**E-Commerce Shipping Analysis**

**Project Overview**  
This project focuses on identifying key patterns and trends in e-commerce shipping data, highlighting areas for operational improvement. SQL queries are utilized for data exploration and preprocessing, while Power BI visualizations bring the insights to life through interactive dashboards.

**Dataset Description**

The dataset used in this project was sourced from Kaggle: EDA - E-commerce Shipping Data. The dataset contains 12 attributes and 10,999 observations. Key features are:

ID: Unique identifier for each customer.  
Warehouse Block: Warehouses are divided into blocks (A, B, C, D, E).  
Mode of Shipment: Delivery methods include Ship, Flight, and Road.  
Customer Care Calls: Number of customer calls regarding shipments.  
Customer Rating: Ratings on a scale of 1 (worst) to 5 (best).  
Cost of the Product: Price of the product in USD.  
Prior Purchases: Number of previous purchases made by the customer.  
Product Importance: Categories include Low, Medium, and High importance.  
Gender: Gender of the customer (Male/Female).  
Discount Offered: Discount applied to the product.  
Weight in Grams: Weight of the product in grams.  
Reached on Time: Target variable:  
0: Delivered on time.  
1: Delivery delayed.

**Step-by-Step guidelines for the project :**

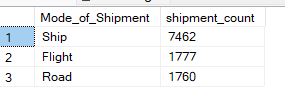
-- check for duplicates

select ID,COUNT(\*) from ecom\_data

group by ID having COUNT(\*)>1

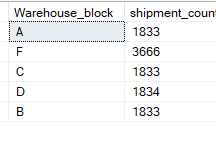
1) --Shipment mode distribution

select Mode\_of\_Shipment,COUNT(\*) as shipment\_count from ecom\_data group by Mode\_of\_Shipment



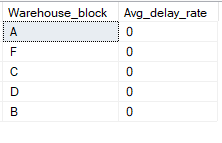
2) -- warehouse block distribution

select Warehouse\_block,COUNT(\*) as shipment\_count from ecom\_data group by Warehouse\_block



3) --Delay rate per warehouse

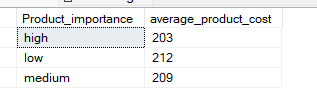
select Warehouse\_block,AVG(1-Reached\_on\_Time\_Y\_N) as Avg\_delay\_rate from ecom\_data group by Warehouse\_block



4) --Avg product cost per importance

select Product\_importance,AVG(Cost\_of\_the\_Product) as average\_product\_cost from ecom\_data

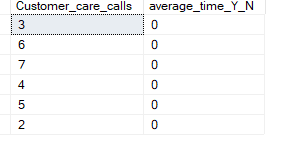
group by Product\_importance



5) --Complaints vs delivery

select Customer\_care\_calls,AVG(Reached\_on\_Time\_Y\_N) as average\_time\_Y\_N from ecom\_data

group by Customer\_care\_calls



6) --Discount vs delay

select Discount\_offered,AVG(1-Reached\_on\_Time\_Y\_N) as average\_discount\_offer from ecom\_data

group by Discount\_offered

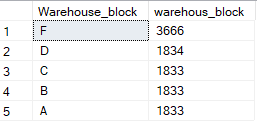
7) ----Analyze the distribution of warehouse blocks:

select Warehouse\_block,COUNT(\*) as warehous\_block

from ecom\_data

group by Warehouse\_block

order by Warehouse\_block desc;



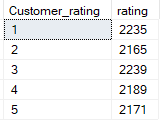
8) ----Customer Rating Distribution

select Customer\_rating,COUNT(\*) as rating

from ecom\_data

group by Customer\_rating

order by Customer\_rating Asc



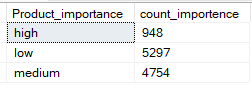
9) ----Product Importance

select Product\_importance,COUNT(\*) as count\_importence

from ecom\_data

group by Product\_importance

order by Product\_importance



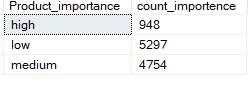
10) -- gender disribution

select Gender,COUNT(\*) as gender\_distribution

from ecom\_data

group by Gender

order by gender\_distribution desc

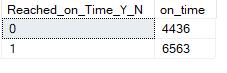


11) -- --On-Time vs. Late Deliveries

select Reached\_on\_Time\_Y\_N,COUNT(\*) as on\_time

from ecom\_data

group by Reached\_on\_Time\_Y\_N

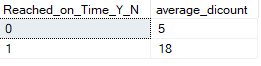


12) ----Average Discount by Delivery Status

select Reached\_on\_Time\_Y\_N,AVG(Discount\_offered) as average\_dicount

from ecom\_data

group by Reached\_on\_Time\_Y\_N

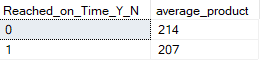


12) ----Average Product Cost by Delivery Status

select Reached\_on\_Time\_Y\_N,AVG(Cost\_of\_the\_Product) as average\_product

from ecom\_data

group by Reached\_on\_Time\_Y\_N



13) --High-Risk Warehouse Blocks

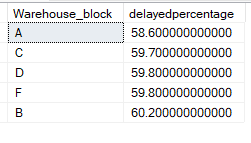
select Warehouse\_block,

round(SUM(Reached\_on\_Time\_Y\_N)\*100.0 /COUNT(\*),1) as delayedpercentage

from ecom\_data

group by Warehouse\_block

order by delayedpercentage



14) --Shipment Mode with Most Delays

select Mode\_of\_Shipment,

round(SUM(Reached\_on\_Time\_Y\_N)\*100.0 /COUNT(\*),1) as shipment\_mode

from ecom\_data

group by Mode\_of\_Shipment

order by shipment\_mode desc



15) --Correlation Between Discounts and Delays

select Reached\_on\_Time\_Y\_N,

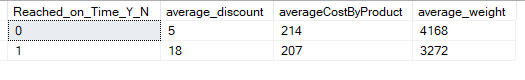
AVG(Discount\_offered) as average\_discount,

AVG(Cost\_of\_the\_Product) as averageCostByProduct,

AVG(Weight\_in\_gms) as average\_weight

from ecom\_data

group by Reached\_on\_Time\_Y\_N



16) --Cost vs. Rating Analysis

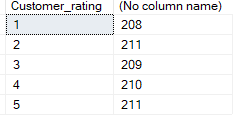
select Customer\_rating,

AVG(Cost\_of\_the\_Product)

from ecom\_data

group by Customer\_rating

order by Customer\_rating



17) --Average Customer Rating by Warehouse

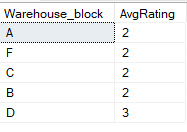
select Warehouse\_block,

AVG(Customer\_rating) as AvgRating

from ecom\_data

group by Warehouse\_block

order by AvgRating



18) --Delays by Shipment Mode

select Mode\_of\_Shipment,

COUNT(\*) NoOfShipment,

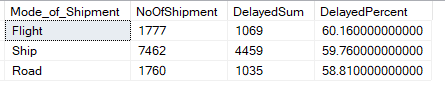
SUM(Reached\_on\_Time\_Y\_N) as DelayedSum,

round(SUM(Reached\_on\_Time\_Y\_N)\*100.0/COUNT(\*),2) as DelayedPercent

from ecom\_data

group by Mode\_of\_Shipment

order by DelayedPercent desc



19) --Correlation Between Customer Care Calls and Delays

select Customer\_care\_calls,

COUNT(\*) as totalorders,

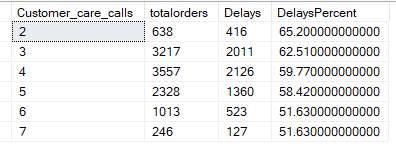
SUM(case when Reached\_on\_Time\_Y\_N=1 then 1 else 0 end) as Delays,

round(SUM(case when Reached\_on\_Time\_Y\_N=1 then 1 else 0 end)\*100.0 /COUNT(\*),2) as DelaysPercent

FROM ecom\_data

group by Customer\_care\_calls

order by Customer\_care\_calls



20) ----Impact of Prior Purchases on Delivery Time

select Prior\_purchases,

COUNT(\*) as totalPrior,

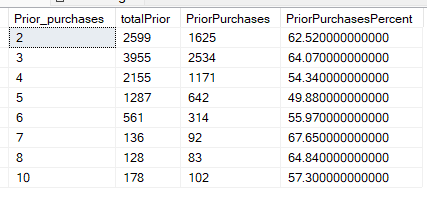
SUM(case when Reached\_on\_Time\_Y\_N=1 then 1 else 0 end) as PriorPurchases,

round(SUM(case when Reached\_on\_Time\_Y\_N=1 then 1 else 0 end)\*100.0/COUNT(\*),2) as PriorPurchasesPercent

from ecom\_data

group by Prior\_purchases

order by Prior\_purchases



21) --Combining Shipment Mode and Product Importance

select

Mode\_of\_Shipment,

Product\_importance,COUNT(\*) as TotalShipments,

SUM(case when Reached\_on\_Time\_Y\_N=1 then 1 else 0 end) as totalDelayed,

round(SUM(case when Reached\_on\_Time\_Y\_N=1 then 1 else 0 end)\*100.0/COUNT(\*),2) as totalDelayedpercent

from ecom\_data

group by Mode\_of\_Shipment,Product\_importance

order by totalDelayedpercent desc

