

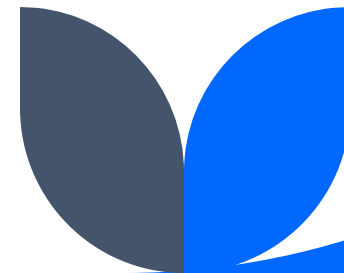
E-Commerce Retail Data Analysis





INTRODUCTION

E-commerce retail data analysis to
make informed decisions
on transactional data, product
purchased and customers behaviour.



OBJECTIVES

- ❑ Analytics for ecommerce merchants helps them determine future trends based on several factors including the size of transactions, the season, category of products and so on.
- ❑ Analytics plays a key role in the detection of fraud.
- ❑ Overall revenue calculation generated by product categories to have an insights on sales in years.
- ❑ Analyse returns help us to minimize the returns and improve customer satisfaction.
- ❑ Determining top selling stores ,which helps in optimizing sales strategies .
- ❑ Discover most popular stores,that helps us to see which store attracts most customers and has high sales.

DATA SETS



CUSTOMERS

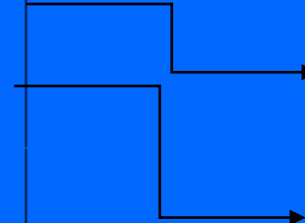
CUSTOMER_ID
DOB
GENDER
CITY_CODE

TRANSACTIONS

Transactions_id
Cust_id
Prod_cat_code
Prod_subcat_code
trans_date
Qty
Rate
Total_amt
Store_type

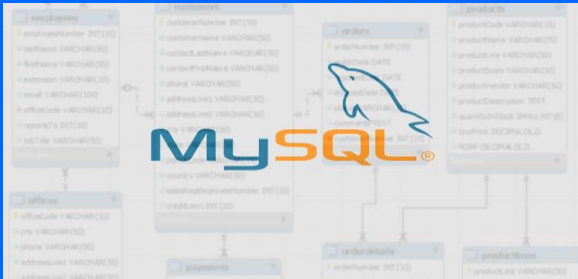
PROD_CAT_INFO

Prod_cat
Prod_cat_code
Prod_subcat
Prod_sub_cat_code



Identified key data sources and databases for the analysis via:

❑ MySQL



1. What is the total number of rows in each of the 3 tables in the database?

```
select count(*) as total_records_cust from customers_new;
```

```
select count(*) as total_records_trans from transaction_new;
```

```
select count(*) as total_records_prod from prod_cat_info;
```

Output :

	total_records_cust
▶	5645

	total_records_trans
▶	23053

	total_records_prod
▶	23

2. what is the total number of transactions that have return?

```
select count(*) as total_returns from transaction_new  
where Qty < 0;
```

Output:

Result Grid	
	total_returns
▶	2177

3. As you would have noticed , the dates provided across the datasets are not in a correct format as first step,please convert the data variables into valid date formats before proceeding ahead.

```
update customers_new
set DOB= str_to_date(DOB,'%d-%m-%Y')
where DOB is not null;

alter table customers_new
change column DOB DOB date not null;

alter table transaction_new
rename column i»¿transaction_id to transaction_id;

update transaction_new
set tran_date= str_to_date(tran_date,'%d-%m-%Y')
where tran_date is not null;

alter table transaction_new
change column tran_date trans_date date not null;
```


4.what is the time range if transaction data available for analysis? show the output in number of days,months,years simultaneously in different columns

```
select  
    MIN(trans_date) AS start_date,  
    MAX(trans_date) AS end_date,  
    DATEDIFF(MAX(trans_date), MIN(trans_date)) AS range_tran_days,  
    TIMESTAMPDIFF(MONTH, MIN(trans_date), MAX(trans_date)) AS range_tran_months,  
    TIMESTAMPDIFF(YEAR, MIN(trans_date), MAX(trans_date)) AS range_tran_years  
from transaction_new;
```

	start_date	end_date	range_tran_days	range_tran_months	range_tran_years
▶	2011-01-02	2014-12-02	1430	47	3

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

```
select prod_cat, prod_subcat from prod_cat_info  
where prod_subcat="DIY";
```

Output :

	prod_cat	prod_subcat
▶	Books	DIY

1. Which channel is most frequently used for transaction?

```
select store_type,count(transaction_id) as most_freq_used_channel from transaction_new  
group by store_type  
order by most_freq_used_channel desc limit 1;
```



Output :

	store_type	most_freq_used_channel
▶	e-Shop	9311

2. What is the count of male and female customers in database?

```
select Gender,count(customer_id) as count_of_gender from customers_new  
where Gender in ('M','F')  
group by Gender;
```



Output :

Result Grid   Filter Rows: <input type="text"/>		
	Gender	count_of_gender
▶	M	2891
	F	2752

3. From which city do we have maximum number of customers and how many?

```
select city_code as CityCode_with_Max_Customers, No_of_customers from  
(select city_code ,count(customer_id) as No_of_customers from customers_new  
group by city_code  
order by No_of_customers desc limit 1) as abc;
```

Output:

Result Grid   Filter Rows: <input type="text"/>		
	CityCode_with_Max_Customers	No_of_customers
▶	3	595

4. How many sub-categories are there under the Books category?

```
select count(*) as total_returns from transaction_new  
where Qty < 0;
```

Output:

Result Grid	
	total_returns
▶	2177

5.What is the maximum quantity of products ever ordered?

```
select max(Qty) as max_Qty_ordered from transaction_new;
```

Output:

	max_Qty_ordered
▶	5

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

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

Output:

	prod_cat	net_total_revenue
►	Electronics	10722463.63
	Books	12822694.04

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

```
select count(*) as No_of_customers_with_trans_more_10 from
(
select cust_id, count(transaction_id) as Total_transactions from transaction_new
where Qty > 0
group by cust_id
) as res
where Total_transactions > 10;
```

Output:

	No_of_customers_with_trans_more_10
▶	6

8.What is the combined revenue earned from the "Electronics" and "Clothing" categories, from "Flagship stores"?

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

Output:

	Store_type	combined_revenue
▶	Flagship store	3851454.3

9. What is the total revenue generated from "MALE" customers in "Electronics" category? output should display total revenue by prod_subcat.

```
select C.Gender,P.prod_cat,P.prod_subcat,sum(T.total_amt) as total_revenue from transaction_new T
join prod_cat_info P on P.prod_cat_code = T.prod_cat_code and T.prod_subcat_code = P.prod_sub_cat_code
join customers_new C on C.customer_id = T.cust_id
where C.Gender = "M" and P.prod_cat = "Electronics" and total_amt > 0
group by P.prod_subcat;
```

Output:

	Gender	prod_cat	prod_subcat	total_revenue
►	M	Electronics	Audio and video	1230539.0500000005
	M	Electronics	Cameras	1307201.7399999984
	M	Electronics	Personal Appliances	1216993.9599999988
	M	Electronics	Mobiles	1351085.7099999993
	M	Electronics	Computers	1204053.3049999992

10. What is the percentage of sales and returns by product sub category ; Display only top 5 sub categories in terms of sales.

Output:

```
with sales_total_cte as
(
select P.prod_subcat ,sum(T.Qty) as total_sales from transaction_new T
inner join prod_cat_info P on P.prod_cat_code = T.prod_cat_code and P.prod_sub_cat_code = T.prod_subcat_code
where T.Qty > 0
group by P.prod_subcat order by total_sales desc
)
select res_subquery.prod_subcat,
round((((total_sales)/(total_sales+total_returns) * 100),2) as sales_percentage,
round((((total_returns)/(total_sales+total_returns) * 100),2) as returns_percentage from (
select S.prod_subcat, S.total_sales,abs(sum(T.Qty)) as total_returns from sales_total_cte S
inner join prod_cat_info P on S.prod_subcat = P.prod_subcat
inner join transaction_new T on P.prod_cat_code = T.prod_cat_code and P.prod_sub_cat_code = T.prod_subcat_code
where T.Qty < 0 group by S.prod_subcat,S.total_sales order by total_returns desc) as res_subquery
group by res_subquery.prod_subcat order by sales_percentage desc limit 5;
```

	prod_subcat	sales_percentage	returns_percentage
▶	Audio and video	92.48	7.52
	Cameras	91.50	8.50
	DIY	91.48	8.52
	Non-Fiction	91.46	8.54
	Mobiles	91.45	8.55

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

```
with max_tran_date as
  (select max(trans_date) as max_date from transaction_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 transaction_new T
  join max_tran_date M on T.trans_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;
```

Output:

	prod_cat	return_amount
►	Home and kitchen	-9840.025

13.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 total_qty_sold from transaction_new
where total_amt > 0
group by Store_type
order by total_sales desc, total_qty_sold desc
limit 1;
```

Output:

	Store_type	total_sales	total_qty_sold
►	e-Shop	22185609.87	8429

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

```
with overall_avg_rev as
(
    select round(avg(total_amt),2) as overall_avg from transaction_new where total_amt > 0
)
select P.prod_cat, round(avg(T.total_amt),2) as avg_rev_categorical from transaction_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 T.total_amt > 0
group by P.prod_cat
having avg_rev_categorical > (select overall_avg from overall_avg_rev);
```

Output:

	prod_cat	avg_rev_categorical
▶	Electronics	2640.69
	Books	2622.09
	Bags	2617.96
	Clothing	2643.68

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

```
with top_5_cat as
```

```
(
```

```
select P.prod_cat, sum(T.Qty) as total_qty from transaction_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 T.Qty > 0
```

```
group by P.prod_cat
```

```
order by total_qty desc limit 5
```

```
)
```

```
select C.prod_cat , P.prod_subcat, round(avg(T.total_amt),2) as Avg_rev, round(sum(T.total_amt),2) as Sum_rev from top_5.
```

```
left join prod_cat_info P on P.prod_cat = C.prod_cat
```

```
left join transaction_new T on T.prod_cat_code = P.prod_cat_code and T.prod_subcat_code = P.prod_sub_cat_code
```

```
group by P.prod_subcat, C.prod_cat;
```


Output:

Continue..

Result Grid				
		Filter Rows:		
		Export:		
	prod_cat	prod_subcat	Avg_rev	Sum_rev
►	Books	DIY	2108.37	2085180.3
	Books	Comics	2037.68	2100848.1
	Books	Children	2136.67	2211450.87
	Books	Non-Fiction	2129.1	2137620.29
	Books	Academic	2125.49	2055344.2
	Books	Fiction	2140.22	2232250.28
	Electronics	Audio and video	2247.96	2140057.92
	Electronics	Cameras	2165.88	2133390.35
	Electronics	Personal Appliances	2170.98	2110196.4
	Electronics	Computers	2181.75	2090116.34
	Electronics	Mobiles	2181.09	2248702.62
	Home and...	Tools	2024.37	2149882.48
	Home and...	Bath	2059.85	2107226.16
	Home and...	Kitchen	2008.96	2083289.65
	Home and...	Furnishing	2084.01	2098595.01
	Footwear	Kids	2125.99	2145126.55
	Footwear	Women	1989.58	2085083.07
	Footwear	Mens	2112.06	1989564.65

CONCLUSION

- In conclusion, E-commerce is a great way for business and consumers to interact internationally.
- Our ecommerce retail data analysis has provided invaluable that can significantly impact our business strategies, by identifying the most popular stores.
- Our analysis of product quantity, total revenue, returns, sales data will help to optimize the inventory management and to improve customer satisfaction.

THANK YOU

Jyoti Dhage

