Project based on Retails sales

By Manasa Devi. K
2024

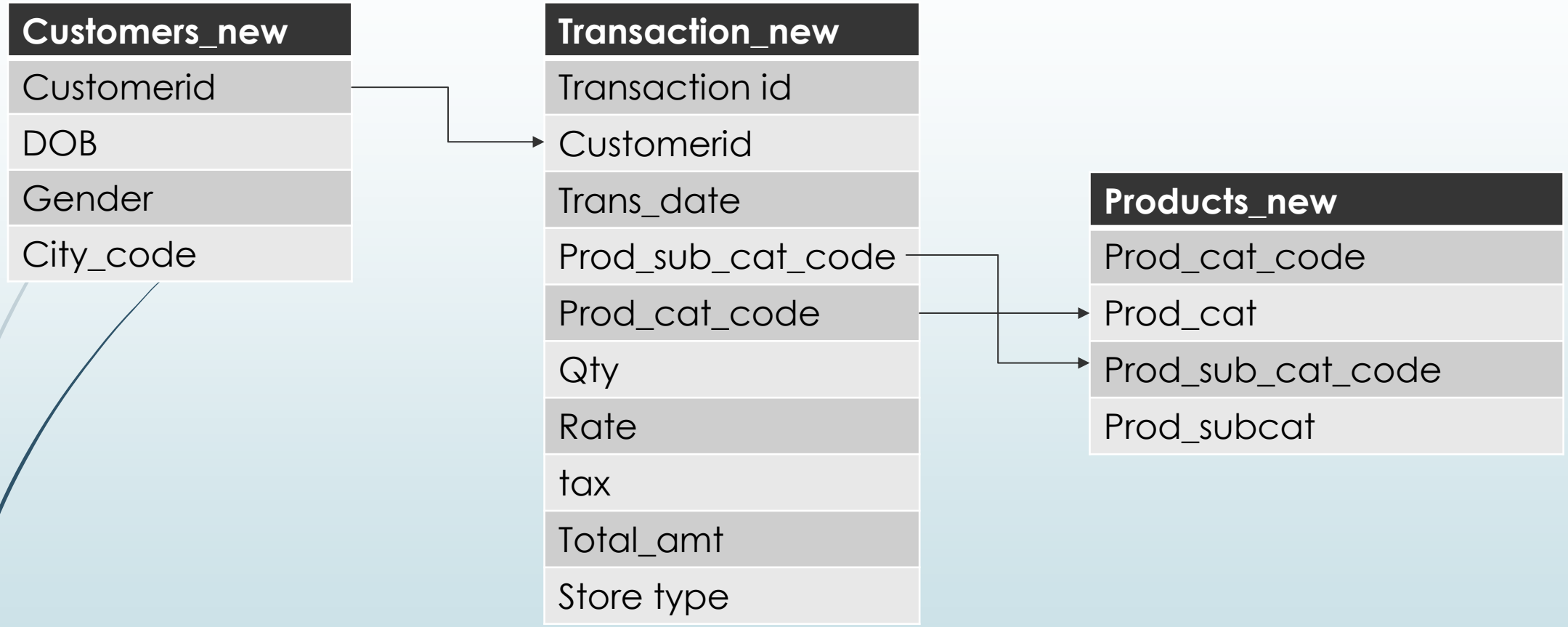
Objective

- To use data analysis method to optimize the performance of business insights for E-commerce databases.
- To predict the sales outcome for current standard, comparison with competitor and predict the futuristic results.
- To collect the data for Retail E-commerce details from the given three tables(customer, transaction and product) and executing the objectives and get the desired output.

Retail Business database overview

- Business database from 3 tables to understand the behavior of customer purchase details and transaction information.
- To understand the project based on customer behavior and understand interest toward the particular product purchase that involves the increase sales and revenue.
- Customers data: The customer table data provides the information about the E-commerce details variable like customer id, DOB(Date of birth), Gender and City_code
- Total variables:5645
- Transaction data: Table contains the transaction sale, transaction date, Qty(Quantity), Customerid, product category code, product subcategories code, Price(Rate), Tax, Total amount.
- Transaction records:23053
- Product category data: The table contains the information about the product categories, product subcategories, product category code, product subcategories code. Total product record: 24

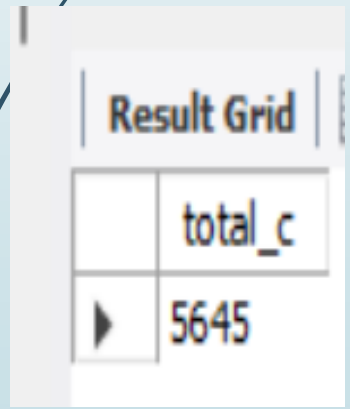
Database flowchart overview

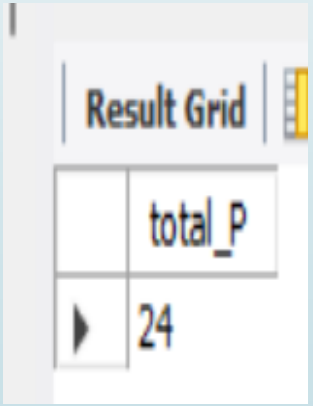


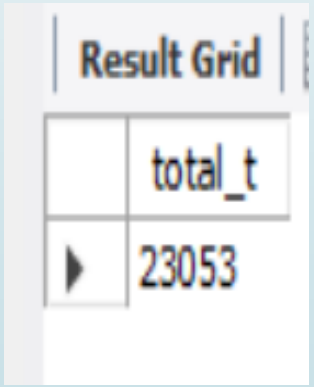
DATA PREPARATION AND UNDERSTANDING

Q1 what is the total no of row in each of the 3 tables in the database

```
select count(*) as total_c from customers_new;
select count(*) as total_P from prod_cat_info;
select count(*) as total_t from transactions_new;
```

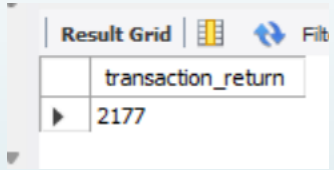






Q2 what is the total number of transaction that have a return

```
select count(transaction_id) as transaction_return from transactions_new
where total_amt < 0;
```



Q3 What is the time range of the transaction data available for analysis? Show the output in number of days, months and years simultaneously in different columns.

```
SET SQL_SAFE_UPDATES = 0;
desc transactions_new;
desc customers_new;
update transactions_new
set tran_date = STR_TO_DATE(tran_date, '%d-%m-%Y')
where tran_date is not null;
alter table transactions_new
change column tran_date tran_date  date not null;
```

```
update customers_new
set DOB = STR_TO_DATE(DOB, '%d-%m-%Y')
where DOB is not null;
select DOB from customers_new;
alter table customers_new
change column DOB DOB  date not null;
```

```
select
     MIN(tran_date) as min_order_date,
     MAX(tran_date) as max_delivery_date,
     DATEDIFF(MAX(tran_date), MIN(tran_date)) AS total_days,
     timestampdiff(MONTH, MIN(tran_date), MAX(tran_date)) as total_months,
     timestampdiff(YEAR, MIN(tran_date),MAX(tran_date)) as total_years
FROM transactions_new;
```

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

min_order_date	max_delivery_date	total_days	total_months	total_years
2011-01-02	2014-12-02	1430	47	3

Q4 which product category does the sub-category "DIY" belongs to?

```
select * from prod_cat_info where prod_subcat = "DIY";
```

Result Grid | Filter Rows: | Export:

prod_cat_code	prod_cat	prod_sub_cat_code	prod_subcat
5	Books	6	DIY

DATA ANALYSIS AND PREPARATION

Q1 which channel is most frequently used for transaction?

```
select S.Store_type from
(select count(distinct(transaction_id)) as Total_count,Store_type from transactions_new
group by Store_type
order by Total_count desc
limit 1 ) S;
```

Result Grid

	Store_type
▶	e-Shop

Q2 what is the count of male and female Customer in the database

```
select count(distinct(customerid)),Gender from customers_new
where Gender in ("M","F")
group by Gender;
```

Result Grid | Filter Rows:

count(distinct(customerid))	Gender
2752	F
2891	M

Q3 From which city do we have the maximum number of customers and how many?

```
select city_code, count(distinct (customerid)) as No_of_Customers from customers_new
group by city_code
order by No_of_Customers desc
limit 1;
```

Result Grid | Filter Rows:

	city_code	No_of_Customers
▶	3	595

Q4 how many sub-categories are there under the books category

```
select prod_cat, count(prod_subcat) as No_of_Subcategories
from prod_cat_info where prod_cat  = 'Books' group by prod_cat;
```

	prod_cat	No_of_Subcategories
▶	Books	6

Q5 What is the Maximum quantity of products ever ordered

```
select max(Qty) as Max_ordered from transactions_new
group by prod_cat_code
order by max(Qty) desc limit 1 ;
```

	Max_ordered
▶	5

Q6 What is the net total revenue generated in categories Electronics and Books?

```
select P.prod_cat, round(sum(T.total_amt),2) as net_total_revenue from transactions_new T
join prod_cat_info P on P.prod_cat_code = T.prod_cat_code and P.prod_sub_cat_code = T.prod_subcat_code
group by P.prod_cat
having P.prod_cat in('Electronics', 'Books');
```

prod_cat	net_total_revenue
Electronics	10722463.63
Books	12822694.04

Q7 How many customers have > 10 transactions with us, excluding returns?

```
with customers as
(
select cust_id, count(transaction_id) as no_of_transactions from transactions_new T
where total_amt > 0
group by cust_id
having count(transaction_id) > 10
)
select count(*) as no_of_customers from customers_new;
```

no_of_customers
5645

Q8 What is the combined revenue earned from the "Electronics" & "Clothing" categories, from "Flagship stores"?

```
select T.Store_type, round(sum(total_amt),2) as combined_revenue from transactions_new T
join prod_cat_info P on T.prod_subcat_code = P.prod_sub_cat_code and T.prod_cat_code = P.prod_cat_code
where P.prod_cat in ('Electronics', 'Clothing') and T.total_amt > 0
group by T.Store_type
having T.Store_type = 'Flagship store';
```

Store_type	combined_revenue
Flagship store	3851454.3

Q9 What is the total revenue generated from "Male" customers in "Electronics" category? # Output should display total revenue by prod sub-cat.

```
select P.prod_cat, P.prod_subcat , round(sum(T.total_amt),2) as total_revenue from transactions_new T
join customers_new C on C.customerid  = T.cust_id
join prod_cat_info P on T.prod_subcat_code = P.prod_sub_cat_code and T.prod_cat_code = P.prod_cat_code
where C.gender = 'M' and T.total_amt > 0
group by P.prod_cat, P.prod_subcat
having P.prod_cat = 'Electronics';
```

prod_cat	prod_subcat	total_revenue
Electronics	Computers	1204053.3
Electronics	Cameras	1307201.74
Electronics	Mobiles	1351085.71
Electronics	Personal Appliances	1216993.96
Electronics	Audio and video	1230539.05

Q10 What is percentage of sales and returns by product sub-category; display only top 5 sub-categories in terms of sales?

```
select P.prod_subcat, round(sum(T.total_amt),2) as total_sales
from transactions_new T join prod_cat_info P on t.prod_subcat_code = p.prod_sub_cat_code
and T.prod_cat_code = P.prod_cat_code
where T.total_amt > 0
group by P.prod_subcat
order by total_sales desc
limit 5;
```

prod_subcat	total_sales
Women	7020079.36
Mens	6905869.88
Kids	4806698.07
Mobiles	2508648.35
Fiction	2492901

Q11 For all customers aged between 25 to 35 years find what is the net total revenue generated by # these consumers in last 30 days of transactions from max transaction date available in the data?

```sql
with max_tran_date as
  (select max(tran_date) as max_date from transactions_new),
last_30days_trans as (
    select T.cust_id, T.tran_date, T.total_amt, M.max_date
    from transactions_new T cross join max_tran_date M
    where T.tran_date between DATE_SUB(M.max_date, interval 30 day) and M.max_date
),
age_25_30 as (
    select C.customerid, year(M.max_date) - year(C.DOB) as age
    from customers_new C
    cross join max_tran_date M
    where year(M.max_date) - year(C.DOB) between 25 and 35
),
```

```sql
net_rev as (
    select sum(T.total_amt) AS net_total_revenue
    from last_30days_trans T
    join age_25_30 A ON T.cust_id = A.customerid
)
select net_total_revenue from net_rev;
```

net_total_revenue
74885.85

Q12 Which product category has seen the max value of returns in the last 3 months of transactions?

```sql
with max_tran_date as
    (select max(tran_date) as max_date from Transactions_new),
last_90days_returns as (
    select P.prod_cat, sum(case when T.total_amt < 0 then T.total_amt else 0 end) as return_amount
    from transactions_new T
    join max_tran_date M on T.tran_date between DATE_SUB(M.max_date, interval 90 day) and M.max_date
    left join prod_cat_info P on T.prod_subcat_code = P.prod_sub_cat_code and T.prod_cat_code = P.prod_cat_code
    group by P.prod_cat
)
select prod_cat, return_amount
from last_90days_returns
order by return_amount
limit 1;
```

Result Grid | Filter Rows:

prod_cat	return_amount
Home and kitchen	-9840.025

Q13 Which store-type sells the maximum products; by value of sales amount and by quantity sold?

```
select Store_type, round(sum(total_amt),2) as total_sales, count(Qty) as qty_sold from transactions_new T
where total_amt > 0
group by Store_type
order by total_sales desc, qty_sold desc
limit 1;
```

Store_type	total_sales	qty_sold
e-Shop	22185609.87	8429

Q14 What are the categories for which average revenue is above the overall average.

```
select  p.prod_cat, avg(t.total_amt) as avg_cat_rev from transactions_new T
join prod_cat_info P on T.prod_cat_code = P.prod_cat_code and T.prod_subcat_code = P.prod_sub_cat_code
where total_amt > 0
group by P.prod_cat
having avg(T.total_amt) > (select avg(total_amt) as overall_avg_rev from transactions_new where total_amt > 0);
```

prod_cat	avg_cat_rev
Electronics	2640.6858840579657
Books	2622.09289099527
Bags	2617.9599468977044
Clothing	2643.6835949177907

Q15 Find the average and total revenue by each subcategory for the categories which are among top 5 categories in terms of quantity sold.

```sql
with TopCategories as (
    select P.prod_cat, T.prod_cat_code, count(Qty) as qty_sold
    from transactions_new T inner join prod_cat_info P
    on T.prod_cat_code = P.prod_cat_code and T.prod_subcat_code = P.prod_sub_cat_code
    where T.total_amt > 0
    group by P.prod_cat, T.prod_cat_code
    order by qty_sold desc
    limit 5
)
select  P.prod_cat, round(avg(T.total_amt),2) as avg_revenue, round(sum(T.total_amt),2) as total_revenue
from transactions_new T join prod_cat_info P
on T.prod_cat_code = P.prod_cat_code and T.prod_subcat_code = P.prod_sub_cat_code
join TopCategories TC on T.prod_cat_code  = TC.prod_cat_code
group by P.prod_cat;
```

prod_cat	avg_revenue	total_revenue
Books	2112.82	12822694.04
Electronics	2189.15	10722463.64
Home and kitchen	2043.83	8438993.29
Footwear	2073.95	6219774.28
Clothing	2111.87	6251137.49

Conclusion

- From the Data- analysis better understanding the customer retention, sales predication for each category Electronics, Flagships, Books, DIY products and predict the upcoming performance growth.
- Challenges faced while understanding the time analysis and total revenue on different period.
- The insights highlight the performance which product category sales tops the revenue and yield. To increase the sales, avail some exclusive offers on weekends and month basis.
- To include practice offering support on multiple channels, enabling self-service systems, and using customer feedback to evaluate your performance.

Thank you