

Parch and Posey EDA

Diva Awanisa Nahdi

Preface

- Parch and posey is a **fictional** poster store
- Queries are made in Microsoft **SQL Server**



Questions

- 1. What is the price of each poster type?
- 2. Which sales rep sell the most amount?
- 3. Which company has the highest total purchase amount?
- 4. Sales rep total sales for each region
- 5. Provide the name of sales rep in each region with the highest total sell
- 6. For the region with the largest (sum) of sales total_amt_usd, how many total (count) orders were placed?
- 7. Who was the Sales Rep associated with the earliest web_event?
- 8. Who was the primary contact associated with the earliest web_event?
- 9. Is there are any accounts associated with more than one region?
- 10. Have any sales reps worked on more than one account?
- 11. How many of the sales reps have more than 5 accounts that they manage?
- 12. How many accounts have more than 20 orders?
- 13. Which account used facebook most as a channel for their web events?

Special case: Provide 10 most prioritized companies for strategy development

What is the price of each poster type?

Query

select distinct std, gloss,poster
from (select (standard_amt_usd/standard_qty)
as std,
 (gloss_amt_usd/gloss_qty) as gloss,
 (poster_amt_usd/poster_qty) as poster
from orders) as yaya
where std is not NULL and gloss is not null
and poster is not null

We used distinct to make sure that the prices are the same for the same poster type (the prices do not have other unique values)

Result

	std	gloss	poster
1	499	749	812

Answer

499 dollar for standard type, 749 dollar for gloss type, and 812 dollar for poster type

Which sales rep sell the most amount?

Query

```
select top 1 s.name, sum(o.total_amt_usd) as
total_sales
from sales_reps as s
join accounts as a on s.id = a.sales_rep_id
join orders as o on a.id = o.account_id
group by (s.name)
order by sum(o.total_amt_usd) desc
```

We used top 1 to limit the number of output

Result

	name	total_sales
1	Earlie Schleusner	109813772

Answer

Earlie Schleusner

Which company has the highest total purchase amount?

Query

```
select top 1 a.name, sum(o.total_amt_usd) a
total_sales
from accounts as a
join orders as o on a.id = o.account_id
group by a.name
order by sum(o.total_amt_usd) desc
```

We used top 1 to limit the number of output

Result

	name	total_sales	
1	EOG Resources	38287330	

Answer

EOG Resources

Sales rep total sales for each region

Query

<pre>select s.name, r.name, sum(o.total_amt_usd)</pre>
as total_sales
from region as r
join sales_reps as s on r.id = s.region_id
join accounts as a on s.id = a.sales_rep_id
join orders as o on a.id = o.account_id
group by s.name, r.name
order by s.name

Result

	name	name	total_sales
1	Akilah Drinkard	Northeast	13661399
2	Arica Stoltzfus	West	81035334
3	Ayesha Monica	Northeast	21714659
4	Babette Soukup	Southeast	21590527
5	Brandie Riva	West	67591764
6	Calvin Ollison	Southeast	59451608
7	Cara Clarke	Northeast	16613865
8	Carletta Kosinski	Midwest	21303245
9	Charles Bidwell	Midwest	67563719
10	Chau Rowles	Midwest	18428260
11	Cliff Meints	Midwest	55610534

.

Provide the name of sales rep in each region with the highest total sell

Query

```
select region, sales, total sales
from (
     select r.name as region, s.name as
     sales,sum(o.total amt usd) as
     total sales, ROW NUMBER()
     over(partition by r.name order by
     sum(o.total amt usd) desc) as row num
     from region as r
     join sales reps as s on r.id =
     s.region id
     join accounts as a on s.id =
     a.sales rep id
     join orders as o on a.id =
     o.account id
     group by r.name, s.name) as Tabell
where row num = 1
order by total sales desc
```

Result

	region	sales	total_sales
1	Southeast	Earlie Schleusner	109813772
2	Northeast	Tia Amato	101069060
3	West	Georgianna Chi	88624412
4	Midwest	Charles Bidwell	67563719

For the region with the largest (sum) of sales total_amt_usd, how many total (count) orders were placed?

select count(*) as order count from region as r join sales reps as s on r.id = s.region id join accounts as a on s.id = a.sales rep id join orders as o on a.id = o.account id from region as r join sales reps as s on r.id = s.region id join accounts as a on s.id = a.sales rep id join orders as o on a.id = o.account id group by r.name

Query

Where subquery to provide the region with the largest total sales



9

Who was the Sales Rep associated with the earliest web_event?

select s.name from accounts as a join web_events as w on a.id = w.account_id join sales_reps as s on a.sales_rep_id = s.id where w.occurred_at in (select min(w.occurred_at) from accounts as a join web_events as w on a.id = w.account_id join sales_reps as s on a.sales_rep_id = s.id)

Query

Where query to provide the earliest web event



Who was the primary contact associated with the earliest web_event?



Where query to provide the earliest web event



Is there are any accounts associated with more than one region?

Query

```
select company, region
from (select a.name as company, r.name as
region, ROW_NUMBER() over(partition by
a.name order by r.name) as row_num
    from region as r
    join sales_reps as s on r.id =
s.region_id
    join accounts as a on a.sales_rep_id =
s.id
    group by a.name, r.name) as tabellll
where row_num > 1
```

Row number is used to mark the number of region each account had

Result

company	region	

Answer

No, each account is only associated with one region

Have any sales reps worked on more than one account?

Query

select count(distinct sales) as
num_of_salesrep
from (select s.name as sales, a.name as
company, ROW_NUMBER() over(partition by
s.name order by a.name) as row_num
from accounts as a
join sales_reps as s on a.sales_rep_id =
s.id
group by s.name, a.name) as tabellll
where row_num > 1

Row number is used to mark the number of account each sales rep had. Distinct sales is used to fetch unique sales rep names

Result (subquery)

	sales	company	row_num
1	Akilah Drinkard	Comcast	1
2	Akilah Drinkard	Fannie Mae	2
3	Akilah Drinkard	FedEx	3
4	Arica Stoltzfus	Applied Materi	1
5	Arica Stoltzfus	Assurant	2

Result



Answer

Yes, there are 50 sales rep that worked on more than one account

How many of the sales reps have more than 5 accounts that they manage?

Query

Select count (distinct sales) as
num_of_salesrep
from (select s.name as sales, a.name as
company, ROW_NUMBER() over(partition by
s.name order by a.name) as row_num
from accounts as a
join sales_reps as s on a.sales_rep_id =
s.id
group by s.name, a.name) as tabellll
where row_num > 5

Row number is used to mark the number of account each sales rep had. Distinct sales is used to fetch unique sales rep names

Result (subquery)

	sales	company	row_num
1	Akilah Drinkard	Comcast	1
2	Akilah Drinkard	Fannie Mae	2
3	Akilah Drinkard	FedEx	3
4	Arica Stoltzfus	Applied Materi	1
5	Arica Stoltzfus	Assurant	2

Result



Answer

There are 34 sales rep that have more than 5 accounts to manage

How many accounts have more than 20 orders?

Query

```
select count(*) as account_num
from (select a.name, count(o.id) as
order_count
   from accounts as a
   join orders as o on a.id = o.account_id
   group by a.name
   having count(o.id) > 20) as tabellll
```

From query to fetch accounts that have more than 20 orders

Result



Answer

There are 120 accounts that have more than 20 orders

Which account used facebook most as a channel for their web events?

select top 1 a.name, count(*) as count from accounts as a join web_events as w on w.account_id = a.id where w.channel = 'facebook' group by a.name order by count(*) desc

Query

We used top 1 to limit the number of output

namecount1Gilead Sciences16

Result

Answer

Gilead sciences

With assumption that Parch and Posey (P&P) does not have specific target market, we will identify the prioritized companies by their Customer Lifetime Value (CLV). CLV is a common metric that can measure how profitable the customer is in current time or in the future.

According to Gupta (2006), CLV can be measured by three factors:

- 1. **Recency** : How recent the customer purchased the products
- 2. Frequency : The frequency of customers' order
- 3. Monetary : The total amount of money the customer spent in the company

Then, each factor is weighted subjectively according to the stakeholder requirements. The following is the formula to measure customers' CLV

CLV = weight * recency + weight * frequency + weight* monetary

1. Recency

For limitation, the recency calculation will be based on the latest date the orders had been made in the dataset. The recency data is then calculated by identifying the date difference between the max date on the whole dataset with the latest date each company had ordered in P&P.

2. Frequency

The frequency data will be identified by counting the number of unique order id for each company

3. Monetary

The monetary data will be identified by calculating the total of purchase amount that each company had made at all time

First thing first, we are making a CTE that will fetch the name of the company, their order frequency, their total monetary value, and the latest date they made an order

with group_1 as (Select a.name as company, count(o.id) as frequency,sum(total_amt_usd) as monetary, max(occurred_at) as max_date from accounts as a join orders as o on a.id = o.account_id group by a.name),

Result (inside the subquery)

	company	frequency	monetary	max_date
1	3 M	28	12794510	2016-09-23T03
2	Abbott Laborat	30	9681992	2015-07-14T16
3	AbbVie	3	1124363	2016-12-13T04
4	ADP	60	16357918	2016-12-20T22
5	Advance Auto	4	696207	2016-12-11T00
6	AECOM	9	1849151	2016-12-17T08
7	AES	3	1303864	2016-12-09T15

Then, we are making the second CTE to standardize the frequency and monetary value. The recency column is made of the difference between the max date of all order with the max date for each company

group_2 as (select company,

(cast(frequency as float)/cast((select sum(frequency) from group_1) as float)) as frequency, (cast(monetary as float)/cast((select sum(monetary) from

group_1) as float() as monetary,

-1 * DATEDIFF(day,(select max(max_date) from group_1), max_date) as recency

from group_1),

Result (inside the subquery)

	company	frequency	monetary	recency
1	3M	0,00405092592	0,00552881336	101
2	Abbott Laborat	0,00434027777	0,00418381999	538
3	AbbVie	0,00043402777	0,00048586410	20
4	ADP	0,00868055555	0,00706864707	13
5	Advance Auto	0,00057870370	0,00030084767	22
6	AECOM	0,00130208333	0,00079906231	16
7	AES	0,00043402777	0,00056343077	24

The last step is to count the CLV. It is assumed that weight for recency is 0.2, frequency is 0.3, and monetary is 0,6. The monetary value is standardized first (in from subquery)

select top 10 company, (recency * 0.2 + frequency * 0.3 +
monetary * 0.6) as CLV
from (select company,
cast(recency as float)/cast((select sum(recency) from
group_2)as float) as recency,
frequency, monetary from group_2) as tabellll
order by CLV desc

Result

	company	CLV
1	EOG Resources	0,01339215008
2	Mosaic	0,01192928120
3	Republic Servic	0,01147597591
4	IBM	0,01110540886
5	Leucadia Natio	0,01080744332
6	General Dynam	0,01078526138
7	Sysco	0,01032622615
8	Arrow Electroni	0,01024937483
9	Supervalu	0,01010273386
10	Archer Daniels	0,01007946314