## horizontal line



Retail Shop Analysis

**─**

Your Name

Daksh Sinha

# Overview

A retail store would like to understand customer behavior using their point of sale data (POS).

# DATA PREPARATION AND UNDERSTANDING

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

→ Transaction table has 2304 rows

→Product category info consists of 24 rows

→Customer table consists of 5648 rows

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

→ Total number of transaction that have a return is 2178

3. As you would have noticed, the dates provided across the datasets are not in a

correct format. As first steps, pls convert the date variables into valid date formats

before proceeding ahead.

→"The Date format has been corrected. The column name is (correct\_tran\_date), and the format is MDY."

4. 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.

→ Extract the number of days, months and years in different columns.

5. Which product category does the sub-category “DIY” belong to?

→Books

# DATA ANALYSIS

1. Which channel is most frequently used for transactions?

→SELECT TOP 1 Store\_type FROM copy\_transcation

GROUP BY Store\_type

ORDER BY COUNT(Store\_type) DESC

2. What is the count of Male and Female customers in the database?

→SELECT Gender, COUNT(Gender) AS CNT FROM copy\_Customer AS CC

WHERE Gender IS NOT NULL

GROUP BY Gender

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

→SELECT TOP 1 city\_code, COUNT(city\_code) AS CNT FROM copy\_Customer AS CC

WHERE city\_code IS NOT NULL

GROUP BY city\_code

ORDER BY CNT DESC

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

→SELECT COUNT(prod\_subcat) AS CNT FROM copy\_prod\_cat\_info

WHERE prod\_cat = 'BOOKS'

GROUP BY prod\_cat

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

→SELECT CPC.prod\_cat, MAX(Qty) FROM copy\_transcation AS CT

JOIN copy\_prod\_cat\_info AS CPC ON CPC.prod\_sub\_cat\_code = CT.prod\_subcat\_code

GROUP BY CPC.prod\_cat

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

→SELECT SUM(CAST(CT.total\_amt AS float)) AS TOTAL\_REVENUE FROM copy\_transcation AS CT

JOIN copy\_prod\_cat\_info AS CPC ON CPC.prod\_sub\_cat\_code = CT.prod\_subcat\_code AND CT.prod\_cat\_code = CPC.prod\_cat\_code

WHERE CPC.prod\_cat = 'ELECTRONICS' OR CPC.prod\_cat = 'BOOKS'

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

→SELECT COUNT(\*) AS CUST\_CNT FROM

(

SELECT cust\_id FROM copy\_transcation

WHERE Status <> 'RETURN'

GROUP BY cust\_id

HAVING COUNT(\*) > 10

) AS T

8. What is the combined revenue earned from the “Electronics” & “Clothing”

categories, from “Flagship stores”?

→SELECT SUM(CT.total\_amt) AS TOTAL\_REVENUE FROM copy\_transcation AS CT

JOIN copy\_prod\_cat\_info AS CPC ON CT.prod\_cat\_code = CPC.prod\_cat\_code AND CPC.prod\_sub\_cat\_code = CT.prod\_subcat\_code

WHERE CT.Store\_type = 'Flagship store' AND ( CPC.prod\_cat = 'Electronics' OR CPC.prod\_cat = 'Clothing' ) AND QTY > 0

9. What is the total revenue generated from “Male” customers in “Electronics”

category? Output should display total revenue by prod sub-cat.

→SELECT CPC.prod\_subcat ,SUM(CAST(CT.total\_amt AS FLOAT)) AS TOTAL\_REVENUE FROM copy\_transcation AS CT

JOIN copy\_prod\_cat\_info AS CPC ON CPC.prod\_sub\_cat\_code = CT.prod\_subcat\_code AND CPC.prod\_cat\_code = CT.prod\_cat\_code

JOIN copy\_Customer AS CC ON CC.customer\_Id = CT.cust\_id

WHERE CPC.prod\_cat = 'Electronics' AND CC.Gender = 'M'

GROUP BY CPC.prod\_subcat

10.What is percentage of sales and returns by product sub category; display only top

5 sub categories in terms of sales?

→SELECT T1.prod\_subcat, PERCENTAGE\_SALES, PERCENTAGE\_RETURNS

FROM

(

SELECT TOP 5 prod\_subcat, (SUM(total\_amt)/(SELECT SUM(total\_amt) FROM copy\_transcation WHERE Qty >0)) AS PERCENTAGE\_SALES

FROM copy\_transcation AS CT

JOIN copy\_prod\_cat\_info AS CPC ON CPC.prod\_cat\_code = CT.prod\_cat\_code AND CPC.prod\_sub\_cat\_code = CT.prod\_subcat\_code

WHERE Qty > 0

Group by prod\_subcat

ORDER BY PERCENTAGE\_SALES DESC

) AS T1

JOIN

(

SELECT prod\_subcat, (SUM(total\_amt)/(SELECT SUM(total\_amt) FROM copy\_transcation WHERE Qty < 0)) AS PERCENTAGE\_RETURNS

FROM copy\_transcation AS CT

JOIN copy\_prod\_cat\_info AS CPC ON CPC.prod\_cat\_code = CT.prod\_cat\_code AND CPC.prod\_sub\_cat\_code = CT.prod\_subcat\_code

WHERE Qty < 0

Group by prod\_subcat

) AS T2

ON T1.prod\_subcat = T2.prod\_subcat

11. 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?

→SELECT cust\_id,DATEDIFF(YY,DOB,(SELECT MAX(correct\_tran\_date) FROM copy\_transcation)) AS AGE,

SUM(total\_amt) AS TOTAL\_REVENUE, correct\_tran\_date, Qty FROM copy\_Customer AS CC

JOIN copy\_transcation AS CT ON CT.cust\_id = CC.customer\_Id

WHERE correct\_tran\_date IN

(

SELECT correct\_tran\_date FROM copy\_transcation

GROUP BY correct\_tran\_date, cust\_id

HAVING correct\_tran\_date>= DATEADD(DD,-30, (SELECT MAX(correct\_tran\_date) FROM copy\_transcation))

) AND Qty >0

GROUP BY cust\_id, DOB, correct\_tran\_date, Qty

HAVING DATEDIFF(YY,DOB,MAX(correct\_tran\_date)) BETWEEN 25 AND 35

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

transactions?

→SELECT TOP 1 CPC.prod\_cat FROM copy\_transcation AS CT

JOIN copy\_prod\_cat\_info AS CPC ON CPC.prod\_cat\_code = CT.prod\_cat\_code

WHERE correct\_tran\_date >= DATEADD(MONTH,-3,(SELECT MAX(correct\_tran\_date) FROM copy\_transcation)) AND Status = 'RETURN'

GROUP BY CPC.prod\_cat

ORDER BY COUNT( CPC.prod\_cat) DESC

13.Which store-type sells the maximum products; by value of sales amount and by

quantity sold?

→SELECT TOP 1 Store\_type FROM copy\_transcation

GROUP BY Store\_type

ORDER BY SUM(Qty) DESC, SUM(total\_amt) DESC

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

→SELECT CT.prod\_cat\_code, CPC.prod\_cat ,AVG(total\_amt) AS AVG\_ FROM copy\_transcation AS CT

JOIN copy\_prod\_cat\_info AS CPC ON CPC.prod\_cat\_code = CT.prod\_cat\_code

WHERE Qty > 0

GROUP BY CT.prod\_cat\_code, CPC.prod\_cat

HAVING AVG(total\_amt) >= (SELECT AVG(total\_amt) FROM copy\_transcation WHERE Qty > 0 )

ORDER BY CT.prod\_cat\_code

15. Find the average and total revenue by each subcategory for the categories which

are among top 5 categories in terms of quantity sold.

→SELECT CPC.prod\_cat,CPC.prod\_sub\_cat\_code, CPC.prod\_subcat, AVG(total\_amt) AS AVG\_, SUM(total\_amt) AS TOTAL\_REV FROM copy\_transcation AS CT

JOIN copy\_prod\_cat\_info AS CPC ON CPC.prod\_sub\_cat\_code = CT.prod\_subcat\_code AND CPC.prod\_cat\_code = CT.prod\_cat\_code

WHERE CPC.prod\_cat\_code IN

(

SELECT TOP 5 prod\_cat\_code FROM copy\_transcation

WHERE Qty > 0

GROUP BY prod\_cat\_code

ORDER BY SUM(Qty) DESC

) AND Qty > 0

GROUP BY CPC.prod\_cat,CPC.prod\_sub\_cat\_code, CPC.prod\_subcat

ORDER BY CPC.prod\_sub\_cat\_code