E-Commerce Data Analysis

(My SQL Workbench)

1. Website Traffic Source Analysis

1) Finding top traffic sources

Scenario:

Date: 2012-04-12

We've been live for almost a month now and sales are starting to generate. Help is needed to understand where the bulk of website sessions are coming from, through yesterday.

A breakdown by UTM source, campaign, and referring domain is requested if possible.

Result:

utm_source	utm_campaign	http_referer	sessions
gsearch	nonbrand	https://www.gsearch.com	3613
NULL	NULL	NULL	28
NULL	NULL	https://www.gsearch.com	27
gsearch	brand	https://www.gsearch.com	26
bsearch	brand	https://www.bsearch.com	7
NULL	NULL	https://www.bsearch.com	7

Most of the traffic is being generated from gsearch nonbrand campaign which can be drilled deeper to explore potential optimization opportunities.

2) Traffic source conversion rate

Scenario: Calculate the conversion rate (CVR) from session to order for gsearch nonbrand as the major traffic source. In performance marketing, measuring the return on advertising spend (ROAS) is crucial. Based on the cost paid for clicks, a CVR of at least 4% is required to make the numbers work. If the CVR is much lower than this threshold, bids need to be reduced. If the CVR is much higher, bids can be increased to drive more volume.

Result:

sessions	orders	conv_rate
3895	112	2.8755

The conversion rate is below the 4% threshold. Based on the analysis, bids need to be reduced a bit to make the economics work.

3) Traffic source trending

Scenario: Based on your conversion rate analysis, we bid down gsearch nonbrand on 2012-04-15. Can you pull gsearch nonbrand trended session volume, by week, to see if the bid changes have caused volume to drop at all?

Result:

year	week	start_of_week_date	sessions
2012	12	2012-03-19	896
2012	13	2012-03-25	956
2012	14	2012-04-01	1152
2012	15	2012-04-08	983
2012	16	2012-04-15	621
2012	17	2012-04-22	594
2012	18	2012-04-29	681
2012	19	2012-05-06	399

gsearch nonbrand sessions volume has declined after the bid changes on 2012-04-15.

4) Bid optimization for paid traffic

Scenario: I was trying to use our site on my mobile device the other day, and the experience was not great. Could you pull conversion rates from session to order, by device type? If desktop performance is better than on mobile, we may be able to bid up for desktop specifically to get more volume?

Result:

device_type	sessions	orders	conv_rate
mobile	2492	24	0.9631
desktop	3911	146	3.7331

Desktop sessions have higher conversion rate than mobile. So, bids on desktop can be increased for sales boost.

5) Traffic Source Segment Trending

Scenario: After your device-level analysis of conversion rates, we realized desktop was doing well, so we bid our gsearch nonbrand desktop campaigns up on 2012-05-19. Could you pull weekly trends for both desktop and mobile so we can see the impact on volume? You can use 2012-04-15 until the bid change as a baseline.

Result:

week_start_date	mob_sessions	desk_sessions	total_sessions
2012-04-15	238	383	621
2012-04-22	234	360	594
2012-04-29	256	425	681
2012-05-06	282	430	712
2012-05-13	214	403	617
2012-05-20	190	661	851
2012-05-27	183	585	768
2012-06-03	157	582	739

After the bid changes the desktop volume has risen which means the changes has affected the business positively.

2. Business Pattern and Seasonality Analysis

1) Analyzing seasonality

Scenario: 2012 was a great year for us. As we continue to grow, we should take a look at 2012's monthly and weekly volume patterns, to see if we can find any seasonal trends, we should plan for in 2013. If you can pull session volume and order volume, that would be excellent.

Result:

year	month	sessions	orders
2012	3	1879	60
2012	4	3734	99
2012	5	3736	108
2012	6	3963	140
2012	7	4249	169
2012	8	6097	228
2012	9	6546	287
2012	10	8183	371
2012	11	14011	618
2012	12	10072	506

There can be seen a steady growth for the whole year with a surge in the months of November and December. This can be due to Black Friday and Cyber Monday weeks which can be seen in the weekly breakdown table below. So, this surge should be kept in mind to cope with the customer service and inventory for the year 2013.

year	week	week_start	sessions	orders
2012	12	2012-03-19	896	25
2012	13	2012-03-25	983	35
2012	14	2012-04-01	1193	29
2012	15	2012-04-08	1029	28
2012	16	2012-04-15	679	22
2012	17	2012-04-22	655	18
2012	18	2012-04-29	770	19
2012	19	2012-05-06	798	17
2012	20	2012-05-13	706	23
2012	21	2012-05-20	965	28
2012	22	2012-05-27	875	31
2012	23	2012-06-03	920	34
2012	24	2012-06-10	994	29
2012	25	2012-06-17	966	37
2012	26	2012-06-24	883	32
2012	27	2012-07-01	892	30
2012	28	2012-07-08	925	36
2012	29	2012-07-15	987	47
2012	30	2012-07-22	954	41
2012	31	2012-07-29	1172	55
2012	32	2012-08-05	1235	48
2012	33	2012-08-12	1181	39
2012	34	2012-08-19	1522	55
2012	35	2012-08-26	1593	52
2012	36	2012-09-02	1418	56
2012	37	2012-09-09	1488	72
2012	38	2012-09-16	1776	76
2012	39	2012-09-23	1624	70
2012	40	2012-09-30	1553	67
2012	41	2012-10-07	1632	73
2012	42	2012-10-14	1955	93
2012	43	2012-10-21	2042	95
2012	44	2012-10-28	1923	82
2012	45	2012-11-04	2086	91
2012	46	2012-11-11	1973	101
2012	47	2012-11-18	5130	223
2012	48	2012-11-25	4172	179
2012	49	2012-12-02	2727	145
2012	50	2012-12-09	2489	123
2012	51	2012-12-16	2718	135
2012	52	2012-12-23	1682	74
2012	53	2012-12-30	309	21

2) Analyzing business pattern

Scenario: We're considering adding live chat support to the website to improve our customer experience. Could you analyze the average website session volume, by hour of day and by day week, so that we can staff appropriately? Let's avoid the holiday time period and use a date range of Sep 15 - Nov 15, 2012.

Result:

hr	avg_mon_sessions	avg_tue_sessions	avg_wed_sessions	avg_thu_sessions	avg_fri_sessions	avg_sat_sessions	avg_sun_sessions
0	8.7	7.7	6.3	7.4	6.8	5	5
1	6.6	6.7	5.3	4.9	7.1	5	3
2	6.1	4.4	4.4	6.1	4.6	3.7	3
3	5.7	4	4.7	4.6	3.6	3.9	3.4
4	5.9	6.3	6	4	6.1	2.8	2.4
5	5	5.4	5.1	5.4	4.6	4.3	3.9
6	5.4	5.6	4.8	6	6.8	4	2.6
7	7.3	7.8	7.4	10.6	7	5.7	4.8
8	12.3	12.2	13	16.5	10.5	4.3	4.1
9	17.6	15.7	19.6	19.3	17.5	7.6	6
10	18.4	17.7	21	18.4	19	8.3	6.3
11	18	19.1	24.9	21.6	20.9	7.2	7.7
12	21.1	23.3	22.8	24.1	19	8.6	6.1
13	17.8	23	20.8	20.6	21.6	8.1	8.4
14	17.9	21.6	22.3	18.5	19.5	8.7	6.7
15	21.6	17.1	25.3	23.5	21.3	6.9	7.1
16	21.1	23.7	23.7	19.6	20.9	7.6	6.6
17	19.4	15.9	20.2	19.8	12.9	6.4	7.6
18	12.7	15	14.8	15.3	10.9	5.3	6.8
19	12.4	14.1	13.3	11.6	14.3	7.1	6.4
20	12.1	12.4	14.2	10.6	10.3	5.7	8.4
21	9.1	12.6	11.4	9.4	7.3	5.7	10.2
22	9.1	10	9.8	12.1	6	5.7	10.2
23	8.8	8.6	9.6	10.6	7.6	5.3	8.3

According to the results, more staffing can be needed during 8 am-5pm for live chat support and the rest of hours can be covered with the regular number of agents.

3. Product-Level Analysis

1) Product-level sales analysis

Scenario: We're about to launch a new product, and Id like to do a deep dive on our current flagship product. Can you please pull monthly trends for year 2012 for number of sales, total revenue, and total margin generated for the business?

Result:

year	month	number_of_sales	total_revenue	total_margin
2012	3	60	2999.4	1830
2012	4	99	4949.01	3019.5
2012	5	108	5398.92	3294
2012	6	140	6998.6	4270
2012	7	169	8448.31	5154.5
2012	8	228	11397.72	6954
2012	9	287	14347.13	8753.5
2012	10	371	18546.29	11315.5
2012	11	618	30893.82	18849
2012	12	506	25294.94	15433

2) Product-launches analysis (Impact of new product launch)

Scenario: We launched our second product back on January 6th. Can you pull together some trended analysis? I'd like to see monthly order volume, overall conversion rates, revenue per session, and a breakdown of sales by product, all for the time period since April 1, 2012.

Result:

year	month	sessions	orders	conv_rate_perc	total_revenue	revenue_per_session	product_1_orders	product_2_orders
2012	4	3734	99	2.6513	4949.01	1.325391	99	0
2012	5	3736	108	2.8908	5398.92	1.445107	108	0
2012	6	3963	140	3.5327	6998.6	1.765985	140	0
2012	7	4249	169	3.9774	8448.31	1.988305	169	0
2012	8	6097	228	3.7395	11397.72	1.869398	228	0
2012	9	6546	287	4.3844	14347.13	2.19174	287	0
2012	10	8183	371	4.5338	18546.29	2.266441	371	0
2012	11	14011	618	4.4108	30893.82	2.204969	618	0
2012	12	10072	506	5.0238	25294.94	2.511412	506	0
2013	1	6401	391	6.1084	20016.09	3.127025	344	47
2013	2	7168	497	6.9336	26465.03	3.692108	335	162
2013	3	6264	385	6.1462	19896.15	3.176269	320	65

Product 2 has got lots of sales in February but then it declined in March. Revenue per session and conversion rate has improved over time.

3) Analyzing product-level website pathing

Scenario: Now that we have a new product, I'm thinking about our user path and conversion funnel. Let's look at sessions which hit the /products page and see where they went next. Could you please pull clickthrough rates from /products since the new product launch on January 6th 2013, by product, and compare to the 3 months leading up to launch as a baseline?

Result:

time_period	sessions	sessions_to_next_pg	pct_sessions_to_next_pg	sessions_to_mrfuzzy	pct_sessions_to_mrfuzzy	sessions_to_lovebear	pct_sessions_to_lovebear
A. pre_product_2_launch	15696	11347	72.2923	11347	72.2923	0	0
B. post_product_2_launch	10709	8200	76.5711	6654	62.1347	1546	14.4365

The percentage of products pageviews that clicked to Mr. Fuzzy has gone down since the launch of the Love Bear, but the overall clickthrough rate has gone up.

4) Building product conversion funnels

Scenario: I'd like to look at our two products since January 6th and analyze the conversion funnels from each product page to conversion. It would be great if you could produce a comparison between the two conversion funnels, for all website traffic.

Result:

product_seen	sessions	to_cart	to_shipping	to_billing	to_thankyou
lovebear	1599	877	603	488	301
mrfuzzy	6985	3038	2084	1710	1088

product_seen	sessions	product_page_click_rate	cart_click_rate	shipping_click_rate	billing_click_rate
lovebear	1599	54.8468	68.7571	80.9287	61.6803
mrfuzzy	6985	43.4932	68.5978	82.0537	63.6257

Adding a second product increased overall click through rate (CTR) from the products page. The second product has comparable rates throughout the rest of funnel. It seems that second product was also a great addition to the business.

5) Cross-sell analysis

Scenario: On September 25th we started giving customers the option to add a 2nd product while on the /cart page. Morgan says this has been positive, but I'd like your take on it.

Could you please compare the month before vs the month after the change? I'd like to see CTR from the /cart page, Avg Products per Order, Average order value (AOV), and overall revenue per /cart page view.

Result:

time_period	cart_sessions	clickthroughs	%cart_ctr	orders_placed	products_purchased	products_per_order	revenue	aov	rev_per_cart_session
A. Pre_Cross_Sell	1830	1229	67.1585	652	652	1	33523.48	51.41638	18.318842
B. Post_Cross_Sell	1975	1351	68.4051	671	701	1.0447	36402.99	54.251848	18.431894

It looks like the CTR from the cart page did not go down after introducing the cross-sell feature and all the metrics are slightly up since the feature was added.

6) Product portfolio expansion

Scenario: On December 12th, 2013, we launched a third product targeting the birthday gift market (Birthday Bear). Could you please run a pre-post analysis comparing the month before vs. the month after, in terms of session-to-order conversion rate, AOV, products per order, and revenue per session?

Result:

time_period	sessions	orders	% conv_rate	total_revenue	total_products_sold	average_order_value	products_per_order	revenue_per_session
A. Pre_Birthday_Bear	17343	1055	6.0831	57208.96	1104	54.226502	1.0464	3.298677
B. Post_Birthday_Bear	13383	940	7.0238	53515.44	1056	56.931319	1.1234	3.998763

It seems all of the metrics have improved a bit after the launch of 3rd product.

7) Analyzing product refund rates

Scenario: Our Mr. Fuzzy supplier had some quality issues which weren't corrected until September 2013. Then they had a major problem where the bears' arms were falling off in Aug/Sep 2014. As a result, we replaced them with a new supplier on September 16, 2014. Can you please pull monthly product refund rates, by product, and confirm our quality issues are now fixed?

Result:

yr	mo	p1_orders	p1_refund_rt	p2_orders	p2_refund_rt	p3_orders	p3_refund_rt	p4_orders	p4_refund_rt
2012	3	60	1.6667	0	NULL	0	NULL	0	NULL
2012	4	99	5.0505	0	NULL	0	NULL	0	NULL
2012	5	108	3.7037	0	NULL	0	NULL	0	NULL
2012	6	140	5.7143	0	NULL	0	NULL	0	NULL
2012	7	169	8.284	0	NULL	0	NULL	0	NULL
2012	8	228	7.4561	0	NULL	0	NULL	0	NULL
2012	9	287	9.0592	0	NULL	0	NULL	0	NULL
2012	10	371	7.2776	0	NULL	0	NULL	0	NULL
2012	11	618	7.4434	0	NULL	0	NULL	0	NULL
2012	12	506	5.9289	0	NULL	0	NULL	0	NULL
2013	1	343	4.9563	47	2.1277	0	NULL	0	NULL
2013	2	336	7.1429	162	1.2346	0	NULL	0	NULL
2013	3	320	5.625	65	4.6154	0	NULL	0	NULL
2013	4	459	4.1394	94	1.0638	0	NULL	0	NULL
2013	5	489	6.3395	82	2.439	0	NULL	0	NULL
2013	6	503	7.7535	90	5.5556	0	NULL	0	NULL
2013	7	509	7.2692	95	3.1579	0	NULL	0	NULL
2013	8	510	5.4902	98	1.0204	0	NULL	0	NULL
2013	9	537	4.2831	98	1.0204	0	NULL	0	NULL
2013	10	603	2.8192	135	1.4815	0	NULL	0	NULL
2013	11	724	3.453	174	2.2989	0	NULL	0	NULL
2013	12	818	2.3227	183	2.1858	139	7.1942	0	NULL
2014	1	728	4.2582	183	2.1858	200	6.5	0	NULL
2014	2	584	3.9384	351	1.7094	211	6.6351	202	0.9901
2014	3	785	3.0573	193	1.5544	244	6.9672	205	0.4878
2014	4	917	3.4896	214	1.8692	267	6.7416	259	1.5444
2014	5	1030	2.9126	246	1.626	299	5.6856	298	0.6711
2014	6	893	5.7111	245	3.6735	288	5.5556	249	2.4096
2014	7	961	4.3704	244	3.6885	276	3.9855	264	1.5152
2014	8	958	13.7787	237	1.6878	294	6.8027	303	0.6601
2014	9	1056	13.2576	251	3.1873	317	6.6246	327	1.2232
2014	10	513	2.729	135	0.7407	165	4.8485	155	3.2258

It seems the refund rates for Product 1 did go down after the initial improvements in September 2013, but they were terribly high from Aug-Sep 2014 (13-14%).

4. User-Level Analysis

1) Identifying repeat visitors

Scenario: We've been thinking about customer value based solely on their first session conversion and revenue. But if customers have repeat sessions, they may be more valuable than we thought. If that's the case, we might be able to spend a bit more to acquire them. Could you please pull data on how many of our website visitors come back for another session? 2014 to date is good.

Result:

repeat_sessions	users
0	126813
1	14086
2	315
3	4686

It seems a fair number of customers do come back after their first visit.

2) Analyzing time to repeat

Scenario: Ok, so the repeat session data was really interesting to see. Now you've got me curious to better understand the behavior of these repeat customers. Could you help me understand the minimum, maximum, and average time between the first and second session for customers who do come back? Again, analyzing 2014 to date is probably the right time period.

Result:

avg_days_first_to_second	min_days_first_to_second	max_days_first_to_second
33.2622	1	69

3) Analyzing repeat channel behaviour

Scenario: Can you help me understand the channels they come back through? Curious if it's all direct type-in, or if we're paying for these customers with paid search ads multiple times. Comparing new vs. repeat sessions by channel would be really valuable, if you're able to pull it! 2014 to date is great.

Result:

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

All the repeat visitors are coming only through organic search, direct type-in and paid brand with 33% of visitors coming through paid channel.

4) Analyzing new and repeat conversion rates

Scenario: I'd love to do a comparison of conversion rates and revenue per session for repeat sessions vs new sessions. Let's continue using data from 2014, year to date.

Result:

is_repeat_session	sessions	orders	% conv_rate	total_revenue	rev_per_session
0	149787	10179	6.7956	650637.91	4.343754
1	33577	2724	8.1127	173553.75	5.168828

It looks like repeat sessions are more likely to convert and produce a little bit more revenue per session.