

May contain typos

| species | | trees | | | | | | |
|---------|---------|-------|---|----|---------|----------|----------|-----|
| code | species | tree | x | y | species | diameter | priority | |
| m | maple | 0 | A | 10 | 4 | m | 8 | 71 |
| p | pine | 1 | B | 20 | 4 | m | 10 | 100 |
| | | 2 | C | 30 | 4 | p | 6 | 30 |
| | | 3 | D | 40 | 4 | p | 8 | 40 |
| | | 4 | E | 50 | 4 | m | 12 | 99 |

import sqlite3
c = sqlite3.connect("worksheet.db")
def qry(sql):
 return pd.read_sql(sql, c)

```
species = qry("SELECT * FROM species")
trees = qry("SELECT * FROM trees")
```

1 trees[trees["priority"] > 90][["x", "y"]] # convert to SQL

| | X | Y |
|---|----|---|
| 0 | 20 | 4 |
| 1 | 50 | 4 |
| 2 | 4 | 4 |

What is this

in SQL?

SELECT X, Y

FROM trees

WHERE priority > 90

2 qry("SELECT x+y FROM trees WHERE species = 'm'") # convert to Pandas

X+Y

| | 0 | 1 | 2 | 3 |
|-----|----|----|----|---|
| X+Y | 14 | 24 | 54 | |

X-col = trees[trees['Species'] == 'm'][['X']]

Y-col = trees[trees['Species'] == 'm'][['Y']]

X-col + Y-col // Series

3 cd = species["code"][species["species"]=="maple"].iloc[0]

trees[trees["species"] == cd][["tree"]]

convert to 2 SQL queries

(d='m')

| | 0 | 1 | 2 | 3 | 4 | 5 |
|------|---|---|---|---|---|---|
| tree | A | B | | | E | |

SELECT code

FROM species

WHERE Species = 'maple'

SELECT tree

FROM trees

WHERE Species = 'm'

4 qry("SELECT species FROM trees ORDER BY priority DESC")

Species

| | 0 | 1 | 2 | 3 | 4 |
|---------|---|---|---|---|---|
| Species | M | M | M | P | |

trees.sort_values("priority", ascending=False)

species

5

| code | species |
|------|---------|
| m | maple |
| p | pine |

trees

| tree | x | y | species | diameter | priority |
|------|----|---|---------|----------|----------|
| A | 10 | 4 | m | 71 | 71 |
| B | 20 | 4 | m | 10 | 100 |
| C | 30 | 4 | p | 6 | 30 |
| D | 40 | 4 | p | 8 | 40 |
| E | 50 | 4 | m | 12 | 99 |

list(qry("SELECT tree, priority FROM trees " +
 "ORDER BY priority DESC LIMIT 1").iloc[0])

0 tree priority → tree B priority 100 → [B, 100]

list(trees.sort_values("priority", ascending=False).iloc[0][["tree", "priority"]])

6 qry("""SELECT COUNT(SPECIES) AS c1,
 COUNT(DISTINCT SPECIES) as c2
 FROM trees""")

0 c1 c2
 5 2

7 qry("""SELECT species, COUNT(SPECIES) AS count,
 AVG(diameter) AS size
 FROM trees
 GROUP BY species ORDER BY count DESC""")

Species Count size
 0 m 3 10
 1 p 2 7

hydrants

| year | color | style | owner | alt | active |
|------|-------|-------|---------|------|--------|
| 1999 | red | K-81 | private | 1179 | 0 |
| 2000 | red | M-3 | public | 1065 | 0 |
| 2001 | green | Pacer | private | 1058 | 1 |
| 2010 | blue | Pacer | public | 1081 | 1 |
| 2014 | blue | Pacer | public | 1052 | 1 |
| 2018 | blue | Pacer | public | 1109 | 1 |

hydrants = qry("""
SELECT * FROM hydrants
""")

8 qry("SELECT color, year FROM hydrants WHERE color = 'blue'")

| | color | year | hydrants[hydrants['color']] == 'blue'] |
|---|-------|------|--|
| 0 | blue | 2010 | [['color', 'year']] |
| 1 | blue | 2014 | |
| 2 | blue | 2018 | |

9 df = qry("SELECT color, year FROM hydrants") df.color

df[df.color == "blue"]

| | color | year | df[['color']] |
|---|-------|------|---------------|
| 0 | red | 1999 | 3 blue 2010 |
| 1 | red | 2000 | 4 blue 2014 |
| 2 | green | 2001 | 5 blue 2018 |
| 3 | blue | 2010 | |
| 4 | blue | 2014 | |
| 5 | blue | 2018 | |

active = 1

10 qry("SELECT year FROM hydrants WHERE owner='private' AND active")

| | year | hydrants[(hydrants['owner'] == 'private') & (hydrants['active'] == 1)][['year']] |
|---|------|--|
| 0 | 2001 | |

11 df = qry("SELECT year, style, active FROM hydrants")
df[df.active == 1][["style"]]

| | year | style | active | |
|---|------|-------|--------|-------------------|
| 0 | 1999 | K-81 | 0 | Series |
| 1 | 2000 | M-3 | 0 | 2 Pacer |
| 2 | 2001 | Pacer | 1 | → 3 Pacer |
| 3 | 2010 | Pacer | 1 | 4 Pacer |
| 4 | 2014 | Pacer | 1 | 5 Pacer |
| 5 | 2018 | Pacer | 1 | |

Data Frame

Series

hydrants

| year | color | style | owner | alt | active |
|------|-------|-------|---------|------|--------|
| 1999 | red | K-81 | private | 1179 | 0 |
| 2000 | red | M-3 | public | 1065 | 0 |
| 2001 | green | Pacer | private | 1058 | 1 |
| 2010 | blue | Pacer | public | 1081 | 1 |
| 2014 | blue | Pacer | public | 1052 | 1 |
| 2018 | blue | Pacer | public | 1109 | 1 |

hydrants = qry("""
 SELECT * FROM hydrants
 """)

returns in descending
 order

12 hydrants["color"].value_counts() # convert to SQL

blue 3
 red 2
 green 1

SELECT color, COUNT(*) as c
 FROM hydrants
 GROUP BY color
 ORDER BY c DESC

13 qry("""SELECT color, COUNT(*) FROM hydrants
WHERE active GROUP BY color""")

| color | count(*) |
|-------|----------|
| green | 1 |
| blue | 3 |

14 qry("""SELECT color, COUNT(*) AS count FROM hydrants
 GROUP BY color HAVING count > 1""")

| color | count |
|-------|-------|
| red | 2 |
| blue | 3 |

15 qry("""SELECT color, COUNT(*) AS count
 FROM hydrants WHERE year >= 2000
 GROUP BY color HAVING count < 2""")

| color | count |
|-------|-------|
| red | 1 |
| green | 1 |
| | 1 |