```
SQL Document:
Creating the table and importing content into the table
/* Creating Table */
Create table MIO(
       Typeofsales varchar,
       Patient_ID Bigint,
       Specialisation varchar,
       Dept varchar,
       Dateofbill varchar,
       Quantity int,
       ReturnQuantity int,
       Final_Cost DOUBLE PRECISION,
       Final_Sales DOUBLE PRECISION,
       RtnMRP DOUBLE PRECISION,
       Formulation varchar,
       DrugName varchar,
       SubCat varchar,
       SubCat1 varchar
)
To import data into the created MIO table. Right Click on table Name on left side of panel -> select
import and export Data -> choose the csv file to import -> cross check the column names in column tab
-> Click ok
Selecting or Viewing data from the table:
select * from MIO
```



Total Data in dataset before cleaning 14218. The data consists of null values in some fields.

Creating view: Creating view to do feature engineering. We are using this view for further Analysis

--Create view from MIO table

create view MioView as

select * from MIO



Data Preprocessing:

Describing Dataset:

/*Describing Dataset */

SELECT

column_name,

data_type,

character_maximum_length,

is_nullable

FROM

information_schema.columns

WHERE

table_name = 'mio' AND table_schema = 'public'

	column_name name	data_type character varying	character_maximum_length integer	is_nullable character varying (3)
1	typeofsales	character varying	[null]	YES
2	patient_id	bigint	[null]	YES
3	specialisation	character varying	[null]	YES
4	dept	character varying	[null]	YES
5	dateofbill	character varying	[null]	YES
6	quantity	integer	[null]	YES
7	returnquantity	integer	[null]	YES
8	final_cost	double precision	[null]	YES
9	final_sales	double precision	[null]	YES
10	rtnmrp	double precision	[null]	YES
11	formulation	character varying	[null]	YES
Tota	l rows: 14 of 14	Query complete 00	0:00:00.069	

Checking duplicate values in patient id fields:

select Dateofbill,patient_id, final_sales from MioVieworder by patient_id;

= + ſ		1 • • • • • • • • • • • • • • • • • • •	
	dateofbill character varying	patient_id bigint	final_sales double precision
1	08-06-2022	12017998218	447.2
2	02-05-2022	12017998218	81
3	7/15/2022	12017998261	160.51
4	11-08-2022	12017998278	43.33
5	11-09-2022	12017998278	415.89
6	12/16/2022	12017998291	0
7	5/29/2022	12017998321	63.9
8	7/28/2022	12017998321	138.63
9	5/15/2022	12017998321	331
10	5/27/2022	12017998321	173.332
11	5/22/2022	12017998321	0
Total	rows: 1000 of 14218	Query com	plete 00:00:00.085

Checking Null values: checking if the column contains null values

-- checking null values in a column

select * from MioView where patient_id is NULL

select * from MioView where specialisation is NULL

select * from MioView where dept is NULL

select * from MioView where dateofbill is NULL

select * from MioView where quantity is NULL

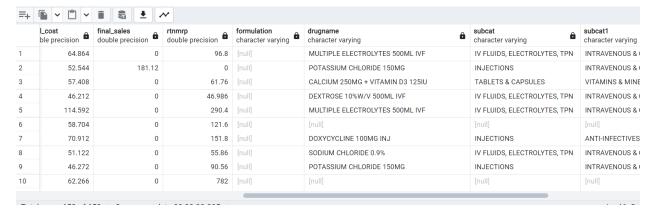
select * from MioView where returnquantity is NULL

select * from MioView where final cost is NULL

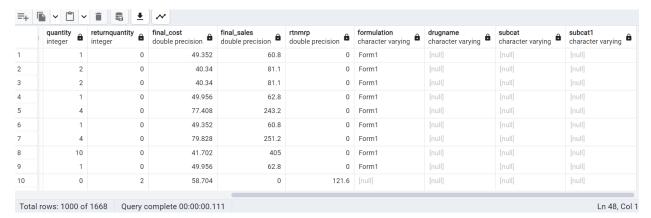
select * from MioView where final sales is NULL

select * from MioView where rtnmrp is NULL

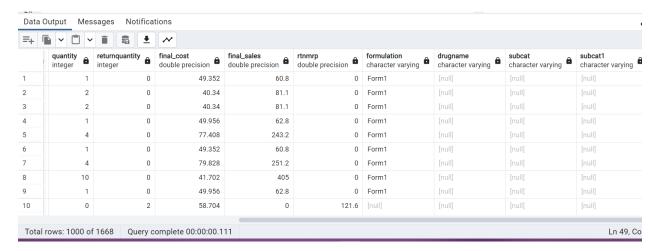
select * from MioView where formulation is NULL -- contains null value



select * from MioView where drugname is NULL -- contains null values



select * from MioView where subcat is NULL -- contains null values



select * from MioView where subcat1 is NULL -- contains null values



From the above queries we can see Columns formulation, Drugname, subcat, subcat1 contains null value so we need to replace the null value with NA Value so it won't affect the remaining data while performing EDA

Replacing Null Values with NA

-- replace NUII values in Formulation column

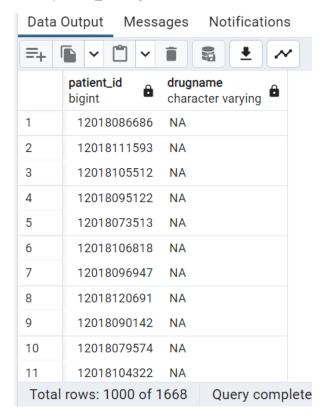
Update MioView set formulation = COALESCE(formulation, 'NA');

-- replace NUII values in drugname, subcat, subcat1 column

Update MioView

--selecting Patient_id columns where durgname is NA

select patient_id,drugname from MioView where drugname = 'NA'



Counting NA Values from columns:

select

from MioView

count(CASE WHEN formulation = 'NA' then 1 END) as Formulation_NA_Count, count(CASE WHEN drugname = 'NA' then 1 END) as Durgname_NA_Count, count(CASE WHEN subcat = 'NA' then 1 END) as Subcat_NA_count, count(case when subcat1 = 'NA' then 1 END) as Subcat1_NA_count

	formulation_na_count bigint	durgname_na_count bigint	subcat_na_count bigint	subcat1_na_count bigint
1	653	1668	1668	1692

From the above table formulation contains 653 null values and drugname and subcat both contains 1668 null values each and subcat1 contains 16692

Data Formatting:

From the observation in date column, the format is in consistant. 08-06-2022 and 7/15/2022 are the two formats we can see in the date column. So we are changing the format uniformly 7/15/2022

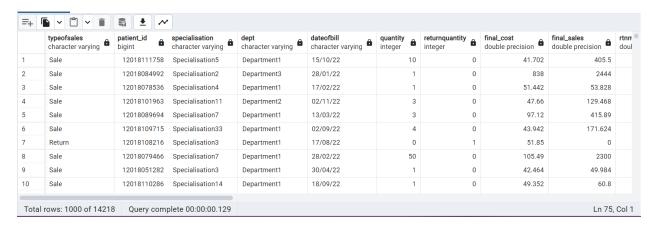
-- Formating date format

Update MioView

Set dateofbill = TO CHAR(TO DATE(dateofbill, 'MM/DD/YY'), 'DD/MM/YY')

WHERE dateofbill IS NOT NULL AND dateofbill != ";

select * from MioView;



Feature engineering:

Create column with month name from date and add that column and remaning cleaned data from MioView to other view name MioView1

-- adding a new column in other view

create view MioView1 as

select *,To_CHAR(to_date(dateofbill,'dd/mm/yy'),'Mon') as month from MioView

--selecting dateofbill and month values from MioView1

select dateofbill, month from MioView1

	dateofbill character varying		onth ext	â
1	15/10/22	C	ct	
2	28/01/22	J	an	
3	17/02/22	F	eb	
4	02/11/22	N	lov	
5	13/03/22	N	1ar	
6	02/09/22	S	ер	
7	17/08/22	A	ug	
8	28/02/22	F	eb	
9	30/04/22	A	pr	
10	18/09/22	S	Sep	
11	01/08/22	A	ug	
Total	Total rows: 1000 of 14218		Que	ry co