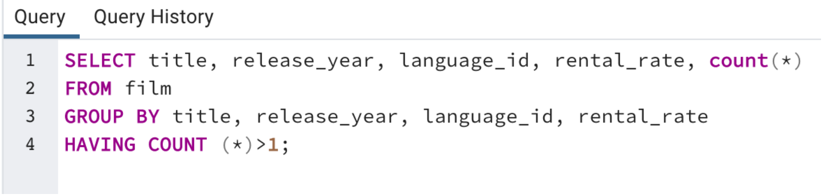
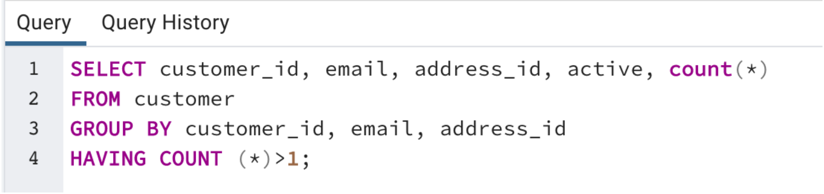
1. Duplicate date

Film table



```
SELECT title, release_year, language_id, rental_rate, count(*)
FROM film
GROUP BY title, release_year, language_id, rental_rate
HAVING COUNT (*)>1;
```

Customer table



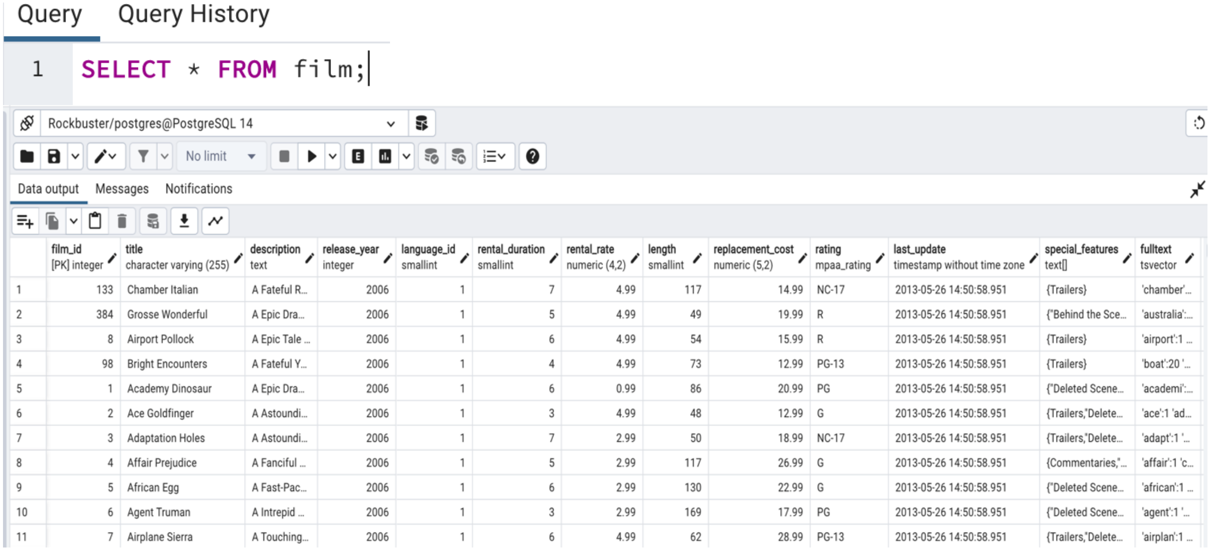
```
SELECT customer_id, email, address_id, active, count(*)
FROM customer
GROUP BY customer_id, email, address_id
HAVING COUNT (*)>1;
```

Solution for the duplicate record (1) create a virtual table, known as a “view”, where you select only unique records, or (2) delete the duplicate record from the table or view.

2. Summarize Data

Film table

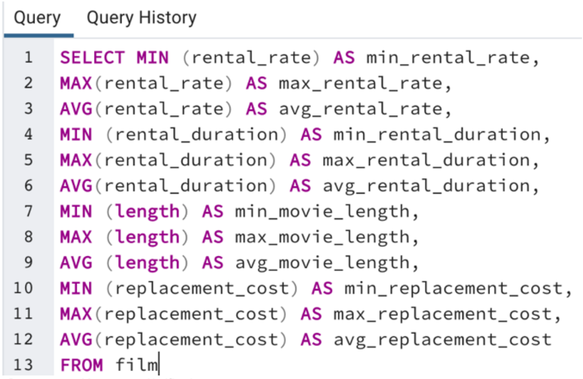
Step1: take a look what the table look like and to determine which are numeric column, which are not.



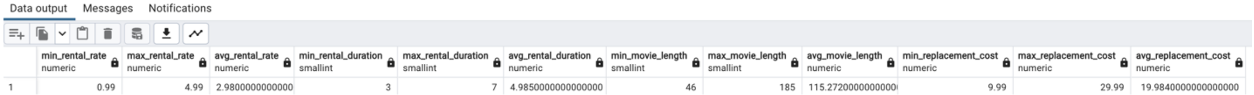
```
SELECT * FROM film;
```

	film_id [PK] integer	title character varying (255)	description text	release_year integer	language_id smallint	rental_duration smallint	rental_rate numeric (4,2)	length smallint	replacement_cost numeric (5,2)	rating mpaa_rating	last_update timestamp without time zone	special_features text[]	fulltext tsvector
1	133	Chamber Italian	A Fateful R...	2006	1	7	4.99	117	14.99	NC-17	2013-05-26 14:50:58.951	{Trailers}	'chamber'...
2	384	Grosse Wonderful	A Epic Dra...	2006	1	5	4.99	49	19.99	R	2013-05-26 14:50:58.951	{"Behind the Sce...	'australia':...
3	8	Airport Pollock	A Epic Tale ...	2006	1	6	4.99	54	15.99	R	2013-05-26 14:50:58.951	{Trailers}	'airport':1 ...
4	98	Bright Encounters	A Fateful Y...	2006	1	4	4.99	73	12.99	PG-13	2013-05-26 14:50:58.951	{Trailers}	'boat':20 '...
5	1	Academy Dinosaur	A Epic Dra...	2006	1	6	0.99	86	20.99	PG	2013-05-26 14:50:58.951	{"Deleted Scene...	'academi':...
6	2	Ace Goldfinger	A Astoundi...	2006	1	3	4.99	48	12.99	G	2013-05-26 14:50:58.951	{Trailers,"Delete...	'ace':1 'ad...
7	3	Adaptation Holes	A Astoundi...	2006	1	7	2.99	50	18.99	NC-17	2013-05-26 14:50:58.951	{Trailers,"Delete...	'adapt':1 '...
8	4	Affair Prejudice	A Fanciful ...	2006	1	5	2.99	117	26.99	G	2013-05-26 14:50:58.951	{Commentaries,"...	'affair':1 'c...
9	5	African Egg	A Fast-Pac...	2006	1	6	2.99	130	22.99	G	2013-05-26 14:50:58.951	{"Deleted Scene...	'african':1 ...
10	6	Agent Truman	A Intrepid ...	2006	1	3	2.99	169	17.99	PG	2013-05-26 14:50:58.951	{"Deleted Scene...	'agent':1 '...
11	7	Airplane Sierra	A Touching...	2006	1	6	4.99	62	28.99	PG-13	2013-05-26 14:50:58.951	{Trailers,"Delete...	'airplan':1 ...

Step2 Summarized the numeric records

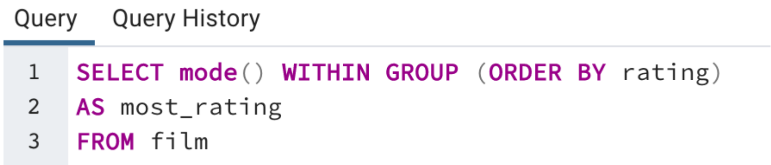


```sql
SELECT MIN (rental_rate) AS min_rental_rate,
MAX(rental_rate) AS max_rental_rate,
AVG(rental_rate) AS avg_rental_rate,
MIN (rental_duration) AS min_rental_duration,
MAX(rental_duration) AS max_rental_duration,
AVG(rental_duration) AS avg_rental_duration,
MIN (length) AS min_movie_length,
MAX (length) AS max_movie_length,
AVG (length) AS avg_movie_length,
MIN (replacement_cost) AS min_replacement_cost,
MAX(replacement_cost) AS max_replacement_cost,
AVG(replacement_cost) AS avg_replacement_cost
FROM film
```

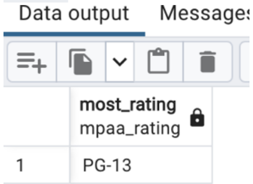


	min_rental_rate numeric	max_rental_rate numeric	avg_rental_rate numeric	min_rental_duration smallint	max_rental_duration smallint	avg_rental_duration numeric	min_movie_length smallint	max_movie_length smallint	avg_movie_length numeric	min_replacement_cost numeric	max_replacement_cost numeric	avg_replacement_cost numeric
1	0.99	4.99	2.9800000000000	3	7	4.9850000000000000	46	185	115.2720000000000	9.99	29.99	19.9840000000000000

Step 3 Summarized non-numeric records



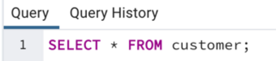
```sql
SELECT mode() WITHIN GROUP (ORDER BY rating)
AS most_rating
FROM film
```



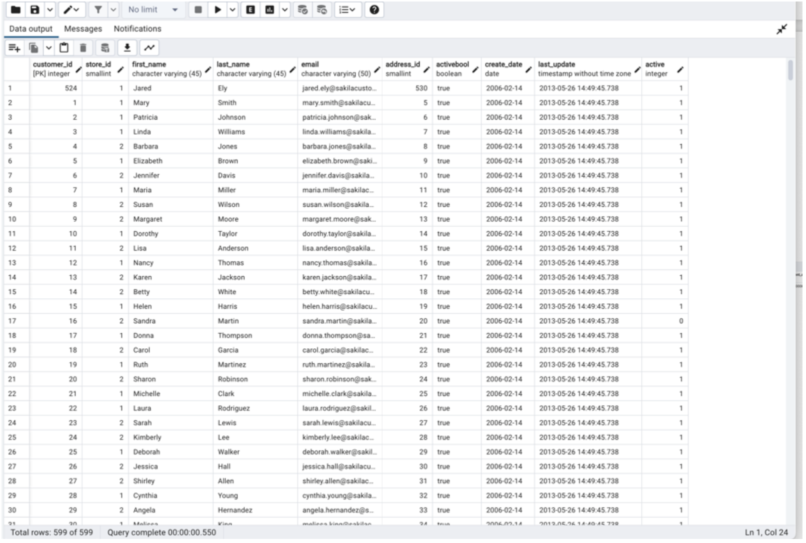
	most_rating mpaa_rating
1	PG-13

Customer table

Step1: take a look what the table look like and to determine which are numeric column, which are not.



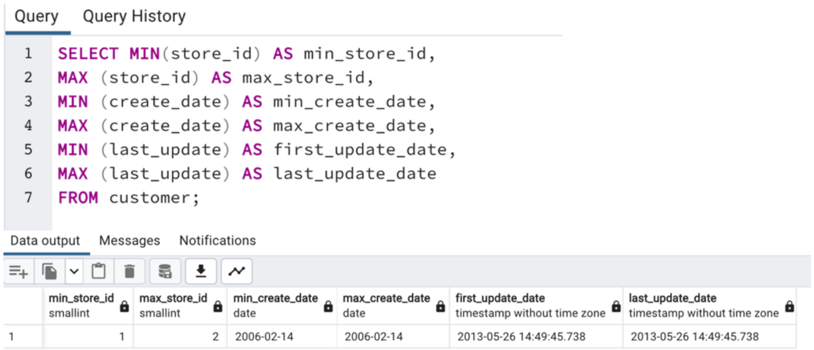
```sql
SELECT * FROM customer;
```



	customer_id [PK] integer	store_id smallint	first_name character varying (45)	last_name character varying (45)	email character varying (50)	address_id smallint	activebool boolean	create_date date	last_update timestamp without time zone	active integer
1	524	1	Jared	Ely	jared.ely@sakilacusto...	530	true	2006-02-14	2013-05-26 14:49:45.738	1
2	1	1	Mary	Smith	mary.smith@sakilacu...	5	true	2006-02-14	2013-05-26 14:49:45.738	1
3	2	1	Patricia	Johnson	patricia.johnson@sak...	6	true	2006-02-14	2013-05-26 14:49:45.738	1
4	3	1	Linda	Williams	linda.williams@sakila...	7	true	2006-02-14	2013-05-26 14:49:45.738	1
5	4	2	Barbara	Jones	barbara.jones@sakila...	8	true	2006-02-14	2013-05-26 14:49:45.738	1
6	5	1	Elizabeth	Brown	elizabeth.brown@saki...	9	true	2006-02-14	2013-05-26 14:49:45.738	1
7	6	2	Jennifer	Davis	jennifer.davis@sakila...	10	true	2006-02-14	2013-05-26 14:49:45.738	1
8	7	1	Maria	Miller	maria.miller@sakilac...	11	true	2006-02-14	2013-05-26 14:49:45.738	1
9	8	2	Susan	Wilson	susan.wilson@sakila...	12	true	2006-02-14	2013-05-26 14:49:45.738	1
10	9	2	Margaret	Moore	margaret.moore@sak...	13	true	2006-02-14	2013-05-26 14:49:45.738	1
11	10	1	Dorothy	Taylor	dorothy.taylor@sakila...	14	true	2006-02-14	2013-05-26 14:49:45.738	1
12	11	2	Lisa	Anderson	lisa.anderson@sakila...	15	true	2006-02-14	2013-05-26 14:49:45.738	1
13	12	1	Nancy	Thomas	nancy.thomas@sakila...	16	true	2006-02-14	2013-05-26 14:49:45.738	1
14	13	2	Karen	Jackson	karen.jackson@sakila...	17	true	2006-02-14	2013-05-26 14:49:45.738	1
15	14	2	Betty	White	betty.white@sakilacu...	18	true	2006-02-14	2013-05-26 14:49:45.738	1
16	15	1	Helen	Harris	helen.harris@sakilacu...	19	true	2006-02-14	2013-05-26 14:49:45.738	1
17	16	2	Sandra	Martin	sandra.martin@sakila...	20	true	2006-02-14	2013-05-26 14:49:45.738	0
18	17	1	Donna	Thompson	donna.thompson@sa...	21	true	2006-02-14	2013-05-26 14:49:45.738	1
19	18	2	Carol	Garcia	carol.garcia@sakilac...	22	true	2006-02-14	2013-05-26 14:49:45.738	1
20	19	1	Ruth	Martinez	ruth.martinez@sakila...	23	true	2006-02-14	2013-05-26 14:49:45.738	1
21	20	2	Sharon	Robinson	sharon.robinson@sak...	24	true	2006-02-14	2013-05-26 14:49:45.738	1
22	21	1	Michelle	Clark	michelle.clark@sakila...	25	true	2006-02-14	2013-05-26 14:49:45.738	1
23	22	1	Laura	Rodriguez	laura.rodriguez@sakil...	26	true	2006-02-14	2013-05-26 14:49:45.738	1
24	23	2	Sarah	Lewis	sarah.lewis@sakilacu...	27	true	2006-02-14	2013-05-26 14:49:45.738	1
25	24	2	Kimberly	Lee	kimberly.lee@sakilac...	28	true	2006-02-14	2013-05-26 14:49:45.738	1
26	25	1	Deborah	Walker	deborah.walker@sakil...	29	true	2006-02-14	2013-05-26 14:49:45.738	1
27	26	2	Jessica	Hall	jessica.hall@sakilacu...	30	true	2006-02-14	2013-05-26 14:49:45.738	1
28	27	2	Shirley	Allen	shirley.allen@sakilac...	31	true	2006-02-14	2013-05-26 14:49:45.738	1
29	28	1	Cynthia	Young	cynthia.young@sakila...	32	true	2006-02-14	2013-05-26 14:49:45.738	1
30	29	2	Angela	Hernandez	angela.hernandez@s...	33	true	2006-02-14	2013-05-26 14:49:45.738	1
31	30	1	Melissa	King	melissa.king@sakilac...	34	true	2006-02-14	2013-05-26 14:49:45.738	1

Total rows: 599 of 599 Query complete 00:00:00.550 Ln 1, Col 24

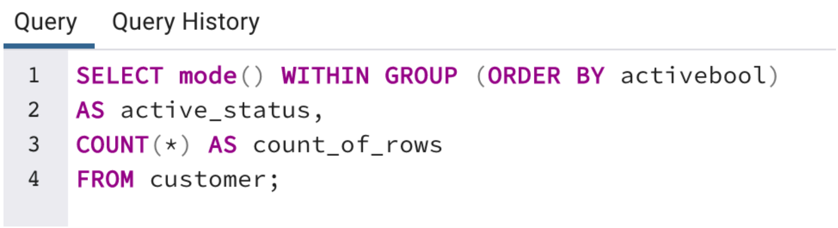
Step2 Summarized the numeric records



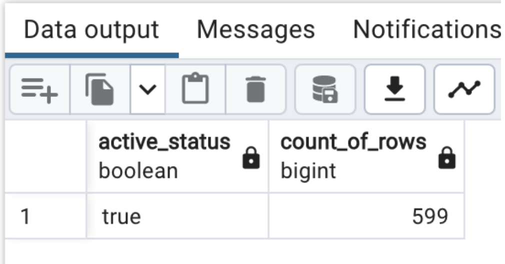
```
SELECT MIN(store_id) AS min_store_id,
MAX (store_id) AS max_store_id,
MIN (create_date) AS min_create_date,
MAX (create_date) AS max_create_date,
MIN (last_update) AS first_update_date,
MAX (last_update) AS last_update_date
FROM customer;
```

	min_store_id smallint	max_store_id smallint	min_create_date date	max_create_date date	first_update_date timestamp without time zone	last_update_date timestamp without time zone
1	1	2	2006-02-14	2006-02-14	2013-05-26 14:49:45.738	2013-05-26 14:49:45.738

Step 3 Summarized non-numeric records



```
SELECT mode() WITHIN GROUP (ORDER BY activebool)
AS active_status,
COUNT(*) AS count_of_rows
FROM customer;
```



	active_status boolean	count_of_rows bigint
1	true	599

3. It depends on the size of the database. Excel works perfectly with small data size. With this it will be easy to view the data using the pivot table. On the other hand, it is easy to work with huge data in SQL. Using SQL data profiling becomes much easier and faster. With SQL once the code has been written it can be applied time and again without much effort.