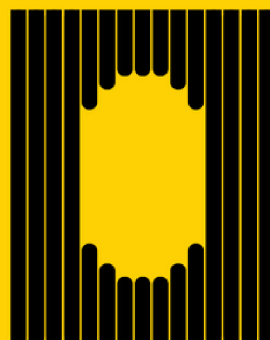




PANDAS HSE  
PYTHON  
ANALYSIS  
DANO 2023  
TINKOFF  
BEST PLOTS  
HYPOTHESIS  
TEAM  
DATA CORRELATION  
STATISTICS  
REGRESSION



D A  
N O

ОТТЕНКОВ АНАЛИТИКИ



# АНАЛИЗ СОВРЕМЕННОГО РОССИЙСКОГО РЫНКА ТРУДА

NEW

БАЗА ДАННЫХ HH.RU



# Этапы исследования

Извлечение данных и их обработка

Формирование гипотезы

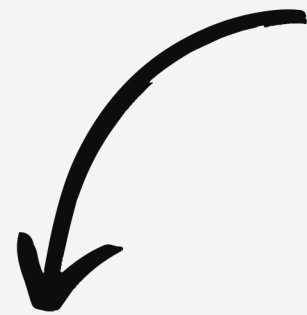
Разработка механизма

Матмодель / визуализации

Интерпретация результатов

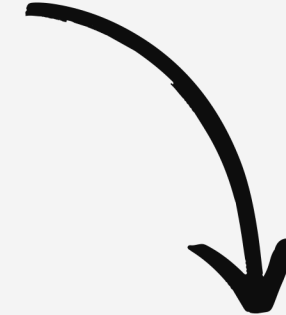
Подведение итогов

# Анализ структуры **данных**



## Количественные данные

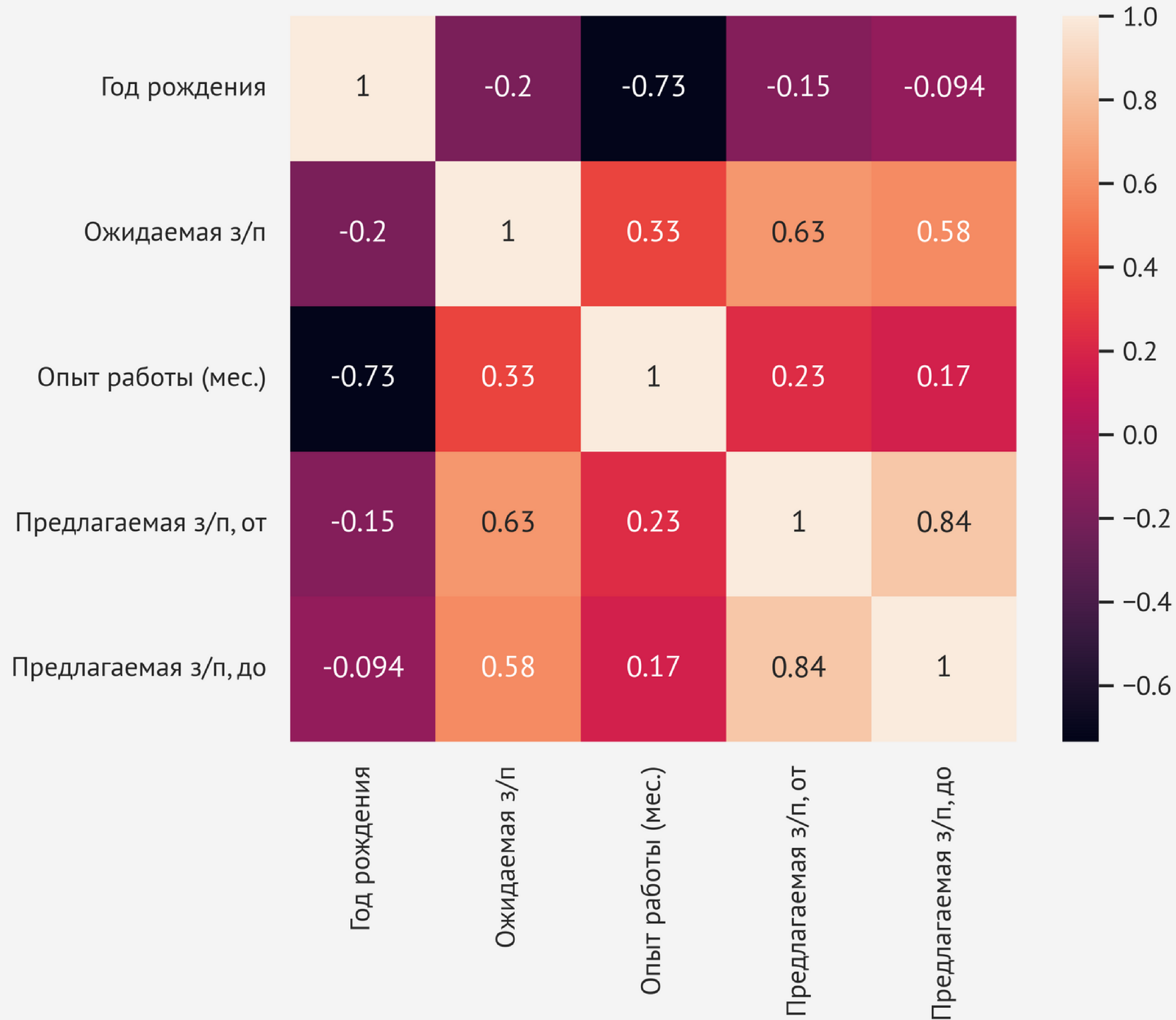
- Год рождения
- Ожидаемая з/п
- Опыт работы
- Предлагаемая з/п (от)
- Предлагаемая з/п (до)



## Номинальные данные

- Регион вакансии
- Финальный статус заявки
- Профессия
- Образование
- ID вакансии и др.

### Корреляции между количественными переменными



**Количественных  
переменных  
всего 5**

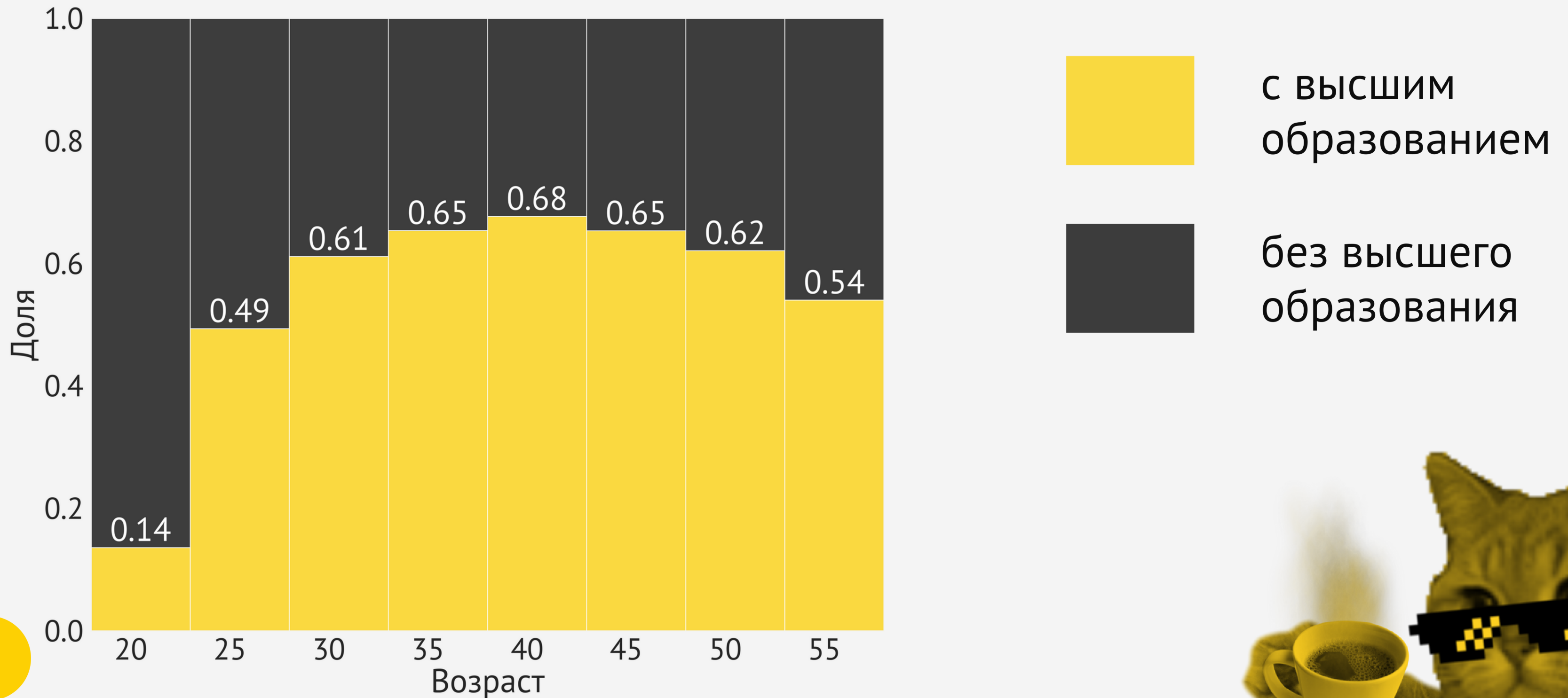


# Предварительный анализ данных

- ✘ Удаляем выбросы
- ✘ Очищаем от пропусков
- ▽ Пропуски в “предлагаемая з/п, от” заполняем значениями из “предлагаемая з/п, до” и наоборот
- ? Рассматриваем только людей от 18 до 60 лет
- ? Убираем жуликов, у которых опыт работы больше продолжительности их жизни
- ? Убираем редко встречающиеся в датасете профессии (5 из 64), суммарно менее 5000 заявок

# Предварительный анализ данных

Доля заявок от людей с высшим образованием



# Исследовательский вопрос

Как регион вакансии влияет на разницу в зарплатах между людьми с высшим образованием и без?

# Гипотеза

В регионах с большей медианной зарплатой  
разница в доходах между людьми с высшим  
образованием и без него больше.

# Механизм

## Человек без высшего образования

- 1) Схожие рабочие места и функции во всех регионах
- 2) Компании не повышают зарплаты выше среднего по региону

## Человек с высшим образованием

- 1) Концентрация крупных компаний в “богатых” регионах
- 2) Потребность в специалистах
- 3) В “богатых” регионах более высокая зарплата для специалистов с высшим образованием
- 4) Средний доход связан не только с уровнем зарплат в регионе, но и с наличием крупных компаний в нем



# Предподготовка данных

Ввели несколько новых переменных на основе старых:

**mean\_compensation** - среднее арифметическое от предлагаемой з/п от и до

**prof\_type** - категория профессии

**region\_type** - категория региона по медианной зарплате

\*распределение в приложении

# Предподготовка данных

Мы разделили профессии на следующие категории:

- Обслуживание
- Информационные технологии
- Инженерия
- Общественные науки
- Управление
- Производство и ремонт

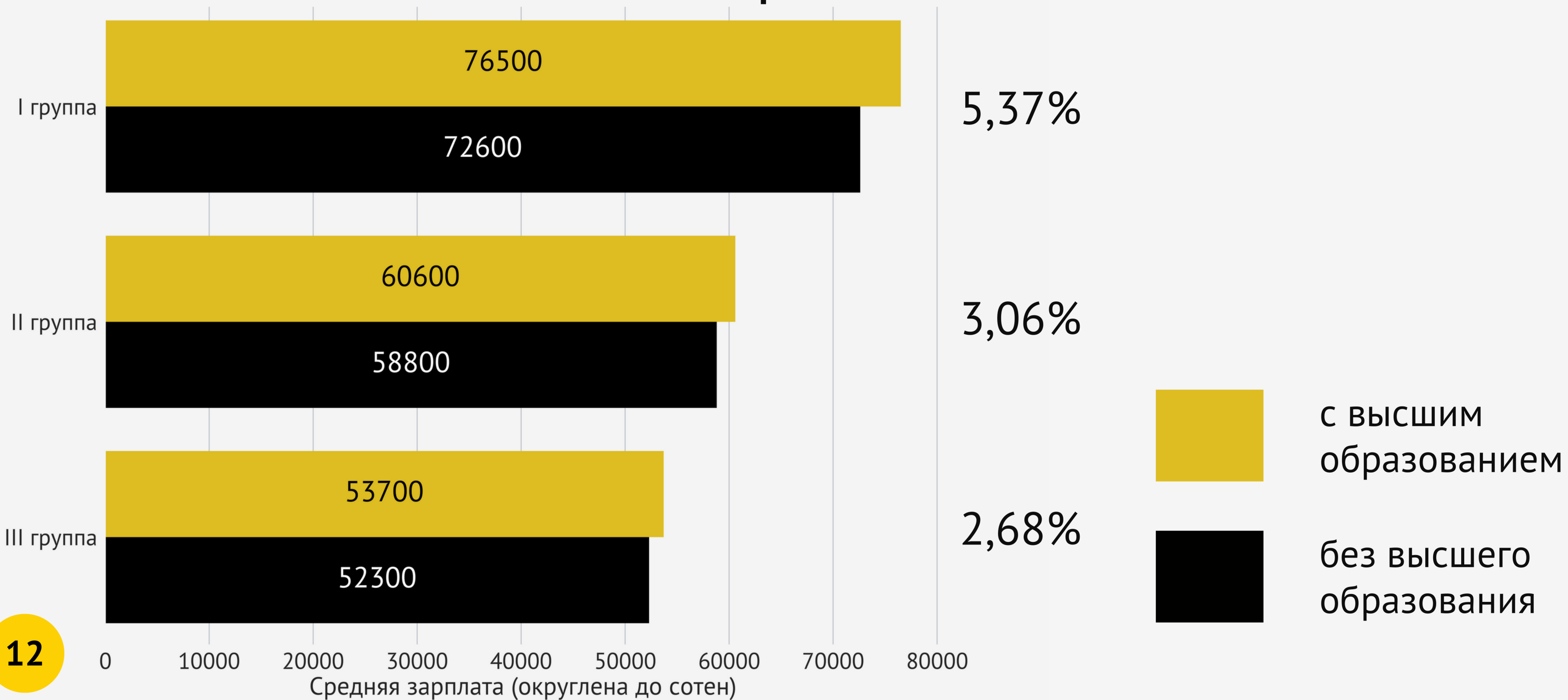
Рассматриваем только уникальные вакансии и заявки со статусом invitation,

итоговая выборка - 108 830 заявок

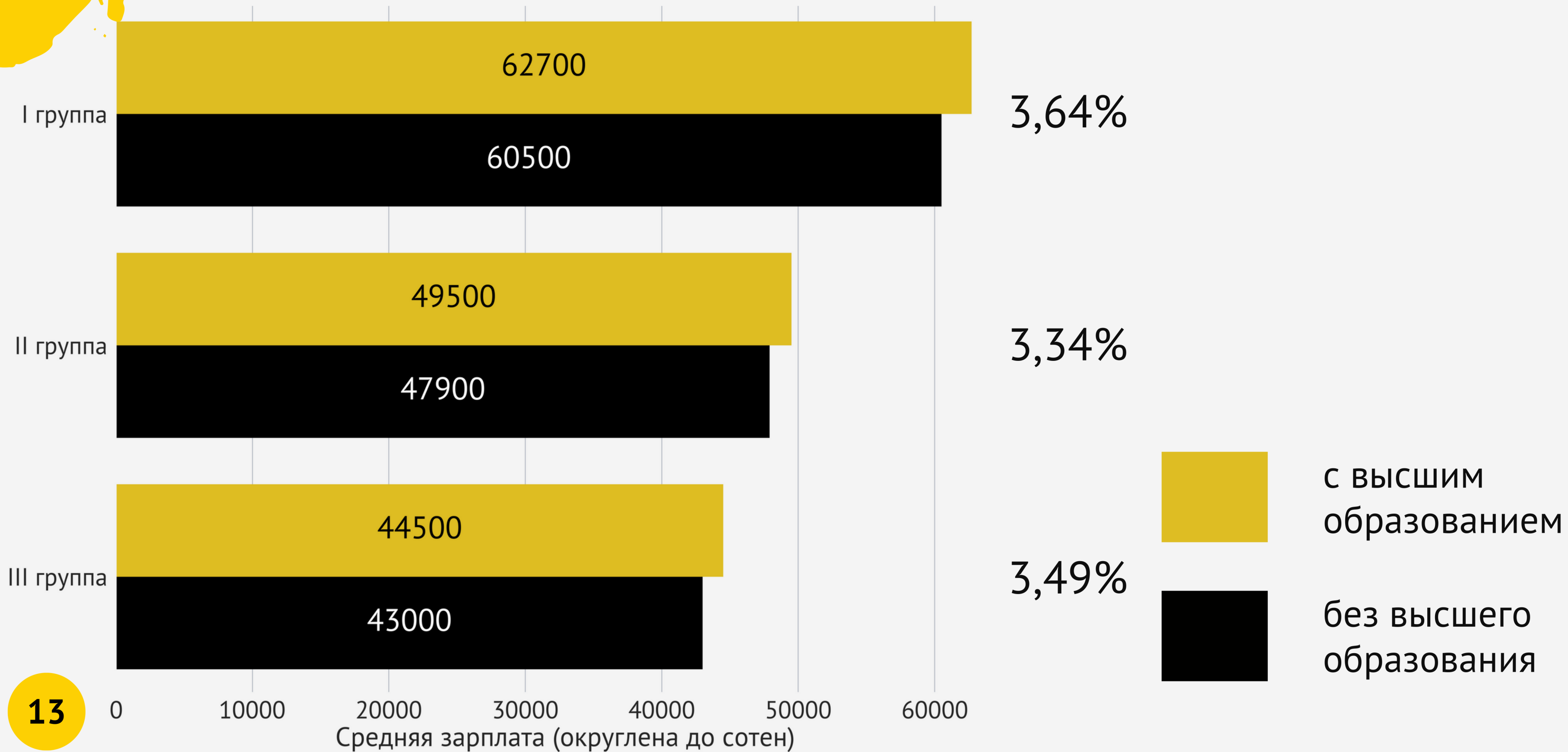


# Визуализации

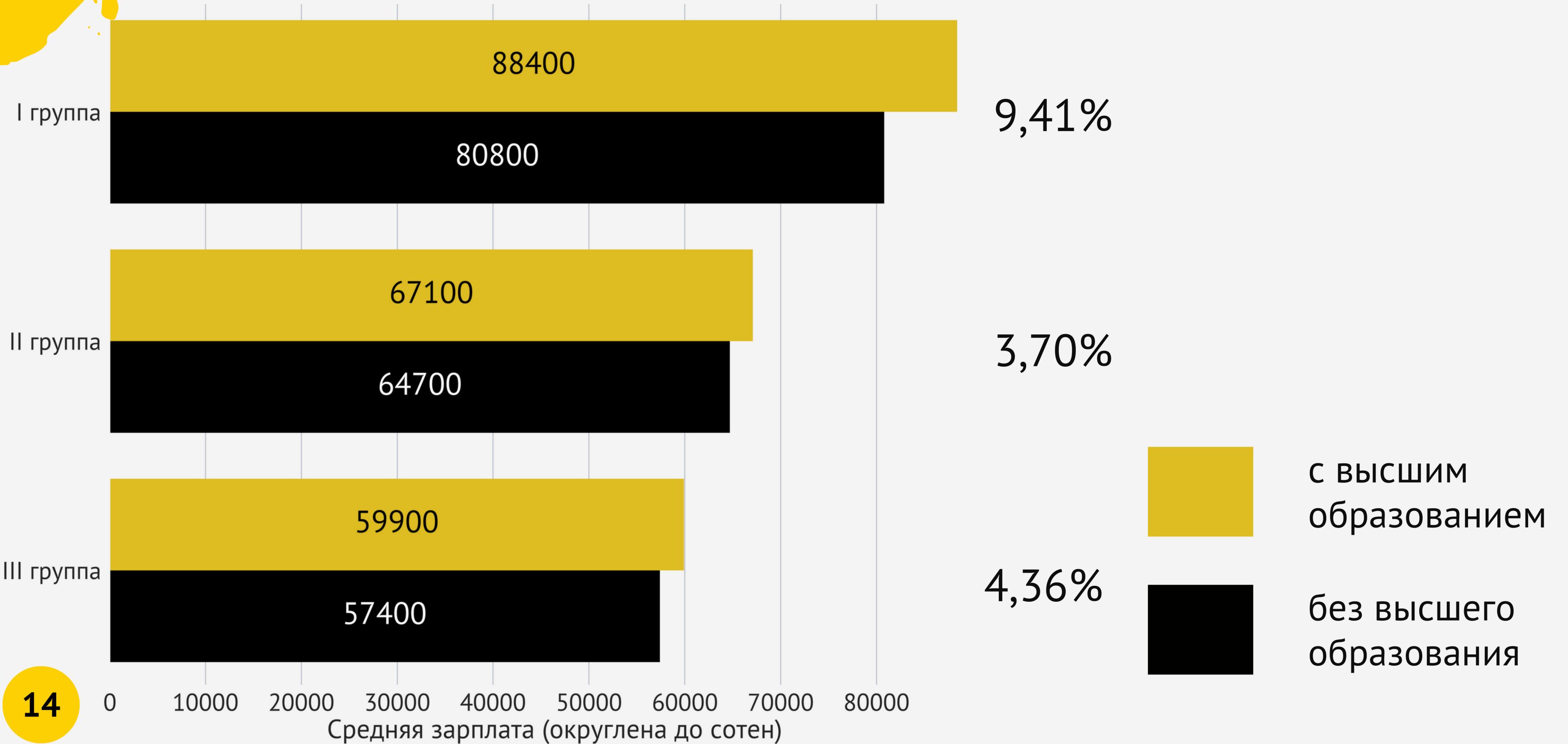
## Управление



# Информационные технологии

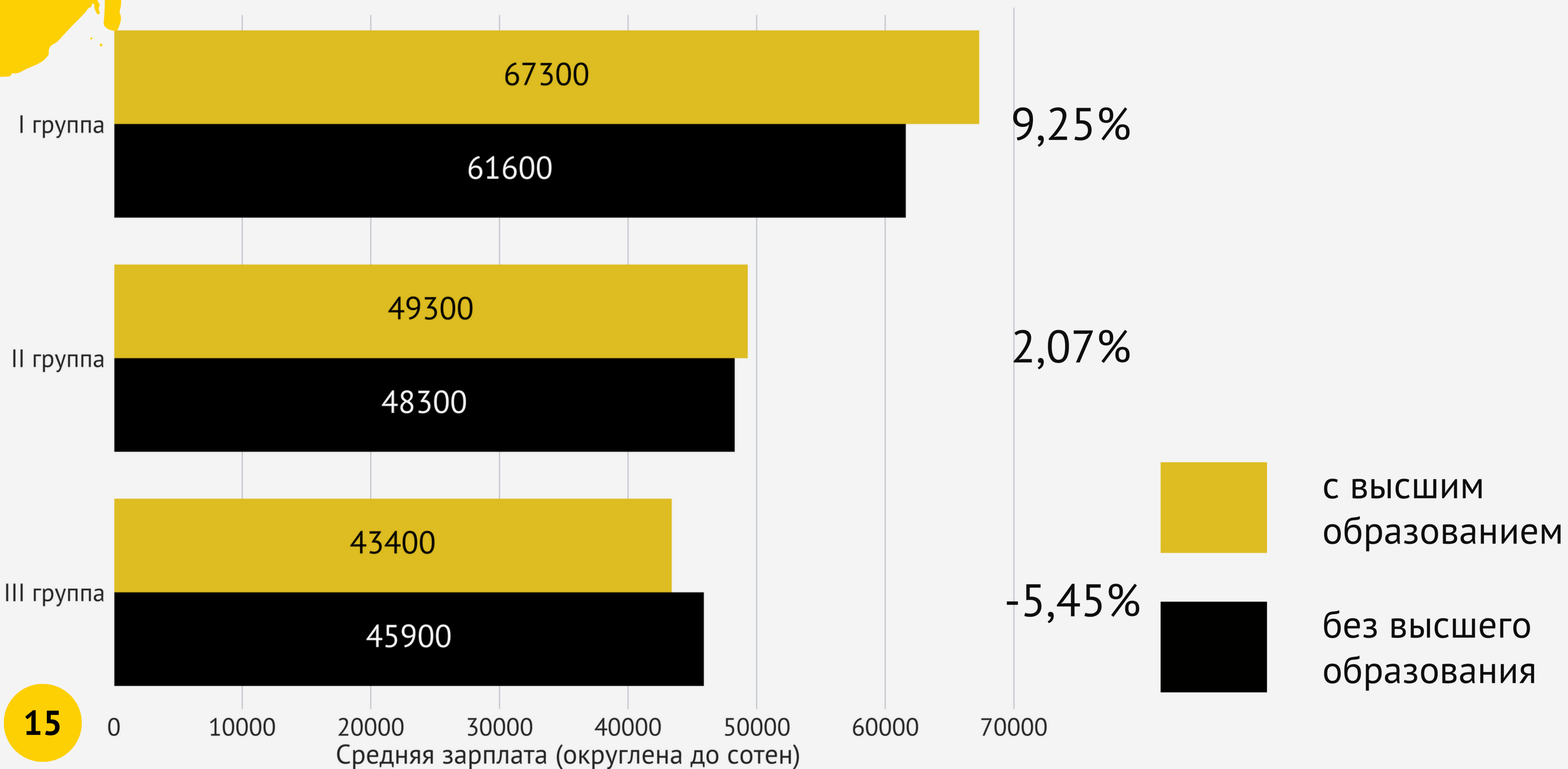


# Инженерия

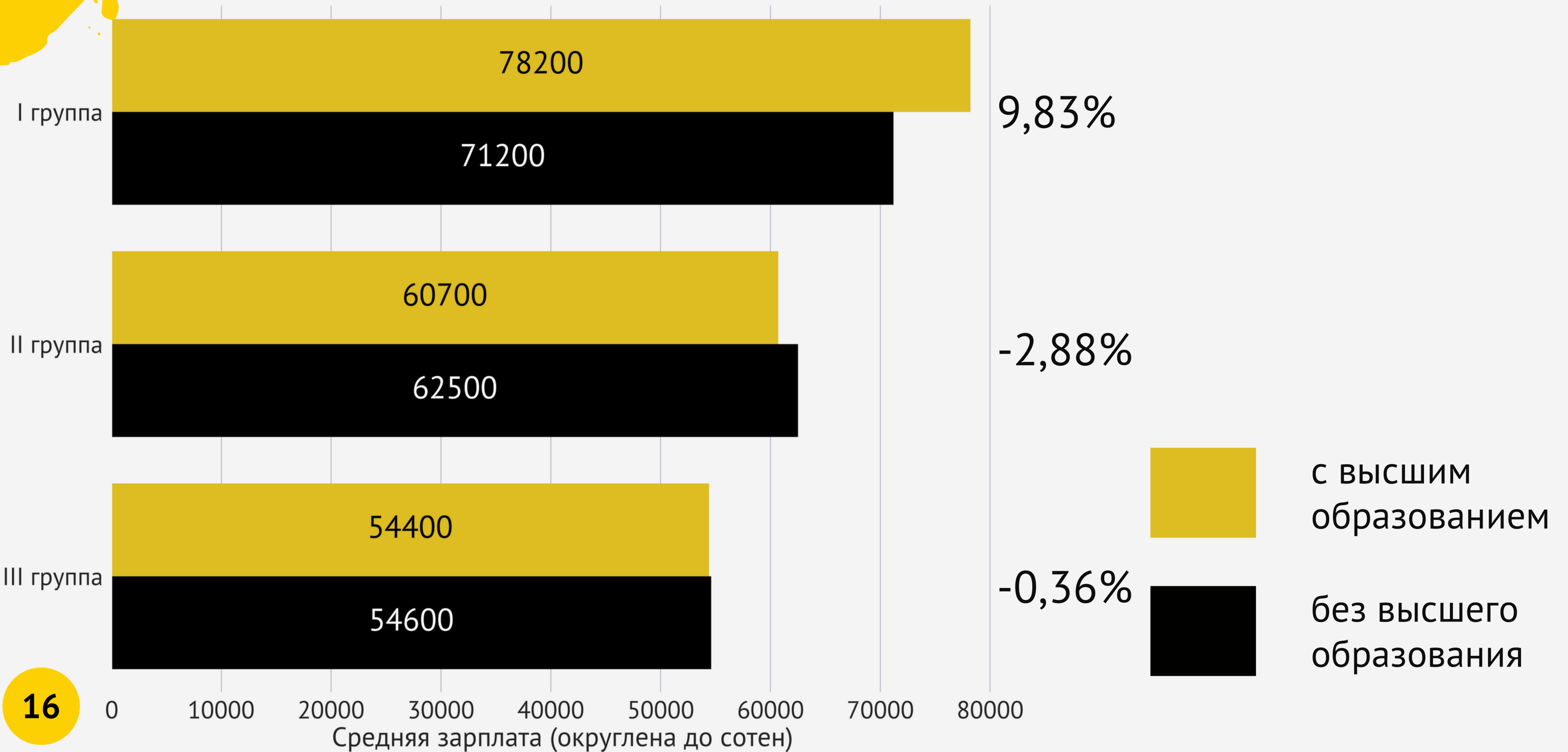




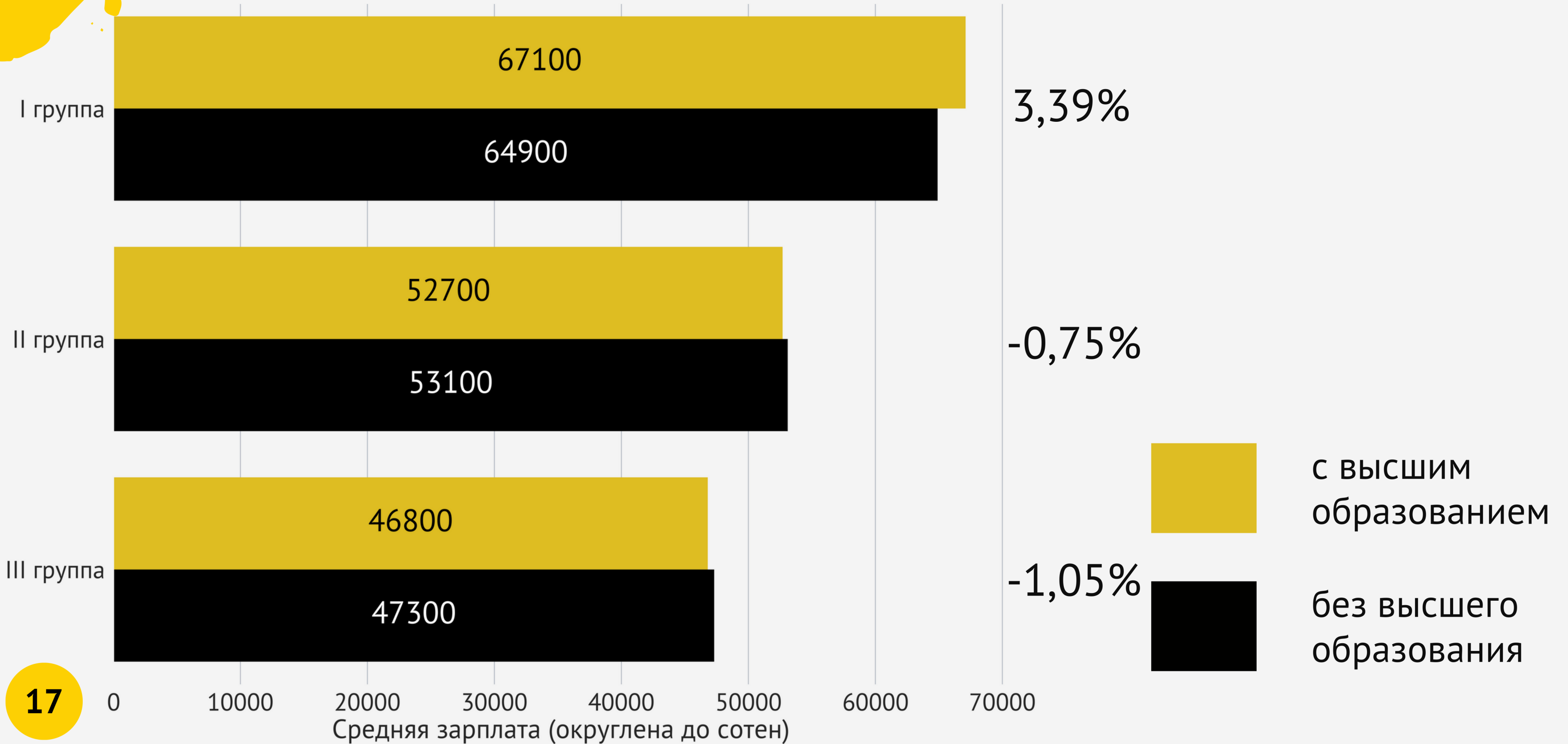
# Общественные науки



# Производство и ремонт



# Обслуживание



# Мат. подтверждение

<b>H<sub>0</sub></b>	<b>Нет различий</b> между средней зарплатой у людей с высшим образованием и без него в зависимости от группы регионов
<b>H<sub>1</sub></b>	<b>Есть различия</b> между средней зарплатой у людей с высшим образованием и без него в зависимости от группы регионов

Используем t-критерий

# Мат. подтверждение



## Управление

	p-value	Выборка	Вывод
I группа	< 0,01	22383	✓
II группа	< 0,01	19772	✓
III группа	0,556	4945	✗

## Производство и ремонт

	p-value	Выборка	Вывод
I группа	< 0,01	3170	✓
II группа	0,049	3813	✓
III группа	0,9267	850	✗

Принимаемый уровень  
значимости 0.05



# Мат. подтверждение

## Инженерия

	p-value	Выборка	Вывод
I группа	< 0,01	2623	✓
II группа	0,0462	2799	✓
III группа	0,392	783	✗

## Информационные технологии

	p-value	Выборка	Вывод
I группа	0,0304	2951	✓
II группа	0,071	2817	✗
III группа	0,28	897	✗

Принимаемый уровень  
значимости 0.05

# Мат. подтверждение

## Обслуживание

	p - value	Выборка	Вывод
I группа	< 0,01	16986	✓
II группа	0,4639	13365	✗
III группа	0,5715	2872	✗

## Общественные науки

	p-value	Выборка	Вывод
I группа	< 0,01	3488	✓
II группа	0,2762	3344	✗
III группа	0,1269	972	✗

Принимаемый уровень  
значимости 0.05

## Выводы

- В более "богатых" регионах разница между зарплатами людей с высшим образованием и без значительно отличается
- В регионах со средней медианной зарплатой влияние высшего образования зависит от сферы деятельности
- В "бедных" регионах нельзя сделать статистически значимый вывод о влиянии образования на предлагаемую зарплату

# Ограничения



Нам предоставлены данные о приглашениях на собеседование, а не о найме на работу



Специфика сайта hh.ru: нельзя обобщить данные нашего исследования на все профессии



Мы не знаем, соответствует ли профиль высшего образования выбранной профессии





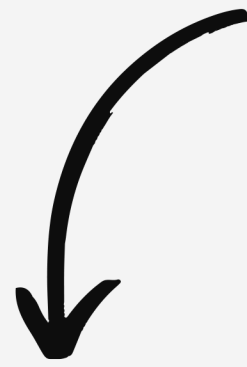
# Практическая польза





# Policy Implication

Создание конкурентноспособной среды в регионах



Постройка офисных помещений  
для новых компаний

Льготное налогообложение  
для новых компаний





**СПАСИБО ЗА  
ВНИМАНИЕ!**

**Наша команда:**

Калашникова Любовь

Сухарева Дарья

Зайцев Сергей

Захаров Арсений

Бетелев Михаил

**Время ответов на вопросы!**

**Приложение**

# Приложение

## Распределение по сферам

### Инженерия

Инженер-энергетик, инженер-электрик, инженер-конструктор, инженер-проектировщик, механик, сервисный инженер, инженер-механик, инженер по охране труда и технике безопасности, инженер-эколог

### Информационные технологии

Программист, разработчик, оператор ПК, оператор базы данных, аналитик, тестировщик

### Производство и ремонт

Сварщик, слесарь, сантехник, оператор производственной линии, технолог, разнорабочий

### Общественные науки

Делопроизводитель, архивариус, юрисконсульт, экономист, юрист, кредитный специалист

# Приложение

## Распределение по сферам

### Управление

Бухгалтер, интернет-маркетолог, специалист по подбору персонала, директор магазина, сети магазинов, руководитель отдела продаж, системный администратор, секретарь, помощник руководителя, ассистент, различные менеджеры, руководитель проектов, супервайзер, торговый представитель, администратор, мерчендайзер, начальник производства, начальник смены, мастер участка, генеральный директор, исполнительный директор, координатор отдела продаж, начальник склада

### Обслуживание

Официант, бармен, бариста, специалист технической поддержки, водитель, охранник, воспитатель, няня, курьер, кассир-операционист, диспетчер, продавец-консультант, продавец-кассир, машинист, фитнес-тренер, инструктор тренажерного зала, оператор call-центра, специалист контактного центра, уборщица, уборщик

# Приложение

## Удалённые профессии

- Удалили 5 профессий, в которых мало заявок и их нельзя добавить ни в одну из сфер или создать новую, из 64: дизайнер, учитель, психолог, врач, повар
- Суммарно было удалено 5769 заявок



# Приложение

## Количество людей в сферах

Управление - 161660 человек

Общественные науки - 37515

человек

Обслуживание - 88400 человек

Производство и ремонт - 17079

человек

Инженерия - 19411 человек

Информационные технологии - 24075

человек

# Приложение

Количество заявок в группах регионов  
(учитывая все финальные статусы  
взаимодействия)

I группа - 190721 человек

II группа - 113432 человек

III группа - 43987 человек

# Приложение

Количество заявок в группах регионов  
(учитывая только приглашенных на  
собеседование)

I группа - 66393 человек

II группа - 52239 человек

III группа - 11884 человек

# Приложение

Удалили 12 регионов, из которых было менее 100 заявок.  
В сумме было удалено 440 строк.

Удалённые регионы: Чеченская республика, Донецкая область, Луганская область, республика Алтай, Кабардино-Балкарская республика, республика Ингушетия, республика Калмыкия, Чукотский АО, Еврейская АО, Ненецкий АО