



# 8 Week SQL Challenge

Start your SQL learning journey today!



Data With Danny Case Studies Getting Started Resources About

## Case Study #8 - Fresh Segments

Danny Ma · July 9, 2021

8WEEKSQLCHALLENGE.COM

# CASE STUDY #8



**FRESH SEGMENTS**  
**EXTRACT MAXIMUM VALUE**

DATAWITHDANNY.COM

Danny created Fresh Segments, a digital marketing agency that helps other businesses analyse trends in online ad click behaviour for their unique customer base.

Clients share their customer lists with the Fresh Segments team who then aggregate interest metrics and generate a single dataset worth of metrics for further analysis.

In particular - the composition and rankings for different interests are provided for each client showing the proportion of their customer list who interacted with online assets related to each interest for each month.

Danny has asked for your assistance to analyse aggregated metrics for an example client and provide some high level insights about the customer list and their interests.

## Available Data

For this case study there is a total of 2 datasets which you will need to use to solve the questions.

### Interest Metrics

This table contains information about aggregated interest metrics for a specific major client of Fresh Segments which makes up a large proportion of their customer base.

Each record in this table represents the performance of a specific **interest\_id** based on the client's customer base interest measured through clicks and interactions with specific targeted advertising content.

<b>_month</b>	<b>_year</b>	<b>month_year</b>	<b>interest_id</b>	<b>composition</b>	<b>index_value</b>	<b>ranki</b>
7	2018	07-2018	32486	11.89	6.19	1
7	2018	07-2018	6106	9.93	5.31	2
7	2018	07-2018	18923	10.85	5.29	3

<b>_month</b>	<b>_year</b>	<b>month_year</b>	<b>interest_id</b>	<b>composition</b>	<b>index_value</b>	<b>ranki</b>
7	2018	07-2018	6344	10.32	5.1	4
7	2018	07-2018	100	10.77	5.04	5
7	2018	07-2018	69	10.82	5.03	6
7	2018	07-2018	79	11.21	4.97	7
7	2018	07-2018	6111	10.71	4.83	8
7	2018	07-2018	6214	9.71	4.83	8
7	2018	07-2018	19422	10.11	4.81	10

◀ ▶

For example - let's interpret the first row of the **interest\_metrics** table together:

<b>_month</b>	<b>_year</b>	<b>month_year</b>	<b>interest_id</b>	<b>composition</b>	<b>index_value</b>	<b>ranki</b>
7	2018	07-2018	32486	11.89	6.19	1

◀ ▶

In July 2018, the **composition** metric is 11.89, meaning that 11.89% of the client's

customer list interacted with the interest **interest\_id = 32486** - we can link

**interest\_id** to a separate mapping table to find the segment name called "Vacation Rental Accommodation Researchers"

The **index\_value** is 6.19, means that the **composition** value is 6.19x the average composition value for all Fresh Segments clients' customer for this particular interest in the month of July 2018.

The **ranking** and **percentage\_ranking** relates to the order of **index\_value** records in each month year.

## Interest Map

This mapping table links the `interest_id` with their relevant interest information. You will need to join this table onto the previous `interest_details` table to obtain the `interest_name` as well as any details about the summary information.

<b>id</b>	<b>interest_name</b>	<b>interest_summary</b>
1	Fitness Enthusiasts	Consumers using fitness tracking apps and websites.
2	Gamers	Consumers researching game reviews and cheat cod
3	Car Enthusiasts	Readers of automotive news and car reviews.
4	Luxury Retail Researchers	Consumers researching luxury product reviews and c
5	Brides & Wedding Planners	People researching wedding ideas and vendors.
6	Vacation Planners	Consumers reading reviews of vacation destinations
7	Motorcycle Enthusiasts	Readers of motorcycle news and reviews.
8	Business News Readers	Readers of online business news content.
12	Thrift Store Shoppers	Consumers shopping online for clothing at thrift stor
13	Advertising Professionals	People who read advertising industry news.

## Interactive SQL Instance

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

You can click on the `Edit on DB Fiddle` link on the top right hand corner of the embedded session below and it will take you to a fully functional SQL editor where you can write your own queries to analyse the data.

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!

Schema SQL    Query SQL    Results    Edit on DB Fiddle

```

CREATE SCHEMA fresh_segments;

DROP TABLE IF EXISTS fresh_segments.json_data;
CREATE TABLE fresh_segments.json_data (raw_data JSON);

CREATE TABLE fresh_segments.interest_map (
    "id" INTEGER,
    "interest_name" TEXT,
    "interest_summary" TEXT,
    "created_at" TIMESTAMP,
    "last_modified" TIMESTAMP
);

INSERT INTO fresh_segments.interest_map
    ("id", "interest_name", "interest_summary", "created_at", "last_modified")
VALUES
    ('1', 'Fitness Enthusiasts', 'Consumers using fitness tracking apps and websites.', '2016-05-26 14:57:59', '2016-05-26 14:57:59'),
    ('2', 'Gamers', 'Consumers researching game reviews and cheat codes.', '2016-05-26 14:57:59', '2016-05-26 14:57:59'),
    ('3', 'Car Enthusiasts', 'Readers of automotive news and car reviews.', '2016-05-26 14:57:59', '2016-05-26 14:57:59'),
    ('4', 'Luxury Retail Researchers', 'Consumers researching luxury product reviews and gift ideas.', '2016-05-26 14:57:59', '2016-05-26 14:57:59'),
    ('5', 'Brides & Wedding Planners', 'People researching wedding ideas and vendors.', '2016-05-26 14:57:59', '2016-05-26 14:57:59'),
    ('6', 'Vacation Planners', 'Consumers reading reviews of vacation destinations and accommodations.', '2016-05-26 14:57:59', '2016-05-26 14:57:59'),
    ('7', 'Motorcycle Enthusiasts', 'Readers of motorcycle news and reviews.', '2016-05-26 14:57:59', '2016-05-26 14:57:59'),
    ('8', 'Business News Readers', 'Readers of online business news content.', '2016-05-26 14:57:59', '2016-05-26 14:57:59'),
    ('12', 'Thrift Store Shoppers', 'Consumers shopping online for clothing at thrift stores and vintage boutiques.', '2016-05-26 14:57:59', '2016-05-26 14:57:59'),
    ('13', 'Advertising Professionals', 'People who read advertising industry news.', '2016-05-26 14:57:59', '2016-05-26 14:57:59'),
    ('14', 'NFL Fans', 'People reading articles and websites about football and the NFL.', '2016-05-26 14:57:59', '2016-05-26 14:57:59'),
    ('15', 'NBA Fans', 'People reading articles and websites about basketball and the NBA.', '2016-05-26 14:57:59', '2016-05-26 14:57:59'),
    ('16', 'NCAA Fans', 'People reading articles and websites about college sports and the NCAA.', '2016-05-26 14:57:59', '2016-05-26 14:57:59'),
    ('17', 'MLB Fans', 'People reading articles and websites about baseball and the MLB.', '2016-05-26 14:57:59', '2016-05-26 14:57:59'),
    ('18', 'Nascar Fans', 'People reading articles and websites about nascar.', '2016-05-26 14:57:59', '2016-05-26 14:57:59'),
    ('19', 'Eco-Conscious Consumers', 'People researching climate change and sustainable living.', '2016-05-26 14:57:59', '2016-05-26 14:57:59'),
    ('20', 'Moviegoers', 'Consumers researching box office listings and buying movie tickets.', '2016-05-26 14:57:59', '2016-05-26 14:57:59'),
    ('21', 'DIYers', 'Consumers researching DIY projects.', '2016-05-26 14:57:59', '2018-05-23 11:45:00'),
    ('22', 'Grill Masters', 'Grill owners researching new recipes.', '2016-05-26 14:57:59', '2018-05-23 11:45:00'),
    ('23', 'Techies', 'Readers of tech news and gadget reviews.', '2016-05-26 14:57:59', '2018-05-23 11:45:00'),
    ('24', 'Home Design Enthusiasts', 'Consumers researching home design styles and trends.', '2016-05-26 14:57:59', '2018-05-23 11:45:00'),
    ('25', 'Doctors', 'People reading medical journals and using online prescribing services.', '2016-05-26 14:57:59', '2018-05-23 11:45:00'),
    ('26', 'Streaming Video Fans', 'Consumers streaming feature length movies and TV shows.', '2016-05-26 14:57:59', '2018-05-23 11:45:00'),
    ('27', 'Auto Insurance Shoppers', 'Consumers shopping for auto insurance.', '2016-05-26 14:57:59', '2018-05-23 11:45:00'),
    ('28', 'Teachers', 'People reading education industry publications and teaching advice websites.', '2016-05-26 14:57:59', '2018-05-23 11:45:00'),
    ('29', 'College Students', 'People visiting university websites.', '2016-05-26 14:57:59', '2018-05-23 11:45:00'),
    ('30', 'Department Store Shoppers', 'Consumers shopping at multi-category department stores.', '2016-05-26 14:57:59', '2018-05-23 11:45:00'),
    ('31', 'Home Decor Shoppers', 'Consumers shopping for home design, art and accessories.', '2016-05-26 14:57:59', '2018-05-23 11:45:00')

```

## Case Study Questions

The following questions can be considered key business questions that are required to be answered for the Fresh Segments team.

Most questions can be answered using a single query however some questions are more open ended and require additional thought and not just a coded solution!

## Data Exploration and Cleansing

1. Update the `fresh_segments.interest_metrics` table by modifying the `month_year` column to be a date data type with the start of the month
2. What is count of records in the `fresh_segments.interest_metrics` for each `month_year` value sorted in chronological order (earliest to latest) with the null values appearing first?
3. What do you think we should do with these null values in the `fresh_segments.interest_metrics`
4. How many `interest_id` values exist in the `fresh_segments.interest_metrics` table but not in the `fresh_segments.interest_map` table? What about the other way around?
5. Summarise the `id` values in the `fresh_segments.interest_map` by its total record count in this table
6. What sort of table join should we perform for our analysis and why? Check your logic by checking the rows where `interest_id = 21246` in your joined output and include all columns from `fresh_segments.interest_metrics` and all columns from `fresh_segments.interest_map` except from the `id` column.
7. Are there any records in your joined table where the `month_year` value is before the `created_at` value from the `fresh_segments.interest_map` table? Do you think these values are valid and why?

## Interest Analysis

1. Which interests have been present in all `month_year` dates in our dataset?
2. Using this same `total_months` measure - calculate the cumulative percentage of all records starting at 14 months - which `total_months` value passes the 90% cumulative percentage value?

3. If we were to remove all `interest_id` values which are lower than the `total_months` value we found in the previous question - how many total data points would we be removing?
4. Does this decision make sense to remove these data points from a business perspective? Use an example where there are all 14 months present to a removed `interest` example for your arguments - think about what it means to have less months present from a segment perspective.
5. After removing these interests - how many unique interests are there for each month?

## Segment Analysis

1. Using our filtered dataset by removing the interests with less than 6 months worth of data, which are the top 10 and bottom 10 interests which have the largest composition values in any `month_year`? Only use the maximum composition value for each interest but you must keep the corresponding `month_year`
2. Which 5 interests had the lowest average `ranking` value?
3. Which 5 interests had the largest standard deviation in their `percentile_ranking` value?
4. For the 5 interests found in the previous question - what was minimum and maximum `percentile_ranking` values for each interest and its corresponding `year_month` value? Can you describe what is happening for these 5 interests?
5. How would you describe our customers in this segment based off their composition and ranking values? What sort of products or services should we show to these customers and what should we avoid?

## Index Analysis

The `index_value` is a measure which can be used to reverse calculate the average composition for Fresh Segments' clients.

Average composition can be calculated by dividing the `composition` column by the `index_value` column rounded to 2 decimal places.

1. What is the top 10 interests by the average composition for each month?

2. For all of these top 10 interests - which interest appears the most often?
3. What is the average of the average composition for the top 10 interests for each month?
4. What is the 3 month rolling average of the max average composition value from September 2018 to August 2019 and include the previous top ranking interests in the same output shown below.
5. Provide a possible reason why the max average composition might change from month to month? Could it signal something is not quite right with the overall business model for Fresh Segments?

Required output for question 4:

<b>month_year</b>	<b>interest_name</b>	<b>max_index_composition</b>	<b>3_month_rolling_avg</b>
2018-09-01	Work Comes First Travelers	8.26	7.61
2018-10-01	Work Comes First Travelers	9.14	8.20
2018-11-01	Work Comes First Travelers	8.28	8.56
2018-12-01	Work Comes First Travelers	8.31	8.58
2019-01-01	Work Comes First Travelers	7.66	8.08
2019-02-01	Work Comes First Travelers	7.66	7.88
2019-03-01	Alabama Trip Planners	6.54	7.29
2019-04-01	Solar Energy Researchers	6.28	6.83
2019-05-01	Readers of Honduran Content	4.41	5.74
2019-06-01	Las Vegas Trip Planners	2.77	4.49
2019-07-01	Las Vegas Trip Planners	2.82	3.33
2019-08-01	Cosmetics and Beauty Shoppers	2.73	2.77

## Conclusion

You have probably come across this concept of customer segments or marketing segments in your everyday life, maybe without you even noticing it!

Segments or audiences are super popular in the digital marketing space and using these interests or traits of customers is a mainstay of massive businesses like Google, Facebook, Instagram, LinkedIn and other social media where there are targeted advertising.

Traditional businesses such as this client for Fresh Segments usually upload their customer emails or matched cookies into various digital marketing systems in order to generate some sort of match, usually using some machine learning methods, to other similar customers with the same interests.

Hopefully this case study helps you think about how these index metrics and compositions can be used for digital marketing!

## 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](#)
- [Case Study #5 - Data Mart](#)
- [Case Study #6 - Clique Bait](#)
- [Case Study #7 - Balanced Tree Clothing Co.](#)
- [Case Study #8 - Fresh Segments](#)

Share: [Twitter](#), [Facebook](#)

