

Regular Expressions

| | |
|----|------------------------------------|
| \d | A digit |
| [] | Any item in brackets |
| ? | previous character 0 or 1 times |
| + | Previous character 1 or more times |
| * | Previous character 0 or more times |

Identifying Missing Data

Missing Completely at Random

```
SELECT * FROM restaurant_inspection WHERE score IS NULL;
```

```
SELECT COUNT(*) FROM restaurant_inspection WHERE score IS NULL;
```

Replacing null values with an average

```
UPDATE TABLE Patient SET Income = (SELECT avg(Income) FROM Patient) WHERE Income IS NULL;
```

Dropping Values where a column is null

```
SELECT count(*) FROM Patient WHERE Weight IS NULL;
```

Dealing with Duplicates

```
WITH DuplicateWithNumber AS ( SELECT Column1, Column2, row_number() OVER ( PARTITION BY Column1, Column2 ORDER BY Column3) AS NthAppearance FROM YourTableName ) DELETE FROM DuplicateWithNumber WHERE NthAppearance > 1
```

Cleaning with Case

```
UPDATE Project..housing SET SoldAsVacant = CASE WHEN SoldAsVacant = 'Y' THEN 'Yes' WHEN SoldAsVacant = 'N' THEN 'No' ELSE SoldAsVacant END
```

Replacing missing values with COALESCE()

```
The COALESCE() function substitutes null values
```

```
SELECT name, COALESCE(score, -1), inspection_type FROM restaurant_inspection
```

SQL Set Operations

| EXCEPT/MINUS | INTERSECT | UNION |
|---|---|---|
| Return the rows that are found in one relation but not the other. | the rows that are found in both relations | e rows that are found in either relation. |

Note that input relations must have the same number of columns and compatible data types for the respective columns.

Detecting out of range values

```
SELECT * FROM series WHERE rating NOT BETWEEN 0 AND 10
```

```
SELECT * FROM series WHERE rating < 0 OR rating > 10
```

Replicate: Repeats string x no of times

```
REPLICATE (string, integer)
```

```
REPLICATE ('0', 9 - LEN(registration_code))
```

Checking if a value is a website

```
select case when email LIKE '%.com' then email else null end AS email from pension_funds
```

Handy Numeric Functions

| | | |
|------------|--|--------------------------------|
| IS NUMERIC | Returns true if the expression has a numeric value, false if not | |
| CONVERT | Converts a character string into a number | SELECT CONVERT(int, 25.65); |
| TRUNCATE | n truncated to d decimal places. If you skip d, then n is truncated to 0 decimal places. If d is a negative number, the function truncates the number n to d digits left to the decimal point. | SELECT TRUNCATE(-123.4567, 2); |

Column values must translate to type to stand

| | |
|---------------|-----------------------------|
| ALTER TABLE | tableName |
| ALTER COLUMN | columnName |
| TYPE smallint | USING columnName::-smallint |

String Operations

```
-- SUBSTR
syntax is:
SUBSTR($string, starting
character position, # of
characters):
SELECT incidn t_num,
       date,
       SUB STR (date, 4,
2) AS day
FROM tutori al.s f_ cri me_ -
inc ide nts _20 14_01
# Extracting date and time
specifically from a column
SELECT incidn t_num,
       date,
       LEF T(date, 10) AS
cleane d_date,
       RIG HT( date, 17)
AS cleane d_time
FROM tutori al.s f_ cri me_ -
inc ide nts _20 14_01
# TRIM, to remove characters
from the beginning and end of a
string.
SELECT location,
       TRI M(both '() '
FROM location)
FROM tutori al.s f_ cri me_ -
inc ide nts _20 14_01
# POSITION allows you to specify
a substring,
then returns a numerical value
equal to the
character number (counting from
left) where that
substring first appears in the
target string.
SELECT incidn t_num,
```

String Operations (cont)

```
des cript,
POS ITI ON('A' IN
descript) AS a_position
FROM tutori al.s f_ cri me_ -
inc ide nts _20 14_01
# COALESCE can replace the null
values
SELECT incidn t_num,
       des cript,
       COA LES CE( des -
cript, 'No Descri ption')
FROM tutori al.s f_ cri me_ -
inc ide nts _cl eandate
ORDER BY descript DESC
```

Left for time

Right for date

Trim to remove chars at beginning and end

Position to return the location of a character!

COALESCE to replace the nulls!



By **datamansam**

cheatography.com/datamansam/

Not published yet.

Last updated 30th September, 2022.

Page 3 of 2.

Sponsored by **Readable.com**

Measure your website readability!

<https://readable.com>