Modern Big Data Analysis with SQL

Coursera Specialisation (Offered by Cloudera)

Course-2: Analysing Big Data with SQL

Week-6: Honours

Question-1: In 2019, a growing hotel chain sent out customer satisfaction surveys to all guests after their stay and recorded the responses in a table named survey2019. The survey includes a field for the location of the recent stay, but some customers neglect to include it in their responses, so sometimes the location field is NULL. (Note that some locations may close, and several new locations were opened in 2019.) The chain also has a table that holds average responses from previous years, for different demographics. Consider this query, which gets information from both tables for comparison.
SELECT this.location, AVG(this.room_service_rating) AS avg_2019,
AVG(s.room_service_rating) AS avg_previous FROM survey2019 this
JOIN survey_summary s ON (this.location <=> s.location)
GROUP BY this.location;
Which best describes the result of this query?
Answer-1: Only rows with a match will be used to calculate averages; there will be one row with NULL in this.location, and it will have values for both avg_2019 and avg_previous

Question-2: Following are two tables, and a query to join them.
uncles

name	age
John	38
Harry	63

aunts

name	age
Ann	24
Mildred	64

SELECT aunts.name AS aunt, uncles.name AS uncle FROM aunts JOIN uncles
ON aunts.age > uncles.age; How many rows will this query return?
Answer-2: 5

Question-3: How many rows will result if you cross join a table that has 10 rows with a table that has 60 rows?
Answer-3: 600

Question-4: Following are the schema for the fly.flights and fly.planes tables on the VM.
fly.flights

name	type
year	smallint
month	tinyint
day	tinyint
dep_time	smallint
sched_dep_time	smallint
dep_delay	smallint
arr_time	smallint
sched_arr_time	smallint
arr_delay	smallint
carrier	string
flight	smallint
tailnum	string
origin	string
dest	string
air_time	smallint

fly.planes

name	type
tailnum	string
year	int
type	string
manufacturer	string
model	string
engines	int
seats	int

Which of the following are valid semi-joins? Check all that apply.
Answer-4: SELECT origin, flight, arr_delay
FROM fly.flights f LEFT SEMI JOIN fly.planes p
ON f.tailnum = p.tailnum AND engines > 1;

SELECT origin, flight, arr_delay
FROM fly.flights f LEFT SEMI JOIN fly.planes p
ON f.tailnum = p.tailnum WHERE arr_delay>60;