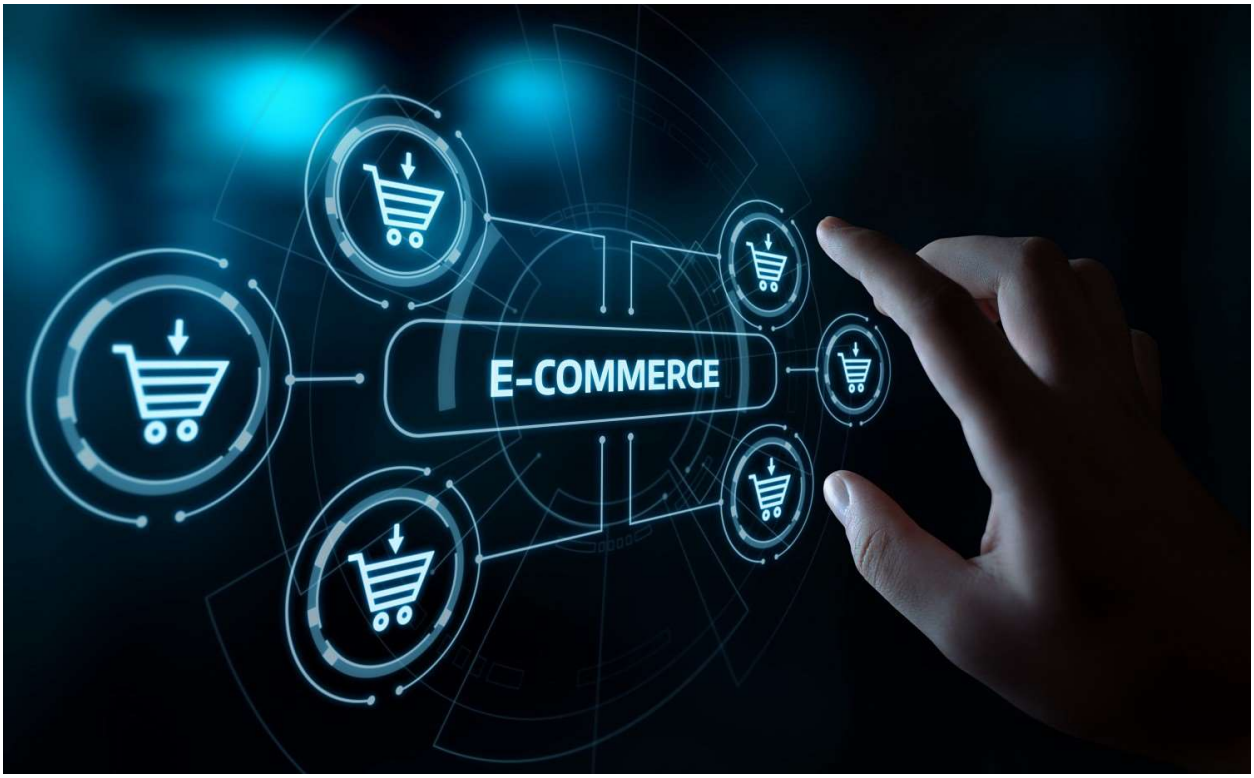
eCommerce Analysis

Optimization for an online store

January 2022

By: Juan Carlos Benavides B



Topics:

- Traffic sources analysis
- Website performance analysis
- Channel portfolio analysis
- Business patterns analysis
- Product analysis

Tools:

- SQL
- Tableau

Introduction

The following report is an exercise about optimization queries in an online store.

The date base has been taken from the course: Advanced SQL for-analytics business intelligence in Udemy.

In the example an eCommerce store will initiate operations in the year 2012 with one product: the *Mr. Fuzzy bear*, and then for the year 2013 will incorporate the second product: the *Love bear*.

The scope of this project is to analyze the following topics throughout the activity of the store:

- Traffic sources analysis
- Website performance analysis
- Channel portfolio analysis
- Business patterns analysis
- Product analysis

The tables contained in the data base are the following:

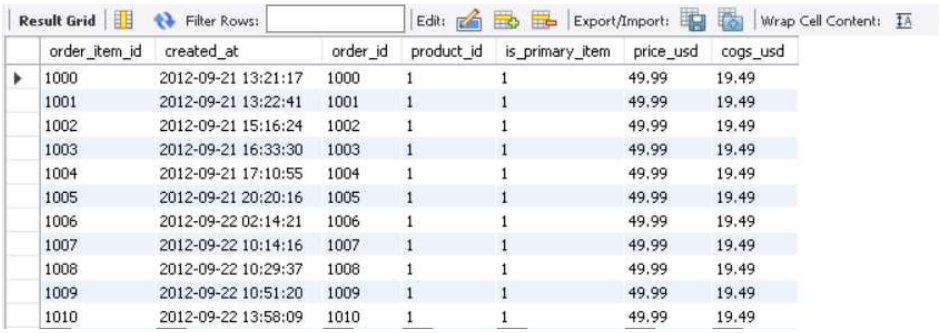
For better detail and understanding of information contained in the tables, a summary of the used tables are shown as follows:

Orders table:

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content:

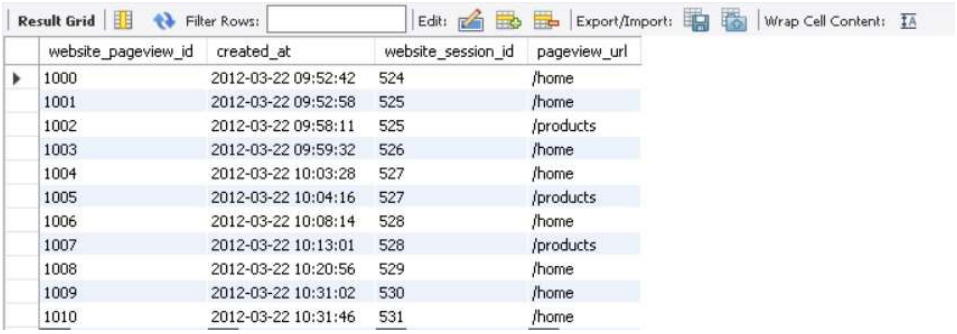
order_id	created_at	website_session_id	user_id	primary_product_id	items_purchased	price_usd	cogs_usd
1000	2012-09-21 13:21:17	28158	25828	1	1	49.99	19.49
1001	2012-09-21 13:22:41	28157	25827	1	1	49.99	19.49
1002	2012-09-21 15:16:24	28190	25855	1	1	49.99	19.49
1003	2012-09-21 16:33:30	28218	25882	1	1	49.99	19.49
1004	2012-09-21 17:10:55	28234	25896	1	1	49.99	19.49
1005	2012-09-21 20:20:16	28276	25935	1	1	49.99	19.49
1006	2012-09-22 02:14:21	28319	25970	1	1	49.99	19.49
1007	2012-09-22 10:14:16	28352	21698	1	1	49.99	19.49
1008	2012-09-22 10:29:37	28354	25999	1	1	49.99	19.49
1009	2012-09-22 10:51:20	28357	20444	1	1	49.99	19.49
1010	2012-09-22 13:58:09	28383	26024	1	1	49.99	19.49

Order items table:



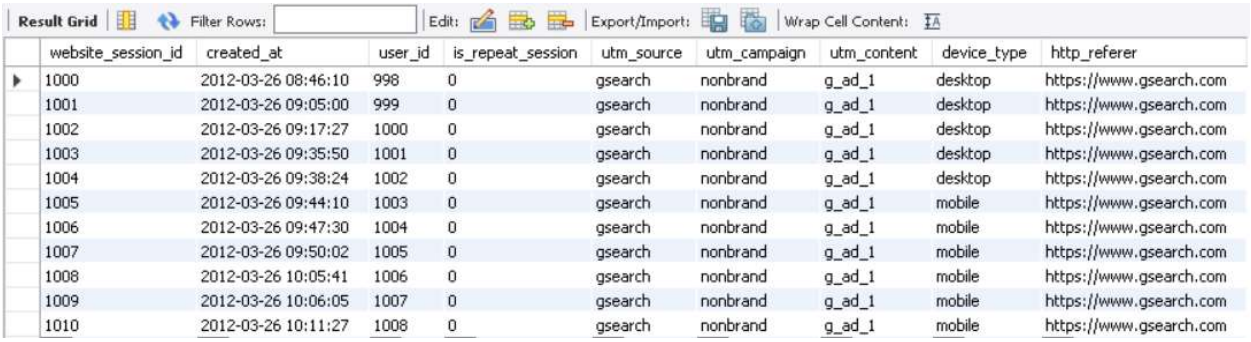
order_item_id	created_at	order_id	product_id	is_primary_item	price_usd	cogs_usd
1000	2012-09-21 13:21:17	1000	1	1	49.99	19.49
1001	2012-09-21 13:22:41	1001	1	1	49.99	19.49
1002	2012-09-21 15:16:24	1002	1	1	49.99	19.49
1003	2012-09-21 16:33:30	1003	1	1	49.99	19.49
1004	2012-09-21 17:10:55	1004	1	1	49.99	19.49
1005	2012-09-21 20:20:16	1005	1	1	49.99	19.49
1006	2012-09-22 02:14:21	1006	1	1	49.99	19.49
1007	2012-09-22 10:14:16	1007	1	1	49.99	19.49
1008	2012-09-22 10:29:37	1008	1	1	49.99	19.49
1009	2012-09-22 10:51:20	1009	1	1	49.99	19.49
1010	2012-09-22 13:58:09	1010	1	1	49.99	19.49

Website pageviews table:



website_pageview_id	created_at	website_session_id	pageview_url
1000	2012-03-22 09:52:42	524	/home
1001	2012-03-22 09:52:58	525	/home
1002	2012-03-22 09:58:11	525	/products
1003	2012-03-22 09:59:32	526	/home
1004	2012-03-22 10:03:28	527	/home
1005	2012-03-22 10:04:16	527	/products
1006	2012-03-22 10:08:14	528	/home
1007	2012-03-22 10:13:01	528	/products
1008	2012-03-22 10:20:56	529	/home
1009	2012-03-22 10:31:02	530	/home
1010	2012-03-22 10:31:46	531	/home

Website sessions table:



website_session_id	created_at	user_id	is_repeat_session	utm_source	utm_campaign	utm_content	device_type	http_referer
1000	2012-03-26 08:46:10	998	0	gsearch	nonbrand	g_ad_1	desktop	https://www.gsearch.com
1001	2012-03-26 09:05:00	999	0	gsearch	nonbrand	g_ad_1	desktop	https://www.gsearch.com
1002	2012-03-26 09:17:27	1000	0	gsearch	nonbrand	g_ad_1	desktop	https://www.gsearch.com
1003	2012-03-26 09:35:50	1001	0	gsearch	nonbrand	g_ad_1	desktop	https://www.gsearch.com
1004	2012-03-26 09:38:24	1002	0	gsearch	nonbrand	g_ad_1	desktop	https://www.gsearch.com
1005	2012-03-26 09:44:10	1003	0	gsearch	nonbrand	g_ad_1	mobile	https://www.gsearch.com
1006	2012-03-26 09:47:30	1004	0	gsearch	nonbrand	g_ad_1	mobile	https://www.gsearch.com
1007	2012-03-26 09:50:02	1005	0	gsearch	nonbrand	g_ad_1	mobile	https://www.gsearch.com
1008	2012-03-26 10:05:41	1006	0	gsearch	nonbrand	g_ad_1	mobile	https://www.gsearch.com
1009	2012-03-26 10:06:05	1007	0	gsearch	nonbrand	g_ad_1	mobile	https://www.gsearch.com
1010	2012-03-26 10:11:27	1008	0	gsearch	nonbrand	g_ad_1	mobile	https://www.gsearch.com

Conclusions and recommendations will be obtained for every ***request***.

In some tasks visualizations made in Tableau will be shown.

For more details for every query see the scrips of codes for *MySQL*.

ANALYZING TRAFFIC SOURCES

REQUEST 1

Request date: 2012-04-12

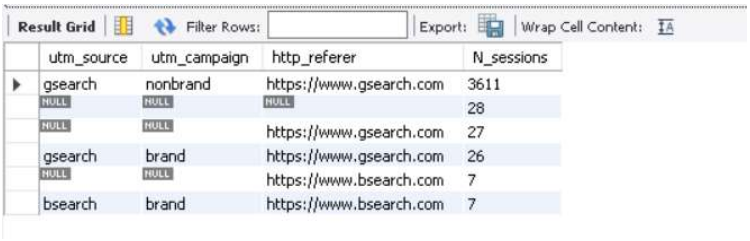
Determine which is the main source of traffic (utm_source and utm_campaign) for the website.

Then for this source: calculate conversion rate **from session to order.**

Based on cost by clicks and metrics the company is expecting to get a conversion rate at least of 4%

Analysis and Output:

The main source of traffic can be obtained from ***website_sessions*** table:



Result Grid | Filter Rows: | Export: | Wrap Cell Content:

utm_source	utm_campaign	http_referer	N_sessions
gsearch	nonbrand	https://www.gsearch.com	3611
NULL	NULL	NULL	28
NULL	NULL	https://www.gsearch.com	27
gsearch	brand	https://www.gsearch.com	26
NULL	NULL	https://www.bsearch.com	7
bsearch	brand	https://www.bsearch.com	7

Then using ***orders*** and ***website_sessions*** tables the conversion rate is equal to 2.96%

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

sessions	orders	cv_rate
3611	107	0.0296

Conclusion:

Based on this conversion rate and the expected value (4%) the company is over spending in this campaigns.

Recommendation:

It is necessary to dial down the search bids.

REQUEST 2

Request date: 2012-05-10

Based on the last recommendation the paid traffic has been bid down since 2012-04-15.

To confirm if the bid changes have caused volume to drop, it is necessary to pull the trending on session volume by week.

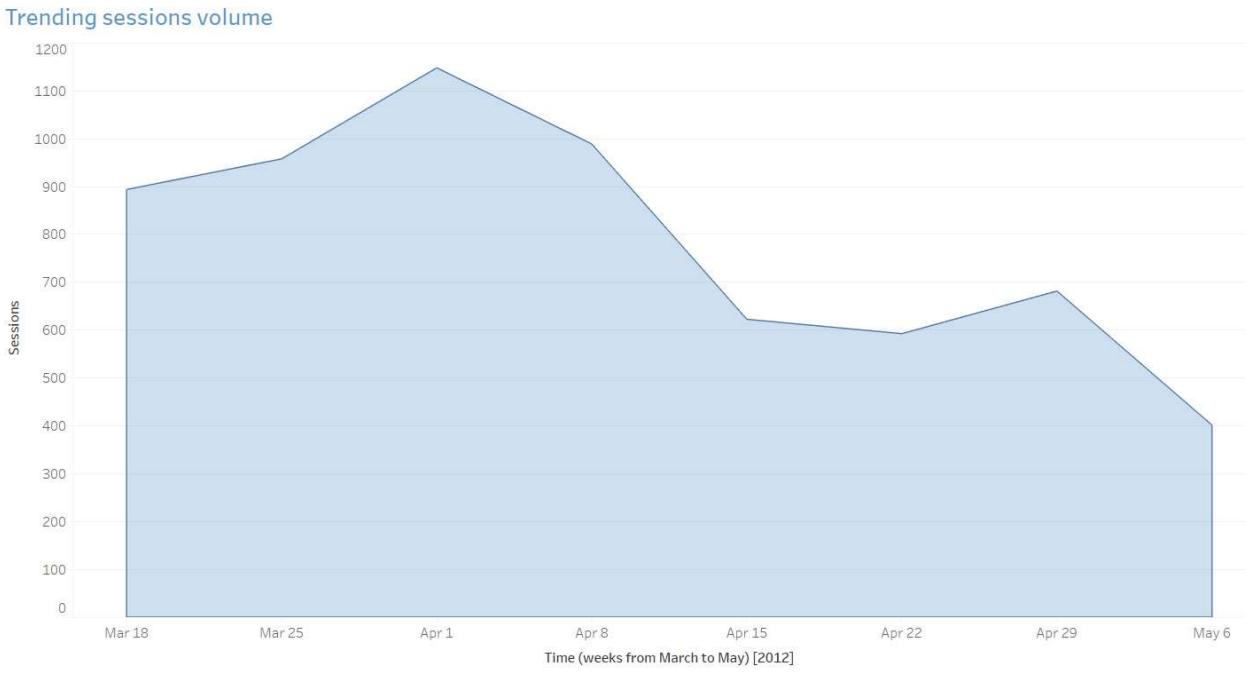
Analysis and Output:

The volume of sessions by week can be summarized from ***website_sessions*** table as follows:

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

week_started_at	sessions
2012-03-19	893
2012-03-25	957
2012-04-01	1147
2012-04-08	988
2012-04-15	622
2012-04-22	592
2012-04-29	681
2012-05-06	401

For a better understanding this result can be visualized as:



Conclusion:

This analysis shows this channel of traffic (gsearch and nonbrand) is fairly sensitive to bid changes.

This chart shows sessions have decreased since the date paid ads were bid down (2012-04-15).

Recommendation:

Because what is looked for is the maximum volume of sessions and do not spend more on advertisements than can be afforded, it is recommended to continue tracking conversion rates and theses trending and to continue with a similar strategy of try and test to obtained the best tuning.

ANALYZING TRAFFIC SOURCES

REQUEST 3

Request date: 2012-06-28

A new home page (lander) has been designed, then it is required to run an A/B test to calculate bouncing rates and to evaluate which one has a better performance.

Analysis and Output:

From the ***website_sessions*** and ***website_pageviews*** tables the analysis can be broken into five parts:

STEP 0: find out when the new page/lander launched.

STEP 1: find the first website_pageview_id for relevant sessions.

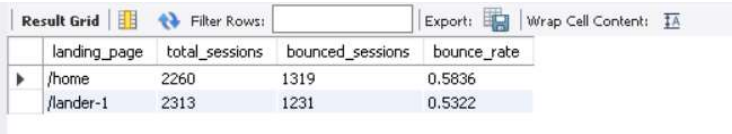
STEP 2: identify the landing page of each session.

STEP 3: counting pageviews for each session, to identify "bounces".

Finally:

STEP 4: summarizing total sessions and bounce sessions, by landing page.

Since all of these the bounce rate can be calculated for the original home page (58.36%) and for the new and tested lander page (53.22)%



landing_page	total_sessions	bounced_sessions	bounce_rate
/home	2260	1319	0.5836
/lander-1	2313	1231	0.5322

Conclusion:

The new home page has a better performance than the original since it has a fewer bounce rate.

Recommendation:

Company has to consider to update the campaigns pointing the traffic to the new page.

REQUEST 4

Request date: 2012-08-31

After a while to confirm all the traffic is routed correctly. It is necessary to pull the trending volume of paid traffic by week for the original home page and the new one.

Analysis and Output:

From the ***website_sessions*** table the analysis can be broken into four parts:

STEP 1: find the first website_pageview_id for relevant sessions.

STEP 2: identify the landing page of each session.

STEP 3: counting pageviews for each session, to identify "bounces".

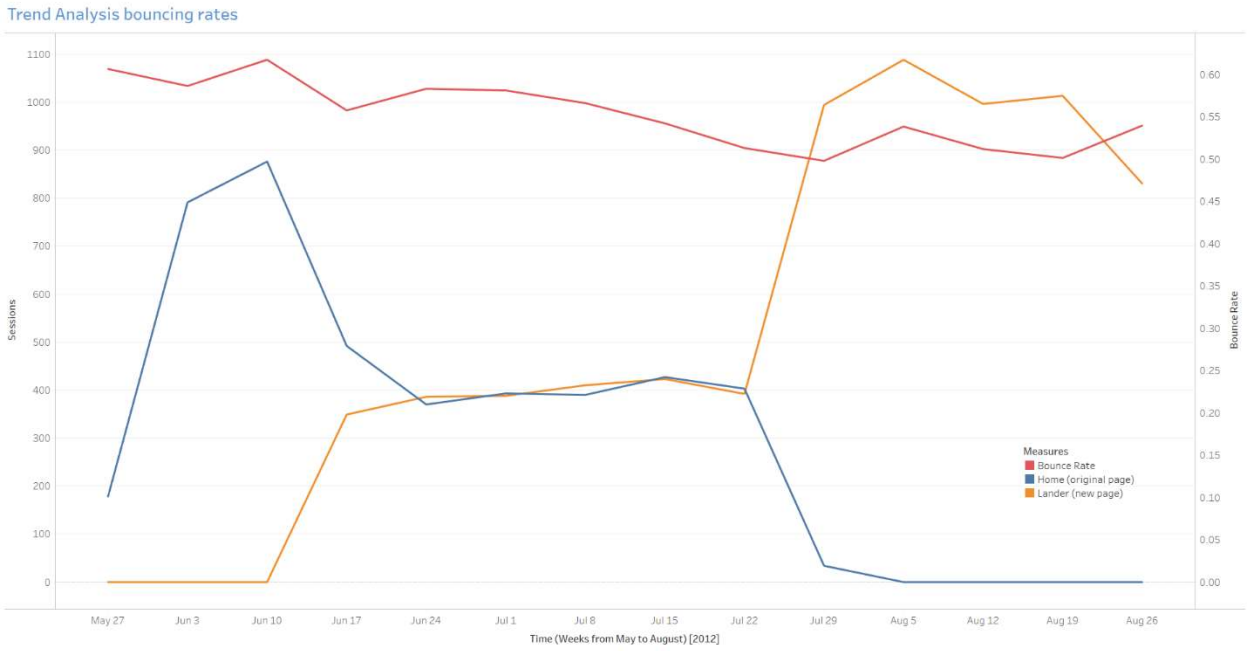
STEP 4: summarizing by week (bounce rate, sessions to each lander).

The trend obtained can be summarized as follows:

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

year_week	week_start_date	total_sessions	bounced_sessions	bounce_rate	home_sessions	lander_sessions
201222	2012-06-01	178	108	0.6067	178	0
201223	2012-06-03	791	464	0.5866	791	0
201224	2012-06-10	876	541	0.6176	876	0
201225	2012-06-17	841	469	0.5577	492	349
201226	2012-06-24	756	441	0.5833	370	386
201227	2012-07-01	781	454	0.5813	393	388
201228	2012-07-08	800	453	0.5663	390	410
201229	2012-07-15	850	461	0.5424	427	423
201230	2012-07-22	795	408	0.5132	403	392
201231	2012-07-29	1028	512	0.4981	34	994
201232	2012-08-05	1088	586	0.5386	0	1088
201233	2012-08-12	996	510	0.5120	0	996
201234	2012-08-19	1013	508	0.5015	0	1013
201235	2012-08-26	830	448	0.5398	0	830

A better understanding can be reached with the following visualization:



Conclusion:

Both pages were getting traffic for a while and then company fully switched to the new home page. And the overall bounce rate has come down over time.

Recommendation:

If this trend continues soon will be necessary to point **all** the traffic to the new home page.

REQUEST 5

Request date: 2012-09-05

Now the new home page is working it is necessary to understand all the path from home to placed orders, and where is it possible the company is losing visitors?

The request is to build a conversion funnel analyzing how many customers make it to each step, from August 5th.

Analysis and Output:

From the tables: ***website_sessions*** and ***website_pageviews*** tables the analysis can be broken into the steps:

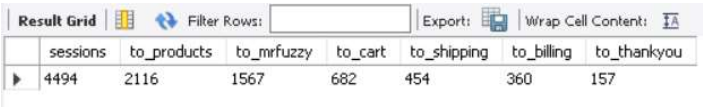
STEP 1: select all pageviews for relevant sessions.

STEP 2: identify each relevant pageviews as the specific funnel step.

STEP 3: create the session-level conversion funnel view.

STEP 4: aggregate the data to assess funnel performance.

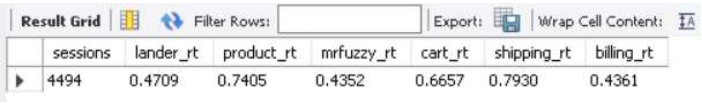
In this way the obtained conversion funnel is:



Result Grid | Filter Rows: | Export: | Wrap Cell Content:

sessions	to_products	to_mrfuzzy	to_cart	to_shipping	to_billing	to_thankyou
4494	2116	1567	682	454	360	157

The last result into rates will be:



Result Grid | Filter Rows: | Export: | Wrap Cell Content:

sessions	lander_rt	product_rt	mrfuzzy_rt	cart_rt	shipping_rt	billing_rt
4494	0.4709	0.7405	0.4352	0.6657	0.7930	0.4361

Conclusion:

Once the customers are visiting the company´s website has low click rates in the pages: product_1 (mrfuzzy) and billing.

Recommendation:

The company must consider to run an A/B test for the */mrfuzzy* and */billing* steps in the process of placing orders, in order to avoid losing purchases.

ANALYSIS FOR CHANNEL PORTFOLIO MANAGMENT

REQUEST 6

Request date: 2012-11-30

The company wants to know the current performance of sources and devices type.

It is required to calculate traffic percentage from mobile and conversion rates for main sources and by device type, considering dates since 2012-08-22.

Analysis and Output:

Using the ***website_sessions*** table can be obtained the traffic coming to website from mobiles for the two main sources:

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

utm_source	sessions	mobile_sessions	pct_mobile
bsearch	6521	562	0.0862
gsearch	20072	4922	0.2452

In a similar way using the ***oders*** table and ***website_sessions*** table can be obtained the conversion rates for sources and device types:

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

device_type	utm_source	sessions	orders	cvRate
desktop	bsearch	5959	303	0.0508
desktop	gsearch	15150	750	0.0495
mobile	bsearch	562	5	0.0089
mobile	gsearch	4922	77	0.0156

Conclusion:

Traffic channels do not perform identically.

bsearch has lower conversion rate for both: mobile and desktop.

Recommendation:

It is recommended to differentiate the bids based on the channel, and bid down for *bsearch*.

REQUEST 7

Request date: 2012-12-22

The company bided down the paid traffic for *bsearch* since December 2nd.

Now to compare the results it is necessary to pull a trend of weekly session volume for *gsearch* and *bsearch*, considering dates since November 4th.

Analysis and Output:

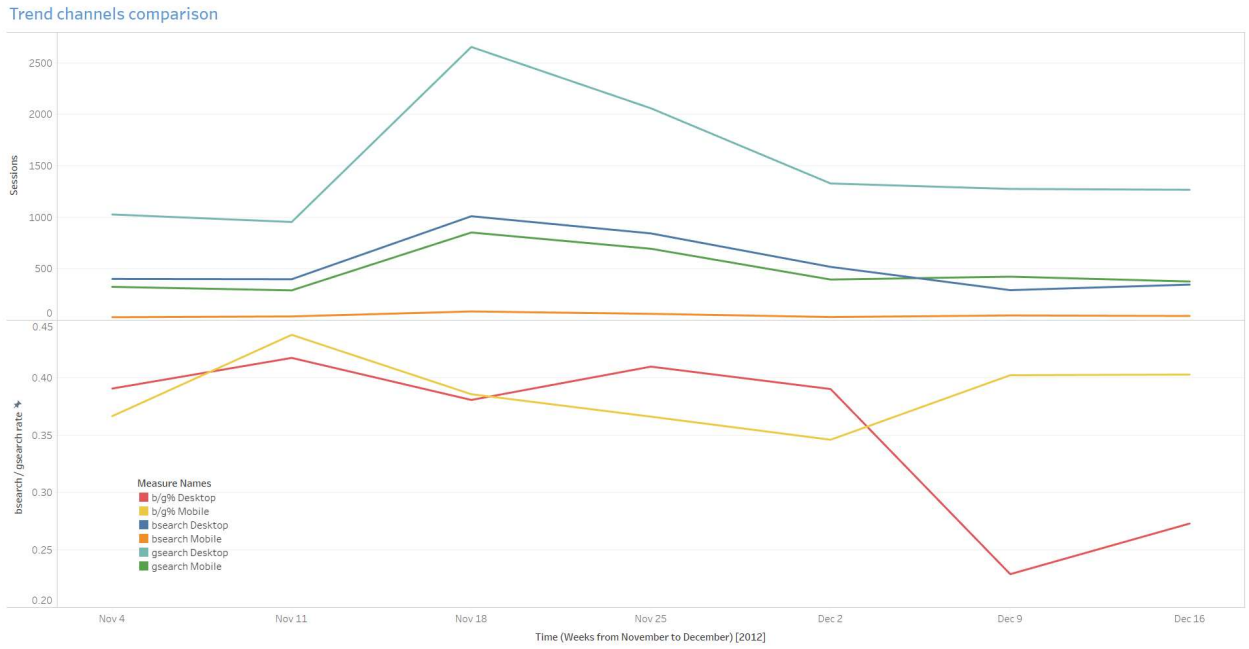
To simplify the analysis and metric comparison the rate will be calculated as *bsearch* percentage of *gsearch*.

The trend con be obtained from the ***website_sessions*** table, as follows:

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

week_start_date	g_dtop_sessions	b_dtop_sessions	b_pct_of_g_dtop	week_start_date	g_mob_sessions	b_mob_sessions	b_pct_of_g_mob
2012-11-04	1027	401	0.3905	2012-11-04	324	29	0.0895
2012-11-11	954	398	0.4172	2012-11-11	290	37	0.1276
2012-11-18	2654	1010	0.3806	2012-11-18	852	85	0.0998
2012-11-25	2058	843	0.4096	2012-11-25	694	62	0.0893
2012-12-02	1328	518	0.3901	2012-12-02	395	31	0.0785
2012-12-09	1275	292	0.2290	2012-12-09	423	46	0.1087
2012-12-16	1267	346	0.2731	2012-12-16	376	41	0.1090

This can be understood better from the following visualization:



Conclusion:

It is necessary to consider extraordinary traffic due to the analyzed dates (Black Friday and Cyber Monday) but it can be concluded that both *bsearch* and *gsearch*, but specially *bsearch* traffic dropped after the bid down.

ANALYSIS FOR CHANNEL PORTFOLIO MANAGMENT

REQUEST 8

Request date: 2013-01-02

The company is considering to improve the customer experience, and it will be necessary to hire new staff for hours with more traffic.

It is needed an analysis of the average website session volume, by hour of day and by day week.

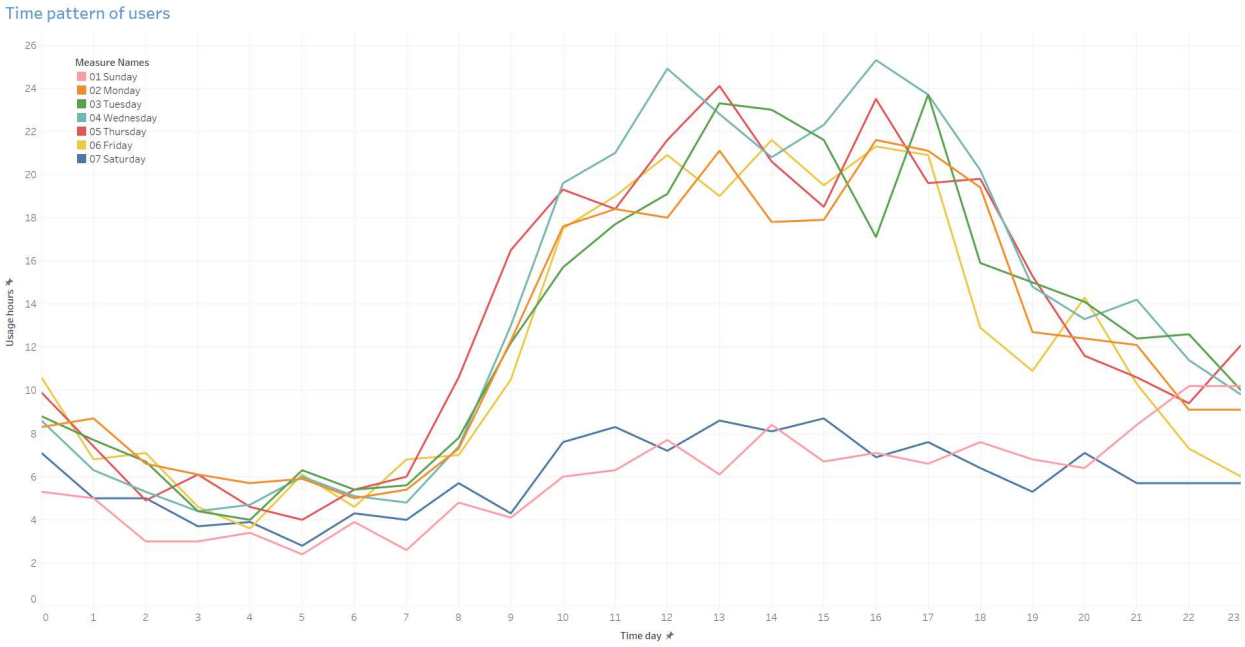
Analysis and Output:

The information from the ***website_sessions*** table can be summarized as follows:

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

hr	mo	tue	wed	thu	fri	sat	sun
0	8.3	8.8	8.6	9.9	10.6	7.1	5.3
1	8.7	7.7	6.3	7.4	6.8	5.0	5.0
2	6.6	6.7	5.3	4.9	7.1	5.0	3.0
3	6.1	4.4	4.4	6.1	4.6	3.7	3.0
4	5.7	4.0	4.7	4.6	3.6	3.9	3.4
5	5.9	6.3	6.0	4.0	6.1	2.8	2.4
6	5.0	5.4	5.1	5.4	4.6	4.3	3.9
7	5.4	5.6	4.8	6.0	6.8	4.0	2.6
8	7.3	7.8	7.4	10.6	7.0	5.7	4.8
9	12.3	12.2	13.0	16.5	10.5	4.3	4.1
10	17.6	15.7	19.6	19.3	17.5	7.6	6.0
11	18.4	17.7	21.0	18.4	19.0	8.3	6.3
12	18.0	19.1	24.9	21.6	20.9	7.2	7.7
13	21.1	23.3	22.8	24.1	19.0	8.6	6.1
14	17.8	23.0	20.8	20.6	21.6	8.1	8.4
15	17.9	21.6	22.3	18.5	19.5	8.7	6.7
16	21.6	17.1	25.3	23.5	21.3	6.9	7.1
17	21.1	23.7	23.7	19.6	20.9	7.6	6.6
18	19.4	15.9	20.2	19.8	12.9	6.4	7.6
19	12.7	15.0	14.8	15.3	10.9	5.3	6.8
20	12.4	14.1	13.3	11.6	14.3	7.1	6.4
21	12.1	12.4	14.2	10.6	10.3	5.7	8.4
22	9.1	12.6	11.4	9.4	7.3	5.7	10.2
23	9.1	10.0	9.8	12.1	6.0	5.7	10.2

For a better analysis this table can be represented in a chart like this:



Conclusion:

The time with more traffic can be considered from 8 am to 5pm.

The days Sundays and Saturdays are days with the fewest traffic.

Recommendation:

The company should considerer to increase the staff from 8am to 5pm on Mondays to Friday.

PRODUCT ANALYSIS

REQUEST 9

Request date: 2013-04-05

The company launched a new product since January 6th, 2013. Now it is necessary if the second product has improved the overall conversion rate and revenues per session. And to compare the performance of each product.

For the analysis to consider time period since April 1st, 2012.

Analysis and Output:

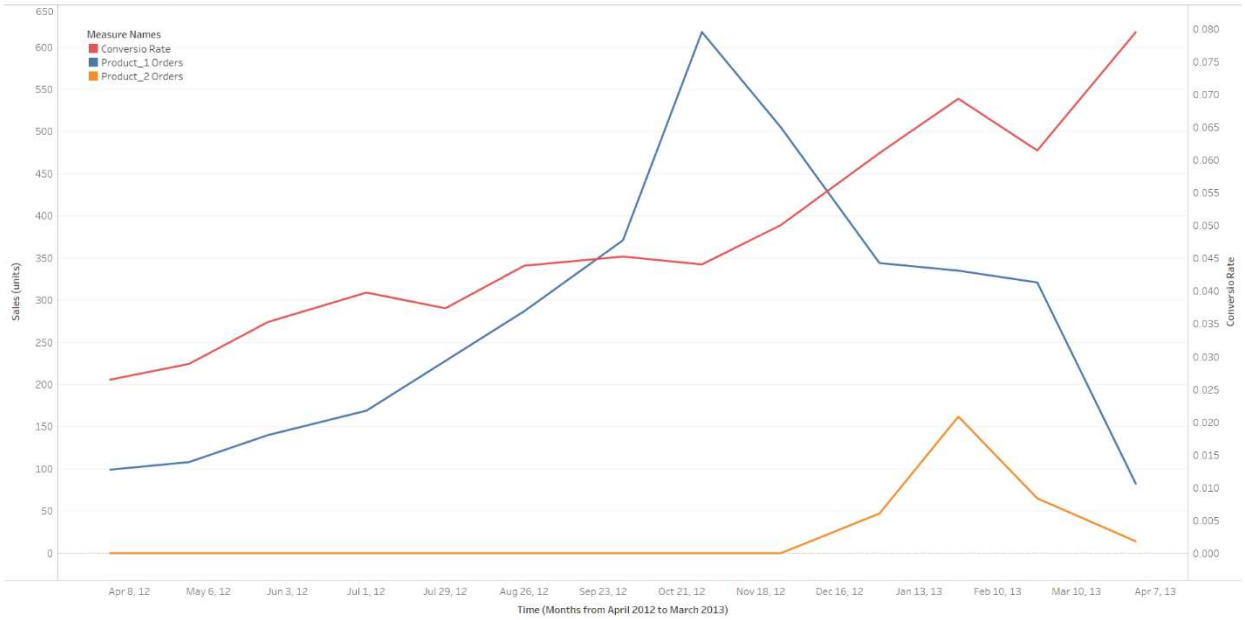
From the table ***website_sessions*** and the table ***orders***, it can be seen the trends of sales:

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

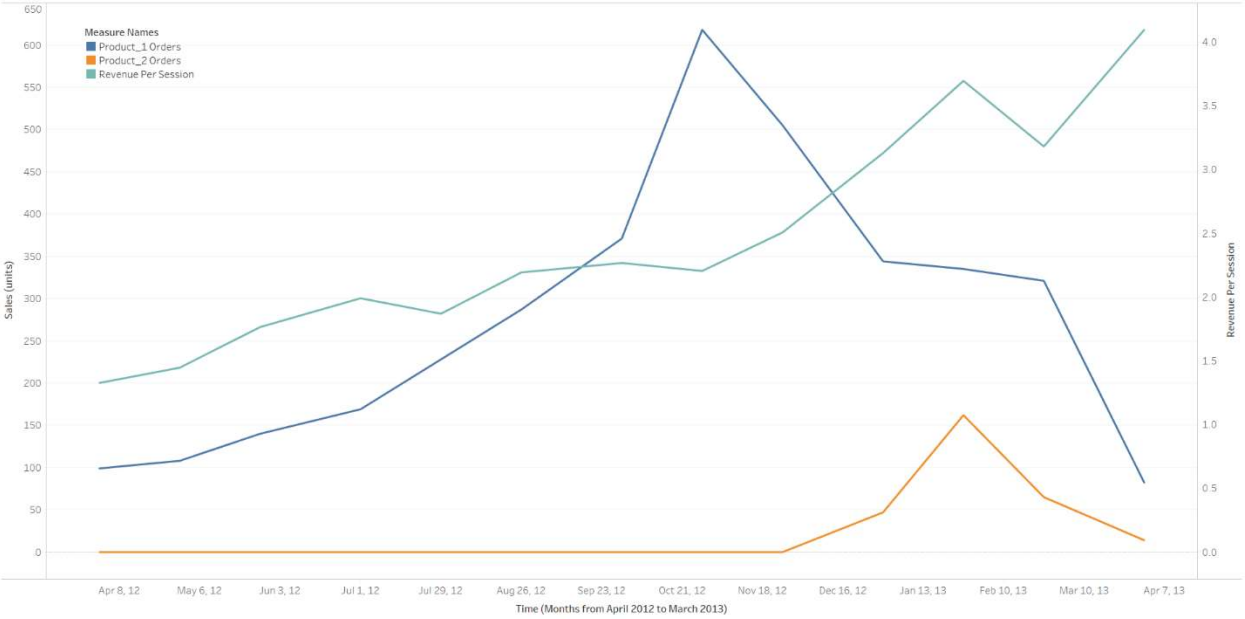
yr	mo	month	sessions	orders	cnv_r	revenue_per_session	product_one_orders	product_two_orders
2012	4	2012-04-01	3732	99	0.0265	1.326101	99	0
2012	5	2012-05-01	3734	108	0.0289	1.445881	108	0
2012	6	2012-06-01	3967	140	0.0353	1.764205	140	0
2012	7	2012-07-01	4245	169	0.0398	1.990179	169	0
2012	8	2012-08-01	6098	228	0.0374	1.869092	228	0
2012	9	2012-09-01	6543	287	0.0439	2.192745	287	0
2012	10	2012-10-01	8182	371	0.0453	2.266718	371	0
2012	11	2012-11-01	14015	618	0.0441	2.204340	618	0
2012	12	2012-12-01	10072	505	0.0501	2.506449	505	0
2013	1	2013-01-01	6398	391	0.0611	3.128492	344	47
2013	2	2013-02-01	7163	497	0.0694	3.694685	335	162
2013	3	2013-03-01	6272	386	0.0615	3.180188	321	65
2013	4	2013-04-01	1206	96	0.0796	4.095390	82	14

This table can be shown in the charts:

Sales product_1 and product_2 vs Conversion Rate



Sales product_1 and product_2 vs Revenue per session



From this analysis is difficult to determine if the growth in sales since January 2013 corresponds to the new product launch or it is a continuation of overall business improvements.

For a better understanding a conversion funnel analysis can be pulled, and calculating clickthrough rates from products page. This can be achieved from the table ***website_pageviews*** and ***products_pageviews***:

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

time_period	sessions	w_next_pg	pct_w_next_pg	to_mrfuzzy	pct_to_mrfuzzy	to_lovebear	pct_to_lovebear
A.Pre_product_2	15694	11348	0.7231	11348	0.7231	0	0.0000
B.Post_product_2	10713	8201	0.7655	6655	0.6212	1546	0.1443

Conclusion:

The percent of products page views that clicked to product_1 (Mr. Fuzzy) has decreased a bit (from 72.3% to 62.1%) since the launch of product_2 (Love Bear), but the overall clickthrough rate has increased (from 72.3% to 76.5%)

Recommendation:

Company should run a complete conversion funnel analysis for each product individually.