Contents

Data Preparation

By analysing the dataset, an anomaly was found where different *event_dates* were recorded for the same session having same *page_type* and *event_type*. If there were only 2 entries for each session, one could think of them being session start and end time but in some cases, there were 100 or more entries (example below) so it seems to be an error in data collection. To our conclusions, only the first entry of each session gives us the correct information while others are misleading.

event_date	session	user	page_type	event_type	product
09/10/2022 18:40	7160321411091116049s	9697135938868880738u	product_page	page_view	0
09/10/2022 20:37	7160321411091116049s	9697135938868880738u	product_page	page_view	0
09/10/2022 22:40	7160321411091116049s	9697135938868880738u	product_page	page_view	0
09/10/2022 21:50	7160321411091116049s	9697135938868880738u	product_page	page_view	0
09/10/2022 19:10	7160321411091116049s	9697135938868880738u	product_page	page_view	0
09/10/2022 17:42	7160321411091116049s	9697135938868880738u	product_page	page_view	0

Figure 1. Repeated entries for the same session at different times (Truncated view)

So, all the repeated entries were removed with preprocessing in python, keeping only the first entry.

Moreover, following calculations were performed to gather all the necessary information before moving to the visualization step.

- Identifying the first session for each user (time, page visited, and action taken in first session).
- Identifying the total number of sessions for each user.
- Identifying the total orders made by each user.

All the information is merged in one dataset to be visualized in PowerBI.

Insights From Visualization

Following key matrices are used to get the insights from the data:

- Website traffic (total visitors on the website)
- Bounce rate (Visitors who only visited the pages)
- ACR, Add to Cart Rate (Visitors who added at least one product to the cart)
- Conversion rate (Visitors who ordered at least one product)

Website Traffic	Bounce Rate	ACR	Conversion Rate
288088	99.8%	3.5%	2.5%

Table 1. Key matrices for users' behaviour on website

If we see the ultimate conversion rate, 2.5% seems low compared to step one but it's around **73%** from the previous step (ACR) which shows that 73% of the users who added the product to cart ended up purchasing compared to 2.5% of those who viewed the pages.
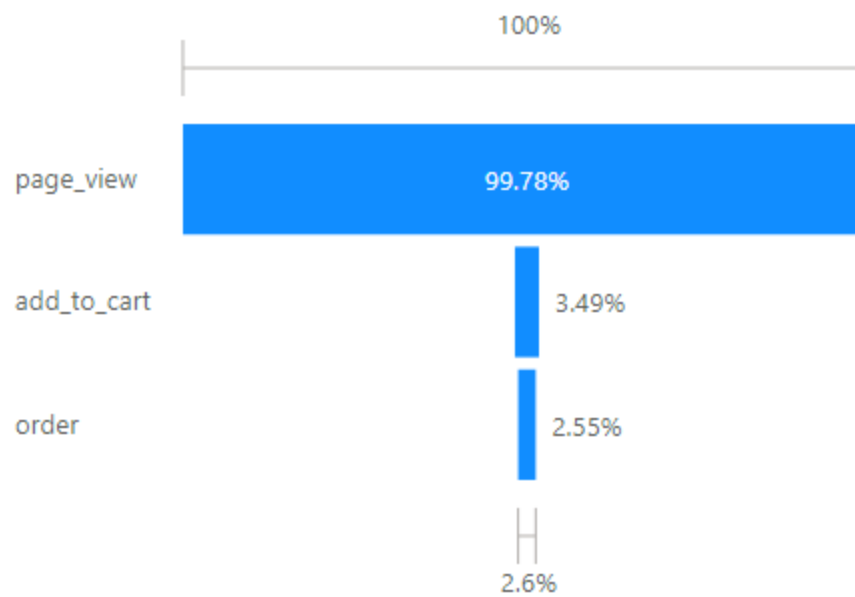
Visitors Bounce & Conversion Rate

Figure 2. Customer Purchase Funnel

The chart below shows that average number sessions on page is directly related to the actions taken. i.e., More sessions on the order page means more products purchased.

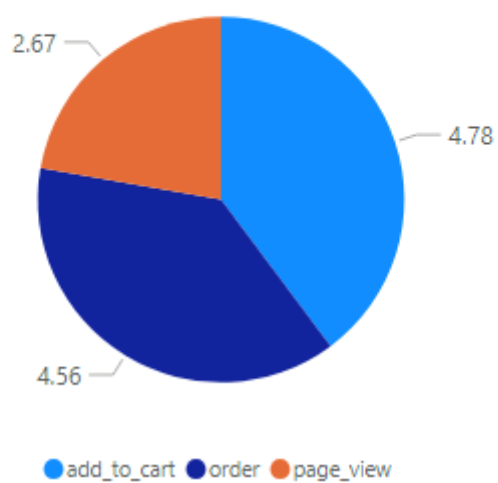## Average Sessions by Visitors for Action ...

Figure 3. Average number of session of visitors for each action taken

How Visitors Journey affected by the first page visited?

It was concluded that visitors who landed on the product page in the first session are most likely to order the product while the ones landed on the *search_listing_page* are less likely to order.
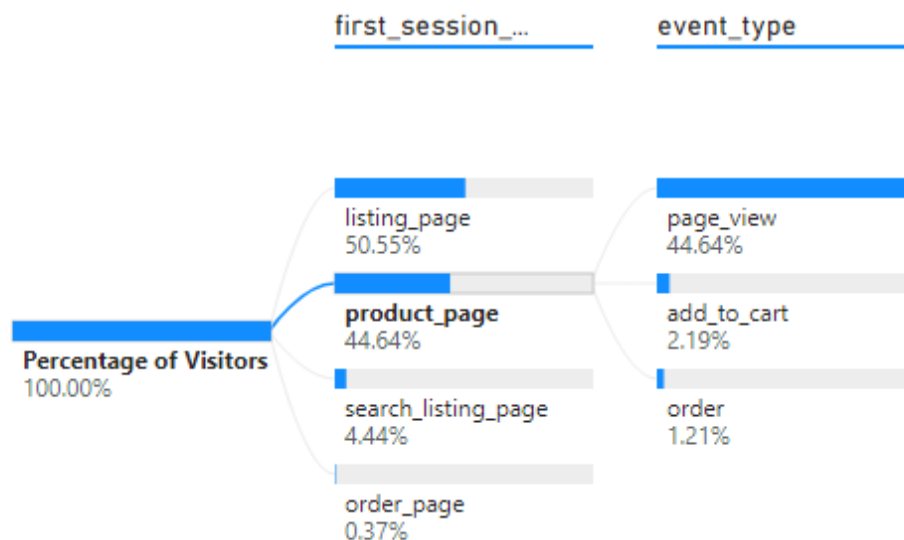

Figure 4. Visitors' journey affected by the first page visited

Custom Marketing Funnel

A basic marketing funnel has these 4 steps.

- Awareness: Customer visits the listing or product page

- Consideration: Product is added to cart

- Conversion: Decision step where product is purchased

- Loyality: Repeated purchases

Figure 5. Basic Marketing Funnel

Analysing this funnel with tracking gives us the following information:

Awareness: Although it's not evident from the data how users arrive at view pages (such as search_listing_page, listing_page, and product_page), it appears that the awareness phase of current marketing strategy is effective, particularly in terms of generating traffic.

Consideration: The product description pages, including details, images, and examples, assist users in considering this as an option and adding items to their cart. The low ACR rate of only 3.5% underscores the importance of concentrating on this stage of the marketing funnel.

Conversion: The user's decision to complete a purchase is influenced by factors such as product reviews, ratings, and the chosen shipping method and delivery time. In this instance, the conversion rate after the *add_to_cart* step stands at a commendable **73%**, indicating the effective performance of this stage in the marketing funnel.

Loyalty: Analysing customer post-purchase behaviour provides valuable insights into the efficacy of this stage. It reveals how customers engage with feedback surveys/emails and whether they return to make additional purchases on the website. The fact that only 3.5% of customers made multiple purchases underscores the need for targeted improvements in this segment of the marketing funnel.

Additional Information Needed for effective marketing Strategy

Following additional information would be needed to understand the effectiveness of the current strategy:

- Product types

- Pricing range of products

- Demographics of the visitors (age, gender, location, etc.)

- Time spent by visitor on each session

SQL Queries for Part 2

Note: Queries are performed on the dataset after removing the duplicates as explained in section 1.

- SQL query that will return the number of clients by day that only viewed products in their first session:

```
WITH FirstSessionViews AS (
   SELECT
     user, MIN(event_date) AS first_session_date
   FROM
     CleanedData
   WHERE
     event_type = 'page_view' AND
     page_type = 'product_page'
   GROUP BY
     user
)

SELECT
   DATE(f.first_session_date) AS day,
   COUNT(DISTINCT f.user) AS clients_count
FROM
   FirstSessionViews f
GROUP BY
   day
ORDER BY
   day;
```

- SQL query that will return the number of clients by day that added only one product to the basket:

```
WITH TotalAddedToCart AS (
   SELECT
     *, COUNT(user) as items_in_basket
   FROM
     CleanedData
   WHERE
      event_type = 'add_to_cart'
   GROUP BY
     user
)

SELECT
   DATE(t.event_date) AS day,
   COUNT(DISTINCT t.user) AS clients_count
FROM
   TotalAddedToCart AS t
WHERE items_in_basket=1
GROUP BY
   day
ORDER BY
   day;
```

- SQL query that will return the number of clients by day that placed an order within two days' time after the first session

```sql
WITH FirstSessionViews AS (
   SELECT
     user,
     MIN(event_date) AS first_session_date
   FROM
     CleanedData
   GROUP BY
     user
),
OrdersWithinTwoDays AS(

   SELECT
     DATE(c.event_date) as order_date,
       f.user
   FROM
     CleanedData c
   JOIN
     FirstSessionViews f
   ON c.user = f.user
   WHERE
       event_type='order'
       AND (JULIANDAY(c.event_date) -
JULIANDAY(f.first_session_date)) <=2
)
SELECT order_date,
COUNT(DISTINCT user) AS clients_ordered_within_two_days
FROM OrdersWithinTwoDays
GROUP BY
   order_date
ORDER BY
   order_date;
```

SQL Queries for Abnormal Behaviour

There are a few ways to identify the abnormal behaviour of the visitors.

Below are a few examples of with SQL queries.

Case 1:

```sql
SELECT user, COUNT(*) as views
FROM CleanedData
WHERE event_type = 'page_view' AND product IS NOT NULL
GROUP BY user
HAVING COUNT(*) > 5 AND user NOT IN (
    SELECT DISTINCT user
    FROM CleanedData
    WHERE event_type IN ('add_to_cart', 'order')
```

This query returns users who have viewed products more than 5 times but have never added a product to their cart or made a purchase. This could indicate indecisiveness or comparison shopping.

Case 2:

```
SELECT user, COUNT(*) as cart_additions
FROM CleanedData
WHERE event_type = 'add_to_cart'
GROUP BY user
HAVING COUNT(*) > 3 AND user NOT IN (
    SELECT DISTINCT user
    FROM CleanedData
    WHERE event_type = 'order'
)
```

This query returns users who have added products to their cart more than 3 times but have never made a purchase. This could indicate window shopping or hesitation to complete the purchase.

Case 3:

```
SELECT DISTINCT user
FROM CleanedData
WHERE event_type = 'order' AND user NOT IN (
    SELECT DISTINCT user
    FROM CleanedData
    WHERE event_type = 'page_view' AND page_type =
'product_page'
)
```

This query returns users who have made a purchase without viewing the product page. This could indicate impulse buying or familiarity with the product.