# PROJECT-2. Подгрузка новых данных. Уточнение анализа.

## 2. Предварительный анализ данных

## Задание 2.1

Код:

select

max(age) /\*находим максимальный возраст\*/

from

hh.candidate

Результат: 100

Вывод:

Максимальный возраст кандидата 100 лет можно считать аномалией.

## Задание 2.2

Код:

select

min(age) /\*находим минимальный возраст\*/

from

hh.candidate

Результат: 14

Вывод:

К минимальному возрасту кандидата в 14 лет как же следует отнестись с осторожностью, т.к. это возможно это искаженные данные.

## Задание 2.3

Код:

select

age, /\*возраст\*/

count(\*) cnt /\*количество человек этого возраста\*/

from

hh.candidate

group by age /\*группируем по полю age\*/

order by age desc /\*copmupуем по полю age, по убыванию\*/

## Результат:

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

#### Вывод:

Исходя из результат, можно утверждать, что в данных присутствуют аномалии, требующие дальнейшей "очистки".

## Задание 2.4

Код:

select

count(\*) cnt /\*количество кандидатов\*/

from

hh.candidate

where age between 41 and 99 /\*возраст в диапазоне от 40 до 99\*/

Результат: 6263

#### Вывод:

Количество кандидатов, которые старше среднего возраста, составляет примерно 14% от общего количества соискателей.

## 3. Глобальный анализ показателей

## Задание 3.1

Код:

select

title city, /\*название города\*/

count(\*) cnt /\*количество кандидатов из каждого города\*/

from

hh.candidate ca

join hh.city ci ON ci.id = ca.city\_id /\*присоединяем таблицу city\*/

group by title /\*группируем по названию города\*/

order by 2 desc /\*coртируем по количеству кандидатов, по убыванию\*/

## Результат:



## Вывод:

Наибольшее количество кандидатов в городах Москва и Санкт-Петербург. Это обусловлено более высоким уровнем зарплат относительно других городов России.

## Задание 3.2

Код:

```
select
ca.gender, /*пол*/
ca.age, /*возраст*/
ca.desirable_occupation, /*желаемая профессия*/
ci.title, /*город*/
ca.employment_type /*вид занятости*/
from
hh.candidate ca
join hh.city ci ON ci.id = ca.city_id /*присоединяем таблицу city*/
```

where ci.title = 'Mocква' and ca.employment\_type like '%проектная работа%' /\*условия выборки\*/ order by ca.id /\*copmupyeм по id кандидата, по возрастанию\*/

#### Результат:

gend	v age desirable_occupation v	title v employment_type v
М	38 Веб-разработчик (HTML/CSS/JS/PHP/базы данных; фреймворки, дизайн, и	нтер Москва частичная занятость, проектная работа, полная занятость
М	31 Специалист	Москва частичная занятость, проектная работа, полная занятость
F	42 pre-sale инженер, pre-sale менеджер	Москва частичная занятость, проектная работа, полная занятость
М	49 Дежурный администратор	Москва частичная занятость, проектная работа, полная занятость

#### Вывод:

Исходя из полученных данных можно утверждать, что средний возраст кандидатов, которых устроит «проектная работа», 33 года, из них количество мужчин(2479) преобладает над женщинами(471) в 5.2 раза. Желаемая профессия кандидатов в основном из ІТ сферы.

## Задание 3.3

Код:

```
select
       ca.gender, /*пол*/
       ca.age, /*eospacm*/
       ca.desirable_occupation, /*желаемая профессия*/
       ci.title, /*zopo∂*/
       ca.employment_type /*вид занятости*/
from
       hh.candidate ca
      join hh.city ci ON ci.id = ca.city_id /*присоединяем таблицу city*/
where
       ci.title = 'Москва' and
       ca.employment_type like '%проектная работа%' and
       (ca.desirable_occupation ilike '%разработчик%' or
       ca.desirable occupation ilike '%аналитик%' or
       ca.desirable_occupation ilike '%программист%') /*условия, по которым выполняем
отбор данных*/
order by ca.id /*copmupyeм по id кандидата, по возрастанию*/
```

## Результат:

gender v	desirable_occupation ~	title v employment_type	
М	Зеб-разработчик (HTML/CSS/JS/PHP/базы да	анных; фреймворки, дизайн, ин Москва частичная заня	тость, проектная работа, полная занятость
М	Программист C++	Москва проектная рабо	ота, частичная занятость
М	rontend- <b>разработчик</b>	Москва стажировка, вод	лонтерство, частичная занятость, проектная раб
М	<b>Трограммист</b>	Москва частичная заня	тость, проектная работа

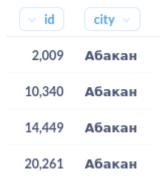
#### Вывод:

Исходя из результата можно утверждать, что наиболее популярные IT профессии это - Программист(54), Программист-разработчик(30), Инженер-программист(29), Аналитик(29). Средний возраст кандидата - 33 года, большинство из них мужчины(687), женщины(91).

## Задание 3.4

```
Код:
```

## Результат:



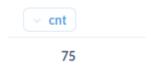
## Вывод:

Наибольшее количество кандидатов у которых занимаемая должность совпадает с желаемой это Москва(1855) и Санкт-Петербург(527).

## Задание 3.5

#### Код:

## Результат:



#### Вывод:

Количество кандидатов пенсионного возраста составляет около 0.2% от общего количества соискателей.

## 4. Анализ кандидатов для заказчиков

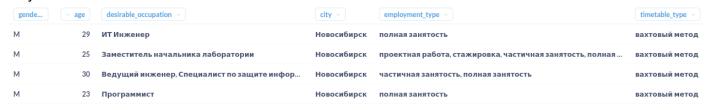
количества кандидатов пенсионного возраста\*/

## Задание 4.1

```
Код:
```

```
select
       ca.gender, /*пол*/
       ca.age, /*eospacm*/
       ca.desirable occupation, /*желаемая профессия*/
       ci.title city, /*zopod*/
       ca.employment_type, /*вид занятости*/
       tt.title timetable_type /*вариант рабочего графика*/
from
       hh.candidate ca
       join hh.city ci on ci.id = ca.city_id /*присоединяем таблицу city*/
      join hh.candidate_timetable_type ctt on ctt.candidate_id = ca.id /*присоединяем таблицу
candidate timetable type*/
       join hh.timetable_type tt on tt.id = ctt.timetable_id /*присоединяем таблицу
timetable type*/
where
       (ci.title = 'Новосибирск' or
       ci.title = 'Омск' or
       ci.title = 'Томск' or
       ci.title = 'Тюмень') and
       tt.title ilike '%вахт%' /*условия отбора кандидатов из Новосибирска, Омска, Томска и
Тюмени, которые готовы работать вахтовым методом*/
order by city, ca.id /*copтируем по городу и по id, по возрастанию*/
```

#### Результат:



#### Вывод:

Для добывающей компании подобрали кандидатов из Новосибирска, Омска, Томска и Тюмени, которые готовы работать вахтовым методом.

## Задание 4.2

#### Код:

union all /\*объединяем с данными\*/
(select
'Total', count(x.\*) /\*количество\*/
from x) /\*выборку производим из исходных данных\*/

## Результат:



#### Вывод:

Для заказчиков из Санкт-Петербурга собрали список из 10 желаемых профессий кандидатов от 16 до 21 года с указанием их возраста и с их общим количеством.

## Вывод.

Исходя из проанализированных данных можно утверждать, то:

- 1. Наибольшее количество кандидатов представлено в городах Москва и Санкт-Петербург. Это обусловлено более высоким уровнем зарплат относительно других городов России.
- 2. Средний возраст кандидата 32 года, из них количество мужчин(36,190) преобладает над женщинами(8,532) в 4.2 раза.
- 3. Из всех представленных кандидатов присутствует и небольшая группа пенсионного возраста, 0.2% от общего числа кандидатов.
- 4. Также предоставили выборку данных для заказчиков из Санкт-Петербурга и добывающей компании.