Ordering, Groups, and Functions

Module 2: 02

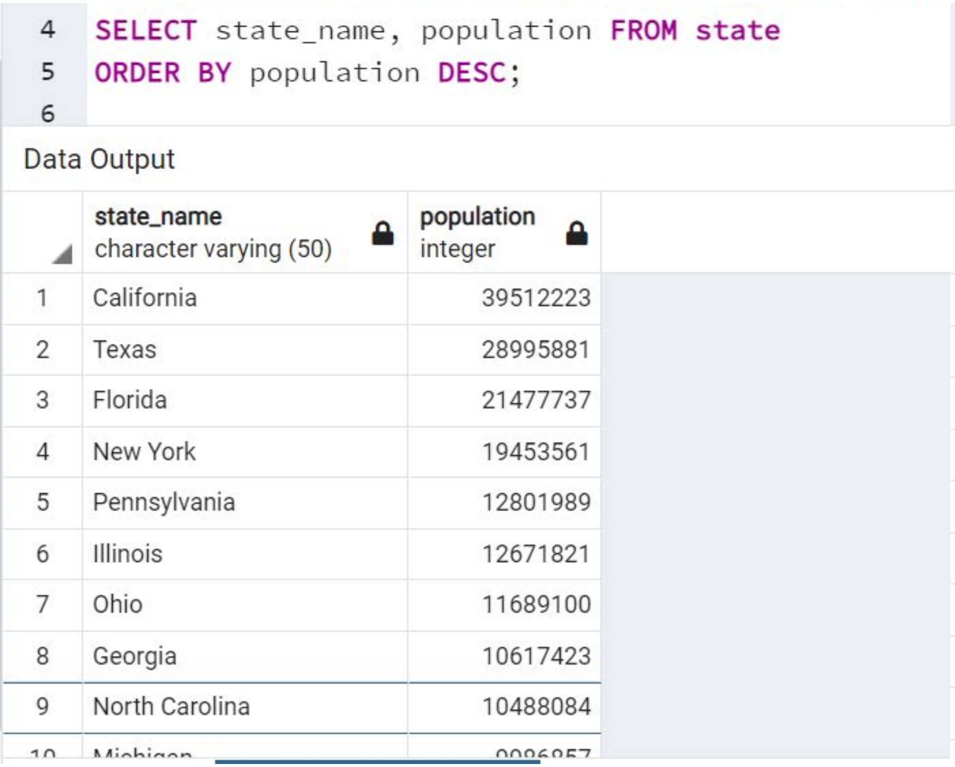
Objectives

- Ordering
- Limiting Results
- Numeric and String operations functions
- Aggregate functions
- Grouping Results
- Subqueries

- Know the structure of queries

```
SELECT      {column_name(s)}
FROM        {table_name(s)}
WHERE       {some condition}
GROUP BY    {aggregate SELECTED columns}
HAVING      {more conditional logic on GROUP}
ORDER BY    {sort order of result set}
```

ORDER BY can be added to a query to order the results by the data in a row. The Order by clause is added after WHERE. Results can be ordered in Ascending (**ASC**) or Descending (**DESC**) order. *The default order is ASC*.

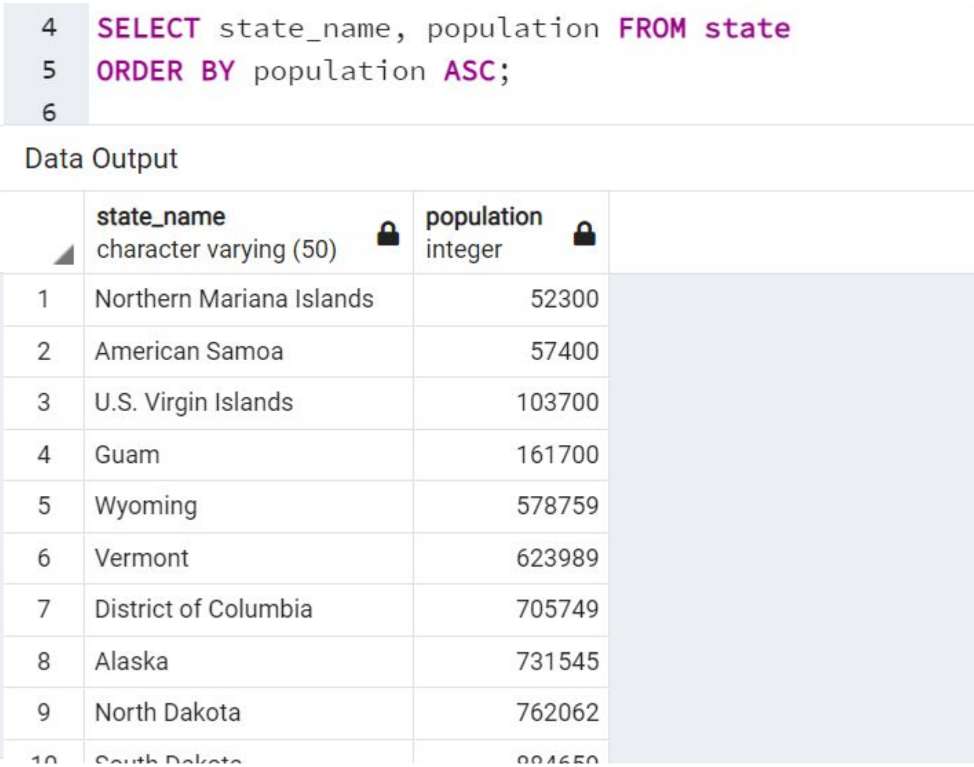


```sql
SELECT state_name, population FROM state
ORDER BY population DESC;

```

Data Output

	state_name character varying (50)	population integer
1	California	39512223
2	Texas	28995881
3	Florida	21477737
4	New York	19453561
5	Pennsylvania	12801989
6	Illinois	12671821
7	Ohio	11689100
8	Georgia	10617423
9	North Carolina	10488084
10	Michigan	9986857

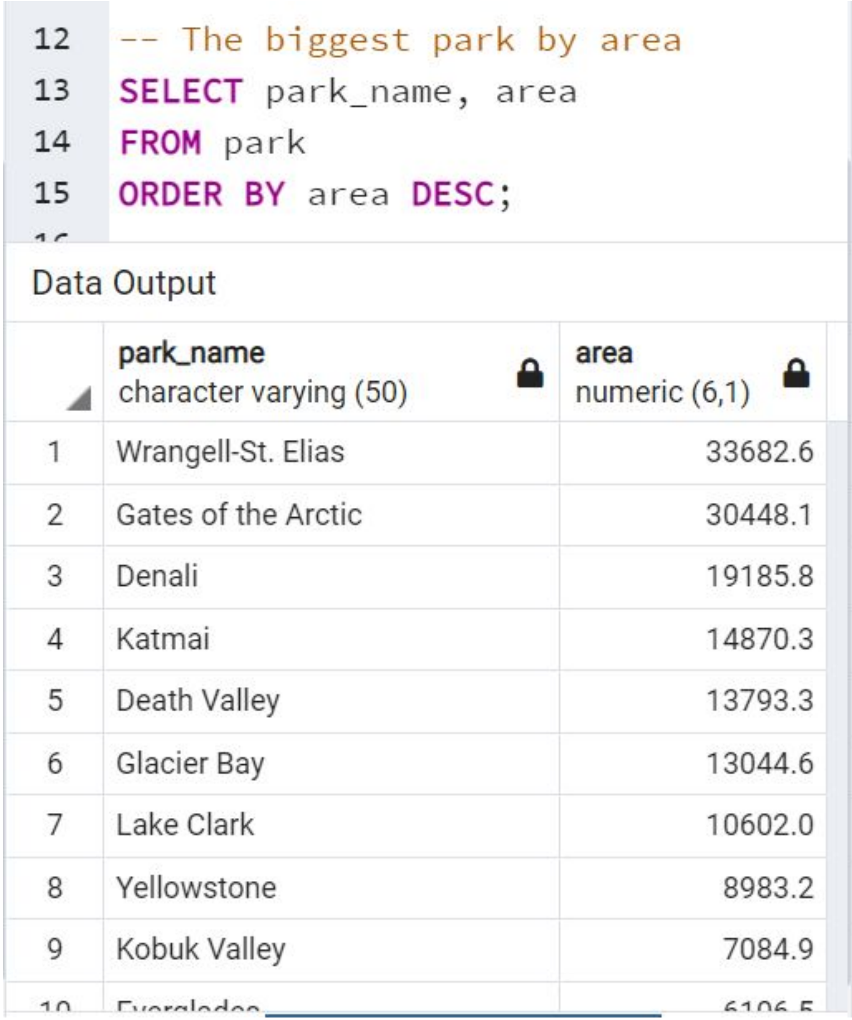


```sql
SELECT state_name, population FROM state
ORDER BY population ASC;

```

Data Output

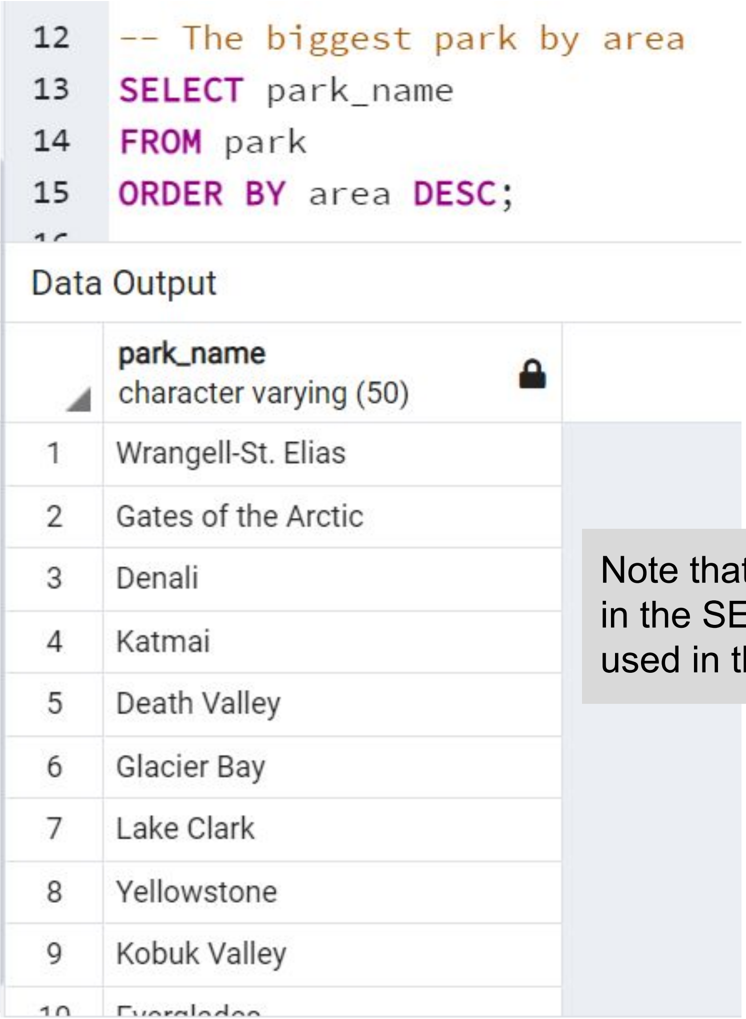
	state_name character varying (50)	population integer
1	Northern Mariana Islands	52300
2	American Samoa	57400
3	U.S. Virgin Islands	103700
4	Guam	161700
5	Wyoming	578759
6	Vermont	623989
7	District of Columbia	705749
8	Alaska	731545
9	North Dakota	762062
10	South Dakota	884659



```
-- The biggest park by area
SELECT park_name, area
FROM park
ORDER BY area DESC;
```

Data Output

	park_name character varying (50)	area numeric (6,1)
1	Wrangell-St. Elias	33682.6
2	Gates of the Arctic	30448.1
3	Denali	19185.8
4	Katmai	14870.3
5	Death Valley	13793.3
6	Glacier Bay	13044.6
7	Lake Clark	10602.0
8	Yellowstone	8983.2
9	Kobuk Valley	7084.9



```
-- The biggest park by area
SELECT park_name
FROM park
ORDER BY area DESC;
```

Data Output

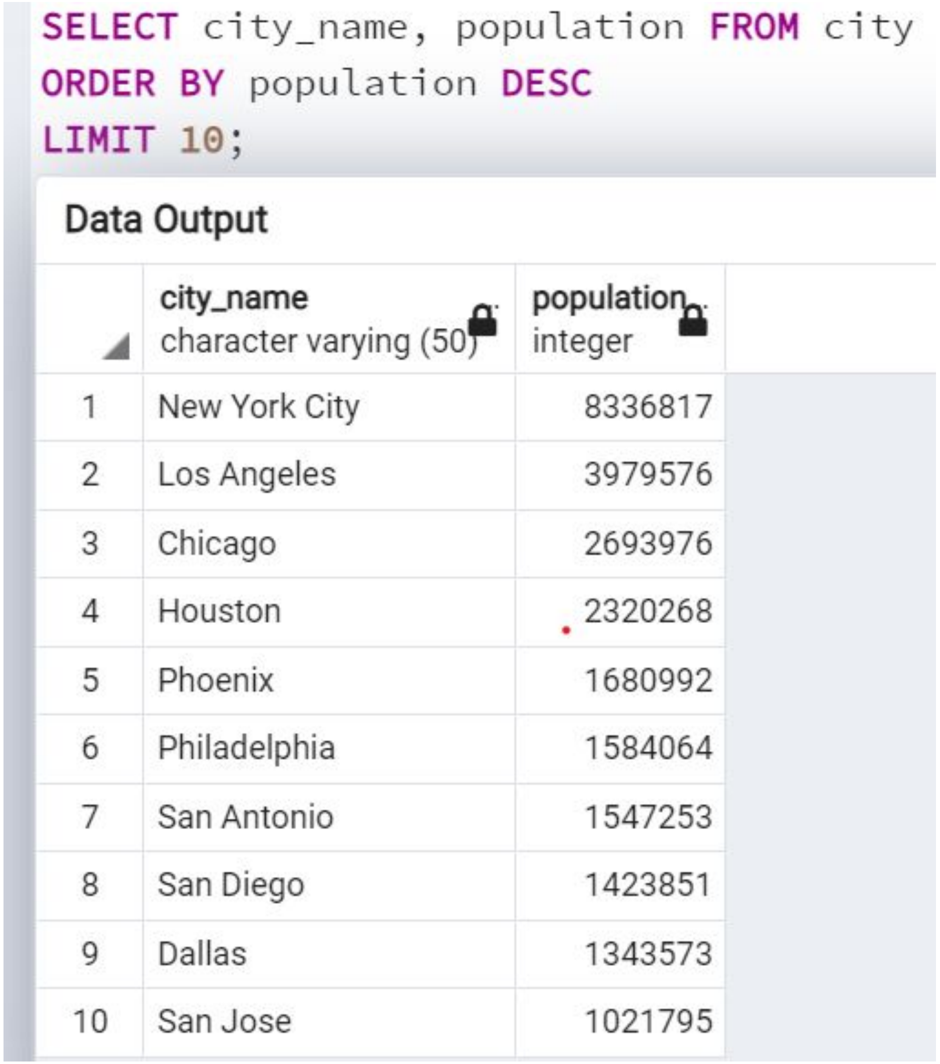
	park_name character varying (50)
1	Wrangell-St. Elias
2	Gates of the Arctic
3	Denali
4	Katmai
5	Death Valley
6	Glacier Bay
7	Lake Clark
8	Yellowstone
9	Kobuk Valley

Note that the area isn’t in the SELECT, but is used in the ORDER BY

Limiting Results

The **LIMIT #** clause can be used to limit the number of rows returned. The LIMIT clause is added at the end of the query.

Note: Limiting the number of rows returned has nothing to do with ordering (or sorting the data).



```
SELECT city_name, population FROM city
ORDER BY population DESC
LIMIT 10;
```

Data Output

	city_name character varying (50)	population integer
1	New York City	8336817
2	Los Angeles	3979576
3	Chicago	2693976
4	Houston	2320268
5	Phoenix	1680992
6	Philadelphia	1584064
7	San Antonio	1547253
8	San Diego	1423851
9	Dallas	1343573
10	San Jose	1021795

Numeric Operations

round(value, scale) rounds a floating point number to a set scale.

```
select area/3 from park;
```

result : 347.4666666666666

```
select round(area/3, 4) from park;
```

result : 347.4667

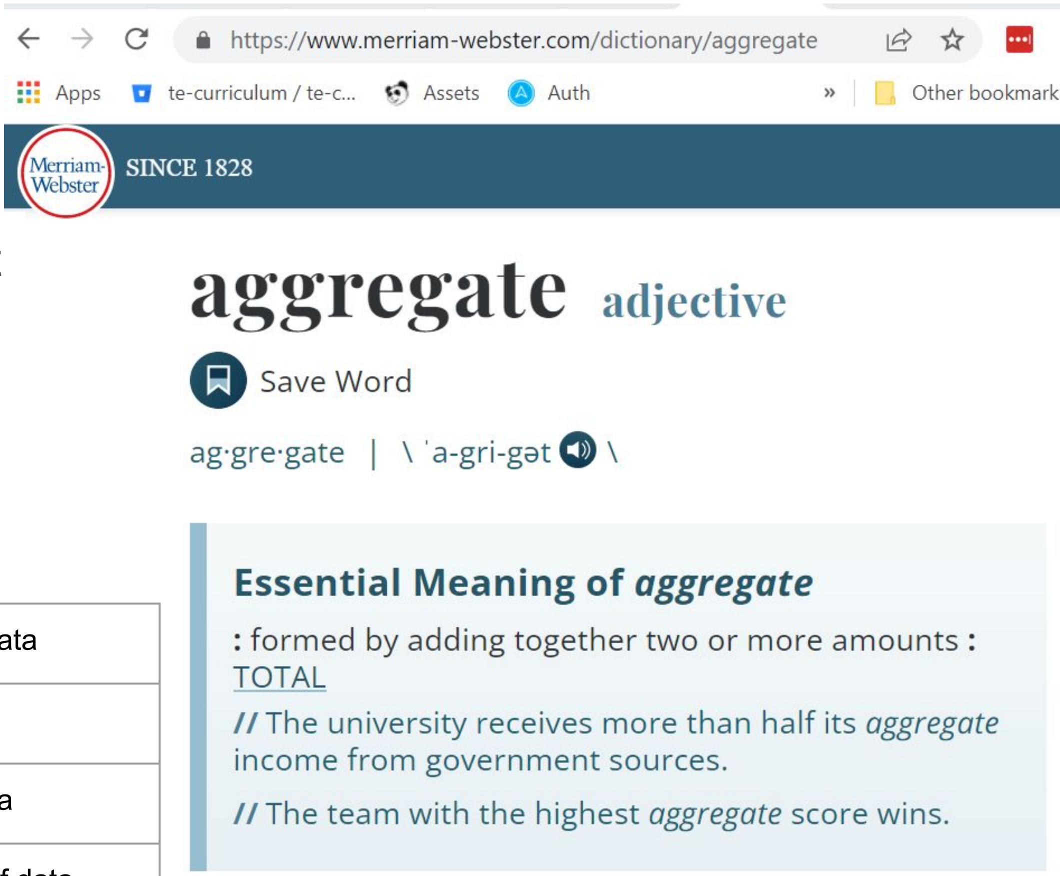
```
select round(area/3, 2) from park;
```

result : 347.47

Aggregate Functions

Aggregate functions collapse a dataset into **1 result**, like an Average, Sum, or Count. The WHERE clause is applied first, which allows for aggregates of subgroups on a table.

AVG()	returns the average value of a numeric set of data
SUM()	returns the total sum of a numeric set of data
COUNT()	returns the number of rows matching the criteria
MIN()	returns the smallest value from a numeric set of data
MAX()	returns the largest value from a numeric set of data



aggregate adjective

ag·gre·gate | \ ˈa-gri-gət \

Essential Meaning of *aggregate*

: formed by adding together two or more amounts : TOTAL

// The university receives more than half its *aggregate* income from government sources.

// The team with the highest *aggregate* score wins.

https://www.postgresqltutorial.com/postgresql-aggregate-functions/

GROUP BY

GROUP BY groups records into summary rows and returns one record for each group.

Used in conjunction with Aggregate Functions to tell SQL how to group non-aggregate values. All non-aggregate columns in the SELECT must be in the GROUP BY clause.

```
SELECT min(population), max(population), region, name FROM country
GROUP BY region, name
ORDER BY region, name
```

Groups are applied in the order listed. So first the data is grouped by region and then by name within each region, and then the min() and max() aggregate function is applied to each group.

GROUP BY

- Rules of GROUP BY
  - SELECT line and GROUP BY include same columns
  - Appears after tables have been specified and JOINs completed
  - If filtering with WHERE clause, GROUP BY follows WHERE
  - HAVING is optional filter on the data after being ‘grouped’

Table: Patients

first_name	last_name	age
Jane	Smith	32
Joe	Smith	15
Dave	Jones	25
Sam	Davies	42
Bill	Smith	72
Jill	Jones	54
Fred	Hart	38

SELECT last_name, AVG(age) FROM patients GROUP BY last_name

first_name	last_name	age
Jane	Smith	32
Joe	Smith	15
Dave	Jones	25
Sam	Davies	42
Bill	Smith	72
Jill	Jones	54
Fred	Hart	38

first_name	last_name	age
Jane	Smith	32
Joe	Smith	15
Bill	Smith	72
Dave	Jones	25
Jill	Jones	54
Sam	Davies	42
Fred	Hart	38

First the rows are grouped by unique values in the column in the GROUP BY.

For this table and data it creates 4 groups by last_name: Smith, Jones, Davies, Hart

first_name	last_name	age		AVG(age)
Jane	Smith	32	>	39.6
Joe	Smith	15		
Bill	Smith	72		
Dave	Jones	25	>	39.5
Jill	Jones	54		
Sam	Davies	42	>	42
Fred	Hart	38	>	38

The Aggregate Function, in this case AVG(), is applied to the values in each GROUP.

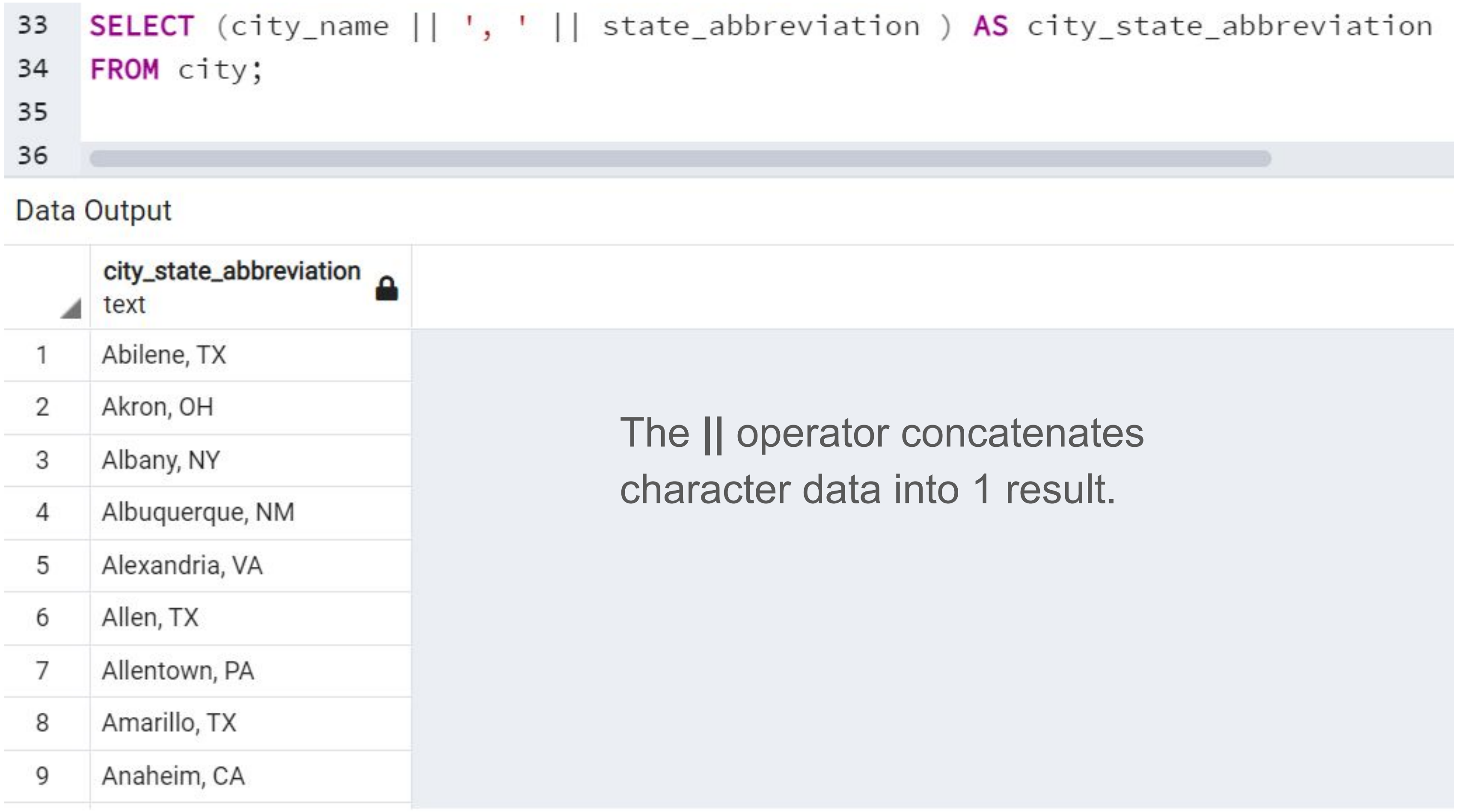
RETURNED RESULT

last_name	AVG(age)
Smith	39.6
Jones	39.5
Davies	42
Hart	38

The return is 1 row for each group with the aggregate (AVG) performed for the data in each group, in this case the age. Since the items are grouped by last_name, then there will be 1 row returned for each unique

last_name in the data set, with the average done for the set of ages associated with the last name.

String Operations



```
SELECT (city_name || ', ' || state_abbreviation ) AS city_state_abbreviation
FROM city;
```

Data Output

	city_state_abbreviation text
1	Abilene, TX
2	Akron, OH
3	Albany, NY
4	Albuquerque, NM
5	Alexandria, VA
6	Allen, TX
7	Allentown, PA
8	Amarillo, TX
9	Anaheim, CA

The || operator concatenates character data into 1 result.

Obnoxious tip to memorize the order

Select	**S**ome
From	**F**rench
Where	**W**aiters
Group by	**G**row
Having	**H**ealthy
Order by	**O**ranges &
Limit	**L**emons

Subqueries

A **SubQuery** is an inner query that can provide results as input to its parent query. A subquery can only return 1 column of data.

```
SELECT * FROM country WHERE continent = 'Europe' AND gnp > 1000000
```

Returns: `'GBR', 'ITA', 'FRA', 'DEU'`

Without SubQuery:
```
SELECT * FROM city WHERE countrycode IN ('GBR', 'ITA', 'FRA', 'DEU');
```

Subquery provides same list for use in the in.

With SubQuery:
```
SELECT * FROM city WHERE countrycode IN (SELECT code FROM COUNTRY WHERE continent =
'Europe' AND gnp > 1000000);
```