greatlearning

Learning for Life

SQL MINI PROJECT

OUR GROUP

1] AMEY SHENDAGE

2] RUSHIKESH GOSAVI

3] ABHIJEET PANSALKAR

4] AARTI KADU

5] NIRAJ CHAUDHARI

QUESTIONS

1. Join all the tables and create a new table called combined_table. (market_fact, cust_dimen, orders_dimen, prod_dimen, shipping_dimen)
2. Find the top 3 customers who have the maximum number of orders
3. Create a new column DaysTakenForDelivery that contains the date difference of Order_Date and Ship_Date.
4. Find the customer whose order took the maximum time to get delivered.
5. Retrieve total sales made by each product from the data (use Windows function)
6. Retrieve total profit made from each product from the data (use windows function)
7. Count the total number of unique customers in January and how many of them came back every month over the entire year in 2011
8. Retrieve month-by-month customer retention rate since the start of the business.(using views) Tips: #1: Create a view where each user's visits are logged by month, allowing for

Q1.Join all the tables and create a new table called combined_table. (market_fact, cust_dimen, orders_dimen, prod_dimen, shipping_dimen)

```
create table combined_table as (
select m.*,c.Customer_Name,c.Province, c.Region, c.Customer_Segment,
o.Order_ID, o.Order_Date,o.Order_Priority,
p.Product_Category,p.Product_Sub_Category,
s.Ship_Mode,s.ship_Date
from cust_dimen c
join market_fact m using(cust_id)
join orders_dimen o using(ord_id)
join prod_dimen p using(prod_id)
join shipping_dimen s using(ship_id));
```

Prod_id	Cust_id	Ord_id	Ship_id	Order_ID	Order_Date	Order_Priority	Ship_Mode	Ship_Date	Sales	Discount	Order_Quantity	Profit	Shipping_Cost	Product_Base_Margin	Customer_Name	Province
Prod_2	Cust_2	Ord_2	SHP_2	293	01-10-2012	HIGH	DELIVERY TRUCK	02-10-2012	10123.02	0.07	49	457.81	68.02	0.58	BARRY FRENCH	NUNAVUT
Prod_3	Cust_2	Ord_2	SHP_3	293	01-10-2012	HIGH	REGULAR AIR	03-10-2012	244.57	0.01	27	46.71	2.99	0.39	BARRY FRENCH	NUNAVUT
Prod_4	Cust_3	Ord_3	SHP_4	483	10-07-2011	HIGH	REGULAR AIR	12-07-2011	4965.7595	0.08	30	1198.97	3.99	0.58	CLAY ROZENDAL	NUNAVUT
Prod_5	Cust_4	Ord_4	SHP_5	515	28-08-2010	NOT SPECIFIED	REGULAR AIR	30-08-2010	146.69	0.05	21	4.43	4.95	0.37	CARLOS SOLTERO	NUNAVUT
Prod_2	Cust_4	Ord_4	SHP_5	515	28-08-2010	NOT SPECIFIED	REGULAR AIR	30-08-2010	394.27	0.08	19	30.94	5.94	0.5	CARLOS SOLTERO	NUNAVUT
Prod_3	Cust_5	Ord_5	SHP_6	613	17-06-2011	HIGH	REGULAR AIR	17-06-2011	93.54	0.03	12	-54.04	7.72	0.38	CARL JACKSON	NUNAVUT
Prod_2	Cust_9	Ord_9	SHP_11	868	08-06-2012	NOT SPECIFIED	REGULAR AIR	09-06-2012	716.84	0	32	134.72	5.94	0.5	CARLOS DALY	NUNAVUT
Prod_8	Cust_9	Ord_9	SHP_12	868	08-06-2012	NOT SPECIFIED	REGULAR AIR	10-06-2012	1474.33	0.04	31	114.46	3.61	0.71	CARLOS DALY	NUNAVUT
Prod_3	Cust_10	Ord_10	SHP_13	933	04-08-2012	NOT SPECIFIED	REGULAR AIR	04-08-2012	80.61	0.02	15	-4.72	2.99	0.37	CLAUDIA MINER	NUNAVUT
Prod_9	Cust_11	Ord_12	SHP_15	998	25-11-2009	NOT SPECIFIED	REGULAR AIR	26-11-2009	248.26	0.07	16	93.8	1.39	0.4	ALLEN ROSENBLATT	NUNAVUT

Q2.Find the top 3 customers who have the maximum number of orders

```
select* from combined_table;
select * from (select *,dense_rank() over(order by orders desc) rnk
 from (select Customer_Name, count(distinct Order_ID) orders from combined_table group by Customer_Name) t) tem where rnk<=3;
```

cust_id	Customer_Name	orders	rnk
Cust_1140	PATRICK JONES	30	1
Cust_1329	JONATHAN DOHERTY	21	2
Cust_444	BILL DONATELLI	21	2
Cust_572	LENA CREIGHTON	21	2
Cust_942	DENNIS KANE	20	3
Cust_999	SALLY HUGHSBY	20	3
Cust_1337	DENISE MONTON	20	3
Cust_1445	ED BRAXTON	20	3
Cust_1799	RAYMOND BOOK	20	3

Q3.Create a new column DaysTakenForDelivery that contains the date difference of Order_Date and Ship_Date.

```
select*, datediff(str_to_date(Ship_Date,'%d-%m-%Y'),str_to_date(Order_Date,'%d-%m-%Y'))
DaysTakenForDelivery
 from orders_dimen o join shipping_dimen s
on o.Order_ID=s.Order_ID;
```

Order_ID	Order_Date	Order_Priority	Ord_id	Order_ID	Ship_Mode	Ship_Date	Ship_id	DaysTakenForDelivery
3	13-10-2010	LOW	Ord_1	3	REGULAR AIR	20-10-2010	SHP_1	7
293	01-10-2012	HIGH	Ord_2	293	DELIVERY TRUCK	02-10-2012	SHP_2	1
293	01-10-2012	HIGH	Ord_2	293	REGULAR AIR	03-10-2012	SHP_3	2
483	10-07-2011	HIGH	Ord_3	483	REGULAR AIR	12-07-2011	SHP_4	2
515	28-08-2010	NOT SPECIFIED	Ord_4	515	REGULAR AIR	30-08-2010	SHP_5	2
613	17-06-2011	HIGH	Ord_5	613	REGULAR AIR	17-06-2011	SHP_6	0
613	17-06-2011	HIGH	Ord_5	613	REGULAR AIR	18-06-2011	SHP_7	1
643	24-03-2011	HIGH	Ord_6	643	EXPRESS AIR	25-03-2011	SHP_8	1
678	26-02-2010	LOW	Ord_7	678	REGULAR AIR	26-02-2010	SHP_9	0
807	23-11-2010	MEDIUM	Ord_8	807	REGULAR AIR	24-11-2010	SHP_10	1
868	08-06-2012	NOT SPECIFIED	Ord_9	868	REGULAR AIR	09-06-2012	SHP_11	1
868	08-06-2012	NOT SPECIFIED	Ord_9	868	REGULAR AIR	10-06-2012	SHP_12	2

Q4.Find the customer whose order took the maximum time to get delivered.

```
select *from(
select c.Customer_Name,c.Cust_id,s.Order_ID, datediff(str_to_date(Ship_Date,'%d-%m-%Y'),str_to_date(Order_Date,'%d-%m-%Y'))
DaysTakenForDelivery
 from orders_dimen o join shipping_dimen s join market_fact m on m.Ship_id=s.Ship_id
join cust_dimen c on m.Cust_id=c.Cust_id
on o.Order_ID=s.Order_ID)temp
order by DaysTakenForDelivery desc limit 1 ;
```

Customer_Name	Cust_id	Order_ID	DaysTakenForDelivery
DEAN PERCER	Cust_1460	353	92

Q5.Retrieve total sales made by each product from the data (use Windows function)

```
select distinct Prod_id,total_sale from(
select*, sum(sales)over(partition by prod_id)total_sale from market_fact)temp;
```

Prod_id	total_sale
Prod_1	1028240.7600000008
Prod_10	814425.9000000001
Prod_11	1786776.752000001
Prod_12	38981.549999999996
Prod_13	167107.2199999999
Prod_14	1130361.2999999998
Prod_15	1652823
Prod_16	80996.30999999997
Prod_17	2168697.139999999
Prod_2	736991.54
Prod_3	1022957.5900000001
Prod_4	1889313.8020000001
Prod_5	698093.8100000003

Q6.Retrieve total profit made from each product from the data (use windows function)

```
select distinct Prod_id,sum(profit)over(partition by prod_id) as profit from market_fact;
```

Prod_id	profit
Prod_16	-7799.250000000002
Prod_17	307712.93
Prod_2	97158.05999999995
Prod_3	307413.38999999996
Prod_4	316951.6200000001
Prod_5	100427.93000000007
Prod_6	45263.20000000007
Prod_7	-102.67000000000004
Prod_8	94287.48000000008

Q7.Count the total number of unique customers in January and how many of them came back every month over the entire year in 2011

```
select count(cust_id) as CustVisitMonth from (
select m.cust_id,count(distinct month(str_to_date(order_date,'%d-%m-%Y'))) as months
from orders_dimen join market_fact m using(ord_id)
where year(str_to_date(order_date,'%d-%m-%Y'))=2011
group by cust_id
having months >= 12) custvisit;
```

CustVisitMonth
0

Q8.Retrieve month-by-month customer retention rate since the start of the business.(using views)

```
-- 1. Creating view for all customers and their visits
create view customer_visits as
select cust_id,str_to_date(order_date,"%d-%m-%Y") cust_visit
from combined_table;
 -- 2.creating view to check previous visit of every customer
 create view customer_visits_timelamp as
 select cust_id ,year(cust_visit),cust_visit, lag(cust_visit) over(partition by cust_id
 order by cust_visit,year(cust_visit)) previous_visit
 from customer_visits order by cust_id,cust_visit;


 -- 3. creating view to check those customer who visited back next month
 create view customer_time_gaps as
 select cust_id, cust_visit, previous_visit
 from customer_visits_timelamp where  datediff(cust_visit,previous_visit)<61 and month(cust_visit)-month(previous_visit) in (1,-11) and
 month(cust_visit)!=month(previous_visit) order by cust_visit;
```

```sql
-- 4. creating view to check number of customer visited every month and checking the customer retained.
create view monthly_cust_visit as
select *,lag(Monthly_customer_visit)over() previous_mnth_cust_visit
from (select year(cust_visit) year,month(cust_visit) month,count(cust_visit) Monthly_customer_visit
 from customer_visits_timelamp
 group by year(cust_visit),month(cust_visit) order by year(cust_visit)) t;

select * from monthly_cust_visit;

create view customer_retention as
select year(cust_visit) year,month(cust_visit) Month, count(cust_visit) customer_retained
from customer_time_gaps group by year,month order by year,month;

-- 5 final answer checking thecustomer retention rate by joining the monthly_cust_visit and customer_retention
select year,month,previous_mnth_cust_visit,customer_retained,(customer_retained/previous_mnth_cust_visit)*100 cust_retn_rate
from customer_retention natural join monthly_cust_visit order by year;
```

year	month	previous_mnth_cust_visit	customer_retained	cust_retn_rate
2009	2	221	11	4.9774
2009	3	151	13	8.6093
2009	4	185	13	7.0270
2009	5	171	13	7.6023
2009	6	179	9	5.0279
2009	7	163	12	7.3620
2009	8	189	7	3.7037
2009	9	192	11	5.7292
2009	10	180	9	5.0000
2009	11	165	5	3.0303
2009	12	167	5	2.9940
2010	1	173	11	6.3584
2010	2	166	7	4.2169

THANK YOU