

PROJECT-2. Подгрузка новых данных. Уточнение анализа

Предварительный анализ данных

Задание 2.1

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

```
select
    max(age) -- функция max для вычисления максимального значения
from
    SKILLFACTORY.HH.candidate
```

Результат запроса: **100**

Выводы: в результате мы получили значение в 100 лет, что маловероятно. Скорее всего данные были некорректно введены соискателем. Тем более, что остальные верхние значения колонки возраста находятся в пределах 70+ лет.

Задание 2.2

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

```
select
    min(age) -- функция min для вычисления минимального значения
from
    SKILLFACTORY.HH.candidate
```

Результат запроса: **14**

Выводы: в результате мы получили значение в 14 лет, что противоречит статье 63 ТК РФ, говорящей о том, что заключение трудового договора допускается с лицами достигшими 16 лет. Это вынуждает нас не рассматривать в анализе эту и другие не подходящие по минимальному возрасту анкеты соискателей.

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
    age,          --вывод столбца возраст
    count(age) --функция подсчёта значений столбца возраст
from
    SKILLFACTORY.HH.candidate
group by
    age -- группировка по возрасту
order by
    age desc -- сортировка по возрасту по убыванию
```

Результат запроса:

<u>age</u>	<u>count</u>
<u>100</u>	<u>1</u>
<u>77</u>	<u>1</u>
<u>76</u>	<u>1</u>
<u>73</u>	<u>4</u>
<u>72</u>	<u>3</u>

Выводы: этот запрос позволяет оценить количество кандидатов от возраста. Судя по итоговой таблице, основному количеству кандидатов от 22 до 38 лет.

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.

```
select
    count(*)    -- подсчёт всех значений и вывод на экран
from
    SKILLFACTORY.HH.candidate
where age <> 100 and age > 40 -- условия вывода значений столбца возраст
```

Результат запроса: **6263**

Выводы: мы нашли число людей среднего возраста и выше, обоего пола в нашей базе, что является 14% от общего количества кандидатов. Это подтверждает вывод из предыдущего задания, что основная масса кандидатов - молодежь.

Глобальный анализ показателей

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.

```
select
    title city,    -- столбец города
    count(*) cnt   -- подсчет значений
from
    SKILLFACTORY.HH.city c
```

join SKILLFACTORY.HH.candidate can on c.id = can.city_id -- соединение таблиц по ключу id
group by title -- группировка по городам
order by
cnt desc -- сортировка по городам по убыванию

Результат запроса:

<u>city</u>	<u>cnt</u>
Москва	16 622
Санкт-Петербург	4 937
Краснодар	1 066
Новосибирск	958
Казань	872

Выводы: ожидаемо результаты показали, что Москва с большим отрывом лидирует среди городов по наличию соискателей. Косвенно можно судить об экономическом благополучии городов в зависимости от количества соискателей, стремящихся найти там работу.

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.

```
select
  gender,
  age,
  desirable_occupation,
  title city,
  employment_type -- названия столбцов для вывода
from
  SKILLFACTORY.HH.city c
  join SKILLFACTORY.HH.candidate can on c.id = can.city_id -- соединение таблиц по ключу id
where employment_type like '%проектная работа%' and title = 'Москва' -- условия вывода
order by
  can.id -- сортировка по id кандидатов
```

Результат запроса:

<u>gender</u>	<u>age</u>	<u>desirable_occupation</u>	<u>city</u>	<u>employment_type</u>
<u>М</u>	<u>38</u>	<u>Веб-разработчик (HTML / CSS / JS / PHP / базы данных: фреймворки, дизайн.</u>	<u>Москва</u>	<u>частичная занятость, проектная работа, полная занятость</u>

		<u>интерфейсы, CMS)</u>		
<u>М</u>	<u>31</u>	<u>Специалист</u>	<u>Москва</u>	<u>частичная занятость, проектная работа, полная занятость</u>
<u>Е</u>	<u>42</u>	<u>pre-sale инженер, pre-sale менеджер</u>	<u>Москва</u>	<u>частичная занятость, проектная работа, полная занятость</u>
<u>М</u>	<u>49</u>	<u>Дежурный администратор</u>	<u>Москва</u>	<u>частичная занятость, проектная работа, полная занятость</u>
<u>М</u>	<u>29</u>	<u>Главный инженер проекта</u>	<u>Москва</u>	<u>частичная занятость, проектная работа, полная занятость</u>

Выводы: выборка показала, что в Москве на проектную работу согласно 2950 человек или примерно 18% от общего количества соискателей. Это может сказать о том, что данный вид занятости не самый популярный у соискателей.

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
  gender,
  age,
  desirable_occupation,
  title city,
  employment_type      -- названия столбцов для вывода
from
  SKILLFACTORY.HH.city c
  join SKILLFACTORY.HH.candidate can on c.id = can.city_id      -- соединение таблиц по ключу id
where employment_type like '%проектная работа%'
  and title = 'Москва'
  and (desirable_occupation like '%разработчик%'
  or desirable_occupation like '%аналитик%'
  or desirable_occupation like '%программист%'
  or desirable_occupation like '%Разработчик%'
  or desirable_occupation like '%Аналитик%'
  or desirable_occupation like '%Программист%')      -- условия вывода
order by
  can.id      -- сортировка по id кандидатов

```

Результат запроса:

gender	age	desirable_occupation	city	employment_type
<u>M</u>	<u>38</u>	<u>Веб-разработчик (HTML / CSS / JS / PHP / базы данных: фреймворки, дизайн, интерфейсы, CMS)</u>	<u>Москва</u>	<u>частичная занятость, проектная работа, полная занятость</u>
<u>M</u>	<u>22</u>	<u>Программист C++</u>	<u>Москва</u>	<u>проектная работа, частичная занятость</u>
<u>M</u>	<u>25</u>	<u>Frontend-разработчик</u>	<u>Москва</u>	<u>стажировка, волонтерство, частичная занятость, проектная работа, полная занятость</u>
<u>M</u>	<u>30</u>	<u>Программист</u>	<u>Москва</u>	<u>частичная занятость, проектная работа</u>
<u>M</u>	<u>35</u>	<u>Ruby / Rails разработчик</u>	<u>Москва</u>	<u>частичная занятость, проектная работа, полная занятость</u>

Выводы: после дополнительной фильтрации данных из предыдущего задания оставшееся количество кандидатов составило 778 человек, что составляет 4,7% от общего количества соискателей в Москве.

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.

```

select
  can.id id,
  title city  -- названия столбцов для вывода
from
  SKILLFACTORY.HH.city c
  join SKILLFACTORY.HH.candidate can on c.id = can.city_id  -- соединение таблиц по ключу id
where
  current_occupation = desirable_occupation  -- условия вывода
order by
  c.title,
  can.id  -- сортировка по городу и id кандидата

```

Результат запроса:

<u>id</u>	<u>city</u>
<u>2 009</u>	<u>Абакан</u>
<u>10 340</u>	<u>Абакан</u>

<u>14 449</u>	<u>Абакан</u>
<u>20 261</u>	<u>Абакан</u>
<u>13 705</u>	<u>Агрыз</u>

Выводы: в результате у 5104 или 11,4% соискателей полностью совпадают их занимаемая и желаемая должности. Однако стоит провести более детальную выборку, поскольку в данных о желаемой должности может попадаться перечень должностей. Что в текущей выборке не учитывалось, а учитывалось только 100% совпадение.

Задание 3.5

Определите количество кандидатов пенсионного возраста.

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

```
select
    count(*)    -- подсчёт всех значений и вывод на экран
from
    SKILLFACTORY.HH.candidate
where
    age <> 100
    and
    (gender = 'M' and age >= 65)
    or
    (gender = 'F' and age >= 60)  -- условия вывода с учетом некорректного возраста "100 лет"
```

Результат запроса: **75**

Выводы: количество соискателей пенсионного возраста всего 75 или 0,17% от общего количество кандидатов. Это может говорить об общей тенденции отрасли к более молодым кадрам (22-38 и менее лет) как мы могли видеть в одном из заданий ранее.

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.

```
select
    gender,
    age,
    desirable_occupation,
    c.title city,
    employment_type,
    tt.title timetable_type  -- названия столбцов для вывода
```

from

SKILLFACTORY.HH.city c -- соединение таблиц по ключам

join SKILLFACTORY.HH.candidate can on c.id = can.city_id

join SKILLFACTORY.HH.candidate_timetable_type ctt on can.id = ctt.candidate_id

join SKILLFACTORY.HH.timetable_type tt on ctt.timetable_id = tt.id

where

tt.title = 'вахтовый метод' and

(c.title = 'Новосибирск' or

c.title = 'Омск' or

c.title = 'Томск' or

c.title = 'Тюмень') -- условия вывода

order by

c.title,

can.id -- сортировка по городу и id кандидата

Результат запроса:

<u>gender</u>	<u>age</u>	<u>desirable occupation</u>	<u>city</u>	<u>employment type</u>	<u>timetable type</u>
<u>M</u>	<u>29</u>	<u>ИТ Инженер</u>	<u>Новосибирск</u>	<u>полная занятость</u>	<u>вахтовый метод</u>
<u>M</u>	<u>25</u>	<u>Заместитель начальника лаборатории</u>	<u>Новосибирск</u>	<u>проектная работа, стажировка, частичная занятость, полная занятость</u>	<u>вахтовый метод</u>
<u>M</u>	<u>30</u>	<u>Ведущий инженер. Специалист по защите информации.</u>	<u>Новосибирск</u>	<u>частичная занятость, полная занятость</u>	<u>вахтовый метод</u>
<u>M</u>	<u>23</u>	<u>Программист</u>	<u>Новосибирск</u>	<u>полная занятость</u>	<u>вахтовый метод</u>
<u>M</u>	<u>35</u>	<u>Инженер АСУТП, инженер-электроник</u>	<u>Омск</u>	<u>полная занятость</u>	<u>вахтовый метод</u>

Выводы: по данному запросу нашлось всего 11 кандидатов, что может сделать поиск подходящей кандидатуры проблематичной. Возможно придется расширять критерии поиска и пытаться подходящим кандидатам не указавшим вахтовый метод работы предлагать рассмотреть другой вариант.

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.

```
(select
  desirable_occupation,
  age                -- названия столбцов для вывода
from
  SKILLFACTORY.HH.city c
  join SKILLFACTORY.HH.candidate can on c.id = can.city_id -- соединение таблиц по ключам
where
  c.title = 'Санкт-Петербург' and
  age between 16 and 21      -- условия вывода
order by
  age                        -- сортировка по возрасту
limit 10)                  -- ограничение вывода

union all                  -- объединение таблиц
```

```
(select
  'Total',
  count(desirable_occupation) -- добавляем внизу строку 'Total'
from
  SKILLFACTORY.HH.city c
  join SKILLFACTORY.HH.candidate can on c.id = can.city_id
where
  c.title = 'Санкт-Петербург' and
  age between 16 and 21)
```

Результат запроса:

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

Выводы: данная выборка показывает количество соискателей из Санкт-Петербурга старшеклассников и студентов. Стоит отметить, что в желаемых попадают вакансии требующие большого релевантного опыта, что для студента может быть проблематично. Также остается открытым вопрос, как такие вакансии совместимы с учёбой. Возможно такие анкеты можно не учитывать в дальнейшем, либо наоборот, более детально проверить.

Общие выводы

Общие выводы таковы:

- по возрастным показателям рынок труда сосредоточен в основном от 20 до 40 лет. Возможно потому, что это либо выпускники вузов (20+), либо люди (30+) повысившие квалификацию (Middle, Senior, Lead) и ищущие работу согласно своим новым навыкам.
- людей возраста выше среднего мало среди соискателей. Пенсионеров - крайне мало. Возможно потому, что работодатели предпочитают более молодых кандидатов.
- выборка показала, что на работу вахтовым методом готово крайне мало кандидатов. Даже если расширить выборку до всей России, их будет всего 279. Возможно придется расширять критерии поиска и пытаться подходящим кандидатам не указавшим вахтовый метод работы предлагать рассмотреть другой вариант.
- полная занятость остается самым популярным типом занятости среди соискателей. Возможно так как ощущается как наиболее надежным.
- количество старшеклассников и студентов из Санкт-Петербурга ищущих работу всего 161, что на общем фоне города в 4937 смотрится крайне мало. Что возможно говорить о непопулярности данного направления деятельности среди студентов.