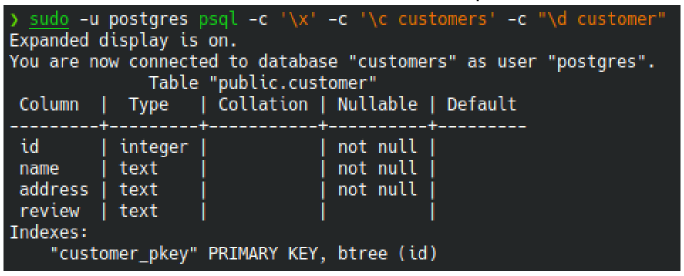
# LAB_8

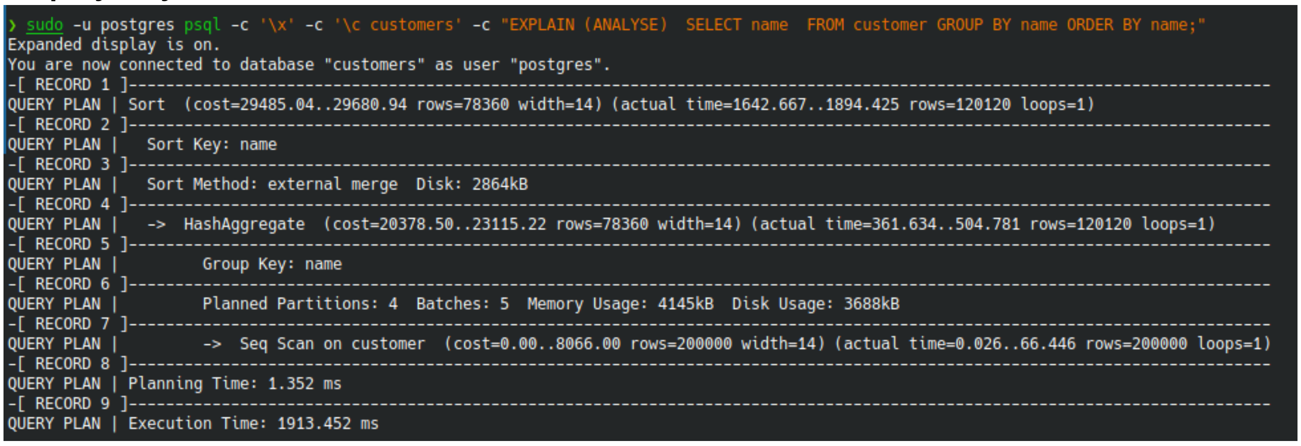
I used the **terminal** to work with the **postgresql**, so here will be the screens of the terminal

## Exercise 1

- Now we do not have additional indecies exept for the one for the primary key
- 

```
❯ sudo -u postgres psql -c '\x' -c '\c customers' -c "\d customer"
Expanded display is on.
You are now connected to database "customers" as user "postgres".
              Table "public.customer"
 Column  |  Type   | Collation | Nullable | Default
---------+---------+-----------+----------+---------
 id      | integer |           | not null |
 name    | text    |           | not null |
 address | text    |           | not null |
 review  | text    |           |          |
Indexes:
    "customer_pkey" PRIMARY KEY, btree (id)
```

1. **1st query analysis:**

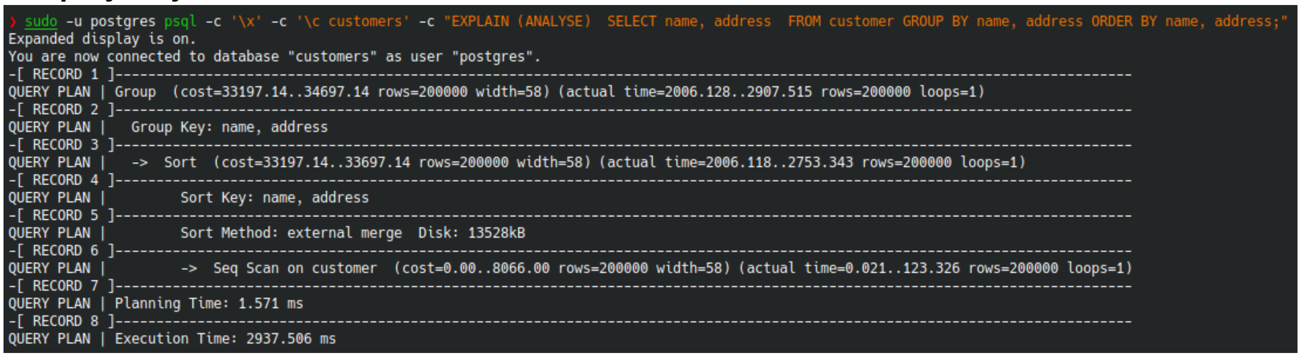


```
❯ sudo -u postgres psql -c '\x' -c '\c customers' -c "EXPLAIN (ANALYSE)  SELECT name  FROM customer GROUP BY name ORDER BY name;"
Expanded display is on.
You are now connected to database "customers" as user "postgres".
-[ RECORD 1 ]-------------------------------------------------------------------------------------------------------------------
QUERY PLAN | Sort  (cost=29485.04..29680.94 rows=78360 width=14) (actual time=1642.667..1894.425 rows=120120 loops=1)
-[ RECORD 2 ]-------------------------------------------------------------------------------------------------------------------
QUERY PLAN |   Sort Key: name
-[ RECORD 3 ]-------------------------------------------------------------------------------------------------------------------
QUERY PLAN |   Sort Method: external merge  Disk: 2864kB
-[ RECORD 4 ]-------------------------------------------------------------------------------------------------------------------
QUERY PLAN |   ->  HashAggregate  (cost=20378.50..23115.22 rows=78360 width=14) (actual time=361.634..504.781 rows=120120 loops=1)
-[ RECORD 5 ]-------------------------------------------------------------------------------------------------------------------
QUERY PLAN |         Group Key: name
-[ RECORD 6 ]-------------------------------------------------------------------------------------------------------------------
QUERY PLAN |         Planned Partitions: 4  Batches: 5  Memory Usage: 4145kB  Disk Usage: 3688kB
-[ RECORD 7 ]-------------------------------------------------------------------------------------------------------------------
QUERY PLAN |         ->  Seq Scan on customer  (cost=0.00..8066.00 rows=200000 width=14) (actual time=0.026..66.446 rows=200000 loops=1)
-[ RECORD 8 ]-------------------------------------------------------------------------------------------------------------------
QUERY PLAN | Planning Time: 1.352 ms
-[ RECORD 9 ]-------------------------------------------------------------------------------------------------------------------
QUERY PLAN | Execution Time: 1913.452 ms
```

$Cost\_1\_no\_index = 60,862.16$

$Execution\_time\_1\_no\_index = 1913.452\ (ms)$

2. **2nd query analysis**

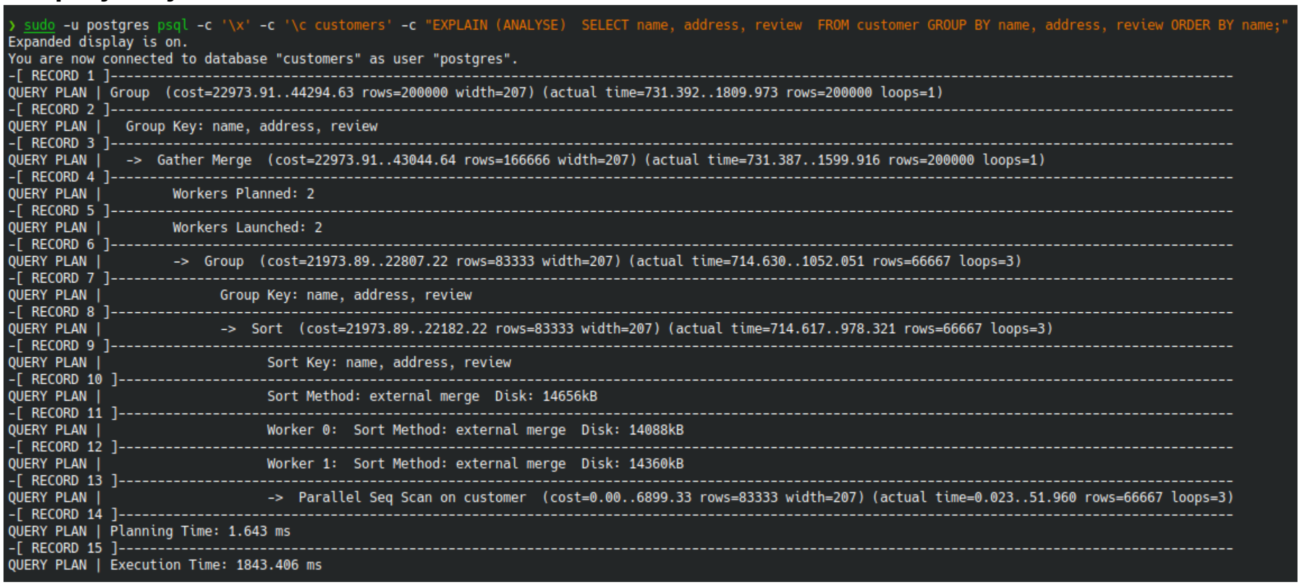


```
❯ sudo -u postgres psql -c '\x' -c '\c customers' -c "EXPLAIN (ANALYSE)  SELECT name, address  FROM customer GROUP BY name, address ORDER BY name, address;"
Expanded display is on.
You are now connected to database "customers" as user "postgres".
-[ RECORD 1 ]-------------------------------------------------------------------------------------------------------------
QUERY PLAN | Group  (cost=33197.14..34697.14 rows=200000 width=58) (actual time=2006.128..2907.515 rows=200000 loops=1)
-[ RECORD 2 ]-------------------------------------------------------------------------------------------------------------
QUERY PLAN |   Group Key: name, address
-[ RECORD 3 ]-------------------------------------------------------------------------------------------------------------
QUERY PLAN |   ->  Sort  (cost=33197.14..33697.14 rows=200000 width=58) (actual time=2006.118..2753.343 rows=200000 loops=1)
-[ RECORD 4 ]-------------------------------------------------------------------------------------------------------------
QUERY PLAN |         Sort Key: name, address
-[ RECORD 5 ]-------------------------------------------------------------------------------------------------------------
QUERY PLAN |         Sort Method: external merge  Disk: 13528kB
-[ RECORD 6 ]-------------------------------------------------------------------------------------------------------------
QUERY PLAN |         ->  Seq Scan on customer  (cost=0.00..8066.00 rows=200000 width=58) (actual time=0.021..123.326 rows=200000 loops=1)
-[ RECORD 7 ]-------------------------------------------------------------------------------------------------------------
QUERY PLAN | Planning Time: 1.571 ms
-[ RECORD 8 ]-------------------------------------------------------------------------------------------------------------
QUERY PLAN | Execution Time: 2937.506 ms
```

$Cost\_2\_no\_index = 70,460.28$

$Execution\_time\_2\_no\_index = 2937.506\ (ms)$

3. **3rd query analysis**

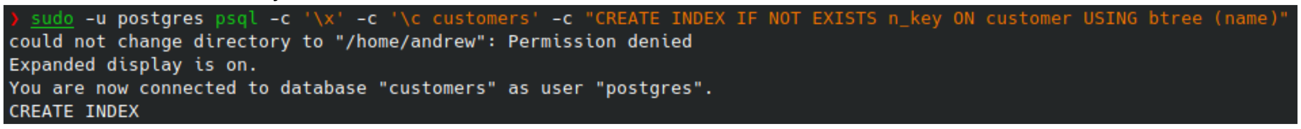


```
❯ sudo -u postgres psql -c '\x' -c '\c customers' -c "EXPLAIN (ANALYSE)  SELECT name, address, review  FROM customer GROUP BY name, address, review ORDER BY name;"
Expanded display is on.
You are now connected to database "customers" as user "postgres".
-[ RECORD 1 ]--------------------------------------------------------------------------------------------------------------------------------------
QUERY PLAN | Group  (cost=22973.91..44294.63 rows=200000 width=207) (actual time=731.392..1809.973 rows=200000 loops=1)
-[ RECORD 2 ]--------------------------------------------------------------------------------------------------------------------------------------
QUERY PLAN |   Group Key: name, address, review
-[ RECORD 3 ]--------------------------------------------------------------------------------------------------------------------------------------
QUERY PLAN |   ->  Gather Merge  (cost=22973.91..43044.64 rows=166666 width=207) (actual time=731.387..1599.916 rows=200000 loops=1)
-[ RECORD 4 ]--------------------------------------------------------------------------------------------------------------------------------------
QUERY PLAN |         Workers Planned: 2
-[ RECORD 5 ]--------------------------------------------------------------------------------------------------------------------------------------
QUERY PLAN |         Workers Launched: 2
-[ RECORD 6 ]--------------------------------------------------------------------------------------------------------------------------------------
QUERY PLAN |         ->  Group  (cost=21973.89..22807.22 rows=83333 width=207) (actual time=714.630..1052.051 rows=66667 loops=3)
-[ RECORD 7 ]--------------------------------------------------------------------------------------------------------------------------------------
QUERY PLAN |               Group Key: name, address, review
-[ RECORD 8 ]--------------------------------------------------------------------------------------------------------------------------------------
QUERY PLAN |               ->  Sort  (cost=21973.89..22182.22 rows=83333 width=207) (actual time=714.617..978.321 rows=66667 loops=3)
-[ RECORD 9 ]--------------------------------------------------------------------------------------------------------------------------------------
QUERY PLAN |                     Sort Key: name, address, review
-[ RECORD 10 ]-------------------------------------------------------------------------------------------------------------------------------------
QUERY PLAN |                     Sort Method: external merge  Disk: 14656kB
-[ RECORD 11 ]-------------------------------------------------------------------------------------------------------------------------------------
QUERY PLAN |                     Worker 0:  Sort Method: external merge  Disk: 14088kB
-[ RECORD 12 ]-------------------------------------------------------------------------------------------------------------------------------------
QUERY PLAN |                     Worker 1:  Sort Method: external merge  Disk: 14360kB
-[ RECORD 13 ]-------------------------------------------------------------------------------------------------------------------------------------
QUERY PLAN |                     ->  Parallel Seq Scan on customer  (cost=0.00..6899.33 rows=83333 width=207) (actual time=0.023..51.960 rows=66667 loops=3)
-[ RECORD 14 ]-------------------------------------------------------------------------------------------------------------------------------------
QUERY PLAN | Planning Time: 1.643 ms
-[ RECORD 15 ]-------------------------------------------------------------------------------------------------------------------------------------
QUERY PLAN | Execution Time: 1843.406 ms
```

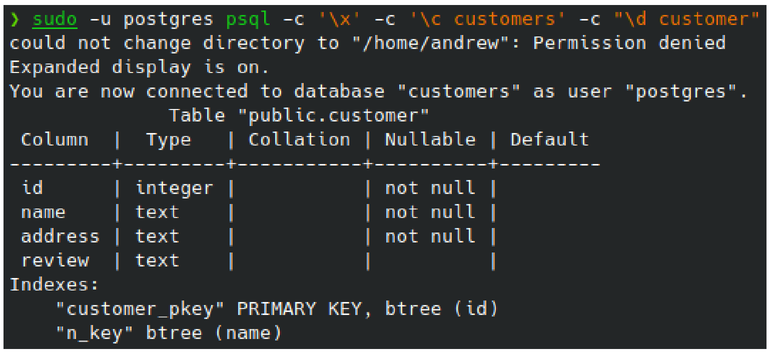
$Cost\_3\_no\_index = 94,933.41$

$Execution\_time\_3\_no\_index = 1843.406\ (ms)$

- I created the index na_key

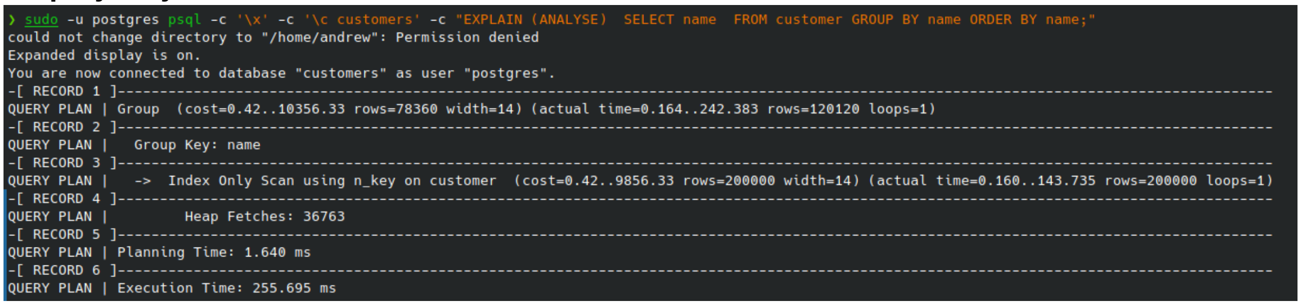


```
❯ sudo -u postgres psql -c '\x' -c '\c customers' -c "CREATE INDEX IF NOT EXISTS n_key ON customer USING btree (name)"
could not change directory to "/home/andrew": Permission denied
Expanded display is on.
You are now connected to database "customers" as user "postgres".
CREATE INDEX
```



```
❯ sudo -u postgres psql -c '\x' -c '\c customers' -c "\d customer"
could not change directory to "/home/andrew": Permission denied
Expanded display is on.
You are now connected to database "customers" as user "postgres".
              Table "public.customer"
 Column  |  Type   | Collation | Nullable | Default
---------+---------+-----------+----------+---------
 id      | integer |           | not null |
 name    | text    |           | not null |
 address | text    |           | not null |
 review  | text    |           |          |
Indexes:
    "customer_pkey" PRIMARY KEY, btree (id)
    "n_key" btree (name)
```

1. **1st query analysis**

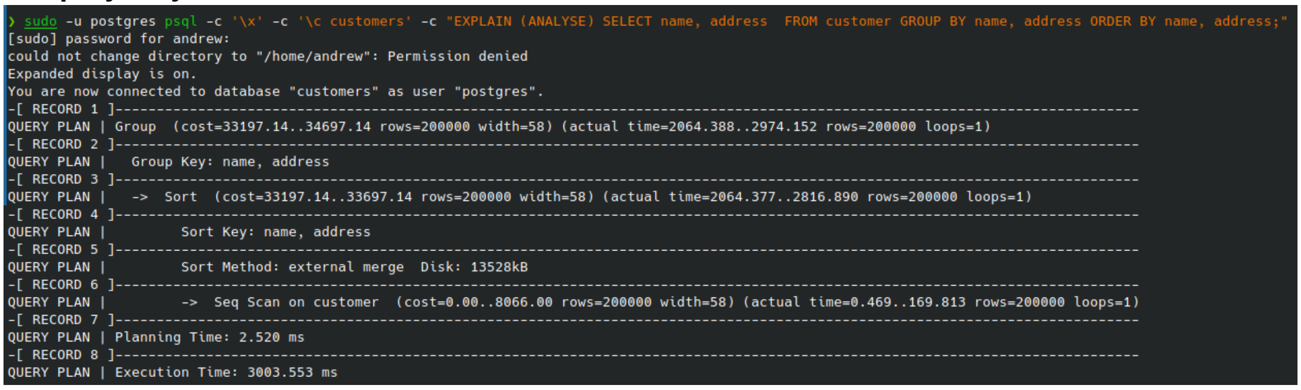


```
❯ sudo -u postgres psql -c '\x' -c '\c customers' -c "EXPLAIN (ANALYSE)  SELECT name  FROM customer GROUP BY name ORDER BY name;"
could not change directory to "/home/andrew": Permission denied
Expanded display is on.
You are now connected to database "customers" as user "postgres".
-[ RECORD 1 ]-------------------------------------------------------------------------------------------------------------------------------------------
QUERY PLAN | Group  (cost=0.42..10356.33 rows=78360 width=14) (actual time=0.164..242.383 rows=120120 loops=1)
-[ RECORD 2 ]-------------------------------------------------------------------------------------------------------------------------------------------
QUERY PLAN |   Group Key: name
-[ RECORD 3 ]-------------------------------------------------------------------------------------------------------------------------------------------
QUERY PLAN |   ->  Index Only Scan using n_key on customer  (cost=0.42..9856.33 rows=200000 width=14) (actual time=0.160..143.735 rows=200000 loops=1)
-[ RECORD 4 ]-------------------------------------------------------------------------------------------------------------------------------------------
QUERY PLAN |         Heap Fetches: 36763
-[ RECORD 5 ]-------------------------------------------------------------------------------------------------------------------------------------------
QUERY PLAN | Planning Time: 1.640 ms
-[ RECORD 6 ]-------------------------------------------------------------------------------------------------------------------------------------------
QUERY PLAN | Execution Time: 255.695 ms
```

$Cost\_1\_with\_index = 20,212.66$

$Execution\_time\_1\_with\_index = 255.695\ (ms)$

2. **2nd query analysis**

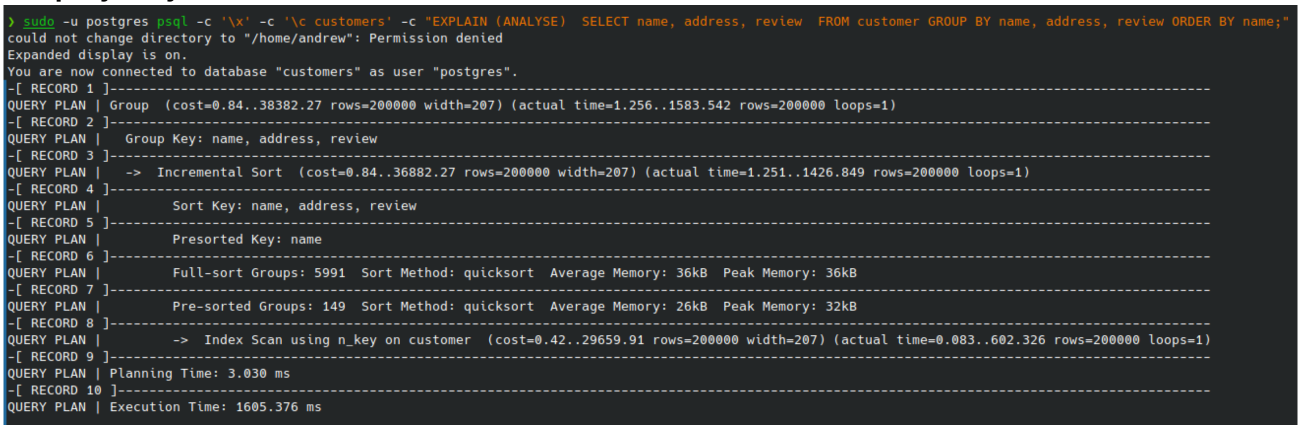


```
❯ sudo -u postgres psql -c '\x' -c '\c customers' -c "EXPLAIN (ANALYSE) SELECT name, address  FROM customer GROUP BY name, address ORDER BY name, address;"
[sudo] password for andrew:
could not change directory to "/home/andrew": Permission denied
Expanded display is on.
You are now connected to database "customers" as user "postgres".
-[ RECORD 1 ]----------------------------------------------------------------------------------------------------------------------
QUERY PLAN | Group  (cost=33197.14..34697.14 rows=200000 width=58) (actual time=2064.388..2974.152 rows=200000 loops=1)
-[ RECORD 2 ]----------------------------------------------------------------------------------------------------------------------
QUERY PLAN |   Group Key: name, address
-[ RECORD 3 ]----------------------------------------------------------------------------------------------------------------------
QUERY PLAN |   ->  Sort  (cost=33197.14..33697.14 rows=200000 width=58) (actual time=2064.377..2816.890 rows=200000 loops=1)
-[ RECORD 4 ]----------------------------------------------------------------------------------------------------------------------
QUERY PLAN |         Sort Key: name, address
-[ RECORD 5 ]----------------------------------------------------------------------------------------------------------------------
QUERY PLAN |         Sort Method: external merge  Disk: 13528kB
-[ RECORD 6 ]----------------------------------------------------------------------------------------------------------------------
QUERY PLAN |         ->  Seq Scan on customer  (cost=0.00..8066.00 rows=200000 width=58) (actual time=0.469..169.813 rows=200000 loops=1)
-[ RECORD 7 ]----------------------------------------------------------------------------------------------------------------------
QUERY PLAN | Planning Time: 2.520 ms
-[ RECORD 8 ]----------------------------------------------------------------------------------------------------------------------
QUERY PLAN | Execution Time: 3003.553 ms
```

$Cost\_2\_wiht\_index = 70,460.28$

$Execution\_time\_2\_with\_index = 3003.553\ (ms)$

3. **3rd query analysis**



```
❯ sudo -u postgres psql -c '\x' -c '\c customers' -c "EXPLAIN (ANALYSE)  SELECT name, address, review  FROM customer GROUP BY name, address, review ORDER BY name;"
could not change directory to "/home/andrew": Permission denied
Expanded display is on.
You are now connected to database "customers" as user "postgres".
-[ RECORD 1 ]-------------------------------------------------------------------------------------------------------------------------------
QUERY PLAN | Group  (cost=0.84..38382.27 rows=200000 width=207) (actual time=1.256..1583.542 rows=200000 loops=1)
-[ RECORD 2 ]-------------------------------------------------------------------------------------------------------------------------------
QUERY PLAN |   Group Key: name, address, review
-[ RECORD 3 ]-------------------------------------------------------------------------------------------------------------------------------
QUERY PLAN |   ->  Incremental Sort  (cost=0.84..36882.27 rows=200000 width=207) (actual time=1.251..1426.849 rows=200000 loops=1)
-[ RECORD 4 ]-------------------------------------------------------------------------------------------------------------------------------
QUERY PLAN |         Sort Key: name, address, review
-[ RECORD 5 ]-------------------------------------------------------------------------------------------------------------------------------
QUERY PLAN |         Presorted Key: name
-[ RECORD 6 ]-------------------------------------------------------------------------------------------------------------------------------
QUERY PLAN |         Full-sort Groups: 5991  Sort Method: quicksort  Average Memory: 36kB  Peak Memory: 36kB
-[ RECORD 7 ]-------------------------------------------------------------------------------------------------------------------------------
QUERY PLAN |         Pre-sorted Groups: 149  Sort Method: quicksort  Average Memory: 26kB  Peak Memory: 32kB
-[ RECORD 8 ]-------------------------------------------------------------------------------------------------------------------------------
QUERY PLAN |         ->  Index Scan using n_key on customer  (cost=0.42..29659.91 rows=200000 width=207) (actual time=0.083..602.326 rows=200000 loops=1)
-[ RECORD 9 ]-------------------------------------------------------------------------------------------------------------------------------
QUERY PLAN | Planning Time: 3.030 ms
-[ RECORD 10 ]------------------------------------------------------------------------------------------------------------------------------
QUERY PLAN | Execution Time: 1605.376 ms
```

$Cost\_3\_no\_index = 104,924.45$

$Execution\_time\_3\_with\_index = 1605.376\ (ms)$

## Analysing of the result

- **1st query**

  $Cost\_1\_no\_index = 60,862.16 > Cost\_1\_with\_index = 20,212.66$

  ***Since we created the index for the column we are work with, we have the smaller result after the creating the index***

  $Execution\_time\_1\_no\_index = 1913.452\ (ms) > Execution\_time\_1\_with\_index = 255.695\ (ms)$

- **2nd query**

  $Cost\_2\_no\_index = 70,460.28 = Cost\_2\_wiht\_index = 70,460.28$

  ***Here we see that the result for the cost is the same, so we could even not use the index***

  $Execution\_time\_2\_no\_index = 2937.506\ (ms) < Execution\_time\_2\_with\_index = 3003.553\ (ms)$

- **3rd query**

  $Cost\_3\_no\_index = 94,933.41 < Cost\_3\_no\_index = 104,924.45$

  ***The result says that in this case it would be better not to use the index for this query***

  $Execution\_time\_3\_no\_index = 1843.406\ (ms) > Execution\_time\_3\_with\_index = 1605.376\ (ms)$

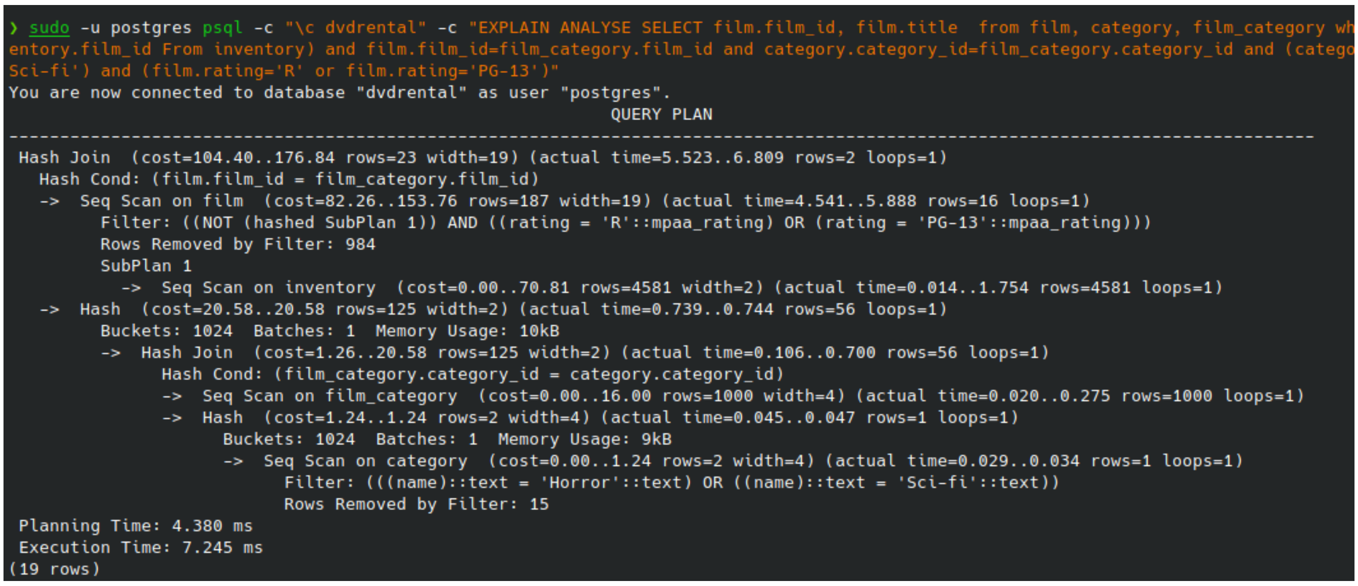
---

# Exercise 2

result

- **1st query**

```
SELECT film.film_id, film.title  from film, category, film_category where film.film_id NOT IN (SELECT
inventory.film_id From inventory) and film.film_id=film_category.film_id and
category.category_id=film_category.categor
y_id and (category.name='Horror' or category.name='Sci-fi') and (film.rating='R' or film.rating='PG-13')
```

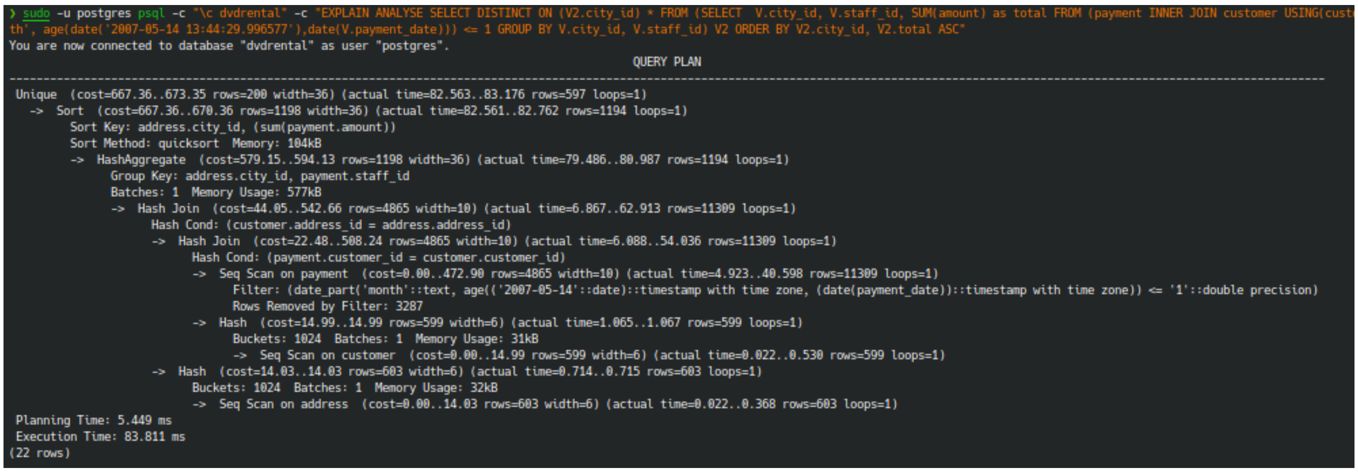


```
❯ sudo -u postgres psql -c "\c dvdrental" -c "EXPLAIN ANALYSE SELECT film.film_id, film.title  from film, category, film_category wh
entory.film_id From inventory) and film.film_id=film_category.film_id and category.category_id=film_category.category_id and (catego
Sci-fi') and (film.rating='R' or film.rating='PG-13')"
You are now connected to database "dvdrental" as user "postgres".
                                                           QUERY PLAN
------------------------------------------------------------------------------------------------------------------------------------
 Hash Join  (cost=104.40..176.84 rows=23 width=19) (actual time=5.523..6.809 rows=2 loops=1)
   Hash Cond: (film.film_id = film_category.film_id)
   ->  Seq Scan on film  (cost=82.26..153.76 rows=187 width=19) (actual time=4.541..5.888 rows=16 loops=1)
         Filter: ((NOT (hashed SubPlan 1)) AND ((rating = 'R'::mpaa_rating) OR (rating = 'PG-13'::mpaa_rating)))
         Rows Removed by Filter: 984
         SubPlan 1
           ->  Seq Scan on inventory  (cost=0.00..70.81 rows=4581 width=2) (actual time=0.014..1.754 rows=4581 loops=1)
   ->  Hash  (cost=20.58..20.58 rows=125 width=2) (actual time=0.739..0.744 rows=56 loops=1)
         Buckets: 1024  Batches: 1  Memory Usage: 10kB
         ->  Hash Join  (cost=1.26..20.58 rows=125 width=2) (actual time=0.106..0.700 rows=56 loops=1)
               Hash Cond: (film_category.category_id = category.category_id)
               ->  Seq Scan on film_category  (cost=0.00..16.00 rows=1000 width=4) (actual time=0.020..0.275 rows=1000 loops=1)
               ->  Hash  (cost=1.24..1.24 rows=2 width=4) (actual time=0.045..0.047 rows=1 loops=1)
                     Buckets: 1024  Batches: 1  Memory Usage: 9kB
                     ->  Seq Scan on category  (cost=0.00..1.24 rows=2 width=4) (actual time=0.029..0.034 rows=1 loops=1)
                           Filter: (((name)::text = 'Horror'::text) OR ((name)::text = 'Sci-fi'::text))
                           Rows Removed by Filter: 15
 Planning Time: 4.380 ms
 Execution Time: 7.245 ms
(19 rows)
```

The most expensive operation was: **Scannin on film**

- **2nd query**

```
SELECT DISTINCT ON (V2.city_id) * FROM (SELECT  V.city_id, V.staff_id, SUM(amount) as total FROM (payment
INNER JOIN customer USING(customer_id) INNER JOIN address USING(address_id)) V WHERE date_part('month',
age(date('2007-05-14 13:44:29.996577'),date(V.payment_date))) <= 1 GROUP BY V.city_id, V.staff_id) V2 ORDER BY
V2.city_id, V2.total ASC
```



```
❯ sudo -u postgres psql -c "\c dvdrental" -c "EXPLAIN ANALYSE SELECT DISTINCT ON (V2.city_id) * FROM (SELECT  V.city_id, V.staff_id, SUM(amount) as total FROM (payment INNER JOIN customer USING(cust
th', age(date('2007-05-14 13:44:29.996577'),date(V.payment_date))) <= 1 GROUP BY V.city_id, V.staff_id) V2 ORDER BY V2.city_id, V2.total ASC"
You are now connected to database "dvdrental" as user "postgres".
                                                                                                      QUERY PLAN
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
 Unique  (cost=667.36..673.35 rows=200 width=36) (actual time=82.563..83.176 rows=597 loops=1)
   ->  Sort  (cost=667.36..670.36 rows=1198 width=36) (actual time=82.561..82.762 rows=1194 loops=1)
         Sort Key: address.city_id, (sum(payment.amount))
         Sort Method: quicksort  Memory: 104kB
         ->  HashAggregate  (cost=579.15..594.13 rows=1198 width=36) (actual time=79.486..80.987 rows=1194 loops=1)
               Group Key: address.city_id, payment.staff_id
               Batches: 1  Memory Usage: 577kB
               ->  Hash Join  (cost=44.05..542.66 rows=4865 width=10) (actual time=6.867..62.913 rows=11309 loops=1)
                     Hash Cond: (customer.address_id = address.address_id)
                     ->  Hash Join  (cost=22.48..508.24 rows=4865 width=10) (actual time=6.088..54.036 rows=11309 loops=1)
                           Hash Cond: (payment.customer_id = customer.customer_id)
                           ->  Seq Scan on payment  (cost=0.00..472.90 rows=4865 width=10) (actual time=4.923..40.598 rows=11309 loops=1)
                                 Filter: (date_part('month'::text, age(('2007-05-14'::date)::timestamp with time zone, (date(payment_date))::timestamp with time zone)) <= '1'::double precision)
                                 Rows Removed by Filter: 3287
                           ->  Hash  (cost=14.99..14.99 rows=599 width=6) (actual time=1.065..1.067 rows=599 loops=1)
                                 Buckets: 1024  Batches: 1  Memory Usage: 31kB
                                 ->  Seq Scan on customer  (cost=0.00..14.99 rows=599 width=6) (actual time=0.022..0.530 rows=599 loops=1)
                     ->  Hash  (cost=14.03..14.03 rows=603 width=6) (actual time=0.714..0.715 rows=603 loops=1)
                           Buckets: 1024  Batches: 1  Memory Usage: 32kB
                           ->  Seq Scan on address  (cost=0.00..14.03 rows=603 width=6) (actual time=0.022..0.368 rows=603 loops=1)
 Planning Time: 5.449 ms
 Execution Time: 83.811 ms
(22 rows)
```

The most expensive operation was: **sort**