



СКОЛЬКО МОЖНО РАБОТАТЬ!?

От команды "За деньги – DA, NO не всегда"

Сусликов Иван
Круглов Ярослав
Петров Михаил





Вступление

Наша база данных

hh.ru:
характеристики вакансий
и резюме

Каждая строка –
взаимодействие
соискателя и работодателя.

Всего – **500 000** значений

Даты всех
взаимодействий – **лето**
2023 года



Вступление

Переменные

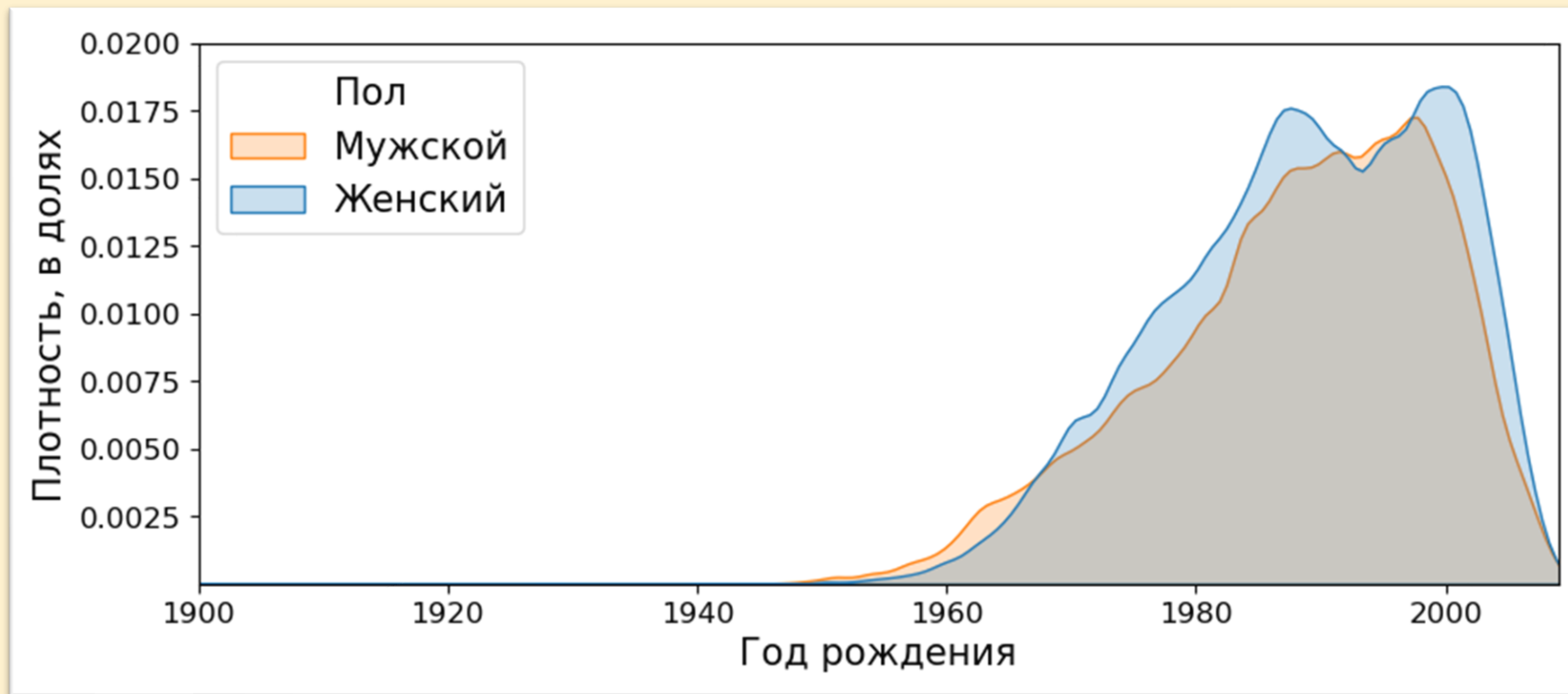
Количественные:

- Год рождения
- Опыт работы в месяцах
- Ожидаемая зарплата
- Зарплата от и до

Качественные:

- Параметры взаимодействия
- Пол
- Степень образования
- Профессия
- Регион вакансии

Распределение количества резюме по году рождения соискателя



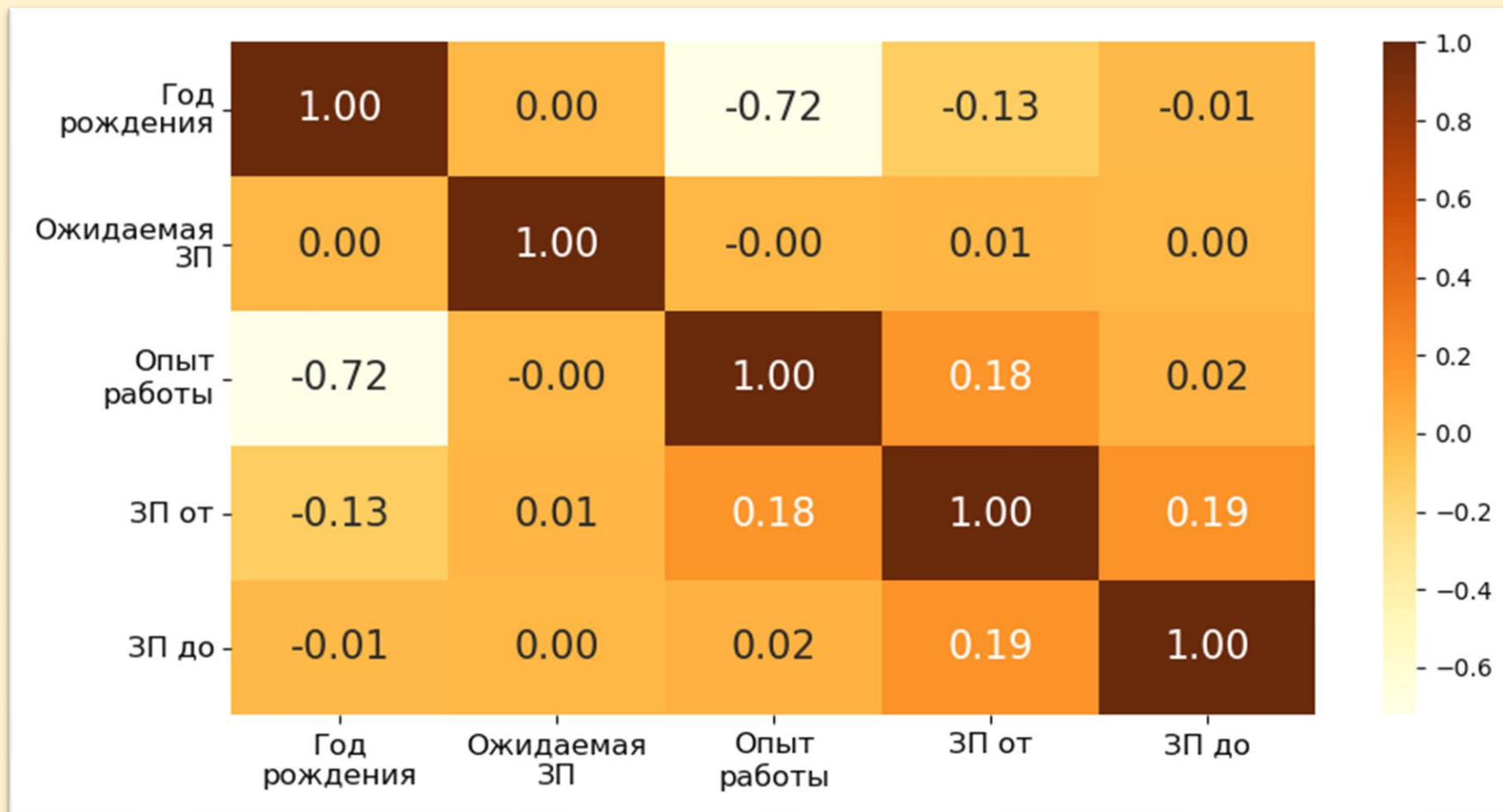
Медиана – 1990 – 33 года

Мода – 1998 – 25 лет

1 2 3 4 5

Вступление

Матрица корреляции



① ② ③ ④ ⑤

Предобработка

Работа с пропущенными значениями

- Удаляем пропущенные значения в **initial_state** и **final_state**

Удалено **936** значений
Осталось **99,8%** датасета

① ② ③ ④ ⑤

Предобработка

Введение новых переменных

`year_of_birth`



`applicant_age` - возраст

`initial_state`

`final_state`



`successful_topic` - ГОТОВНОСТЬ
работодателя пригласить
соискателя на собеседование

① ② ③ ④ ⑤

Предобработка

Работа с выбросами. Очистка

- Удаляем **applicant_age > 74**
- Удаляем выбросы **work_experience_months**
- Удаляем волонтеров

Удалено **469** значений
Осталось **99,7%** датасета

① ② ③ ④ ⑤

Гипотеза и механизм



Как **регион** влияет на **требование** работодателя в разных **профессиях**?

Требование работодателя -
статистически значимые переменные для **successful_topic**



Чем **больше*** регион, тем в среднем **больше** значимость **рабочего опыта** для работодателя

*Регион с большим кол-вом уникальных резюме

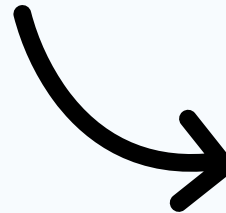
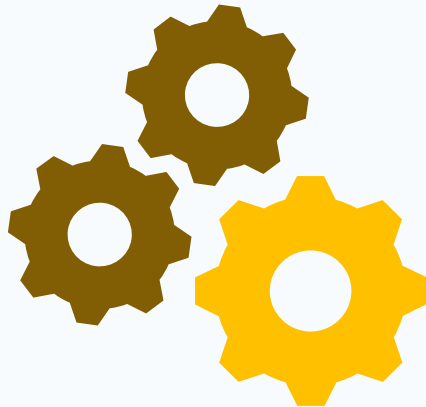
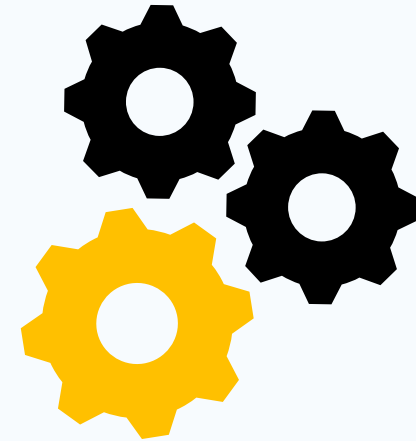
1 2 3 4 5

Гипотеза и механизм

Больше регион -
больше предложение
труда



Работодатель имеет
большой выбор
кандидатов



Он выбирает более
опытных кандидатов



Гипотеза и механизм

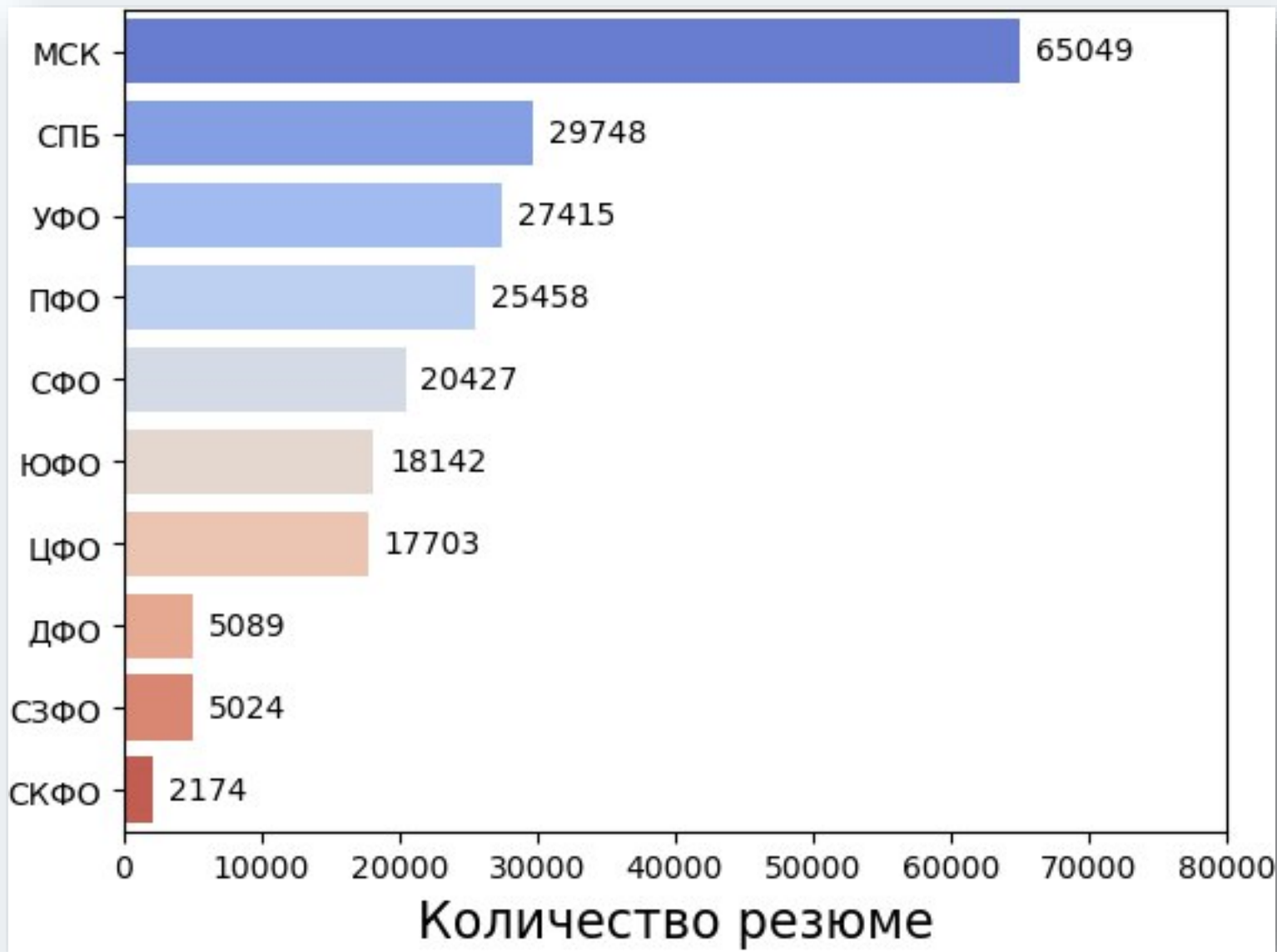
Группировка по регионам и профессиям

Группируем **vacancy_region** по федеральным округам
Отделяем **Москву** и **Санкт-Петербург** в **отдельные** группы

Группируем **profession** по специализациям, указанным на
сайте hh.ru

① ② ③ ④ ⑤

Гипотеза и механизм



①②③④⑤

Статистическая проверка

Проведем **t-test**

H₀

В генеральной совокупности **нет различий** в вероятности приглашения на собеседование среди людей с разным **опытом работы**

H₁

В генеральной совокупности **есть различия** в вероятности приглашения на собеседование среди людей с разным **опытом работы**

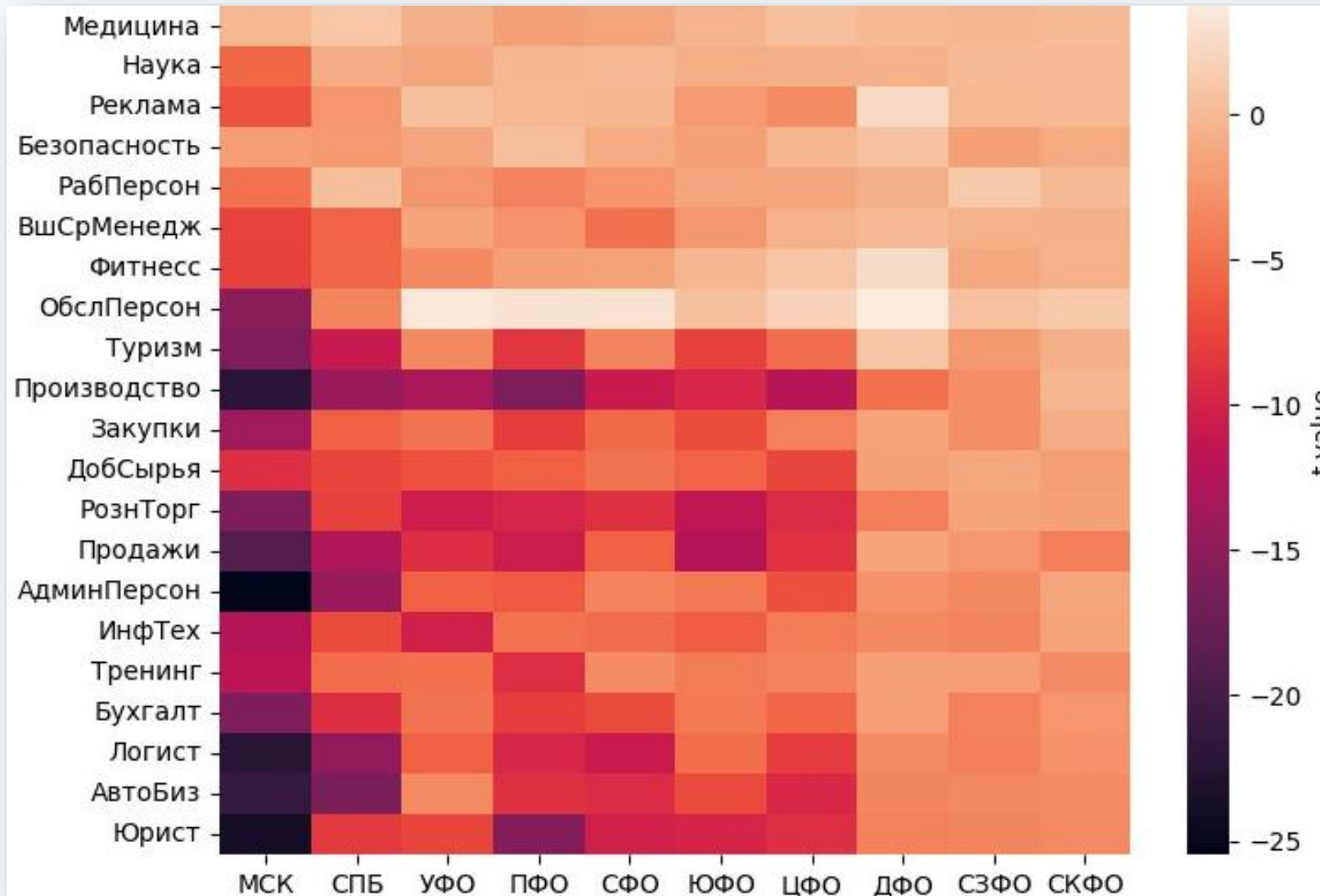


Принимаемый **уровень значимости - 5%**

① ② ③ ④ ⑤

Статистическая проверка

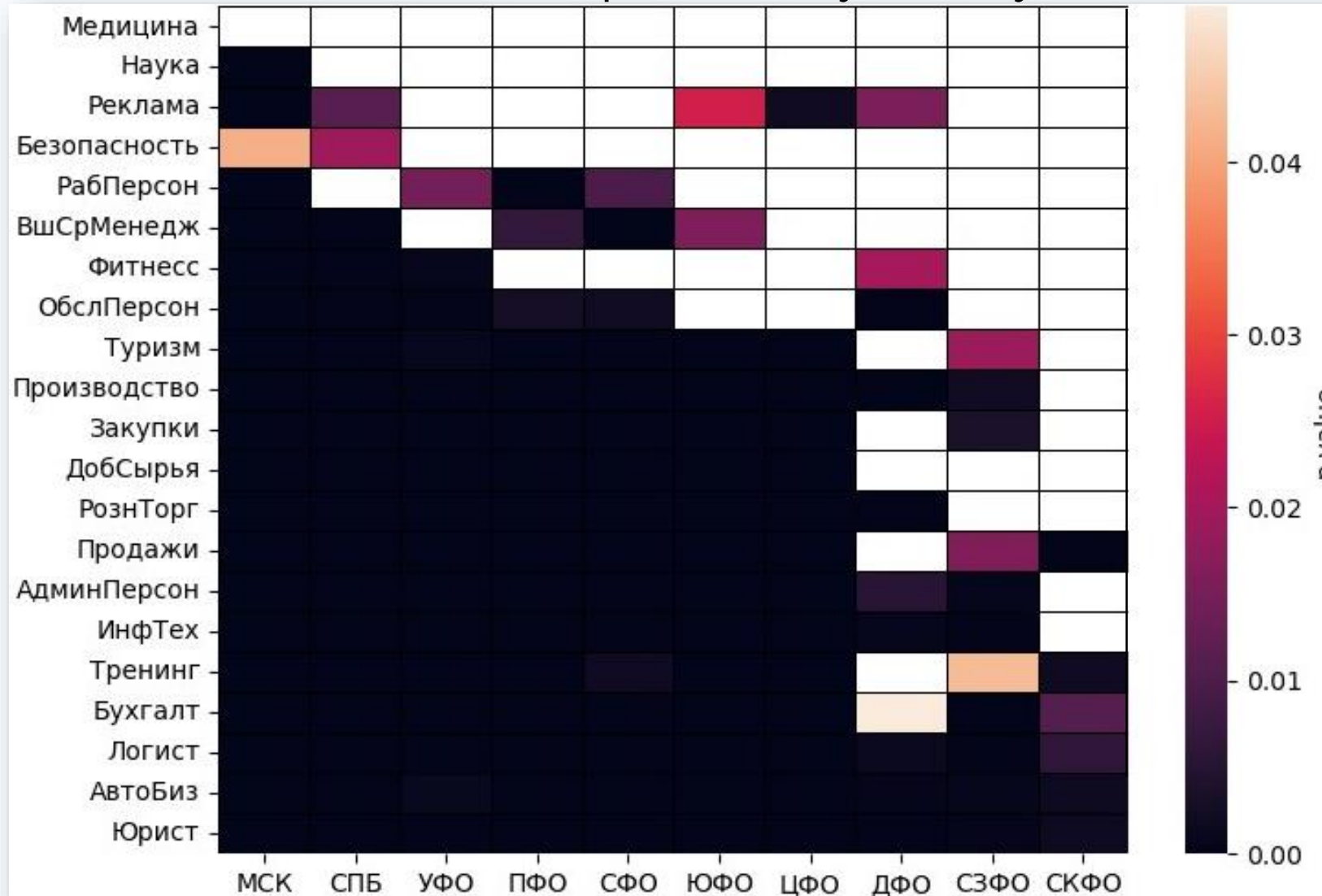
T-value по рабочему опыту



① ② ③ ④ ⑤

Статистическая проверка

P-value по рабочему опыту





Выводы



В больших регионах **значимость рабочего опыта** для работодателя в среднем больше, чем в меньших регионах

Гипотеза **подтвердилась**, но не во всех профессиях и не во всех регионах

Есть статистически значимые различия в требованиях работодателей в разных регионах

① ② ③ ④ ⑤

Выводы

Наша работа

1

Посещение сайта

2

Отклик на вакансию

3

Приглашение на собеседование

Стек ML

разработки hh.ru*

*Взято с ai-russia.ru/library/hh



Выводы

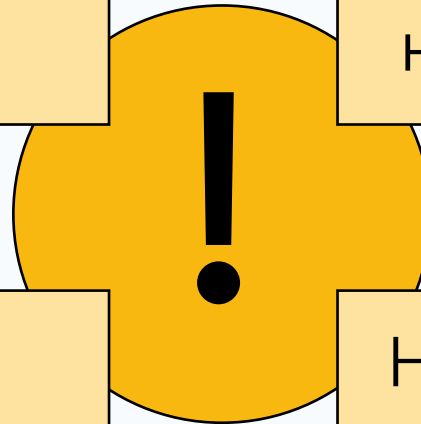
Ограничения исследования

Данные за короткий промежуток времени

Нельзя делать выводы насчет трудоустройства

Результаты нельзя обобщить на весь мир

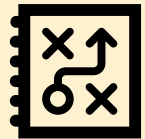
Не во всех регионах есть все профессии





Выводы

Перспективы исследования



Собрать данные о результатах собеседований



Собрать данные о прошлых трудоустройствах



Отследить эту же выборку соискателей
через год



Выводы

Применение

Соискатели без рабочего опыта на hh.ru **смогут найти** подходящую себе работу не только во вкладке «Нет опыта», но и **среди найденных нами** профессий и регионов

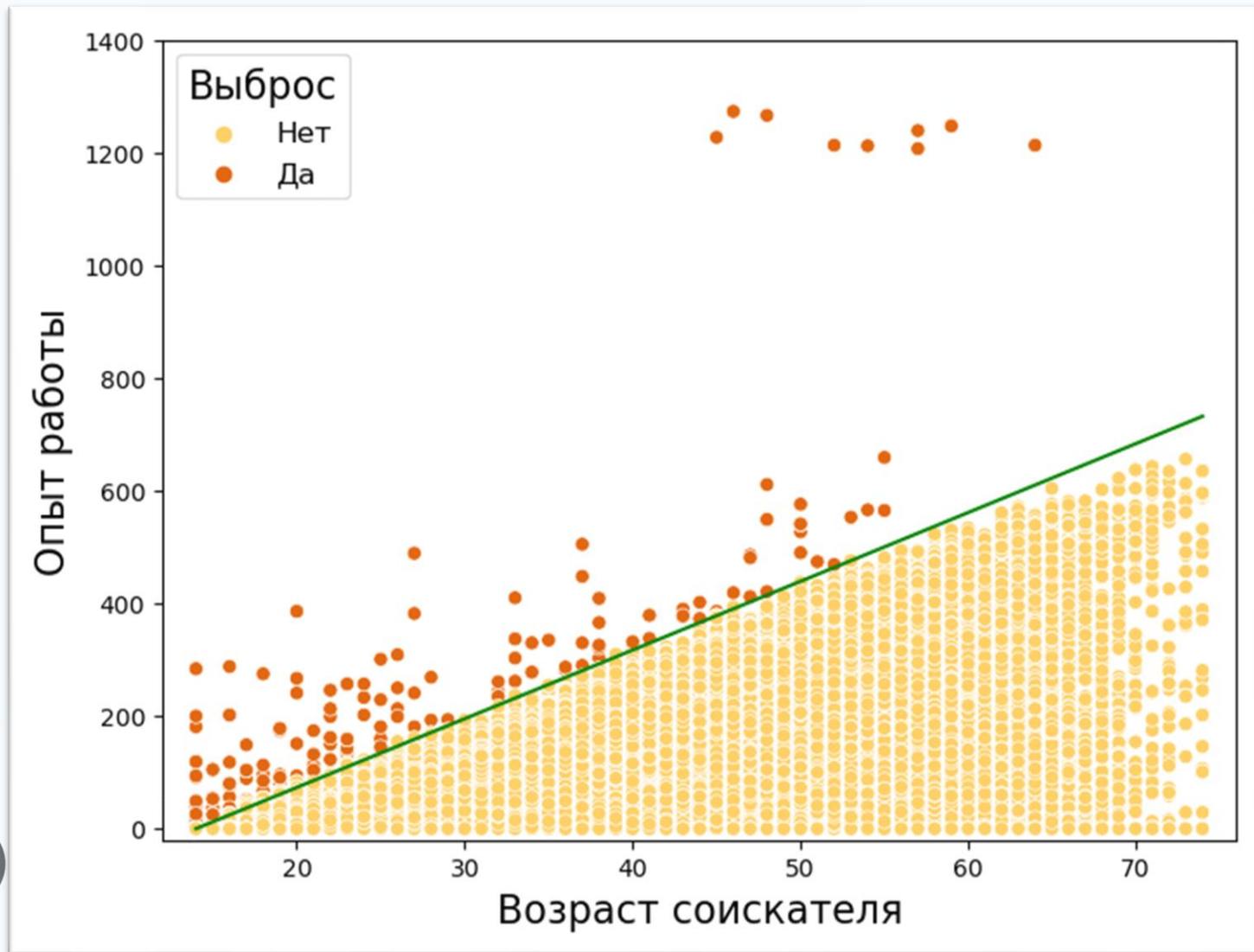
Приложение к слайду 7

successful_topic

описание	initial_state	final_state	successful_topic
отказ	discard	discard	0
приглашение на собеседование	invitation	discard	1
приглашение на собеседование	invitation	invitation	1
отклик	response	discard	0
приглашение на собеседование	response	invitation	1
отклик	response	response	0

Приложение к слайду 8

work_experience_months



Вводим вспомогательные переменные:

- ***max_possible_exp*** - максимальный опыт работы, который соискатель может иметь (зависит от возраста)
- ***diff_in_exp*** - разница между ***max_possible_exp*** и ***work_experience_months***

Удаляем всех, у кого

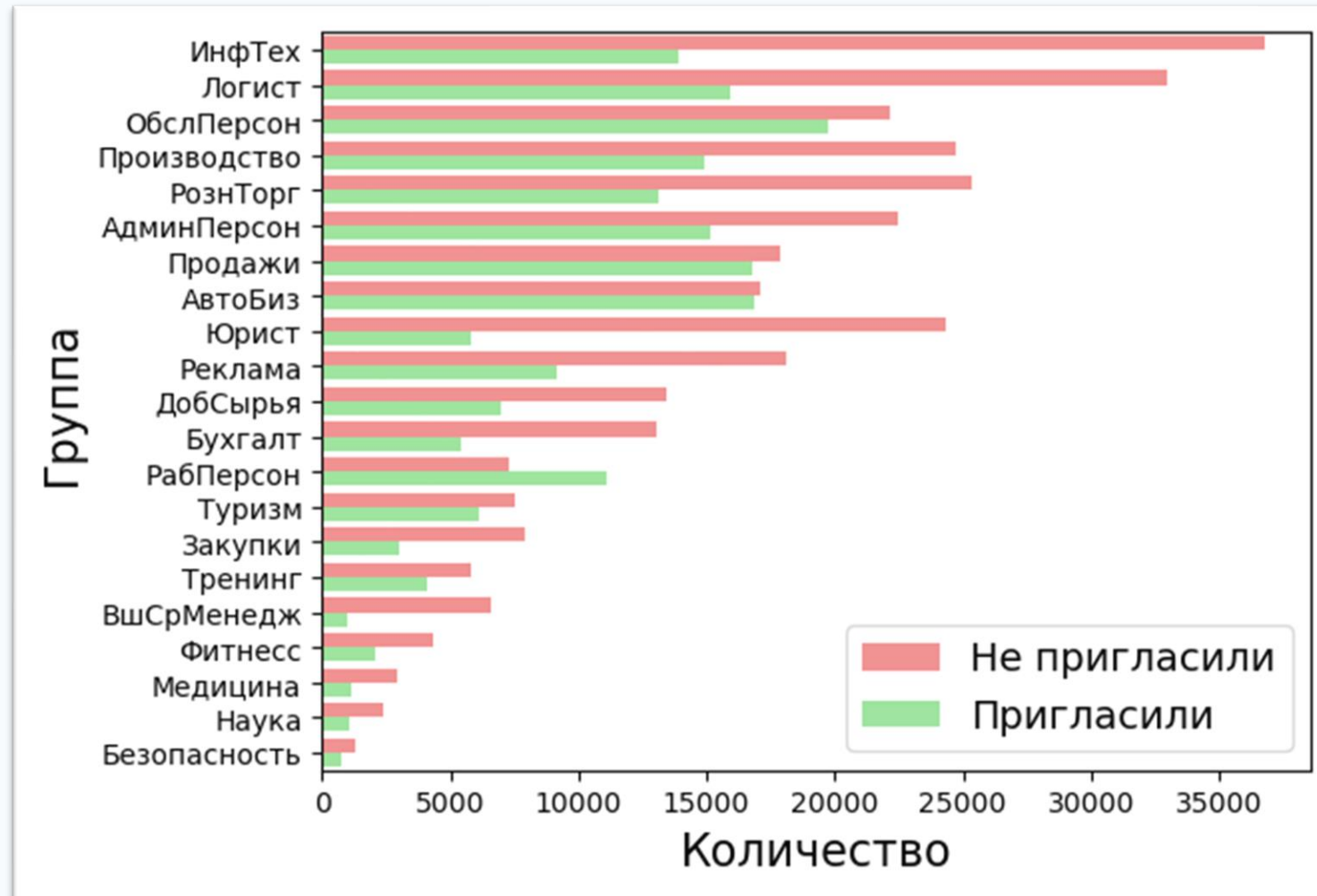
$$\mathbf{diff_in_exp} < -12$$

(погрешность измерения возраста)

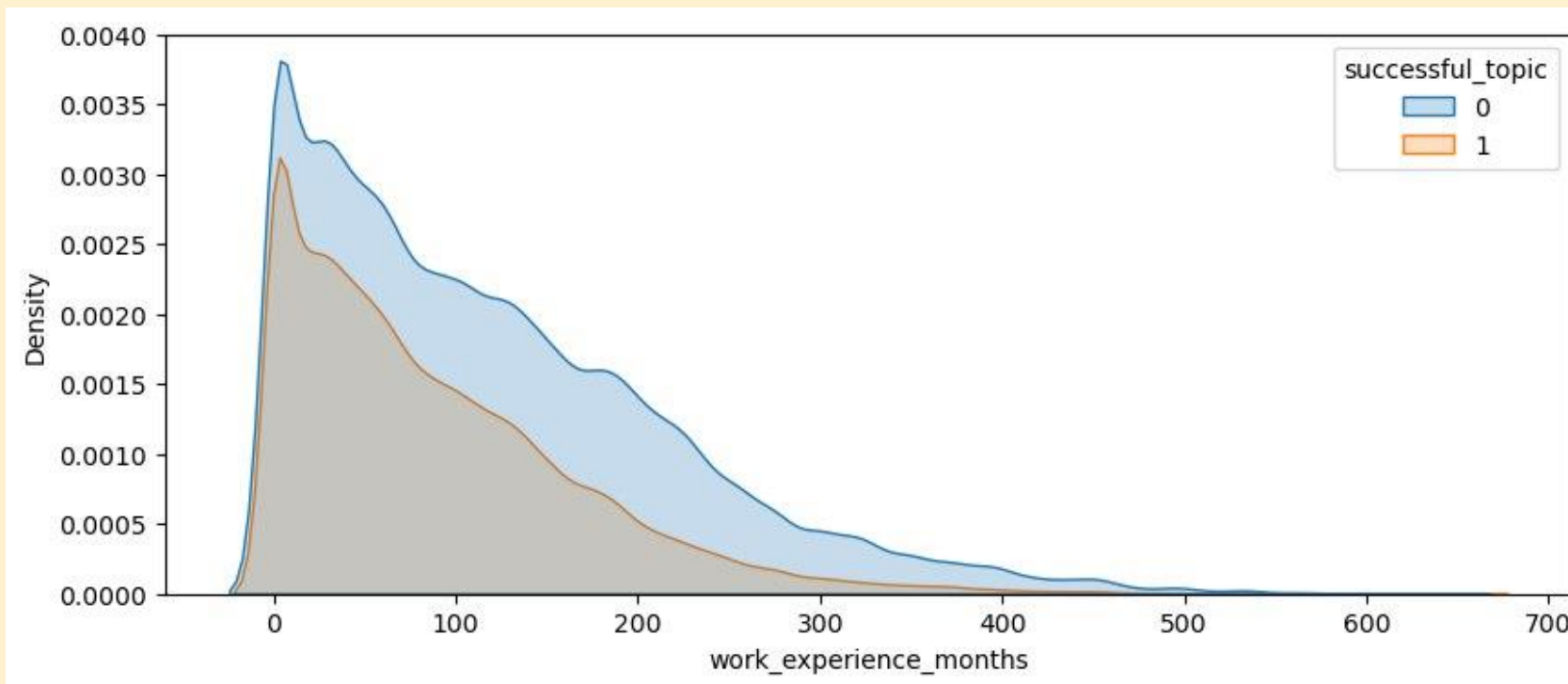
Приложение к слайду 11

profession

График количества взаимодействий по каждой группе профессий



Распределение рабочего опыта

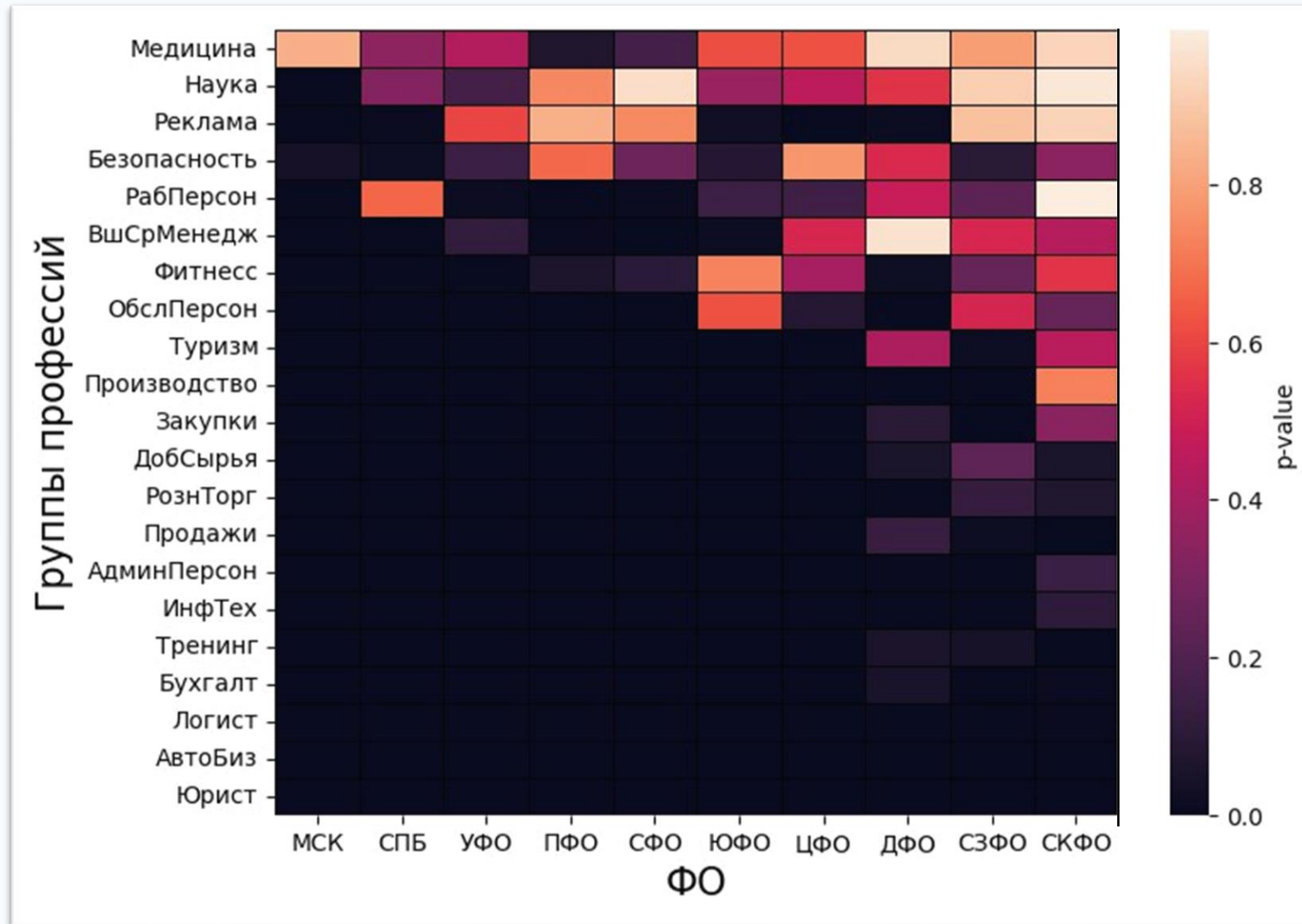


Медиана – **85** месяцев

Мода – **0** месяцев

Приложение к слайду 15

heatmap





Спасибо за внимание!

Работа лежит на странице команды
github.com/listens-to-spotify/dano2023

