WEB PERFORMANCE ANALYTICS USING SQL

-Long Bui-

1. Introduction

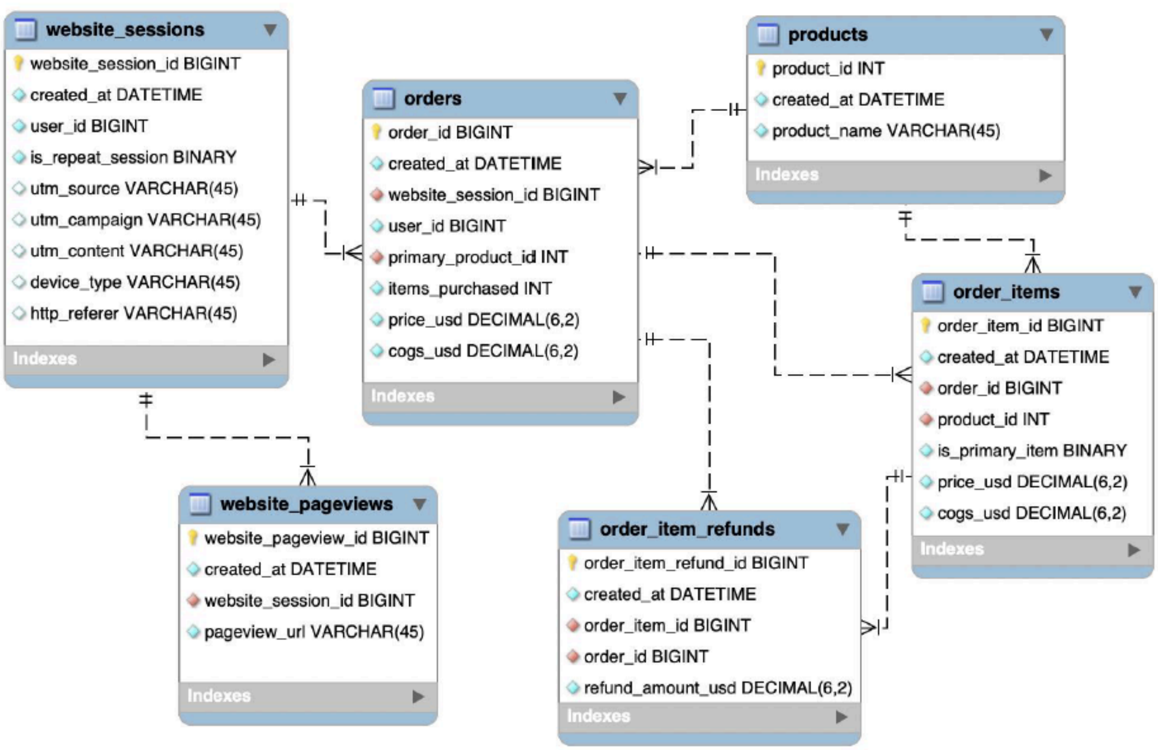
This project optimizes SQL Queries and statistical testing to analyze and optimize ecommerce product's performance including:

- Web analysis: analyzing web performance
  - Traffic Source Analysis: analyze accessing rate of each pages in the web
  - Website Performance Analysis: analyze how webpages corresponding to others
  - Channel Portfolio Analysis: analyze how efficiency of these webpages

- Product Analysis: analyzing product sales efficiency
  - Product Sales Analysis: analyze overall product sales
  - Cross-sell analysis: analyze patterns that products that being bought together
  - Product Refund Rate Analysis: analyze refund rate by product
- User Analysis: analyzing how customers interact with websites
  - Repeat Channel Behaviors: how frequently users access a certain page
  - New vs Repeat Comparison: comparing different types of pages

2. Tools And Applications

Languages: *SQL*
Tools: MySQLWorkbench
Applications: Statistics (A/B Testing, Hypothesis Testing), Ecommerce (Webpages analysis, marketing analysis, product analysis)
Data: Mavenfuzzy Factory



3. Web Performance Analysis

a. Conversion Rate Analysis

<u>Definition:</u>

- **utm_source**: This will be the field to store the source of session information, for example, if you run an ad via Facebook, then the user clicks on the ad and is redirected to your website, then utm_source=facebook. If you perform traffic analysis and see a field called utm_source in the database, please understand the meaning of this field.
- **utm_campaign**: Website owners will often have many advertising campaigns, each such campaign will be considered a campaign.
- **http_referer**: This is the full domain name of each utm_source
- **Conversion rate:** a concept that indicates the EFFECTIVENESS of a website through the actions that customers perform.
    - For example, your website receives 1,000 visitors in a month and 90 customers sign up for consultation. The conversion rate will be 90/1000 or 9%.
    - A conversion can be any customer action that a business wants them to take. It could be the ratio of people registering for consulting/visitors to the website. It can also be the ratio of people clicking to register for the service/web visitors. Each goal that a business sets will have its own conversion rate.
-
- **CVR** = $\frac{numer\ of\ order\_id}{total\ website\_session\_id}$
- **Bid:** In the marketing process, optimizing the marketing budget is very important. Marketing campaigns will have many ways to charge. For example, when running ads on Facebook, you will have to set prices for advertising campaigns. So the price here will be understood as a bid. The higher the campaign price, the more customers the campaign will reach (avoid the confusion between reaching more customers and definitely increasing the conversion rate).

– Website session by source

utm_source	utm_campaign	http_referer	number_session
gsearch	nonbrand	https://www.gsearch.com	3485
NULL	NULL	NULL	26
NULL	NULL	https://www.gsearch.com	25
gsearch	brand	https://www.gsearch.com	23
NULL	NULL	https://www.bsearch.com	7
bsearch	brand	https://www.bsearch.com	7

We realize gsearch and nonbrand are big traffic sources, but we need to validate whether sessions from gsearch and nonbrand are driving sales. To find this, we will calculate the conversion rate (CVR) from the orders (order_id). The requirement is that the CVR ratio must be greater than 4%. If we calculate a CVR lower than 4%, it means we have to reduce your bid. If we calculate a higher CVR, it means you can increase your bid.

– Total conversion Rate

sessions	orders	session_to_order_conv_rt
3816	111	0.0291

Based on our conversion rate analysis in request 2, the company reduced the price of gsearch's nonbrand items on April 15, 2012. Our job is to find the number of trending sessions (nonbrand and gsearch) by week to see if price changes cause the number of sessions to decrease at all. The figures will be calculated before May 15, 2012.

– Weekly website sessions

week_started_at	sessions
2012-03-19	854
2012-03-25	949
2012-04-01	1147
2012-04-08	984
2012-04-15	645
2012-04-22	598
2012-04-29	684
2012-05-06	595

We received a response from the marketing director about a poor experience using the mobile website before May 11, 2012. And this director wants to calculate CVR on devices. If CVR on Desktop is better than on mobile, we can increase bids on sessions accessed by Desktop. Write a query to fulfill the director's request, then comment on the returned results.

– Conversion Rate by Device Type

device_type	orders	conv_rt
desktop	144	0.0372
mobile	24	0.0098

After analyzing the conversion rate by device, we will notice that Desktop gives higher CVR. So the company increased bids for Desktop campaigns (nonbrand and gsearch) on May 19, 2012. O ur job is to write a query to track weekly trends for both Desktop and mobile sessions to The company can see the impact on the total number of sessions, then comment on the results returned.

(search period will start from April 15, 2012)

– BID Optimization

week_start_date	desktop_sessions	mobile_sessions
2012-04-15	397	248
2012-04-22	364	234
2012-04-29	429	255
2012-05-06	414	276
2012-05-13	410	228
2012-05-20	656	185
2012-05-27	583	184
2012-06-03	553	152

b. Website Performance Analysis

A new manager wants to find the website pages with the most views ranked Ranked by number of sessions up to June 9, 2012. Our job is to write query to make this request, then comment on the returned results.

– sessions by pageview url

pageview_url	pvs
/home	10322
/products	4194
/the-original-mr-fuzzy	3004
/cart	1293
/shipping	860
/billing	710
/thank-you-for-your-order	305

– top page with highest session

landing_page_url	sessions_hitting_page
/home	10614

In the previous request, we learned that all traffic is currently concentrated on the home page (/home) first. Now the admin wants you to check that the home page is working how. The manager asked to find the home page's session bounce rate as shown by three session quantities, bounced session (number of sessions bounced), and % bounced session on the total number of sessions. Our job is to write queries like an administrator process the request and provide comments on the returned results.

– bounced sessions

sessions	bounced_sessions	bounced_rate
10942	6475	0.5918

Based on the analysis of our session bounce rate, the company ran a custom landing page new page (/lander-1), this new page and the old home page (/home) will be divided equally 50/50 in traffic access (nonbrand and gsearch). This is also known as A/B testing. Manager time wants to see the session bounce rate of two new and old websites and please make sure to review during times when /lander-1 is experiencing traffic to keep it fair.

– A/B testing with lander-1

landing_page	sessions	bonced_sessions	bonce_rate
/home	2227	1300	0.5837
/lander-1	2278	1210	0.5312

week_start_date	total_sessions	bounce_sessions	home_sessions	lander_sessions
2012-06-01	229	0.60262	229	0
2012-06-03	796	0.58920	796	0
2012-06-10	869	0.61565	869	0
2012-06-17	843	0.55991	512	331
2012-06-24	754	0.58621	367	387
2012-07-01	789	0.57795	401	388
2012-07-08	794	0.56927	384	410
2012-07-15	853	0.54045	428	425
2012-07-22	788	0.50635	397	391
2012-07-29	1028	0.50000	54	974
2012-08-05	1080	0.54722	0	1080
2012-08-12	992	0.50605	0	992
2012-08-19	1022	0.50391	0	1022
2012-08-26	765	0.54641	0	765

Conclusion: Lander-1 successfully decreased bouncing rates

c. Channel Portfolio Management

Analyzing a portfolio of marketing channels is about bidding efficiently and using data to maximize the effectiveness of our marketing budgets. Our goals are:

- Understanding with marketing channels are driving the mose sessions and orders through our website
- Understanding differences in user characteristics and conversion performance across marketing channels

- Optimizing bids and allocating marketing spend across multi-channel portfolio to achieve maximum performance
- **Free Traffic Analysis:** To identify traffic to our site that we are not paying for with marketing campaigns.

With gsearch doing well and the site performing better, we launched a second paid search channel, bsearch. Let's discover the weekly stranded sessions column and compare this to gsearch nonbrand.

–gsearch vs bsearch sessions

week_start_date	gsearch	bsearch
2012-08-22	593	196
2012-08-26	1059	343
2012-09-02	920	290
2012-09-09	950	329
2012-09-16	1154	362
2012-09-23	1052	323
2012-09-30	999	318
2012-10-07	996	330

It looks like bsearch tends to get roughly a this the traffic of gsearch. Let's compare the percentage of traffic coming on Mobile with gsearch.

utm_source	sessions	mobile_session	pct_mobile
bsearch	6460	558	0.0864
gsearch	19890	4869	0.2448

Let's dig in a bit more so that we can getting our bids. We will pull nonbrand conversion rates from session to order for gsearch and bsearch, slicing the data by device type.

	device_type	utm_source	sessions	conv_rt
▶	desktop	bsearch	5902	0.0513
	desktop	gsearch	15021	0.0493
	mobile	bsearch	558	0.0090
	mobile	gsearch	4869	0.0156

The channels don't perform identically, so we should differentiate our bids in order to optimize our overall paid marketing budgets.
⇒ Bid down bsearch based on its under-performance.

Let's compare gsearch and bsearch nonbrand broken down by device and see how it changes through each week.

	week_start_date	gsearch	bsearch	b_pct_g_dtop
▶	2012-11-04	1028	401	0.3901
	2012-11-11	956	401	0.4195
	2012-11-18	2655	1008	0.3797
	2012-11-25	2058	843	0.4096
	2012-12-02	1326	517	0.3899
	2012-12-09	1277	293	0.2294
	2012-12-16	1270	348	0.2740

Looks like bsearch traffic dropped off a bit after the bid down. Seems like gseach was down too after Black Friday and Cyber Monday, but bsearch dropped even more.

– Free-traffic analysis
Let's pull organic search, direct type in , and paid brand seach sessions by month as % of paid search non-brand

yr	mo	nonbrand	nonbrand	brand_pct_of_nonbrand	direct	direct_pct_of_nonbrand	organic	organic_pct_of_nonbrand
2012	3	1850	10	0.0054	9	0.0049	8	0.0043
2012	4	3507	76	0.0217	71	0.0202	78	0.0222
2012	5	3294	139	0.0422	151	0.0458	150	0.0455
2012	6	3442	165	0.0479	170	0.0494	190	0.0552
2012	7	3656	195	0.0533	187	0.0511	207	0.0566
2012	8	5320	263	0.0494	250	0.0470	265	0.0498
2012	9	5588	339	0.0607	285	0.0510	331	0.0592
2012	10	6883	431	0.0626	440	0.0639	428	0.0622

Conclusion: All are growing ass a pct of our paid traffic volume. → paid traffic volume is doing well

d. Seasonality Analysis

Analyzing business patterns is about generating insights to help us maximize efficiency and anticipate future trends. Goals:

- Day-parting analysis to understand how much support staff you should have at different times of day or days of the week
- Analyzing seasonality to better prepare for upcoming spikes or slowdown in demands

Let's take a look at 2012's monthly and weekly volume pattern

yr	mo	week_start	sessions	orders
2012	3	2012-03-19	1877	60
2012	4	2012-04-01	3732	99
2012	5	2012-05-01	3734	108
2012	6	2012-06-01	3967	140
2012	7	2012-07-01	4245	169
2012	8	2012-08-01	6098	228
2012	9	2012-09-01	6543	287
2012	10	2012-10-01	8182	371

Looks like we grew fairly steadily all year, and saw significant volume around the holiday months (especially the weeks of Black Friday and Cyber Monday). We will want to keep this in mind in 2013 as we think about customer support and inventory management.

Let’s consider the average website session by volume, by hour of day and by day of week.

hr	ave_sesssions	mon	tues
0	8.1	9.3333	7.8889
1	6.3	7.8889	7.1111
2	5.6	5.8889	6.3333
3	4.5	6.4444	4.8889
4	4.7	5.4444	4.8889
5	4.4	5.2500	5.2222
6	5.1	5.3333	6.0000

Conclusion: We can plan on one support staff around the clock and then we should double up 2 staffs members from 8 AM to 5 PM Monday through Friday

4. Product Analysis

a. Product Sales Analysis

Definition:

- **Orders:** Number of orders created by the user
- **Revenue:** Revenue
- **Margin:** Can be understood as profit (revenue - minus fees)
- **AOV:** Average revenue per order.
- **Product-Level Website Pathing:** product-focused website analysis such as finding out how customers interact with each product and then each of those products on how well our customers convert.

The company is about to launch a new product and it wants to learn about the products in depth The company wants to calculate total sales revenue and total profits until January 4, 2013

– Product-level Sales

yr	mo	number_of_states	total_revenue	total_margin
2012	3	59	2949.41	1799.50
2012	4	99	4949.01	3019.50
2012	5	107	5348.93	3263.50
2012	6	139	6948.61	4239.50
2012	7	169	8448.31	5154.50
2012	8	230	11497.70	7015.00
2012	9	283	14147.17	8631.50
2012	10	361	18046.39	11010.50
2012	11	619	30943.81	18879.50
2012	12	514	25694.86	15677.00
2013	1	40	1999.60	1220.00

The company introduced a second product on January 6, we'll need to compile some analyze trends by writing queries. The indicators we need to look for are the number of orders monthly, conversion rate. Let's analyze the period from April 1, 2012 to April 1, 2013.

yr	mo	sessions	orders	conv_rate	revenue_per_session
2012	4	3708	100	0.0270	1.348166
2012	5	3707	106	0.0286	1.429442
2012	6	4023	139	0.0346	1.727221
2012	7	4164	169	0.0406	2.028893
2012	8	6097	230	0.0377	1.885796
2012	9	6591	283	0.0429	2.146438
2012	10	8042	362	0.0450	2.250234
2012	11	14066	618	0.0439	2.196347
2012	12	10117	514	0.0508	2.539771
2013	1	6362	388	0.0610	3.119478
2013	2	7157	486	0.0679	3.616758
2013	3	6353	402	0.0633	3.271837

On January 6, 2013, the Company just had a new product and the company wanted to know what page they will look at next during their visits to the /products page, and just take it during the period from 3 months before the product launch until now. The date we receive is announced by the company on April 6, 2013.

	time_period	sessions	w_next_pg	pct_w_next_pg	to_mrfuzzy	pct_to_mrfuzzy	to_lovebear	pct_to_lovebear
▶	A. Pre_Product_2	15725	11374	0.7233	11374	0.7233	0	0.0000
	B. Post_Product_2	10669	8169	0.7657	6629	0.6213	1540	0.1443

Let's see the conversion funnels from each product page to conversion. We would like to compare between the two conversion funnels of the 2 product above for all website traffic.

b. Cross-sell Analysis

On September 25th we start giving customers the option to add 2nd product while on the /cart page. We can compare the month before vs the month after the change to see CTR from the /cart page, Avg Products per Order, AOV, and overall revenue per /cart pageview.

Now let's run a pre-post analysis comparing the month before vs the month after in term of session-to-order conversion rate, AOV, products per order, and revenue per session

time_period	cart_sessions	clickthroughs	cart_ctr	orders_placed	products_purchased	products_per_order	revenue	aov	rev_per_cart_session
A. Pre_Cross_Sell	1829	1227	0.6709	651	651	1.0000	33473.49	51.418571	18.301525
B. Post_Cross_Sell	1976	1353	0.6847	671	701	1.0447	36402.99	54.251848	18.422566

Products per order, AOV, and revenue per /cart session are all up slightly since the cross-sell feature was added. It is not a game changer but the trend looks positive.

time_period	conv_rate	average_order_value	products_per_order	revenue_per_session
A. Pre_Birthday_Bear	0.0608	54.226502	1.0464	3.298677
B. Post_Birthday_Bear	0.0702	56.931319	1.1234	3.998763

It looks like all of our critical metrics have improved since we launched the third product. This seems fantastic.

c. Product Refund Rate

Let's investigate further on monthly product refund rates by product, and confirm whether our quality issues have been fixed.

order_id	order_item_id	price_paid_usd	created_at	order_item_refund_id	refund_amount_usd	created_at
3489	3489	49.99	2013-03-03 09:51:10	NULL	NULL	NULL
27061	33000	49.99	2015-01-03 16:47:12	1505	49.99	2015-01-12 11:47:12
27061	33001	45.99	2015-01-03 16:47:12	1526	45.99	2015-01-19 13:47:12
32049	39671	49.99	2015-03-15 15:33:51	1728	49.99	2015-03-30 21:33:51
32049	39672	45.99	2015-03-15 15:33:51	NULL	NULL	NULL

Conclusion: The refund rates for Mr.Fuzzy did go down after initial improvements in September 2013 but refund rates were terrible in August and September, as expected.

5. User Analysis

Analyzing repeat visits help us to understand user behavior and identify some of our most valuable customers. Goals:

- Analyzing repeat activity to see how often customers are coming back to visit your site
- Understanding which channels they use when they come back, and whether or not you are paying for them again through paid channels
- Using your repeat visit activity to build a better understanding of the value of a customer in order to better optimize marketing channels

a. Repeat Channel Behavior

We have been thinking about customer value based solely on their first session conversion and revenue. But if customers have repeated sessions, they may be more valuable than we may thought. In that case, we might be able to spend a bit more to acquire them.

Let’s pull data on how many of our websites visitors come back for another session

repeat_sessions	users
0	126804
1	14084
2	316
3	4685

It Looks like a fair number of our customers do come back to our site after the first sessions.

Let's investigate the minimum, maximum, and average time between first and second session for customers who come back.

avg_days_first_to_second	min_days_first_to_second	max_days_first_to_second
33.2651	1	69

It is interesting to see that our repeat visitors are coming back about a month later, on average. We should investigate the channels that these visitors are using.

b. Analyzing New & Repeat Channels

Comparing new vs. repeat sessions by channels.

channel_group	new_sessions	repeat_sessions
organic_search	7139	11507
paid_brand	6432	11027
direct_type_in	6591	10564
paid_nonbrand	119950	0
paid_social	7652	0

It looks like when customers come back for repeat visits, they come mainly through organic search, direct type-in, and paid brand. Only about ⅓ comes through a paid channel, and brand clicks are cheaper than non brand. So all in all, we're not paying very much for these subsequent visits.

Let's compare the conversion rate and revenue per session for repeat sessions vs new sessions.

	is_repeat_session	sessions	conv_rt	rev_per_session
▸	0	149787	0.0680	4.343754
	1	33577	0.0811	5.168828

Looks like repeat sessions are more likely ton convert, and produce more revenue per session.

6. **Conclusion**

In conclusion, this web performance analytics project leverages SQL queries and statistical testing to optimize e-commerce product performance. The analysis spans various dimensions, including web, product, user, and seasonality, providing valuable insights for decision-making.

The Conversion Rate Analysis delves into the effectiveness of the website by examining various sources and campaigns, guiding bid adjustments for optimal marketing budget allocation. Website Performance Analysis identifies high-traffic pages, ensuring the home page's effectiveness through bounce rate analysis and A/B testing.

Channel Portfolio Management involves optimizing bids and marketing spend across different channels, distinguishing between paid and free traffic. Seasonality Analysis uncovers patterns in monthly and weekly volumes, aiding in resource planning. Product Analysis evaluates sales, new product launches, and cross-sell features, contributing to revenue optimization.

User Analysis explores repeat visits, revealing valuable insights into customer behavior and channel preferences. The project concludes with a comprehensive overview of the findings, offering actionable recommendations for further optimization.

Through SQL queries and data-driven insights, this project empowers businesses to make informed decisions, enhance user experience, and maximize the efficiency of their web performance and marketing strategies.