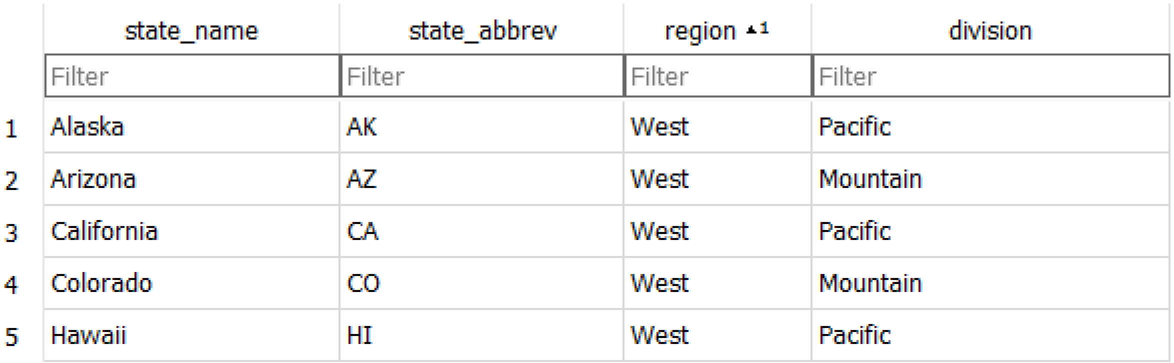
Karan Rawat

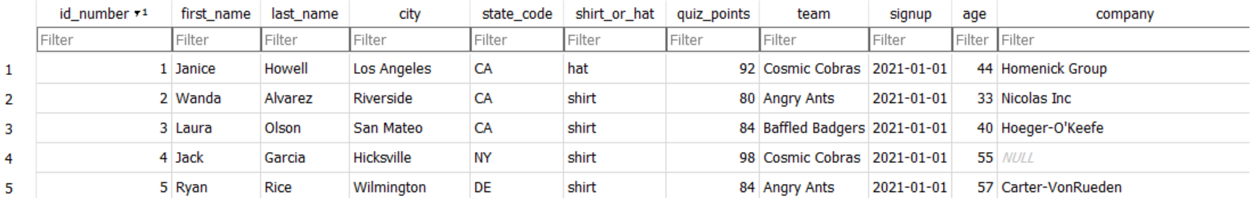
SQL PROJECT

In this project, I had two tables: **States** and **People.** I used SQL on SQLite to write queries and explore the datasets.
Table **States** contains the attributes of US states: state_name, state_abbrev, region, and division.



	state_name	state_abbrev	region ▴1	division
	Filter	Filter	Filter	Filter
1	Alaska	AK	West	Pacific
2	Arizona	AZ	West	Mountain
3	California	CA	West	Pacific
4	Colorado	CO	West	Mountain
5	Hawaii	HI	West	Pacific

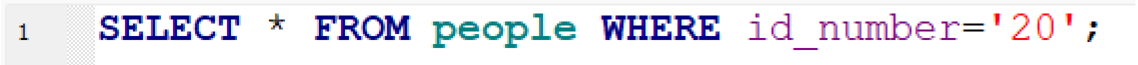
Table **people** contains the following attributes:



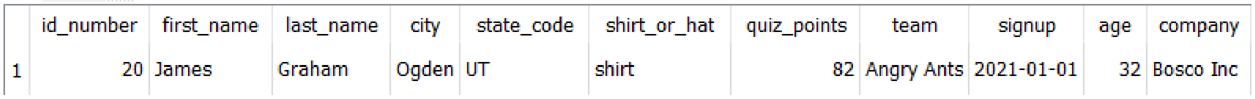
	id_number ▾1	first_name	last_name	city	state_code	shirt_or_hat	quiz_points	team	signup	age	company
	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter
1	1	Janice	Howell	Los Angeles	CA	hat	92	Cosmic Cobras	2021-01-01	44	Homenick Group
2	2	Wanda	Alvarez	Riverside	CA	shirt	80	Angry Ants	2021-01-01	33	Nicolas Inc
3	3	Laura	Olson	San Mateo	CA	shirt	84	Baffled Badgers	2021-01-01	40	Hoeger-O'Keefe
4	4	Jack	Garcia	Hicksville	NY	shirt	98	Cosmic Cobras	2021-01-01	55	NULL
5	5	Ryan	Rice	Wilmington	DE	shirt	84	Angry Ants	2021-01-01	57	Carter-VonRueden

I generated my own questions on the data in order to formulate SQL commands and see the results:

1) *Display all the attributes of a person from a random id_number*

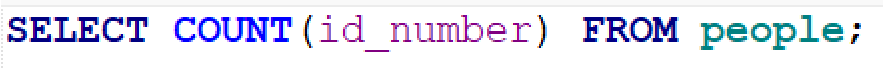


```
SELECT * FROM people WHERE id_number='20';
```

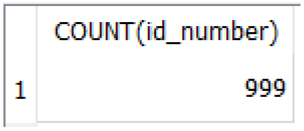


	id_number	first_name	last_name	city	state_code	shirt_or_hat	quiz_points	team	signup	age	company
1	20	James	Graham	Ogden	UT	shirt	82	Angry Ants	2021-01-01	32	Bosco Inc

2) *Count the total number of people signed up for the quiz*

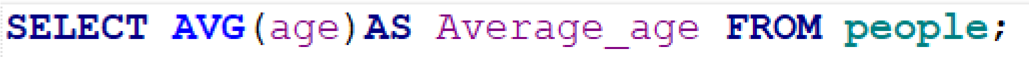


```
SELECT COUNT(id_number) FROM people;
```

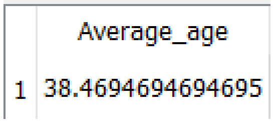


	COUNT(id_number)
1	999

3) *Find the average age of participants*

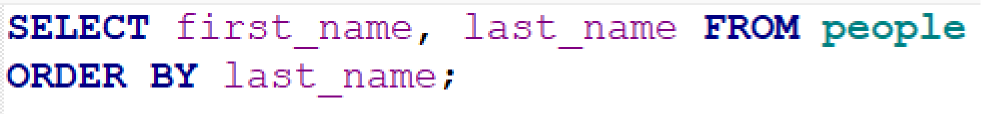


```
SELECT AVG(age)AS Average_age FROM people;
```

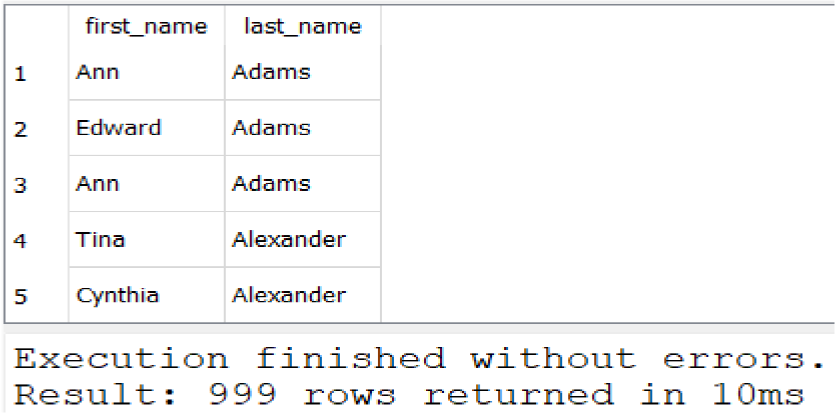


	Average_age
1	38.4694694694695

4) *Retrieve the first_name and last_name of people sorted alphabetically by their last names*



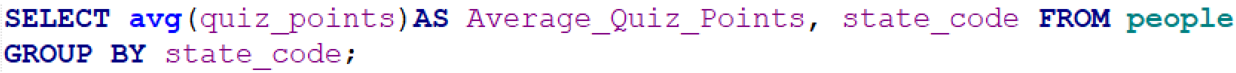
```
SELECT first_name, last_name FROM people
ORDER BY last_name;
```



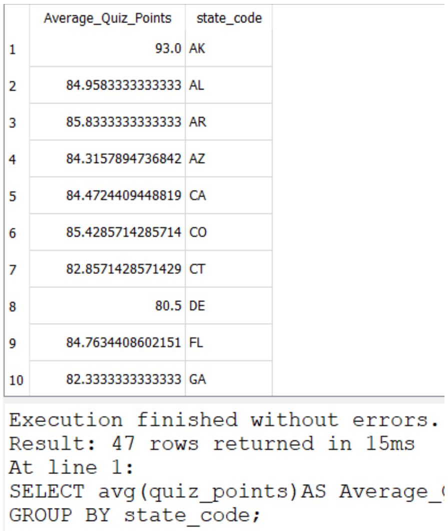
	first_name	last_name
1	Ann	Adams
2	Edward	Adams
3	Ann	Adams
4	Tina	Alexander
5	Cynthia	Alexander

```
Execution finished without errors.
Result: 999 rows returned in 10ms
```

5) *Find the average quiz points per state.*



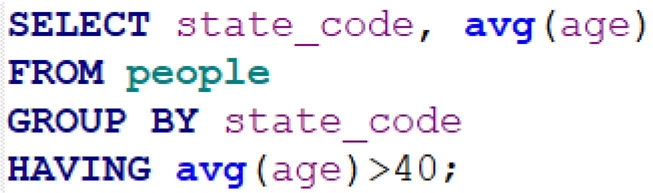
```
SELECT avg(quiz_points)AS Average_Quiz_Points, state_code FROM people
GROUP BY state_code;
```



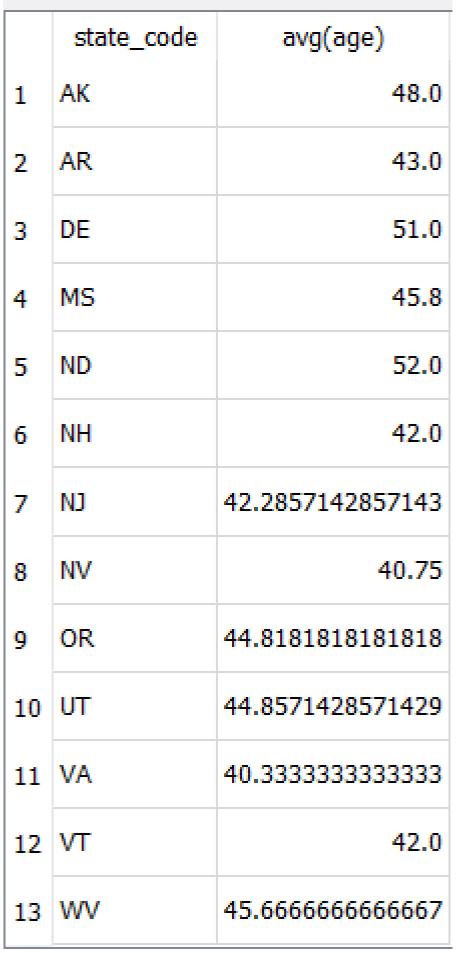
	Average_Quiz_Points	state_code
1	93.0	AK
2	84.9583333333333	AL
3	85.8333333333333	AR
4	84.3157894736842	AZ
5	84.4724409448819	CA
6	85.4285714285714	CO
7	82.8571428571429	CT
8	80.5	DE
9	84.7634408602151	FL
10	82.3333333333333	GA

```
Execution finished without errors.
Result: 47 rows returned in 15ms
At line 1:
SELECT avg(quiz_points)AS Average_
GROUP BY state_code;
```

6) *Display states where the average age of participants is above 40.*

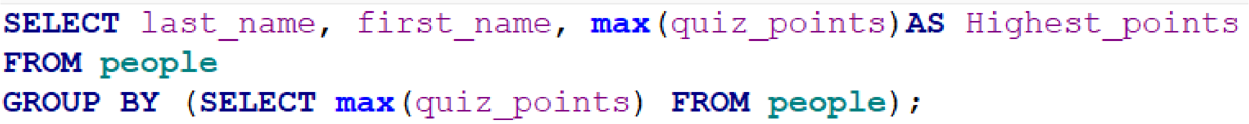


```
SELECT state_code, avg(age)
FROM people
GROUP BY state_code
HAVING avg(age)>40;
```

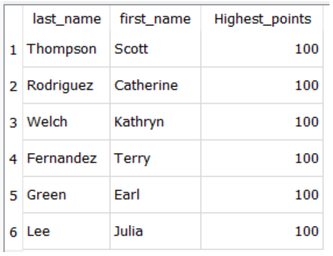


	state_code	avg(age)
1	AK	48.0
2	AR	43.0
3	DE	51.0
4	MS	45.8
5	ND	52.0
6	NH	42.0
7	NJ	42.2857142857143
8	NV	40.75
9	OR	44.8181818181818
10	UT	44.8571428571429
11	VA	40.3333333333333
12	VT	42.0
13	WV	45.6666666666667

7) *Find the person with the highest quiz points.*

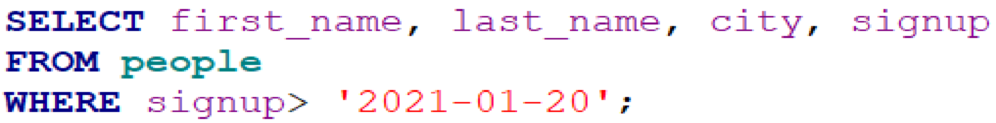


```
SELECT last_name, first_name, max(quiz_points)AS Highest_points
FROM people
GROUP BY (SELECT max(quiz_points) FROM people);
```

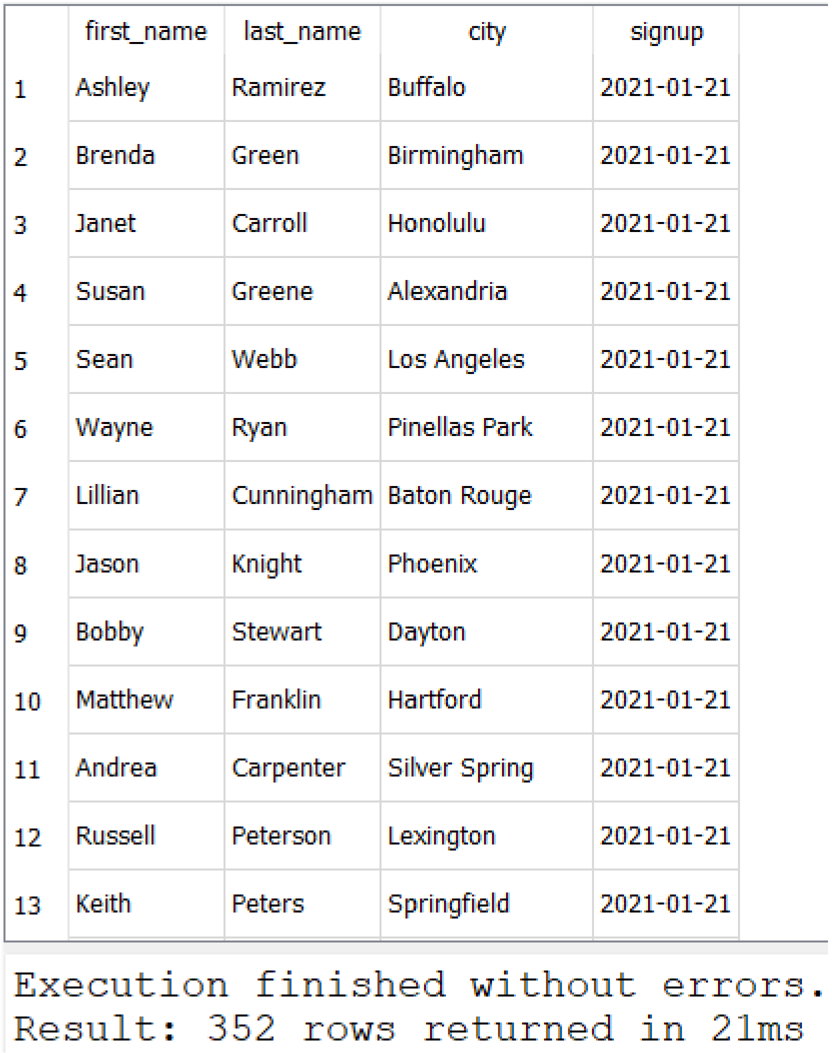


	last_name	first_name	Highest_points
1	Thompson	Scott	100
2	Rodriguez	Catherine	100
3	Welch	Kathryn	100
4	Fernandez	Terry	100
5	Green	Earl	100
6	Lee	Julia	100

8) *Retrieve people who signed up after a certain date.*



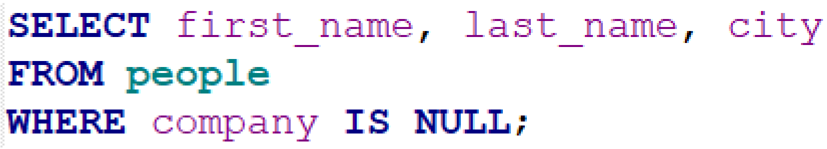
```
SELECT first_name, last_name, city, signup
FROM people
WHERE signup> '2021-01-20';
```



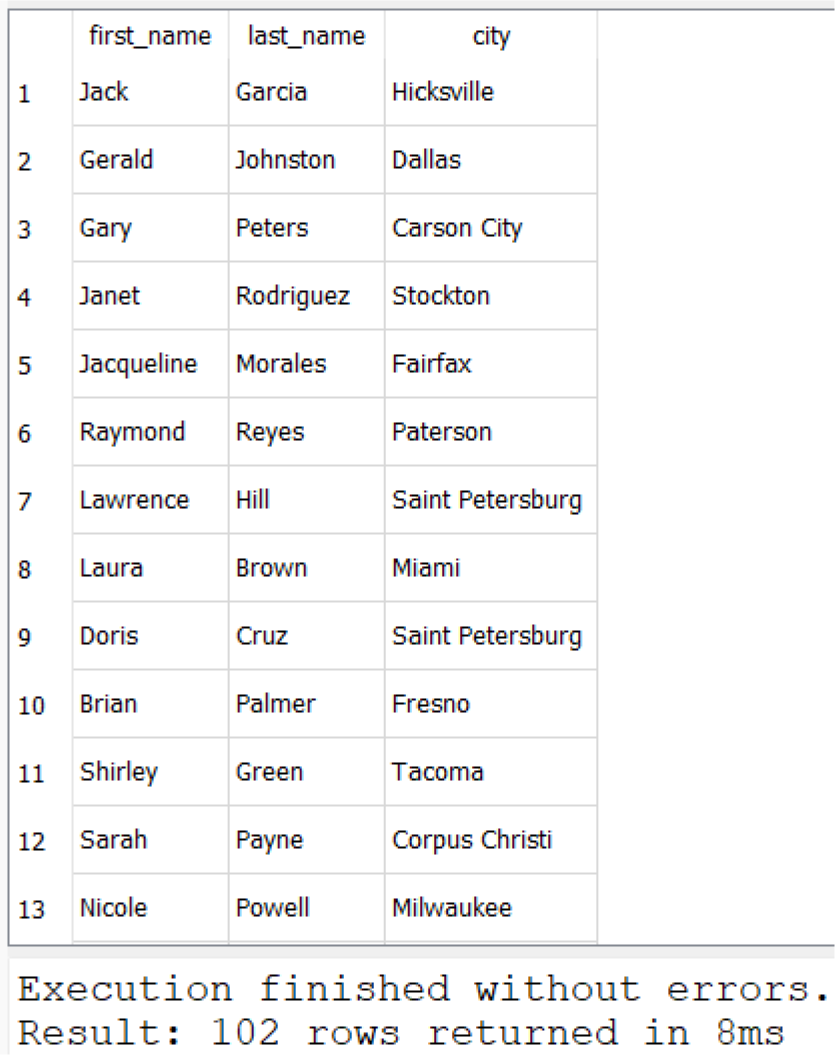
	first_name	last_name	city	signup
1	Ashley	Ramirez	Buffalo	2021-01-21
2	Brenda	Green	Birmingham	2021-01-21
3	Janet	Carroll	Honolulu	2021-01-21
4	Susan	Greene	Alexandria	2021-01-21
5	Sean	Webb	Los Angeles	2021-01-21
6	Wayne	Ryan	Pinellas Park	2021-01-21
7	Lillian	Cunningham	Baton Rouge	2021-01-21
8	Jason	Knight	Phoenix	2021-01-21
9	Bobby	Stewart	Dayton	2021-01-21
10	Matthew	Franklin	Hartford	2021-01-21
11	Andrea	Carpenter	Silver Spring	2021-01-21
12	Russell	Peterson	Lexington	2021-01-21
13	Keith	Peters	Springfield	2021-01-21

```
Execution finished without errors.
Result: 352 rows returned in 21ms
```

9) *Identify people whose company information is missing.*



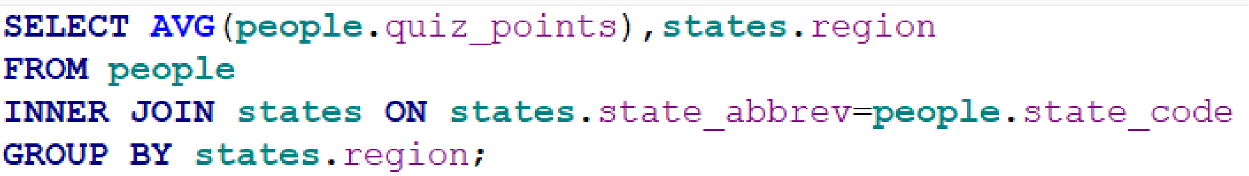
```
SELECT first_name, last_name, city
FROM people
WHERE company IS NULL;
```



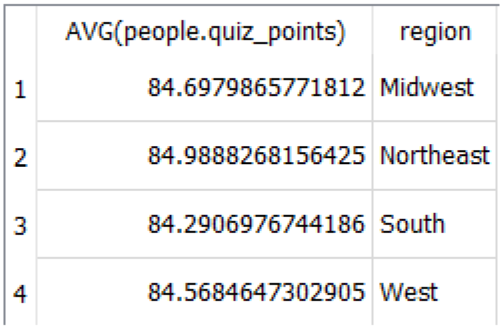
	first_name	last_name	city
1	Jack	Garcia	Hicksville
2	Gerald	Johnston	Dallas
3	Gary	Peters	Carson City
4	Janet	Rodriguez	Stockton
5	Jacqueline	Morales	Fairfax
6	Raymond	Reyes	Paterson
7	Lawrence	Hill	Saint Petersburg
8	Laura	Brown	Miami
9	Doris	Cruz	Saint Petersburg
10	Brian	Palmer	Fresno
11	Shirley	Green	Tacoma
12	Sarah	Payne	Corpus Christi
13	Nicole	Powell	Milwaukee

```
Execution finished without errors.
Result: 102 rows returned in 8ms
```

10) Calculate the average quiz points by region.



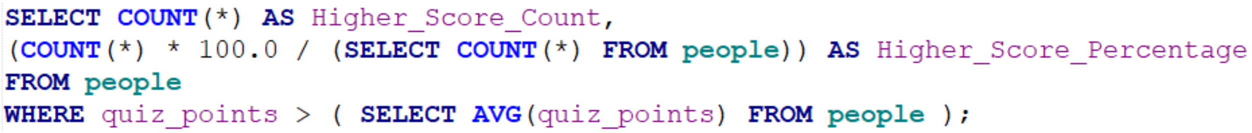
```sql
SELECT AVG(people.quiz_points),states.region
FROM people
INNER JOIN states ON states.state_abbrev=people.state_code
GROUP BY states.region;
```



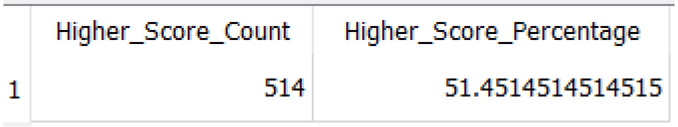
	AVG(people.quiz_points)	region
1	84.6979865771812	Midwest
2	84.9888268156425	Northeast
3	84.2906976744186	South
4	84.5684647302905	West

Karan Rawat

11) Calculate the count and percentage of people who scored higher than average.

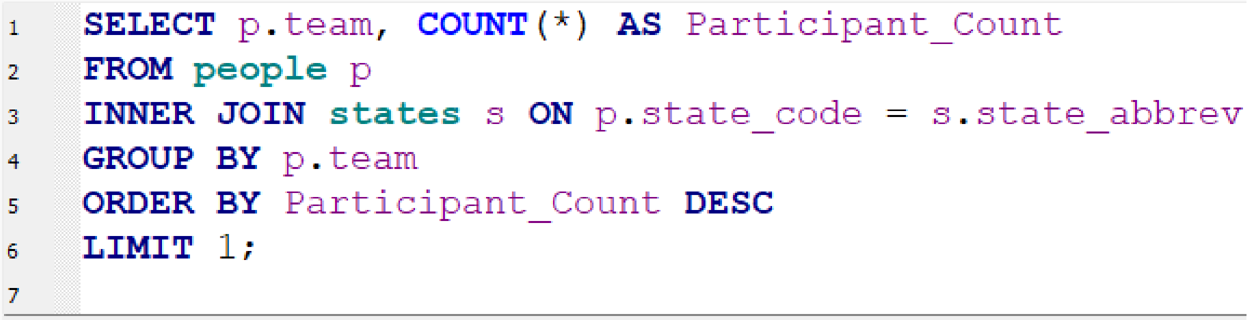


```sql
SELECT COUNT(*) AS Higher_Score_Count,
(COUNT(*) * 100.0 / (SELECT COUNT(*) FROM people)) AS Higher_Score_Percentage
FROM people
WHERE quiz_points > ( SELECT AVG(quiz_points) FROM people );
```

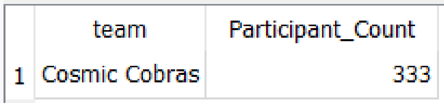


	Higher_Score_Count	Higher_Score_Percentage
1	514	51.4514514514515

12) Which team has the highest number of participants across all states?

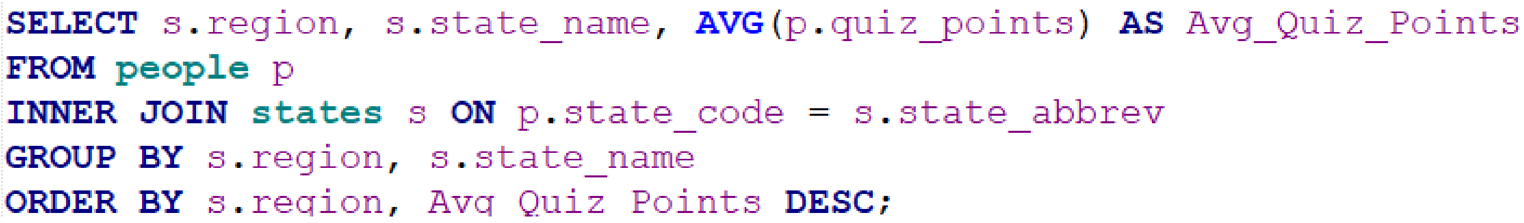


```sql
SELECT p.team, COUNT(*) AS Participant_Count
FROM people p
INNER JOIN states s ON p.state_code = s.state_abbrev
GROUP BY p.team
ORDER BY Participant_Count DESC
LIMIT 1;
```

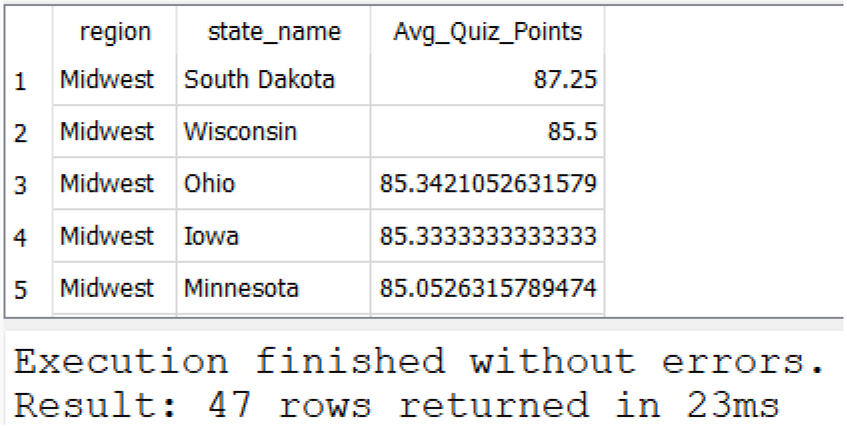


	team	Participant_Count
1	Cosmic Cobras	333

13) How do the average quiz points differ between states within the same region?



```sql
SELECT s.region, s.state_name, AVG(p.quiz_points) AS Avg_Quiz_Points
FROM people p
INNER JOIN states s ON p.state_code = s.state_abbrev
GROUP BY s.region, s.state_name
ORDER BY s.region, Avg_Quiz_Points DESC;
```



	region	state_name	Avg_Quiz_Points
1	Midwest	South Dakota	87.25
2	Midwest	Wisconsin	85.5
3	Midwest	Ohio	85.3421052631579
4	Midwest	Iowa	85.3333333333333
5	Midwest	Minnesota	85.0526315789474

```
Execution finished without errors.
Result: 47 rows returned in 23ms
```