**PROJECT-1 SQL**

Query 1

Create database project1\_sales\_store

Query 2

create table sales\_store(

transaction\_id text,

customer\_id text,

customer\_name text,

customer\_age text,

gender text,

product\_id text,

product\_name text,

product\_category text,

quantiy text,

prce text,

payment\_mode text,

purchase\_date text,

time\_of\_purchase text,

status text

);

Query3—import csv file/data--

copy sales\_store(transaction\_id,

customer\_id, customer\_name, customer\_age,

gender, product\_id, product\_name, product\_category,

quantiy, prce, payment\_mode, purchase\_date, time\_of\_purchase, status

)

from 'G:\SQL\project1\_sales\_store.csv'

delimiter ','

csv header;

Query4—change data type text to int, varchar etc—

alter table sales\_store

alter column customer\_age type integer

using customer\_age:: integer

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

alter table sales\_store

alter column customer\_name type varchar(80)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

alter table sales\_store

alter column customer\_id type varchar

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

alter table sales\_store

alter column transaction\_id type varchar

alter table sales\_store\_copy

add constraint sales\_store\_copy\_pkey

primary key (transaction\_id);

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

alter table sales\_store

alter column gender type varchar

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

alter table sales\_store

alter column product\_id type varchar

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

alter table sales\_store

alter column product\_name type varchar

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

alter table sales\_store

alter column product\_category type varchar

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

alter table sales\_store

alter column quantiy type integer

using quantiy:: integer;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

alter table sales\_store

alter column prce type float

using prce:: float;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

alter table sales\_store

alter column payment\_mode type varchar(80)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

alter table sales\_store

alter column purchase\_date type date

using purchase\_date:: date;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

alter table sales\_store

alter column time\_of\_purchase type time

using time\_of\_purchase:: time;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

alter table sales\_store

alter column status type varchar(80)

Query5– make copy of data set—

select \* from sales\_store –original table—

--duplicate/copy--

select \* into sales\_store\_copy from sales\_store

select \* from sales\_store\_copy

Query6–data cleaning steps—

--step 1--check duplicates and delete duplicates--

select transaction\_id , count(\*)

from sales\_store\_copy

group by transaction\_id

having count(transaction\_id)>1

--'TXN855235'

--'TXN240646'

--'TXN342128'

--'TXN981773'

--anotgher detailed way-

begin;

with mycte as (select \*, ctid,

row\_number() over(partition by transaction\_id order by transaction\_id) as rownum

from sales\_store\_copy)

delete from sales\_store\_copy as c

using mycte

where c.ctid=mycte.ctid

and mycte.rownum>1;

-- select \* from mycte

-- where rownum>1

--step 2--correction of headings--wrong spelled--

alter table sales\_store\_copy

rename column quantiy to quantity

alter table sales\_store\_copy

rename column prce to price

--step 3--to check data type--

select column\_name, data\_type

from information\_schema.columns

where table\_name='sales\_store\_copy'

--step 4--to check nulls and treat nulls--

select \* from sales\_store\_copy

where transaction\_id is null

or customer\_id is null

or customer\_name is null

or customer\_age is null

or gender is null

or product\_id is null

or product\_name is null

or product\_category is null

or quantity is null

or price is null

or payment\_mode is null

or purchase\_date is null

or time\_of\_purchase is null

or status is null

--transaction\_id--

begin;

delete from sales\_store\_copy

where transaction\_id is null

rollback;

commit;

--customer\_id--

select \* from sales\_store\_copy

where customer\_name = 'Ehsaan Ram'

update sales\_store\_copy

set customer\_id= 'CUST9494'

where transaction\_id= 'TXN977900'

-- \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

select \* from sales\_store\_copy

where customer\_name= 'Damini Raju'

update sales\_store\_copy

set customer\_id='CUST1401'

where transaction\_id='TXN985663'

--customer\_name--

select \* from sales\_store\_copy

where customer\_id= 'CUST1003'

update sales\_store\_copy

set customer\_name='Mahika Saini', customer\_age=35, gender='Male'

where transaction\_id='TXN432798'

--step 5--data cleaning--

select distinct gender

from sales\_store\_copy

update sales\_store\_copy

set gender= 'M'

where gender='Male'

update sales\_store\_copy

set gender='F'

where gender='Female'

-- \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

select distinct product\_name

from sales\_store\_copy

--\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

select distinct payment\_mode

from sales\_store\_copy

update sales\_store\_copy

set payment\_mode= 'Credit Card'

where payment\_mode='CC'

Query7--solving business insights questions--

--data analysis--

--q1 top 5 most selliing products by quantity—

-- business problem= which products are most in demand--

--impact= helps prioritize stock and boost sales through targeted promotions--

select distinct status

from sales\_store\_copy

select product\_name, sum(quantity)as total\_quantity

from sales\_store\_copy

where status='delivered'

group by product\_name

order by total\_quantity desc

limit 5

--q2 which products are frequently cancelled--

--business problem= frequently cancellations affect revenue and cust trust--

--impact=identify poor-performing product to improve quality or remove from catlog--

select product\_name, count(status) as freq\_cancelled

from sales\_store\_copy

where status ='cancelled'

group by product\_name

order by freq\_cancelled desc

limit 5

--q3 what time of the day has the highest no of purchaces--

--business problem= find pick sales time--

--impact=optimize staffing, promotions and server loads --

select

case

when extract(hour from time\_of\_purchase) between 0 and 5 then 'Night'

when extract(hour from time\_of\_purchase) between 6 and 12 then 'Morning'

when extract(hour from time\_of\_purchase) between 12 and 17 then 'Afternoon'

when extract(hour from time\_of\_purchase) between 17 and 24 then 'Evening'

end as time\_of\_day,

count(\*) as total\_order

from sales\_store\_copy

group by

case

when extract(hour from time\_of\_purchase) between 0 and 5 then 'Night'

when extract(hour from time\_of\_purchase) between 6 and 12 then 'Morning'

when extract(hour from time\_of\_purchase) between 12 and 17 then 'Afternoon'

when extract(hour from time\_of\_purchase) between 17 and 24 then 'Evening'

end

order by total\_order desc

--q4 top 5 highest spending customer--

--business problem= identify VIP cust--

--impact= personalized offeres, loyalty rewards and retention--

select customer\_id, customer\_name,

'$'||to\_char(sum(price\*quantity),'999,99,99,999')as highest\_spent

from sales\_store\_copy

where status= 'delivered'

group by customer\_id, customer\_name

order by highest\_spent desc

limit 5

--q5 highest revenue generating product\_category--

--business problem=identify top performing category--

--impact= refine product strategy, supply chain and promotions,allows bsiness to invest more in specific categories --

select product\_category,

'$'||to\_char(sum(price\*quantity), '999,99,99,999') as highest\_revenue

from sales\_store\_copy

group by product\_category

order by highest\_revenue desc

--q6 what is the returned/cancellation rate per preoduct category --

----business problem=monitor disatisfaction trend per category

--impact= reduce returns, improve product discriptions/expectations,helps identify and fix product issues--

select product\_category,

round(count(case when status='returned' then 1 end)\*100.0/count(\*), 2)|| '%'as r\_rate

from sales\_store\_copy

group by product\_category

order by r\_rate desc

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

select product\_category,

round(count(case when status='cancelled' then 1 end)\*100.0/count(\*), 2) || '%'as c\_rate

from sales\_store\_copy

group by product\_category

order by c\_rate desc

--q7most preferred payment mode --

----business problem= which payment option customer preffer

--impact= stremline payment processing, prioterize payment mode --

select payment\_mode,count(payment\_mode) AS pmode

from sales\_store\_copy

group by payment\_mode

order by pmode desc

--q8 how does age group affect business behaviour--

----business problem= understand cust demographics/understand most spoending age group

--impact=target that group to generate revenue--

select max(customer\_age), min(customer\_age)

from sales\_store\_copy

select

case

when customer\_age between 18 and 25 then 'teenagers'

when customer\_age between 26 and 35 then 'youngster'

when customer\_age between 36 and 50 then 'adult'

else 'senior'

end as agegroup,

'₹'||to\_char(sum(price\*quantity), '999,99,99,999') as total\_purchase

from sales\_store\_copy

group by agegroup

order by total\_purchase desc

--q9 monthly sales trend--

--business problemsa=sales fluctuation go unnoticed

--impact=plan inventory and marketing according to season--

select

extract(year from purchase\_date) as year,

extract(month from purchase\_date) as month,

'₹'||to\_char(sum(price\*quantity):: numeric, 'FM999,99,99,999') as totalsale

from sales\_store\_copy

group by year,month

order by year,month asc

--q10 certain gender buying specific product category--

--business problem= GENDER BASED PRODUCT PREFERENCE--

--impact= personalized ads, gender focus campaign to increase sales--

select gender, product\_category, count(product\_category)as countp

from sales\_store\_copy

group by gender, product\_category

order by gender