8 Week SQL Challenge

Start your SQL learning journey today!

Data With Danny   Case Studies   Getting Started   Resources   About
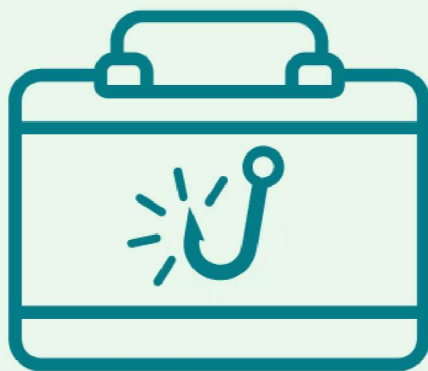
Case Study #6 - Clique Bait

Danny Ma · June 29, 2021

8WEEKSQLCHALLENGE.COM
CASE STUDY #6

CliqueBait

ATTENTION CAPTURING

DATAWITHDANNY.COM

Clique Bait is not like your regular online seafood store - the founder and CEO Danny, was also a part of a digital data analytics team and wanted to expand his knowledge into the seafood industry!

In this case study - you are required to support Danny's vision and analyse his dataset and come up with creative solutions to calculate funnel fallout rates for the Clique Bait online store.

Available Data

For this case study there is a total of 5 datasets which you will need to combine to solve all of the questions.

Users

Customers who visit the Clique Bait website are tagged via their `cookie_id`.

user_id	cookie_id	start_date
397	3759ff	2020-03-30 00:00:00
215	863329	2020-01-26 00:00:00
191	eefca9	2020-03-15 00:00:00
89	764796	2020-01-07 00:00:00
127	17ccc5	2020-01-22 00:00:00
81	b0b666	2020-03-01 00:00:00
260	a4f236	2020-01-08 00:00:00
203	d1182f	2020-04-18 00:00:00
23	12dbc8	2020-01-18 00:00:00

user_id	cookie_id	start_date
375	f61d69	2020-01-03 00:00:00

Events

Customer visits are logged in this `events` table at a `cookie_id` level and the `event_type` and `page_id` values can be used to join onto relevant satellite tables to obtain further information about each event.

The sequence_number is used to order the events within each visit.

visit_id	cookie_id	page_id	event_type	sequence_number	event_time
719fd3	3d83d3	5	1	4	2020-03-02 00:
fb1eb1	c5ff25	5	2	8	2020-01-22 07:
23fe81	1e8c2d	10	1	9	2020-03-21 13:
ad91aa	648115	6	1	3	2020-04-27 16:
5576d7	ac418c	6	1	4	2020-01-18 04:
48308b	c686c1	8	1	5	2020-01-29 06:
46b17d	78f9b3	7	1	12	2020-02-16 09:
9fd196	ccf057	4	1	5	2020-02-14 08:
edf853	f85454	1	1	1	2020-02-22 12:
3c6716	02e74f	3	2	5	2020-01-31 17:

Event Identifier

The `event_identifier` table shows the types of events which are captured by Clique Bait's digital data systems.

event_type	event_name
1	Page View
2	Add to Cart
3	Purchase
4	Ad Impression
5	Ad Click

Campaign Identifier

This table shows information for the 3 campaigns that Clique Bait has ran on their website so far in 2020.

campaign_id	products	campaign_name	start_date
1	1-3	BOGOF - Fishing For Compliments	2020-01-01 00:00:00
2	4-5	25% Off - Living The Lux Life	2020-01-15 00:00:00
3	6-8	Half Off - Treat Your Shellf(ish)	2020-02-01 00:00:00

Page Hierarchy

This table lists all of the pages on the Clique Bait website which are tagged and have data passing through from user interaction events.

page_id	page_name	product_category	product_id
1	Home Page	null	null
2	All Products	null	null
3	Salmon	Fish	1
4	Kingfish	Fish	2

page_id	page_name	product_category	product_id
5	Tuna	Fish	3
6	Russian Caviar	Luxury	4
7	Black Truffle	Luxury	5
8	Abalone	Shellfish	6
9	Lobster	Shellfish	7
10	Crab	Shellfish	8
11	Oyster	Shellfish	9
12	Checkout	null	null
13	Confirmation	null	null

Interactive SQL Instance

You can use the provided DB Fiddle instance to easily access these example datasets - this interactive session has everything you need to start solving these questions using SQL.

Click here to open a fully functional SQL editor where you can write your own queries to analyse the data.

In past case studies - you would have seen an embedded DB Fiddle instance, but this time it seems like this week's case study datasets might have been too big and broke the API for the embedded iframe code - whoops!!!

You can feel free to choose any SQL dialect you'd like to use, the existing Fiddle is using PostgreSQL 13 as default.

Serious SQL students will have access to the same relevant schema SQL and example solutions which they can use with their Docker setup from within the course player!

Case Study Questions

1. Enterprise Relationship Diagram

Using the following DDL schema details to create an ERD for all the Clique Bait datasets.

`Click here` to access the DB Diagram tool to create the ERD.

```sql
CREATE TABLE clique_bait.event_identifier (
  "event_type" INTEGER,
  "event_name" VARCHAR(13)
);


CREATE TABLE clique_bait.campaign_identifier (
  "campaign_id" INTEGER,
  "products" VARCHAR(3),
  "campaign_name" VARCHAR(33),
  "start_date" TIMESTAMP,
  "end_date" TIMESTAMP
);


CREATE TABLE clique_bait.page_hierarchy (
  "page_id" INTEGER,
  "page_name" VARCHAR(14),
  "product_category" VARCHAR(9),
  "product_id" INTEGER
);


CREATE TABLE clique_bait.users (
  "user_id" INTEGER,
  "cookie_id" VARCHAR(6),
  "start_date" TIMESTAMP
);


CREATE TABLE clique_bait.events (
  "visit_id" VARCHAR(6),
```

```
  "cookie_id" VARCHAR(6),

  "page_id" INTEGER,

  "event_type" INTEGER,

  "sequence_number" INTEGER,

  "event_time" TIMESTAMP

);
```

2. Digital Analysis

Using the available datasets - answer the following questions using a single query for each one:

1. How many users are there?
2. How many cookies does each user have on average?
3. What is the unique number of visits by all users per month?
4. What is the number of events for each event type?
5. What is the percentage of visits which have a purchase event?
6. What is the percentage of visits which view the checkout page but do not have a purchase event?
7. What are the top 3 pages by number of views?
8. What is the number of views and cart adds for each product category?
9. What are the top 3 products by purchases?

3. Product Funnel Analysis

Using a single SQL query - create a new output table which has the following details:

- How many times was each product viewed?
- How many times was each product added to cart?
- How many times was each product added to a cart but not purchased (abandoned)?
- How many times was each product purchased?

Additionally, create another table which further aggregates the data for the above points but this time for each product category instead of individual products.

Use your 2 new output tables - answer the following questions:

1. Which product had the most views, cart adds and purchases?
2. Which product was most likely to be abandoned?
3. Which product had the highest view to purchase percentage?
4. What is the average conversion rate from view to cart add?
5. What is the average conversion rate from cart add to purchase?

3. Campaigns Analysis

Generate a table that has 1 single row for every unique `visit_id` record and has the following columns:

- `user_id`
- `visit_id`
- `visit_start_time`: the earliest `event_time` for each visit
- `page_views`: count of page views for each visit
- `cart_adds`: count of product cart add events for each visit
- `purchase`: 1/0 flag if a purchase event exists for each visit
- `campaign_name`: map the visit to a campaign if the `visit_start_time` falls between the `start_date` and `end_date`
- `impression`: count of ad impressions for each visit
- `click`: count of ad clicks for each visit
- **(Optional column)** `cart_products`: a comma separated text value with products added to the cart sorted by the order they were added to the cart (hint: use the `sequence_number`)

Use the subsequent dataset to generate at least 5 insights for the Clique Bait team - bonus: prepare a single A4 infographic that the team can use for their management reporting sessions, be sure to emphasise the most important points from your findings.

Some ideas you might want to investigate further include:

- Identifying users who have received impressions during each campaign period and comparing each metric with other users who did not have an impression event
- Does clicking on an impression lead to higher purchase rates?
- What is the uplift in purchase rate when comparing users who click on a campaign impression versus users who do not receive an impression? What if we compare them with users who just an impression but do not click?
- What metrics can you use to quantify the success or failure of each campaign compared to eachother?

Conclusion

This case study is based off my many years working with Digital datasets in consumer banking and retail supermarkets - all of the datasets are designed based off real datasets I've come across in challenging problem solving scenarios and the questions reflect similar problems which I worked on.

Campaign analysis is almost everywhere in the data world, especially in marketing, digital, UX and retail industries - and being able to analyse views, clicks and other digital behaviour is a critical skill to have in your toolbelt as a data professional!

Ready for the next 8 Week SQL challenge case study? Click on the banner below to get started with case study #7!

Official Solutions

If you'd like to see the official code solutions and explanations for this case study and a whole lot more, please consider joining me for the Serious SQL course - you'll get access to all course materials and I'm on hand to answer all of your additional SQL questions directly!

Serious SQL is priced at $49USD and $29 for students and includes access to all written course content, community events as well as live and recorded SQL training videos!

Please send an email to support@datawithdanny.com from your educational email or include your enrolment details or student identification for a speedy response!

Community Solutions

This section will be updated in the future with any community member solutions with a link to their respective GitHub repos!

Final Thoughts

The 8 Week SQL Challenge is proudly brought to you by me - Danny Ma and the Data With Danny virtual data apprenticeship program.

Students or anyone undertaking further studies are eligible for a $20USD student discount off the price of Serious SQL please send an email to support@datawithdanny.com from your education email or include information about your enrolment for a fast response!

We have a large student community active on the official DWD Discord server with regular live events, trainings and workshops available to all Data With Danny students, plus early discounted access to all future paid courses.

There are also opportunities for 1:1 mentoring, resume reviews, interview training and more from myself or others in the DWD Mentor Team.

From your friendly data mentor, Danny :)

All 8 Week SQL Challenge Case Studies

All of the 8 Week SQL Challenge case studies can be found below:

- Case Study #1 - Danny's Diner
- Case Study #2 - Pizza Runner
- Case Study #3 - Foodie-Fi
- Case Study #4 - Data Bank

Share: Twitter, Facebook