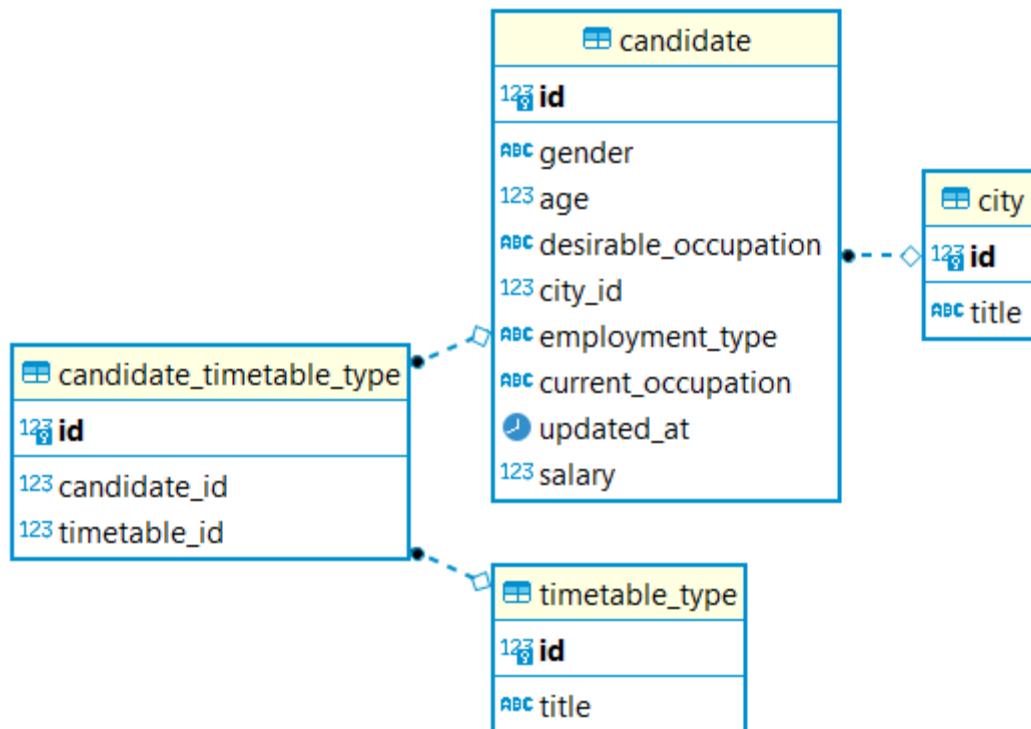


SQL Project 2

1. Введение. Знакомство с датасетом.



2. Предварительный анализ данных

Задание 2.1

Рассчитайте максимальный возраст (**max_age**) кандидата в таблице.

```
select
  max(can.age)
from hh.candidate can
```

100

Выводы по результатам отчета: Максимальный возраст кандидата в 100 лет выглядит довольно подозрительно, в данных по возрасту кандидатов содержатся ошибочные записи, их мы будем игнорировать в процессе обработке возраста кандидатов.

Задание 2.2

Теперь давайте рассчитаем минимальный возраст (**min_age**) кандидата в таблице.

```
select
  min(can.age)
from hh.candidate can
```

14

Выводы по результатам отчета: минимальный возраст в 14 лет соответствует текущим реалиям нашего рынка труда.

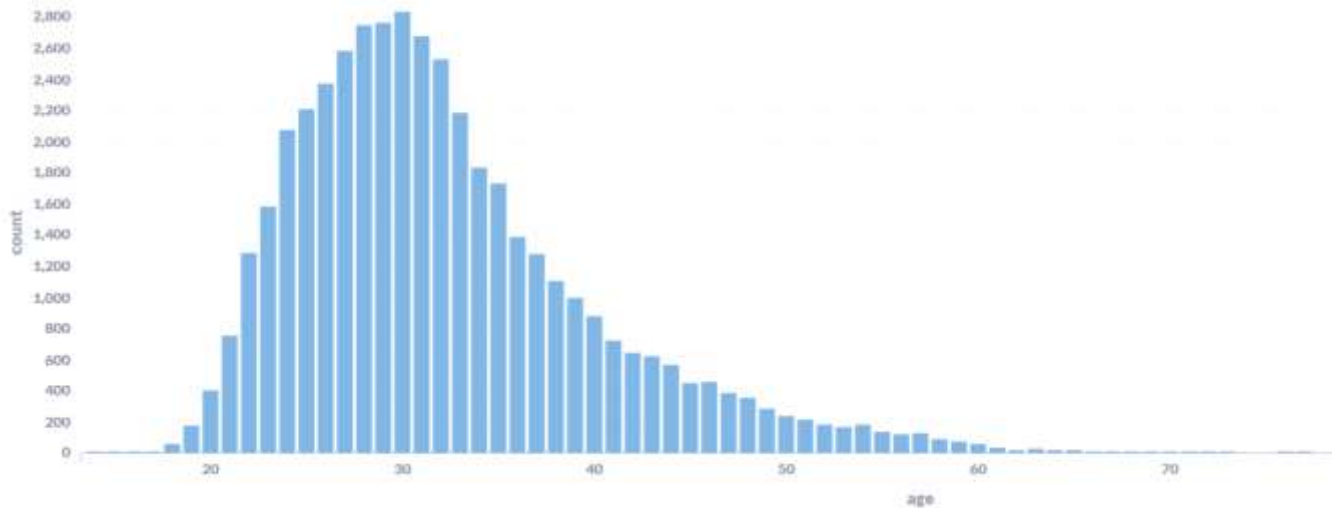
Задание 2.3

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

```
select
  can.age age,
  count(can.id)
from hh.candidate can
group by can.age
order by age desc
```

age	Количество
100	1
77	1
76	1

Выводы по результатам отчета: в текущем датасете мы видим логнормальное распределение возрастного состава соискателей, что говорит о качестве выборки.



Задание 2.4

Найдите количество кандидатов, которые старше данного возраста (40). Не забудьте отфильтровать «ошибочный» возраст 100.

```
select
  count(can.id)
from hh.candidate can
where can.age > 40 and can.age <> 100 --
убираем недостоверные данные см п. 2.1.
```

6,263

3. Глобальный анализ показателей

Задание 3.1

Сколько (cnt) у нас кандидатов из каждого города (city).
Отсортируйте результат по количеству в обратном порядке.

```
select
  ci.title city,
  count(can.id) cnt
from hh.candidate can
left join hh.city ci on can.city_id = ci.id
group by ci.title
order by cnt desc
```

city	cnt
Москва	16,622
Санкт-Петербург	4,937
Краснодар	1,066

Выводы по результатам отчета: более 2/3(около 28 тыс. из немногим более 44 тыс. соискателей) проживает в 10 крупнейших городах миллиониках, распределение между городами отражает размеры экономик данных городов.

Задание 3.2

Запрос: каких кандидатов из Москвы устроит «проектная работа».
Формат выборки: gender, age, desirable_occupation, city, employment_type.
Отсортируйте результат по id кандидата.

```
select
  can.gender gender,
  can.age age,
  can.desirable_occupation desirable_occupation,
  ci.title city,
  can.employment_type employment_type
from hh.candidate can
left join hh.city ci on can.city_id = ci.id
where ci.title like 'Москва' and can.employment_type like '%проектная работа%'
order by can.id
```

gender	age	desirable_occupation	city	employment_type
M	36	Веб-разработчик (HTML...	Москва	частичная занятость, проектная работа, полная занятость
M	31	Специалист	Москва	частичная занятость, проектная работа, полная занятость
F	42	pre-sale инженер, pre-sal...	Москва	частичная занятость, проектная работа, полная занятость
M	49	Дежурный администра...	Москва	частичная занятость, проектная работа, полная занятость

Выводы по результатам отчета: наиболее гибкая страта, соискатели из IT отрасли, в основном они и готовы работать на проектной работе, дополнительно накладывается ситуация высокого спроса, это позволяет быть более гибким при выборе задач.

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.

```
select
  can.gender gender,
  can.age age,
  can.desirable_occupation desirable_occupation,
  ci.title city,
  can.employment_type employment_type
from hh.candidate can
  left join hh.city ci on can.city_id = ci.id
where ci.title ilike 'Москва'
  and can.employment_type ilike '%проектная работа%'
  and can.desirable_occupation ilike any (array['%разработчик%', '%аналитик%', '%программист%']) -- использую конструкцию ilike any которая возвращает True/False при
order by ci.title, can.id
```

gender	age	desirable_occupation	city	employment_type
M	38	Веб-разработчик (HTML / CSS / JS / PHP / баз...	Москва	частичная занятость, проектная работа, полная занятость
M	22	Программист C++	Москва	проектная работа, частичная занятость
M	25	Frontend-разработчик	Москва	стажировка, волонтерство, частичная занятость, проектная работа

Выводы по результатам отчета: 27% соискателей среди тех, кто готов работать на проектной работе являются представители 3х наиболее популярных профессий разработчик, аналитик, программист.

Задание 3.4

Выбрать номера и города кандидатов, у которых занимаемая должность совпадает с желаемой.

Формат выборки: `id`, `city`.

Отсортируйте результат по городу и *id* кандидата.

```
select
  can.id id,
  ci.title city
from hh.candidate can
  left join hh.city ci on can.city_id = ci.id
where can.desirable_occupation =
  can.current_occupation
order by ci.title, can.id
```

id	city
2,009	Абакан
10,340	Абакан
14,449	Абакан

Выводы по результатам отчета: 5104 соискателя (более 11%) находятся в поиске работы не по причине того что они выросли из своей позиции, а скорее всего, для улучшения тех или иных условий труда.

Задание 3.5

Определите количество кандидатов пенсионного возраста.

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

```
select
  count(can.id)
from hh.candidate can
where can.age <> 100 -- очищаем выборку
от некорректных данных см. п. 2.1
  and ((can.gender = 'M' and can.age >= 65)
  or (can.gender = 'F' and can.age >= 60))
```

75

Выводы по результатам отчета: только 75 соискателей пенсионного возраста пользуются услугами HH.ru, наводит на мысль HH.ru принципиально не работает с соискателями пенсионного возраста, при текущей демографической ситуации, HH.ru рискует упустить развитие рынка труда синьоров(людей пенсионного возраста)

4. Анализ кандидатов для заказчиков

Задание 4.1

Выбор 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.

```
select
  can.gender gender,
  can.age age,
  can.desirable_occupation desirable_occupation,
  ci.title city,
  can.employment_type employment_type,
  tt.title timetable_type
from hh.candidate can
  left join hh.city ci on (ci.id = can.city_id)
  left join hh.candidate_timetable_type ctt on (can.id = ctt.candidate_id)
  left join hh.timetable_type tt on (ctt.timetable_id = tt.id)
where ci.title ilike any (array['Новосибирск', 'Омск', 'Томск', 'Тюмень'])
  and tt.title ilike 'вахтовый метод'
order by city, can.id
```

gender	age	desirable_occupation	city	employment_type	timetable_type
M	29	ИТ Инженер	Новосибирск	полная занятость	вахтовый метод
M	25	Заместитель начальника лаборатории	Новосибирск	проектная работа, стажировка, частичная занятость, полная занятость	вахтовый метод
M	30	Ведущий инженер, Специалист по защите информации	Новосибирск	частичная занятость, полная занятость	вахтовый метод

Выводы по результатам отчета: 11 соискателей готовы из обозначенных городов работать вахтовым методом. не наша целевая аудитория.

Задание 4.2

1 point possible (graded)

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.

```

with x_gen as(
select
    c.desirable_occupation
desirable_occupation,
    c.age age
from hh.candidate c
    left join hh.city ci on (ci.id = c.city_id)
where ci.title = 'Санкт-Петербург'
    and c.age between 16 and 21)

(select *
from x_gen
order by age
limit 10)

union all

(select
    'Total',
    count(age)
from x_gen)

```

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

Подход к решению задачи: предварительно сделали выборку (x_gen)с помощью конструкции with-as-select среди соискателей Санкт-Петербурга в возрасте от 16 до 21 включительно, после этого во второй части кода отсортировали по возрасту и взяли первые 10 позиций, при помощи конструкции union all добавили строку итогов по выборке (x_gen) .

Выводы из данных: подавляющее большинство соискателей в сделанной выборке, предпочитают развиваться в рамках IT отрасли.

Главными выводами в рамках этого отчета:

- необходимо принять одно из двух решений или концентрация на нашей целевой аудитории и минимизация обработки запросов не свойственных ей (соискатели пенсионного возраста(тенденции в демографии), вахтовики), или развитие данных сегментов для более полного и адекватного предложения рынку.
- аудитория от 16 до 21 года включительно выбирает IT профессии(более 90%)
- подавляющее большинство соискателей среди готовых работать на проектной основе являются представителями IT сектора (начиная от разработчиков и тестировщиков заканчивая руководителями проектов)
- предшествующие 2 пункта позволяют говорить о возможном росте повторных продаж наших сервисов, что ведет к необходимости разработки стратегии привлечения и удержания данной категории соискателей.

