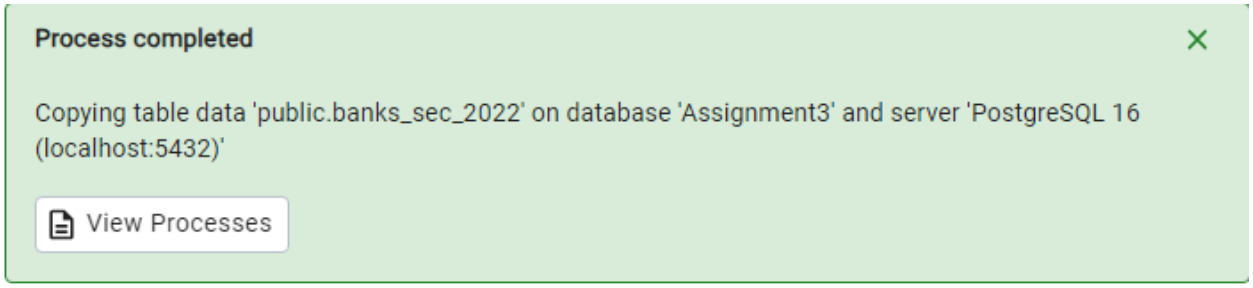
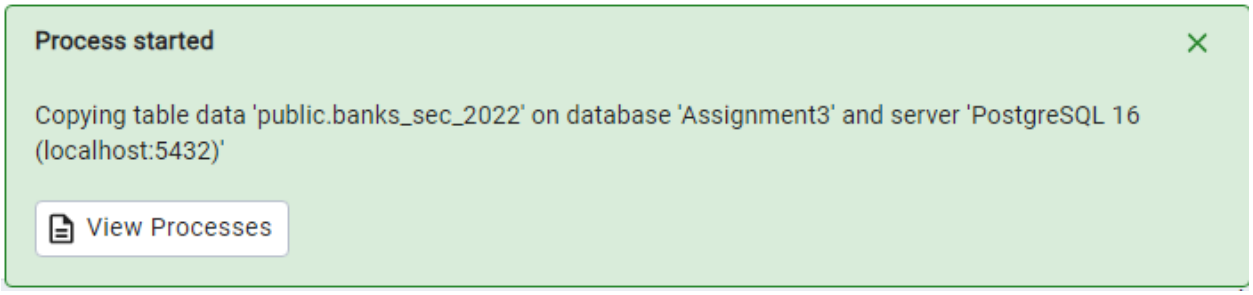
Q1) Querying Multiple Tables

1) Import data from banks sec 2002 and banks al 2002. Delete duplicate rows from banks sec 2002

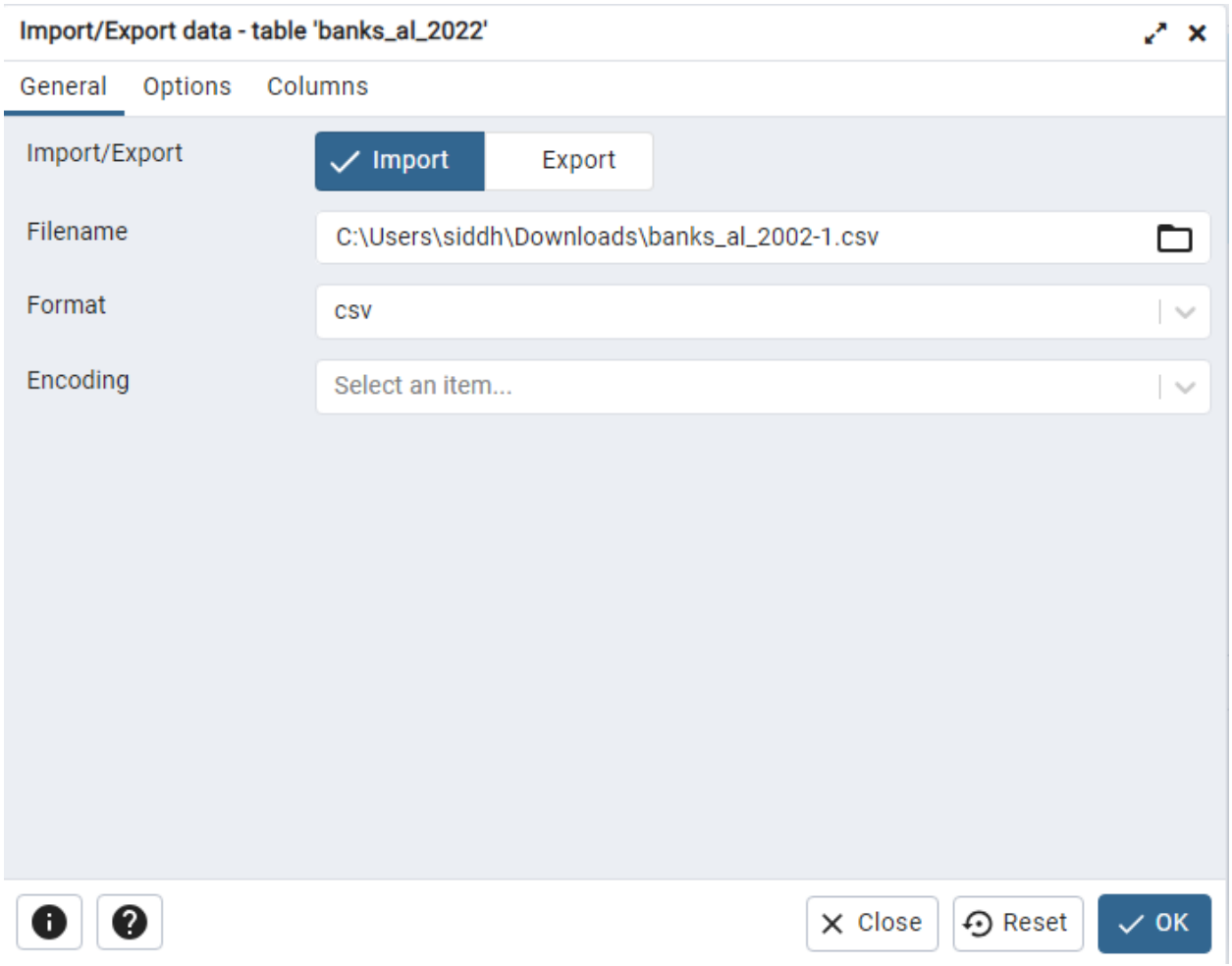
Creating both tables and uploading the data

```
CREATE TABLE banks_sec_2022 (
    id INT,
    date DATE,
    security INT
);
```






```
CREATE TABLE banks_al_2022 (
    id INT,
    date DATE,
    asset INT,
    liability INT
);
```



Process completed

Copying table data 'public.banks_al_2022' on database 'Assignment3' and server 'PostgreSQL 16 (localhost:5432)'

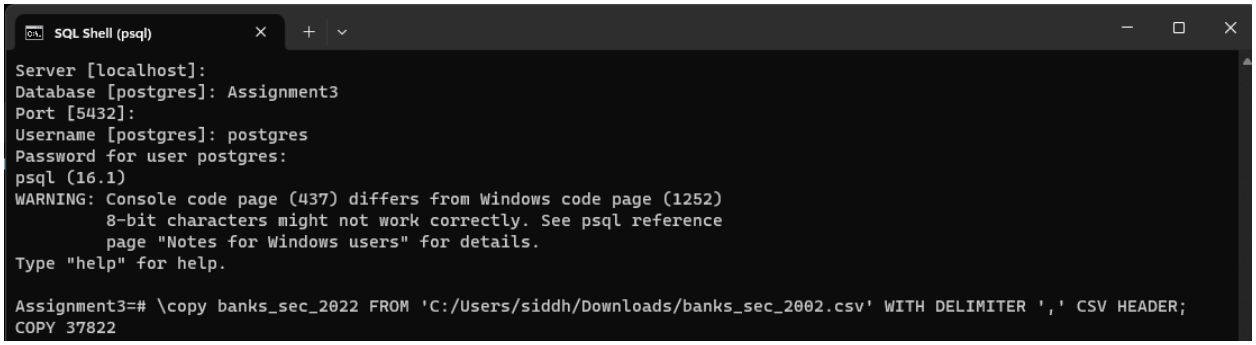
View Processes

Process started

Copying table data 'public.banks_al_2022' on database 'Assignment3' and server 'PostgreSQL 16 (localhost:5432)'

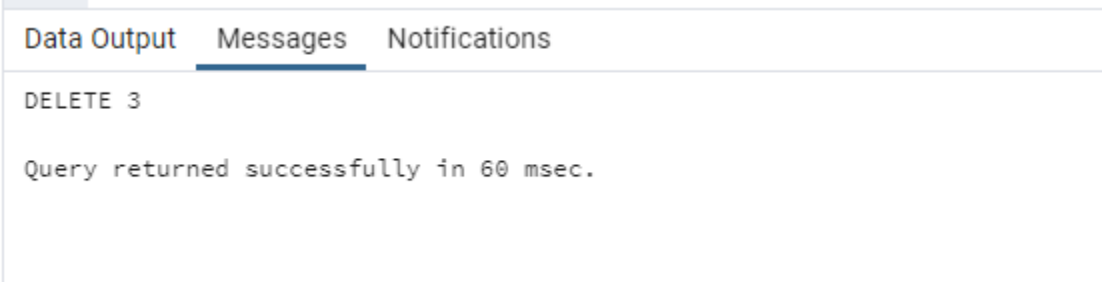
View Processes

Copying using psql due to permissions issue



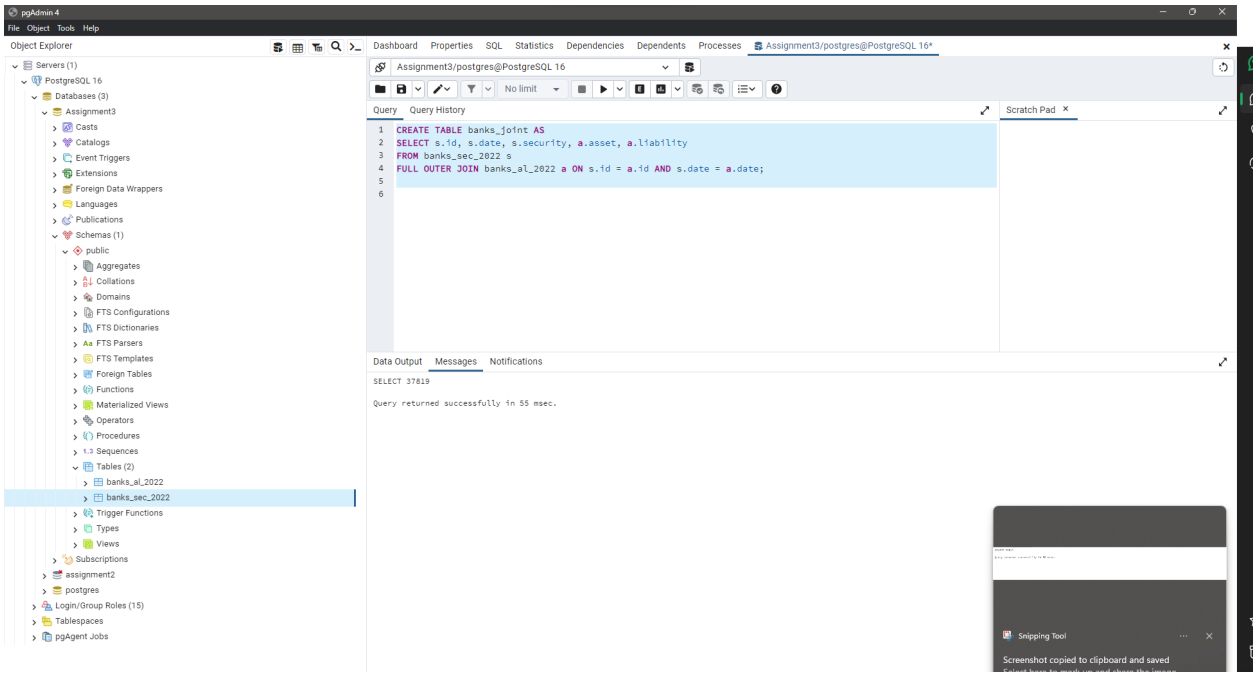
Deleting duplicate rows

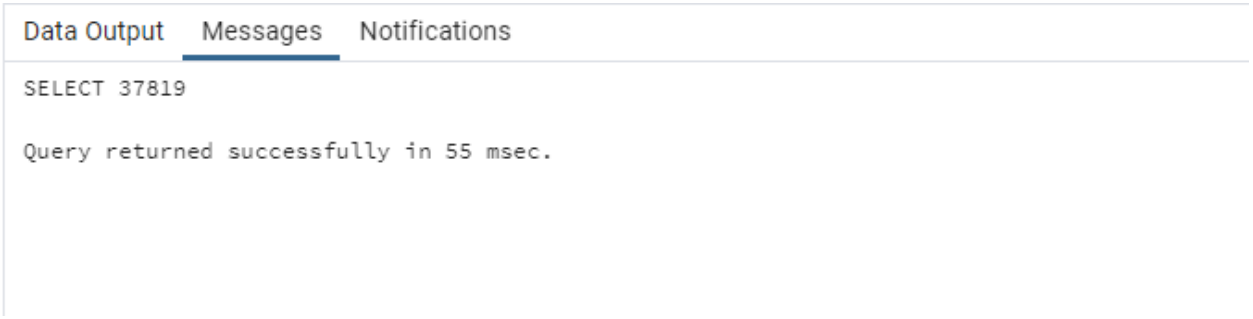
```
DELETE FROM banks_sec_2022
WHERE ctid NOT IN (
   SELECT MAX(ctid)
   FROM banks_sec_2022
   GROUP BY id, date, security
);
```



2)Select proper join manner to join banks sec 2002 and banks al 2002. Make sure that all data from banks sec 2002 are kept in the joint table. Report the first 10 observations

```
CREATE TABLE banks_joint AS
SELECT s.id, s.date, s.security, a.asset, a.liability
FROM banks_sec_2022 s
FULL OUTER JOIN banks_al_2022 a ON s.id = a.id AND s.date = a.date;
```





Reporting 10 observations

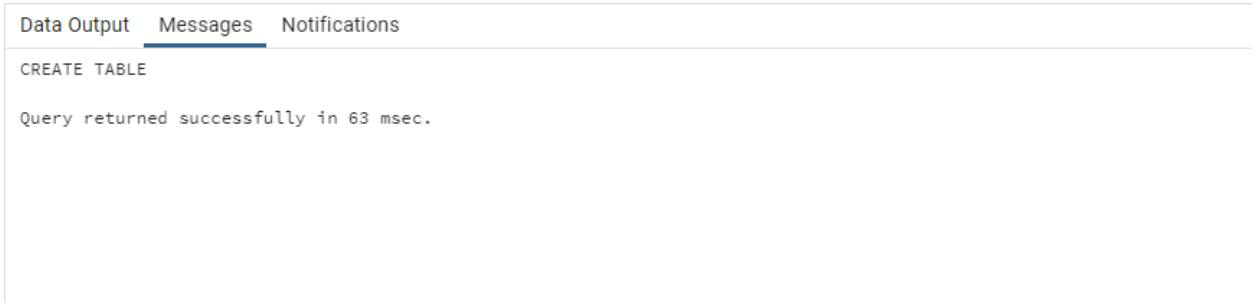
```
SELECT *
FROM banks_joint
LIMIT 10;
```

Data Output | Messages | Notifications

	id integer	date date	security integer	asset integer	liability integer
1	32307	2002-09-30	0	53714	49350
2	22598	2002-03-31	0	57360	53205
3	15879	2002-06-30	5357	76960	62945
4	35373	2002-06-30	0	46551	38721
5	5226	2002-09-30	7960	53873	48146
6	22092	2002-12-31	0	147828	135596
7	13749	2002-03-31	17476	214733	189972
8	823	2002-12-31	44972	271961	241138
9	29831	2002-03-31	6505	58300	43689
10	10203	2002-09-30	3889	320853	293453

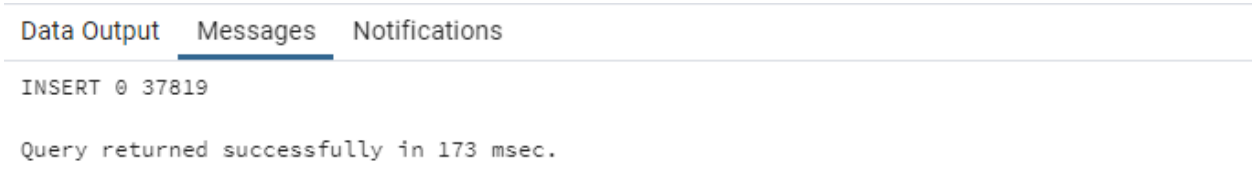
3)Create a new table banks total. Insert the values from previous joint table into this new one. And set a primary key for the table.

```
CREATE TABLE banks_total (
    id INTEGER,
    date DATE,
    security NUMERIC,
    asset NUMERIC,
    liability NUMERIC,
    PRIMARY KEY (id, date)
);
```



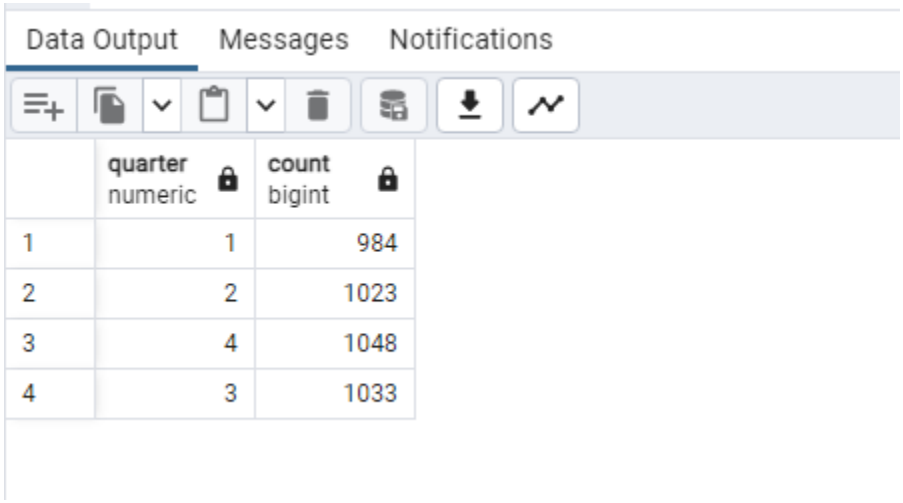

```
INSERT INTO banks_total (id, date, security, asset, liability)
SELECT id, date, security, asset, liability
```

FROM banks_joint;



4)For each quarter of the year 2002 count how many banks have security over 20% of its' asset.

```
SELECT EXTRACT(QUARTER FROM date) AS quarter, COUNT(*)
FROM banks_total
WHERE security > 0.2 * asset
GROUP BY quarter;
```

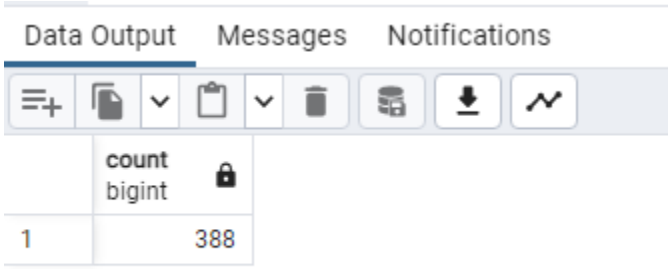


	quarter numeric	count bigint
1	1	984
2	2	1023
3	4	1048
4	3	1033

5)How many banks have liability over 90% of assets in first quarter of 2002 but goes below 90% in the second quarter of 2002.

```
SELECT COUNT(*)
FROM banks_total
WHERE EXTRACT(QUARTER FROM date) = 1 AND liability > 0.9 * asset
AND id IN (
    SELECT id
    FROM banks_total
    WHERE EXTRACT(QUARTER FROM date) = 2 AND liability < 0.9 * asset
```

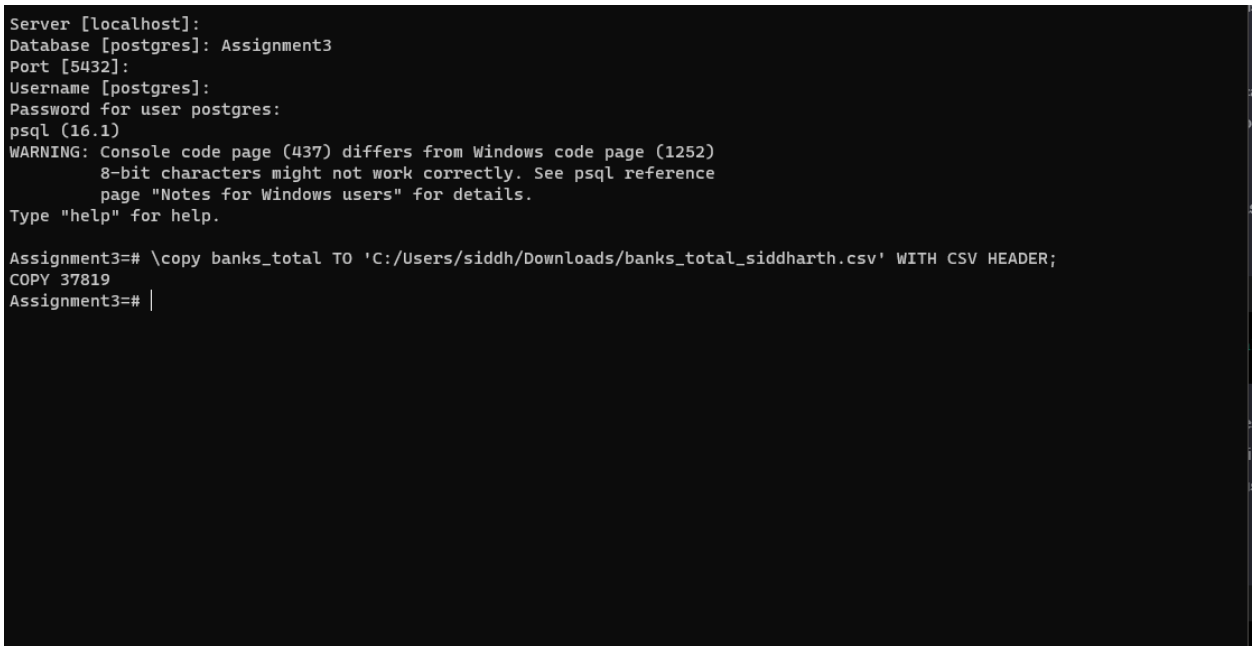
);



6)Export the joint table (banks total) to a csv file

Using psql to export the banks_total table

\copy banks_total TO 'C:/Users/siddh/Downloads/banks_total.csv' WITH CSV HEADER;



QUESTION 2

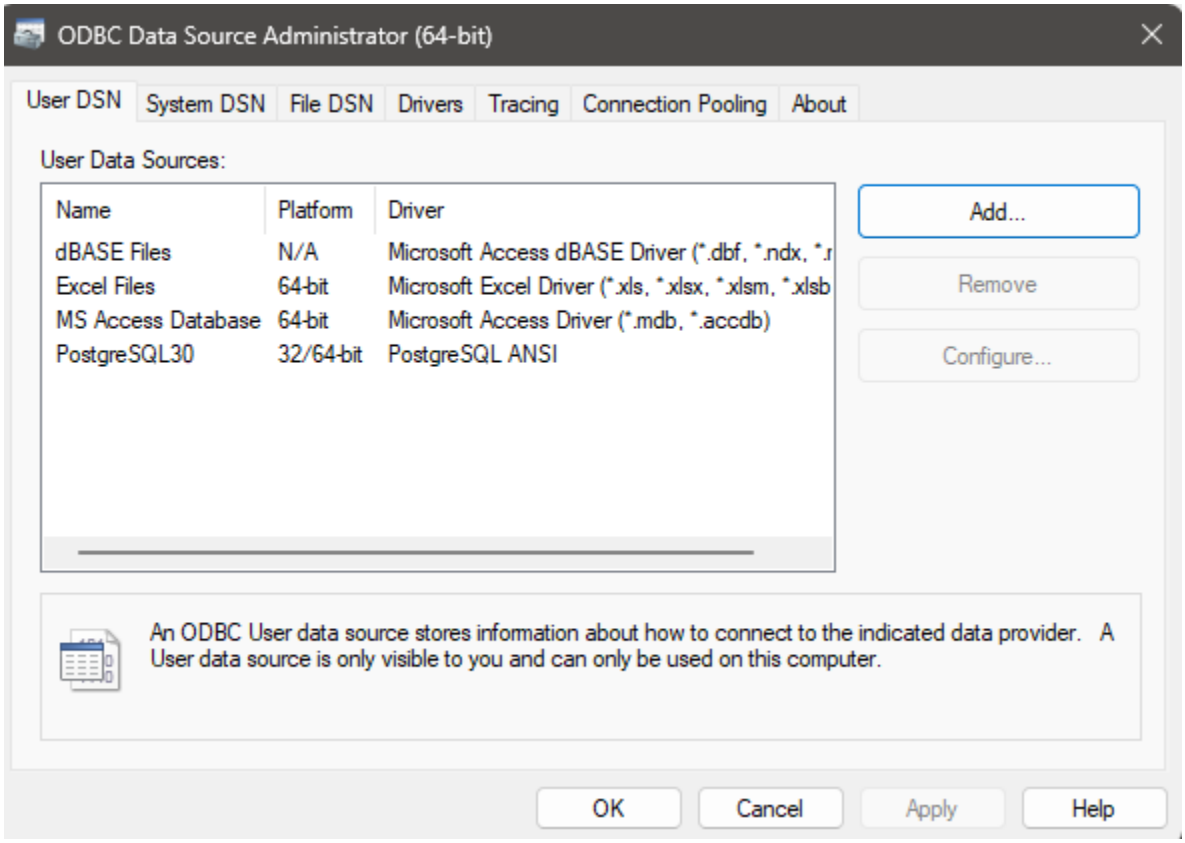
1)Make a connection to your local PostgreSQL database using API

Installed the required packages in RStudio

```
> install.packages("RODBC")
> install.packages("DBI")
```

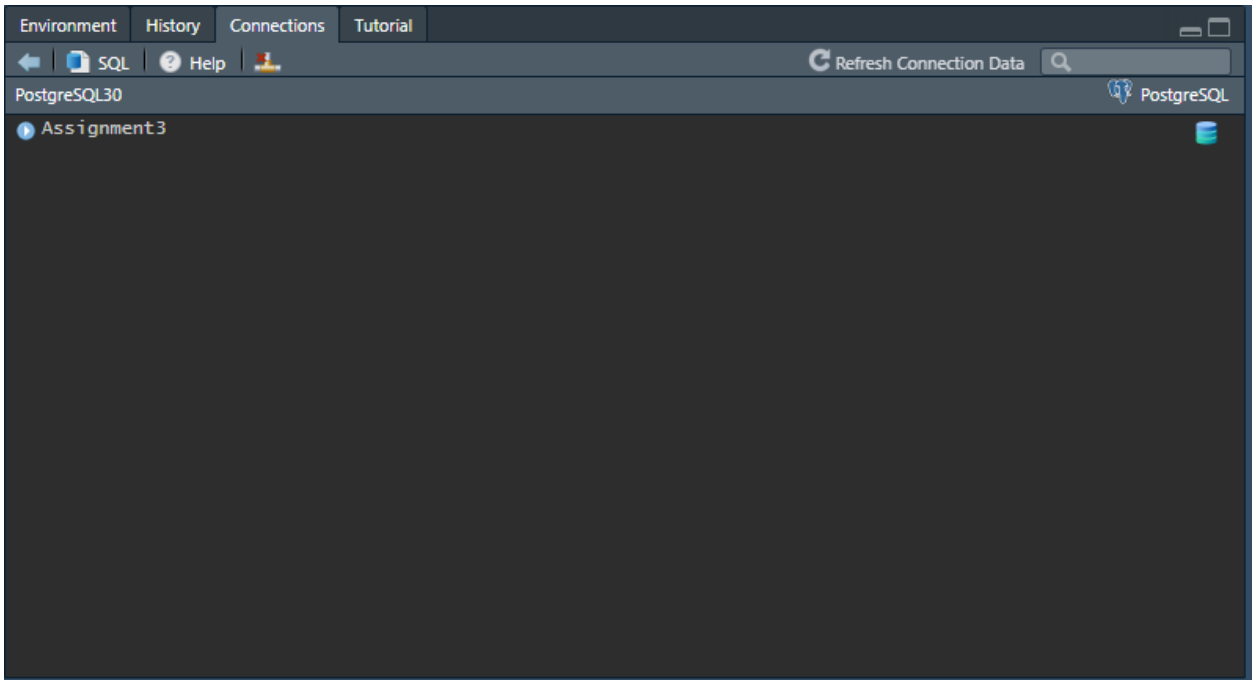
```
> install.packages("odbc")
```

Downloaded odbc driver for postgresql using https://www.postgresql.org/ftp/odbc/versions/msi/
Made the necessary additions in Data Source Administrator



RSCRIPT

```
library(DBI)
library(RODBC)
library(odbc)
con <- dbConnect(odbc::odbc(),"PostgreSQL30")
```



Connection was established

2)Import the csv file you got from Problem 1 (banks total) into a new table in the database using API. (Hint. Please give the table a new name if table ‘banks total‘ exists in the database)

```
csv_file_path <- "C:/Users/siddh/Downloads/banks_total_siddharth.csv"
banks_total <- read.csv(csv_file_path)
new_table_name <- "banks_total_new"

dbWriteTable(con, new_table_name, banks_total, overwrite = TRUE, row.names = FALSE)
```

3)Retrieve the data of table ‘banks total‘ using API. Count how many rows in the table.

```
banks_total_data <- dbReadTable(con, new_table_name)
num_rows <- nrow(banks_total_data)
print(num_rows)
```

```
[1] 37819
```

```
Console  Terminal  Background Jobs
R 4.2.3 · ~/
> library(DBI)
> library(RODBC)
> library(odbc)
> con <- dbConnect(odbc::odbc(),"PostgreSQL30")
> csv_file_path <- "C:/Users/siddh/Downloads/banks_total.csv"
> banks_total <- read.csv(csv_file_path)
Error in file(file, "rt") : cannot open the connection
In addition: Warning message:
In file(file, "rt") :
  cannot open file 'C:/Users/siddh/Downloads/banks_total.csv': No such file or directory
> csv_file_path <- "C:/Users/siddh/Downloads/banks_total_siddharth.csv"
> banks_total <- read.csv(csv_file_path)
> new_table_name <- "banks_total_new"
> dbWriteTable(con, new_table_name, banks_total, overwrite = TRUE, row.names = FALSE)
> banks_total_data <- dbReadTable(con, new_table_name)
> num_rows <- nrow(banks_total_data)
> print(num_rows)
[1] 37819
>
```

```
> num_rows <- nrow(banks_total_data)
> print(num_rows)
[1] 37819
>
```