Ecommerce Database Analysis

Project Case Study

Srikrishna S - PGPDE

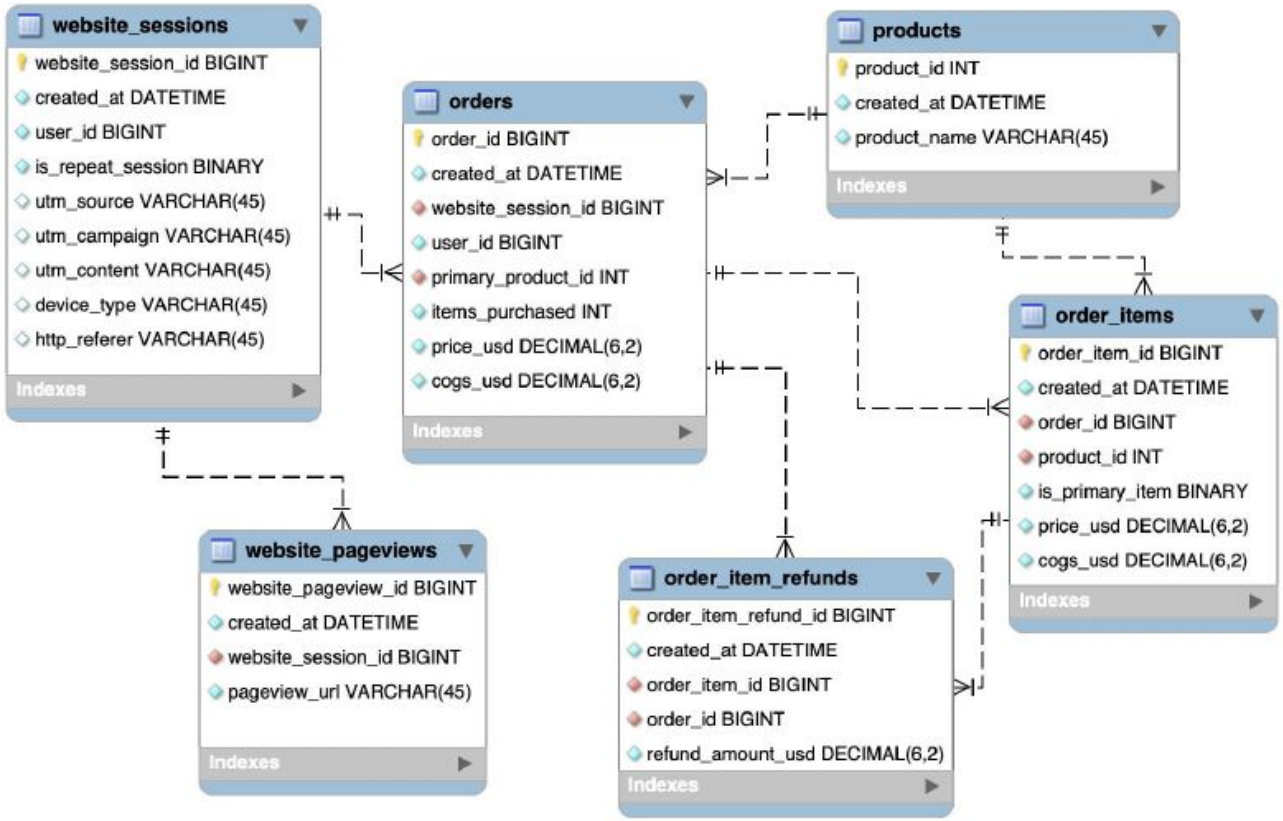
Case Study Infographics

OVERVIEW	As an Ecommerce database analyst for an online retailer that has just launched its first product, we will be working with stakeholders to steer the business.
CHALLENGES	Optimize marketing channels, measure and test website conversion performance, and use data to understand the impact of new product launches.
OBJECTIVE	Explore the database and perform mission critical analyses through SQL querying(MySQL Workbench).
RESULTS	Display result of query under each use case, for next steps to be taken with stakeholders.

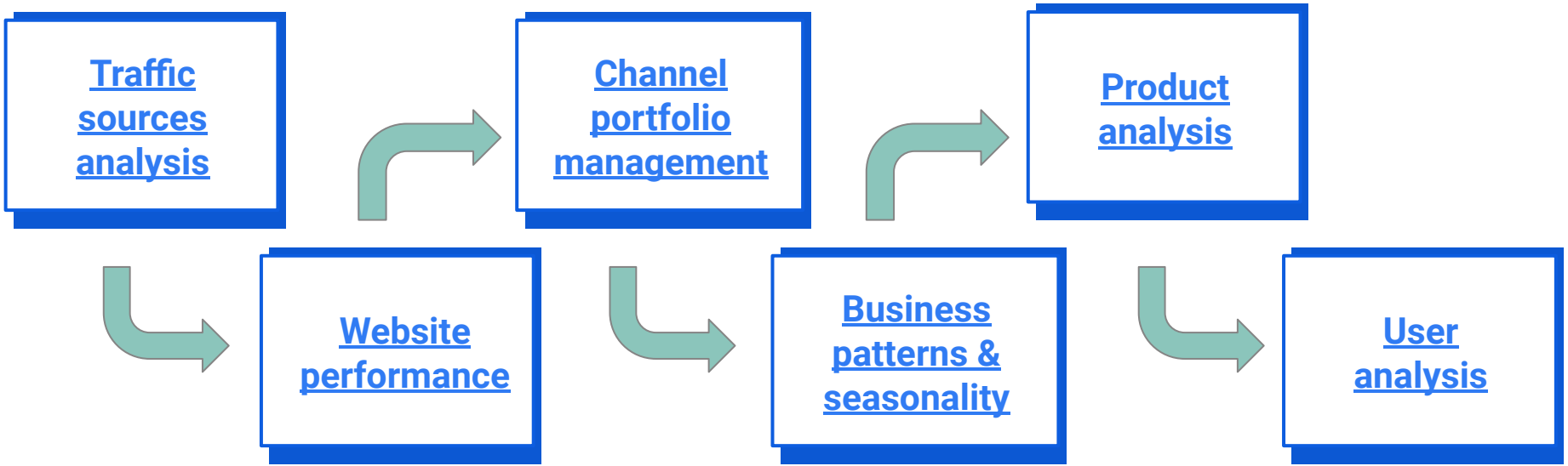
DATABASE OVERVIEW

6 related tables, which contain eCommerce data about:
- Website Activity
- Products
- Orders & Refunds

Understand how customers access and interact with the site, analyze landing page performance & Conversion etc..



BUSINESS CONCEPTS COVERED



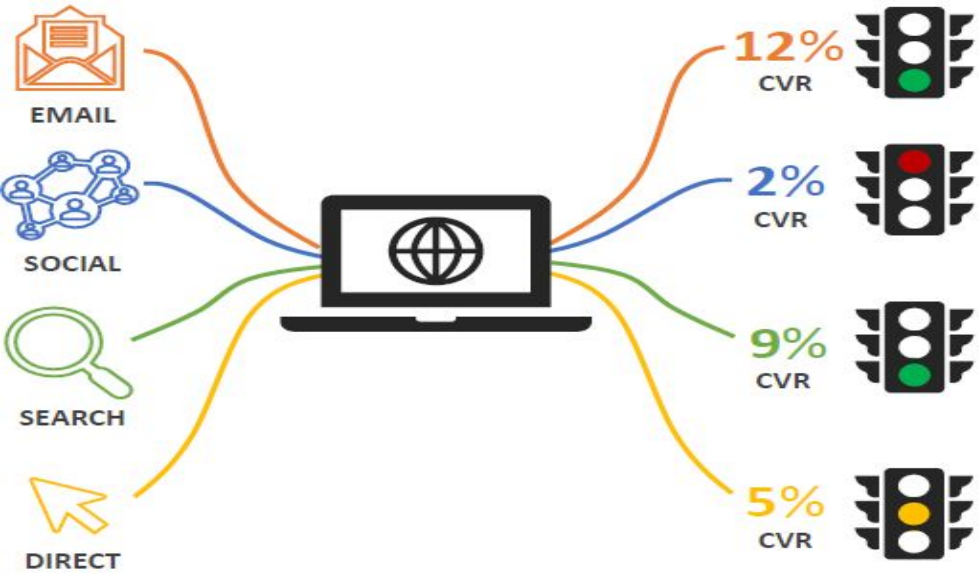
In the project we will go through few use cases from each challenge/concept and analyze the results with the stakeholder to make improvements to the business.

Challenge 1

Analyze Traffic sources

Concept : Traffic source analysis

Traffic source analysis is about understanding where **customers are coming from and which channels are driving the highest quality traffic.**



COMMON USE CASES:

- Analyzing search data and shifting budget towards the engines, campaigns or keywords driving the strongest conversion rates
- Comparing user behavior patterns across traffic sources to inform creative and messaging strategy
- Identifying opportunities to eliminate wasted spend or scale high-converting traffic

KEY TABLES

WEBSITE_SESSIONS

```
SELECT * FROM website_sessions WHERE website_session_id = 1059
```

website_session_id	created_at	user_id	is_repeat_session	utm_source	utm_campaign	utm_content	device_type	http_referer
1059	2012-03-26 13:51:37	1055	0	gsearch	nonbrand	g_ad_1	desktop	https://www.gsearch.com

WEBSITE_PAGEVIEWS

```
SELECT * FROM website_pageviews WHERE website_session_id = 1059
```

website_pageview_id	created_at	website_session_id	pageview_url
2039	2012-03-26 13:51:37	1059	/home
2040	2012-03-26 13:54:27	1059	/products
2041	2012-03-26 13:56:48	1059	/the-original-mr-fuzzy
2042	2012-03-26 14:00:14	1059	/cart
2043	2012-03-26 14:04:06	1059	/shipping
2044	2012-03-26 14:05:47	1059	/billing
2045	2012-03-26 14:13:56	1059	/thank-you-for-your-order

ORDERS

```
SELECT * FROM orders WHERE website_session_id = 1059
```

order_id	created_at	website_session_id	user_id	primary_product_id	items_purchased	price_usd	cogs_usd
32	2012-03-26 14:13:56	1059	1055	1	1	49.99	19.49

UTM TRACKING PARAMETERS

When businesses run paid marketing campaigns, they often obsess over performance and measure everything ; how much they spend, how well traffic converts to sales,

Paid traffic is commonly tagged with tracking (UTM) parameters, which are appended to URLs and allow us to tie website activity back to specific traffic sources and campaigns.

www.abcwebsite.com?utm_source=trafficSource&utm_campaign=campaignName

WEBSITE_SESSIONS

```
SELECT DISTINCT
        utm_source,
        utm_campaign
FROM website_sessions
```

utm_source	utm_campaign
NULL	NULL
bsearch	brand
bsearch	nonbrand
gsearch	brand
gsearch	nonbrand
socialbook	desktop_targeted
socialbook	pilot

FINDING TOP TRAFFIC SOURCES

Problem statement

We need to understand **where the bulk of website sessions are coming from, through before April 12, 2012**.
The result view should be a breakdown by UTM source , campaign and referring domain.

RESULT GRID

Result Grid | Filter Rows: Search | Export:

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

SOLUTION QUERY

CLICK HERE

NEXT STEPS ??

Drill deeper into gsearch nonbrand campaign traffic to explore potential optimization opportunities.

TRAFFIC CONVERSION RATES

Problem statement

Now that **gsearch nonbrand is the major traffic source**, but we need to understand if those sessions are driving sales.
We need to **calculate the conversion rate (CVR)** from session to order. Based on what we're paying for clicks, **we'll need a CVR of at least 4%**.
If we're much lower, we'll need to reduce bids. If we're higher, we can increase bids to drive more volume. Consider study **before April 14, 2012**.

SOLUTION QUERY

CLICK HERE

RESULT GRID

sessions	orders	session_to_order_conv_rt
3895	112	0.0288

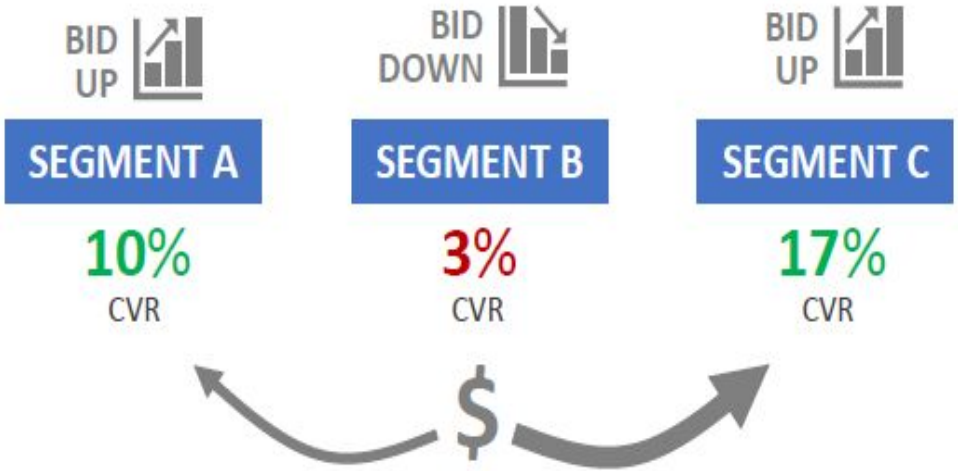
NEXT STEPS ??

Based on this analysis we are over spending based on the current conversion rate.

- *Monitor the impact of bid reductions*
- *Analyze performance trending by device type in order to refine bidding strategy*

Concept : BID OPTIMIZATION

Analyzing for bid optimization is about **understanding the value of various segments of paid traffic, so that you can optimize your marketing budget.**



COMMON USE CASES:

- Using conversion rate and revenue per click analyses to figure out how much you should spend per click to acquire customers
- Understanding how your website and products perform for various subsegments of traffic (i.e.mobile vs desktop) to optimize within channels
- Analyzing the impact that bid changes have on your ranking in the auctions, and the volume of customers driven to your site

TRAFFIC SOURCE TRENDING

Problem statement

Based on the **conversion rate analysis**, we **bid down** gsearch nonbrand on 2012-04-15.
Pull **gsearch nonbrand trended session volume**, by week, to see if the bid changes have caused **volume to drop** at all?
Consider dates **before May 10, 2012.**

RESULT GRID

Result Grid Filter

week_start_date	sessions
2012-03-19	896
2012-03-25	956
2012-04-01	1152
2012-04-08	983
2012-04-15	621
2012-04-22	594
2012-04-29	681
2012-05-06	399

SOLUTION QUERY

CLICK HERE

NEXT STEPS ??

Based on this, gsearch nonbrand is fairly sensitive to bid changes. We want maximum volume, but don't want to spend more on ads than we can afford.

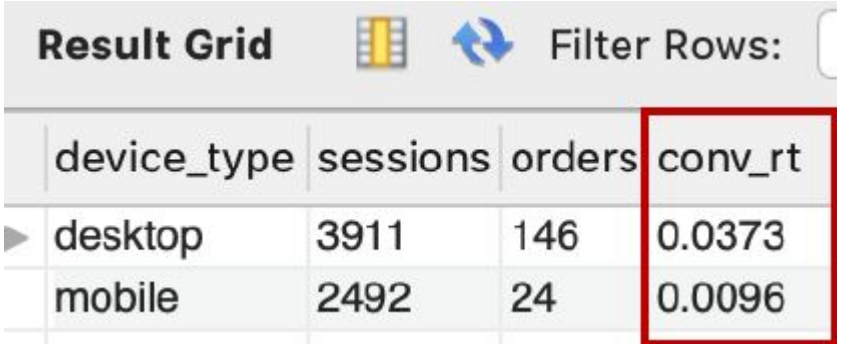
- *Continue to monitor volume levels.*
- *Think about how we could make the campaigns more efficient so that we can increase volume again*

TRAFFIC SOURCE BID OPTIMIZATION

Problem statement

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**?
Consider dates **before May 11, 2012**.

RESULT GRID


Result Grid Filter Rows:

device_type	sessions	orders	conv_rt
desktop	3911	146	0.0373
mobile	2492	24	0.0096

SOLUTION QUERY

CLICK HERE

NEXT STEPS ??

When we bid higher, we'll rank higher in the auctions, so the insights here should lead to a sales boost.

- *Analyze volume by device type to see if the bid changes make a material impact.*
- *Continue to look for ways to optimize campaigns*

TRAFFIC SOURCE SEGMENT TRENDING

Problem statement

After 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**.
Pull **weekly trends for both desktop and mobile** so we can see the impact on volume?
Use **2012-04-15** until the bid change as a baseline. Consider dates **before June 09, 2012**.

RESULT GRID

week_start_date	dtop_sessions	mob_sessions
2012-04-15	383	238
2012-04-22	360	234
2012-04-29	425	256
2012-05-06	430	282
2012-05-13	403	214
2012-05-20	661	190
2012-05-27	585	183
2012-06-03	582	157

SOLUTION QUERY

CLICK HERE

NEXT STEPS ??

It looks like mobile has been pretty flat or a little down, but desktop is looking strong thanks to the bid changes we made based on your previous conversion analysis.

- *Continue to monitor device level volume and be aware of the impact bid levels has.*
- *monitor conversion performance at the device level to optimize spend*

Challenge 2

ANALYZING WEBSITE PERFORMANCE

Concept : ANALYZING TOP WEBSITE CONTENT

Website content analysis is about understanding which pages are seen the most by your users, **to identify where to focus on improving your business**

COMMON USE CASES:

- Finding the most-viewed pages that customers view on your site
- Identifying the most common entry pages to your website – the first thing a user sees
- For most-viewed pages and most common entry pages, understanding how those pages perform for your business objectives

IDENTIFYING TOP WEBSITE PAGES

Problem statement

Get an understanding of the site by pulling the **most viewed website pages, ranked by session volume** before **June 09, 2012.**

RESULT GRID

pageview_url	sessions
/home	10398
/products	4238
/the-original-mr-fuzzy	3036
/cart	1305
/shipping	869
/billing	716
/thank-you-for-your-order	306

SOLUTION QUERY

CLICK HERE

NEXT STEPS ??

It definitely seems like the homepage, the products page, and the Mr. Fuzzy page get the bulk of our traffic .

- *Dig into whether this list is also representative of our top entry pages*
- *Analyze the performance of each of our top pages to look for improvement opportunities*

IDENTIFYING TOP ENTRY PAGES

Problem statement

Pull a list of **the top entry pages** to confirm where our users are hitting the site.
Pull all entry pages and **rank them on entry volume**.
Consider analysis before **June 12, 2012.**

RESULT GRID

Result Grid Filter Rows:

landing_page_url	sessions_hitting_page
/home	10711

SOLUTION QUERY

CLICK HERE

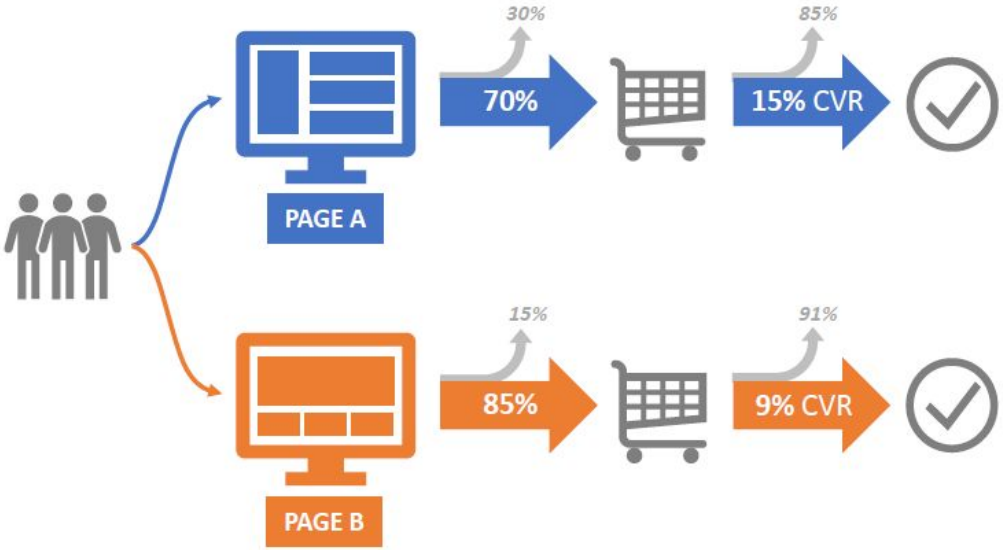
NEXT STEPS ??

Looks like our traffic all comes in through the homepage right now.

- Analyze landing page performance, for the homepage specifically.
- Think about whether or not the homepage is the best initial experience for all customers.

Concept : LANDING PAGE PERFORMANCE & TESTING

Landing page analysis and testing is about understanding the **performance of your key landing pages and then testing to improve your results.**



COMMON USE CASES:

- Identifying your top opportunities for landing pages – high volume pages with higher than expected bounce rates or low conversion rates

- Setting up A/B experiments on your live traffic to see if you can improve your bounce rates and conversion rates

- Analyzing test results and making recommendations on which version of landing pages you should use going forward

CALCULATING BOUNCE RATES

Problem statement

Since all of our traffic is landing on the **homepage** right now, We should check how that landing page is performing.
Pull **bounce rates** for traffic landing on the homepage?
Show Sessions , Bounced Sessions , and % of Sessions which Bounced(aka “Bounce rate”)
Consider analysis before **June 14, 2012.**

SOLUTION QUERY

CLICK HERE

RESULT GRID

Result Grid Filter Rows:

sessions	bounced_sessions	bounce_rate
11044	6536	0.5918

NEXT STEPS??

Almost a 60% bounce rate ! That’s pretty high, especially for paid search, which should be high quality traffic.

- Keep an eye on bounce rates, which represent a major area of improvement.
- Help stakeholder measure and analyze a new page that will improve performance, and analyze the results of an A/B split test against the homepage.

ANALYZING LANDING PAGE TESTS

Problem statement

Based on the bounce rate analysis, we ran a new **custom landing page (/lander 1) in a 50/50 test against the homepage (/home)** for our gsearch nonbrand traffic.
Pull bounce rates for the two groups so we can evaluate the new page. Make sure to just **look at the time period where /lander 1 was getting traffic** , so that it is a fair comparison.
Consider analysis before **July 28, 2012.**

SOLUTION QUERY

CLICK HERE

RESULT GRID

first_created_at	first_pageview_id
2012-06-19 00:35:54	23504

landing_page	total_sessions	bounced_sessions	bounce_rate
/home	2261	1319	0.58337
/lander-1	2315	1232	0.53218

NEXT STEPS??

It looks like the custom lander has a lower bounce rate, which is good.

- Confirm that traffic is all running to the new custom lander after campaign updates
- Keep an eye on bounce rates and help the team look for other areas to test and optimize

LANDING PAGE TREND ANALYSIS

Problem statement

Pull the **volume of paid search nonbrand traffic landing on /home and /lander 1, trended weekly since June 1st**. I want to confirm the traffic is all routed correctly.
Also pull overall paid search bounce rate trended weekly **to make sure the lander change has improved the overall picture**.
Consider analysis before **Aug 31, 2012.**

RESULT GRID

week_start_date	bounce_rate	home_sessions	lander_sessions
2012-06-01	0.60674	178	0
2012-06-03	0.58660	791	0
2012-06-10	0.61758	876	0
2012-06-17	0.55767	492	349
2012-06-24	0.58333	370	386
2012-07-01	0.58131	393	388
2012-07-08	0.56625	390	410
2012-07-15	0.54235	427	423
2012-07-22	0.51321	403	392
2012-07-29	0.49805	34	994
2012-08-05	0.53860	0	1088
2012-08-12	0.51205	0	996
2012-08-19	0.50148	0	1013
2012-08-26	0.53976	0	830

SOLUTION QUERY

CLICK HERE

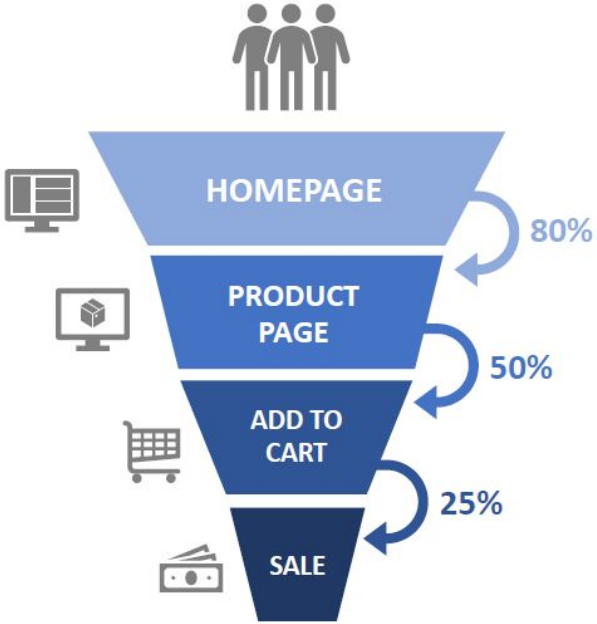
NEXT STEPS??

After switching over to the custom lander , it looks like our overall bounce rate has come down over time.

- Given the successful findings we may move on to the next wave of analysis.

Concept : ANALYZING CONVERSION FUNNELS

Conversion funnel analysis is about **understanding and optimizing each step of your user's experience on their journey toward purchasing your products.**



COMMON USE CASES:

- Identifying the most common paths customers take before purchasing your products.
- Identifying how many of your users continue on to each next step in your conversion flow, and how many users abandon at each step
- Optimizing critical pain points where users are abandoning, so that you can convert more users and sell more products

KEY TABLES

When we perform conversion funnel analysis, we will look at each step in our conversion flow to **see how many customers drop off and how many continue on at each step.**

WEBSITE_PAGEVIEWS

```
SELECT * FROM website_pageviews WHERE website_session_id = 1059
```

website_pageview_id	created_at	website_session_id	pageview_url
2039	2012-03-26 13:51:37	1059	/home
2040	2012-03-26 13:54:27	1059	/products
2041	2012-03-26 13:56:48	1059	/the-original-mr-fuzzy
2042	2012-03-26 14:00:14	1059	/cart
2043	2012-03-26 14:04:06	1059	/shipping
2044	2012-03-26 14:05:47	1059	/billing
2045	2012-03-26 14:13:56	1059	/thank-you-for-your-order

BUILDING CONVERSION FUNNELS

Problem statement

We need to understand where we lose our gsearch visitors between the new /lander 1 page and placing an order.
Build a **full conversion funnel, analyzing how many customers make it to each step**
Start with /lander 1 and build the funnel all the way to the thank you page .
Please use data since **August 5th** and before **September 05, 2012.**

SOLUTION QUERY

CLICK HERE

RESULT GRID

first_pv_id
53550

billing_version_seen	sessions	orders	billing_to_order_rt
/billing	657	300	0.4566
/billing-2	654	410	0.6269

NEXT STEPS??

Looks like we should focus on the lander, Mr. Fuzzy page, and the billing page, which have the lowest click rates.
We need ideas for the billing page that will make customers more comfortable entering their credit card info.

- Help stakeholder analyze a new billing page test.
- Continue to look for opportunities to improve customer conversion rates

ANALYZING CONVERSION FUNNEL TESTS

Problem statement

The team tested an updated billing page based on the funnel analysis.
Take a look and see whether **/billing 2 is doing any better** than the original /billing page?
What **% of sessions on those pages** end up placing an order . Test was ran for all traffic, **not just for our search visitors.**
Consider analysis before **November 10, 2012.**

SOLUTION QUERY

CLICK HERE

RESULT GRID

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

lander_click_rt	products_click_rt	mrfuzzy_click_rt	cart_click_rt	shipping_click_rt	billing_click_rt
0.4709	0.7405	0.4352	0.6657	0.7930	0.4361

NEXT STEPS??

Looks like the new version of the billing page is doing a much better job converting customers.

- Get Engineering team to roll out the new version to 100% of traffic, use the data to confirm they have done so correctly
- Monitor overall sales performance to see the impact this change produces

Challenge 3

CHANNEL MANAGEMENT ANALYSIS

Concept : CHANNEL PORTFOLIO OPTIMIZATION

Analyzing a portfolio of marketing channels is about **bidding efficiently and using data to maximize the effectiveness of your marketing budget**.

COMMON USE CASES:

- Understanding which marketing channels are driving the most sessions and orders through the website

.

- Understanding differences in user characteristics and conversion performance across marketing channels

- Optimizing bids and allocating marketing spend across a multi-channel portfolio to achieve maximum performance

WEBSITE_SESSIONS: SESSION DATA

In addition to measuring and analyzing where traffic is coming from, we can use a business' session level data to understand user characteristics and behaviors,

For example, we can see if the user is new to our site or if they are a repeat visitor, and which type of device they used during the session (mobile or desktop)

WEBSITE_SESSIONS

```
SELECT DISTINCT
        is_repeat_session,
        device_type
FROM website_sessions
```

is_repeat_session	device_type
0	mobile
0	desktop
1	desktop
1	mobile

ANALYZING CHANNEL PORTFOLIOS

Problem statement

With gsearch doing well and the site performing better, **a second paid search channel (bsearch), was launched around August 22**.
Pull **weekly trended session volume** since then and **compare to gsearch nonbrand** so we can get a sense for how important this will be for the business.
Consider analysis before **November 29, 2012.**

RESULT GRID

week_start_date	gsearch_sessions	bsearch_sessions
2012-08-22	590	197
2012-08-26	1056	343
2012-09-02	925	290
2012-09-09	951	329
2012-09-16	1151	365
2012-09-23	1050	321
2012-09-30	999	316
2012-10-07	1002	330
2012-10-14	1257	420
2012-10-21	1302	431
2012-10-28	1211	384
2012-11-04	1350	429
2012-11-11	1246	438
2012-11-18	3508	1093
2012-11-25	2286	774

SOLUTION QUERY

CLICK HERE

NEXT STEPS??

It looks like bsearch tends to get roughly a third the traffic of gsearch. This is big enough that we should really get to know the channel better.

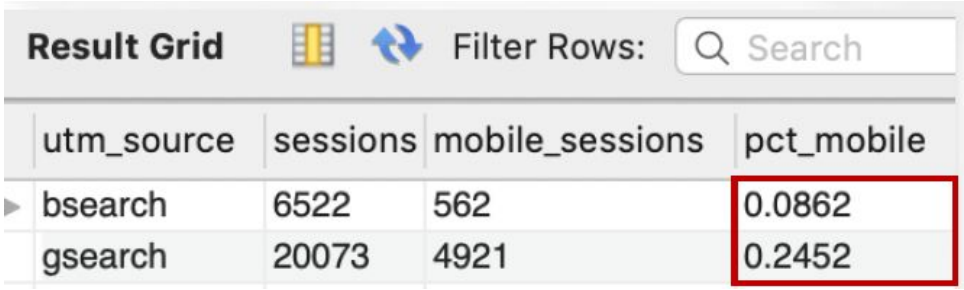
- Keep an eye out for channel specific requests from stakeholders.
- Proactively slice and dice the data to see what you can find.

COMPARING CHANNEL CHARACTERISTICS

Problem statement

We need to learn more about the **bsearch nonbrand** campaign.
Pull the **percentage of traffic coming on Mobile** , and **compare that to gsearch**.
Aggregate data since August 22nd is great, no need to show trending at this point.
Consider analysis before **November 30, 2012.**

RESULT GRID


Result Grid Filter Rows: Search

utm_source	sessions	mobile_sessions	pct_mobile
bsearch	6522	562	0.0862
gsearch	20073	4921	0.2452

SOLUTION QUERY

CLICK HERE

NEXT STEPS??

Now that we know these channels are pretty different from a device standpoint, we would need further digging so that we can get our bids right.

- Anticipate what stakeholders will be asking for next. As an analyst, it's better we understand the types of data points P&L owners need.

CROSS CHANNEL BID OPTIMIZATION

Problem statement

Is it required for bsearch nonbrand to have the same bids as gsearch?
Pull **nonbrand conversion rates from session to order for gsearch and bsearch, and slice the data by device type**.
Analyze data from **August 22 to September 18**.
We ran a special pre holiday campaign for gsearch starting on **September 19th** , so the data after that isn't fair game.

SOLUTION QUERY

CLICK HERE

RESULT GRID

Result Grid Filter Rows: Search

device_type	utm_source	sessions	orders	conv_rate
desktop	bsearch	1162	44	0.0379
desktop	gsearch	3011	136	0.0452
mobile	bsearch	130	1	0.0077
mobile	gsearch	1015	13	0.0128

NEXT STEPS??

As suspected, the channels don't perform identically, so we should differentiate our bids in order to optimize our overall paid marketing budget.

- Understand the impact of the bid changes based on this analysis
- Keep your eye out for new bid optimization opportunities.

CHANNEL PORTFOLIO TRENDS

Problem statement

Based on your last analysis, we bid down bsearch nonbrand on **December 2nd**
Pull **weekly session volume for gsearch and bsearch nonbrand, broken down by device, since November 4th**
Include a **comparison metric to show bsearch as a percent of gsearch** for each device.
Consider analysis before **December 22, 2012.**

SOLUTION QUERY

CLICK HERE

RESULT GRID

Result Grid Filter Rows: Search Export:

week_start_date	g_dtop_sessions	b__dtop_sessions	b_pct_of_g_dtop	g_mob_sessions	b_mob_sessions	b_pct_of_g_mob
2012-11-04	1027	400	0.3895	323	29	0.0898
2012-11-11	956	401	0.4195	290	37	0.1276
2012-11-18	2655	1008	0.3797	853	85	0.0996
2012-11-25	2058	843	0.4096	692	62	0.0896
2012-12-02	1326	517	0.3899	396	31	0.0783
2012-12-09	1277	293	0.2294	424	46	0.1085
2012-12-16	1270	348	0.2740	376	41	0.1090

NEXT STEPS??

Looks like bsearch traffic dropped off a bit after the bid down. Seems like gsearch was down too after Black Friday and Cyber Monday, but bsearch dropped even more. This is okay given the low conversion rate.

- Try to fully grasp the results of this data
- Think about which of these metrics best controls for the seasonality and isolates the impact of the bsearch bid changes.

Concept : ANALYZING DIRECT TRAFFIC

Analyzing your branded or direct traffic is about **keeping a pulse on how well your brand is doing with consumers, and how well your brand drives business.**

COMMON USE CASES:

• Identifying how much revenue you are generating from direct traffic – this is high margin revenue without a direct cost of customer acquisition

.

• Understanding whether or not your paid traffic is generating a “halo” effect, and promoting additional direct traffic

• Assessing the impact of various initiatives on how many customers seek out your business

ANALYZING FREE CHANNELS

Problem statement

A potential investor is asking if we're building any momentum with our brand or if we'll need to keep relying on paid traffic.
Pull **organic search, direct type in, and paid brand search sessions by month**, and show those sessions as a % of paid search nonbrand.
Consider analysis before **December 23, 2012.**

RESULT GRID

Result Grid Filter Rows: Search Export:

yr	mo	nonbrand	brand	brand_pct_of_nonbrand	direct	direct_pct_of_nonbrand	organic	organic_pct_of_nonbrand
2012	3	1852	10	0.0054	9	0.0049	8	0.0043
2012	4	3509	76	0.0217	71	0.0202	78	0.0222
2012	5	3295	140	0.0425	151	0.0458	150	0.0455
2012	6	3439	164	0.0477	170	0.0494	190	0.0552
2012	7	3660	195	0.0533	187	0.0511	207	0.0566
2012	8	5318	264	0.0496	250	0.0470	265	0.0498
2012	9	5591	339	0.0606	285	0.0510	331	0.0592
2012	10	6883	432	0.0628	440	0.0639	428	0.0622
2012	11	12260	556	0.0454	571	0.0466	624	0.0509
2012	12	6643	464	0.0698	482	0.0726	492	0.0741

SOLUTION QUERY

CLICK HERE

NEXT STEPS??

Looks like not only are our brand, direct, and organic volumes growing, but they are growing as a percentage of our paid traffic volume .

- Spend some time thinking about what other data points one could pull that would make the company look good

Challenge 4

BUSINESS PATTERNS & SEASONALITY

Concept : ANALYZING PATTERNS

Analyzing business patterns is about **generating insights to help you maximize efficiency and anticipate future trends.**

COMMON USE CASES:

- 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 slowdowns in demand

ANALYZING SEASONALITY

Problem statement

2012 was a great year for the company. 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.

Pull session volume and order volume.

SOLUTION QUERY

CLICK HERE

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

RESULT GRID

yr	mo	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

NEXT STEPS??

Looks like the company grew fairly steadily all year, and saw significant volume around the holiday months. We'll want to keep this in mind in 2013 as we think about customer support and inventory management.

ANALYZING BUSINESS PATTERNS

Problem statement

We're considering adding live chat support to the website to improve our customer experience.
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, 2013**.

RESULT GRID

hr	mon	tue	wed	thu	fri	sat	sun
0	8.7	7.7	6.3	7.4	6.8	5.0	5.0
1	6.6	6.7	5.3	4.9	7.1	5.0	3.0
2	6.1	4.4	4.4	6.1	4.6	3.7	3.0
3	5.7	4.0	4.7	4.6	3.6	3.9	3.4
4	5.9	6.3	6.0	4.0	6.1	2.8	2.4
5	5.0	5.4	5.1	5.4	4.6	4.3	3.9
6	5.4	5.6	4.8	6.0	6.8	4.0	2.6
7	7.3	7.8	7.4	10.6	7.0	5.7	4.8
8	12.3	12.2	13.0	16.5	10.5	4.3	4.1
9	17.6	15.7	19.6	19.3	17.5	7.6	6.0
10	18.4	17.7	21.0	18.4	19.0	8.3	6.3
11	18.0	19.1	24.9	21.6	20.9	7.2	7.7
12	21.1	23.3	22.8	24.1	19.0	8.6	6.1
13	17.8	23.0	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	15.0	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	10.0	9.8	12.1	6.0	5.7	10.2
23	8.8	8.6	9.6	10.6	7.6	5.3	8.3

SOLUTION QUERY

CLICK HERE

NEXT STEPS??

It looks like ~10 sessions per hour per employee staffed is about right. We can plan on one support staff around the clock and then we should double up to two staff members from 8am to 5pm Monday through Friday.

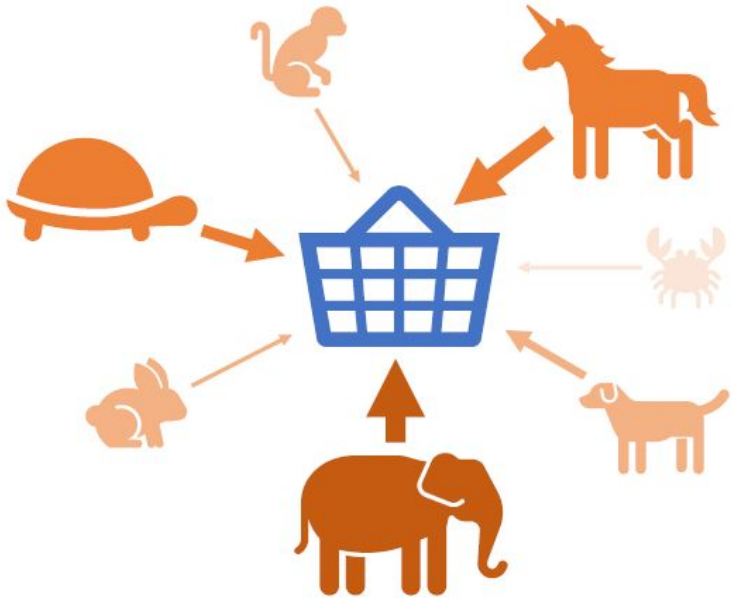
- Think proactively about how this type of trended analysis could be applied to other areas of the business

Challenge 5

PRODUCT ANALYSIS

Concept : PRODUCT SALES ANALYSIS

Analyzing product sales helps you **understand how each product contributes to your business, and how product launches impact the overall portfolio**.



COMMON USE CASES:

- Analyzing sales and revenue by product
- Monitoring the impact of adding a new product to your product portfolio
- Watching product sales trends to understand the overall health of your business

KEY TERMS

ORDERS	Number of orders placed by customers	COUNT(order_id)
REVENUE	Money business brings in from orders	SUM(price_usd)
MARGIN	Revenue less the cost of good sold	SUM(price_usd cogs_usd)
AOV	Average revenue generated per order	AVG(price_usd)

```
SELECT
        COUNT(order_id ) AS orders
        SUM(price_usd ) AS revenue
        SUM(price_usd cogs_usd ) AS margin
        AVG(price_usd ) AS average_order_value
FROM orders
WHERE order_id BETWEEN 100 AND 200
```

orders	revenue	margin	average_order_value
101	5048.99	3080.50	49.990000

PRODUCT LEVEL SALES ANALYSIS

Problem statement

We're about to launch a new product, and we would like to do a deep dive on our current flagship product.
Pull **monthly trends to date for number of sales , total revenue , and total margin** generated for the Business?
Consider analysis before **January 04, 2013.**

RESULT GRID

yr	mo	number_of_sales	total_revenue	total_margin
2012	3	60	2999.40	1830.00
2012	4	99	4949.01	3019.50
2012	5	108	5398.92	3294.00
2012	6	140	6998.60	4270.00
2012	7	169	8448.31	5154.50
2012	8	228	11397.72	6954.00
2012	9	287	14347.13	8753.50
2012	10	371	18546.29	11315.50
2012	11	618	30893.82	18849.00
2012	12	506	25294.94	15433.00
2013	1	42	2099.58	1281.00

SOLUTION QUERY

CLICK HERE

NEXT STEPS??

This will serve as great baseline data so that we can see how our revenue and margin evolve as we roll out the new product.

- Keep an eye on performance as the new product launches

PRODUCT LAUNCH SALES ANALYSIS

Problem statement

We launched our second product back on January 6th. Pull together some trended analysis.
Depict **monthly order volume , overall conversion rates , revenue per session , and a breakdown of sales by product** , all for the time period since **April 1, 2013**
Consider analysis before **April 05, 2013.**

RESULT GRID

yr	mo	orders	conv_rate	revenue_per_session	product_one_orders	product_two_orders
2012	4	99	0.0265	1.325391	99	0
2012	5	108	0.0289	1.445107	108	0
2012	6	140	0.0353	1.765985	140	0
2012	7	169	0.0398	1.988305	169	0
2012	8	228	0.0374	1.869398	228	0
2012	9	287	0.0438	2.191740	287	0
2012	10	371	0.0453	2.266441	371	0
2012	11	618	0.0441	2.204969	618	0
2012	12	506	0.0502	2.511412	506	0
2013	1	391	0.0611	3.127025	344	47
2013	2	497	0.0693	3.692108	335	162
2013	3	385	0.0615	3.176269	320	65
2013	4	96	0.0794	4.085227	82	14

SOLUTION QUERY

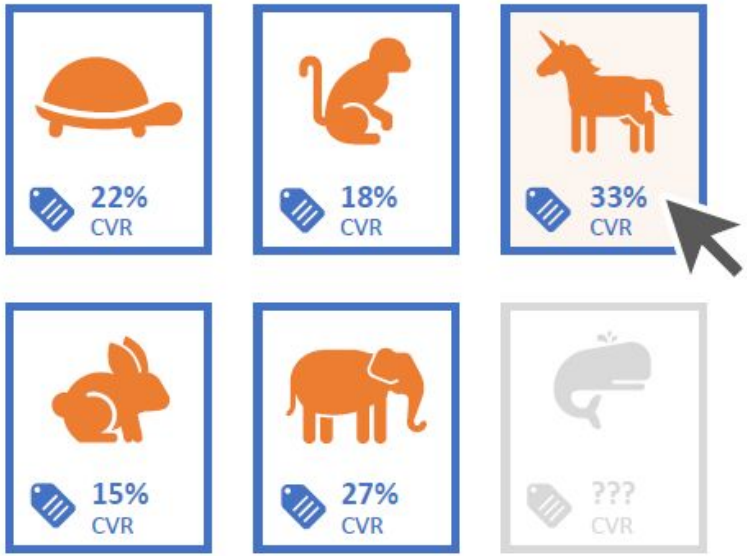
CLICK HERE

NEXT STEPS??

This confirms that our conversion rate and revenue per session are improving over time.
But it is hard to understand if the growth since January is due to our new product launch or just a continuation of our overall business improvements.

Concept : PRODUCT LEVEL WEBSITE ANALYSIS

Product-focused website analysis is about **learning how customers interact with each of your products, and how well each product converts customers.**



COMMON USE CASES:

- Understanding which of your products generate the most interest on multi-product showcase pages
- Analyzing the impact on website conversion rates when you add a new product
- Building product-specific conversion funnels to understand whether certain products convert better than others

PRODUCT PATHING ANALYSIS

Problem statement

Now that we have a new product, let's look at **sessions which hit the /products page and see where they went next** .
Pull clickthrough rates from /products **since the new product launch on January 6 th 2013** , by product, and compare to the **3 months** leading up to launch as a baseline.
Consider analysis before **April 06, 2013.**

RESULT GRID

Result Grid | Filter Rows: Search | Export:

time_period	sessions	w_next_pg	pct_w_next_pg	to_mrfuzzy	pct_to_mrfuzzy	to_lovebear	pct_to_lovebear
A. Pre_Product_2	15696	11347	0.7229	11347	0.7229	0	0.0000
B. Post_Product_2	10709	8200	0.7657	6654	0.6213	1546	0.1444

SOLUTION QUERY

CLICK HERE

NEXT STEPS??

Looks like the percent 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, so it seems to be generating additional product interest overall. As a follow up, we should probably look at the conversion funnels for each product individually.

PRODUCT CONVERSION FUNNELS

Problem statement

Take a look at our two products since January 6th and **analyze the conversion funnels from each product page to conversion.**
Produce a **comparison between the two conversion funnels, for all website traffic.**
Consider analysis before **April 10, 2013.**

RESULT GRID

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	product_page_click_rt	cart_click_rt	shipping_click_rt	billing_click_rt
lovebear	0.5485	0.6876	0.8093	0.6168
mrfuzzy	0.4349	0.6860	0.8205	0.6363

SOLUTION QUERY

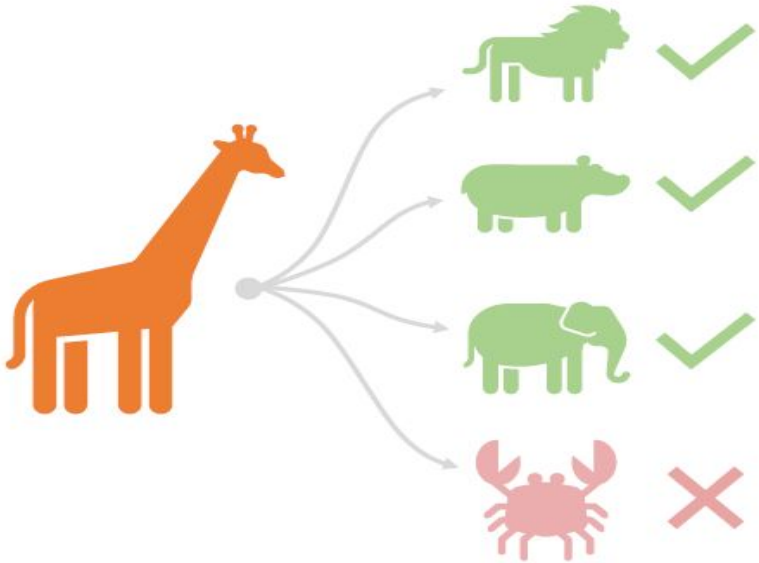
CLICK HERE

NEXT STEPS??

We had found that adding a second product increased overall CTR from the /products page, and this analysis shows that the Love Bear has a better click rate to the /cart page and comparable rates throughout the rest of the funnel. Seems like the second product was a great addition for our business. I wonder if we should add a third…

Concept : CROSS-SELLING PRODUCTS

Cross-sell analysis is about **understanding which products users are most likely to purchase together, and offering smart product recommendations.**



COMMON USE CASES:

- Understanding which products are often purchased together
- Testing and optimizing the way you cross-sell products on your website
- Understanding the conversion rate impact and the overall revenue impact of trying to crosssell additional products

CROSS SELL ANALYSIS

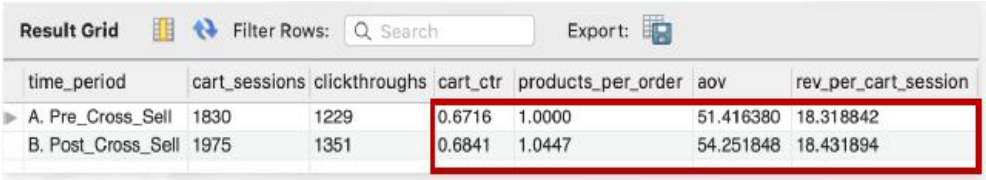
Problem statement

On September 25th we started giving customers the **option to add a 2nd product while on the /cart page**.
Compare the **month before vs the month** after the change. Depict **CTR from the /cart page, Avg Products per Order , AOV , and overall revenue** per /cart page view

SOLUTION QUERY

CLICK HERE

RESULT GRID


Result Grid Filter Rows: Search Export:

time_period	cart_sessions	clickthroughs	cart_ctr	products_per_order	aov	rev_per_cart_session
A. Pre_Cross_Sell	1830	1229	0.6716	1.0000	51.416380	18.318842
B. Post_Cross_Sell	1975	1351	0.6841	1.0447	54.251848	18.431894

NEXT STEPS??

It looks like the CTR from the /cart page didn't go down, and that our products per order, AOV, and revenue per /cart session are all up slightly since the cross sell feature was added. Doesn't look like a game changer, but the trend looks positive.

- Ask whether you think adding a second product and attempting cross sell was a good idea for the business. Do you think more products should be added?

PORTFOLIO EXPANSION ANALYSIS

Problem statement

On December 12th 2013, we launched a third product targeting the birthday gift market (Birthday Bear).
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.
Consider analysis before **January 12, 2014.**

SOLUTION QUERY

CLICK HERE

RESULT GRID

Result Grid | Filter Rows: Search | Export:

time_period	conv_rate	aov	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

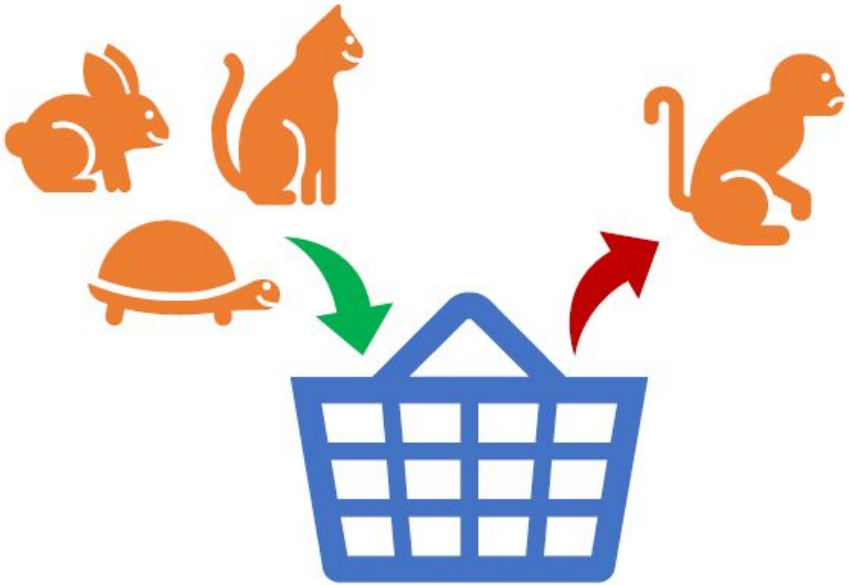
NEXT STEPS??

Looks like all of our critical metrics have improved since we launched the third product.
Discuss with stakeholder about increasing our ad spend now that we're driving more revenue per session, and we may also consider adding a fourth product.

- Keep an eye on paid traffic volume if ad spend increases.

Concept : PRODUCT REFUND ANALYSIS

Analyzing product refund rates is about **controlling for quality and understanding where you might have problems to address.**



COMMON USE CASES:

- Monitoring products from different suppliers
- Understanding refund rates for products at different price points
- Taking product refund rates and the associated costs into account when assessing the overall performance of your business

PRODUCT REFUNDS

Keeping a eye on on refund rates is a great way to **analyze the relative quality of your products, track customer satisfaction , and keep a pulse on overall business health.**

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

PRODUCT REFUND RATES

Problem statement

Our 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.**

Pull **monthly product refund rates, by product, and confirm our quality issues are now fixed.**

Consider analysis before **October 15, 2014.**

RESULT GRID

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	0.0167	0	NULL	0	NULL	0	NULL
2012	4	99	0.0505	0	NULL	0	NULL	0	NULL
2012	5	108	0.0370	0	NULL	0	NULL	0	NULL
2012	6	140	0.0571	0	NULL	0	NULL	0	NULL
2012	7	169	0.0828	0	NULL	0	NULL	0	NULL
2012	8	228	0.0746	0	NULL	0	NULL	0	NULL
2012	9	287	0.0906	0	NULL	0	NULL	0	NULL
2012	10	371	0.0728	0	NULL	0	NULL	0	NULL
2012	11	618	0.0744	0	NULL	0	NULL	0	NULL
2012	12	506	0.0593	0	NULL	0	NULL	0	NULL
2013	1	343	0.0496	47	0.0213	0	NULL	0	NULL
2013	2	336	0.0714	162	0.0123	0	NULL	0	NULL
2013	3	320	0.0563	65	0.0462	0	NULL	0	NULL
2013	4	459	0.0414	94	0.0106	0	NULL	0	NULL
2013	5	489	0.0634	82	0.0244	0	NULL	0	NULL
2013	6	503	0.0775	90	0.0556	0	NULL	0	NULL
2013	7	509	0.0727	95	0.0316	0	NULL	0	NULL
2013	8	510	0.0549	98	0.0102	0	NULL	0	NULL
2013	9	537	0.0428	96	0.0102	0	NULL	0	NULL
2013	10	603	0.0282	135	0.0148	0	NULL	0	NULL
2013	11	724	0.0345	174	0.0230	0	NULL	0	NULL
2013	12	818	0.0232	183	0.0219	139	0.0719	0	NULL
2014	1	728	0.0426	183	0.0219	200	0.0650	0	NULL
2014	2	584	0.0394	351	0.0171	211	0.0664	202	0.0099
2014	3	785	0.0306	193	0.0155	244	0.0697	205	0.0049
2014	4	917	0.0349	214	0.0187	267	0.0674	259	0.0154
2014	5	1030	0.0291	246	0.0163	299	0.0569	298	0.0067
2014	6	893	0.0571	245	0.0367	288	0.0556	249	0.0241
2014	7	961	0.0437	244	0.0369	276	0.0399	264	0.0152
2014	8	958	0.1378	237	0.0169	294	0.0680	303	0.0066
2014	9	1056	0.1326	251	0.0319	317	0.0662	327	0.0122
2014	10	513	0.0273	135	0.0074	165	0.0485	155	0.0323

SOLUTION QUERY

CLICK HERE

NEXT STEPS??

Looks like the refund rates for old supplier did go down after the initial improvements in September 2013, but refund rates were terrible in August and September, as expected (13 14%). Seems like the new supplier is doing much better so far, and the other products look okay too.

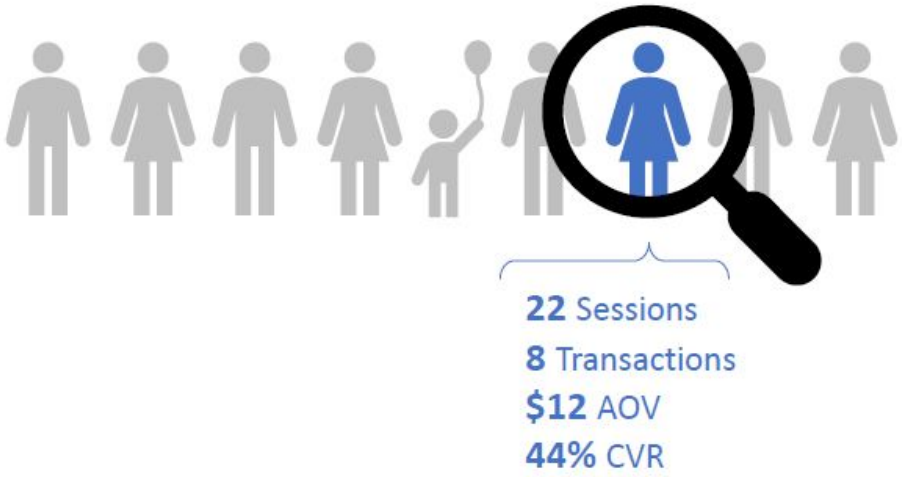
- Keep an eye on product refund rates in case there are more quality issues in the future

Challenge 6

USER ANALYSIS

Concept : ANALYZE REPEAT BEHAVIOR

Analyzing repeat visits helps you **understand user behavior and identify some of your most valuable customers.**



COMMON USE CASES:

- 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

TRACKING REPEAT CUSTOMERS

Businesses track customer behavior across multiple sessions using **browser cookies.**

Cookies have unique ID values associated with them, which allows us to recognize a customer when they come back and track their behavior over time.

website_session_id	created_at	user_id	is_repeat_session	utm_source	utm_campaign	utm_content	device_type	http_referer
237966	2014-04-28 23:02:55	204524	0	gsearch	nonbrand	g_ad_1	desktop	https://www.gsearch.com
319940	2014-09-11 12:38:39	271374	0	bsearch	nonbrand	b_ad_1	desktop	https://www.bsearch.com
326645	2014-09-22 13:50:39	271374	1	gsearch	brand	g_ad_2	desktop	https://www.gsearch.com
325116	2014-09-19 11:42:44	275579	0	socialbook	desktop_targeted	social_ad_2	desktop	https://www.socialbook.com
349691	2014-10-26 19:24:17	275579	1	gsearch	brand	g_ad_2	desktop	https://www.gsearch.com
357769	2014-11-06 22:10:25	275579	1	NULL	NULL	NULL	mobile	https://www.gsearch.com
367395	2014-11-19 14:56:33	275579	1	NULL	NULL	NULL	mobile	NULL

IDENTIFYING REPEAT VISITORS

Problem statement

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.
Pull **data on how many of our website visitors come back for another session. 2014 to date is good**.
Consider analysis before **November 01, 2014.**

RESULT GRID

repeat_sessions	users
0	126813
1	14086
2	315
3	4686

SOLUTION QUERY

CLICK HERE

NEXT STEPS??

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

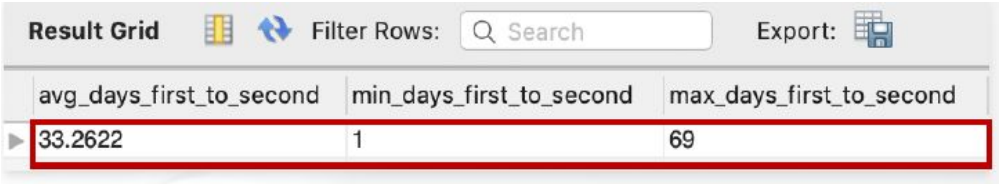
- Dig further into repeat customer behavior.

ANALYZING REPEAT BEHAVIOR

Problem statement

Now we need to better understand the behavior of these repeat customers.
Pull the **minimum, maximum, and average time between the first and second session** for customers who do come back. Again, analyzing 2014 to date is the right time period.
Consider analysis before **November 03, 2014.**

RESULT GRID


Result Grid Filter Rows: Search Export:

avg_days_first_to_second	min_days_first_to_second	max_days_first_to_second
33.2622	1	69

SOLUTION QUERY

CLICK HERE

NEXT STEPS??

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.

NEW VS REPEAT CHANNEL PATTERNS

Problem statement

Let's do a bit more digging into our repeat customers. We need to **understand the channels they come back through**, 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, 2014 to date is great. Consider analysis before **November 05, 2014.**

RESULT GRID

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

SOLUTION QUERY

CLICK HERE

NEXT STEPS??

- It looks like when customers come back for repeat visits, they come mainly through organic search, direct type in, and paid brand.
- Only about 1/3 come through a paid channel, and brand clicks are cheaper than nonbrand. So all in all, we're not paying very much for these subsequent visits.
- But we need to see whether these convert to orders.

NEW VS REPEAT PERFORMANCE

Problem statement

Perform 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.
Consider analysis before **November 08, 2014.**

RESULT GRID

is_repeat_session	sessions	conv_rate	rev_per_session
0	149787	0.0680	4.343754
1	33577	0.0811	5.168828

SOLUTION QUERY

CLICK HERE

NEXT STEPS??

Looks like repeat sessions are more likely to convert , and produce more revenue per session.
Since we aren't paying much for repeat sessions, we should probably take them into account when bidding on paid traffic.

- We could try to incorporate the value of these repeat sessions into figuring out the value of your paid clicks.

There are further more ways of interpreting the database and measuring the performance of a company. Looking forward to add more usecases within the project in the future.

Project master github link -

Special thanks to my instructors Mr S Saurabh and Mr John Pauler.