**Data Cleaning Techniques:**

**1.Replace null values with default value or delete them**.

**(ISNULL OR COALESCE)**

\*SELECT \* FROM table\_name WHERE column\_name IS NULL;

\*UPDATE table\_name

SET column\_name= COALESCE(COLUMN\_NAME,’Default value’)

WHERE column\_name IS NULL

**2.Identify and delete duplicate rows based on key columns.**

**\*** SELECT column\_name ,

COUNT(\*) FROM table\_name

GROUP BY column\_name

HAVING COUNT(\*)>1;

**\***DELETE FROM table\_name

WHERE id NOT IN ( SELECT MIN(id) FROM table\_name GROUP BY column1, column2 );

\*WITH CTE AS(

SELECT

id,

Column1,

Column2,

ROW\_NUMBER() OVER(PARTITION BY column1,column2 ORDER BY id) AS row\_num

FROM table\_name

)

DELETE FROM table\_name

WHERE id IN (

SELECT id

FROM CTE

WHERE row\_num>1

);

**3. Convert text to lower or upper case to ensure consistency**

\*SELECT LOWER(column\_name) AS cleaned\_column FROM table\_name

**4.Remove extra spaces from text fields.**

**TRIM([LEADING | TRAILING | BOTH] [characters] FROM string)**

**TRIM(TRAILING [characters] FROM string)**

**SELECT TRIM(TRAILING '!' FROM 'Hello World!!!') AS result;**

\*SELECT TRIM(column\_name) FROM table\_name

\*UPDATE products

SET product\_code = TRIM(TRAILING '-' FROM product\_code) WHERE product\_code LIKE '%-';

\*UPDATE layoffs\_staging2

SET country = TRIM(TRAILING '.' FROM country);

**5.Convert date strings into a consistent date format**

**Standardize data formats:update data formats :Ensure consistency in data formats ,such as date formats or phone numbers.**

\*SELECT STR\_TO\_DATE(column\_name,’%m/%d/%Y) FROM table\_name

\*UPDATE table\_name

SET date\_column=TO\_DATE(date\_column,’YYYY-MM-DD’)

WHERE date\_column IS NOT NULL

**6. Identify and manage outliers in numerical data**

\*SELECT \* FROM table\_name WHERE column\_name BETWEEN lower\_limit AND upper\_limit;

\* SELECT \* FROM table\_name WHERE

column\_name>(SELECT AVG(column\_name)+3\*STDDEV(column\_name) FROM table\_name)

OR column\_name<(SELECT AVG(column\_name)-3\*STDDEV(column\_name) FROM table\_name);

**7.Repalce or remove special characters in text fields**

SELECT REGEXP\_REPLACE(column\_name,’[^a-zA-Z0-9]’,’’) AS cleaned\_column

FROM table\_name

**8.Standardize values in categorical columns**

UPDATE table\_name

SET column\_name=’Male’

WHERE column\_name in (‘M’,’male’)

**9 use statistical methods to fill in missing numerical data**

UPDATE table\_name

SET column\_name=(SELECT avg(column\_name) FROM table\_name)

Where column\_name IS NULL

10 Delete unnecessary data

DELETE FROM table\_name

WHERE condition\_to\_remove\_rows

11.Normalize Data: create new tables for normalization: Break down large tables into smaller related tables.

\*CREATE TABLE new\_table AS

SELECT DISTINCT column\_name

FROM table\_name