SQL övningstenta

I provet ingår multiple choice frågor och frågor där du behöver utveckla mer. I multiple choice frågorna gäller att 1 alternativ är korrekt, och svarar man på flera alternativ samtidigt, räknas det som noll poäng på den frågan.

Tillåtna hjälpmedel:

Ett A4-papper handskrivna anteckningar, båda sidorna får användas. Får skriva vad som helst på detta A4-papper.

Tid: 4 timmar

Betygssättning

Godkänt: 50% dvs 15 poäng

Väl godkänt: 80% dvs 24 poäng

Question 1

Which SQL keyword is used to retrieve unique values from a column? (1p)

a) SELECT DISTINCT

b) SELECT UNIQUE

c) SELECT DIFFERENT

d) SELECT SINGLE

e) SELECT VARIED

Question 2

Which SQL function is used to return the total number of rows in a table? (1p)

a) COUNT

b) SUM

c) TOTAL

d) ROWS

e) LENGTH

Question 3

What is the correct SQL statement to select all the columns from a table named `employees`? (1p)

a) SELECT * FROM employees;

b) SELECT ALL COLUMNS FROM employees;

c) SELECT COLUMNS FROM employees;

d) SELECT ALL employees;

e) SELECT EVERYTHING FROM employees;

Question 4

This is part of a table called hemnet_data in the main schema of a database.

address varchar	asked_price double	final_price int64	rooms double	price_per_area double	area double
Markviksvägen 15	6475000.0	6950000	5.0	59402.0	117.0
Mälarhöjdsvägen 38	9950000.0	11300000	5.0	57949.0	195.0
Häradsvägen 202	5795000.0	6200000	6.0	68889.0	90.0
Hemmansvägen 18	11950000.0	12700000	5.0	115455.0	110.0
Alholmsbacken 8	7950000.0	8100000	6.0	51592.0	157.0

What will the following SQL query return? (2p)

```sql
SELECT address FROM hemnet_data WHERE rooms > 5;
```

a) All addresses.

b) All addresses with 5 or more rooms.

c) All addresses with 6 or more rooms.

d) All addresses with less than 5 rooms.

e) Empty set.

Question 5

We use the same table as in question 4.

```sql
SELECT
    rooms,
    AVG(asked_price)
FROM
    hemnet_data
GROUP BY
    rooms;
```

Describe the result set you get from this query. (2p)

Question 6

We use the same table as in question 4. Write a query to get all addresses, their sold price and the number of rooms for homes with 1,2 or 3 rooms. (2p)

Question 7

In duckdb CLI you can use .mode line in order to list all the columns in lines instead of table format. This is a record of a table in hemnet data. From a critical point of view, find out the problem(s) with this record and motivate what could be a correct fix. (2p)

```
              column00 = 8.8
               address = Norlindsvägen 25
          asked_price = 8495000.0
           coordinate = [59.3553520745678, 17.9147884910278]
          final_price = 8700000
pourcentage_difference = 2.0
            land_area = 612.0
                 area = 65.0
              commune = Bromma Kyrka, Stockholms kommun
      price_per_area = 133846.0
               rooms = 45.0
           sale_date = Såld 2022-03-28
    supplemental_area = 70.0
```

Question 8

You have a table called salaries that looks like this and is stored under main schema.

employment_type	job_title	salary_sek_yearly	salary_sek_monthly	salary_level
Full time	AI Engineer	2145000	179000	insanely high
Full time	AI Engineer	975000	81000	low
Full time	Data Engineer	1381000	115000	medium
Full time	Data Engineer	1016000	85000	low
Full time	Machine Learning Engineer	2010000	168000	insanely high
Full time	Machine Learning Engineer	1693000	141000	high
Full time	ML Engineer	4232000	353000	insanely high
Full time	ML Engineer	688000	57000	low
Full time	Data Analyst	1074000	90000	medium
Full time	Data Analyst	485000	40000	low

a) What is the result after the following query (2p)

```sql
SELECT DISTINCT salary_level FROM salaries;
```

b) What is the result after the following query (2p)

```sql
SELECT salary_level FROM staging.salaries;
```

c) What is the result after the following query (2p)

```sql
SELECT COUNT(DISTINCT salary_level) AS salaries FROM main.salaries;
```

d) What is the result after the following query (2p)

```
SELECT salary_level, COUNT(*) FROM main.salaries GROUP BY salary_level;
```

e) Describe what the following query does, you don't need to calculate the exact values. (2p)

```
SELECT job_title, AVG(salary_sek_monthly) FROM main.salaries GROUP BY
job_title;
```

Question 9

You have two tables in a schema called synthetic:

sales_jan

id	product_name	amount	sale_date
1	Laptop	1200.50	2024-01-05
2	Headphones	200.00	2024-01-12
3	Monitor	300.00	2024-01-15

sales_feb

id	product_name	amount	sale_date
3	Monitor	300.00	2024-02-10
4	Keyboard	100.00	2024-02-11
5	Mouse	50.00	2024-02-18

---

a) What is the result after the following query (2p)

```
SELECT product_name, amount FROM synthetic.sales_jan
UNION
SELECT product_name, amount FROM synthetic.sales_feb;
```

b) If I run the query in a) several times, do you expect to get the same result set each time, motivate your answer. (2p)

c) What is the result after the following query (2p)

```
SELECT product_name, amount FROM synthetic.sales_feb
EXCEPT
SELECT product_name, amount FROM synthetic.sales_jan;
```

Question 10

You have two tables in a schema called synthetic:

products

product_id	product_name	category	supplier	price
1	Monitor	Electronics	TechCorp	300.00
2	Keyboard	Accessories	InputMasters	100.00
3	Mouse	Accessories	InputMasters	50.00
4	Laptop	Electronics	GigaTech	1000.00
5	Headphones	Audio	SoundPro	150.00

sales

sale_id	product_id	sale_date	quantity	total_amount
2	2	2024-02-11	5	500.00
4	4	2024-02-20	1	1000.00
5	5	2024-02-21	4	600.00

a) Explain what this code does, you don't need to write out the whole result set (2p)

```sql
SELECT
    *
FROM
    synthetic.sales s
LEFT JOIN synthetic.products p ON
    p.product_id = s.product_id;
```

b) Write a query to get this result set (3p)

product_name	category	supplier	total_amount
Keyboard	Accessories	InputMasters	500.00
Laptop	Electronics	GigaTech	1000.00
Headphones	Audio	SoundPro	600.00