SQL Queries

1.SELECT \* FROM raw\_data.`finaldata.csv`;

--->This query is used to fetch all the columns from the data.

2. select \* from raw\_data.`finaldata.csv` limit 5;

--->Used to fetch the first five rows of the data.

3. SELECT COUNT(\*) as number\_of\_rows FROM raw\_data.`finaldata.csv`;

--->Used to fetch the total number of rows present in the data.

4 .SELECT COUNT(DISTINCT Patient\_ID) AS unique\_patients FROM

raw\_data.`finaldata.csv`;

🡪It fetches total number of unique patients from the data.

5. select

count(case when Typeofsales ="" then 1 end ) as Null\_values\_in\_Typesofsales,

count(case when Patient\_ID ="" then 1 end ) as Null\_values\_in\_patientid,

count(case when Specialisation ="" then 1 end ) asNull\_values\_in\_specialisation,

count(case when Dept ="" then 1 end ) as Null\_values\_in\_dept,

count(case when Dateofbill ="" then 1 end ) as Null\_values\_in\_dateofbill,

count(case when Quantity ="" then 1 end ) as Null\_values\_in\_quantity,

count(case when ReturnQuantity ="" then 1 end ) as Null\_values\_in\_returnquantity,

count(case when Final\_Cost ="" then 1 end ) as Null\_values\_in\_finalcost,

count(case when Final\_Sales ="" then 1 end ) as Null\_values\_in\_finalsales,

count(case when RtnMRP ="" then 1 end ) as Null\_values\_in\_rtnmrp,

count(case when Formulation ="" then 1 end ) as Null\_values\_in\_formulation,

count(case when DrugName ="" then 1 end ) as Null\_values\_in\_drugname,

count(case when SubCat ="" then 1 end ) as Null\_values\_in\_subcat,

count(case when SubCat1 ="" then 1 end ) as Null\_values\_in\_subcat1

from raw\_data.`finaldata.csv`;

---> It fetches the count of null values present in each column

6. SELECT

Specialisation, AVG(Quantity) AS avg\_Quantity,

AVG(ReturnQuantity) AS avg\_ReturnQuantity,

AVG(Final\_Cost) AS avg\_Final\_Cost,

AVG(Final\_Sales) AS avg\_Final\_Sales,

AVG(RtnMRP) AS avg\_RtnMRP

FROM raw\_data.`finaldata.csv`

GROUP BY Specialisation order by Final\_Sales limit 5 ;

-🡪This query fetches the avg quantity, avg returns , average final cost final return MRP by each specialisation.

7. update raw\_data.`finaldata.csv` set Dateofbill=replace(Dateofbill,'/','-');

-🡪This query replaces the / by – in the date of bill column.

8. SELECT DATE\_FORMAT(Dateofbill, '%d-%m-%y') AS Formatted\_Date

FROM raw\_data.`finaldata.csv` ;

🡪 It changes the date of bill into date-moth-year format.

9. select DrugName , avg(Quantity) ,

avg(ReturnQuantity) ,

avg(Final\_Cost),

avg(Final\_Sales)

from raw\_data.`finaldata.csv` group by DrugName

order by Final\_Sales limit 5;

-🡪 It fetches the top 5 selling medicines along with their average quantity buyed, average final cost and average final sales.

10. SELECT \* FROM raw\_data.`finaldata.csv` WHERE Final\_Cost > 100;

-🡪 It fetches the all the transactions with the final cost >100.

11. SELECT SubCat, COUNT(\*) AS count

FROM raw\_data.finaldata

GROUP BY SubCat

ORDER BY count DESC LIMIT 5;

🡪 It gives top 5 sub categories with the highest count .

12. SELECT

DrugName,

SUM(Quantity) AS total\_quantity,

SUM(ReturnQuantity) AS total\_return\_quantity

FROM

raw\_data.`finaldata.csv`

GROUP BY DrugName;

-🡪 I fetches the drug name along with the quantity buyed and quantity returned.

13. SELECT

DrugName, SUM(ReturnQuantity) AS total\_return\_quantity

FROM

raw\_data.`finaldata.csv`

GROUP BY DrugName

ORDER BY total\_return\_quantity DESC

LIMIT 5;

🡪It fetches the Top 5 drug names with highest return quantity.

14. SELECT COUNT(DISTINCT Dept) AS unique\_depts FROM raw\_data.`finaldata.csv`;

🡪 It fetches the count of distinct departments

15. SELECT

Specialisation, AVG(Final\_Cost) AS avg\_final\_cost

FROM

raw\_data.`finaldata.csv`

GROUP BY Specialisation

🡪It gives the top 5 specialisations with highest average cost.