**Часть 1. Написание SQL-запросов**

В базе есть различные данные о вакансиях и резюме. Пример того, как эти данные устроены, можно посмотреть в excel-файле ([бд\_sql](https://docs.google.com/spreadsheets/d/1AqzBK1bMIsBAAo3qdUQh7i6rh8AWl-TlWW-ackPyN00/edit?usp=sharing)): там есть и описание данных в том числе. Изучите этот файл и на основе таблиц, указанных в примере, напишите 3 sql-запроса, которые «достанут» из этих таблиц нужные срезы данных:

1. Выгрузить число созданных вакансий (в динамике по месяцам), опубликованных в России, в названии которых встречается слово «водитель», предлагающих «Гибкий график» работы, за 2020-2021 годы. Важно, чтобы вакансии на момент сбора данных не были удаленными / заблокированными.

*Запросы писал под PostgreSQL. Использовал разные вариации написания (алиасы, преобразование типов данных, условия джоинов). Во 2 запросе вывел топ-10.*

**Запрос №1**

SELECT DISTINCT DATE\_TRUNC('MONTH', v.creation\_time::TIMESTAMP) year\_month,

COUNT(vacancy\_id) vacancy\_amount

FROM vacancy v

JOIN area a ON v.area\_id = a.area\_id

WHERE a.country\_name = 'Россия'

AND v.creation\_time::DATE BETWEEN '2020-01-01' AND '2021-12-31'

AND v.name = 'Водитель'

AND v.work\_schedule = 'Гибкий график'

AND v.disabled IS FALSE

GROUP BY 1

ORDER BY 1

2. Выяснить, в каких регионах РФ (85 штук) выше всего *доля* вакансий, предлагающих удаленную работу. Вакансии должны быть не заархивированными, не заблокированными и не удаленными, и быть созданными в 2021-2022 годах.

**Запрос №2**

WITH t1 AS (

SELECT DISTINCT a.region\_name rn,

COUNT(v.vacancy\_id) vct

FROM vacancy v

JOIN area a ON v.area\_id = a.area\_id

WHERE CAST(v.creation\_time AS DATE) BETWEEN '2021-01-01' AND '2022-12-31'

AND v.disabled IS FALSE

AND v.archived IS FALSE

GROUP BY 1

),

t2 AS (

SELECT DISTINCT a.region\_name rn,

COUNT(v.vacancy\_id) vcr

FROM vacancy v

JOIN area a ON v.area\_id = a.area\_id

WHERE CAST(v.creation\_time AS DATE) BETWEEN '2021-01-01' AND '2022-12-31'

AND v.disabled IS FALSE

AND v.archived IS FALSE

AND v.work\_schedule = 'Удаленная работа'

GROUP BY 1

)

SELECT t1.rn region,

ROUND(t2.vcr / t1.vct \* 100, 2) remote\_vacancy\_percent

FROM t1

JOIN t2 ON t1.rn = t2.rn

ORDER BY 2 DESC

LIMIT 10

3. Подсчитать «вилку» (10,25,50,75 и 90 процентиль) ожидаемых зарплат (в рублях) из московских и питерских резюме, имеющих роль «разработчик» (id роли — 91), по городам и возрастным группам (группы сделать как в примере таблицы ниже, не учитывать резюме без указания даты рождения — такие тоже есть). Возрастные группы считать на дату составления запроса. Резюме должно быть не удалено и иметь статус «завершено». Дополнительно выяснить (при помощи того же запроса) долю резюме по каждой группе, в которых указана ожидаемая зарплата. Пример таблицы, которая должна получиться на выходе:

| Город | Возрастная группа | Доля резюме с указанной зарплатой | 10 процентиль | 25 процентиль | 50 процентиль (медиана) | 75 процентиль | 90 процентиль |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Москва | 17 лет и младше |  |  |  |  |  |  |
| Москва | 18–24 |  |  |  |  |  |  |
| Москва | 25–34 |  |  |  |  |  |  |
| Москва | 35–44 |  |  |  |  |  |  |
| Москва | 45–54 |  |  |  |  |  |  |
| Москва | 55 и старше |  |  |  |  |  |  |
| … | … |  |  |  |  |  |  |

**Запрос №3**

WITH t1 AS (

SELECT DISTINCT a.area\_name city,

r.resume\_id emp,

EXTRACT (YEAR FROM AGE(CURRENT\_DATE(), r.birth\_day::DATE)) emp\_age,

ROUND(r.compensation/c.rate) exp\_salary

FROM resume r

JOIN area a USING(area\_id)

JOIN currency c ON r.currency=c.code

WHERE a.area\_name IN ('Москва','Санкт-Петербург')

AND 91 = ANY (r.role\_id\_list)

AND r.birth\_day IS NOT NULL

AND r.disabled IS FALSE

AND r.is\_finished = 1

)

SELECT city,

CASE

WHEN emp\_age <= 17 THEN '17\_and\_younger'

WHEN 17 < emp\_age AND emp\_age <= 24 THEN '18\_to\_24'

WHEN 24 < emp\_age AND emp\_age <= 34 THEN '25\_to\_34'

WHEN 34 < emp\_age AND emp\_age <= 44 THEN '35\_to\_44'

WHEN 44 < emp\_age AND emp\_age <= 54 THEN '45\_to\_54'

ELSE '55\_and\_older'

END AS age\_group,

ROUND(COUNT(exp\_salary)/COUNT()\*100, 2) stated\_salary\_percent,

PERCENTILE\_DISC(0.1) WITHIN GROUP (ORDER BY exp\_salary) percentile\_10,

PERCENTILE\_DISC(0.25) WITHIN GROUP (ORDER BY exp\_salary) percentile\_25,

PERCENTILE\_DISC(0.5) WITHIN GROUP (ORDER BY exp\_salary) percentile\_50,

PERCENTILE\_DISC(0.75) WITHIN GROUP (ORDER BY exp\_salary) percentile\_75,

PERCENTILE\_DISC(0.9) WITHIN GROUP (ORDER BY exp\_salary) percentile\_90

FROM t1

GROUP BY 1, 2

ORDER BY 1, 2

**Часть 2. Визуализация данных и выводы**

У вас есть набор таблиц (файл [таблицы\_для\_дашборда](https://docs.google.com/spreadsheets/d/194N2boyPof0zvTWf0p57OZwGNOw2tYxuaAWrHEM9RK0/edit?usp=sharing)) с данными о вакансиях и резюме на тему командировок. Сделайте дашборд в любом удобном для Вас BI-сервисе с визуализацией этих данных и выводами.

**Ссылка на дашборд -** <https://datalens.yandex.cloud/k50a0imujure7>

Выводы:

* Исходя из данных за 2019-2021гг. видно, что доли вакансий с упоминанием обязательных командировок по отраслям компаний меняются от года к году независимо. То есть нельзя сказать, что они становятся меньше или растут с течением времени. Отрасли в которых наибольший процент (>3%) вакансий с упоминанием обязательных командировок - “Сельское хозяйство”, “Химическое производство, удобрения”, “Электроника, приборостроение, бытовая техника, компьютеры и оргтехника” и “Энергетика”.

Если брать вакансии с упоминанием обязательных командировок в целом по России, то видим, что процент снижается с каждым годом. С 1.53% в 2019 году упал до 1.28% в 2021 году. При этом предлагаемая зарплата (медиана) с каждым годом растет и вакансии с обязательными командировками оплачиваются лучше в среднем на 12тр.

* По данным за 2021 год в 55.34% резюме указано “Не готовы к командировкам”. Если смотреть в разрезе профобластей, то самый высокий процент (67-82%) в следующих областях - “Начало карьеры, студенты”, “Рабочий персонал”, “Спортивные клубы, фитнес, салоны красоты”, “Домашний персонал”.

По возрасту сильные различия наблюдаются лишь в группах до 17 лет и в группе 18-24, не готовы 98% и 74% соответственно. В остальных возрастных группах распределение практически одинаковое, только люди старше 55 охотнее согласны на командировки (на 7-10%).

Мужчины чаще соглашаются на командировки, чем женщины, 38.5% и 22.3% соответственно. Хотя к редким командировкам женщины чуть более благосклонны (на 2.4%).