EXPLORING THE PARCH AND POSEY DATABASE (SQL ANALYSIS)

Data Analyst: Odudu Ochuko Richie

INTRODUCTION

In this analysis, I will be exploring the Parch and Posey database. I will carry out an analysis of this database to get insights that will be beneficial to the growth of the Parch and Posey company.

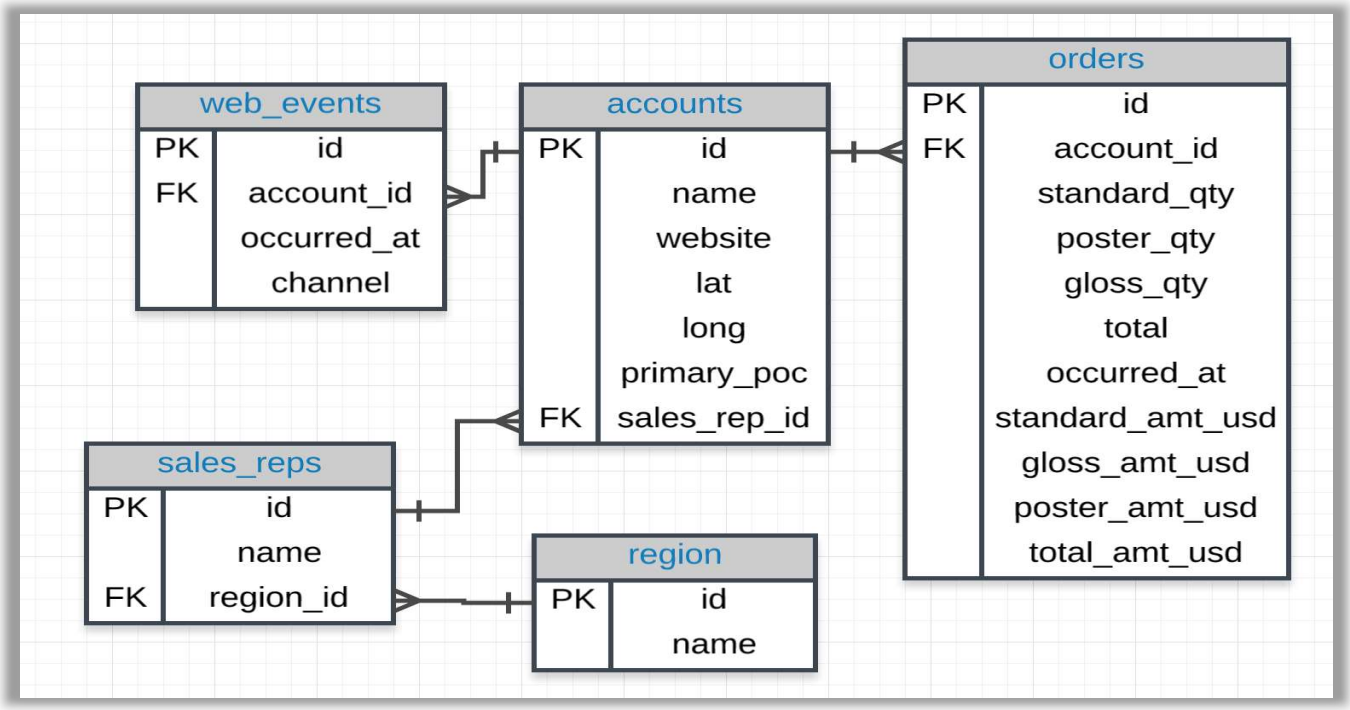
Firstly, Parch and Posey is a hypothetical company that sells paper. They have 50 sales reps spread across 4 regions in America. They sell 3 types of paper.

1. Standard
2. Gloss
3. Poster

Their customers are large companies they attract by advertising on Google, Facebook, and Twitter. In the Parch & Posey database, there are five tables web_events, accounts, orders, sales_reps, and region. The analysis will help understand the progress growth of Parch and Posey. In this analysis I will be looking at:

1. The account that placed the most orders
2. The account that placed the least orders
3. The account that placed the earliest order
4. The account that placed the latest order
5. Rank the total amount of paper ordered (from highest to lowest) for each account using a partition.
6. What kind of paper was ordered the most
7. Find the mean (average) amount spent per order on each paper type, as well as the mean amount of each paper type purchased per order.
8. The region for each sales rep along with their associated accounts. Sort the accounts alphabetically (A-Z) according to the account name.
9. Which year did Parch and Posey have the greatest sales in terms of total number of orders
10. Which month did Parch and Posey have the greatest sales in terms of total number of orders
11. The names of the sales reps in each region with the largest amount of sales (USD)
12. The region with the largest amount of total sales (USD), and how many total orders were placed?
13. The region with the smallest amount of total sales (USD), and how many total orders were placed?
14. The customer that spent the most (in total over their lifetime as a customer) in terms of total amount (USD). How many web events did they have for each channel?
15. Provide the name of the sales rep in each region with the largest amount of total amount (USD).
16. The account that purchased the most (in total over their lifetime as a customer) in terms of standard quantity paper, how many accounts still had more in total?

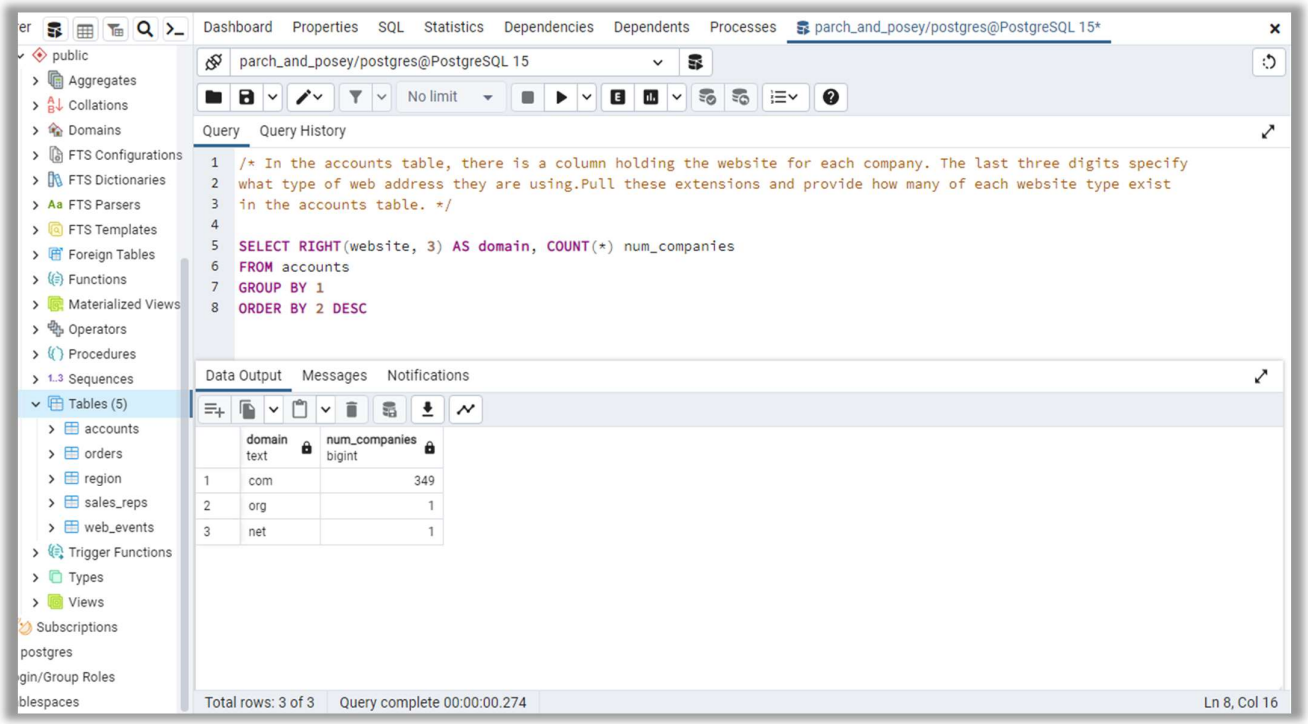
THE PARCH AND POSEY ERD DIAGRAM



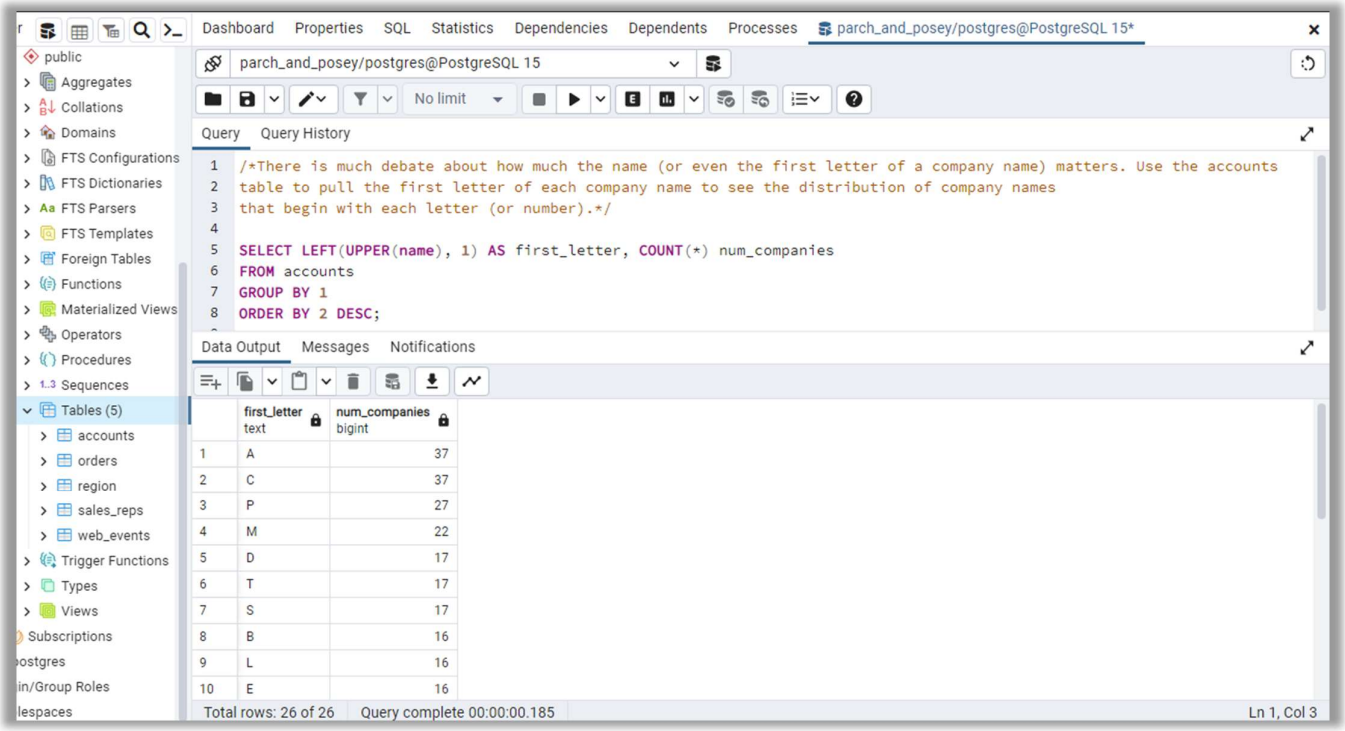
DATA CLEANING

1. Clean and restructure messy data
2. Convert columns to different data types.

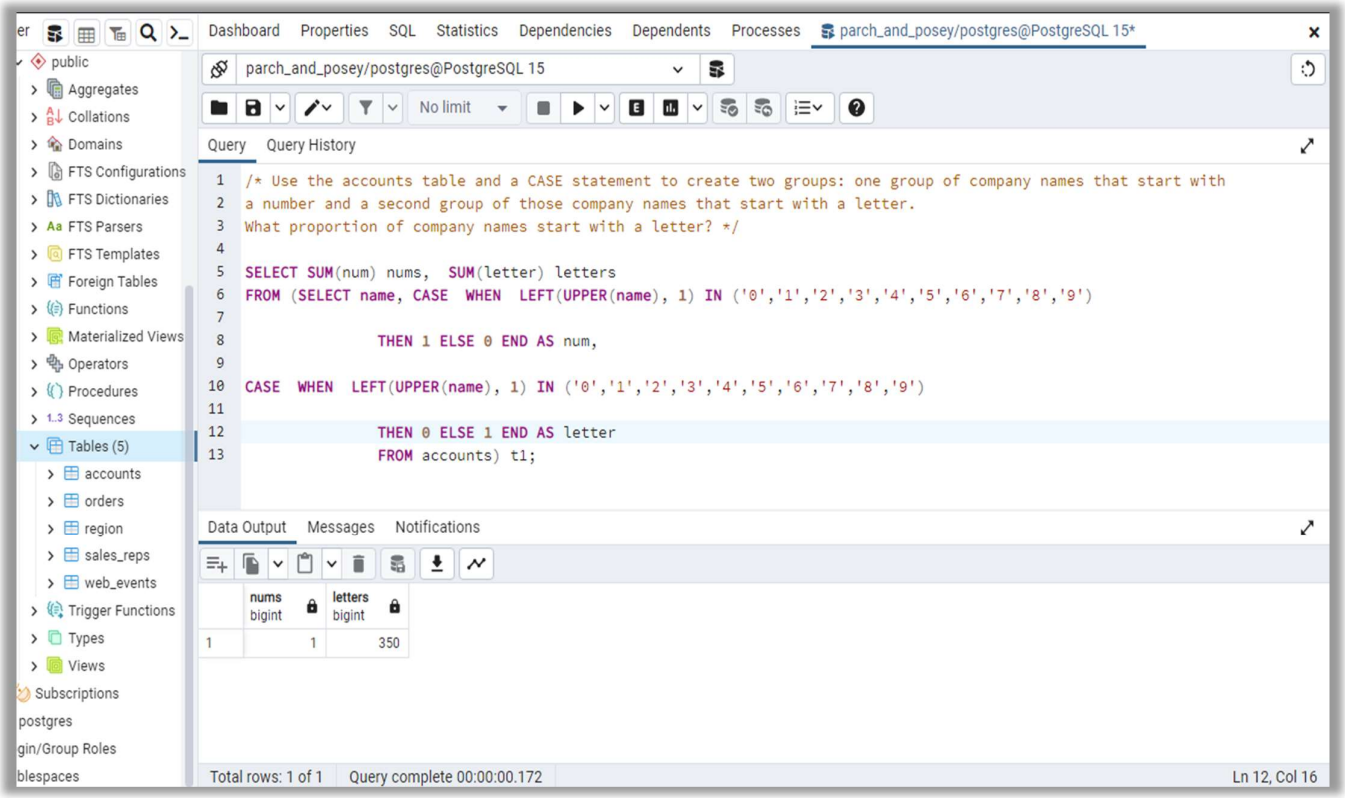
A. In the **accounts** table, there is a column holding the **website** for each company. The last three digits specify what type of web address they are using. Pull these extensions and provide how many of each website type exist in the **accounts** table.



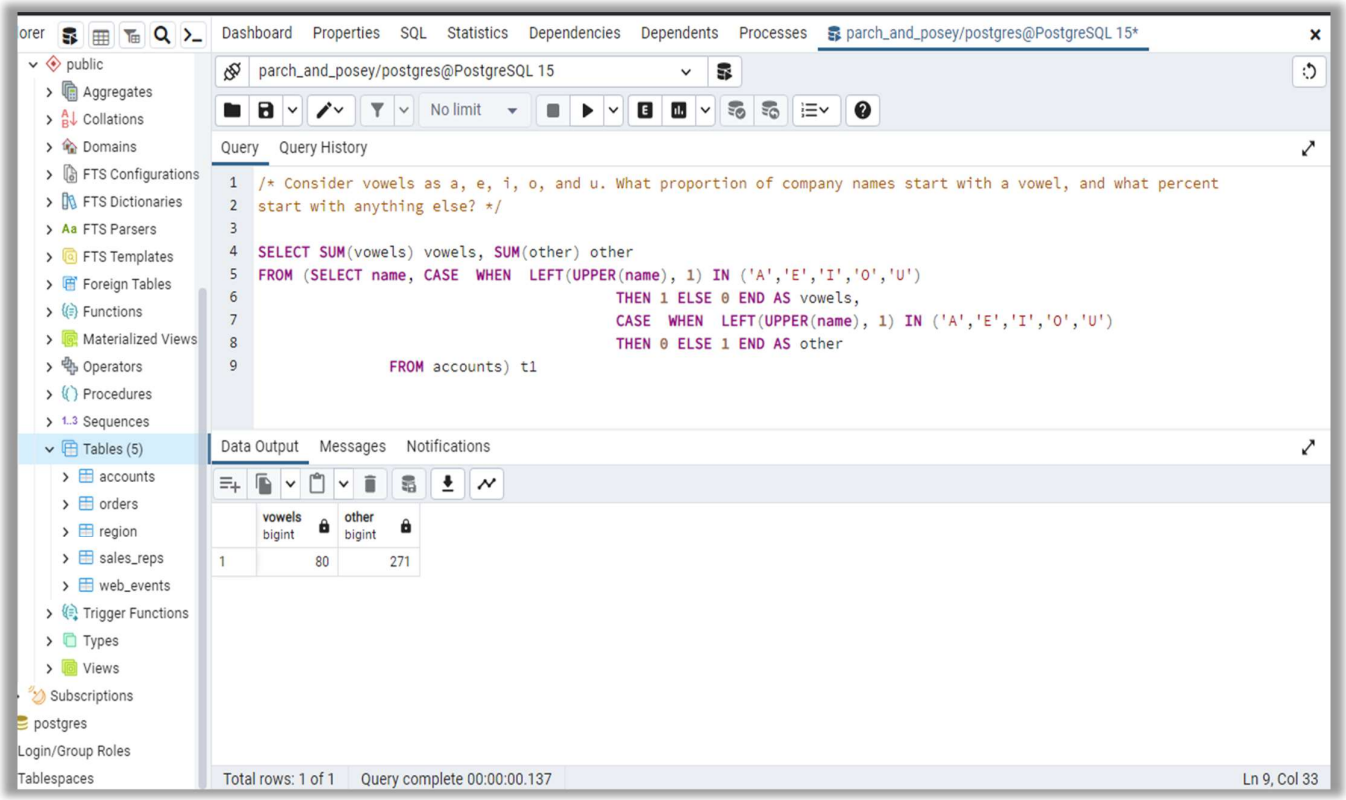
B. There is much debate about how much the name **(or even the first letter of a company name)** matters. Use the **accounts** table to pull the first letter of each company name to see the distribution of company names that begin with each letter (or number).



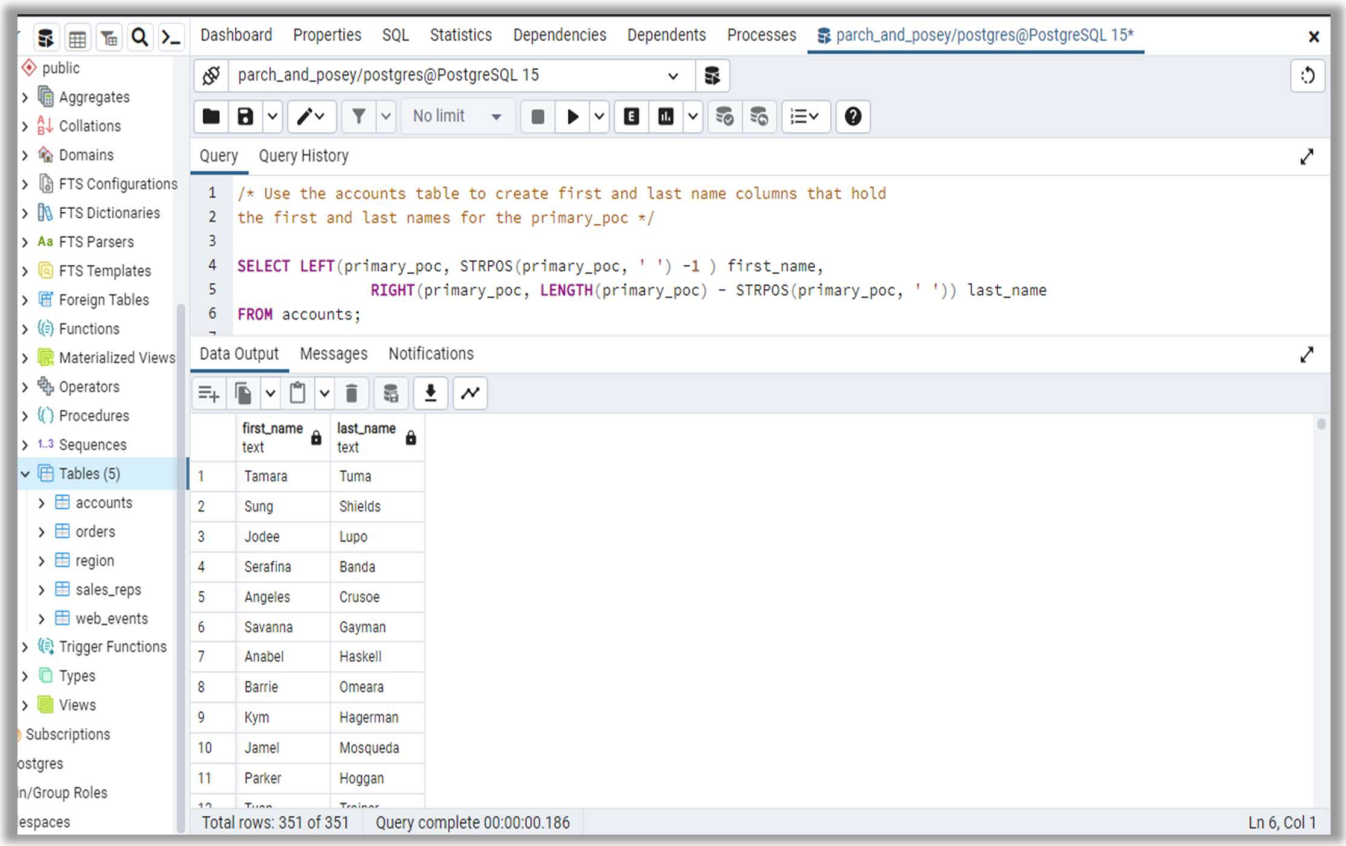
C. Use the **accounts** table and a **CASE** statement to create two groups: one group of company names that start with a number and a second group of those company names that start with a letter. What proportion of company names start with a letter?



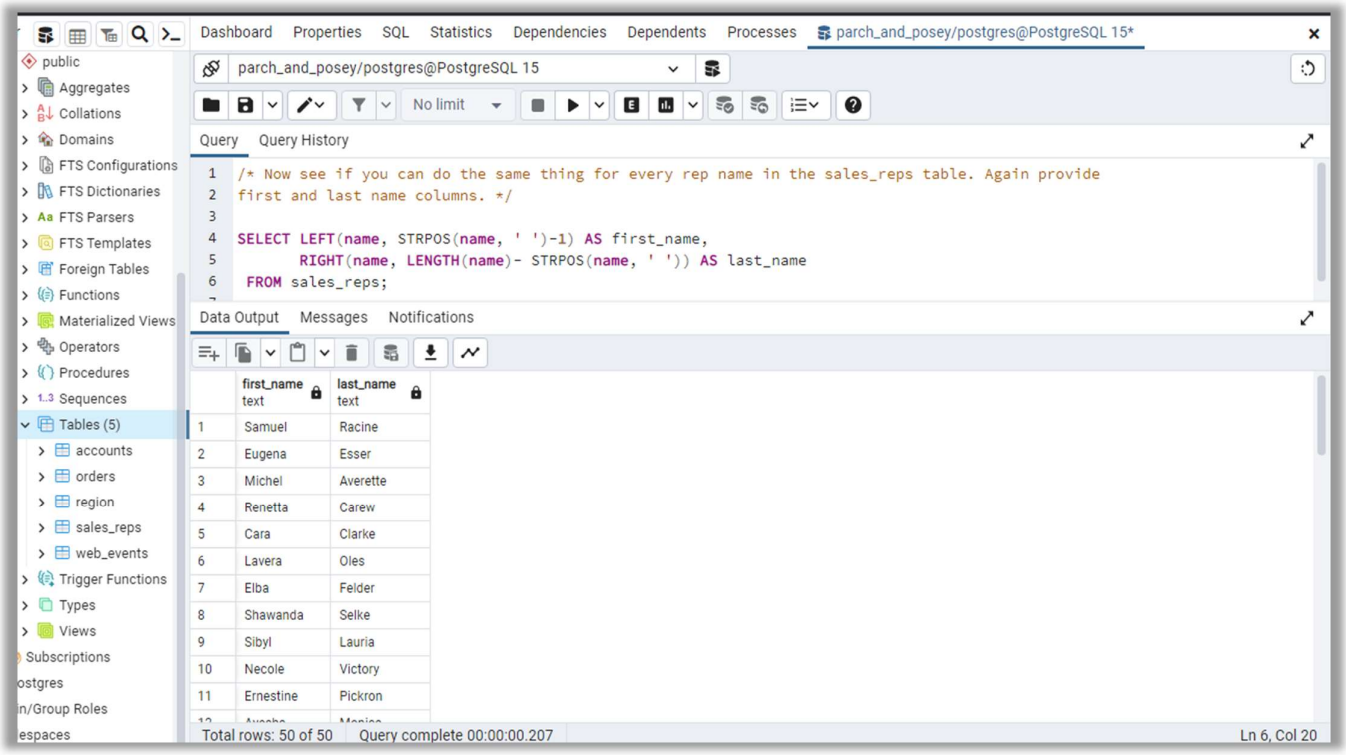
D. Consider vowels as a, e, i, o, and u. What proportion of company names start with a vowel, and what percent start with anything else?



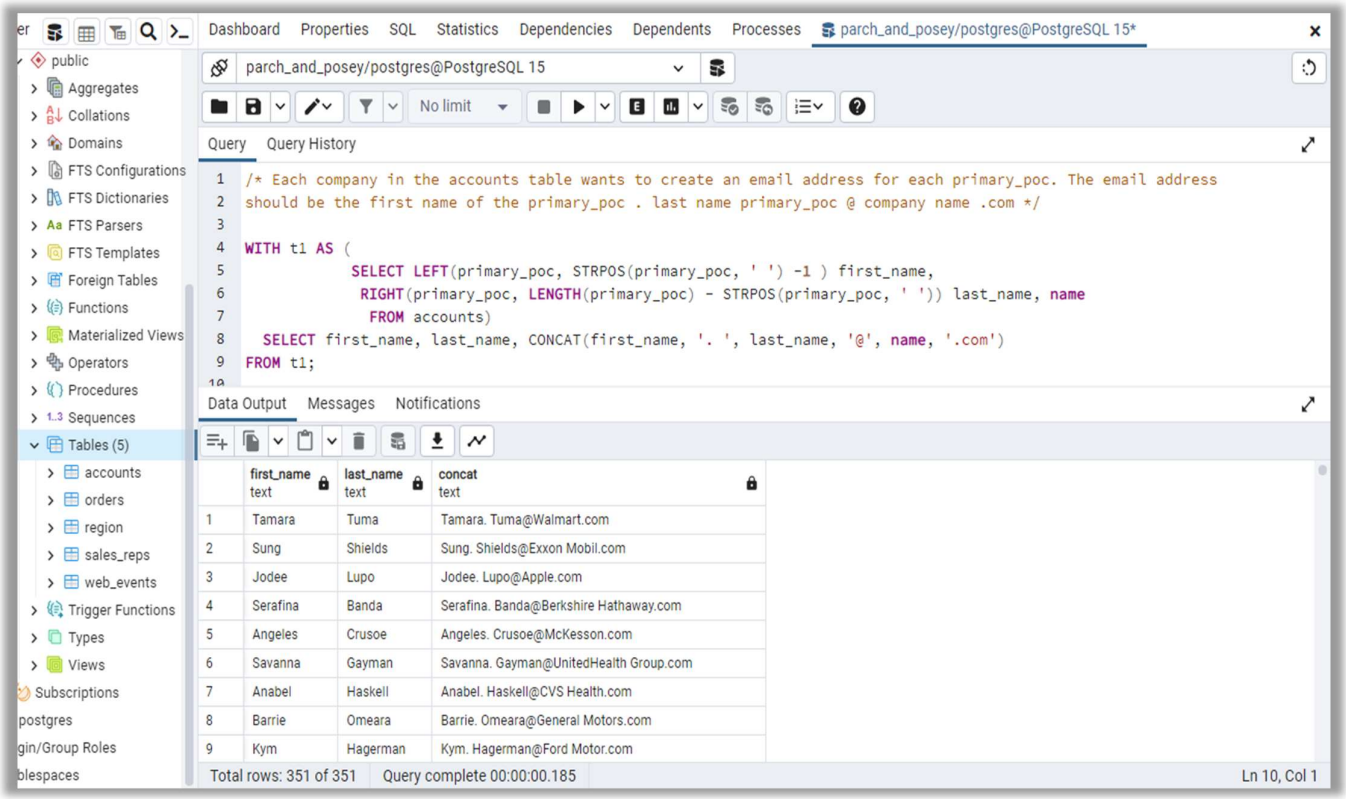
E. Use the **accounts** table to create first and last name columns that hold the first and last names for the **primary_poc**.



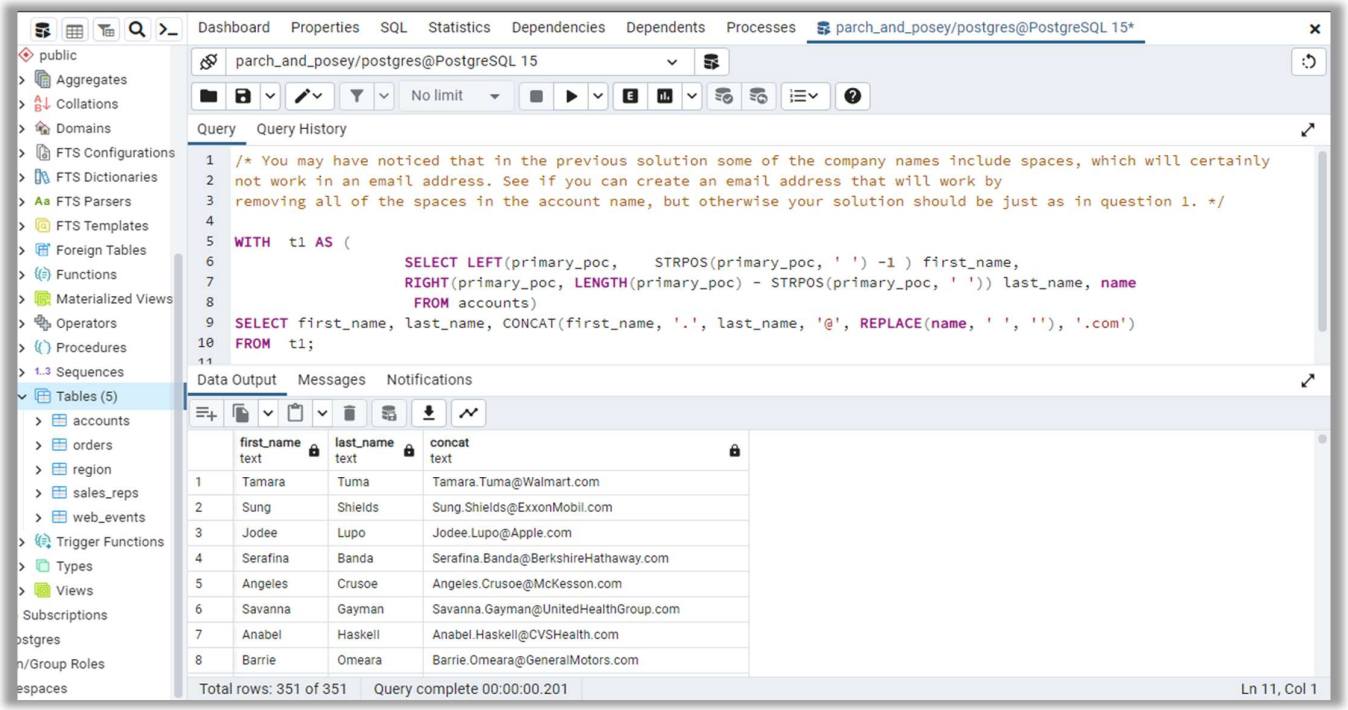
F. Now see if you can do the same thing for every rep **name** in the **sales_reps** table. Again, provide first and last name columns.



G. Each company in the accounts table wants to create an email address for each primary_poc. The email address should be the first name of the **primary_poc**. last name **primary_poc** @ company name .com.



H. You may have noticed that in the previous solution some of the company names include spaces, which will certainly not work in an email address. See if you can create an email address that will work by removing all of the spaces in the account name, but otherwise your solution should be just as in question 1



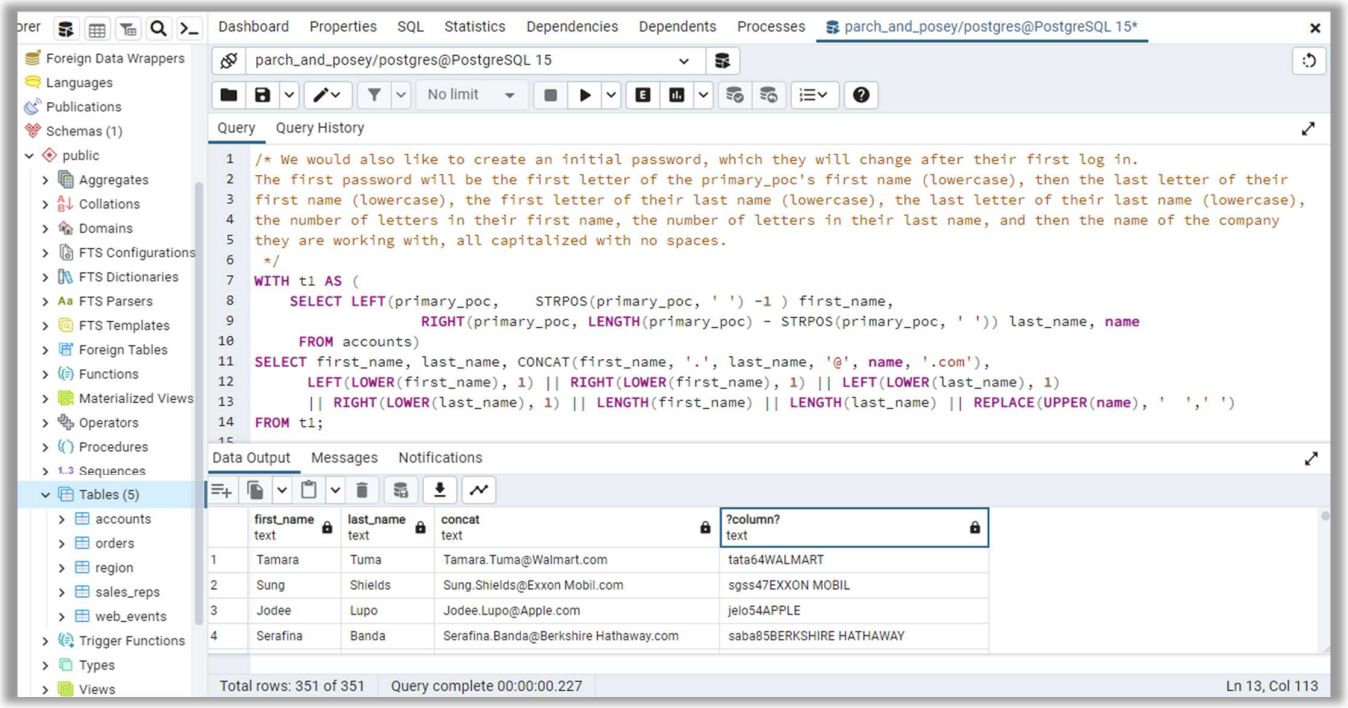
```
/* You may have noticed that in the previous solution some of the company names include spaces, which will certainly
not work in an email address. See if you can create an email address that will work by
removing all of the spaces in the account name, but otherwise your solution should be just as in question 1. */

WITH  t1 AS (
                    SELECT LEFT(primary_poc,    STRPOS(primary_poc, ' ') -1 ) first_name,
                    RIGHT(primary_poc, LENGTH(primary_poc) - STRPOS(primary_poc, ' ')) last_name, name
                     FROM accounts)
SELECT first_name, last_name, CONCAT(first_name, '.', last_name, '@', REPLACE(name, ' ', ''), '.com')
FROM  t1;
```

	first_name text	last_name text	concat text
1	Tamara	Tuma	Tamara.Tuma@Walmart.com
2	Sung	Shields	Sung.Shields@ExxonMobil.com
3	Jodee	Lupo	Jodee.Lupo@Apple.com
4	Serafina	Banda	Serafina.Banda@BerkshireHathaway.com
5	Angeles	Crusoe	Angeles.Crusoe@McKesson.com
6	Savanna	Gayman	Savanna.Gayman@UnitedHealthGroup.com
7	Anabel	Haskell	Anabel.Haskell@CVSHealth.com
8	Barrie	Omeara	Barrie.Omeara@GeneralMotors.com

Total rows: 351 of 351 | Query complete 00:00:00.201 | Ln 11, Col 1

I. We would also like to create an initial password, which they will change after their first log in. The first password will be the first letter of the primary_poc's first name (lowercase), then the last letter of their first name (lowercase), the first letter of their last name (lowercase), the last letter of their last name (lowercase), the number of letters in their first name, the number of letters in their last name, and then the name of the company they are working with, all capitalized with no spaces



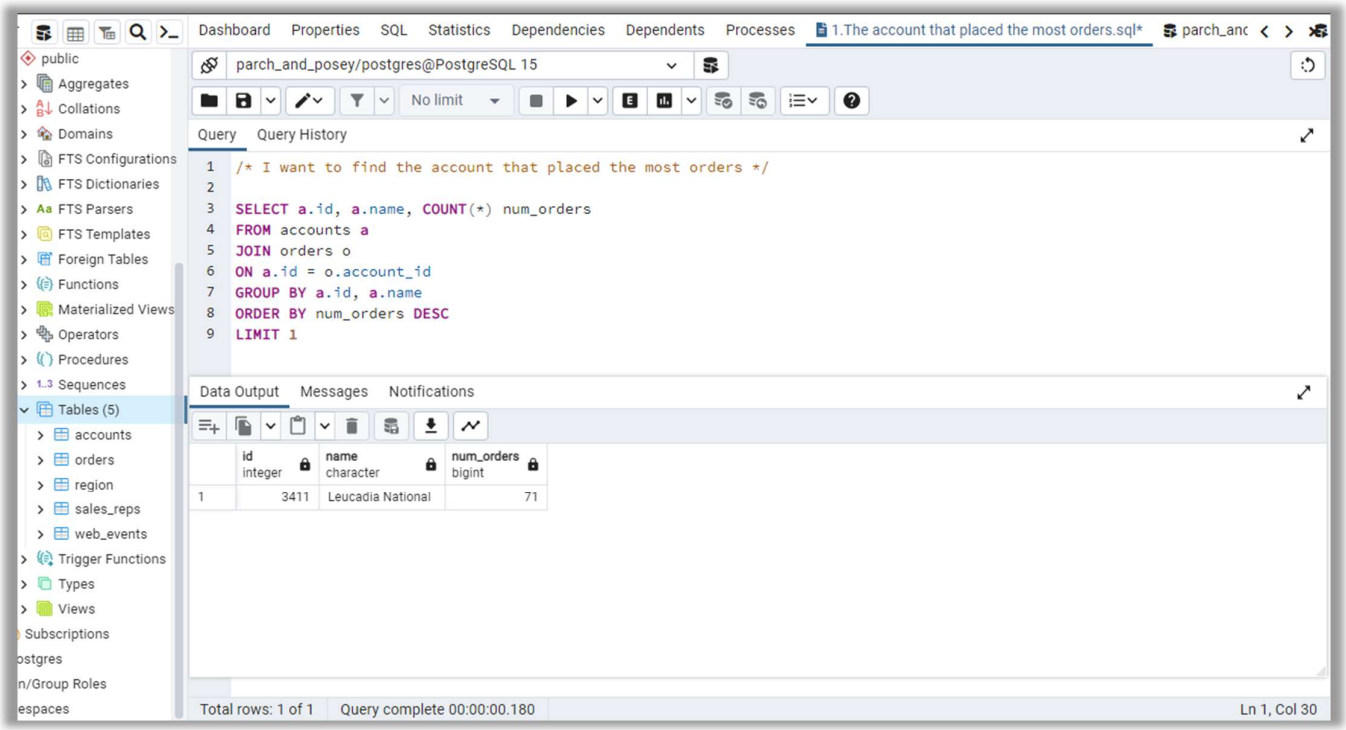
```
/* We would also like to create an initial password, which they will change after their first log in.
The first password will be the first letter of the primary_poc's first name (lowercase), then the last letter of their
first name (lowercase), the first letter of their last name (lowercase), the last letter of their last name (lowercase),
the number of letters in their first name, the number of letters in their last name, and then the name of the company
they are working with, all capitalized with no spaces.
 */
WITH t1 AS (
    SELECT LEFT(primary_poc,    STRPOS(primary_poc, ' ') -1 ) first_name,
                RIGHT(primary_poc, LENGTH(primary_poc) - STRPOS(primary_poc, ' ')) last_name, name
     FROM accounts)
SELECT first_name, last_name, CONCAT(first_name, '.', last_name, '@', name, '.com'),
      LEFT(LOWER(first_name), 1) || RIGHT(LOWER(first_name), 1) || LEFT(LOWER(last_name), 1)
      || RIGHT(LOWER(last_name), 1) || LENGTH(first_name) || LENGTH(last_name) || REPLACE(UPPER(name), '  ',' ')
FROM t1;
```

	first_name text	last_name text	concat text	?column? text
1	Tamara	Tuma	Tamara.Tuma@Walmart.com	tata64WALMART
2	Sung	Shields	Sung.Shields@Exxon Mobil.com	sgss47EXXON MOBIL
3	Jodee	Lupo	Jodee.Lupo@Apple.com	jelo54APPLE
4	Serafina	Banda	Serafina.Banda@Berkshire Hathaway.com	saba85BERKSHIRE HATHAWAY

Total rows: 351 of 351 | Query complete 00:00:00.227 | Ln 13, Col 113

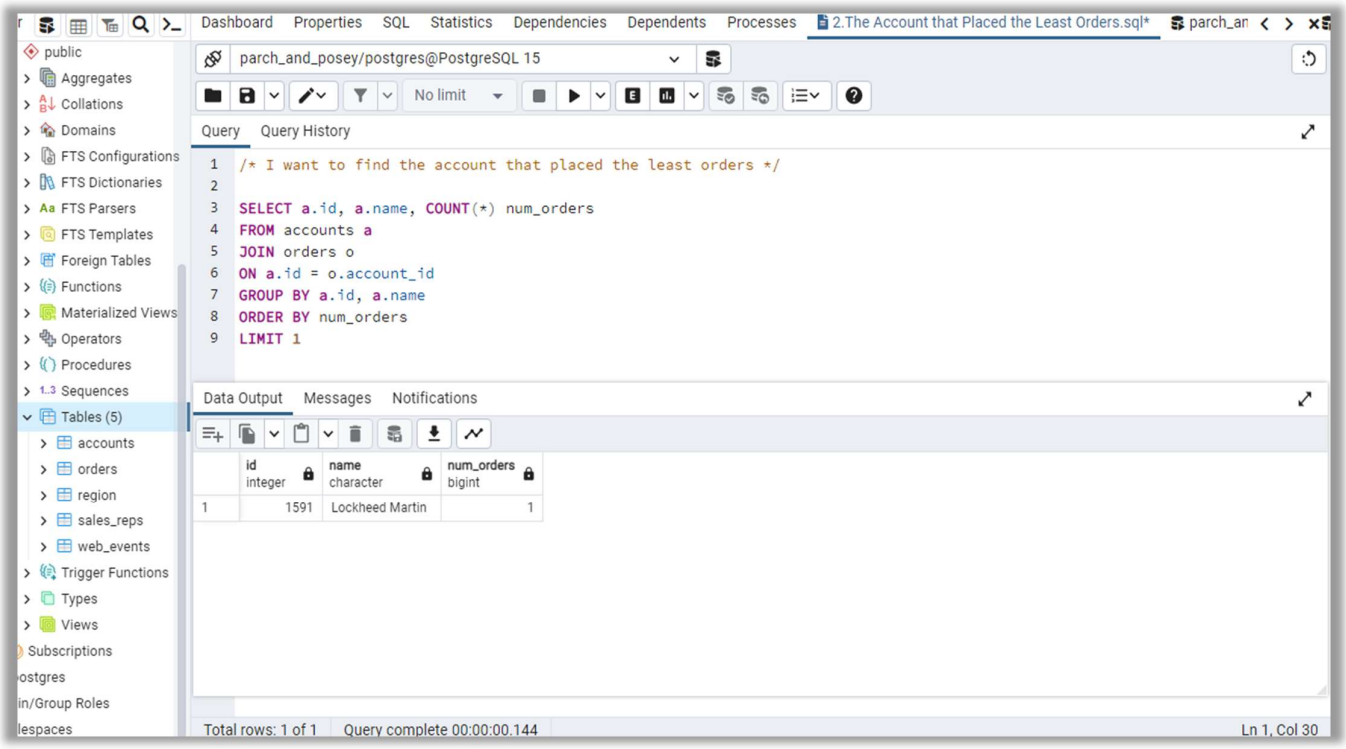
DATA ANALYSIS

1. The Account that Placed the Most Orders:



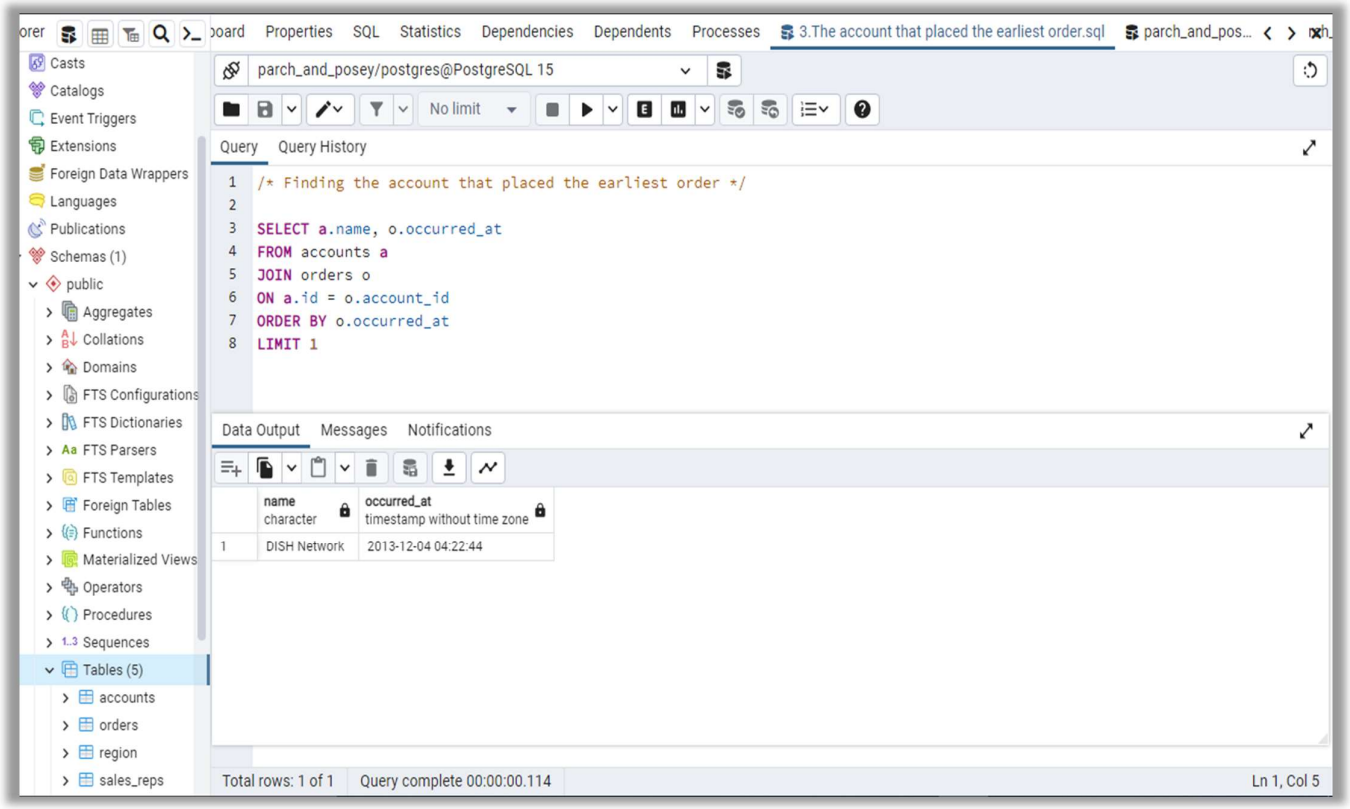
"Leucadia National" placed the most orders. This account placed 71 orders in total.

2. The Account that Placed the Least Orders:



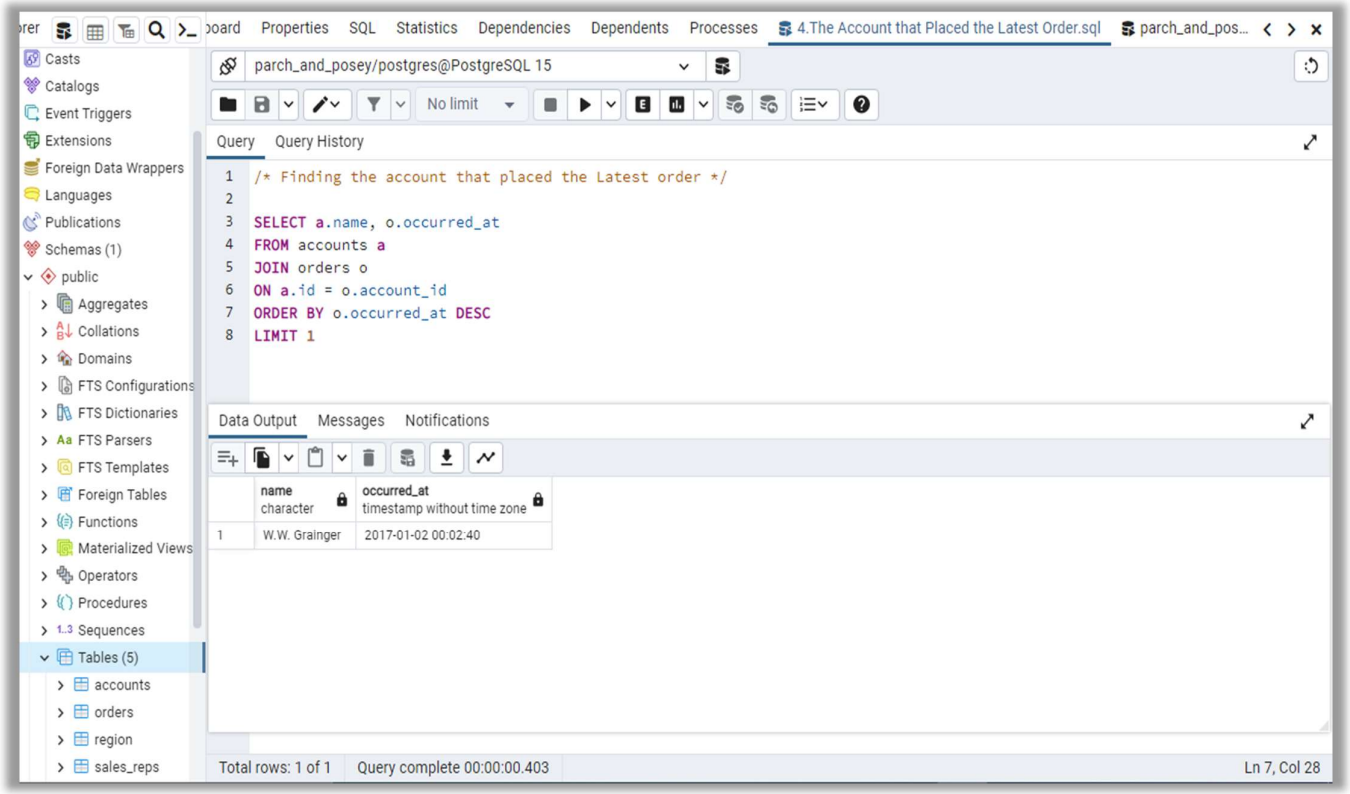
"Lockheed Martin" placed the least orders. This account placed 1 order in total.

3. The Account that Placed the Earliest Order:



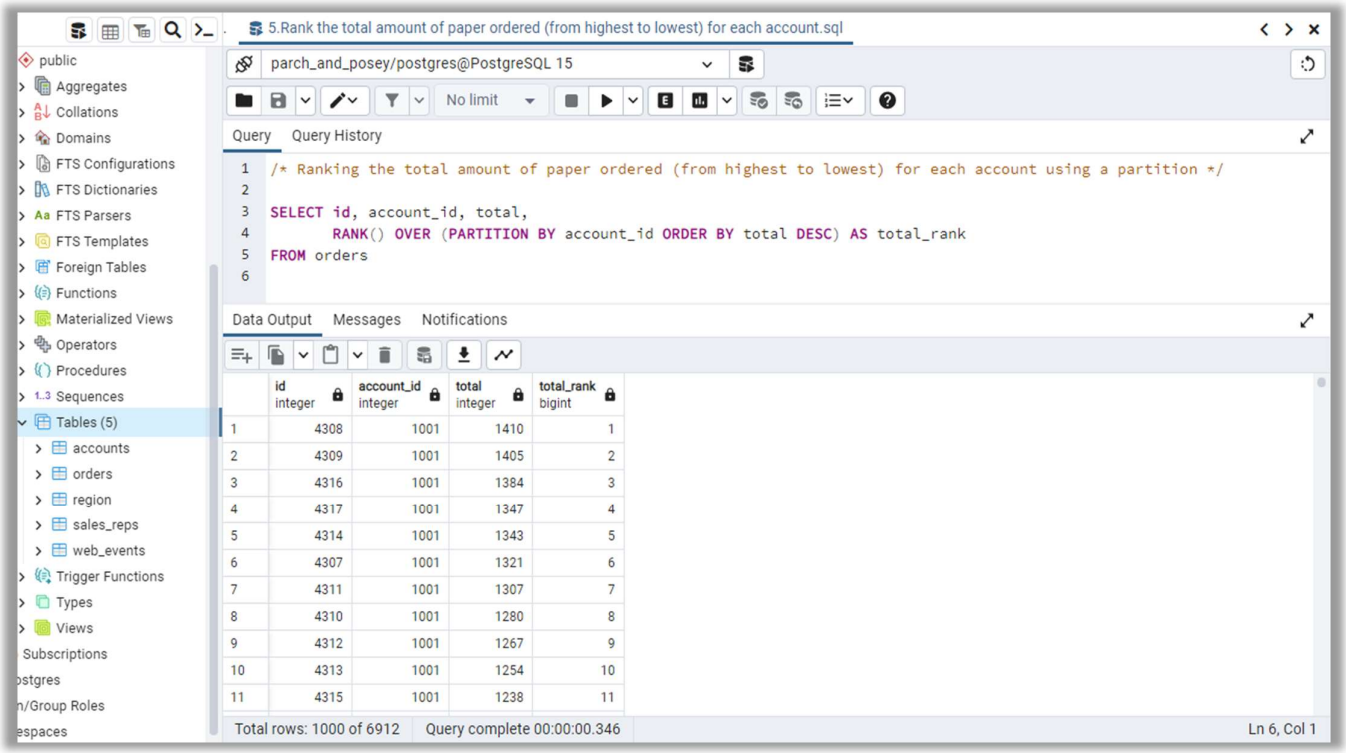
"DISH Network" placed the earliest order in **"2013-12-04 04:22:44"**

4. The Account that Placed the Latest Order:

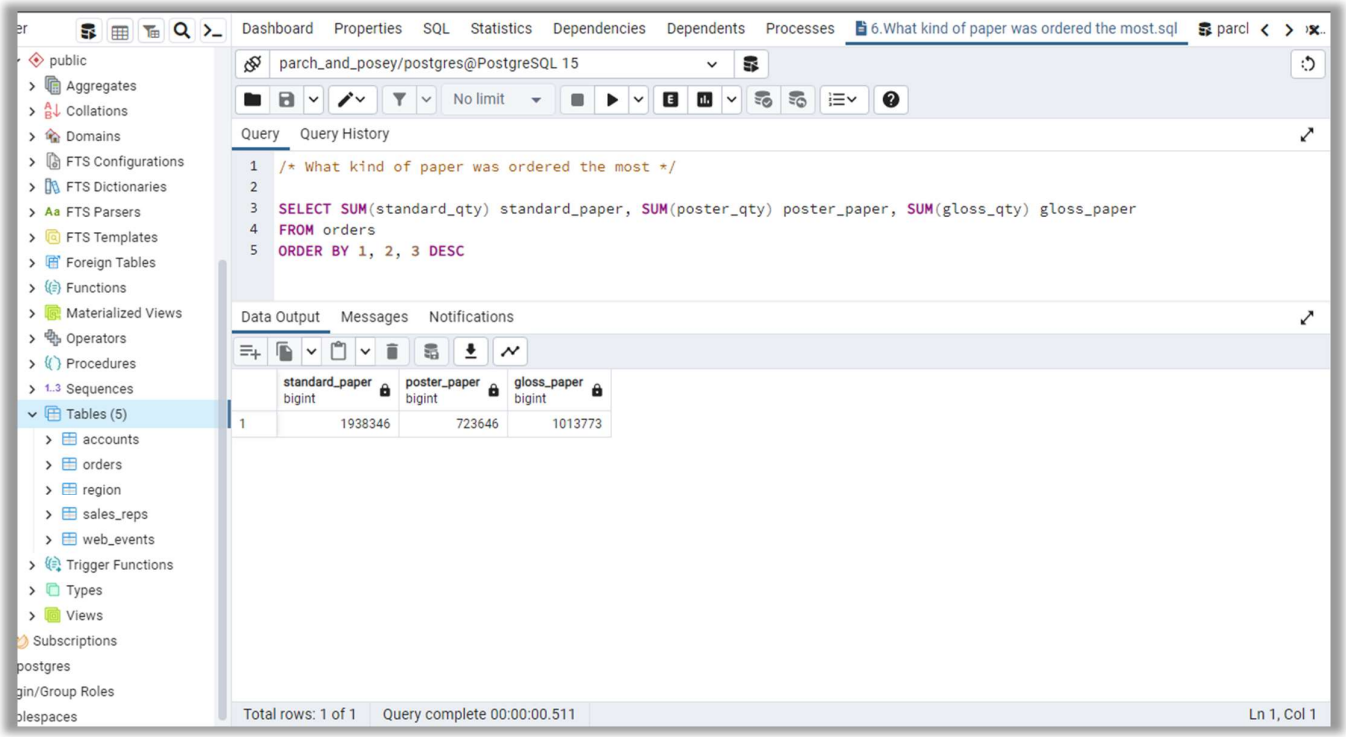


"W.W. Grainger" placed the latest order in **"2017-01-02 00:02:40"**

5. **Rank the Total Amount of Paper Ordered (From Highest to Lowest) for Each Account using a Partition:**

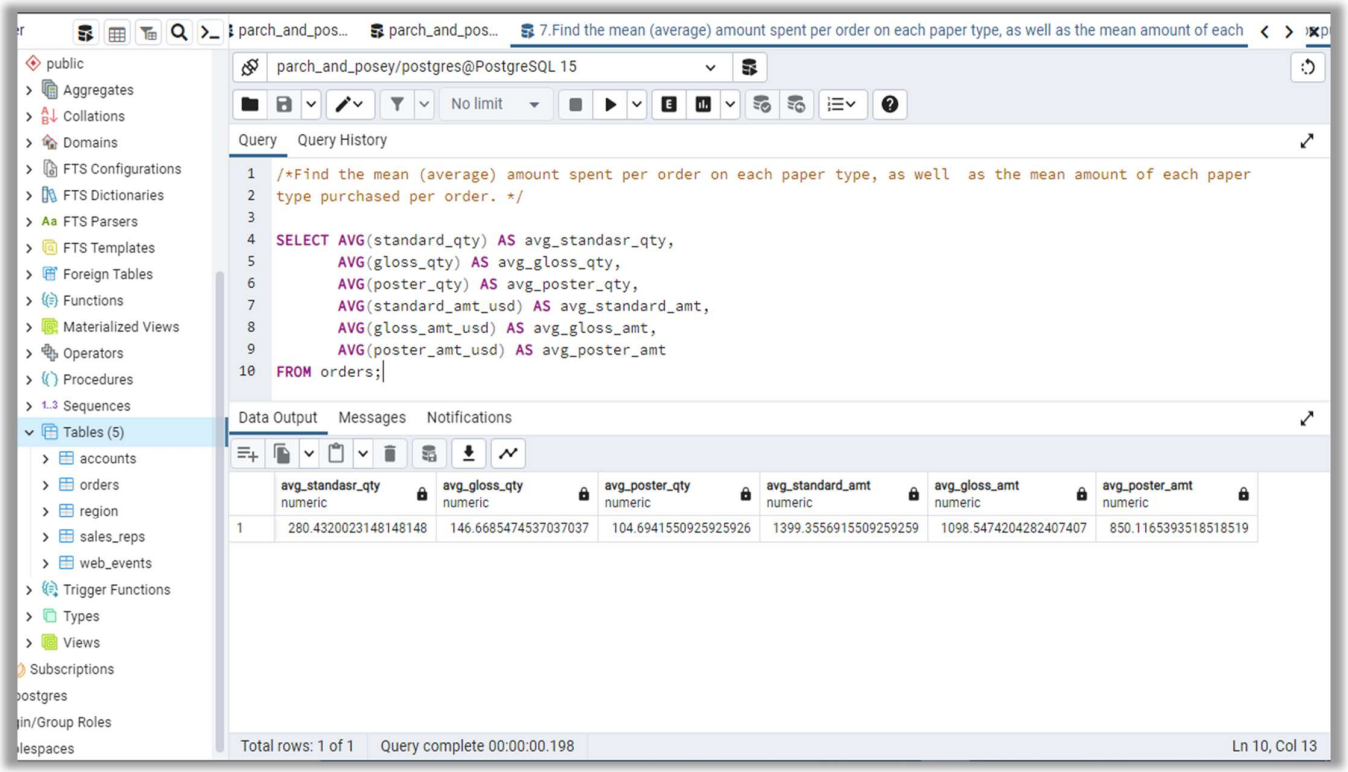


6. **What Kind of Paper was Ordered the Most:**

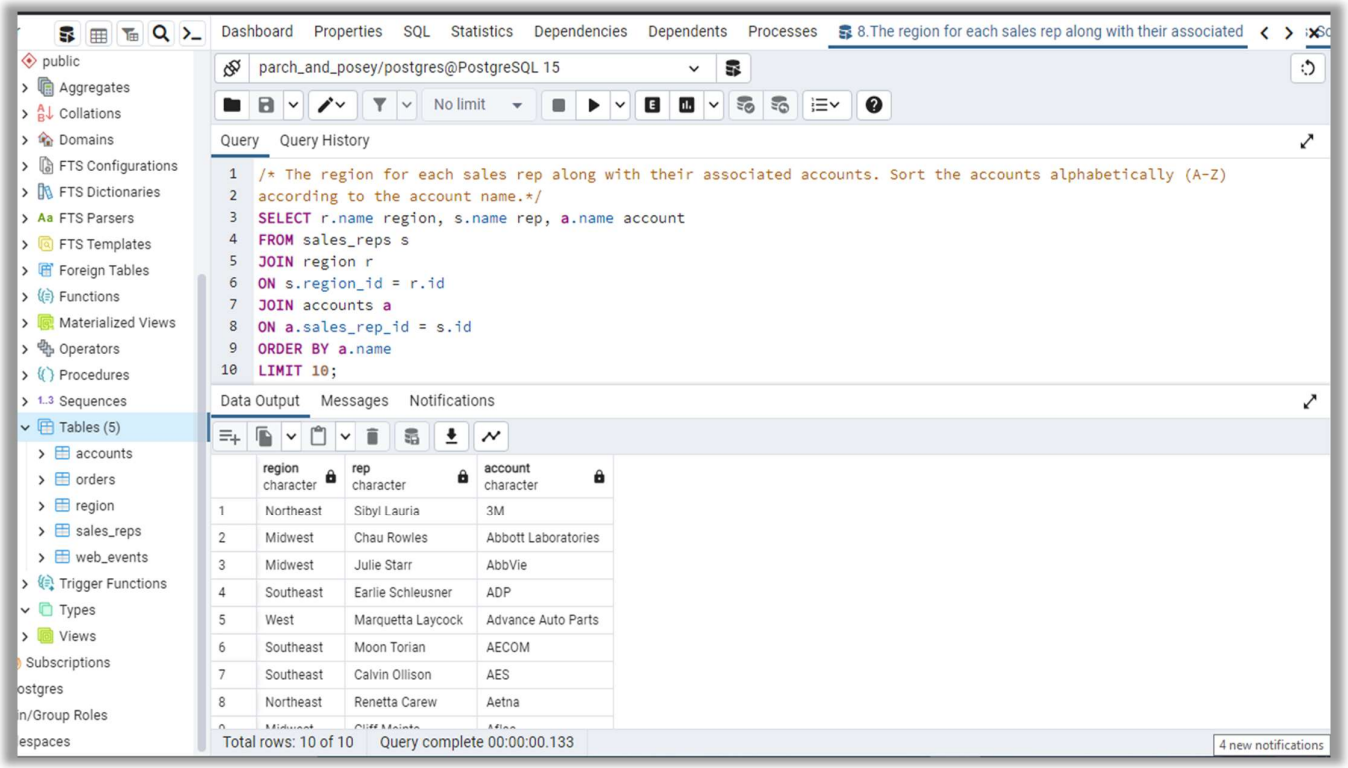


Standard paper had **1938346** orders, **Post paper** had **723646** orders, **Gloss paper** had **1013773** orders. This analysis showed that **Standard paper** had the highest orders which was one million nine hundred thirty-eight thousand three hundred forty-six.

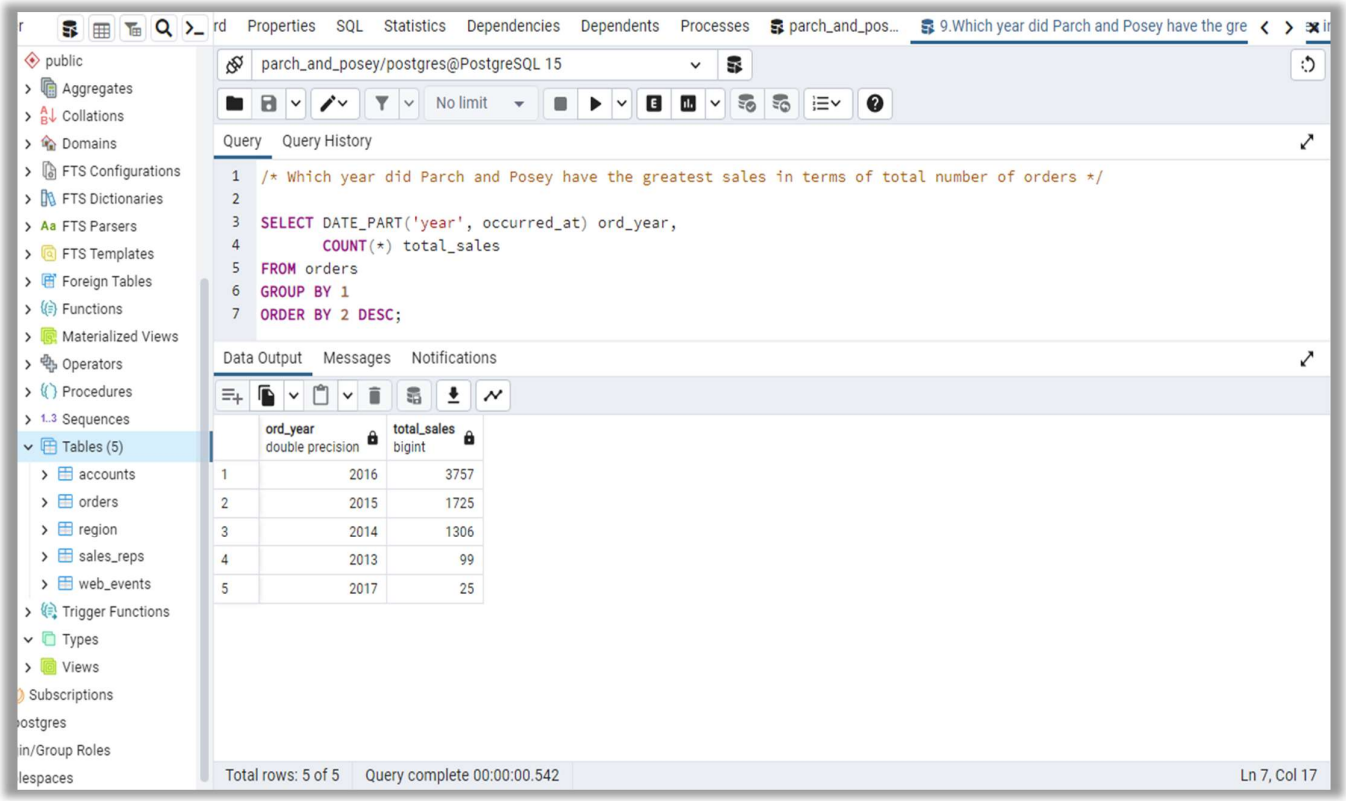
7. **Find the Mean (Average) Amount Spent Per Order on Each Paper Type, as well as the Mean Amount of Each Paper type Purchased Per Order:**



8. **The Region for Each Sales Rep Along with their Associated Accounts. Sort the Accounts Alphabetically (A-Z) According to the Account Name:**

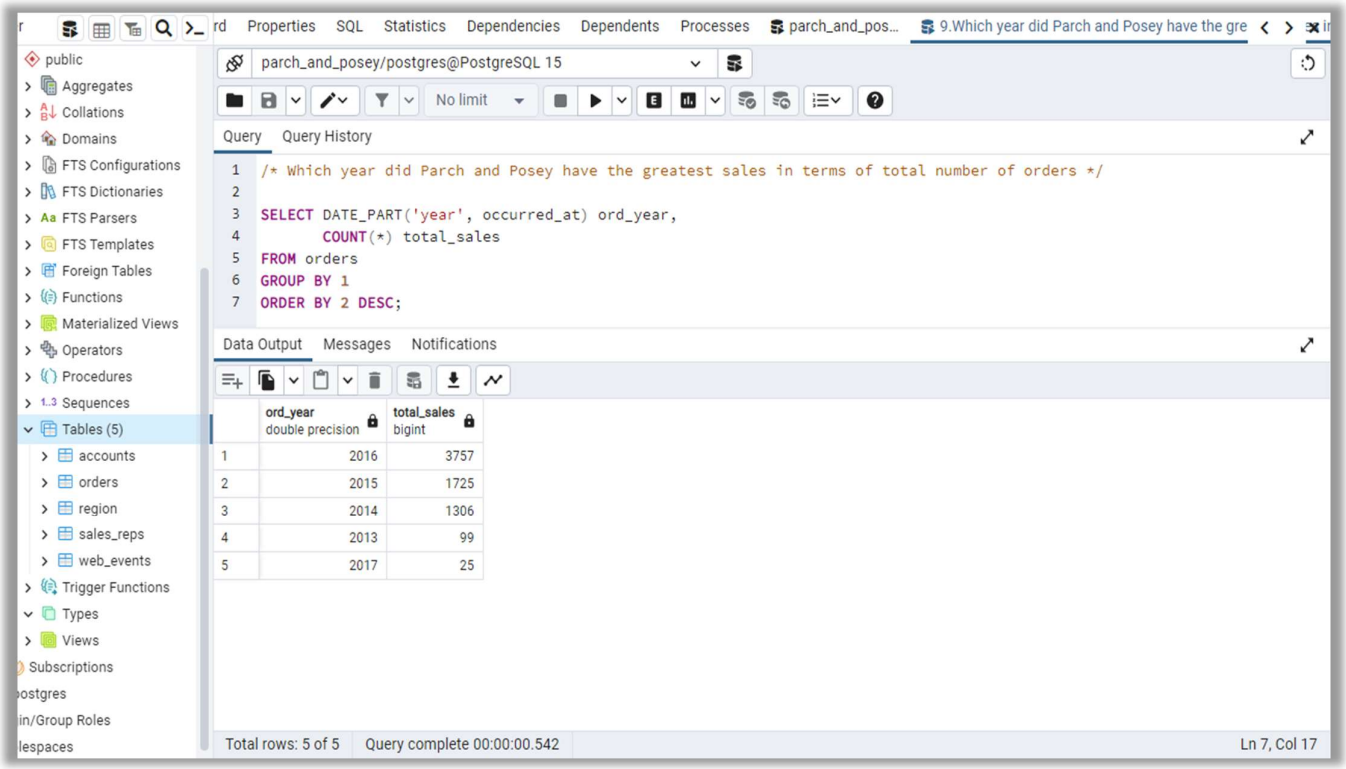


9. Which Year Did Parch and Posey have the Greatest Sales in Terms of Total Number of Orders:



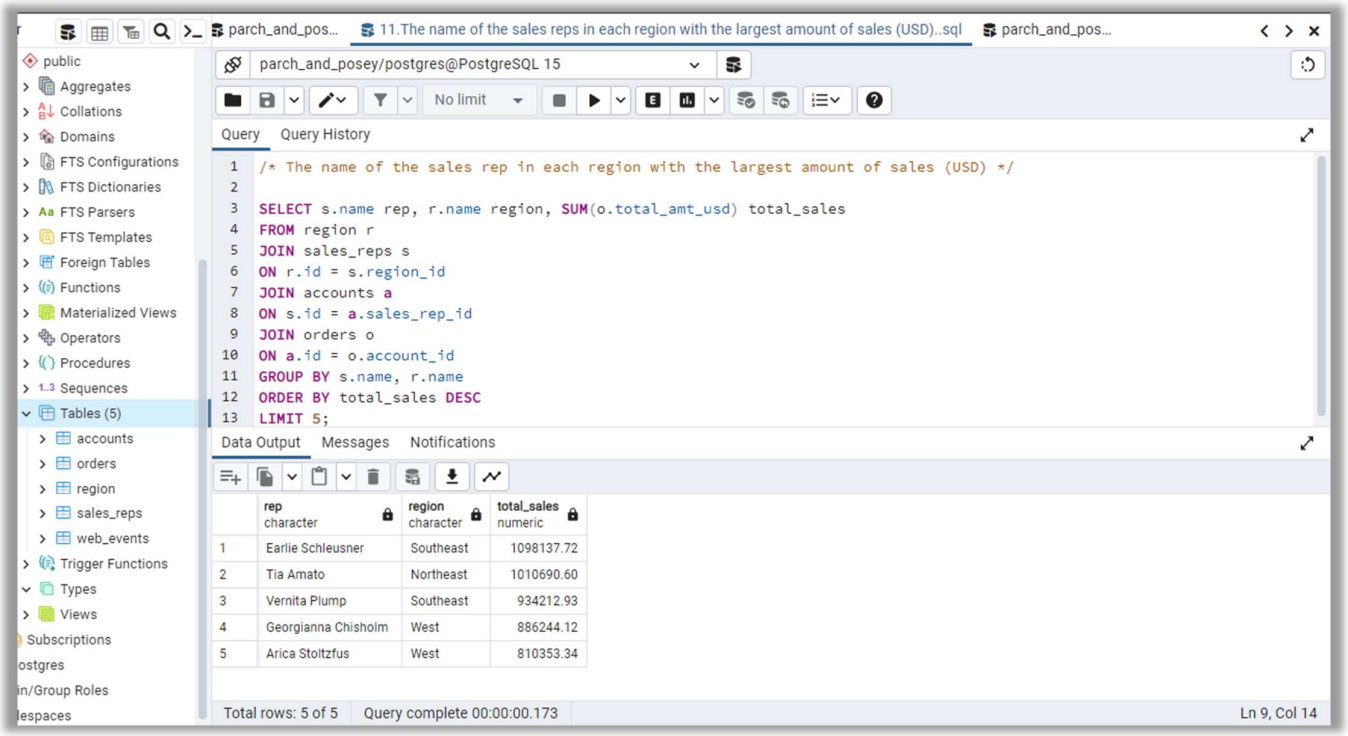
Parch and Posey had their greatest sales in **2016** with over **3757 total sales**.

10. Which Month did Parch and Posey have the Greatest Sales in Terms of Total Number of Orders:



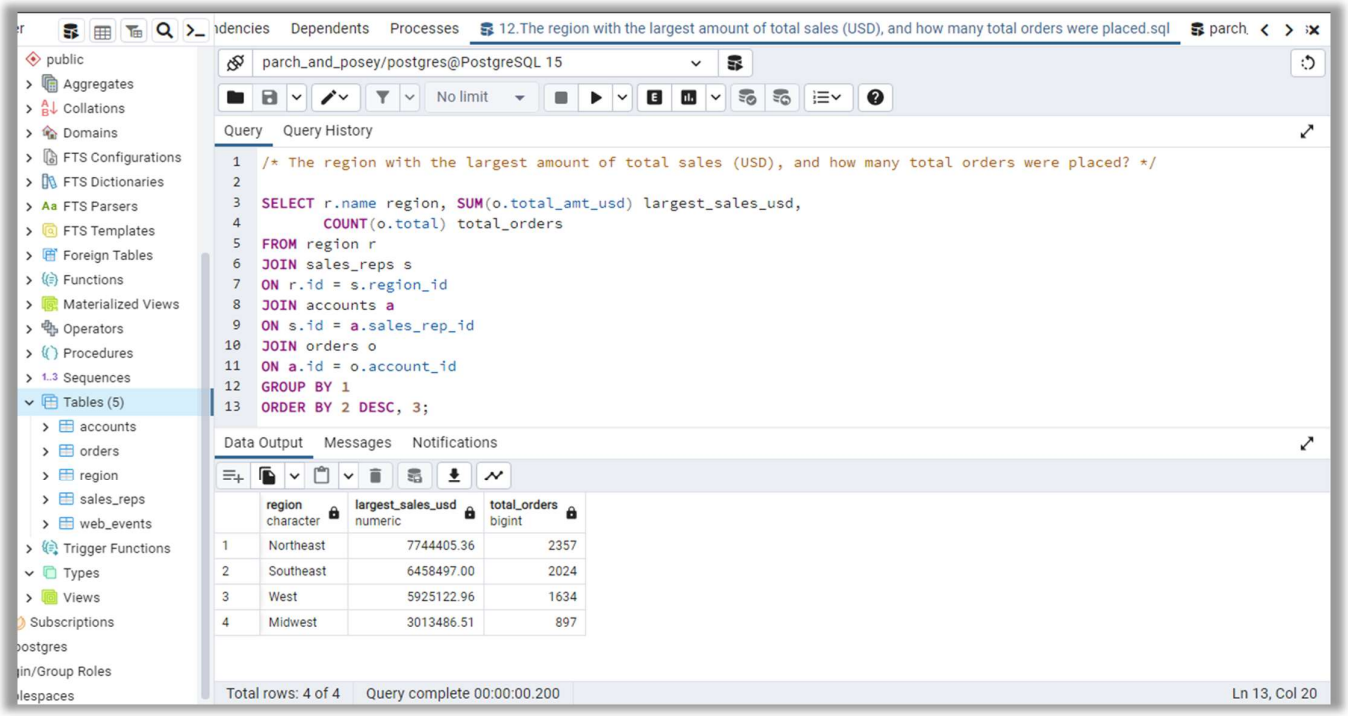
Parch and Posey had their greatest sales in the month of **December** with over **882 total sales**

11. The Names of the Sales Reps in each Region with the Largest Amount of Sales (USD):



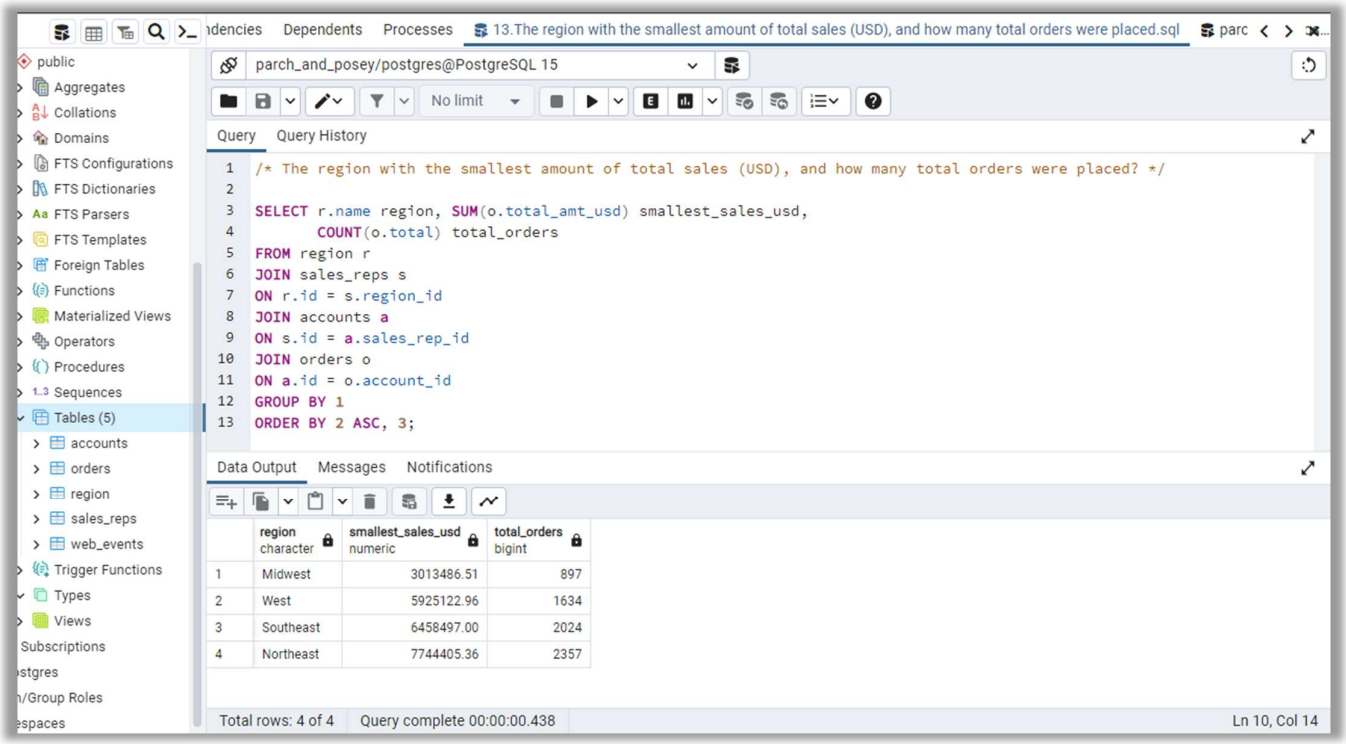
"Earlie Schleusner" from the **"Southeast"** region had the largest amount of sales **(1098137.72)**

12. The Region with the Largest Amount of Total Sales (USD), and How Many Total Orders were Placed:



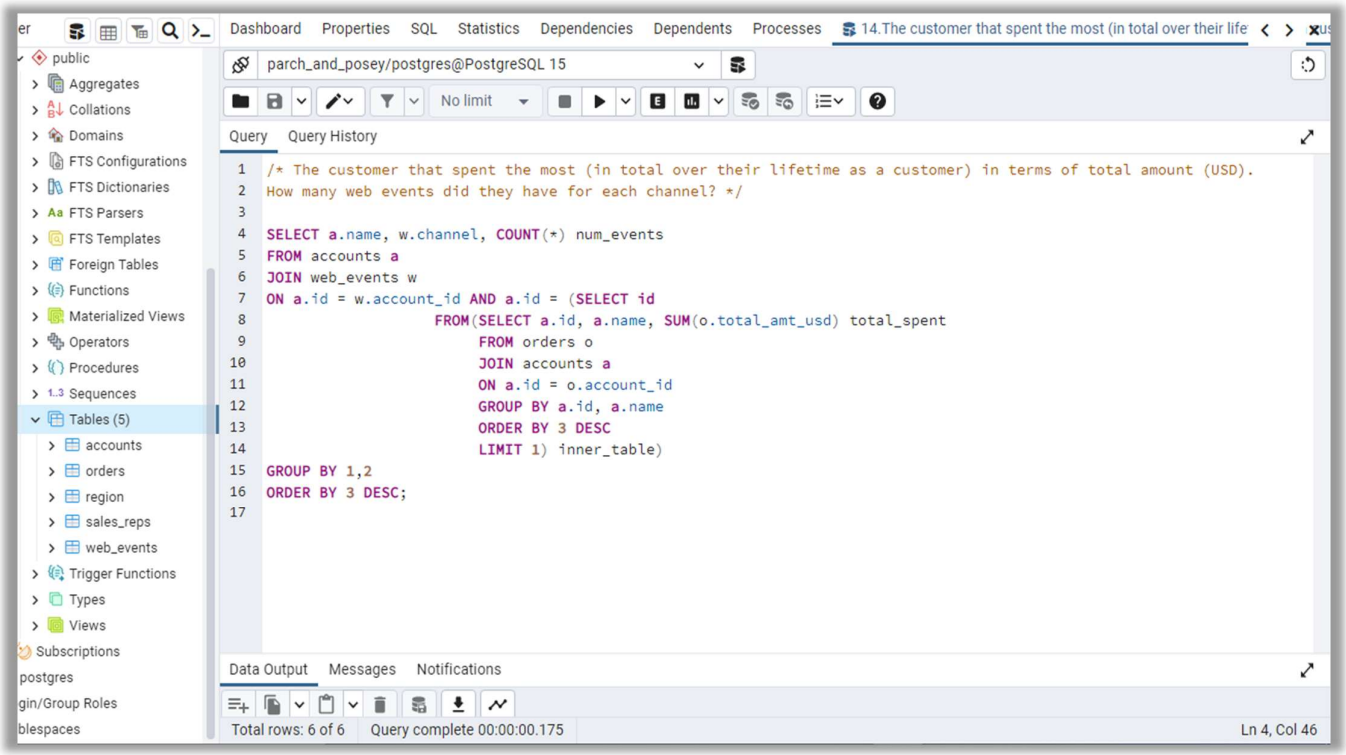
"Northeast" had the largest amount of total sales **(7744405.36)** in USD and **2357** orders were placed from that region.

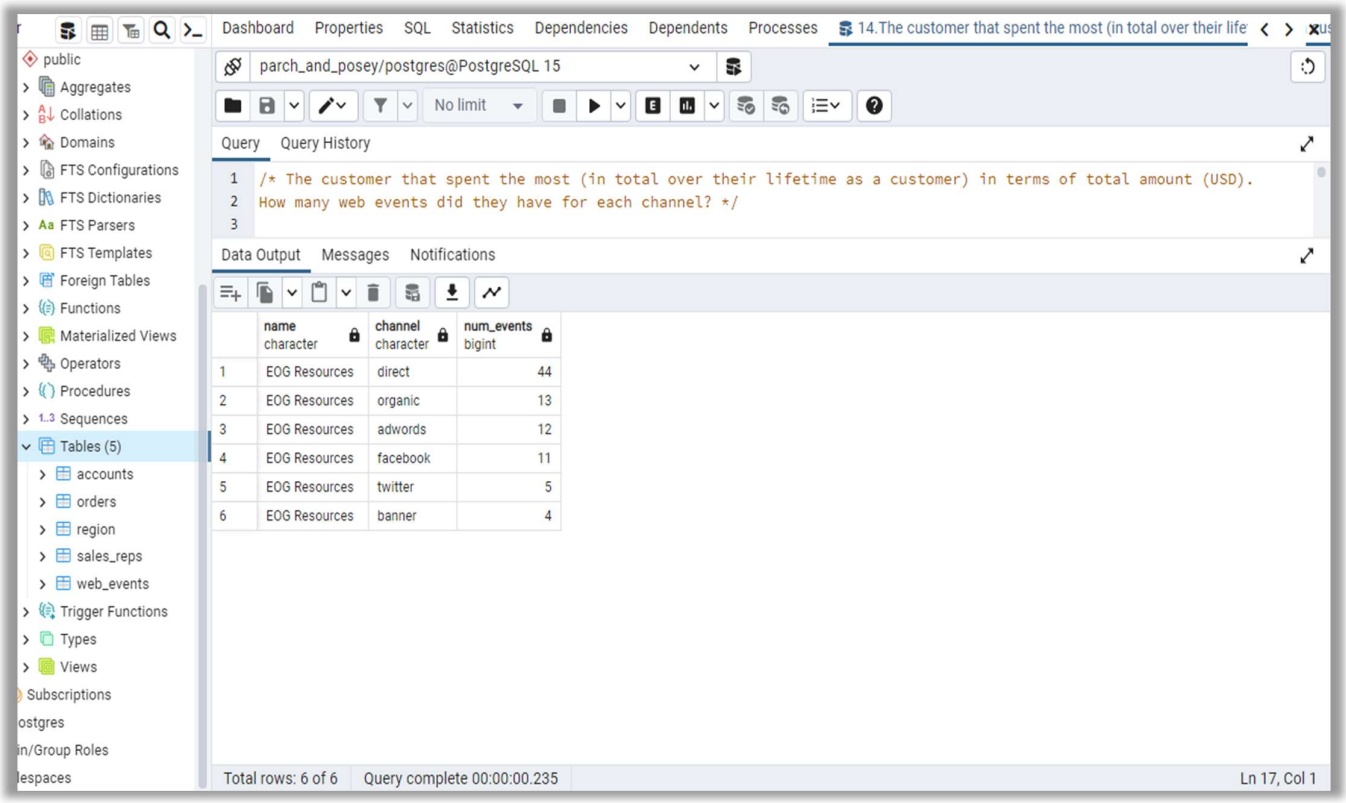
13. **The Region with the Smallest Amount of Total Sales (USD), and How Many Total Orders were Placed:**



"Midwest" had the smallest amount of total sales **(3013486.51)** in USD and **897** orders were placed from that region

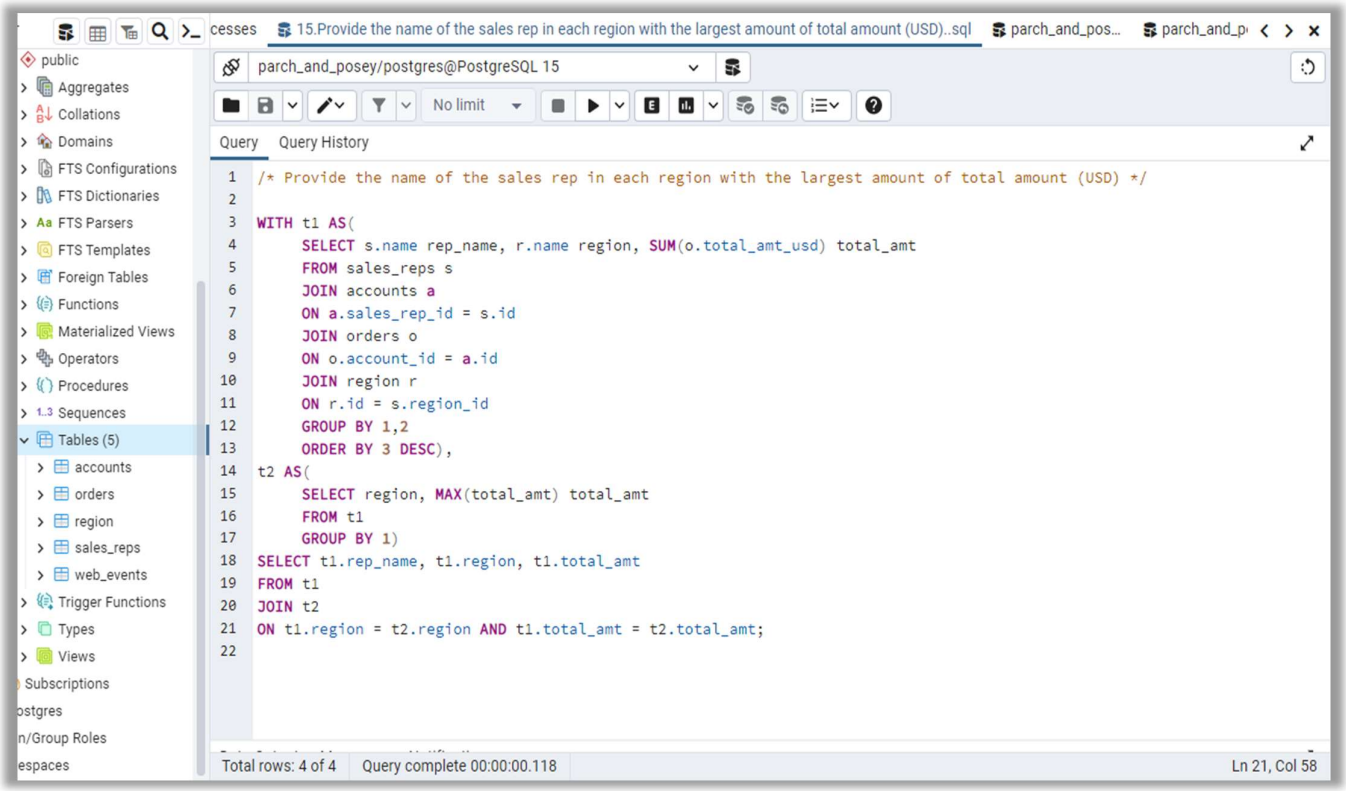
14. **The Customer that Spent the Most (In Total Over their Lifetime as a Customer) in Terms of Total Amount (USD). How Many Web Events did they Have for Each Channel:**

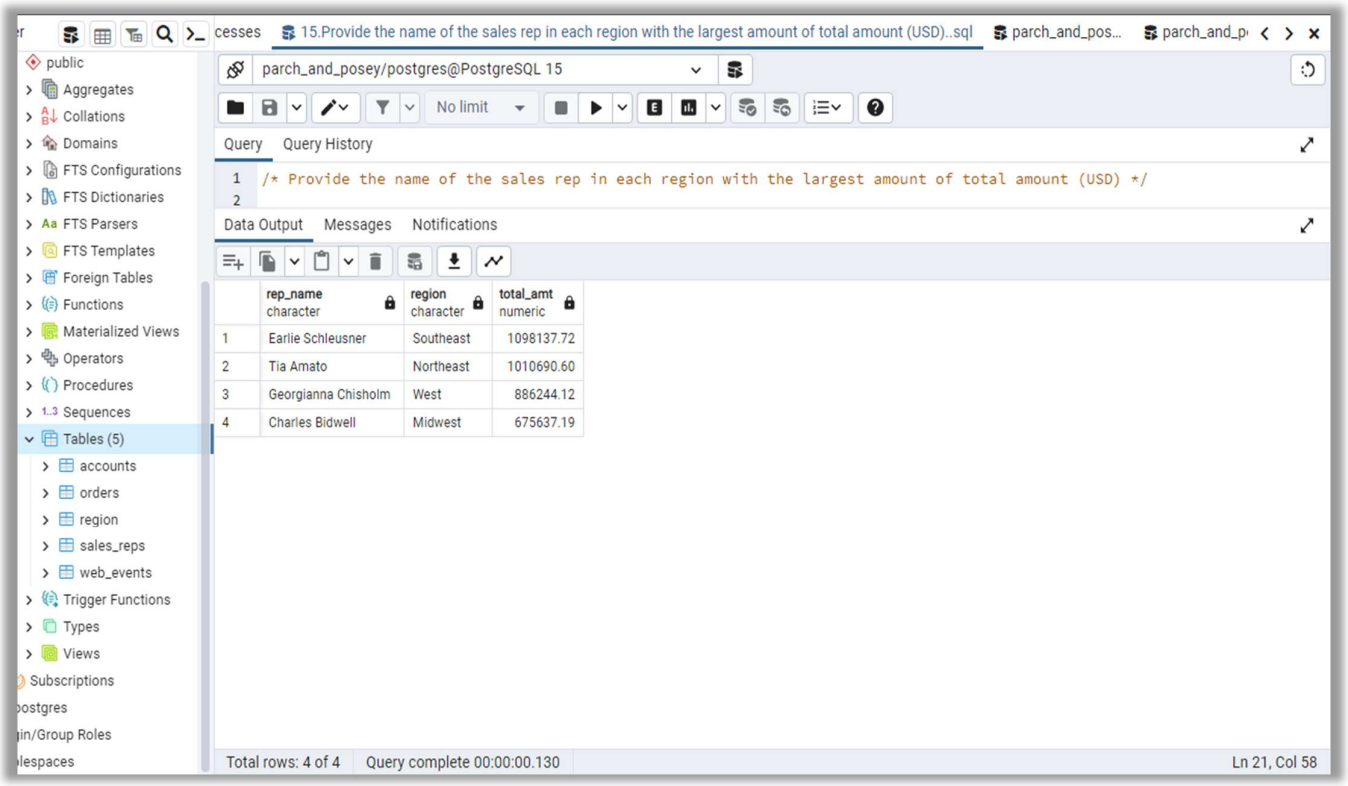




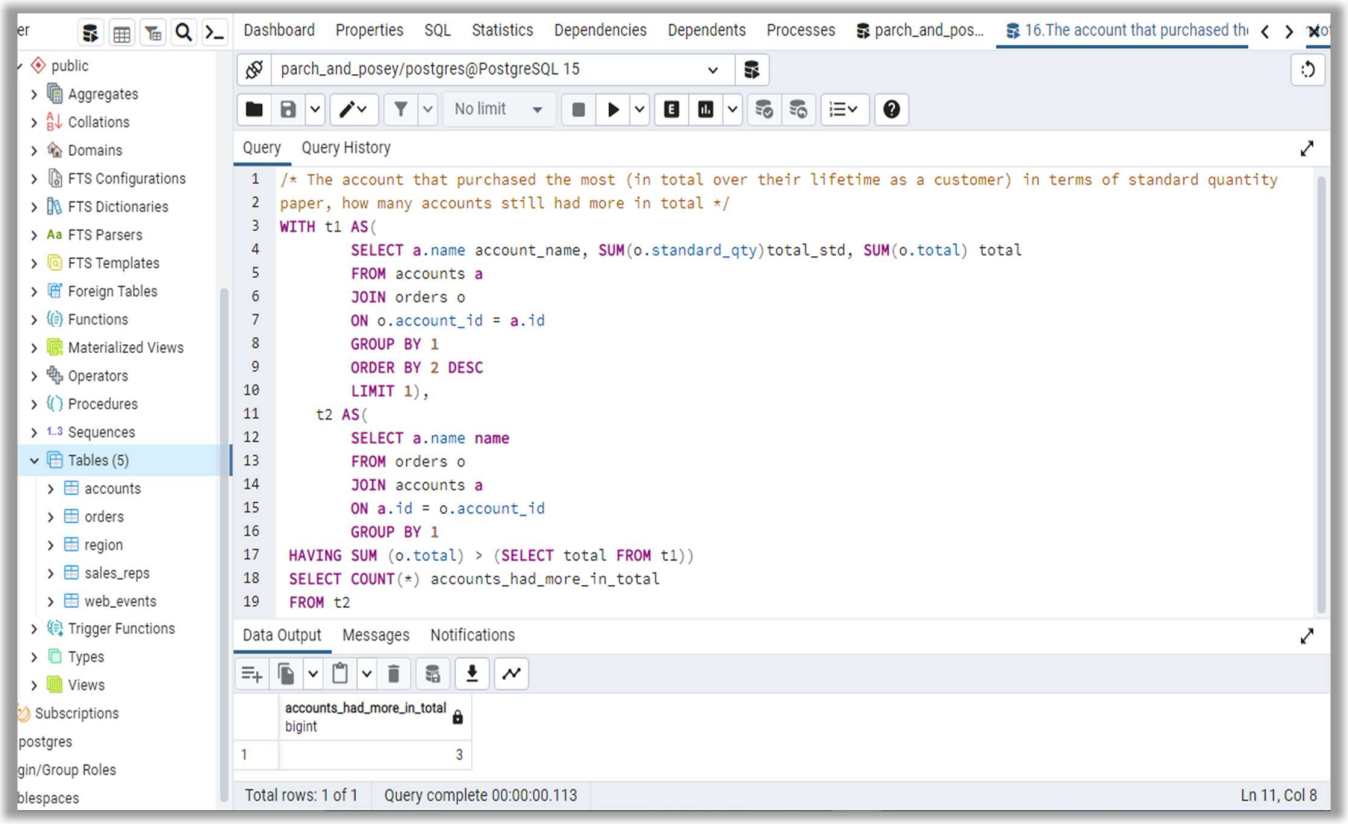
EOG Resources spent the most in total over their lifetime as a customer terms of total amount in us dollars, thy had 89 number of web events all together.

15. Provide the Name of the Sales Rep in Each Region with the Largest Amount of Total Amount (USD):





16. The Account that Purchased the Most (In Total Over their Lifetime as a Customer) in Terms of Standard Quantity Paper, How Many Accounts still had More in Total:



In terms of standard paper, 3 accounts still had more in total.

CONCLUSIONS

- In my conclusion, I found out that **Leucadia National"** placed the most orders, and this account placed 71 orders in total.
- I also found that **"Lockheed Martin"** placed the least orders, and this account placed 1 order in total.
- **DISH Network** placed the earliest order in **2013-12-04 04:22:44** and **W.W. Grainger"** placed the latest order in **2017-01-02 00:02:40.**
- **Standard paper** had **1938346** orders, **Post paper** had **723646** orders, **Gloss paper** had **1013773** orders. This analysis showed that **Standard paper** had the highest orders which was one million nine hundred thirty-eight thousand three hundred forty-six.
- I found out that Parch and Posey had their greatest sales in **2016** with over **3757 total sales** and precisely had their greatest sales in the month of **December** with over **882 total sales**
- **Earlie Schleusner** from the **"Southeast"** region had the largest amount of sales **(1098137.72)**
- **Northeast** had the largest amount of total sales **(7744405.36)** in USD and **2357** orders were placed from that region.
- **Midwest** had the smallest amount of total sales **(3013486.51)** in USD and **897** orders were placed from that region
- **EOG Resources** spent the most in total over their lifetime as a customer in terms of total amount in US dollars, they had 89 number of web events all together.
- In terms of standard paper, 3 accounts still had more in total.