Колобков Александр

Project-2

## **Предварительный анализ данных**

### **Задание 2.1**

select

max(age) --*используем агрегирующую функцию max(), чтобы определить максимальный возраст кандидата*

from hh.candidate

**Результат: 100**

**Вывод:** Максимальный возраст 100 вероятно указан ошибочно, либо это уникальный кандидат.

### **Задание 2.2**

select

min(age) --*используем агрегирующую функцию min(), чтобы определить минимальный возраст кандидата*

from hh.candidate

**Результат: 14**

**Вывод:** вероятно кандидаты этого возраста школьники, ищущие подработку, либо шутники.

### **Задание 2.3**

select

age, --*колонка возраста*

count(id) cnt --*подсчет кандидатов и формирование колонки cnt*

from hh.candidate

group by age --*группирование по колонке возраста*

order by 1 DESC --*формирование колонки в убывающем порядке*

**Результат: таблица 2.3**

**Таблица 2.3 (начало таблицы)**

| **age** | **cnt** |
| --- | --- |
| 100 | 1 |
| 77 | 1 |
| 76 | 1 |
| 73 | 4 |
| 72 | 3 |
| 71 | 4 |
| 70 | 3 |
| 69 | 4 |
| 68 | 4 |
| 67 | 9 |
| 66 | 11 |
| 65 | 20 |
| 64 | 17 |
| 63 | 29 |
| 62 | 22 |

**Вывод:** наибольшее количество кандидатов находится в возрасте от 22 до 40 лет

### **Задание 2.4**

select

count(id) --*подсчет с помощью агрегирующей функции количества кандидатов*

from hh.candidate

where age > 40 and age < 100 --*условия фильтрации для подсчет кандидатов*

**Результат: 6263**

**Вывод:** данное количество кандидатов составляет примерно 15% от общего количества соискателей, поэтому действительно можно сказать, что средний возраст соискателей составляет 39,7 лет.

## **3. Глобальный анализ показателей**

### **Задание 3.1**

select

a.title city, --*колонка названия города*

count(b.city\_id) cnt --*подсчет кандидатов с помощью агрегирующей функции*

from hh.city a

left join hh.candidate b on a.id = b.city\_id --*объединение двух таблиц по id города*

group by 1 --*группирование итоговой таблицы по названию города*

order by 2 desc --*формирование колонки в убывающем порядке*

**Результат: таблица 3.1**

**Таблица 3.1 (начало таблицы)**

| **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 |
| Уфа | 565 |
| Алматы | 556 |
| Воронеж | 538 |
| Пермь | 445 |
| Красноярск | 407 |
| Тюмень | 355 |
| Минск | 353 |
| Челябинск | 330 |
| Омск | 303 |
| Томск | 285 |
| Ярославль | 282 |
| Волгоград | 272 |
| Саратов | 260 |
| Калининград | 219 |
| Иркутск | 213 |
| Владивосток | 210 |

**Вывод:** наибольшее количество кандидатов сосредоточено в Москве и Санкт-Петербурге, что является закономерностей, так как это самые крупные города России по количеству жителей.

### **Задание 3.2**

select

a.gender, --*колонка пола соискателя*

a.age, --*колонка возраста*

a.desirable\_occupation, --*колонка желаемой должности*

b.title city, --*колонка названия города*

a.employment\_type --*колонка типа занятости*

from hh.city b

left join hh.candidate a on a.city\_id = b.id --*объединение двух таблиц по id города*

where b.title = 'Москва' and employment\_type like '%проектная работа%' --*условия фильтрации для поиска кандидатов в Москве, готовых заниматься проектной работой*

group by 4, a.id --*группирование итоговой таблицы по названию города*

order by a.id asc --*сортировка выборки по id кандидатов*

**Результат: таблица 3.2**

**Таблица 3.2 (начало таблицы)**

| **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 |

**Вывод:** согласно данной выборки проектную работу ищут преимущественно на должность связанную с IT, программным обеспечением, и руководством.

### **Задание 3.3**

select

a.gender, --*колонка пола соискателя*

a.age, --*колонка возраста*

a.desirable\_occupation, --*колонка желаемой должности*

b.title city, --*колонка названия города*

a.employment\_type --*колонка типа занятости*

from hh.city b

left join hh.candidate a on a.city\_id = b.id --*объединение двух таблиц по id города*

where b.title = 'Москва' --*условия фильтрации - город Москва*

and a.employment\_type like '%проектная работа%' --*условия фильтрации - проектная работа*

and a.desirable\_occupation ilike any(array['%разработчик%', '%аналитик%','%программист%']) --*условия фильтрации - названия желаемой должности*

group by 4, a.id --*группирование итоговой таблицы по названию города*

order by a.id asc --*сортировка выборки по id кандидатов в прямом порядке*

**Результат: таблица 3.3**

**Таблица 3.3 (начало таблицы)**

| **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 | 22 | Программист С++ | Москва | проектная работа, частичная занятость |
| M | 25 | Frontend-разработчик | Москва | стажировка, волонтерство, частичная занятость, проектная работа, полная занятость |
| M | 30 | Программист | Москва | частичная занятость, проектная работа |

**Вывод:**

### **Задание 3.4**

select

a.id, --*колонка id соискателя*

b.title city --*колонка названия города*

from hh.candidate a

join hh.city b on a.city\_id = b.id --*объединение двух таблиц по id города*

where a.desirable\_occupation = a.current\_occupation --*условия фильтрации - желаемая должность совпадает с действительной*

order by 2,1 --*сортировка выборки по городу и id кандидатов*

**Результат: таблица 3.4**

**Таблица 3.4 (начало таблицы)**

| **id** | **city** |
| --- | --- |
| 2 009 | Абакан |
| 10 340 | Абакан |
| 14 449 | Абакан |
| 20 261 | Абакан |
| 13 705 | Агрыз |
| 967 | Адлер |

### **Вывод:** очень большое количество кандидатов уже занимают ту, должность которую ищут, видимо причиной поиска нового места работы является повышение ЗП.

### **Задание 3.5**

select

gender, --*колонка пола соискателя*

count(\*) --*подсчет кандидатов разного пола с помощью агрегирующей функции*

from hh.candidate

where (gender = 'M' and age >= 65 and age < 100) --*условия фильтрации - для мужского пола возраст указывается с учетом пенсионного возраста 65 лет и отклонением - 100 лет*

or (gender = 'F' and age >=60) -*условия фильтрации - для женского пола возраст указывается с учетом пенсионного возраста 60 лет*

group by 1

**Результат: 75**

**gender count**

F 18

M 57

**Вывод:** общее количество кандидатов пенсионного возраста невелико и составляет примерно 0,002%. Также количество соискателей женского пола пенсионного возраста в 2,5 раза меньше, соискателей мужского возраста пенсионного возраста. Вероятно данная группа соискателей слабо знакома с площадкой HH или не заинтересованы в смене места работы, или не заинтересованы в поисках.

### **Задание 4.1**

select

a.gender, --*колонка пола соискателя*

a.age, --*колонка возраста*

a.desirable\_occupation, --*колонка желаемой должности*

b.title city, --*колонка названия города*

a.employment\_type, --*колонка типа занятости*

d.title timetable\_type --*колонка варианта рабочего графика*

from hh.candidate\_timetable\_type e

join hh.candidate a on a.id = e.candidate\_id --*объединение двух таблиц по id*

join hh.city b on a.city\_id = b.id --*объединение двух таблиц по id города*

join hh.timetable\_type d on e.timetable\_id = d.id --*объединение двух таблиц по id города*

where (b.title in ('Новосибирск','Омск', 'Томск', 'Тюмень')) --*условия фильтрации по городам*

and d.title ='вахтовый метод' --*условия фильтрации по графику работы*

order by b.title, a.id --*сортировка выборки по городу и id кандидатов*

**Результат: таблица 4.1**

**Таблица 4.1**

| **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 |

**Вывод:** среди перечисленных город наблюдается незначительное количество специалистов, готовых работать вахтовым методом, вероятно большая часть соискателей ориентируются на поиск работы в родном городе.

### **Задание 4.2**

with SPb as --*формирование таблицы для дальнейшего запроса*

(select

a.desirable\_occupation desirable\_occupation, --*колонка желаемой должности*

a.age age --*колонка возраста кандидата*

from hh.candidate a

join hh.city b on a.city\_id = b.id --*объединение двух таблиц по id города*

where lower(b.title) = 'санкт-петербург' --*условия фильтрации - город Санкт-Петербург*

and age between 16 and 21) --*условия фильтрации по возрасту*

(select --*запрос к ранее созданной таблице SPb*

desirable\_occupation, --*колонка желаемой должности*

age --*колонка возраста кандидата*

from SPb

order by 2 --*сортировка по колонке возраст*

limit 10) --*ограничение вывода 10 строк от всей таблицы*

union all --*объединение запроса из 10 строк по желаемым должностям и строки с общим количеством*

select

'Total', count(\*) --*строка подсчета общего количества кандидатов*

from SPb

**Результат: таблица 4.2**

**Таблица 4.2**

| **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** |

**Вывод:** в Санкт-Петербурге первую десятку по возрасту соискателей входят должности, относящиеся к IT. В целом среди кандидатов данных возрастов преимущественно преимущественно ищут должности в сфере IT, незначительно количество начинающие специалисты (видимо студенты подрабатывают во время учебы), а также инженеры и маркетологи.

**Общий вывод:** проведена общая оценка рынка труда, наибольшее количество находится в Москве, а также других мегаполисах по убыванию, что логично, из-за большего количества жителей и действующих предприятий.

Основная группа соискателей находится в возрасте от 22 до 40 лет. В крупных городах у молодых специалистов часто желаемая должность связана с IT сферой.

Вахтовый метод среди соискателей в крупных городах не актуален, основные соискатели рабочих мест с вахтовым методом находятся в регионах, из-за меньшего количества вакансий по желаемым должностям в регионе проживания.

Как указано выше, пенсионеры проявляют крайне малую активность, вероятно по нескольким основным причинам: 1) не умеют (либо плохо знакомы) пользоваться платформой HH; 2) не стремятся менять текущее рабочее место.

Также одна из закономерностей данной выборки, большая часть соискателей ищет новые места работы с такой же должностью, вероятно это связано с возможностью увеличить ЗП или поиском более комфортных условий труда.