#### PROJECT - 2

# Задание 2.1

Рассчитайте максимальный возраст (max age) кандидата в таблице.

#### Текст запроса:

SELECT
age max\_age
FROM hh.candidate
ORDER BY age DESC
LIMIT 1

Результат: 100

**Выводы:** Скорее всего, это ошибка при вводе данных. Если это так, то при запросах, не касающихся возраста, данные этого кандидата можно использовать. Было бы полезно проверить, есть ли еще кандидаты с возрастом более 75 лет.

# Задание 2.2

Теперь давайте рассчитаем минимальный возраст (min\_age) кандидата в таблице.

## Текст запроса:

SELECT
age min\_age
FROM hh.candidate
ORDER BY age ASC
LIMIT 1

Результат: 14

**Выводы:** В данном случае возможна как ошибка при вводе данных, так и наличие реального кандидата с таким возрастом. Данные об образовании и опыте работе позволили бы уточнить это.

# Задание 2.3

Попробуем «почистить» данные. Напишите запрос, который позволит посчитать для каждого возраста (age) сколько (cnt) человек этого возраста у нас есть.

Отсортируйте результат по возрасту в обратном порядке.

#### Текст запроса:

SELECT
age,
count(age) cnt
FROM hh.candidate
GROUP BY (age)
ORDER BY age DESC

#### Результат:

v age	v cnt
100	1
77	1
76	1
73	4
72	3

**Выводы:** Результат подтверждает предположение о том, что значение возраста 100 - это выброс и подлежит фильтрации. Значений от 77 до 72 лет мало, и статистического значения они не имеют.

# Задание 2.4

По данным Росстата, средний возраст занятых в экономике России составляет 39.7 лет. Мы округлим это значение до 40. Найдите количество кандидатов, которые старше данного возраста. Не забудьте отфильтровать «ошибочный» возраст 100.

## Текст запроса:

SELECT count(\*)
FROM hh.candidate
WHERE age > 40 and age <100
--исключаем выброс по возрасту 100 лет

## Результат:

# 6,263

**Выводы:** Количество кандидатов с возрастом более 40 лет значительно. Интересно определить моду распределения кандидатов по возрасту и сравнить с данными Росстата.

# Задание 3.1

Для начала напишите запрос, который позволит узнать, сколько (cnt) у нас кандидатов из каждого города (city).

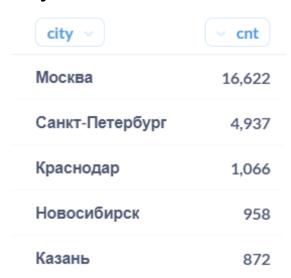
Формат выборки: city, cnt.

Группировку таблицы необходимо провести по столбцу title, результат отсортируйте по количеству в обратном порядке.

#### Текст запроса:

SELECT
ct.title city,
count(\*) cnt
FROM hh.candidate cn
JOIN hh.city ct on cn.city\_id = ct.id
GROUP BY ct.title
ORDER BY cnt DESC

## Результат:



**Выводы:** Данные показывают многократное превышение количества кандидатов из Москвы над количеством кандидатов из Санкт-Петербурга и, соответственно, кандидатов из Санкт-Петербурга над количеством кандидатов из других городов.

## Задание 3.2

Москва бросается в глаза как, пожалуй, самый активный рынок труда. 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.

#### Текст запроса:

```
SELECT
cn.gender,
cn.age,
cn.desirable_occupation,
ct.title city,
cn.employment_type
FROM hh.candidate cn
JOIN hh.city ct on cn.city_id = ct.id
WHERE ct.title = 'Mocква' and cn.employment_type like '%проектная работа%'
ORDER BY cn.id
```

## Результат:

gender v	${\sf desirable\_occupation} \ \lor \\ \\ {\sf city} \ \lor \\ \\ {\sf employment\_type} \ \lor \\ \\ \\ {\sf employment\_type} \ \lor \\ \\ \\ \\ {\sf employment\_type} \ \lor \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\$	
М	Веб-разработчик (HTML / CSS / JS / PHP / базы данных; фреймворки, дизайн, интерфейсы, CMS) Москва частичная занятость, проектная	работа, полная занятость
М	Специалист Москва частичная занятость, проектная	а работа, полная занятость
F	ore-sale инженер, pre-sale менеджер Москва частичная занятость, проектная	работа, полная занятость
М	<b>]</b> ежурный администратор Москва частичная занятость, проектная	а работа, полная занятость
М	лавный инженер проекта Москва частичная занятость, проектная	а работа, полная занятость
М	Программист С++ Москва проектная работа, частичная за	нятость

**Выводы:** В первой части проекта в качестве дополнительного задания я анализировал влияние признака наличия типа занятости "проектная работа" на величину желаемой заработной платы. Было выявлено, что наличие типа занятости "проектная работа" в резюме повышает зарплатные ожидания кандидатов. Интересно проанализировать наличие этой зависимости для Москвы, как вариант, посчитать среднюю желаемую зарплату для кандидатов с

признаком "проектная работа" в типе занятости и для кандидатов без этого признака.

# Задание 3.3

Данных оказалось многовато. 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.

#### Текст запроса:

```
SELECT
cn.gender,
cn.age,
cn.desirable_occupation,
ct.title city,
cn.employment_type
FROM hh.candidate cn
JOIN hh.city ct on cn.city_id = ct.id
WHERE ct.title = 'Mockba' and cn.employment_type LIKE '%проектная работа%'
AND ((lower(cn.desirable_occupation) LIKE '%разработчик%')
OR (lower(cn.desirable_occupation) LIKE '%аналитик%')
OR (lower(cn.desirable_occupation) LIKE '%программист%'))
ORDER BY cn.id
```

#### Результат:

gender v	age	desirable_occupation ∨	city ~	employment_type ~
М	38	Beб-разработчик (HTML / CSS / JS / PHP / базы данных; фреймворки, дизайн, интерфейсы, CMS)	Москва	частичная занятость, проектная работа, полная занятость
М	22	Программист С++	Москва	проектная работа, частичная занятость
М	25	Frontend-разработчик	Москва	стажировка, волонтерство, частичная занятость, проект
М	30	Программист	Москва	частичная занятость, проектная работа
М	35	Ruby / Rails разработчик	Москва	частичная занятость, проектная работа, полная занятость
М	28	Программист микроконтроллеров	Москва	стажировка, частичная занятость, проектная работа, пол

**Выводы:** Может оказаться интересным проанализировать средний возраст кандидатов и среднюю желаемую зарплату по категориям самых популярных *IT*-профессий.

# Задание 3.4

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

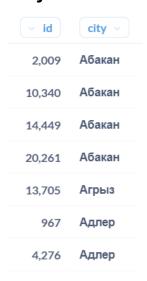
Формат выборки: id, city.

Отсортируйте результат по городу и *id* кандидата.

#### Текст запроса:

```
SELECT
cn.id,
ct.title city
FROM hh.candidate cn
JOIN hh.city ct on cn.city_id = ct.id
WHERE cn.desirable_occupation = cn.current_occupation
ORDER BY city, id
```

#### Результат:



**Выводы:** В результате представлены кандидаты, которые хотят продолжать карьеру в выбранном направлении и не собираются его менять. Возможно, это люди со значительным опытом. Возможно, часть кандидатов хочет продолжать карьеру за счет релокации.

## Задание 3.5

Определите количество кандидатов пенсионного возраста.

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

#### Текст запроса:

SELECT
COUNT(\*)
FROM hh.candidate cn
WHERE cn.age BETWEEN 65 and 99 OR cn.age BETWEEN 60 and 64 AND
cn.gender = 'F'
--исключаем выброс по возрасту 100 лет

#### Результат:

**75** 

Выводы: Количество кандидатов пенсионного возраста весьма невелико.

# 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.

#### Текст запроса:

```
SELECT
  cn.gender,
  cn.age,
  cn.desirable occupation,
  ct.title city,
  cn.employment_type,
  tt.title timetable type
  FROM hh.candidate cn
  JOIN hh.city ct on cn.city id = ct.id
  JOIN hh.candidate timetable type ctt on ctt.candidate id = cn.id
--candidate timetable type - доп. таблица для организации связи
многие-ко-многим
  JOIN hh.timetable type tt on tt.id = ctt.timetable id
--timetable type - справочник вариантов рабочего графика
WHERE ct.title IN ('Новосибирск', 'Омск', 'Томск', 'Тюмень') and tt.title = 'вахтовый
метод'
ORDER BY ct.title, cn.id
```

#### Результат:

gender	v age	desirable_occupation >	city v	employment_type ∨	timetable_type >
М	29	ИТИнженер	Новосибирск	полная занятость	вахтовый метод
М	25	Заместитель начальника лаборатории	Новосибирск	проектная работа, стажировка, частичная занятость, полная занятость	вахтовый метод
М	30	Ведущий инженер, Специалист по защите информации,	Новосибирск	частичная занятость, полная занятость	вахтовый метод
М	23	Программист	Новосибирск	полная занятость	вахтовый метод
М	35	Инженер АСУТП, инженер-электроник	Омск	полная занятость	вахтовый метод
М	25	Тестировщик ПО	Омск	стажировка, полная занятость	вахтовый метод
М	26	Специалист технической поддержки	Томск	частичная занятость, полная занятость	вахтовый метод

**Выводы:** Интересно выяснить, зависят ли зарплатные ожидания кандидатов от города проживания. Кроме того, могут оказаться полезными данные по желаемой заработной плате в зависимости от возраста, разбитого на категории 20-24, 25-29, 30-34, 35-39, 40-50, свыше 50.

# 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:

<sup>ABC</sup> desirable_occupation	<sup>123</sup> age	V:
Системный администратор		16
Junior Разработчик C++/C#		18
3D-дизайнер		18
Unity3D developer Junior/middle		18
Специалист по IT		18
Java-разработчик		18
Программист		18
Руководитель web-разработки		18
HTML-верстальщик		18
Junior Data Scientist		18
Total		88

## Текст запроса:

```
(SELECT cn.desirable_occupation, cn.age FROM hh.candidate cn JOIN hh.city ct on cn.city_id = ct.id WHERE ct.title like 'Санкт-Петербург' AND cn.age BETWEEN 16 and 21 ORDER BY cn.age
```

```
LIMIT 10)
--Добавим итоговую строку
UNION ALL
SELECT
'Total',
COUNT(cn.id)
FROM hh.candidate cn
JOIN hh.city ct on cn.city_id = ct.id
WHERE ct.title like 'Санкт-Петербург' AND cn.age BETWEEN 16 and 21
ORDER BY age
```

#### Результат:

desirable_occupation v	v age
Системный администратор	16
Junior Разработчик C++/C#	18
Программист	18
Junior Data Scientist	18
Руководитель web-разработки	18
Специалист по ІТ	18
Unity3D developer Junior/middle	18
HTML-верстальщик	18
3D-дизайнер	18
Java-разработчик	18
Total	161

**Выводы:** Кандидатов из Санкт-Петербурга от 16 до 21 года 161 человек, при общем количестве кандидатов 4937, что составляет около 3%.

#### Общий вывод по проекту:

Количество кандидатов пенсионного возраста весьма невелико. Можно проанализировать, относятся ли эти кандидаты к самым популярным *IT*-профессиям или являются редкими (уникальными) специалистами.

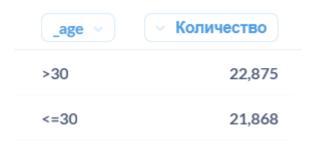
Интересно посмотреть общее количество кандидатов с возрастом от 16 до 21 года и их распределение по городам.

Можно определить моду распределения кандидатов по возрасту и сравнить с данными Росстата. Мода распределения данных hh.ru относится к кандидатам, ищущим работу, а данные Росстата - ко всем работающим.

#### Текст запроса:

-- МОДА РАСПРЕДЕЛЕНИЯ КАНДИДАТОВ ПО ВОЗРАСТУ SELECT '>30' \_age, count(\*) FROM hh.candidate WHERE age > 30 and age != 100 -- исключаем выброс по возрасту 100 лет UNION all SELECT ' <=30', count(\*) FROM hh.candidate WHERE age <= 30 and age != 100

#### Результат:



**Вывод:** Средний возраст кандидатов, ищущих работу, меньше среднего возраста занятых в экономике России

Интересно проанализировать признак "Готовность к переезду" в разрезе городов и желаемую заработную заработную плату по городам.

В первой части проекта в качестве дополнительного задания я анализировал влияние признака наличия типа занятости "проектная работа" на величину желаемой заработной платы. Было выявлено, что наличие типа занятости "проектная работа" в резюме повышает зарплатные ожидания кандидатов. Интересно проанализировать наличие этой зависимости для Москвы и Санкт-Петербурга.

#### по Заданию 3.4

Для кандидатов, которые хотят продолжать карьеру в выбранном направлении и не собираются его менять можно проанализировать, либо это люди не

имеющие карьерных амбиций, либо они хотят продолжать карьеру за счет релокации.

Можно вычислить средний возраст кандидатов, которые хотят продолжать карьеру в выбранном направлении и не собираются его менять, и остальных, готовых поменять должность, а также долю кандидатов, готовых к переезду в этих двух категориях и их средний возраст.