

1.) Check For and Clean Dirty Data

Film Table

Duplicate Data

DashboardPropertiesSQLStatisticsDependenciesDependentsProcessesRockbuster/postgres@PostgreSQL 15*

Rockbuster/postgres@PostgreSQL 15

The session is idle and there is no current transaction.

QueryQuery History

1SELECT

2film_id,

3title,

4description,

5release_year,

6language_id,

7rental_duration,

8rental_rate,

9length,

10replacement_cost,

11rating,

12last_update,

13special_features,

14fulltext,

15COUNT(*)

16FROM film

17GROUP BY

18film_id,

19title,

20description,

21release_year,

22language_id,

23rental_duration,

24rental_rate,

25length,

26replacement_cost,

27rating,

28last_update,

29special_features,

30fulltext,

31COUNT(*)

32HAVING COUNT (*) > 1;

Scratch Pad

Data OutputMessagesNotifications

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

count

bigint

The above query returned no duplicate values, however, I could have used the CREATE VIEW query down below to pull up a view that would return any existing duplicate values:

Rockbuster/postgres@PostgreSQL 15



Query Query History

```
1 CREATE VIEW film_duplicate_records AS
2 SELECT
3     film_id,
4     title,
5     description,
6     release_year,
7     language_id,
8     rental_duration,
9     rental_rate,
10    length,
11    replacement_cost,
12    rating,
13    last_update,
14    special_features,
15    fulltext
16 FROM film
17 GROUP BY film_id,
18     title,
19     description,
20     release_year,
21     language_id,
22     rental_duration,
23     rental_rate,
24     length,
25     replacement_cost,
26     rating,
27     last_update,
28     special_features,
29     fulltext
```

Data Output Messages Notifications

CREATE VIEW

Query returned successfully in 78 msec.

Non-Uniform Data

Looking through the only category that could have non-uniform entries for the same value (i.e. activebool), we find that there are no such entries. In fact, there were the same number of distinct values found as there total rows, as demonstrated in the two queries below.

DashboardPropertiesSQLStatisticsDependenciesDependentsProcessesRockbuster/postgres@PostgreSQL 15*

Rockbuster/postgres@PostgreSQL 15

Query Query History

Scratch Pad

```
1 SELECT DISTINCT file_id,
2 title,
3 description,
4 release_year,
5 language_id,
6 rental_duration,
7 rental_rate,
8 length,
9 replacement_cost,
10 rating,
11 last_update,
12 special_features,
13 fulltext
14 FROM film;
```

Data OutputMessagesNotifications

file_id	title	description	release_year	language_id	rental_duration	rental_rate	length	replacement_cost	rating	last_update	special_features
[PK] integer	character varying (255)	text	integer	smallint	smallint	numeric (4,2)	smallint	numeric (12,2)	enum_rating	timestamp without time zone	text[]
1	Academy Dinosaur	A Epic Drama of a Feminist And a Mad Scientist who must Battle a Teacher in The Canadian Rockies	2006	1	6	0.99	86	20.99	PG	2013-05-26 14:50:58.951	["Deleted Scenes","Behind the Scenes"]
2	Ace Goldfinger	A Surrounding Exsile of a Database Administrator And a Explorer who must Find a Car in Ancient China	2006	1	3	4.99	48	12.99	G	2013-05-26 14:50:58.951	["Deleted Scenes","Behind the Scenes"]
3	Adaptation Holes	A Astounding Reflection of a Lumberjack And a Car who must Sink a Lumberjack in A Baloon Factory	2006	1	7	2.99	50	18.99	NC-17	2013-05-26 14:50:58.951	["Trailers","Deleted Scenes"]
4	Affair Prejudice	A Fanciful Documentary of a Frisbee And a Lumberjack who must Chase a Monkey in A Shark Tank	2006	1	5	2.99	117	26.99	G	2013-05-26 14:50:58.951	["Commentaries","Behind the Scenes"]
5	African Egg	A Fast-Paced Documentary of a Pastry Chef And a Dentist who must Pursue a Forensic Psychologist in The Gulf of Mexico	2006	1	6	2.99	130	22.99	G	2013-05-26 14:50:58.951	["Deleted Scenes"]
6	Agent Truman	A Intrepid Panorama of a Robot And a Boy who must Escape a Sarno Wrestler in Ancient China	2006	1	3	2.99	169	17.99	PG	2013-05-26 14:50:58.951	["Trailers","Deleted Scenes"]
7	Ajaccio Serna	A Touching Saga of a Hunter And a Butler who must Discover a Butler in A Jet Boat	2006	1	6	4.99	67	28.99	PG-13	2013-05-26 14:50:58.951	["Trailers","Deleted Scenes"]
8	Airport Pollock	A Epic Tale of a Moose And a Girl who must Confront a Monkey in Ancient India	2006	1	6	4.99	54	15.99	R	2013-05-26 14:50:58.951	["Trailers"]
9	Alabama Devil	A Thoughtful Panorama of a Database Administrator And a Mad Scientist who must Outgun a Mad Scientist in A Jet Boat	2006	1	3	2.99	114	21.99	PG-13	2013-05-26 14:50:58.951	["Trailers","Deleted Scenes"]
10	Aladdin Calendar	A Action-Packed Tale of a Man And a Lumberjack who must Reach a Feminist in Ancient China	2006	1	6	4.99	63	24.99	NC-17	2013-05-26 14:50:58.951	["Trailers","Deleted Scenes"]
11	Alamo Videotape	A Boring Epistle of a Butler And a Cat who must Fight a Pastry Chef in A Mythical Convention	2006	1	6	0.99	126	16.99	G	2013-05-26 14:50:58.951	["Commentaries","Behind the Scenes"]
12	Alaska Phantom	A Fanciful Saga of a Hunter And a Pastry Chef who must Vanquish a Boy in Australia	2006	1	6	0.99	136	22.99	PG	2013-05-26 14:50:58.951	["Commentaries","Deleted Scenes"]
13	All Forever	A Action-Packed Drama of a Dentist And a Crocodile who must Battle a Feminist in The Canadian Rockies	2006	1	4	4.99	150	21.99	PG	2013-05-26 14:50:58.951	["Deleted Scenes","Behind the Scenes"]
14	Alice Fantasia	A Emotional Drama of a A Shark And a Database Administrator who must Vanquish a Pioneer in Soviet Georgia	2006	1	6	0.99	84	23.99	NC-17	2013-05-26 14:50:58.951	["Trailers","Deleted Scenes","Behind the Scenes"]
15	Alien Center	A Brilliant Drama of a Car And a Mad Scientist who must Battle a Feminist in A Mythical Convention	2006	1	5	2.99	46	10.99	NC-17	2013-05-26 14:50:58.951	["Trailers","Commentaries","Behind the Scenes"]
16	Alley Evolution	A Fast-Paced Drama of a Robot And a Composer who must Battle a Astronaut in New Orleans	2006	1	6	2.99	180	23.99	NC-17	2013-05-26 14:50:58.951	["Trailers","Commentaries"]
17	Alone Trip	A Fast-Paced Character Study of a Composer And a Dog who must Outgun a Boat in An Abandoned Fun House	2006	1	3	0.99	82	14.99	R	2013-05-26 14:50:58.951	["Trailers","Behind the Scenes"]
18	Alter Victory	A Thoughtful Drama of a Composer And a Feminist who must Meet a Secret Agent in The Canadian Rockies	2006	1	6	0.99	57	27.99	PG-13	2013-05-26 14:50:58.951	["Trailers","Behind the Scenes"]
19	Amadeus Holy	A Emotional Display of a Pioneer And a Technical Writer who must Battle a Man in A Baloon	2006	1	6	0.99	113	20.99	PG	2013-05-26 14:50:58.951	["Commentaries","Deleted Scenes","Behind the Scenes"]
20	Ametie Heffluggers	A Boring Drama of a Woman And a Squirrel who must Conquer a Student in A Baloon	2006	1	4	4.99	79	23.99	R	2013-05-26 14:50:58.951	["Commentaries","Deleted Scenes","Behind the Scenes"]
21	American Circus	A Insightful Drama of a Girl And a Astronaut who must Face a Database Administrator in A Shark Tank	2006	1	3	4.99	129	17.99	R	2013-05-26 14:50:58.951	["Commentaries","Behind the Scenes"]
22	Amisrael Midwinter	A Emotional Character Study of a Dentist And a Crocodile who must Meet a Sarno Wrestler in California	2006	1	6	2.99	85	10.99	G	2013-05-26 14:50:58.951	["Commentaries","Behind the Scenes"]
23	Aracanda Confessions	A Lackluster Display of a Dentist And a Dentist who must Fight a Girl in Australia	2006	1	3	0.99	92	9.99	R	2013-05-26 14:50:58.951	["Trailers","Deleted Scenes"]
24	Arayake Heisters	A Thoughtful Display of a Explorer And a Pastry Chef who must Overcome a Feminist in The Sahara Desert	2006	1	6	2.99	181	19.99	R	2013-05-26 14:50:58.951	["Trailers","Behind the Scenes"]
25	Angels Life	A Thoughtful Display of a Woman And a Astronaut who must Battle a Robot in Berlin	2006	1	3	2.99	74	15.99	G	2013-05-26 14:50:58.951	["Trailers"]
26	Annie Identity	A Amazing Panorama of a Pastry Chef And a Boat who must Escape a Woman in An Abandoned Amusement Park	2006	1	3	0.99	86	15.99	G	2013-05-26 14:50:58.951	["Commentaries","Deleted Scenes"]
27	Anonymous Human	A Amazing Reflection of a Database Administrator And a Astronaut who must Outrace a Database Administrator in A Shark Tank	2006	1	7	0.99	179	12.99	NC-17	2013-05-26 14:50:58.951	["Deleted Scenes","Behind the Scenes"]
28	Anthem Luke	A Touching Panorama of a Walrus And a Woman who must Outrace a Dog in An Abandoned Amusement Park	2006	1	5	4.99	91	16.99	PG-13	2013-05-26 14:50:58.951	["Deleted Scenes","Behind the Scenes"]
29	Antarctic Tension	A Fanciful Saga of a Hunter And a Feminist who must Pursue a Database Administrator in Ancient India	2006	1	4	4.99	149	14.99	NC-17	2013-05-26 14:50:58.951	["Deleted Scenes","Behind the Scenes"]

Total rows: 1000 of 1000Query complete 00:00:00.163Ln 13, Col 13

Missing Data

This query did not return any null values in any of the selected columns, meaning there are no missing values.

Dashboard Properties SQL Statistics Dependencies Processes Rockbuster/postgres@PostgreSQL 15*

Rockbuster/postgres@PostgreSQL 15

Query Query History Scratch Pad x

```

1 SELECT film_id,
2        title,
3        description,
4        release_year,
5        language_id,
6        rental_duration,
7        rental_rate,
8        length,
9        replacement_cost,
10       rating,
11       last_update,
12       special_features,
13       fulltext
14 FROM film
15 WHERE film_id IS NULL OR
16        title IS NULL OR
17        description IS NULL OR
18        release_year IS NULL OR
19        language_id IS NULL OR
20        rental_duration IS NULL OR
21        rental_rate IS NULL OR
22        length IS NULL OR
23        replacement_cost IS NULL OR
24        rating IS NULL OR
25        last_update IS NULL OR
26        special_features IS NULL OR
27        fulltext IS NULL
    
```

Data Output Messages Notifications

film_id	title	description	release_year	language_id	rental_duration	rental_rate	length	replacement_cost	rating	last_update	special_features	fulltext
[PK] integer	character varying (255)	text	integer	smallint	smallint	numeric (4,2)	smallint	numeric (5,2)	mpaa_rating	timestamp without time zone	text[]	tvector

Customer Table

Duplicate Data

Dashboard
Statistics
Dependencies
Processes
Rockbuster/postgres@PostgreSQL 15*

Rockbuster/postgres@PostgreSQL 15

No limit

Query
Query History

Scratch Pad

```

1  SELECT customer_id,
2         store_id,
3         first_name,
4         last_name,
5         email,
6         address_id,
7         activebool,
8         create_date,
9         last_update,
10        active,
11        COUNT (*)
12 FROM customer
13 GROUP BY customer_id,
14          store_id,
15          first_name,
16          last_name,
17          email,
18          address_id,
19          activebool,
20          create_date,
21          last_update,
22          active
23 HAVING COUNT(*) > 1

```

Data Output
Messages
Notifications

customer_id
store_id
first_name
last_name
email
address_id
activebool
create_date
last_update
active
count

[PK] integer
smallint
character varying (45)
character varying (45)
character varying (50)
smallint
boolean
date
timestamp without time zone
integer
bigint

The above query returned no duplicate values, however, I could have used the CREATE VIEW query down below to pull up a view that would return any existing duplicate values:

Dashboard Properties SQL Statistics Dependencies Dependents Processes Rockbuster/postgres@PostgreSQL 15*

Rockbuster/postgres@PostgreSQL 15

No limit

Query Query History Scratch Pad x

```

1 CREATE VIEW customer_duplicate_records AS
2 SELECT
3     customer_id,
4     store_id,
5     first_name,
6     last_name,
7     email,
8     address_id,
9     activebool,
10    create_date,
11    last_update,
12    active
13 FROM customer
14 GROUP BY
15     customer_id,
16     store_id,
17     first_name,
18     last_name,
19     email,
20     address_id,
21     activebool,
22     create_date,
23     last_update,
24     active
    
```

Data Output Messages Notifications

customer_id	store_id	first_name	last_name	email	address_id	activebool	create_date	last_update	active	count
[PK] integer	smallint	character varying (45)	character varying (45)	character varying (50)	smallint	boolean	date	timestamp without time zone	integer	bigint

Non-Uniform Data

Looking through the only category that could have non-uniform entries for the same value (i.e. `activebool`), we find that there are no such entries. In fact, there were the same number of distinct values found as there total rows, as demonstrated in the two queries below.

Rockbuster/postgres@PostgreSQL 15*

Query History

```

1 SELECT DISTINCT customer_id,
2                 store_id,
3                 first_name,
4                 last_name,
5                 email,
6                 address_id,
7                 activebool,
8                 create_date,
9                 last_update,
10                active
11 FROM customer;

```

Data Output Messages Notifications

	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	357	1	Keith	Rico	keith.rico@sakilacustomer.org	362	true	2006-02-14	2013-05-26 14:49:45.738	1
2	171	2	Dolores	Wagner	dolores.wagner@sakilacustomer.org	175	true	2006-02-14	2013-05-26 14:49:45.738	1
3	139	1	Amber	Dixon	amber.dixon@sakilacustomer.org	143	true	2006-02-14	2013-05-26 14:49:45.738	1
4	471	1	Dean	Sauer	dean.sauer@sakilacustomer.org	476	true	2006-02-14	2013-05-26 14:49:45.738	1
5	594	1	Eduardo	Hiatt	eduardo.hiatt@sakilacustomer.org	600	true	2006-02-14	2013-05-26 14:49:45.738	1
6	401	2	Tony	Carranza	tony.carranza@sakilacustomer.org	406	true	2006-02-14	2013-05-26 14:49:45.738	1
7	157	2	Darlene	Rose	darlene.rose@sakilacustomer.org	161	true	2006-02-14	2013-05-26 14:49:45.738	1
8	154	2	Michele	Grant	michele.grant@sakilacustomer.org	158	true	2006-02-14	2013-05-26 14:49:45.738	1
9	530	2	Darryl	Ashcraft	darryl.ashcraft@sakilacustomer.org	536	true	2006-02-14	2013-05-26 14:49:45.738	1
10	493	1	Brent	Harkins	brent.harkins@sakilacustomer.org	498	true	2006-02-14	2013-05-26 14:49:45.738	1
11	542	2	Lonnies	Tirado	lonnie.tirado@sakilacustomer.org	548	true	2006-02-14	2013-05-26 14:49:45.738	1
12	566	1	Casey	Mena	casey.mena@sakilacustomer.org	572	true	2006-02-14	2013-05-26 14:49:45.738	1
13	186	2	Holly	Fox	holly.fox@sakilacustomer.org	190	true	2006-02-14	2013-05-26 14:49:45.738	1
14	128	1	Marjorie	Tucker	marjorie.tucker@sakilacustomer.org	132	true	2006-02-14	2013-05-26 14:49:45.738	1
15	466	1	Leo	Ebert	leo.ebert@sakilacustomer.org	471	true	2006-02-14	2013-05-26 14:49:45.738	1
16	494	2	Ramon	Choate	ramon.choate@sakilacustomer.org	499	true	2006-02-14	2013-05-26 14:49:45.738	1
17	178	2	Marion	Snyder	marion.snyder@sakilacustomer.org	182	true	2006-02-14	2013-05-26 14:49:45.738	1
18	65	2	Rose	Howard	rose.howard@sakilacustomer.org	69	true	2006-02-14	2013-05-26 14:49:45.738	1
19	450	1	Jay	Robb	jay.robb@sakilacustomer.org	455	true	2006-02-14	2013-05-26 14:49:45.738	1
20	234	1	Claudia	Fuller	claudia.fuller@sakilacustomer.org	238	true	2006-02-14	2013-05-26 14:49:45.738	1

Total rows: 599 of 599 Query complete 00:00:00.249

Missing Data

This query did not return any null values in any of the selected columns, meaning there are no missing values.

Dashboard Properties SQL Statistics Dependencies Dependent Processes **Rockbuster/postgres@PostgreSQL 15***

Rockbuster/postgres@PostgreSQL 15

No limit

Query Query History

```

1 SELECT customer_id,
2         store_id,
3         first_name,
4         last_name,
5         email,
6         address_id,
7         activebool,
8         create_date,
9         last_update,
10        active
11 FROM customer
12 WHERE customer_id IS NULL OR
13        store_id IS NULL OR
14        first_name IS NULL OR
15        last_name IS NULL OR
16        email IS NULL OR
17        address_id IS NULL OR
18        activebool IS NULL OR
19        create_date IS NULL OR
20        last_update IS NULL OR
21        active IS NULL
22

```

Data Output Messages Notifications

customer_id	store_id	first_name	last_name	email	address_id	activebool	create_date	last_update	active
[PK] integer	smallint	character varying (45)	character varying (45)	character varying (50)	smallint	boolean	date	timestamp without time zone	integer

2.) Summarize Your Data

Film Table

```
Rockbuster/postgres@PostgreSQL 15
Query History
1 SELECT
2     MIN(film_id),
3     MAX(film_id),
4     AVG(film_id),
5
6 MODE () WITHIN GROUP (ORDER BY title)
7     AS title_mode,
8
9 MODE () WITHIN GROUP (ORDER BY title)
10    AS title_mode,
11
12    MIN(release_year),
13    MAX(release_year),
14    AVG(release_year),
15
16 MODE () WITHIN GROUP (ORDER BY language_id)
17    AS language_mode,
18
19    MIN(rental_duration),
20    MAX(rental_duration),
21    AVG(rental_duration),
22
23    MIN(rental_rate),
24    MAX(rental_rate),
25    AVG(rental_rate),
26
27    MIN(length),
28    MAX(length),
29    AVG(length),
30
31    MIN(replacement_cost),
32    MAX(replacement_cost),
33    AVG(replacement_cost),
34
35 MODE () WITHIN GROUP (ORDER BY rating)
36    AS rating_mode,
37
38    MIN(last_update),
39    MAX(last_update),
40
41 MODE () WITHIN GROUP (ORDER BY special_features)
42    AS special_features_mode,
43
44 MODE () WITHIN GROUP (ORDER BY fulltext)
45    AS fulltext_mode
46 FROM film;
```

	min integer	max integer	avg numeric	title_mode character varying	title_mode character varying	min integer	max integer	avg numeric	language_mode smallint	min smallint	max smallint	avg numeric	min numeric	max numeric	avg numeric
1	1	1000	500.5	Academy Dinosaur	Academy Dinosaur	2006	2006	2006	1	3	7	4.985	0.99	4.99	2.98
min smallint	max smallint	avg numeric	min numeric	max numeric	avg numeric	rating_mode mpaa_rating	min timestamp without time zone	max timestamp without time zone	special_features_mode text[]	fulltext_mode tsvector					
46	185	115.272	9.99	29.99	19.984	PG-13	2013-05-26 14:50:58.951	2013-05-26 14:50:58.951	(Trailers,Commentaries,'Behind the Scenes')	'baloon';'19 'confront':14 'documentari':5 'feminist':8;11;'6 'mle':2 'must':13 'sp!':1 thrl...					

Postgre would not allow me to run the `AVG()` function on the `last_update` field as it is a date function.

Customer Table

Dashboard
Properties
SQL
Statistics
Dependencies
Dependents
Processes
Rockbuster/postgres@PostgreSQL 15*

Rockbuster/postgres@PostgreSQL 15

No limit

Query
Query History

```

1  SELECT
2      MIN(customer_id),
3      MAX(customer_id),
4      AVG(customer_id),
5
6      MIN(store_id),
7      MAX(store_id),
8      AVG(store_id),
9
10     MODE () WITHIN GROUP (ORDER BY first_name)
11     AS mode_first_name,
12
13     MODE () WITHIN GROUP (ORDER BY last_name)
14     AS mode_last_name,
15
16     MODE () WITHIN GROUP (ORDER BY email)
17     AS mode_email,
18
19     MIN(address_id),
20     MAX(address_id),
21     AVG(address_id),
22
23     MODE () WITHIN GROUP (ORDER BY activebool)
24     AS mode_active,
25
26     MIN(create_date),
27     MAX(create_date),
28
29     MIN(last_update),
30     MAX(last_update)
31
32 FROM customer;

```

Data Output
Messages
Notifications

	min integer	max integer	avg numeric	min smallint	max smallint	avg numeric	mode_first_name character varying	mode_last_name character varying	mode_email character varying	min smallint	max smallint
1	1	599	300	1	2	1.4557595993322203	Jamie	Abney	aaron.selby@sakilacustomer.org	5	605

Postgre would not allow me to run the `AVG()` function on the `create_date` and `last_update` fields as they are in date formats.

3.) Reflect

When adjusting for the learning curve as it relates to SQL functions, I found SQL to be a quicker method for profiling the data of these two, respective tables because I was able to write queries for both in the same window (i.e. the Postgre Query writing area) and I was able to create these queries without creating Pivot Tables, which would have required me to go into each file, separately and create unique Pivot Tables.