

Selection of Data Wrangling Techniques

Checking for Duplicates

Non-Uniform and Incorrect Values

Finding Missing Values

Checking for Duplicates

--Check for duplicate films

SELECT title,

release_year,

language_id,

rental_duration,

COUNT(*)

FROM film


GROUP BY title,

release_year,

language_id,

rental_duration

HAVING COUNT(*) >1;

Data Output	Explain	Messages	Notifications
 employee_id [PK] integer		name character varying (50)	 contact_number character varying (30)
			 designation_id smallint
			 last_update timestamp without time zone

Non-Uniform and Incorrect Values

--Looking for non-uniform and incorrect values in the length column of the film table




SELECT DISTINCT length AS film_length,

COUNT(*) AS count_rows

FROM film

GROUP BY length

ORDER BY length ASC

Data Output	Explain	Messages	Notifications
 film_length smallint		count_rows bigint	
1		46	5
2		47	7
3		48	11
4		49	5
5		50	9
6		51	7

Finding Missing Values

--Find the null values in the Film Table

SELECT *

FROM film

WHERE title IS NULL

--Returns only records from the film table where the email does not have the @ character



SELECT customer_id,

email AS customer_email

FROM customer

WHERE email

NOT LIKE '%@%'

Data Output	Explain	Messages	Notifications
 customer_id [PK] integer		customer_email character varying (50)	