Week 1 Lab: Introduction to SQL

Parch & Posey Database

Parch & Posey is a fabricated non-real company that sells paper:

- There are 50 sales reps spread across the United States in 4 regions.
- There are 3 types of paper. Regular, Poster and Glossy.
- The clients are primarily large Fortune's 100 companies.

The Parch & Posey database is provided by Mode Analytics (https://mode.com/).

Questions answered using Parch & Posey data are meant to simulate real word problems. Using SQL, we will help Parch & Posey answer tricky questions like: Which of their product lines is worst performing? Which of their market channels they should make a great investment in.

In the Parch & Posey database there are five tables (essentially 5 spreadsheets):

- web_events
- accounts
- orders
- sales_reps
- Region

Figure 1 shows the ERD (entity relationship diagram) for Parch and Posey.

Most of the variables in each table are self-explanatory, but some are not. Here is a description of them:

- channel: marketing channel (twitter, adwords, organic, banner, facebook, direct)
- lat: latitude of the company location
- long: longtitude of the company location
- primary_poc: primary point of contact
- sales_rep_id: sales representative id
- standard_qty: quantity or standard papers ordered
- poster_qty: quantity of poster papers ordered
- glossy_qty: quantity of glossy papers ordered
- total: total quantity of papers ordered (standard + poster + glossy)
- standard_amt_usd: dollar amount paid for the standard papers
- poster_amt_usd: dollar amount paid for the poster papers
- gloss_amt_usd: dollar amount paid for the gloss papers
- total_amt_usd: dollar amount paid for all paper orders

Creating the Database

- Open pgAdmin 4
- Create a new database, and give it a name. In our case, let's call it "Parch".
- Right click on the Parch database, and choose "Qurey Tool…"
- Load the database file "parch.sql" and run all the commands (We will discuss these commands later during the semester, but for now, just think of this step as a database creation and connection)
- Right click on the Parch database, and choose "Qurey Tool…" to create an empty SQL file to query the Parch database. Don't forget to save the file.

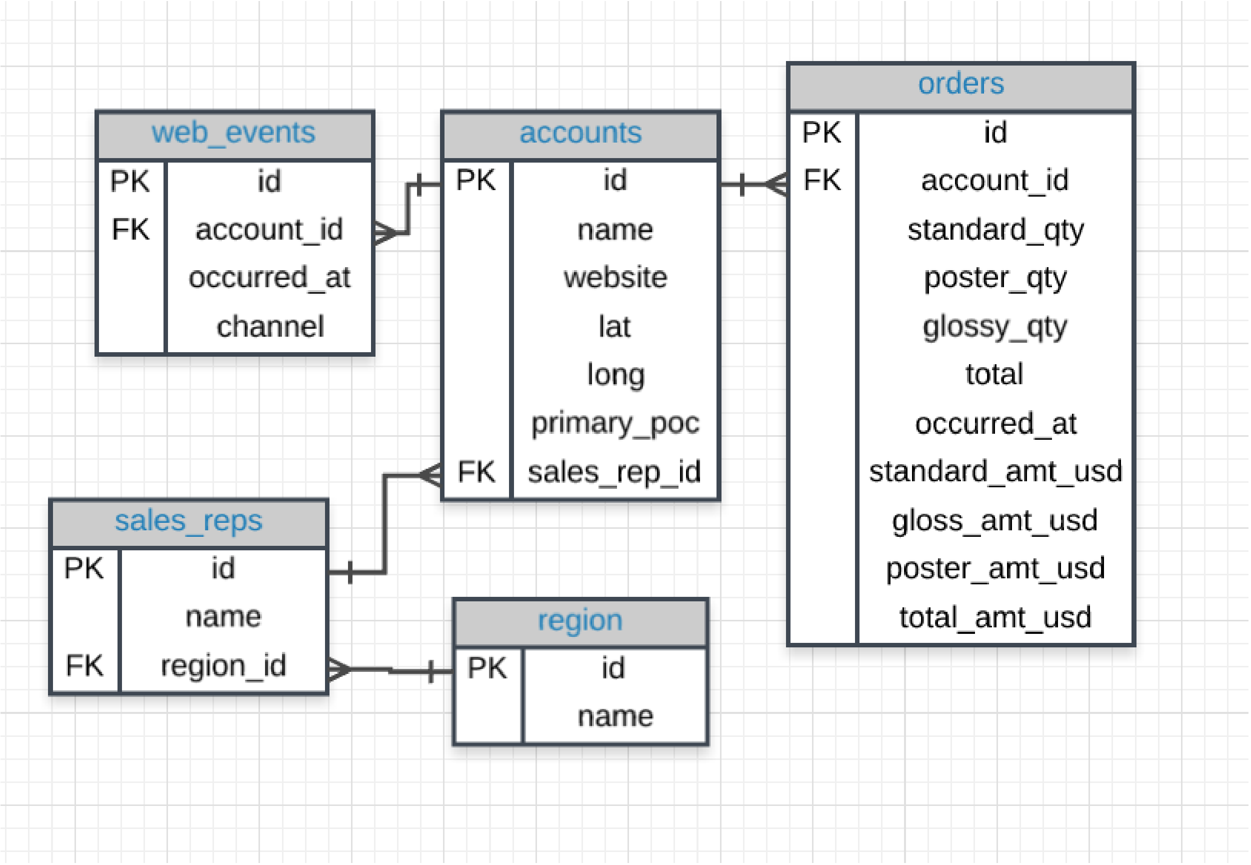

Figure 1: Parch and Posey ERD

Questions

SELECT & FROM

1. Take a quick look at all the 5 tables in the database.
2. Generate a list of all the names of the sales representatives and their IDs that Parch and Posey havs in their database.
3. **YOUR TURN** Write your own query to select only the `id`, `account_id`, and `occurred_at` columns for all orders in the `orders` table.

LIMIT

4. Showing just the first 10 observations of the `sales_reps` table with all of the columns.

ORDER BY

5. Write a SQL query to look up the most 10 recent orders.
6. **YOUR TURN** Write a query to return the 10 earliest orders in the orders table. Include the `id`, `occurred_at`, and `total_amt_usd`.
7. **YOUR TURN** Write a query to return the date of the 5 most expensive orders of papers. Make sure to include the order id, and the total dollar amount.
8. Write a query that that returns all accounts sorted by account id. For each account, list the orders sorted by total dollar amount in descending order.
9. **YOUR TURN** Write a query that returns the top 5 rows with the highest total number of orders. If two rows have the same number of total orders, then use the total dollar amount (highest to lowest) to break the tie. (Return the ID, the total number of orders, and the total dollar amount)

WHERE

10. Write a query to show only orders from our customer with an account ID 4251.
11. **YOUR TURN** Write a query to pull the first 5 rows and all columns from the orders table that have spent a total amount less than or equal to $500.
12. For the account `Exxon Mobil`, return the the company `name`, `website`, and the primary point of contact (`primary_poc`).

Derived Columns

13. For each order, return the quantity of non-standard papers (poster and gloss). Include the account_id and date.
14. **YOUR TURN** Find the percentatge of standard papers ordered for each order. Limit the results to the first 5 orders, and include the id and account_id fields.

Logical Operators

15. Find the website for the Whole Foods account.
16. **YOUR TURN** Find all companies whose names contain the string ‘one’ somewhere in the name.

17. **YOUR TURN** Use the accounts table to find all the companies whose names start with ‘C’.
18. **YOUR TURN** Use the accounts table to find all companies whose names end with ‘s’.
19. Find the account ID for both Apple and Walmart.
20. **YOUR TURN** Get all account IDs for those comapnies who were contacted via ads displayed on twitter or via google adwords.
21. Find the account name, primary poc, and sales rep id for all stores except Walmart, Target, and Nordstrom.
22. **YOUR TURN** Find all company account ids who were contacted via any method except using twitter or adwords methods.
23. **YOUR TURN** Find all the companies whose names do not start with ‘C’.
24. Pull all orders that occurred between April 1, 2016 and September 1, 2016.
25. **YOUR TURN** Write a query that returns all the orders where the standard_qty is over 1000, the poster_qty is 0, and the gloss_qty is 0.
26. **YOUR TURN** Which companies who were contacted via twitter or adwords channel and started their account at any point in 2016 sorted from newest to oldest.
27. Find all customers whose orders did not have at least one type of paper.
28. Find all customers whose orders odid not have at least one type of paper and the order occurred after September 1, 2016. Sort the result from older transactions to the newest.
29. **YOUR TURN** Find list of orders ids where either gloss_qty or poster_qty is greater than 4000. Only include the id field in the resulting table.
30. **YOUR TURN** Write a query that returns a list of orders where the standard_qty is zero and either the gloss_qty or poster_qty is over 1000.