**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,**

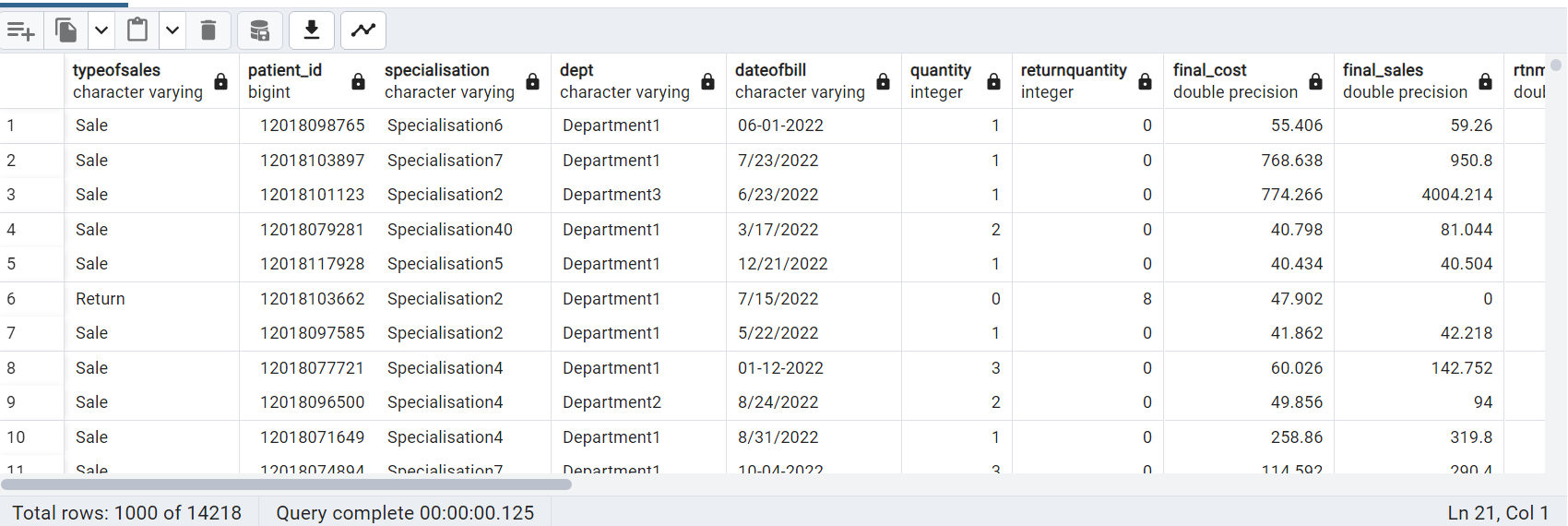
**SubCat1 varchar**

**)**

**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**

****

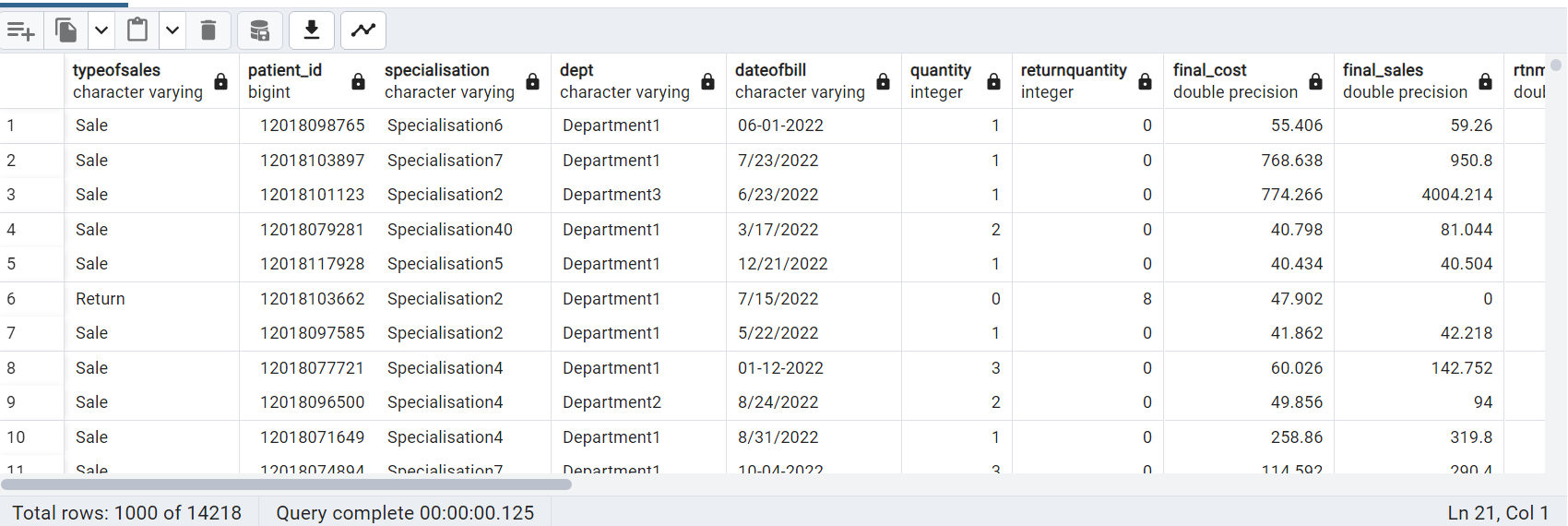
Total Data in dataset before cleaning 14218. The data consists of null values in some fields.

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,

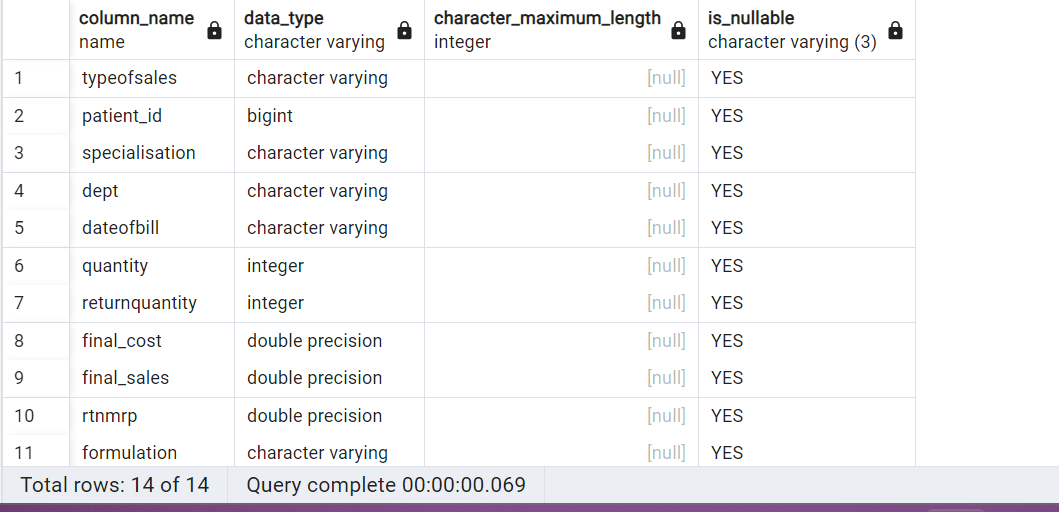
is\_nullable

FROM

information\_schema.columns

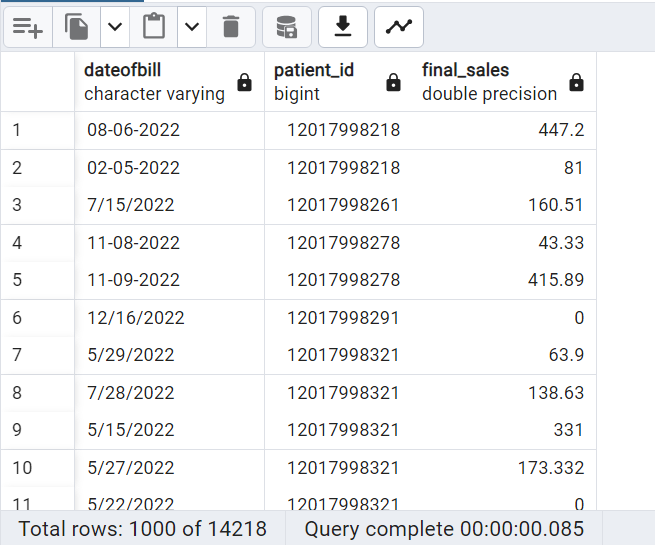
WHERE

table\_name = 'mio' AND table\_schema = 'public'



**Checking duplicate values in patient id fields:**

select Dateofbill,patient\_id, final\_sales from MioVieworder by patient\_id;



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

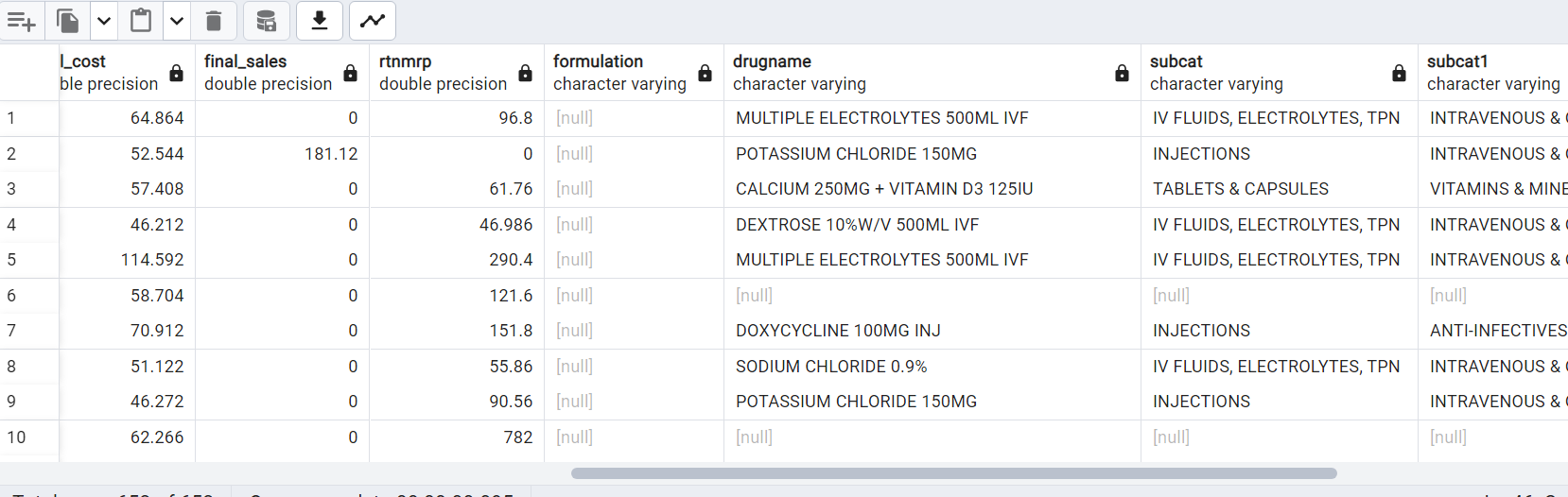
select \* from MioView where returnquantity is NULL

select \* from MioView where final\_cost is NULL

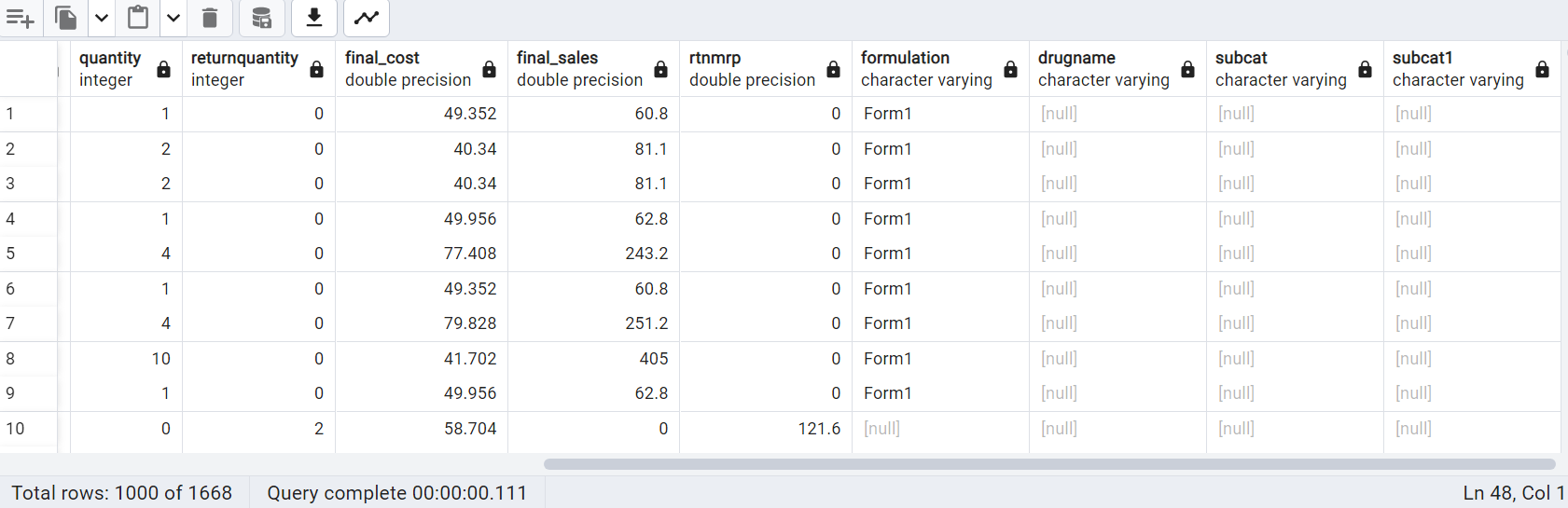
select \* from MioView where final\_sales is NULL

select \* from MioView where rtnmrp is NULL

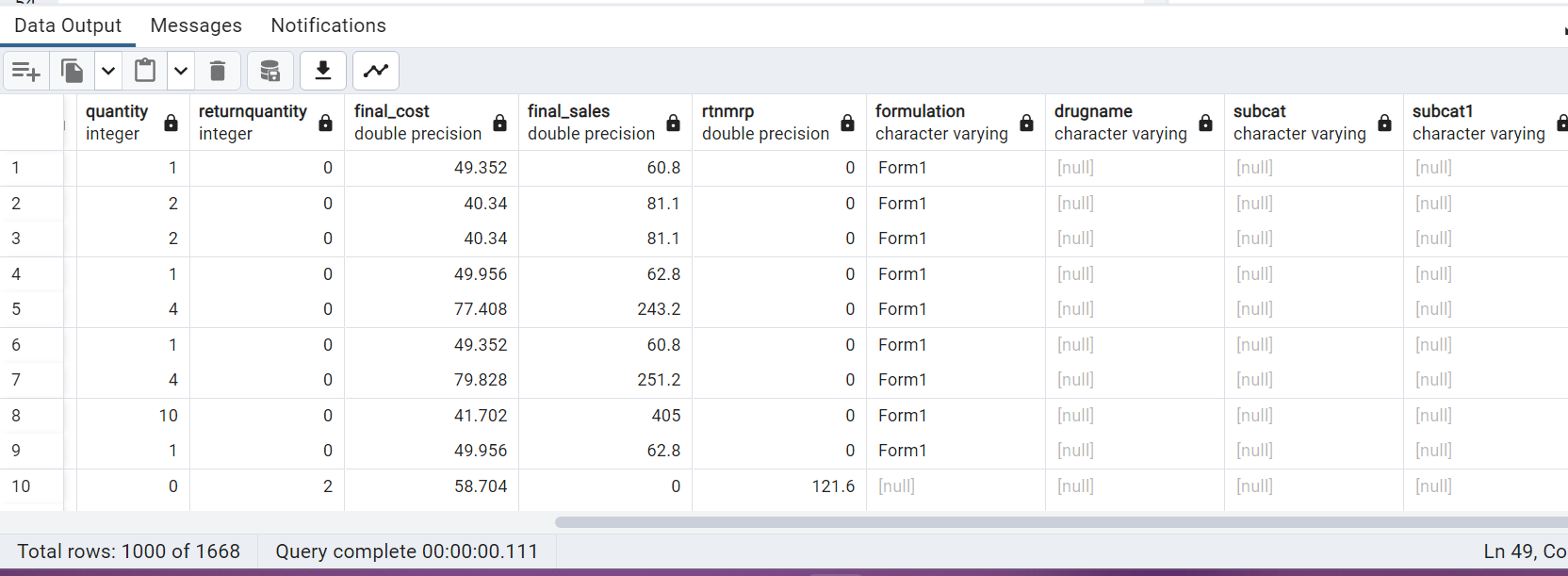
select \* from MioView where formulation is NULL -- contains null value



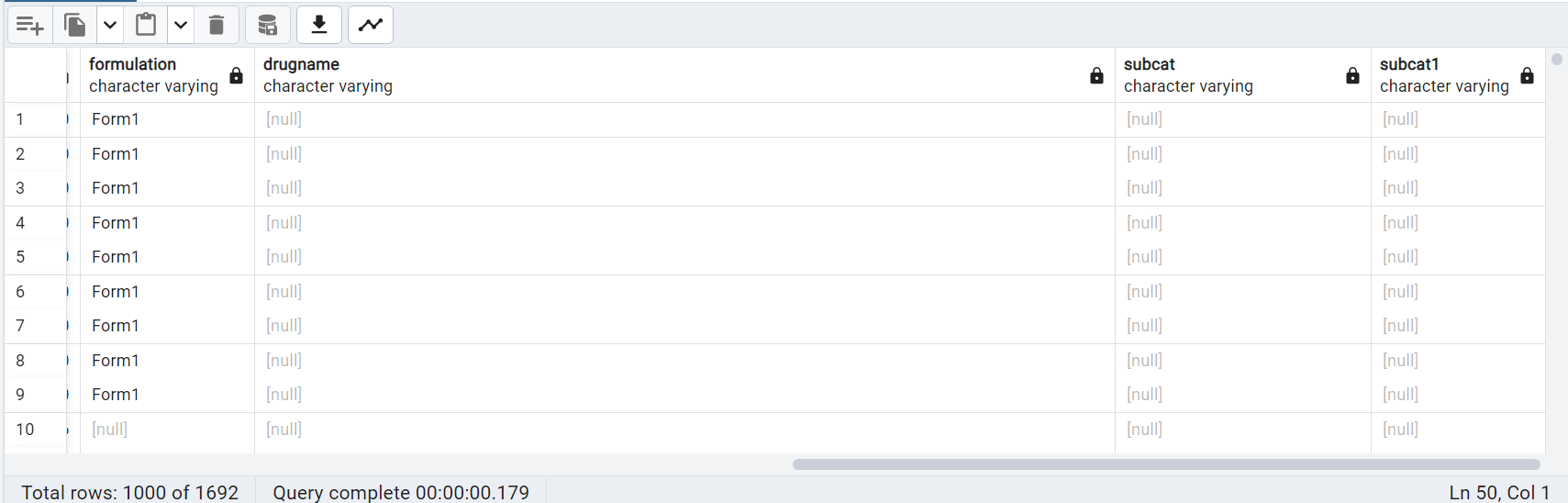
select \* from MioView where drugname is NULL -- contains null values



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 NUll values in Formulation column

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

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

Update MioView

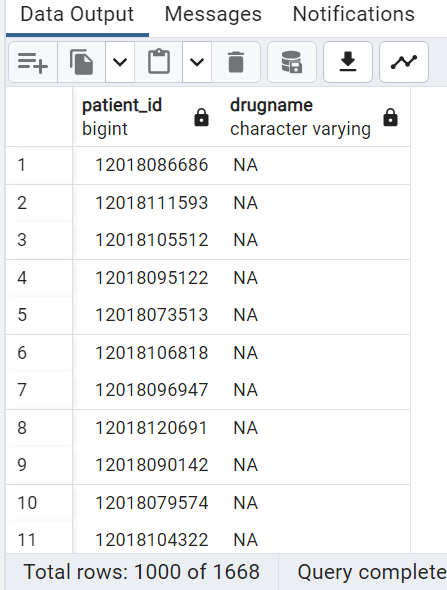
set drugname = COALESCE(drugname, 'NA'),

subcat = COALESCE(subcat,'NA'),

subcat1 = COALESCE(subcat1,'NA');

--selecting Patient\_id columns where durgname is NA

select patient\_id,drugname from MioView where drugname = 'NA'



Counting NA Values from columns:

select

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

from MioView



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

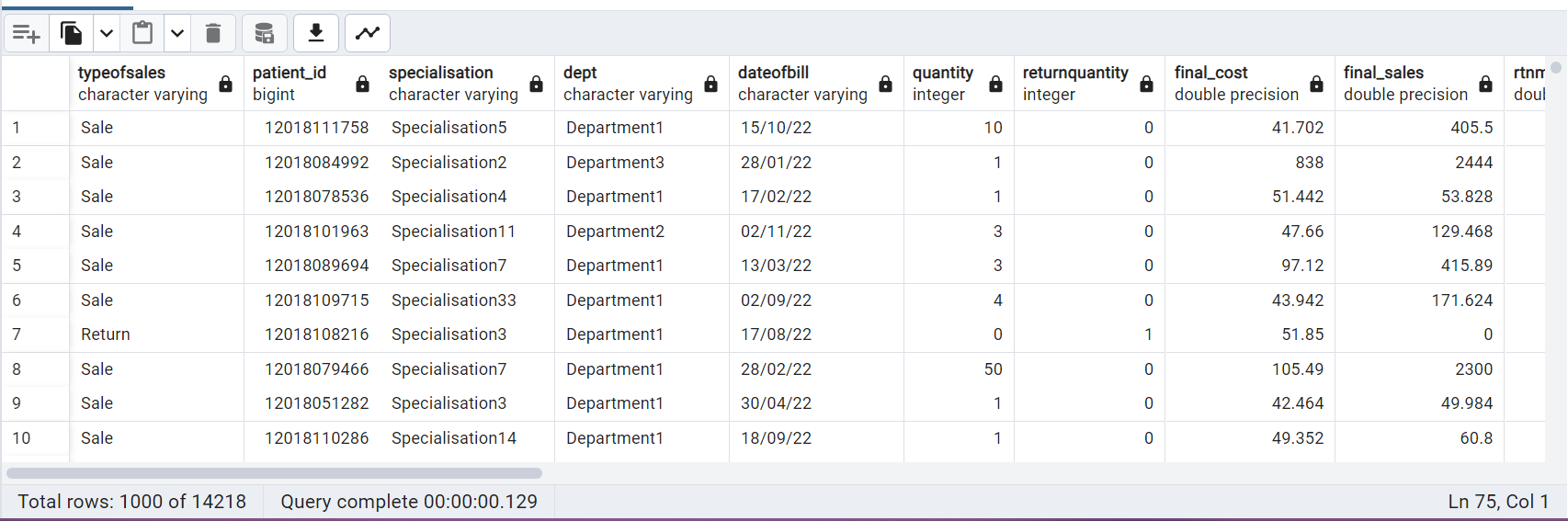
-- Formating date format

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

