

***Проект 2. Подгрузка новых данных. Уточнение анализа.***

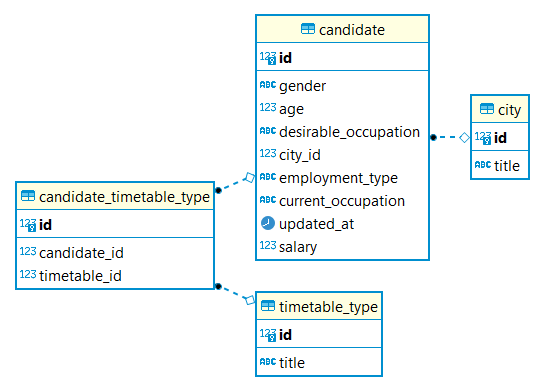
### **Описание проекта.**

Представьте, что вы работаете в некотором дата-центре по трудоустройству, куда стекаются данные с различных платформ для найма. Вам необходимо познакомиться с данными, понять, с какими резюме вы имеете дело, а дальше собрать статистику для различных кадровых агентств и статистических центров.

**Этапы проекта:**

1. знакомство с датасетом;
2. предварительный анализ данных;
3. анализ кандидатов;
4. глобальный анализ показателей.

## ***1. Знакомство с датасетом.***

Датасет включает в себя 4 таблицы представленные на диаграмме.

### 

### **2. Предварительный анализ данных.**

***Задание 2.1***

*Рассчитайте максимальный возраст (max\_age) кандидата в таблице.*

*Запрос:*

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*SELECT MAX(ca.age) max\_age

FROM hh.candidate ca

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Результат запроса:*

**

*Вывод:* Максимальный возраст кандидата - 100 лет, что явно является ошибкой в резюме. При анализе данных данный возраст необходимо исключить.

***Задание 2.2***

*Рассчитайте минимальный возраст (min\_age) кандидата в таблице.*

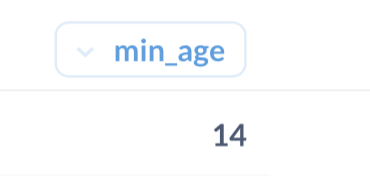
*Запрос:*

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*SELECT MIN(ca.age) min\_age

FROM hh.candidate ca

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

*Результат запроса:*

**

*Вывод:* Минимальный возраст составляет 14 лет. Достаточно молодой возраст для работы, но допустимый.

***Задание 2.3***

*Напишите запрос, который позволит рассчитать для каждого возраста (age) сколько (cnt) человек этого возраста у нас есть. Отсортируйте результат по возрасту в обратном порядке.*

*Запрос:*

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

SELECT c.age,

COUNT(c.age) cnt

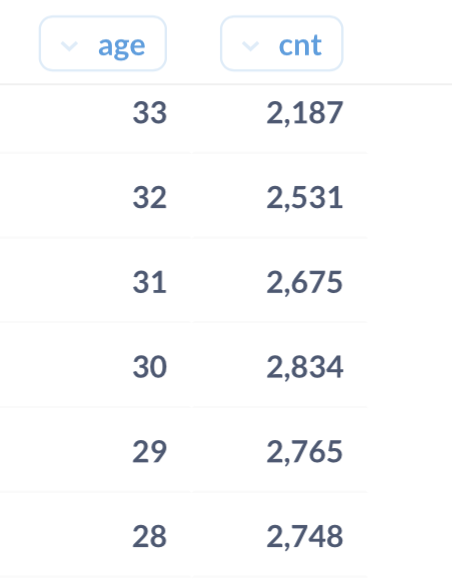
FROM hh.candidate c

GROUP BY c.age

ORDER BY 1 DESC

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

*Результат запроса:*

* *

*Вывод:* В возрасте от 24 до 33 лет находится наибольшее число соискателей, более 2000 человек на каждый возраст, где модой является возраст 30 лет - 2834 человека. Возможно связано с тем, что молодые люди не имеют опыта и им сложнее найти работу или еще ищут себя и находятся в постоянном поиске.

***Задание 2.4***

*Найдите количество кандидатов, которые старше 40 лет. Не забудьте отфильтровать «ошибочный» возраст 100.*

*Запрос:*

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

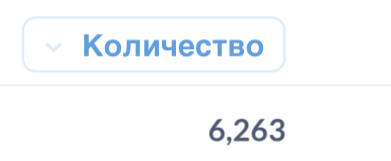
SELECT COUNT(c.age)

FROM hh.candidate c

WHERE c.age BETWEEN 41 AND 99

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

*Результат запроса:*



*Вывод:* 6263 соискателя находятся в возрасте более 40 лет. Они составляют около 14% от всех заявок.

## **3. Глобальный анализ показателей.**

***Задание 3.1***

*Напишите запрос, который позволит узнать, сколько (cnt) у нас кандидатов из каждого города (city). Формат выборки: city, cnt. Группировку таблицы необходимо провести по столбцу title, результат отсортируйте по количеству в обратном порядке.*

*Запрос:*

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

SELECT ci.title city,

COUNT(ca.id) cnt

FROM hh.candidate ca

LEFT JOIN hh.city ci ON ca.city\_id = ci.id

GROUP BY ci.title

ORDER BY 2 DESC

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

*Результат запроса:*



*Вывод:* Чем крупнее город, тем больше кандидатов ищут работу. В столице наибольшее число соискателей - 16622 человека.

***Задание 3.2***

*Напишите запрос, который позволит понять, каких кандидатов из Москвы устроит «проектная работа». Формат выборки: gender, age, desirable\_occupation, city, employment\_type.*

*Отсортируйте результат по id кандидата.*

*Запрос:*

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

SELECT ca.gender,

ca.age,

ca.desirable\_occupation,

ci.title city,

ca.employment\_type

FROM hh.candidate ca

LEFT JOIN hh.city ci ON ca.city\_id = ci.id

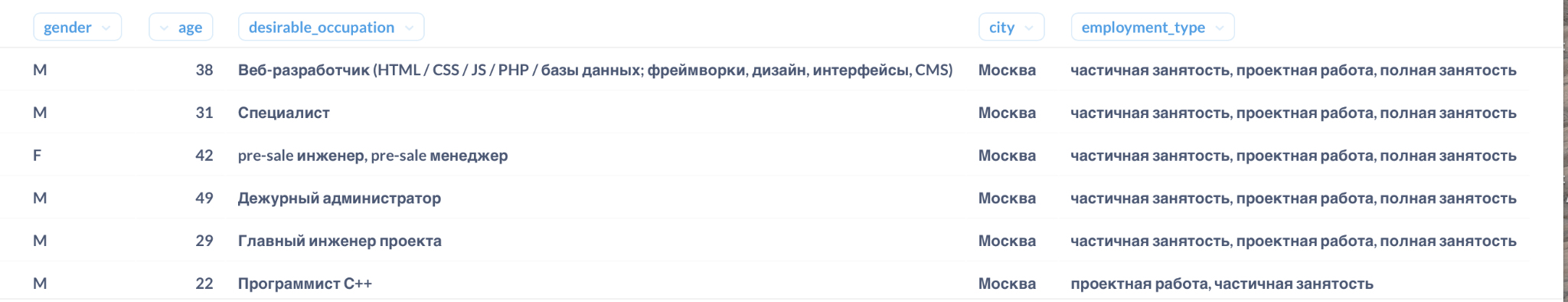
WHERE ci.title = 'Москва'

AND ca.employment\_type LIKE '%проектная работа%'

ORDER BY ca.id

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

*Результат запроса:*



*Вывод:* Выборка получилась достаточно большой и получить информацию из нее достаточно сложно.

***Задание 3.3***

*Отфильтруйте только самые популярные IT-профессии — разработчик, аналитик, программист. Обратите внимание, что данные названия могут быть написаны как с большой, так и с маленькой буквы. Отсортируйте результат по id кандидата.*

*Запрос:*

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

SELECT ca.gender,

ca.age,

ca.desirable\_occupation,

ci.title city,

ca.employment\_type

FROM hh.candidate ca

LEFT JOIN hh.city ci ON ca.city\_id = ci.id

WHERE ci.title = 'Москва'

AND ca.employment\_type LIKE '%проектная работа%'

AND (LOWER(ca.desirable\_occupation) LIKE '%разработчик%'

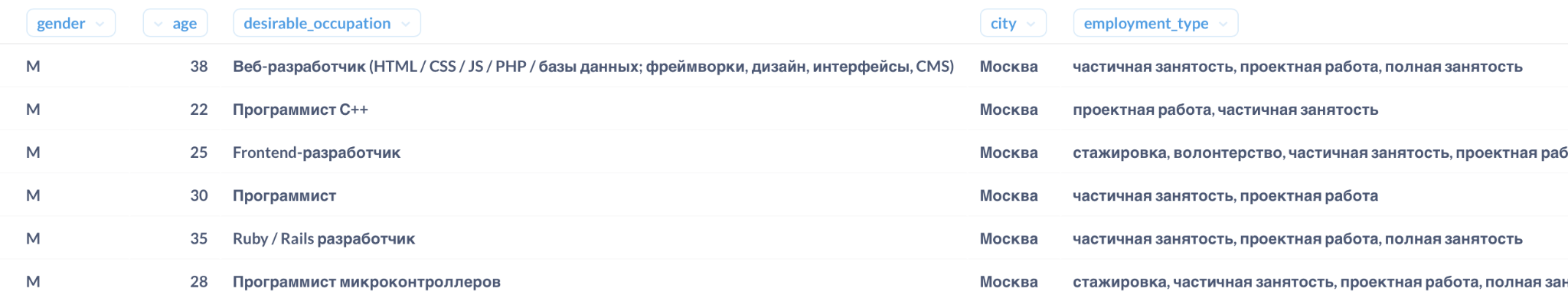
OR LOWER(ca.desirable\_occupation) LIKE '%аналитик%'

OR LOWER(ca.desirable\_occupation) LIKE '%программист%')

ORDER BY ca.id

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

*Результат запроса:*



*Вывод:* Получили выборку самых популярных IT-профессий в городе Москва. В дальнейшем из этих данных можно отбирать кандидатов в данной сфере исходя из дополнительных параметров. Например по желаемой заработной плате и возрасту.

***Задание 3.4***

*Для общей информации попробуйте выбрать номера и города кандидатов, у которых занимаемая должность совпадает с желаемой.*

*Формат выборки: id, city. Отсортируйте результат по городу и id кандидата.*

*Запрос:*

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

SELECT ca.id,

ci.title city

FROM hh.candidate ca

LEFT JOIN hh.city ci ON ca.city\_id = ci.id

WHERE ca.current\_occupation = ca.desirable\_occupation

ORDER BY 2, 1

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

*Результат запроса:*



*Вывод:* Из выборки можем предположить, что данные кандидаты имеют опыт работы в желаемой должности и могут быть в приоритете при трудоустройстве на работу.

***Задание 3.5***

*Определите количество кандидатов пенсионного возраста.*

*Пенсионный возраст для мужчин наступает в 65 лет, для женщин — в 60 лет.*

*Запрос:*

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

SELECT COUNT(ca.id)

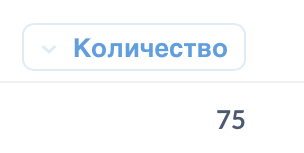
FROM hh.candidate ca

WHERE (ca.gender = 'M' AND ca.age BETWEEN 65 and 99)

OR (ca.gender = 'F' AND ca.age BETWEEN 60 and 99)

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

*Результат запроса:*



*Вывод:* 75 соискателей находятся в пенсионном возрасте. Если отсортировать их по городам, то большинство находится в Москве и Санкт-Петербурге. Возможно это связано с дорогой жизнью в крупных городах, что заставляет людей продолжать работать.

## **4. Анализ кандидатов для заказчиков.**

***Задание 4.1***

*Для добывающей компании нам необходимо подобрать кандидатов из Новосибирска, Омска, Томска и Тюмени, которые готовы работать вахтовым методом.*

*Формат выборки: gender, age, desirable\_occupation, city, employment\_type, timetable\_type.*

*Отсортируйте результат по городу и номеру кандидата.*

*Запрос:*

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

SELECT ca.gender,

ca.age,

ca.desirable\_occupation,

ci.title city,

ca.employment\_type,

tt.title timetable\_type

FROM hh.candidate ca

LEFT JOIN hh.city ci ON ca.city\_id = ci.id

LEFT JOIN hh.candidate\_timetable\_type ctt ON ca.id = ctt.candidate\_id

LEFT JOIN hh.timetable\_type tt ON tt.id = ctt.timetable\_id

WHERE ci.title IN ('Новосибирск', 'Омск', 'Томск', 'Тюмень')

AND tt.title = 'вахтовый метод'

ORDER BY ci.title, ca.id

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

*Результат запроса:*



*Вывод:* В данных городах соискатели готовые работать вахтовым методом только мужчины в молодом возрасте. В основном на должность инженера и программиста. Вероятно это связано с тем, что вахтовый метод очень специфический и не каждая женщина готова к такому графику работы, а также мужчины старше 40 и те у которых есть семья предпочитают другой метод работы.

***Задание 4.2***

*Для заказчиков из Санкт-Петербурга нам необходимо собрать список из 10 желаемых профессий кандидатов из того же города от 16 до 21 года (в выборку включается 16 и 21, сортировка производится по возрасту) с указанием их возраста, а также добавить строку Total с общим количеством таких кандидатов. Напишите запрос, который позволит получить выборку вида:*

*Запрос:*

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

(SELECT ca.desirable\_occupation,

ca.age

FROM hh.candidate ca

LEFT JOIN hh.city ci ON ca.city\_id = ci.id

WHERE ci.title = 'Санкт-Петербург'

AND ca.age BETWEEN 16 AND 21

ORDER BY 2

LIMIT 10)

UNION ALL

(SELECT 'Total', COUNT(\*)

FROM hh.candidate ca

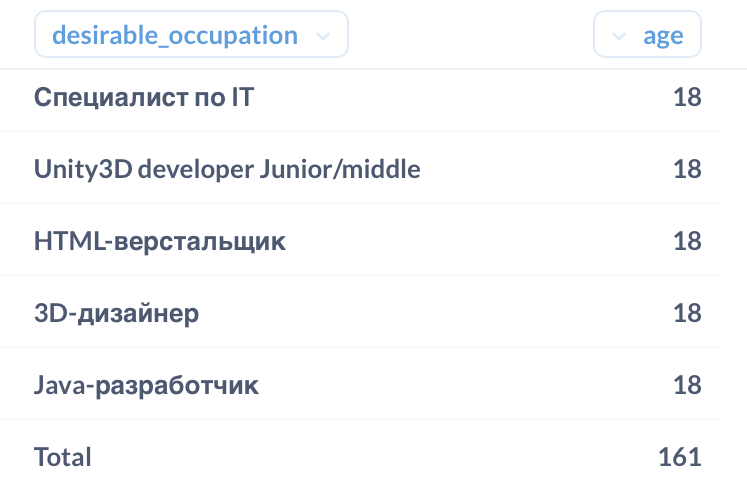
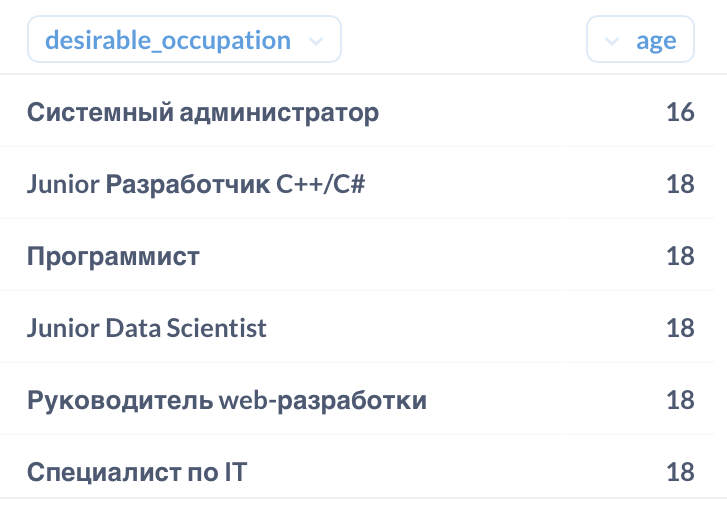
LEFT JOIN hh.city ci ON ca.city\_id = ci.id

WHERE ci.title = 'Санкт-Петербург'

AND ca.age BETWEEN 16 AND 21)

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

*Результат запроса:*



*Вывод:* Из данной выборки мы получили самые востребованные профессии среди молодых людей от 16 до 21 года. Как можно заметить - это IT-профессии. Наш современный мир - это век технологий. Поэтому большинство людей, особенно среди молодого поколения интересуются данными профессиями. Есть множество различных направлений в данной области и возможность выбрать то, что интересно. И плюс к тому - это одна из самых высокооплачиваемых профессий.

*Общий вывод: Познакомившись с предоставленными данными и проведя анализ мы видим, что люди в любом возрасте ищут работу. Мы получили резюме начиная от 14 лет до 77 лет. А это 44744 заявки с 984 различных города. Наибольшее число заявок сосредоточено в крупных городах, таких как Москва, Санкт-Петербург и Новосибирск. На долю столицы приходится 14% всех заявок - 16622. Что является логичным, население города большое, и люди едут в столицу с маленьких городов, деревень и из других стран с целью заработать. 50% всех заявок приходится на возраст 27-36 лет.*

*Мы рассмотрели самые популярные IT-профессии в городе Москва и если не учитывать желаемый график работы, то 2926 соискателей интересует данная сфера деятельности, а это 18% соискателей. Если узнать на сколько данные профессии востребованы среди всех городов, то это тоже составляет 18%. Достаточно высокий показатель, что говорит о востребованности IT-профессий в современном мире. И это мы рассмотрели лишь самые востребованные IT-профессий, а область IT достаточно обширная. Иначе мы получим намного выше процент востребованности IT-профессий. Что касается возраста данных соискателей, то медианный возраст - 30 лет. Это предсказуемо, ведь молодежь интересует данная сфера все больше и больше, тем более зарплаты достаточно высокие.*

*Также можно провести дополнительный анализ и узнать на сколько отличается средняя желаемая заработная плата в области IT в различных городах.*

*Сейчас очень распространена удаленная работа. Можно отобрать список кандидатов, которых это интересует.*