

# ДЗ #3

---

## Предисловие

---

Данные выглядят так, будто отдел с id '00-0000' ('All Occupations') должен иметь численность сотрудников (employee), равную суммарной численности отделов с id вида '%-0000%' (кроме '00-0000'), а также равной суммарной численности всех отделов, имеющих id отличный от '%-0000%'. Однако это не выполняется:

```
SELECT employee
FROM sample_07
WHERE id = '00-0000';
```

даёт 134354250,

```
SELECT SUM(employee)
FROM sample_07
WHERE id LIKE '%-0000%' AND id <> '00-0000';
```

даёт 134354280,

```
SELECT SUM(employee)
FROM sample_07
WHERE NOT id LIKE '%-0000%';
```

даёт 134309410.

Не сходится. Аналогично и для таблицы sample\_08. Поэтому далее каждая запись трактуется как самостоятельный отдел.

Также в задании не разъясняется, является ли 'salary' средней или суммарной зарплатой по отделу (или чем-то ещё). Далее предполагается, что это средняя зарплата по отделу.

## Задание 1

---

Вывести название компании с максимальным количеством сотрудников и зарплатой по отделам. Названия компаний - названия файлов датасетов (20 баллов)

Это задание можно понять по-разному, поэтому ответу на несколько более конкретных вопросов.

1. Вывести название компании с максимальным суммарным количеством сотрудников по всем отделам

```

WITH employee_sums AS
(
    SELECT SUM(employee) AS total_employees, 'sample_07' AS company_name
    FROM sample_07
    UNION
    SELECT SUM(employee) AS total_employees, 'sample_08' AS company_name
    FROM sample_08
),
max_row AS
(
    SELECT total_employees, company_name
    FROM employee_sums
    ORDER BY total_employees DESC
    LIMIT 1
)
SELECT company_name
FROM max_row;

```

Ответ:

2. Вывести название компании с максимальной суммарной зарплатой по всем отделам

```

WITH total_salaries AS
(
    -- суммарная зарплата отдела - произведение числа сотрудников на
    -- среднюю зарплату
    SELECT SUM(employee * salary) AS total_salary, 'sample_07' AS
company_name
    FROM sample_07
    UNION
    SELECT SUM(employee * salary) AS total_salary, 'sample_08' AS
company_name
    FROM sample_08
),
max_row AS
(
    SELECT total_salary, company_name
    FROM total_salaries
    ORDER BY total_salary DESC
    LIMIT 1
)
SELECT company_name
FROM max_row;

```

Ответ: `sample_08`

## Задание 2

Вывести список ID отделов обеих компаний, количество сотрудников в которых превышает 50000 человек (20 баллов)

Опять же задание можно понять по-разному. В обеих компаниях есть отделы с одинаковыми `id`. Считать ли их одним и тем же отделом (и суммировать число сотрудников), или же это разные отделы? Если разные, то нужно ли указывать компанию, нужно ли удалять дубли из вывода?

1. Вывести список ID отделов обеих компаний, количество сотрудников которых *в каждой компании* превышает 50000 человек

```
SELECT id
FROM sample_07
WHERE employee > 50000
INTERSECT
SELECT id
FROM sample_08
WHERE employee > 50000;
```

Ответ:

```
1    00-0000
2    11-0000
3    11-1011
4    11-1021
5    11-1031
6    11-2021
...
```

(всего 407 строк)

2. Вывести список ID отделов обеих компаний (без повторений), количество сотрудников которых *хотя бы в одной компании* превышает 50000 человек

```
SELECT id
FROM sample_07
WHERE employee > 50000
UNION
SELECT id
FROM sample_08
WHERE employee > 50000;
```

Ответ:

```
1  00-0000
2  11-0000
3  11-1011
4  11-1021
5  11-1031
6  11-2021
...
```

(всего 414 строк)

3. Вывести список ID отделов обеих компаний (без повторений), количество сотрудников которых *суммарно по обеим компаниям* превышает 50000 человек

```
WITH department_sizes AS
(
    -- Получаем размеры всех отделов обеих компаний (с повторениями id)
    SELECT id, employee
    FROM sample_07
    UNION ALL
    SELECT id, employee
    FROM sample_08
),
combined_departments AS
(
    -- Суммируем число сотрудников для одинаковых id
    SELECT id, SUM(employee) AS sum_employee
    FROM department_sizes
    GROUP BY id
)
SELECT id
FROM combined_departments
WHERE sum_employee > 50000;
```

Ответ:

```
1  00-0000
2  11-0000
3  11-1011
4  11-1021
5  11-1031
6  11-2011
...
```

(всего 541 строка)

4. Вывести список ID отделов обеих компаний (с указанием компании, т.к. это разные отделы), количество сотрудников которых превышает 50000 человек

```
WITH unique_department_sizes AS
(
    SELECT id, 'sample_07' AS company_name
    FROM sample_07
    WHERE employee > 50000
    UNION
    SELECT id, 'sample_08' AS company_name
    FROM sample_08
    WHERE employee > 50000
)
SELECT id, company_name
FROM unique_department_sizes;
```

Ответ:

```
1    00-0000 sample_07
2    00-0000 sample_08
3    11-0000 sample_07
4    11-0000 sample_08
5    11-1011 sample_07
6    11-1011 sample_08
...
```

(всего 821 строка)

## Задание 3

Для прошлого задания посчитать долю сотрудников отдела от всех сотрудников конкретной компании (20 баллов)

Видимо, в прошлом задании всё же имелся в виду четвёртый вариант.

```
WITH all_departments AS
(
    SELECT id, employee, 'sample_07' AS company_name
    FROM sample_07
    WHERE employee > 50000
    UNION
    SELECT id, employee, 'sample_08' AS company_name
```

```

        FROM sample_08
        WHERE employee > 50000
    ),
    company_sizes AS
    (
        SELECT SUM(employee) AS total_employees, 'sample_07' AS company_name
        FROM sample_07
        UNION
        SELECT SUM(employee) AS total_employees, 'sample_08' AS company_name
        FROM sample_08
    ),
    all_departments_with_company_sizes AS
    (
        SELECT all_departments.id, all_departments.company_name,
        all_departments.employee, company_sizes.total_employees
        FROM all_departments
        INNER JOIN company_sizes
        ON all_departments.company_name = company_sizes.company_name
    )
    SELECT id, company_name, employee / total_employees AS fraction
    FROM all_departments_with_company_sizes;

```

Ответ:

```

1  00-0000 sample_07  0.333370395372474
2  11-0000 sample_07  0.01489742615427988
3  11-1011 sample_07  0.0007422994619048472
4  11-1021 sample_07  0.004107534270062519
5  11-1031 sample_07  0.00015163096709789147
6  11-2021 sample_07  0.00041000656199076396
...

```

(всего 821 строка)

## Задание 4

Проверить, существуют ли отделы, работающие в одной компании, но отсутствующие в другой (20 баллов)

```

WITH all_departments AS
(
    SELECT id, 'sample_07' AS company_name
    FROM sample_07

```

```

        UNION
        SELECT id, 'sample_08' AS company_name
        FROM sample_08
    ),
    common_departments AS
    (
        SELECT id
        FROM sample_07
        INTERSECT
        SELECT id
        FROM sample_08
    ),
    sym_diff_departments AS
    (
        -- Симметрическая разность ID департаментов
        SELECT id
        FROM all_departments
        EXCEPT
        SELECT id
        FROM common_departments
    )
    SELECT id, company_name
    FROM all_departments
    WHERE id IN (SELECT id FROM sym_diff_departments);

```

Ответ:

```

1      27-2011 sample_08

```

Да, в компании `sample_08` есть отдел `27-2011`, отсутствующий в `sample_07`

## Задание 5

Отсортировать отделы компаний по суммарной зарплате по убыванию. Использовать информацию по отделам обеих компаний суммарно (20 баллов)

```

WITH department_salaries AS
(
    -- Получаем суммарные зарплаты отделов обеих компаний (с повторениями id)
    -- перемножая число сотрудников отдела и среднюю зарплату по отделу
    SELECT id, description, employee * salary AS department_salary
    FROM sample_07

```

```

    UNION ALL
    SELECT id, description, employee * salary AS department_salary
    FROM sample_08
),
combined_departments AS
(
    -- Объединяем одинаковые отделы
    SELECT id, description, SUM(department_salary) AS department_salary
    FROM department_salaries
    GROUP BY id, description
)
SELECT id, description, department_salary
FROM combined_departments
ORDER BY department_salary DESC;

```

Ответ:

```

1    00-0000 All Occupations 11181154104600
2    43-0000 Office and administrative support occupations    1474576257000
3    11-0000 Management occupations    1194450191000
4    41-0000 Sales and related occupations    1022318779200
5    29-0000 Healthcare practitioners and technical occupations
927630705600
6    25-0000 Education, training, and library occupations    797173114600
...

```