**PROJECT-2**

**Задание 2.1**

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

    SELECT

        MAX(c.age) max\_age /\*выбираем максимальное значение столбца с названием age, с алиасом max\_age\*/

    FROM

        hh.CANDIDATE c /\*из схемы hh и таблицы CANDIDATE c алиасом с\*/

|  |
| --- |
| max\_age |
| 100 |

**Максимальный возраст кандидата составляет 100 лет, это значение явно является выбросом.**

**Задание 2.2**

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

    SELECT

        MIN(c.age) min\_age /\*выбираем минимальное значение столбца с названием age, с алиасом min\_age\*/

    FROM

        hh.CANDIDATE c /\*из схемы hh и таблицы CANDIDATE c алиасом с\*/

|  |
| --- |
| min\_age |
| 14 |

**Минимальный возраст кандидата составляет 14 лет, однако, выбросом данное значение не является т.к. использование подростками сервиса HH для поиска подработки или временной работы вполне вероятно.**

**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.

    SELECT

        c.age,

        COUNT (c.id) cnt /\*выбираем значение столбца age и количество значений столбца id с алиасом cnt\*/

    FROM

        hh.CANDIDATE c /\*из схемы hh и таблицы CANDIDATE c алиасом с\*/

    GROUP BY c.age /\*группировка по столбцу age таблицы c\*/

    ORDER BY 1 DESC /\*сортировка по age таблицы c по убыванию \*/

|  |  |
| --- | --- |
| age | cnt |
| 100 | 1 |
| 77 | 1 |
| 76 | 1 |
| 73 | 4 |
| 72 | 3 |

**Мы видим, что предположение о том, что 100-летний возраст кандидата является выбросом подтверждается (от следующего значения его отделяет 23 года), также подтверждаются и выводы предыдущего задания о 14-ти летнем кандидате, так как начиная с этого значения начинается плавный рост количества соискателей в зависимости от возраста (без разрывов), достигая максимума в возрасте 30 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 (c.id) cnt /\*выбираем количество значений столбца id cnt\*/

    FROM

        hh.CANDIDATE c /\*из схемы hh и таблицы CANDIDATE c алиасом с\*/

    WHERE

        c.age > 40

        AND c.age !=100 /\*age таблицы c имеет значение более 40 и не равняется 100\*/

|  |
| --- |
| cnt |
| 6 263 |

**Таким образом количество кандидатов в возрасте** **старше среднего возраста занятых в экономике России составляет 6263 чел., что составляет 14% от общего количества соискателей в исследуемой базе HH (всего 44744 чел.). Вероятно, этот результат можно объяснить следующими причинами: люди старше среднего возраста менее склонны к смене места работы, кроме того, они чаще используют для этого знакомства и связи, а не цифровые сервисы кадровых агентств.**

**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

        ct.title city,

        COUNT(cd.id) cnt /\*выбираем значение столбца title с алиасом city и количество значений столбца id с алиасом cnt\*/

    FROM

        hh.CANDIDATE cd /\*из схемы hh и таблицы CANDIDATE c алиасом сd\*/

        JOIN hh.city ct ON cd.city\_id = ct.id

    GROUP BY city /\*группировка по столбцу city таблицы ct\*/

    ORDER BY 2 DESC /\*сортировка по cnt таблицы cd по убыванию\*/

|  |  |
| --- | --- |
| city | cnt |
| Москва | 16 622 |
| Санкт-Петербург | 4 937 |
| Краснодар | 1 066 |
| Новосибирск | 958 |
| Казань | 872 |
| Екатеринбург | 734 |
| Самара | 703 |
| Ростов-на-Дону | 607 |
| Нижний Новгород | 598 |

**Результат распределения количества кандидатов по городам не является неожиданным и в целом соответствует распределению численности населения данных городов. Исключением является Краснодар, численность населения которого ниже, чем в Новосибирске, Екатеринбурге, Казани, Самаре. Возможно, это связано с более низким уровнем промышленного производства в Краснодаре по сравнению с данными городами-миллионниками.**

**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

        cd.gender,

        cd.age,

        cd.desirable\_occupation,

        ct.title city,

        cd.employment\_type /\*выбираем значение столбцов gender, age, desirable\_occupation, title с алиасом city, employment\_type\*/

    FROM hh.CANDIDATE cd /\*из схемы hh и таблицы CANDIDATE c алиасом сd\*/

        JOIN hh.city ct ON cd.city\_id = ct.id /\*оператор соединения таблиц; таблица city; условие: id таблицы city равен city\_id таблицы CANDIDATE\*/

    WHERE (ct.title = 'Москва') /\* title таблицы ct имеет значение 'Москва'\*/

AND (cd.employment\_type LIKE '%проектная работа%') /\* employment\_type таблицы cd включает в себя значение ' проектная работа '\*/

    ORDER BY cd.id /\*сортировка по id таблицы cd\*/

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| gender | age | desirable\_occupation | city | employment\_type |
| M | 38 | Веб-разработчик (HTML / CSS / JS / PHP / базы данных; фреймворки, дизайн, интерфейсы, CMS) | Москва | частичная занятость, проектная работа, полная занятость |
| M | 31 | Специалист | Москва | частичная занятость, проектная работа, полная занятость |
| F | 42 | pre-sale инженер, pre-sale менеджер | Москва | частичная занятость, проектная работа, полная занятость |
| M | 49 | Дежурный администратор | Москва | частичная занятость, проектная работа, полная занятость |
| M | 29 | Главный инженер проекта | Москва | частичная занятость, проектная работа, полная занятость |
| M | 22 | Программист С++ | Москва | проектная работа, частичная занятость |
| F | 29 | Технический специалист | Москва | частичная занятость, проектная работа, полная занятость |
| M | 32 | IT Operations Coordinator | Москва | частичная занятость, проектная работа, полная занятость |
| M | 23 | Инженер-связист,системный администратор | Москва | частичная занятость, проектная работа, полная занятость |
| M | 31 | Менеджер | Москва | частичная занятость, проектная работа, полная занятость |
| M | 44 | Директор по развитию бизнеса / Директор по продажам | Москва | проектная работа, полная занятость |

**Судя по результатам запроса, соискатели из Москвы, которых устроит «проектная работа» в основном связаны с IT отраслью. Такой результат логичен, т.к. применение данного вида занятости для других направлений деятельности (например производство) ограничено по естественным причинам.**

**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.

    SELECT

        cd.gender,

        cd.age,

        cd.desirable\_occupation,

        ct.title city,

        cd.employment\_type /\*выбираем значение столбцов gender, age, desirable\_occupation, title с алиасом city, employment\_type\*/

    FROM hh.CANDIDATE cd /\*из схемы hh и таблицы CANDIDATE c алиасом сd\*/

        JOIN hh.city ct ON cd.city\_id = ct.id /\*оператор соединения таблиц; таблица city; условие: id таблицы city равен city\_id таблицы CANDIDATE\*/

    WHERE (ct.title = 'Москва')

AND (cd.employment\_type LIKE '%проектная работа%')

AND (lower(desirable\_occupation) LIKE '%разработчик%' OR desirable\_occupation LIKE '%аналитик%' OR desirable\_occupation LIKE '%программист%') /\* title таблицы ct имеет значение 'Москва', employment\_type таблицы cd включает в себя 'проектная работа', desirable\_occupation таблицы cd содержит 'разработчик' или 'аналитик' или 'программист'\*/

    ORDER BY cd.id /\*сортировка по id таблицы cd\*/

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| gender | age | desirable\_occupation | city | employment\_type |
| M | 38 | Веб-разработчик (HTML / CSS / JS / PHP / базы данных; фреймворки, дизайн, интерфейсы, CMS) | Москва | частичная занятость, проектная работа, полная занятость |
| M | 25 | Frontend-разработчик | Москва | стажировка, волонтерство, частичная занятость, проектная работа, полная занятость |
| M | 35 | Ruby / Rails разработчик | Москва | частичная занятость, проектная работа, полная занятость |
| M | 36 | Программист-разработчик | Москва | частичная занятость, проектная работа, полная занятость |
| M | 38 | Инженер, программист C/C++, разработчик ПО | Москва | частичная занятость, проектная работа, полная занятость |
| F | 54 | Ведущий инженер программист | Москва | частичная занятость, проектная работа, полная занятость |
| M | 41 | Ведущий разработчик | Москва | частичная занятость, проектная работа, полная занятость |
| M | 40 | Программист-разработчик | Москва | частичная занятость, проектная работа, полная занятость |
| M | 22 | IOS разработчик | Москва | проектная работа, частичная занятость, полная занятость |

**Результаты показывают, что самыми популярными среди IT-профессий являются разработчики web-приложений, что не удивительно т.к. эта работа зачатую имеет именно «проектный» характер.**

**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

        cd.id,

        ct.title city /\*выбираем значение столбца id, значение столбца title с алиасом city\*/

    FROM hh.CANDIDATE cd /\*из схемы hh и таблицы CANDIDATE c алиасом сd\*/

        JOIN hh.city ct ON cd.city\_id = ct.id /\*оператор соединения таблиц; таблица city; условие: id таблицы city равен city\_id таблицы CANDIDATE\*/

    WHERE cd.current\_occupation = cd.desirable\_occupation /\* current\_occupation таблицы cd совпадает со значением desirable\_occupation таблицы cd \*/

    ORDER BY 2,1 /\*сортировка по id таблицы cd и title таблицы ct\*/

|  |  |
| --- | --- |
| id | city |
| 2 009 | Абакан |
| 10 340 | Абакан |
| 14 449 | Абакан |
| 20 261 | Абакан |
| 13 705 | Агрыз |
| 967 | Адлер |
| 4 276 | Адлер |
| 26 878 | Адлер |
| 27 717 | Адлер |

**Сам по себе данный запрос, мне кажется, мало информативен. Полезной является, пожалуй, информация о количестве таких соискателей (их 5104 чел. – 11% от общего количества соискателей в базе). Возможно такие соискатели не ищут карьерного роста и их целью является рост заработной платы, либо лучшие условия труда.**

**Задание 3.5**

Определите количество кандидатов пенсионного возраста.

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

    SELECT

        COUNT(c.id) /\*выбираем количество значений столбца id cnt\*/

    FROM

        hh.CANDIDATE c /\*из схемы hh и таблицы CANDIDATE c алиасом с\*/

    WHERE (c.gender = 'F' AND c.age>=60) OR (c.gender = 'M' AND c.age>=65) AND c.age !=100 /\* gender имеет значение ‘F’ и age таблицы c имеет значение более 60 или gender имеет значение ‘M’ и age таблицы c имеет значение более 65\*/

|  |
| --- |
| count |
| 75 |

**Количество кандидатов пенсионного возраста не велико – 75 человек. Вероятно, это связано с тем, что люди пенсионного возраста для поиска работы не прибегают к услугам цифровых сервисов, либо работающие пенсионеры после выхода на пенсию продолжают работу на текущем месте работы и предпочитают его не менять.**

**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

        cd.gender,

        cd.age,

        cd.desirable\_occupation,

        ct.title city,

        cd.employment\_type,

        tt.title timetable\_type /\*выбираем значение столбцов gender, age, desirable\_occupation, title с алиасом city, employment\_type, title с алиасом timetable\_type\*/

    FROM hh.CANDIDATE cd /\*из схемы hh и таблицы CANDIDATE c алиасом сd\*/

        JOIN hh.city ct ON cd.city\_id = ct.id /\*оператор соединения таблиц; таблица city; условие: id таблицы city равен city\_id таблицы CANDIDATE\*/

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

/\*оператор соединения таблиц; таблица candidate\_timetable\_type; условие: candidate\_id таблицы candidate\_timetable\_type равен id таблицы CANDIDATE\*/

        JOIN hh.timetable\_type tt ON ctt.timetable\_id = tt.id /\*оператор соединения таблиц; таблица timetable\_type; условие: id таблицы timetable\_type равен timetable\_id таблицы candidate\_timetable\_type \*/

    WHERE (ct.title IN ('Новосибирск', 'Омск', 'Томск', 'Тюмень')) AND (tt.title ='вахтовый метод') /\* title таблицы ct равен одному из значений: 'Новосибирск', 'Омск', 'Томск', 'Тюмень'; и title таблицы tt имеет значение 'вахтовый метод' \*/

    ORDER BY 4, cd.id /\*сортировка по title таблицы ct и id таблицы cd\*/

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 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 |
| M | 23 | Программист | Новосибирск | полная занятость | вахтовый метод |
| M | 35 | Инженер АСУТП, инженер-электроник | Омск | полная занятость | вахтовый метод |
| M | 25 | Тестировщик ПО | Омск | стажировка, полная занятость | вахтовый метод |
| M | 26 | Специалист технической поддержки | Томск | частичная занятость, полная занятость | вахтовый метод |
| M | 30 | Менеджер проектов | Томск | проектная работа, частичная занятость, полная занятость | вахтовый метод |
| M | 42 | Инженер | Томск | проектная работа, частичная занятость, полная занятость | вахтовый метод |
| M | 31 | Инженер связи | Тюмень | полная занятость | вахтовый метод |
| M | 31 | Инженер АСУ ТП, АСУ, Мастер КИП, Программист АСУ | Тюмень | полная занятость | вахтовый метод |

**В результатах выборки представлено 11 кандидатов, что свидетельствует о том, кандидаты на должности IT направления и инженерные работники из данных городов не склонны к работе вахтовым методом, соответственно выбор у добывающей компании не большой.**

**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

        cd.desirable\_occupation,

        cd.age /\*выбираем колонки desirable\_occupation, age \*/

    FROM hh.CANDIDATE cd /\*из схемы hh и таблицы CANDIDATE c алиасом сd\*/

        JOIN hh.city ct ON cd.city\_id = ct.id  /\*оператор соединения таблиц; таблица city; условие: id таблицы city равен city\_id таблицы CANDIDATE\*/

    WHERE (ct.title = 'Санкт-Петербург') AND (cd.age BETWEEN 16 AND 21) /\* title таблицы ct имеет значение 'Санкт-Петербург' и age таблицы cd находится в интервале от 16 до 21\*/

    ORDER BY 2 /\*сортировка по age таблицы cd\*/

    LIMIT 10) /\*ограничение выборки первыми 10 значениями\*/

    UNION ALL /\*оператор присоединения\*/

    (SELECT

        'Total', COUNT(cd.id) /\*создаем колонку 'Total', выбираем количество значений колонки id \*/

    FROM

        hh.CANDIDATE cd /\*из схемы hh и таблицы CANDIDATE c алиасом сd\*/

        JOIN hh.city ct ON cd.city\_id = ct.id /\*оператор соединения таблиц; таблица city; условие: id таблицы city равен city\_id таблицы CANDIDATE\*/

    WHERE (ct.title = 'Санкт-Петербург') AND (cd.age BETWEEN 16 AND 21)) /\* title таблицы ct имеет значение 'Санкт-Петербург' и age таблицы cd находится в интервале от 16 до 21\*/

|  |  |
| --- | --- |
| 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 |

**Общее количество кандидатов, удовлетворяющих заданным критериям, составило 161 чел. При этом претенденты в возрасте 18 лет (скорее всего не имеющие опыта) претендуют на вакансии, к которым обычно предъявляются довольно высокие требования (вплоть до руководителя web-разработки). Возможно их требования являются завышенными.**

Общие выводы:

1. Судя по результатам запросов, основная доля соискателей в данной базе связана с направлением IT, причем значительная доля кандидатов в базе (около 37%) из г. Москва. Не велика доля соискателей в возрасте старше среднего возраста занятых в экономике России (согласно Росстата) – 14% от общего количества соискателей в базе., что в т.ч. связано с представленным диапазоном профессий (IT).
2. О характеристиках рынка труда в различных городах можно судить только по результатам запроса из задания 3.1, результаты которого указывают на прямую пропорциональность количества соискателей в разрезе городов численности населения данных городов (некоторым исключением является г. Краснодар).
3. На мой взгляд, было бы интересно провести анализ доли кандидатов готовых к переезду в общем количестве соискателей из данного города.