

# Summarizing & Cleaning – Rockbuster Stealth Data Analysis (SQL)

## Project

Ideal Kuci

- “Film” table searched for *duplicates*:

The screenshot shows a SQL query editor with the following query:

```
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 film_imdb,  
15 COUNT(*)  
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,  
30 film_imdb  
31 HAVING COUNT(*) > 1;
```

The query is designed to find rows in the 'film' table that have identical values across all specified columns, indicating duplicates. The 'COUNT(\*)' function is used to count the number of rows for each group of identical rows, and the 'HAVING COUNT(\*) > 1' clause filters out groups with only one row.

- “Customer” table searched for *duplicates*:

The screenshot shows a SQL query editor with the following query:

```
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;
```

The query is designed to find rows in the 'customer' table that have identical values across all specified columns, indicating duplicates. The 'COUNT(\*)' function is used to count the number of rows for each group of identical rows, and the 'HAVING COUNT(\*) > 1' clause filters out groups with only one row.

- “Film” table, query used for *non – uniform* check:

GROUP BY (I also used DISTINCT below for the sake of comparison):

Dashboard Properties SQL Statistics Dependencies [Rockbuster/postgres@PostgreSQL 15\\*](#)

Rockbuster/postgres@PostgreSQL 15

Query Query History

```

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      film_imdb
15 FROM film
16 GROUP BY film_id,
17         title,
18         description,
19         release_year,
20         language_id,
21         rental_duration,
22         rental_rate,
23         length,
24         replacement_cost,
25         rating,
26         last_update,
27         special_features,
28         fulltext,
29         film_imdb

```

Data Output Messages Notifications

	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	652	Pajama Jawbreaker	A Emotional Drama of a ...	2006	1	3	0.99	126	14.99	R	2013-05-26 14:50:58.951	(Trailers,'Deleted S...	'boy'8 'california'20 'dra
2	273	Effect Gladiator	A Beautiful Display of a ...	2006	1	6	0.99	107	14.99	PG	2013-05-26 14:50:58.951	(Commentaries)	'beauti'4 'chef'9.13 'disp
3	51	Balloon Homeward	A Insightful Panorama o...	2006	1	5	2.99	75	10.99	NC-17	2013-05-26 14:50:58.951	('Deleted Scenes')	'balloon'1 'build'16 'cow
4	951	Voyage Legally	A Epic Tale of a Squirrel ...	2006	1	6	0.99	78	28.99	PG-13	2013-05-26 14:50:58.951	(Commentaries,'Be...	'abandon'19 'boy'16 'co
5	839	Stallion Sundance	A Fast-Paced Tale of a C...	2006	1	5	0.99	130	23.99	PG-13	2013-05-26 14:50:58.951	('Deleted Scenes',...	'australia'21 'car'10 'dog
6	70	Bikini Borrowers	A Astounding Drama of ...	2006	1	7	4.99	142	26.99	NC-17	2013-05-26 14:50:58.951	(Commentaries,'De...	'astound'4 'astronaut'8
7	350	Garden Island	A Unbelievable Charact...	2006	1	3	4.99	80	21.99	G	2013-05-26 14:50:58.951	(Trailers,'Behind th...	'character'5 'garden'1 'isla
8	758	Saints Bride	A Fateful Tale of a Tech...	2006	1	5	2.99	125	11.99	G	2013-05-26 14:50:58.951	('Deleted Scenes')	'bride'2 'compos'12 'exp
9	539	Luck Opus	A Boring Display of a Mo...	2006	1	7	2.99	152	21.99	NC-17	2013-05-26 14:50:58.951	(Trailers,Comment...	'bore'4 'display'5 'luck'1
10	874	Tadpole Park	A Beautiful Tale of a Fris...	2006	1	6	2.99	155	13.99	PG	2013-05-26 14:50:58.951	(Trailers,Comment...	'abandon'19 'amus'20 't
11	278	Erf Murder	A Action-Packed Story o...	2006	1	4	4.99	155	19.99	NC-17	2013-05-26 14:50:58.951	(Trailers,Comment...	'abandon'21 'action'5 'a
12	946	Virtual Spoilers	A Fateful Tale of a Datab...	2006	1	3	4.99	144	14.99	NC-17	2013-05-26 14:50:58.951	(Commentaries,'De...	'administr'9 'databas'8
13	176	Conspiracy Court	A Troubling Documentar...	2006	1	6	0.99	87	31.00	PG-13	2013-05-26 14:50:58.951	(Commentaries,'De...	'insulat'20 'sat'5 'chef'1

DISTINCT:

Dashboard Properties SQL Statistics Dependencies [Rockbuster/postgres@PostgreSQL 15\\*](#)

Rockbuster/postgres@PostgreSQL 15

Query Query History

```

1 SELECT DISTINCT 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      film_imdb
15 FROM film;

```

Data Output Messages Notifications

	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	film_imdb integer
1	1	Academy Dinosaur	A Epic Drama of a Femin...	2006	1	6	0.99	86	20.99	PG	2013-05-26 14:50...	('Deleted Scenes...	'academi'1 'battl'...	
2	2	Ace Goldfinger	A Astounding Epistle of ...	2006	1	3	4.99	48	12.99	G	2013-05-26 14:50...	(Trailers,'Deleted...	'ace'1 'administr'9...	
3	3	Adaptation Holes	A Astounding Reflection...	2006	1	7	2.99	50	18.99	NC-17	2013-05-26 14:50...	(Trailers,'Deleted...	'adapt'1 'astound'...	
4	4	Affair Prejudice	A Fanciful Documentary ...	2006	1	5	2.99	117	26.99	G	2013-05-26 14:50...	(Commentaries,'...	'affair'1 'chase'14...	
5	5	African Egg	A Fast-Paced Document...	2006	1	6	2.99	130	22.99	G	2013-05-26 14:50...	('Deleted Scenes')	'african'1 'chef'11...	
6	6	Agent Truman	A Intrepid Panorama of ...	2006	1	3	2.99	169	17.99	PG	2013-05-26 14:50...	('Deleted Scenes')	'agent'1 'ancient'1...	
7	7	Airplane Sierra	A Touching Saga of a Hu...	2006	1	6	4.99	62	28.99	PG-13	2013-05-26 14:50...	(Trailers,'Deleted...	'airplan'1 'boat'20...	
8	8	Airport Pollock	A Epic Tale of a Moose...	2006	1	6	4.99	54	15.99	R	2013-05-26 14:50...	(Trailers)	'airport'1 'ancient'...	
9	9	Alabama Devil	A Thoughtful Panorama ...	2006	1	3	2.99	114	21.99	PG-13	2013-05-26 14:50...	(Trailers,'Deleted...	'administ'9 'alaba...	
10	10	Aladdin Calendar	A Action-Packed Tale of ...	2006	1	6	4.99	63	24.99	NC-17	2013-05-26 14:50...	(Trailers,'Deleted...	'action'5 'action-p...	
11	11	Alamo Videotape	A Boring Epistle of a But...	2006	1	6	0.99	126	16.99	G	2013-05-26 14:50...	(Commentaries,'...	'alamo'1 'bore'4 b...	
12	12	Alaska Phantom	A Fanciful Saga of a Hu...	2006	1	6	0.99	136	22.99	PG	2013-05-26 14:50...	(Commentaries,'...	'alaska'1 'australia...	
13	13	All Forever	A Action-Packed Drama ...	2006	1	4	4.99	150	21.99	PG	2013-05-26 14:50...	('Deleted Scenes...	'action'5 'action-p...	
14	14	Alice Fantasia	A Emotional Drama of a ...	2006	1	6	0.99	94	23.99	NC-17	2013-05-26 14:50...	(Trailers,'Deleted...	'administ'13 'alic'...	
15	15	Alien Center	A Brilliant Drama of a Ca...	2006	1	5	2.99	46	10.99	NC-17	2013-05-26 14:50...	(Trailers,Comme...	'alien'1 'battl'15 b...	
16	16	Alien Evolution	A Fast-Paced Drama of ...	2006	1	6	2.99	180	23.99	NC-17	2013-05-26 14:50...	(Trailers,Comme...	'alien'1 'astronaut'...	

- “Customer” table, query used for *non – uniform* check:

DISTINCT:

Rockbuster/postgres@PostgreSQL 15

Query 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	Lonnie	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

GROUP BY:

Rockbuster/postgres@PostgreSQL 15

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 GROUP BY customer_id,
13     store_id,
14     first_name,
15     last_name,
16     email,
17     address_id,
18     activebool,
19     create_date,
20     last_update,
21    active;

```

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	184	1	Vivian	Ruiz	vivian.ruiz@sakilacustomer.org	188	true	2006-02-14	2013-05-26 14:49:45.738	1
2	87	1	Wanda	Patterson	wanda.patterson@sakilacustomer.org	91	true	2006-02-14	2013-05-26 14:49:45.738	1
3	477	1	Dan	Paine	dan.paine@sakilacustomer.org	482	true	2006-02-14	2013-05-26 14:49:45.738	1
4	273	2	Priscilla	Lowe	priscilla.lowe@sakilacustomer.org	278	true	2006-02-14	2013-05-26 14:49:45.738	1
5	550	2	Guy	Brownlee	guy.brownlee@sakilacustomer.org	556	true	2006-02-14	2013-05-26 14:49:45.738	1
6	394	2	Chris	Brothers	chris.brothers@sakilacustomer.org	399	true	2006-02-14	2013-05-26 14:49:45.738	1
7	51	1	Alice	Stewart	alice.stewart@sakilacustomer.org	55	true	2006-02-14	2013-05-26 14:49:45.738	1
8	272	1	Kay	Caldwell	kay.caldwell@sakilacustomer.org	277	true	2006-02-14	2013-05-26 14:49:45.738	1
9	70	2	Christina	Ramirez	christina.ramirez@sakilacustomer.org	74	true	2006-02-14	2013-05-26 14:49:45.738	1

- Film table descriptive statistics (numeric):

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

Rockbuster/postgres@PostgreSQL 15

Query Query History

```

1 SELECT MIN(rental_duration) AS min_rental_duration,
2        MAX(rental_duration) AS max_rental_duration,
3        AVG(rental_duration) AS avg_rental_duration,
4
5        MIN(rental_rate) AS min_rental_rate,
6        MAX(rental_rate) AS max_rental_rate,
7        AVG(rental_rate) AS avg_rental_rate,
8
9        MIN(length) AS min_length,
10       MAX(length) AS max_length,
11       AVG(length) AS avg_length,
12
13       MIN(replacement_cost) AS min_replacement_cost,
14       MAX(replacement_cost) AS max_replacement_cost,
15       AVG(replacement_cost) AS avg_replacement_cost
16
17 FROM film
18

```

Data Output Messages Notifications

	min_rental_duration smallint	max_rental_duration smallint	avg_rental_duration numeric	min_rental_rate numeric	max_rental_rate numeric	avg_rental_rate numeric	min_length smallint	max_length smallint	avg_length numeric	min_replacement_cost numeric	max_replacement_cost numeric	avg_replacement_cost numeric
1	3	7	4.9850000000000000	0.99	4.99	2.9800000000000000	46	185	115.2720000	9.99	29.99	19.9840000000000000

- Film table descriptive statistics (non-numeric):

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

Rockbuster/postgres@PostgreSQL 15

Query Query History

```

1 SELECT MODE () WITHIN GROUP (ORDER BY film_id) AS film_id_mode,
2        MODE () WITHIN GROUP (ORDER BY title) AS movie_title_mode,
3        MODE () WITHIN GROUP (ORDER BY description) AS movie_description_mode,
4        MODE () WITHIN GROUP (ORDER BY release_year) AS release_year_mode,
5        MODE () WITHIN GROUP (ORDER BY language_id) AS language_id_mode,
6        MODE () WITHIN GROUP (ORDER BY rating) AS movie_rating_mode,
7        MODE () WITHIN GROUP (ORDER BY last_update) AS last_update_mode,
8        MODE () WITHIN GROUP (ORDER BY special_features) AS special_features_mode,
9        MODE () WITHIN GROUP (ORDER BY fulltext) AS fulltext_mode_mode,
10       MODE () WITHIN GROUP (ORDER BY film_imdb) AS film_imdb_mode
11 FROM film;

```

Data Output Messages Notifications

	film_id_mode integer	movie_title_mode character varying	movie_description_mode text	release_year_mode integer	language_id_mode smallint
1	1	Academy Dinosaur	A Action-Packed Character ...	2006	1

movie_rating_mode mpaa_rating	last_update_mode timestamp without time zone	special_features_mode text[]	fulltext_mode_mode tsvector	film_imdb_mode integer
PG-13	2013-05-26 14:50:58.951	(Trailers,Commentaries,'Behind the Scenes')	'baloon':19 'confront':14 'documentari':5 'feminist':8,11,16 'mile':2 'must':13 'spi':1 'thri...	[null]

None.

- Customer table descriptive statistics (non-numeric):

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

Rockbuster/postgres@PostgreSQL 15

No limit

Query
Query History

```

1 SELECT MODE() WITHIN GROUP (ORDER BY customer_id) AS customer_id_mode,
2        MODE() WITHIN GROUP (ORDER BY store_id) AS store_id_mode,
3        MODE() WITHIN GROUP (ORDER BY first_name) AS first_name_mode,
4        MODE() WITHIN GROUP (ORDER BY last_name) AS last_name_mode,
5        MODE() WITHIN GROUP (ORDER BY email) AS email_mode,
6        MODE() WITHIN GROUP (ORDER BY address_id) AS address_id_mode,
7        MODE() WITHIN GROUP (ORDER BY activebool) AS activebool_mode,
8        MODE() WITHIN GROUP (ORDER BY create_date) AS create_date_mode,
9        MODE() WITHIN GROUP (ORDER BY last_update) AS last_update_mode,
10       MODE() WITHIN GROUP (ORDER BY active) AS active_mode
11 FROM customer
12

```

Data Output
Messages
Notifications

	customer_id_mode integer	store_id_mode smallint	first_name_mode character varying	last_name_mode character varying	email_mode character varying	address_id_mode smallint	activebool_mode boolean	create_date_mode date	last_update_mode timestamp without time zone	active_mode integer
1	1	1	Jamie	Abney	aaron.selby@sakilacustomer.org	5	true	2006-02-14	2013-05-26 14:49:45.738	