**RFM ANALYSIS**

**/\* RFM analysis, is a type of customer segmentation and behavioral targeting used to**

**help businesses rank and segment customers based on the recency,**

**frequency, and monetary value of a transaction.**

**— mailchimp.com/resources/rfm-analysis/ \*/**

select \*

from customer\_shopping\_data;



**-- inspect data**

**-- check how many records**

select count(\*)

from customer\_shopping\_data;



**-- customer id has to be unique, check if it's correct**

select count(distinct(customer\_id))

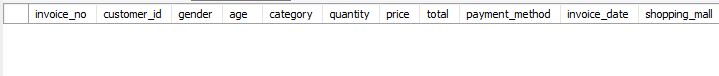
from customer\_shopping\_data;



**-- check if there are null values**

select \*

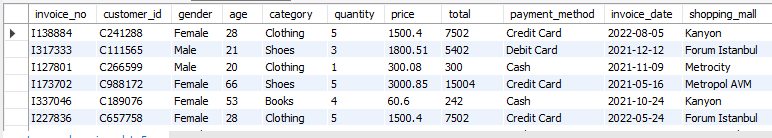
from customer\_shopping\_data where shopping\_mall is null;



**-- let's add the total column**

ALTER TABLE customer\_shopping\_data ADD COLUMN total INT AFTER price;

UPDATE customer\_shopping\_data SET total = price \* quantity;



**-- spending by gender**

select gender, count(distinct(invoice\_no)) orders, round(sum(total),2) sales

from customer\_shopping\_data

group by gender

order by 3 desc;



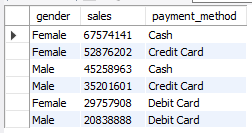
**-- most popular payment method among women and men**

select gender, round(sum(total),2) sales, payment\_method

from customer\_shopping\_data

group by gender, payment\_method

order by sales desc;



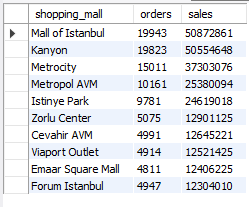
**-- spending by mall**

select shopping\_mall, count(distinct(invoice\_no)) orders, ROUND(sum(total),2) sales

from customer\_shopping\_data

group by shopping\_mall

order by sales desc;



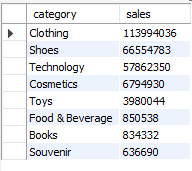
**-- spending by category**

select category , round(sum(total),2) sales

from customer\_shopping\_data

group by category

order by sales desc;



**-- spending by customer age groups**

SELECT

CASE

WHEN age < 20 THEN 'Under 20'

WHEN age BETWEEN 20 and 29 THEN '20 - 29'

WHEN age BETWEEN 30 and 39 THEN '30 - 39'

WHEN age BETWEEN 40 and 49 THEN '40 - 49'

WHEN age BETWEEN 50 and 59 THEN '50 - 59'

WHEN age BETWEEN 60 and 69 THEN '60 - 69'

WHEN age BETWEEN 70 and 79 THEN '70 - 79'

WHEN age >= 80 THEN 'Over 80'

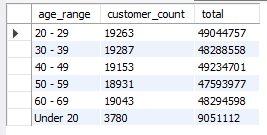
END as age\_range,

COUNT(\*) AS customer\_count, sum(total) as total

FROM customer\_shopping\_data

GROUP BY age\_range

ORDER BY age\_range,total desc;



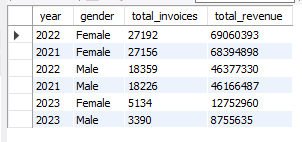
**-- top spending by year, gender wise**

select distinct(year(invoice\_date)) year, gender, count(distinct(invoice\_no)) total\_invoices, round(sum(total),0) total\_revenue

from customer\_shopping\_data

group by year, gender

order by total\_revenue desc;



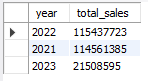
**-- top sales by year**

select distinct(year(invoice\_date)) as year, sum(total) total\_sales

from customer\_shopping\_data

group by year

order by total\_sales desc;



**-- top sales by months, year=2021**

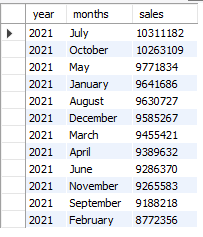
select year(invoice\_date) as year, monthname(invoice\_date) as months, sum(total) as sales

from customer\_shopping\_data

group by year,months

having year=2021

order by sales desc;



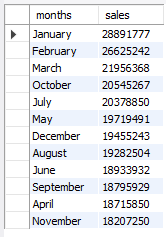
**-- top sales by months**

select monthname(invoice\_date) as months, sum(total) as sales

from customer\_shopping\_data

group by months

order by sales desc;



**-- different days between the last customer's purchase**

select

customer\_id, gender, age, payment\_method, shopping\_mall,

datediff('2023-03-10', invoice\_date) last\_date\_order,

count(distinct(invoice\_no)) total\_orders,

sum(price) spending

from customer\_shopping\_data

group by customer\_id, gender, age, payment\_method, shopping\_mall, invoice\_date

order by last\_date\_order;

with rfm as (

select

customer\_id, gender, age, payment\_method, shopping\_mall,

datediff('2023-03-08', invoice\_date) last\_date\_order,

count(distinct(invoice\_no)) total\_orders,

sum(price) spending

from customer\_shopping\_data

group by customer\_id, gender, age, payment\_method, shopping\_mall, invoice\_date

order by last\_date\_order

)

select \*,

ntile(3) over (order by last\_date\_order) rfm\_recency,

ntile(3) over (order by total\_orders) rfm\_frequency,

ntile(3) over (order by spending) rfm\_monetary

from rfm;

with rfm as (

select

customer\_id, gender, age, payment\_method, shopping\_mall,

datediff('2023-03-08', invoice\_date) last\_date\_order,

count(distinct(invoice\_no)) total\_orders,

sum(price) spending

from customer\_shopping\_data

group by customer\_id, gender, age, payment\_method, shopping\_mall, invoice\_date

order by last\_date\_order

),

rfm\_calc as (

select \*,

ntile(3) over (order by last\_date\_order) rfm\_recency,

ntile(3) over (order by total\_orders) rfm\_frequency,

ntile(3) over (order by spending) rfm\_monetary

from rfm

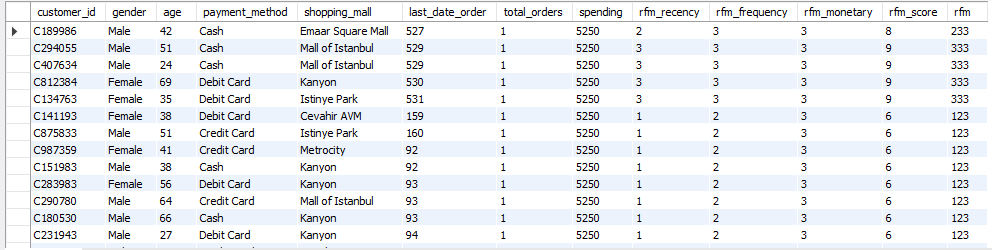
order by rfm\_monetary desc

)

select \*, rfm\_recency + rfm\_frequency + rfm\_monetary as rfm\_score,

concat(rfm\_recency, rfm\_frequency, rfm\_monetary) as rfm

from rfm\_calc;



**-- RFM**

select \*, case

when rfm in (311, 312, 311) then 'new customers'

when rfm in (111, 121, 131, 122, 133, 113, 112, 132) then 'lost customers'

when rfm in (212, 313, 123, 221, 211, 232) then 'regular customers'

when rfm in (223, 222, 213, 322, 231, 321, 331) then 'loyal customers'

when rfm in (333, 332, 323, 233) then 'champion customers'

end rfm\_segment

from

(

with rfm as (

select

customer\_id, gender, age, payment\_method, shopping\_mall,

datediff('2023-03-08', invoice\_date) last\_date\_order,

count(distinct(invoice\_no)) total\_orders,

sum(price) spending

from customer\_shopping\_data

group by customer\_id, gender, age, payment\_method, shopping\_mall, invoice\_date

order by last\_date\_order

),

rfm\_calc as (

select \*,

ntile(3) over (order by last\_date\_order) rfm\_recency,

ntile(3) over (order by total\_orders) rfm\_frequency,

ntile(3) over (order by spending) rfm\_monetary

from rfm

order by rfm\_monetary desc

)

select \*, rfm\_recency + rfm\_frequency + rfm\_monetary as rfm\_score,

concat(rfm\_recency, rfm\_frequency, rfm\_monetary) as rfm

from rfm\_calc

) rfm\_tb;

