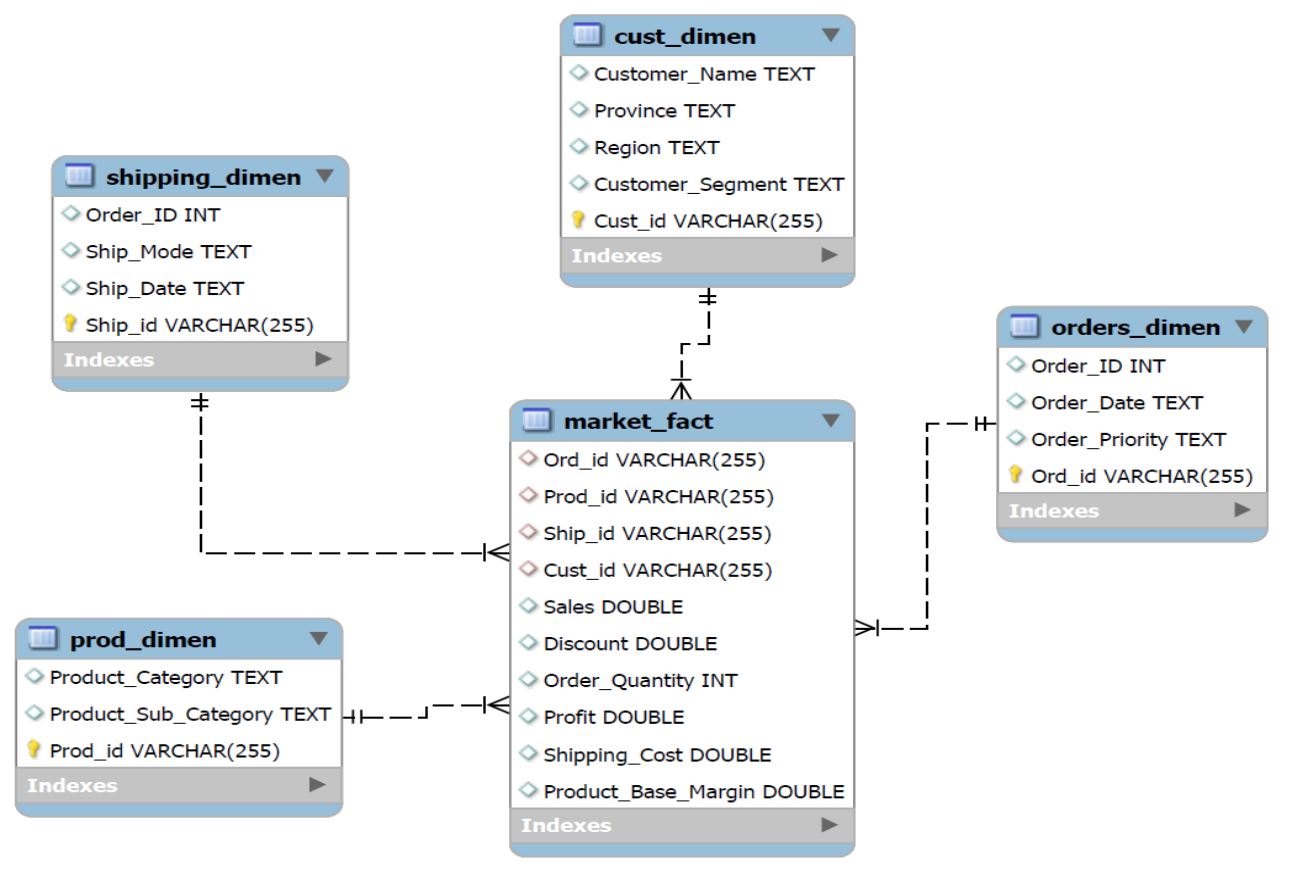
SQL PROJECT

DOMAIN:SALES AND DELIVERY

Problem Statement:

Composite data of a business organization, confined to ‘sales and delivery’ domain is given for the period of last decade. From the given data retrieve solutions for the given scenario.

ERD


cust_dimen
Customer_Name TEXT
Province TEXT
Region TEXT
Customer_Segment TEXT
Cust_id VARCHAR(255)
Indexes
shipping_dimen
Order_ID INT
Ship_Mode TEXT
Ship_Date TEXT
Ship_id VARCHAR(255)
Indexes
orders_dimen
Order_ID INT
Order_Date TEXT
Order_Priority TEXT
Ord_id VARCHAR(255)
Indexes
market_fact
Ord_id VARCHAR(255)
Prod_id VARCHAR(255)
Ship_id VARCHAR(255)
Cust_id VARCHAR(255)
Sales DOUBLE
Discount DOUBLE
Order_Quantity INT
Profit DOUBLE
Shipping_Cost DOUBLE
Product_Base_Margin DOUBLE
Indexes
prod_dimen
Product_Category TEXT
Product_Sub_Category TEXT
Prod_id VARCHAR(255)
Indexes

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 * from cust_dimen natural join
market_fact natural join orders_dimen natural join prod_dimen natural join
shipping_dimen;
```

RESULT:

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;
```

RESULT:

	Customer_Name	orders	rnk
▶	DARREN BUDD	26	1
	BRAD THOMAS	24	2
	ED BRAXTON	22	3

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

```
alter table combined_table add column Days_taken_for_delivery int;

update combined_table set Days_taken_for_delivery=datediff(str_to_date(ship_date,'%d-%m-%Y'),str_to_date(order_date,'%d-%m-%Y')) ;
```

Sales	Discount	Order_Quantity	Profit	Shipping_Cost	Product_Base_Margin	Customer_Name	Province	Region	Customer_Segment	Product_Category	Product_Sub_Category	Days_taken_for_delivery
10123.02	0.07	49	457.81	68.02	0.58	BARRY FRENCH	NUNAVUT	NUNAVUT	CONSUMER	OFFICE SUPPLIES	APPLIANCES	1
244.57	0.01	27	46.71	2.99	0.39	BARRY FRENCH	NUNAVUT	NUNAVUT	CONSUMER	OFFICE SUPPLIES	BINDERS AND BINDER ACCESSORIES	2
4965.7595	0.08	30	1198.97	3.99	0.58	CLAY ROZENDAL	NUNAVUT	NUNAVUT	CORPORATE	TECHNOLOGY	TELEPHONES AND COMMUNICATION	2
146.69	0.05	21	4.43	4.95	0.37	CARLOS SOLTERO	NUNAVUT	NUNAVUT	CONSUMER	FURNITURE	OFFICE FURNISHINGS	2
394.27	0.08	19	30.94	5.94	0.5	CARLOS SOLTERO	NUNAVUT	NUNAVUT	CONSUMER	OFFICE SUPPLIES	APPLIANCES	2
93.54	0.03	12	-54.04	7.72	0.38	CARL JACKSON	NUNAVUT	NUNAVUT	CORPORATE	OFFICE SUPPLIES	BINDERS AND BINDER ACCESSORIES	0
716.84	0	32	134.72	5.94	0.5	CARLOS DALY	NUNAVUT	NUNAVUT	HOME OFFICE	OFFICE SUPPLIES	APPLIANCES	1
1474.33	0.04	31	114.46	3.61	0.71	CARLOS DALY	NUNAVUT	NUNAVUT	HOME OFFICE	TECHNOLOGY	COMPUTER PERIPHERALS	2
80.61	0.02	15	-4.72	2.99	0.37	CLAUDIA MINER	NUNAVUT	NUNAVUT	SMALL BUSINESS	OFFICE SUPPLIES	BINDERS AND BINDER ACCESSORIES	0
248.26	0.07	16	93.8	1.39	0.4	ALLEN ROSENBLATT	NUNAVUT	NUNAVUT	SMALL BUSINESS	OFFICE SUPPLIES	ENVELOPES	1

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

```
select Customer_Name,Days_taken_for_delivery from combined_table
where Days_taken_for_delivery =(select max(Days_taken_for_delivery) from combined_table);
```

RESULT:

	Customer_Name	Days_taken_for_delivery
▶	DEAN PERCER	92

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

```
select distinct Prod_id,sum(Sales) over(partition by Prod_id) total_sales from combined_table ;
```

RESULT:

Prod_id	total_sales
Prod_1	1028240.76
Prod_10	814425.8999999997
Prod_11	1786776.7520000017
Prod_12	38981.55
Prod_13	167107.22000000003
Prod_14	1130361.3000000003
Prod_15	1652822.9999999988
Prod_16	80996.31
Prod_17	2168697.1399999997
Prod_2	736991.5399999993
Prod_3	1022957.5900000002

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

```
select distinct Prod_id,Product_Category,Product_Sub_Category,sum(Profit) over (partition by Prod_id)totalPROFIT from combined_table;
```

RESULT:

Prod_id	Product_Category	Product_Sub_Category	totalPROFIT
Prod_1	OFFICE SUPPLIES	STORAGE & ORGANIZATION	13599.490000000005
Prod_10	FURNITURE	BOOKCASES	-33729.08999999998
Prod_11	FURNITURE	TABLES	-113468.18000000001
Prod_12	OFFICE SUPPLIES	LABELS	13677.169999999996
Prod_13	OFFICE SUPPLIES	PENS & ART SUPPLIES	7564.780000000002
Prod_14	TECHNOLOGY	COPIERS AND FAX	167361.48999999993
Prod_15	FURNITURE	CHAIRS & CHAIRMATS	122738.0699999999
Prod_16	OFFICE SUPPLIES	SCISSORS, RULERS AND TRIMMERS	-7799.250000000003
Prod_17	TECHNOLOGY	OFFICE MACHINES	307712.92999999964
Prod_2	OFFICE SUPPLIES	APPLIANCES	97158.06
Prod_3	OFFICE SUPPLIES	BINDERS AND BINDER ACCESSORIES	307413.39000000025
Prod_4	TECHNOLOGY	TELEPHONES AND COMMUNICATION	316951.62000000005
Prod_5	FURNITURE	OFFICE FURNISHINGS	100427.93
Prod_6	OFFICE SUPPLIES	PAPER	45263.199999999975
Prod_7	OFFICE SUPPLIES	RUBBER BANDS	-102.67000000000017
Prod_8	TECHNOLOGY	COMPUTER PERIPHERALS	94287.48000000001
Prod_9	OFFICE SUPPLIES	ENVELOPES	48182.60000000002

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
(select count(distinct Cust_id)
from combined_table
where (
year(str_to_date(Order_Date,'%d-%m-%Y'))=2011 and month(str_to_date(Order_Date,'%d-%m-%Y'))=1))count_of_jan_customers
,
(SELECT COUNT(distinct Cust_id) AS count_of_customers
FROM (
    SELECT Cust_id, COUNT(DISTINCT MONTH(STR_TO_DATE(Order_Date, '%d-%m-%Y'))) AS month_count
    FROM combined_table
    where year(str_to_date(Order_Date,'%d-%m-%Y'))=2011
    GROUP BY Cust_id
    HAVING month_count = 12)AS customers_in_all_months_of_2011 where
    Cust_id in (select distinct Cust_id
from combined_table where (year(str_to_date(Order_Date,'%d-%m-%Y'))=2011 and month(str_to_date(Order_Date,'%d-%m-%Y'))=1)))jan_repeating_customers_in_year;
```

RESULT:

	count_of_jan_customers	jan_repeating_customers_in_year
▶	99	0

Q8:

(i)Create a view where each user's visits are logged by month, allowing for the possibility that these will have occurred over multiple # years since whenever business started operations.

```
create view customer_visits as
select cust_id,str_to_date(order_date,"%d-%m-%Y") cust_visit
from combined_table;
```

RESULT:

cust_id	cust_visit
Cust_2	2012-10-01
Cust_2	2012-10-01
Cust_3	2011-07-10
Cust_4	2010-08-28
Cust_4	2010-08-28
Cust_5	2011-06-17
Cust_9	2012-06-08
Cust_9	2012-06-08
Cust_10	2012-08-04

Q8:

(ii) Identify the time lapse between each visit. So, for each person and for each month, we see when the next visit is.

```
create view customer_monthly_visit_timelamp as
select cust_id,cust_visit, lag(cust_visit) over(partition by cust_id
order by cust_visit) previous_visit_month
from customer_visit cm ;
```

RESULT:

cust_id	cust_visit	previous_visit_month
Cust_1	2010-10-13	NULL
Cust_10	2012-08-04	NULL
Cust_100	2009-06-21	NULL
Cust_100	2009-11-29	2009-06-21
Cust_100	2010-05-07	2009-11-29
Cust_100	2010-05-07	2010-05-07
Cust_100	2010-06-09	2010-05-07
Cust_100	2011-09-22	2010-06-09

Q8:

(iii) Calculate the time gaps between visits.

```
create view customer_time_gaps as
select cust_id, cust_visit, previous_visit_month,round(datediff(cust_visit,previous_visit_month)/30) month_diff
from customer_monthly_visit_timelamp;
```

RESULT:

cust_id	cust_visit	previous_visit_month	month_diff
Cust_1	2010-10-13	NULL	NULL
Cust_10	2012-08-04	NULL	NULL
Cust_100	2009-06-21	NULL	NULL
Cust_100	2009-11-29	2009-06-21	5
Cust_100	2010-05-07	2009-11-29	5
Cust_100	2010-05-07	2010-05-07	0
Cust_100	2010-06-09	2010-05-07	1
Cust_100	2011-09-22	2010-06-09	16

Q8:

(iv) Categorize the customer with time gap 1 as retained, >1 as irregular and NULL as churned.

```
create view customer_categorise as
select cust_id, cust_visit,
case when month_diff = 1 then "retained"
     when month_diff  >1 then "irregular"
     else "churned"
end retention_status
from customer_time_gaps  ;
```

RESULT:

cust_id	cust_visit	retention_status
Cust_1	2010-10-13	churned
Cust_10	2012-08-04	churned
Cust_100	2009-06-21	churned
Cust_100	2009-11-29	irregular
Cust_100	2010-05-07	irregular
Cust_100	2010-05-07	churned
Cust_100	2010-06-09	retained
Cust_100	2011-09-22	irregular

Q8:

(iv) calculate the retention month wise

```
select year(cust_visit),month(cust_visit), count(cust_id) total_customer,
sum(case when retention_status = "retained" then 1 else 0 end) as retained_customer
from customer_categorise group by year(cust_visit),
month(cust_visit) order by year(cust_visit);
```

RESULT:

year(cust_visit)	month(cust_visit)	total_customer	retained_customer
2009	1	221	2
2009	2	151	7
2009	3	185	8
2009	4	171	10
2009	5	179	10
2009	6	163	12
2009	7	189	15
2009	8	192	10
2009	9	180	9
2009	10	165	7
2009	11	167	5
2009	12	173	5
2010	1	166	7
2010	2	178	10
2010	3	158	7
2010	4	149	5
2010	5	199	10
2010	6	168	9
2010	7	169	11
2010	8	181	9
2010	9	192	12
2010	10	195	12
2010	11	175	7
2010	12	205	10
2011	1	154	4
2011	2	169	6
2011	3	152	6
2011	4	165	3
2011	5	181	12
2011	6	160	7
2011	7	170	9
2011	8	146	1
2011	9	161	3
2011	10	167	9
2011	11	166	7
2011	12	185	7
2012	1	172	8
2012	2	151	6
2012	3	195	7
2012	4	180	10
2012	5	212	14
2012	6	151	5
2012	7	166	4
2012	8	180	5
2012	9	196	9
2012	10	198	10
2012	11	132	11
2012	12	156	6

THANK YOU