

### Remove duplicates

#### Create CTE with added row\_num column

```
WITH dups AS
(SELECT *,
ROW_NUMBER() OVER (
PARTITION BY [column names] AS row_n
um
FROM [original_table] )
SELECT *
FROM dups
WHERE row_num > 1;
```

#### Deleting duplicates

```
DELETE FROM
t USING t
INNER JOIN dups ON t.id = dups.id
WHERE
dups.row_num > 1;
```

### Standardise data

#### Inspect column

```
SELECT DISTINCT c
FROM t
ORDER BY 1;
```

#### Delete spaces

```
UPDATE t
SET c = TRIM(c)
```

#### Delete other character (.)

```
UPDATE t
SET c = TRIM(TRAILING '.' FROM c
)
```

WHERE c LIKE '[problem]

Use either TRAILING or LEADING  
depending on where the character is

#### Change string to date format

```
STR_TO_DATE(c, [format string])
%m = month, %d = day, %Y = year
```

#### Change the data type

```
ALTER TABLE t
MODIFY COLUMN c DATE;
```

