SQL Project

On

Global+Electronics+Retailer

Dr.Trupti Yadav

20-07-2024

Project Description:

This project involves analysing global electronics retailer data to identify data-driven insights.

Dataset Description:

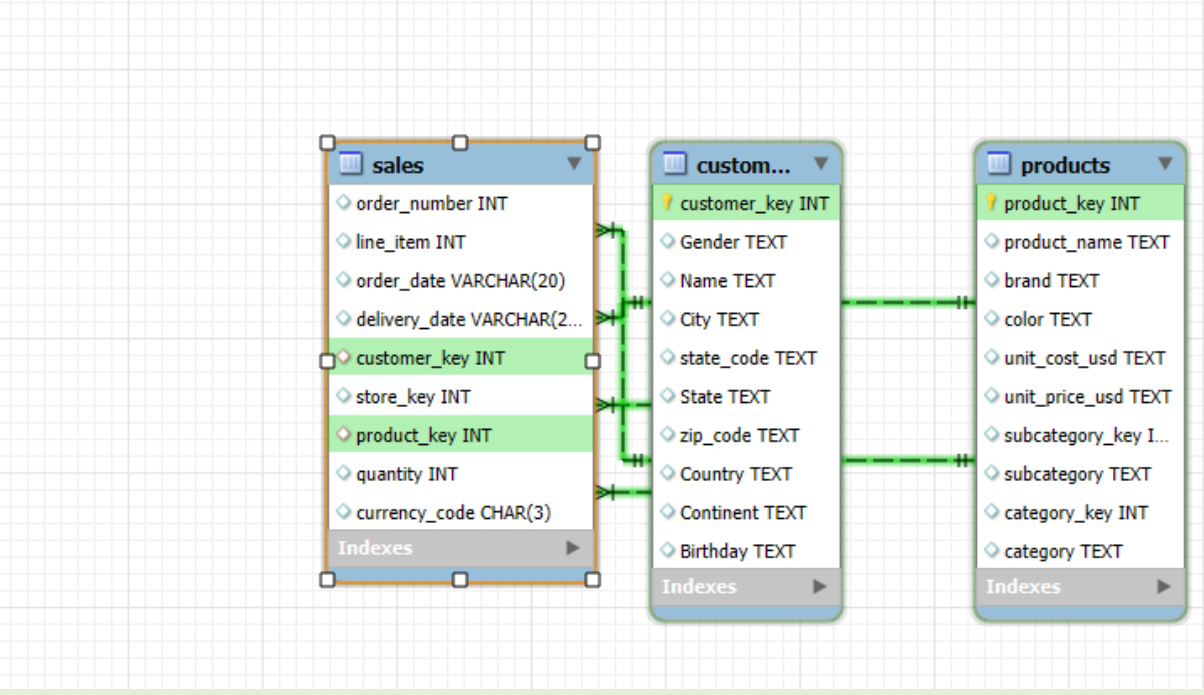
This data has 6 CSV files:

1) **Customers**.csv: It has information about customers.

   customer_key,gender,name,city,state_code,state,zip_code,country,continent,birthday

2) **Sales**.csv: Contains sales transaction details
   order_number,line_item,order_date,delivery_date,customer_key,store_key,product_key,quantity,currency_code
3) **Products**.csv: Lists product details such as product name, brand, colour, cost, and category.
   product_key,product_name,brand,color,unit_cost_usd,unit_price_usd,subcategory_key,subcategory,category_key,category
4) **Exchange_Rates**.csv: Provides currency exchange rates for different currencies
5) **Stores**.csv: Includes store-related information such as location, size, and open date.
6) **Data_Dictionary**.csv: Describes the fields in each table, its data structure and relationships between tables.

ER diagram



Create Database.

Step1:- Create Database

Create database GLOBAL;

use GLOBAL;

Uploading Data

1) Table Import wizard: Customer.csv, Products.csv
2) LOAD DATA LOCAL INFILE: Sales.csv

Check for the number of records in each table

Example: Select count(*) from customers;

```
72 •  select count(*) from customers;
```

Result Grid | Filter Rows:

count(*)
13677

```
72 •  select count(*) from products;
```

Result Grid | Filter Rows:

count(*)
2517

Check for duplicate rows in all tables

Limit to 5000 rows

```
2   -- we have to find how many rows are duplicated.
3
4 •  select COUNT(*) from customers
5   group by customer_key,gender,name,city,state_code,state,zip_code,country,continent,birthday
6   having count(*)>1;
7
8 •  select COUNT(*) from products
9   group by product_key,product_name,brand,color,unit_cost_usd,unit_price_usd,subcategory_key,
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

COUNT(*)

--Analysis from customer table

```
1 •  SELECT * FROM global.customers;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

ï»¿customer_key	Gender	Name	City	state_code	State	zip_code	Country	Continent	Birthday
301	Female	Lilly Harding	WANDEARAH EAST	SA	South Australia	5523	Australia	Australia	07-03-1939
325	Female	Madison Hull	MOUNT BUDD	WA	Western Australia	6522	Australia	Australia	9/27/1979
554	Female	Claire Ferres	WINJALLOK	VIC	Victoria	3380	Australia	Australia	5/26/1947
786	Male	Jai Poltpalingada	MIDDLE RIVER	SA	South Australia	5223	Australia	Australia	9/17/1957
1042	Male	Aidan Pankhurst	TAWONGA SOUTH	VIC	Victoria	3698	Australia	Australia	11/19/1965
1086	Male	Hayden Clegg	TEMPLERS	SA	South Australia	5371	Australia	Australia	1/20/1954
1133	Male	Nicholas Caffyn	JUBILEE POCKET	QLD	Queensland	4802	Australia	Australia	11/22/1969
1256	Male	Lincoln Jenks	KULLOGLUM	QLD	Queensland	4660	Australia	Australia	03-12-1950

```
3  alter table customers
4  rename column ï»¿customer_key to customer_key;
5
6  set sql_safe_updates=0;
7  -- USE OF REPLACE FUNCTION
8  update customers
9  set birthday=replace(Birthday,'-','/');
10
11 -- CORRECTING FORMAT FOR DATE COLUMN
12 update customers
13 set birthday=STR_TO_DATE(Birthday, '%m/%d/%Y');
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

customer_key	Gender	Name	City	state_code	State	zip_code	Country	Continent	Birthday
301	Female	Lilly Harding	WANDEARAH EAST	SA	South Australia	5523	Australia	Australia	1939-07-03
325	Female	Madison Hull	MOUNT BUDD	WA	Western Australia	6522	Australia	Australia	1979-09-27
554	Female	Claire Ferres	WINJALLOK	VIC	Victoria	3380	Australia	Australia	1947-05-26
786	Male	Jai Poltpalingada	MIDDLE RIVER	SA	South Australia	5223	Australia	Australia	1957-09-17
1042	Male	Aidan Pankhurst	TAWONGA SOUTH	VIC	Victoria	3698	Australia	Australia	1965-11-19

*The STR_TO_DATE() converts a string into a date value based on a specified format string.

TO ADD PRIMARY KEY CONSTRAINT

```
14
15 ALTER TABLE CUSTOMERS
16 ADD CONSTRAINT pk_customers PRIMARY KEY(customer_key);
17
18 describe CUSTOMERS;
19
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Field	Type	Null	Key	Default	Extra
customer_key	int	NO	PRI	NULL	
Gender	text	YES		NULL	
Name	text	YES		NULL	
City	text	YES		NULL	
state_code	text	YES		NULL	

Result 7

Customer Count Per Country

```
38
39      -- CUSTOMER COUNT PER COUNTRY
40      SELECT Country,
41          COUNT( Customer_Key) AS CUSTOMER_COUNT
42      FROM Customers
43      GROUP BY Country
44      ORDER BY CUSTOMER_COUNT DESC;
45
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Country	CUSTOMER_COUNT
United States	6828
United Kingdom	1944
Canada	1553
Germany	1473
Australia	1420
Netherlands	733

Result 2

--Analysis from Products table

```
20
21      alter table products
22      rename column ï»¿product_key to product_key;
23
24      ALTER TABLE products
25      ADD CONSTRAINT pk_products PRIMARY KEY(product_key);
26      describe products;
27
28
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Field	Type	Null	Key	Default	Extra
product_key	int	NO	PRI	NULL	
product_name	text	YES		NULL	
brand	text	YES		NULL	
color	text	YES		NULL	
unit_cost_usd	text	YES		NULL	
unit_price_usd	text	YES		NULL	
subcategory_key	int	YES		NULL	
subcategory	text	YES		NULL	
category_key	int	YES		NULL	

Result 8

Types and Quantity of Products Sold

Create A view

```
56
57 •  create view quantity as
58    SELECT DISTINCT  Category, Subcategory, COUNT(category) AS Quantity
59    FROM products
60    GROUP BY Category,Subcategory
61    ORDER BY Category ;
62 •  select * from quantity;
63
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Category	Subcategory	Quantity
Audio	Bluetooth Headphones	50
Audio	MP4&MP3	45
Audio	Recording Pen	20
Cameras and camcorders	Camcorders	103
Cameras and camcorders	Cameras & Camcorders Accessories	69

Quantity sold above 100

```
63
64 •  select quantity,Category,Subcategory
65    from quantity
66    where quantity>100;
67
68
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

quantity	Category	Subcategory
103	Cameras and camcorders	Camcorders
101	Cell phones	Smart phones & PDAs
201	Computers	Computers Accessories
101	Computers	Printers, Scanners & Fax
103	Computers	Projectors & Screens
120	Games and Toys	Download Games
158	Home Appliances	Lamps
102	Home Appliances	Microwaves

--Analysis from Sales table

```
137 • show global variables like 'local_infile';
138 • set global local_infile=true;
139 • set global local_infile=0;
140 • set global local_infile=1;
141 • LOAD DATA LOCAL INFILE 'C:/Users/trupt/Documents/sales.csv' INTO TABLE sales fields terminated by ','
142   LINES TERMINATED BY '\n'
143   ignore 1 LINES;
144 • select count(*) from sales;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

count(*)
55006

```
alter table sales
add constraint fk_sales
FOREIGN KEY (customer_key) REFERENCES Customers(customer_key),
add constraint fk_sales1
FOREIGN KEY (product_key) REFERENCES products(product_key);
```

118	20:10:43	SELECT * FROM global.sales LIMIT 0, 5000	5000 row(s) returned	0.016 sec / 0.01
119	20:11:23	alter table sales add constraint fk_sales FOREIGN KEY (customer_key) REFERENCE...	55006 row(s) affected Records: 55006 Duplicates: 0 Warnings: 0	1.594 sec

```
21 • describe sales;
22
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Field	Type	Null	Key	Default	Extra
order_number	int	YES		NULL	
line_item	int	YES		NULL	
order_date	varchar(20)	YES		NULL	
delivery_date	varchar(20)	YES		NULL	
customer_key	int	YES	MUL	NULL	
store_key	int	YES		NULL	
product_key	int	YES	MUL	NULL	
quantity	int	YES		NULL	
currency_code	char(3)	YES		NULL	

Use of Replace, Str_to_date

```
-- USE OF REPLACE FUNCTION
update sales
set order_date=replace(order_date,'-','/');

-- CORRECTING FORMAT FOR DATE COLUMN
update sales
set order_date=STR_TO_DATE(order_date, '%m/%d/%Y');

select * from sales;
```

ult Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

order_number	line_item	order_date	delivery_date	customer_key	store_key	product_key	quantity	currency_code
366000	1	2016-01-01		265598	10	1304	1	CAD
366001	1	2016-01-01	1/13/2016	1269051	0	1048	2	USD
366001	2	2016-01-01	1/13/2016	1269051	0	2007	1	USD
366002	1	2016-01-01	01-12-2016	266019	0	1106	7	CAD
366002	2	2016-01-01	01-12-2016	266019	0	373	1	CAD
366002	3	2016-01-01	01-12-2016	266019	0	1080	4	CAD
366004	1	2016-01-01		1107461	38	163	6	GBP

Inner Join(customers and sales)

```
155  select c.customer_key,c.gender,c.country,s.order_date,s.quantity
156  from customers as c
157  inner join sales as s
158  on
159  c.customer_key=s.customer_key
160  where gender="Female" and country like "A%";
161
162
163
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

customer_key	gender	country	order_date	quantity
301	Female	Australia	2019-11-11	2
325	Female	Australia	2018-01-13	4
325	Female	Australia	2018-01-13	4
325	Female	Australia	2018-01-13	3
325	Female	Australia	2019-11-16	3
325	Female	Australia	2019-11-16	3
325	Female	Australia	2020-01-04	1
325	Female	Australia	2020-01-04	1

Result 2

LEFT JOIN

```
2 •  select distinct p.category,p.subcategory,s.order_date
3    FROM Sales as S
4    LEFT JOIN Products as P ON S.Product_Key=P.Product_Key
5    group by  p.category,p.subcategory,s.order_date;
6
```

sult Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch row

category	subcategory	order_date
Cameras and camcorders	Cameras & Camcorders Accessories	2016-01-01
Cameras and camcorders	Digital SLR Cameras	2016-01-01
Home Appliances	Microwaves	2016-01-01
Computers	Laptops	2016-01-01
TV and Video	Televisions	2016-01-01
Cell phones	Smart phones & PDAs	2016-01-01

ult 3 ×

Use of substring Function.

```
22 •  select substring(Unit_Price_USD,2) as unitprice,Unit_Price_USD
23    from products;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

unitprice	Unit_Price_USD
12.99	$12.99
12.99	$12.99
14.52	$14.52
21.57	$21.57
21.57	$21.57
21.57	$21.57
21.57	$21.57

**********END**********