

# Modern Big Data Analysis with SQL

Coursera Specialisation (Offered by Cloudera)

Course-2: Analysing Big Data with SQL

Week-4: Core

Question-1: Below is a table with three rows. What is the value of AVG(items) for this table?

order_id	items	total
829	3	38.92
220	7	107.06
1043	2	19.98

Answer-1: 4

Question-2: Which of the following statements are valid? (The column color is a string column, and both red and blue are integer columns.) Check all that apply.

Answer-2: `SELECT MIN(blue + red) FROM wax.crayons;`

`SELECT -20 + MIN(red) FROM wax.crayons;`

`SELECT MIN(-20 + red) FROM wax.crayons;`

Question-3: The flights dataset includes the departure delay (in minutes) and the scheduled time of departure (as an integer, for example 3:14 in the afternoon is 1514). Write and run a query to find the average delay of only those flights that were scheduled to depart after 1:00 in the afternoon. Do not include those scheduled for exactly 1:00. Report to the nearest minute. Note: There are two columns related to departure time—be sure you're using the scheduled departure time.

Answer-3: 13

Question-4: Here is the default.orders table:

order_id	cust_id	empl_id	total
1	c	1	24.78
2	a	4	28.54
3	b	3	48.69
4	b	3	-16.39
5	z	2	29.92

How many columns and rows does the result of this query have?

`SELECT cust_id, COUNT(*), SUM(total) FROM default.orders`

`GROUP BY cust_id;`

Answer-4: 3 columns, 4 rows

Question-5: In the fly.flights table, the air time of each flight is given in minutes by the air\_time column. Write and run a query to find the average air\_time of the flights, in hours, to the nearest tenth of an hour.

Answer-5: 1.8

Question-6: Write and run a query on the fly.planes table that would answer the question, "How many different manufacturers are there for each type of aircraft?" Then use the results to enter the number of balloon manufacturers are included in the table.

Answer-6: 528

Question-7: For a table of students enrolled at a college, the query SELECT MIN(age) FROM students; gave one row in the results, with only one column. The value was 16. The query SELECT COUNT(\*) FROM students WHERE age IS NULL returned the value 2827. Choose which of the following statements is most accurate and informative:

Answer-7: The lowest known age of a student in the students table is 16.

Question-8: Which SELECT statements will return the same result as SELECT COUNT(type) AS num\_types FROM fly.planes; Check all that apply.

Answer-8: SELECT COUNT(ALL type) AS num\_types FROM fly.planes;  
SELECT COUNT(\*) AS num\_types FROM fly.planes WHERE type IS NOT NULL;

Question-9: Write and run a query in the VM to find all the airports with average departure delays of more than 30 minutes. (Note that you want the origin airports, not the destinations.) How many airports have more than 30 minutes for their average departure delay?

Answer-9: 5

Question-10: Choose the SELECT statement that returns a result set describing, for each carrier, the average air time for the flights that have a departure delay longer than the flight's air time, and only for carriers with more than 70,000 of those flights.

Answer-10: SELECT carrier, AVG(air\_time) FROM flights  
WHERE dep\_delay > air\_time  
GROUP BY carrier HAVING count(\*) > 70000;