

✔ Milestone 7 | FastKitchen Customers

INTRODUCTION: In this SkillBuilder, you've learned additional ways of joining tables together, with three different types of outer join: the left join, right join, and the full outer join. While an inner join retains only information when there's a match between the joined tables, an outer join will also output information that can only be found in one table.

FastKitchen is a fictitious restaurant and the dataset you'll be working with is constructed. While this dataset might not represent real data, it emulates characteristics of real data. When you're interviewing for a job, you might be asked to look at this kind of data to show off your skills in a context related to the company and the job position!

HOW IT WORKS: Follow the prompts in the questions below to investigate your data. Post your answers in the provided boxes: the **yellow boxes** for the queries you write, and **blue boxes** for text-based answers. When you're done, export your document as a pdf file and submit it on the Milestone page – see instructions for creating a PDF at the end of the Milestone.

RESOURCES: If you need hints on the Milestone or are feeling stuck, there are multiple ways of getting help. Attend Drop-In Hours to work on these problems with your peers, or reach out to the HelpHub if you have questions. Good luck!

PROMPT: In this Milestone, you'll step into the shoes of a data contractor who is helping a new fast-food restaurant understand their customer base. You will need to make use of one type of outer join to help the restaurant manager combine information about their customers. These customers include registered customers who have accounts on the restaurant's website, and guest customers who do not register for accounts.

– Data Set **Description**

The data in this Milestone (`fastkitchen.*`) depicts orders made at a fictional takeout-only fast food restaurant in the Midwestern United States. The restaurant has an online site where customers can put in orders for carryout or delivery; customers can also make orders offline at the restaurant's storefront. You will be working with two tables in this Milestone: `orders` and `users`.

Each row in the `orders` table is a single order that was placed at the restaurant. This table has seven columns:

- **order_id** - unique order id, primary key
- **timestamp** - when the order was made
- **user_id** - user_id for registered accounts, blank if guest customer
- **order_type** - whether the order was made onsite, online carryout, or online delivery
- **subtotal** - base amount for the order
- **tip** - amount of tip, if any, left by the customer
- **total** - subtotal + tip

Customers have the option of creating a user account, which can be used both in person and online. The `users` table has five columns:

- **user_id** - unique user_id value, primary key
 - **reg_timestamp** - when the user registered their account
 - **city** - user city
 - **state** - two-letter code for state
 - **zip** - zip code
-

– **Task 1:** Explore information about orders.

To start off, let's warm up with some questions on the individual tables, before we ask questions that require joining the two tables together. Let's look at the `orders` table first.

A. What is the average total amount (including tips) spent per order?

(paste your query below 📌)

```
SELECT AVG(total) as avg_total  
FROM fastkitchen.orders
```

(write your **answer** below 📌)

The average spent per order is \$22.21.

B. Compare the average subtotals, tips, and totals spent by each order type (onsite, carryout, delivery). Are there any major differences between order types?

(paste your query below 📌)

```
SELECT  
  order_type,  
  AVG(subtotal) as avg_subtotal,  
  AVG(tip) as avg_tip,  
  AVG(total) as avg_total  
FROM fastkitchen.orders  
GROUP BY order_type
```

(write your **answer** below 📌)

There are only fairly small differences between order types in terms of the overall average bill size. Onsite orders are smallest, at \$22.11 on average, close to carryout orders at \$22.17, with online orders largest at \$22.48. Interestingly, the subtotal average is highest for online orders, tips are also lowest for online orders.

C. Write a query to count the number of orders made by **registered users**.

(paste your query below 📌)

```
SELECT
  COUNT(order_id)
FROM fastkitchen.orders
WHERE user_id IS NOT NULL
```

D. Now, write a query to return the number of orders made by **non-registered customers**. Remember, non-registered customers don't have a user id. Compared alongside the returned count from the query in C, which group is larger?

(paste your query below 📌)

```
SELECT
  COUNT(order_id)
FROM fastkitchen.orders
WHERE user_id IS NULL
```

(write your **answer** below 📌)

There are somewhat more orders created by guests: 2088 guest orders versus 1932 user orders.

– Task 2: Explore information about registered users.

Next, we'll check out the `users` table.

A. Write a query that counts the number of users by city. Which city has the highest number of users, and how many users are there?

(paste your query below 📌)

```
SELECT
  city,
  COUNT(user_id) AS n_users
FROM fastkitchen.users
GROUP BY city
```

(write your **answer** below 📌)

Most user accounts are based in Allen, with 212 users. There are only 98 users in total for Maebe and Nulle Pointe combined.

- B. Expand the query so that you group by zip code as well. Does this help explain what you found in part 2A?

(paste your query below 📌)

```
SELECT
  city,
  zip,
  COUNT(user_id) AS n_users
FROM fastkitchen.users
GROUP BY
  city,
  zip
```

(write your **answer** below 📌)

When we group by zip codes in addition to city, it's pretty clear to see why Allen has about three times as many users as Nulle Pointe: the city limits cover three different zip codes instead of just one. This grouping also shows that Maebe's zip code, 63225, is quite a bit smaller than the others in terms of user counts. This might suggest that Maebe is a smaller city overall, or is just too far

out from the restaurant's physical location to warrant more regular customers.

– Task 3: How do orders compare between zip codes and cities?

Finally, we'll combine the `user` and `orders` tables into a single, joined table.

- A. To start, simply write a query that returns all of the columns, joining the two tables on the `user_id` column. Make sure that you choose a join that keeps all of the orders, even when there isn't a matching registered user.

(paste your query below 📌)

```
SELECT *
FROM fastkitchen.orders as o
LEFT JOIN fastkitchen.users as u
  ON o.user_id = u.user_id
```

Note: You could also `RIGHT JOIN` the tables in the opposite order or use a `FULL OUTER JOIN`, but the `LEFT JOIN` used in the example is useful for implying the most meaningful relationship between the two tables.

- B. Add to the query from 3A to answer the following question: in which zip code is the user with the highest amount of money spent?

(paste your query below 📌)

```
SELECT
  o.user_id,
  u.zip,
  SUM(o.total) as total_spent
FROM fastkitchen.orders as o
LEFT JOIN fastkitchen.users as u
  ON o.user_id = u.user_id
```

```
GROUP BY
  o.user_id,
  u.zip
ORDER BY total_spent DESC
```

You could also add a WHERE statement to the query to filter out guest users (WHERE o.user_id IS NOT NULL), but it's not especially required since they only take up one line in the output.

(write your **answer** below 📌)

The user with the highest spending (\$1103.11) hails from zip code 63222.

You might also notice from the output that the highest spender isn't also the user with the highest amount of tips. That distinction belongs to the user with the 4th-highest amount spent, who had a much lower subtotal, but was consistent in tipping at 20%.

- C. Write a query that returns the average total amount spent per order by zip code. How many of the zip codes spend more on average than non-registered guest customers?

(paste your query below 📌)

```
SELECT
  u.zip,
  AVG(o.total) as avg_total
FROM fastkitchen.orders as o
LEFT JOIN fastkitchen.users as u
  ON o.user_id = u.user_id
GROUP BY
  u.zip
ORDER BY avg_total DESC
```

(Sorting is not strictly required, but helps answer the question.)

(write your **answer** below 📌)

There is quite the variability in how much each zip code spends on average. On average, guests have orders totaling \$21.98. Two zip codes are smaller than this, each around \$19.50, while three zip codes are larger than the guest average, ranging from \$23.18 in 63222 to \$24.22 in 63216.

– Submission

Great work completing this Milestone! To submit your completed Milestone, you will need to download / export this document as a PDF and then upload it to the Milestone submission page. You can find the option to download as a PDF from the File menu in the upper-left corner of the Google Doc interface.