Copee

Brandon Savero

https://github.com/Brandonsvro

Pacmann

Website Analysis

Business Intelligence Project

PostgreSQL

tableau

Copee

Pacmann

Overview

The company Copee launched its e-commerce website on March 19, 2012. To build market awareness, they ran advertisements across several digital channels. Over the coming period, they plan to analyze the traffic and performance of their website.

As a data analyst, the manager have assigned me to handle the required analyses on an ongoing basis.

Objective

- Analyze key metrics to assess overall platform effectiveness.
- Investigate traffic trends across various sources, devices, and campaigns, identifying areas of growth or concern.
- Deliver recommendations to improve user engagement, reduce bounce rates, and enhance the user journey, particularly for key traffic sources and devices.
- Design an intuitive and detailed dashboard to visualize traffic patterns, user behavior, and the performance of implemented strategies for stakeholders.

Next

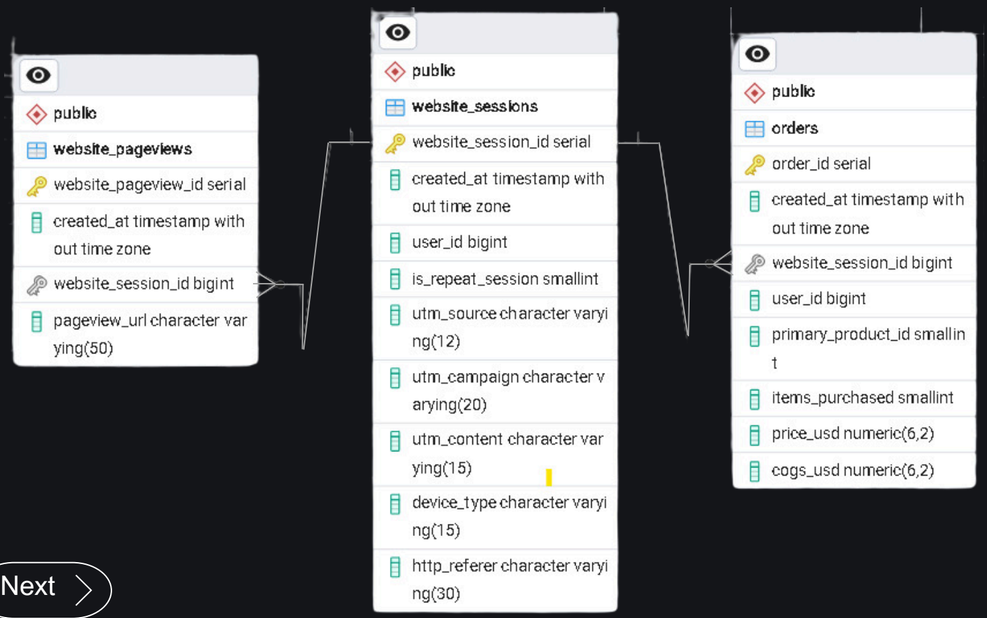
Copee

Pacmann

Data Overview

This is an internal dataset from Pacmann AI related to e-commerce websites. This dataset includes data on website traffic and user transactions from **March 19, 2012 to December 31, 2012**. This data consists of three tables, namely:

- **website_sessions:** Contains session data for each user
- **website_pageviews:** Contains page data for each session
- **orders:** Contains transaction data for each session



Next

Detail of each table

website_sessions (62264 rows)

- **website_session_id:** user's website session id
- **created_at:** the time the session was created
- **utm_source:** the channel source of the session
- **utm_campaign:** the ads source of the session
- **user_id:** user id
- **utm_content:** the ads content source of the session
- **device_type:** session device type
- **http_referer:** The URL referred the session
- **is_repeat_session:** identifier for user's repeat sessions (0/1)

website_pageviews (135.987 rows)

- **website_pageview_id:** session's website page id
- **website_session_id:** user's website session id
- **created_at:** the time the page was accessed
- **pageview_url:** description of the page accessed

orders (2.574 rows)

- **order_id:** order id
- **created_at:** the time the order was created
- **website_session_id:** user's website session id
- **primary_product_id:** primary product id
- **user_id:** user id
- **items_purchased:** quantity of items purchased
- **price_usd:** price of items purchased
- **cogs_usd:** cost of items purchased

56.399 users | 62.264 sessions | 7 pages | 1 product | 2.574 trx

Copee

Pacmann

The Data Cleaning & Transformation step focuses on identifying null values, duplicate data, and performing necessary transformations to support future analysis.

The Data Cleaning and Transformation process was performed using Python due to its efficiency in handling these tasks.

Data Analysis Steps

Data Cleaning and Transformation

Initial Month Analysis

Post-Improvement Analysis

A/B Testing Result

Funnel & Traffic Peak Time Analysis

Next

Copee

Null Values Identification

website_sessions table

```
website_session_id       0
created_at               0
user_id                  0
is_repeat_session        0
utm_source            5705
utm_campaign          5705
utm_content           5705
device_type              0
http_referer          2759
dtype: int64
```

Null values were detected only in the website_session table, specifically in the columns utm_source, utm_campaign, utm_content, and http_referer.

```
      website_session_id          created_at  user_id  is_repeat_session  \
4140                4224 2012-04-16 13:10:16     3313                  1
4282                4417 2012-04-18 10:06:47     1910                  1
5344                5539 2012-04-30 12:03:16     2301                  1
5562                5715 2012-05-01 16:51:28     5073                  1
5679                5873 2012-05-02 17:08:26     5691                  0

     utm_source utm_campaign utm_content device_type http_referer
4140        NaN          NaN         NaN     desktop          NaN
4282        NaN          NaN         NaN      mobile          NaN
5344        NaN          NaN         NaN     desktop          NaN
5562        NaN          NaN         NaN     desktop          NaN
5679        NaN          NaN         NaN     desktop          NaN
```

Null values refer to direct users

Discussions with the team indicated that the null values in utm_source, utm_campaign, utm_content, and http_referer refer to users who accessed the website directly, rather than through a specific campaign.

To ensure that the analysis results are not affected by null values, it is necessary to transform the data by replacing null values with more relevant values.

Transformed null value data

```
      website_session_id          created_at  user_id  is_repeat_session  \
4140                4224 2012-04-16 13:10:16     3313                  1
4282                4417 2012-04-18 10:06:47     1910                  1
5344                5539 2012-04-30 12:03:16     2301                  1
5562                5715 2012-05-01 16:51:28     5073                  1
5679                5873 2012-05-02 17:08:26     5691                  0

     utm_source utm_campaign utm_content device_type http_referer
4140     direct  noncampaign  noncontent     desktop   nonreferer
4282     direct  noncampaign  noncontent      mobile   nonreferer
5344     direct  noncampaign  noncontent     desktop   nonreferer
5562     direct  noncampaign  noncontent     desktop   nonreferer
5679     direct  noncampaign  noncontent     desktop   nonreferer
```

*Duplicate data was not found in the three tables, so the analysis process can continue to the next step.

Next

Copee

Pacmann

One month after the website's launch, **the manager wants to evaluate its initial performance by analyzing the weekly traffic trends, top traffic sources, and the analysis of those sources**.

Data Analysis Steps

Data Cleaning and Transformation

Initial Month Analysis

Post-Improvement Analysis

A/B Testing Result

Funnel & Traffic Peak Time Analysis

Next

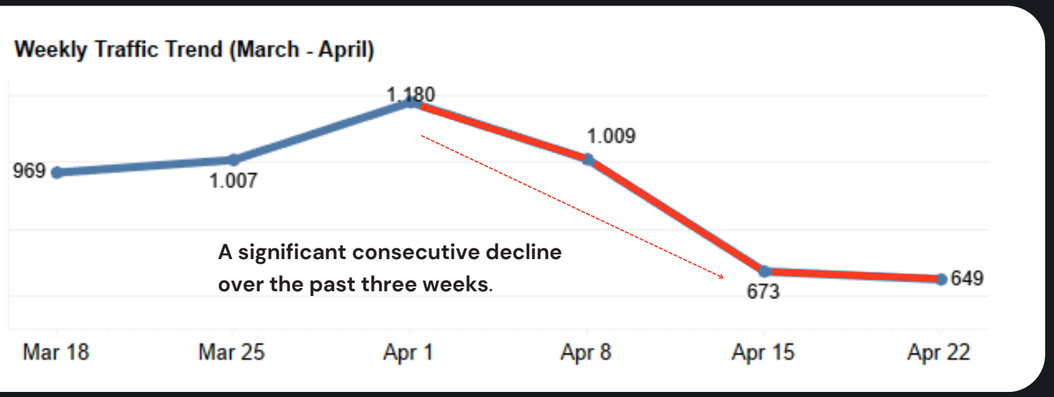
Copee

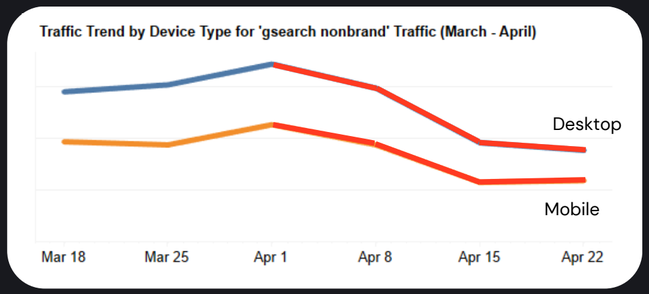
Initial Month Traffic Trend

Traffic dropped significantly during the last three weeks of April (avg -17%).

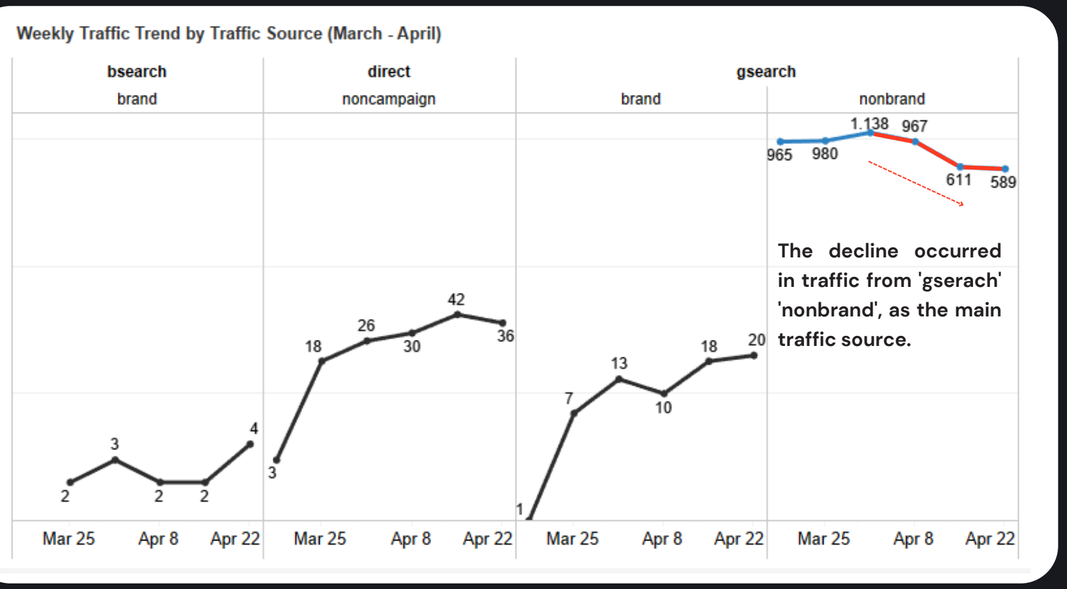
A decline was identified in traffic originating from 'gsearch nonbrand' (avg -18%).

This drop affects all types of user devices for traffic from 'gsearch nonbrand'.





The traffic drop, which was limited to 'gsearch nonbrand', was most likely due to issues with the campaign on this channel, which caused a significant drop in overall user traffic. This initial assumption should be validated by the relevant teams.



Suggestions

Evaluated the traffic source 'gsearch nonbrand' on the effectiveness of the campaign in April.

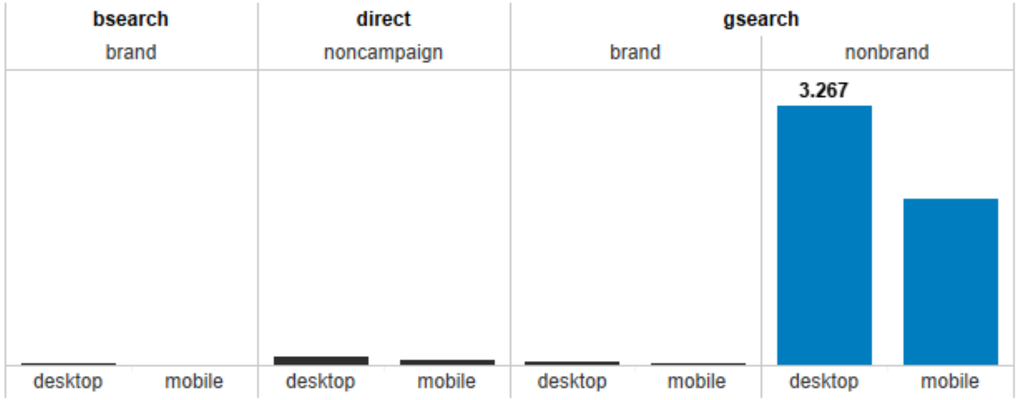
Is the ad no longer running? Has there been any change in targeting (audience or keywords) ? Is the ad copy optimized? Has the ad budget been adjusted? Or could there be technical issues, such as problems with the URL link in the ad?

March 19 - April 30, 2012 **using 'start of the week'*

Next

Copee

Top Traffic Source by Sessions (March - April)



Traffic Source Analysis (March - April)

		Total Session	Total Order	Conversion Rate	Total Sales
bsearch	brand	13	0	0,00%	
direct	noncampaign	166	7	4,22%	350
gsearch	brand	73	6	8,22%	300
	nonbrand	5.361	146	2,72%	7.299

'gsearch nonbrand' Analysis by Device Type (March - April)

	Total Session	Total Order	Conversion Rate	Total Sales
desktop	3.267	125	3,83%	6.249
mobile	2.094	21	1,00%	1.050

Next

March 19 - April 30, 2012

Initial Month Top Traffic Source

Total Session: **5613**

'gsearch nonbrand' has become a primary traffic source, contributing more than 90% of the total website traffic, with desktop users dominating.

Since most of the traffic comes from desktop users, its crucial to focus on optimizing the user experience for desktop devices.

Despite its high traffic, traffic from 'gsearch nonbrand' has not shown strong quality, as reflected by its low conversion rate (<3%).

Greater attention is needed for mobile users, as their traffic is relatively high but the conversion rate is extremely low (1%).

Low Conversion Rate Investigation

With the available data, we can investigate the reasons behind the low conversion rate by examining the performance of the landing page. The initial assumption is that the landing page's performance aligns with the traffic conversion performance.
There are two steps to be taken:

1. **Identify the landing page traffic from 'gsearch nonbrand'**
2. **Analyze the landing page performance based on the bounce rate**

Copee

March 19 - April 30, 2012

Top Landing Page for 'gsearch nonbrand' Traffic (March - April)

Landing Page	
/home	5.361

Landing Page '/home' Analysis (March - April)

Landing Page	Total Session	Bounce Session	Bounce Rate
/home	5.361	3.314	61,82%

Landing Page '/home' Analysis by Device Type (March - April)

Device Type..	Total Session	Bounce Session	Bounce Rate
desktop	3.267	1.812	55,46%
mobile	2.094	1.502	71,73%

Improving the landing page '/home' is an effort to address our initial assumption that the landing page high bounce rate affects the low conversion rate (CR). The results of this effort will be evaluated in the next month to determine the next steps to further increase the CR.

Next

Low CR Investigations

All traffic from 'gsearch nonbrand' is directed to the '/home' landing page.

A high bounce rate indicates that the '/home' landing page does not have enough value to encourage users to continue browsing the website.

A high bounce rate on mobile deivces indicates potential usability issues.

A thorough evaluation of the '/home' landing page is needed to assess the relevance of its content to user needs, as well as identify barriers experienced by users, especially mobile users.

Suggestions

1. Conduct a comprehensive evaluation of the '/home' page to ensure that the content is relevant, engaging, and aligned with the users' needs and expectations.
2. Ensure the '/home' page provides an optimal user experience, focusing on design, ease of navigation, and fast loading times.
3. Optimize pages for mobile users by improving responsiveness and ensuring a smooth and intuitive experience across all device types.

Copee

Pacmann

Based on the first month analysis, optimizations for the 'gsearch nonbrand' campaign and the '/home' landing page were implemented. **The manager has requested a re-evaluation of the optimization results after one month**. These results will determine the next steps for further optimization.

Data Analysis Steps

Data Cleaning and Transformation

Initial Month Analysis

Post-Improvement Analysis

A/B Testing Result

Funnel & Traffic Peak Time Analysis

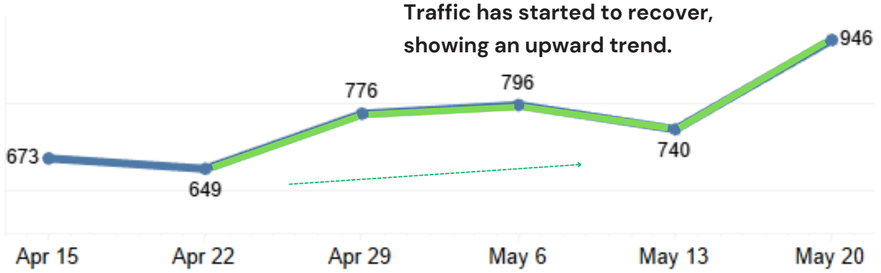
Next

Copee

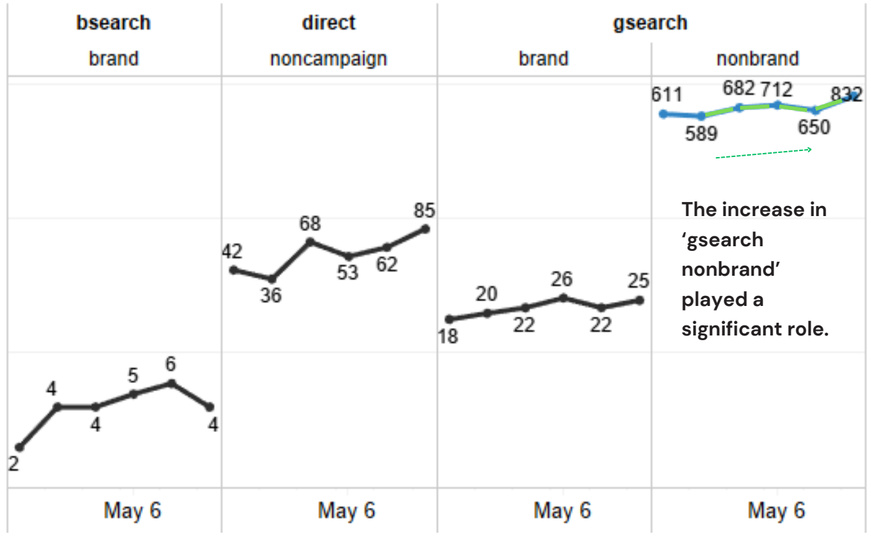
Post-Improvement Analysis

Next

Weekly Traffic Trend (April - May)



Weekly Traffic Trend by Traffic Source (April - May)



*April 16 - May 31, 2012 *using 'start of the week'

Traffic Source Analysis (April - May)

		Total Session	Total Order	Conversion Ra..	Total Sales
gsearch	brand	119	6	5,04%	300
	nonbrand	3.406	92	2,70%	4.599

Top Landing Page 'gsearch nonbrand' (April - May)

Landing Page	Total Session	Bounce Session	Bounce Rate
/home	3.406	2.055	60,33%

The optimization of the 'gsearch nonbrand' campaign has proven effective in recovering the traffic trend, with an increase in the 'gsearch nonbrand' trend followed by an overall traffic growth.

However, optimizing the 'home' landing page did not yield any results, as the bounce rate remained high (>50%) and there was also no impact on improving the conversion rate (<3%).

Suggestions

Due to the lack of results from the optimization of the '/home' landing page for 'gsearch nonbrand' traffic, testing with a new landing page is necessary. A/B testing should be conducted for each landing page, and the results of this testing will serve as a basis for the next optimization decisions.

Copee

Pacmann

With the '/home' landing page optimization showing no results, **the manager decided to conduct A/B testing for a new landing page named '/lander-1' for 'gsearch nonbrand' traffic**. This landing page was launched on June 19, 2012, and the A/B test was run for one month.

Data Analysis Steps

Data Cleaning and Transformation

Initial Month Analysis

Post-Improvement Analysis

A/B Testing Result

Funnel & Traffic Peak Time Analysis

Next

Copee

A/B Testing Result

June 19 - July 31, 2012

A/B Testing Result for '/home' and '/lander-1' (June - July)

Landing Page	Total Session	Bounce Session	Bounce Rate	Total Order	Conversion Rate
/home	2.328	1.365	58,63%	75	3,22%
/lander-1	2.750	1.450	52,73%	106	3,85%

A/B Testing Result for '/home' and '/lander-1' by Device Type (June - July)

Landing Page	Device Type..	Total Session	Bounce Session	Bounce Rate
/home	desktop	1.766	975	55,21%
	mobile	562	390	69,40%
/lander-1	desktop	2.092	1.017	48,61%
	mobile	658	433	65,81%

'gsearch nonbrand' Analysis after A/B Testing

		Total Session	Total Order	Conversion Rate	Total Sales
gsearch	nonbrand	5.078	181	3,56%	9.048

Page '/lander-1' does not replace '/home' on our website; it is only used as the main landing page for the 'gsearch nonbrand' campaign to encourage users to explore the website further.

The landing page '/lander-1' proved to be more effective in keeping users engaged and exploring the website, with a lower bounce rate.

The traffic coming from the '/lander-1' landing page was also of higher quality, with a higher conversion rate compared to '/home'.

Page optimization on mobile devices shows results with a decreasing bounce rate, but is still relatively high and needs periodic optimization.

The overall conversion rate of traffic from 'gsearch nonbrand' also improved, increasing to over 3%.

Suggestions

It is more effective to use '/lander-1' as the primary landing page for the 'gsearch nonbrand' campaign. This page has proven to better engage users and drive higher-quality traffic, making it the ideal choice to optimize campaign outcomes and overall user experience.

Next

Copee

Pacmann

After optimizing the performance of the 'gsearch nonbrand' campaign and landing page, **the manager directed us to conduct a comprehensive funnel analysis** for all traffic sources to identify pages with high exit rates. **The manager also wants to understand when traffic starts to peak on the website** to gain better insights into traffic behavior.

Data Analysis Steps

Data Cleaning and Transformation

Initial Month Analysis

Post-Improvement Analysis

A/B Testing Result

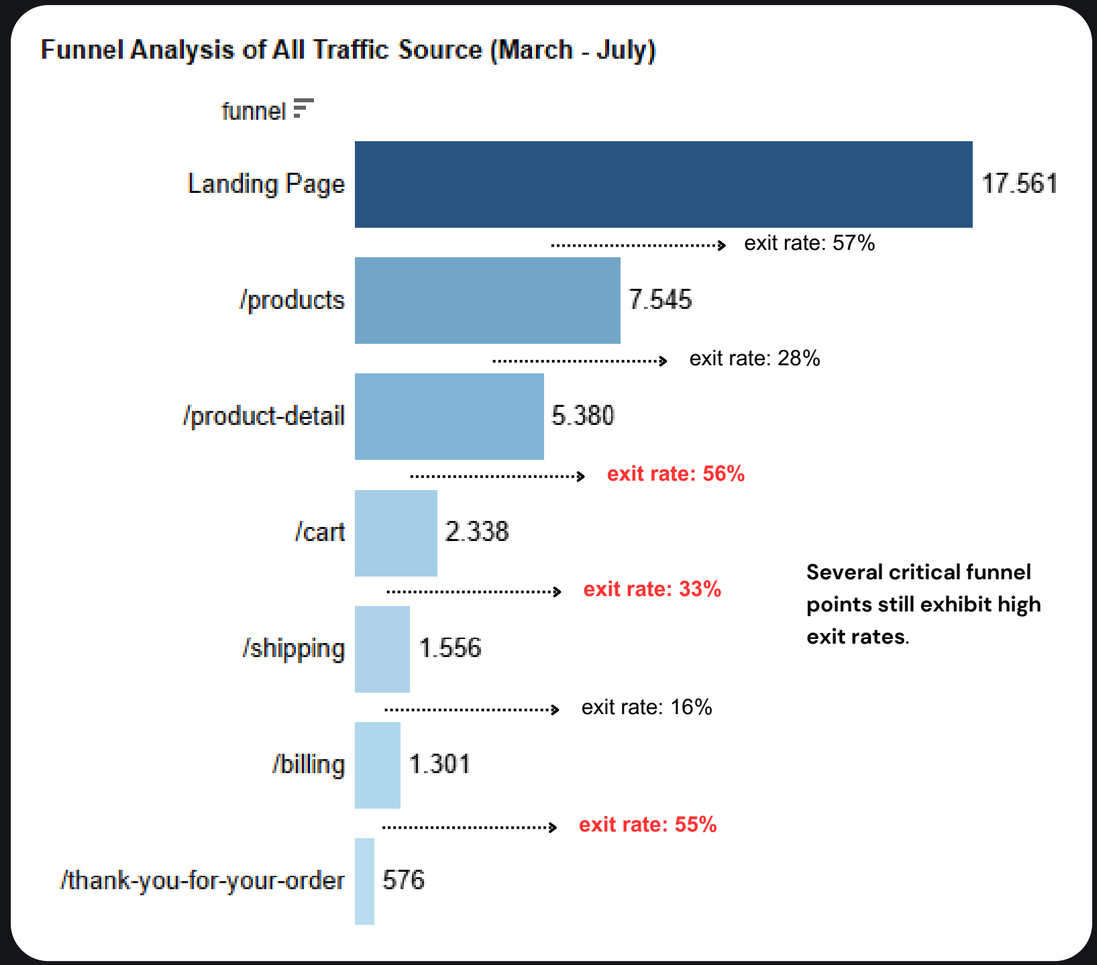
Funnel & Traffic Peak Time Analysis

Next

Copee

Next

Traffic Funnel Analysis



March 19 – July 31, 2012

The high exit rate on the landing page has been addressed by replacing the main landing page, '/lander-1.' The results should be evaluated periodically to ensure effectiveness.

Some key funnels, such as '/product-detail,' '/cart,' and '/billing,' experience very high exit rates. Conducting an in-depth analysis on each funnel is essential to identify pain points, optimize user experience, and effectively encourage users to proceed to the next stage.

Suggestions

1. Reviewing the clarity of product information, ensuring trust signals such as customer reviews are sufficiently strong, and testing product fit to anticipate potential window shoppers.
2. Evaluating the complexity of the checkout process and displaying clear CTAs to encourage users to move from the cart to checkout as quickly as possible.
3. Ensuring no unexpected fees during the billing process, offering a variety of payment methods, and streamlining the payment process for ease of use.
4. Conducting surveys with users about their journey on the website to identify friction points and potential technical issues across different device types.

Copee

Next

Analyzing User Sessions by Day and Hour (March - July)

Hour	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
0	35	64	68	59	79	67	61
1	34	60	49	45	61	46	37
2	17	55	46	56	47	37	34
3	26	53	52	54	36	38	33
4	38	38	49	44	42	43	21
5	31	45	50	40	52	36	25
6	39	58	53	59	72	62	29
7	30	79	75	74	86	87	49
8	56	140	108	116	128	126	38
9	44	150	138	168	189	151	58
10	66	209	208	228	182	205	73
11	62	189	235	212	174	181	80
12	69	211	207	225	210	173	72
13	68	213	231	222	204	191	87
14	61	176	214	192	205	188	80
15	75	185	210	194	219	215	79
16	77	160	244	197	189	194	61
17	71	191	193	178	190	115	45
18	45	145	168	156	155	139	52
19	64	124	148	113	109	97	62
20	83	115	117	113	114	75	47
21	90	119	107	125	126	81	49
22	83	103	106	104	92	61	60
23	100	92	92	59	102	65	54

March 19 – July 31, 2012

Traffic Peak Time Analysis

The website experiences the highest traffic on weekdays (Monday-Friday), with activity starting to occur from 08:00 am to 18:00 pm.

Peak hours occur between 10:00 AM and 15:00 PM, showing the times when users are most active.

Surprisingly, traffic dropped significantly over the weekend (Saturday-Sunday). Indicating that there was a lack of browsing activity during the holidays.

Suggestions

By identifying the days and hours when traffic activity peaks, we can leverage this information to optimize ads, promotions, and digital activations, as well as website and server performance, during peak hours only. Optimizing during these high-traffic periods can help reduce unnecessary costs during off-peak hours, while boosting engagement and potentially increasing revenue.

Copee

Pacmann

Recommendations for Further Strategic Actions

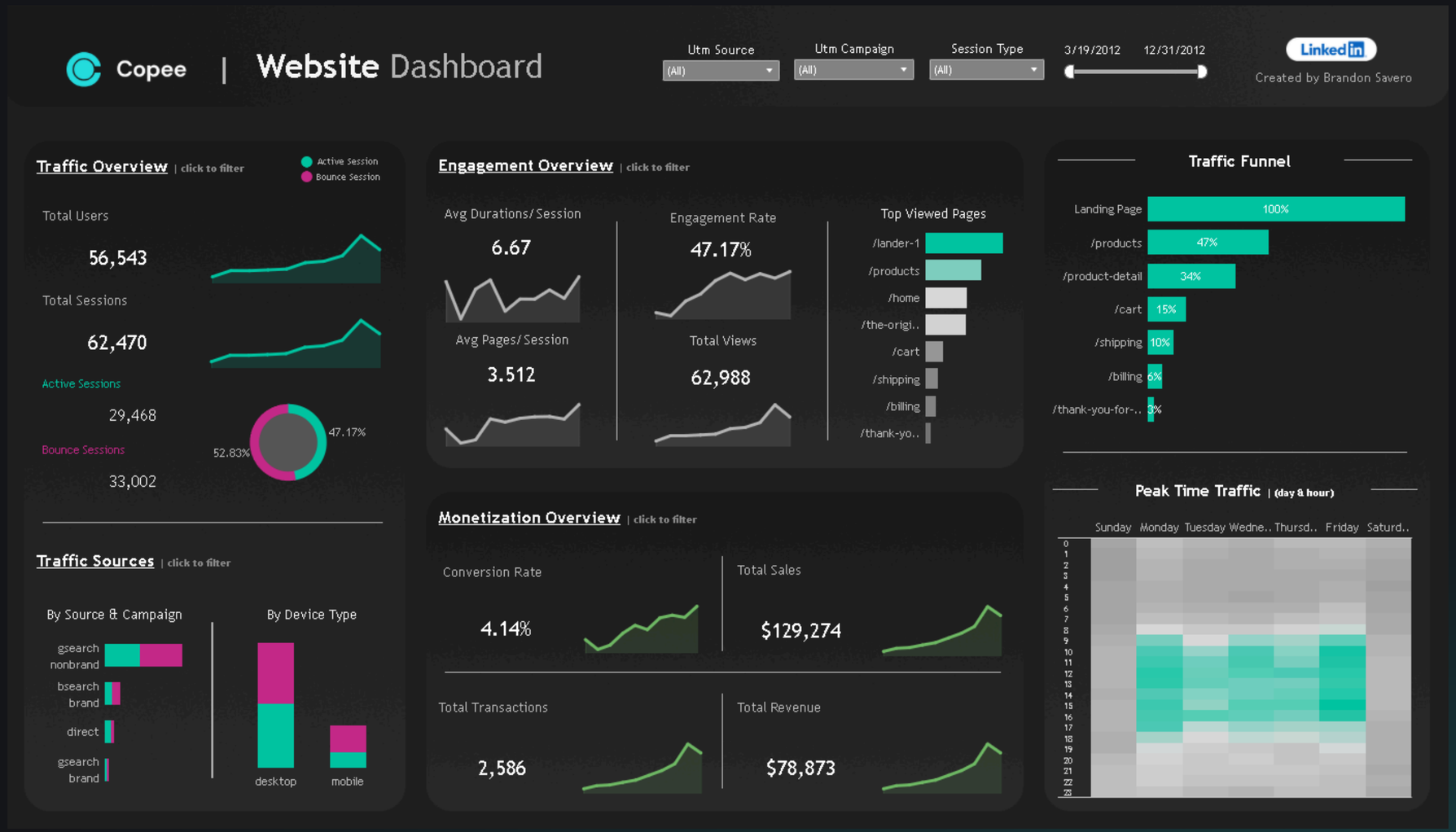
- That the majority of traffic sources come from 'non-branded gsearch', it would be more efficient to focus the campaign budget on that channel, thus saving on spending on underperforming channels.
- Switching the '/home' page to '/lander-1' as the main landing page, with proven better performance data
- Perform periodic optimization on the landing page '/lander-1' to improve performance, taking into account various types of devices with various screen sizes
- Evaluate and optimize the '/product-detail', '/cart', and '/billing' pages by considering the clarity of information, the willingness of CTAs to encourage users to move, and the ease of the user journey on these pages to reduce the exit rate.
- Prioritize the scheduling of campaigns, ads, or other digital activations, at times with the most traffic activity to increase engagement.
- Optimize the entire website ecosystem during times of high traffic to prevent technical issues from occurring, and reduce costs during off-peak times.

Copee

Pacmann

Website Dashboard

Click for Access Dashboard



Copee

Pacmann

Thank You

Brandon Savero

https://github.com/Brandonsvro

Brandon Savero