDATE purchase_frequency SET ranges=('20-30') WHERE times_purchased BETWEEN 20 AND 30;
UPDATE purchase_frequency SET ranges=('30-40') WHERE times_purchased BETWEEN 30 AND 40;
UPDATE purchase_frequency SET ranges=('40-50') WHERE times_purchased BETWEEN 40 AND 50;
UPDATE purchase_frequency SET ranges=('50-60') WHERE times_purchased BETWEEN 50 AND 60;
SET SQL_SAFE_UPDATES = 1;

SELECT ranges, count(ranges) as no_of_customers
FROM purchase_frequency
group by ranges
order by ranges;
```

ranges	no_of_customers
1	13598
10-20	118
2-10	7566
20-30	14
30-40	3
40-50	4
50-60	1

Observations:

- The most clients only made one purchase being the ~64% of total clients.
- ~35% of clients made 2 to 10 purchases.
- Customer recurrence can be better.

Diagnosis analysis

For the Diagnosis analysis ill explore some factor in order to find the effects of the huge difference in frequency of purchase by customers. I'm sure that the key is on the variables that contain customer's behavior data, to find insights regarding to purchase rate. The main variable with this data is event_type, since it defines the actions made by the visitors on the website.

7) How many viewed products where actually purchased?

Query	Output

```sql
/*purchasing rate from viewed products*/
SELECT
    COUNT(a.event_type) AS purchases,
    b.views,
    ROUND(COUNT(a.event_type) / (b.views + COUNT(a.event_type)) * 100,
            1) AS percentage
FROM
    events a,
    (SELECT
        COUNT(event_type) AS views
    FROM
        events
    WHERE
        event_type = 'view'
    GROUP BY event_type) b
WHERE
    a.event_type = 'purchase'
GROUP BY b.views;
```

purchases	views	percentage
37346	793748	4.5

Observations:

- Only the 4.5% of viewed products where actually purchased.

8) What percentage of products in client's carts where actually purchased?

Query

```sql
/*purchasing rate from viewed products*/

SELECT
    COUNT(a.event_type) AS purchases,
    b.cart,
    ROUND(COUNT(a.event_type) / (b.cart + COUNT(a.event_type)) * 100,
            1) AS percentage
FROM
    events a,
    (SELECT
        COUNT(event_type) AS cart
    FROM
        events
    WHERE
        event_type = 'cart'
    GROUP BY event_type) b
WHERE
    a.event_type = 'purchase'
GROUP BY b.cart;
```

Output

purchases	cart	percentage
37346	54035	40.9

Observations:

- 40.9% of products in carts where actually purchased.

Conclusions:

1) In order to gain more clients, it is recommended to **run the *marketing campaign* during January, February and March**, given those are the months with most sales.
2) To complement the marketing campaign**, implement a "*Countdown Timer*".** Countdown timers allow to improve conversions on special occasions (offers, Black Friday, Christmas...) when there is a limited time to buy, or constantly showing the remaining time to receive the order the next day, as Amazon does.
3) Implement **"*Cross-Selling*" purchase suggestion on the website**. Cross-selling consist in offer complementary products to those a customer is purchasing. For example, if customer is adding a Motherboard to the cart, suggest good options for CPU or Cooler fans, this way the client will be engaged to the suggestions.
4) To increase the percentage of products purchased from the cart, there are several options:
    a. **Pay attention to the *reviews and ratings*.** Many customers change their minds when products have bad reviews, in which case, products with bad reviews should be reduced or eliminated from inventory once they are sold.
    b. **Apply a *CRO (Conversion Rate Optimization).*** A CRO is a set of techniques or strategies that aim, as its name suggests, to improve the conversion rate and increase sales in ecommerce, optimizing different elements and processes. One of those strategies is *Optimization of forms*. Up to 34% of online buyers say they have abandoned a purchase because the form or checkout process was too long. Improve this process or even choose to implement 'Pay with PayPal' or 'Pay with Google' buttons that allow users to automatically enter their data and check out in a single click.

5) **More availability of top selling products** like MSI and Gigabyte, so they don't get out of stock.