

Надо было уходить после 9... Или нет?



dano(н)айти:

Мирошникова Александра Владимировна

Малакшанидзе Анна Давидовна

Казакова Евгения Яковлевна

Ковальчук Артём Игоревич

План исследования

1. Предварительный анализ данных
2. Предобработка данных
3. Постановка исследовательского вопроса и гипотезы
4. Математическое подтверждение
5. Интерпретация результата
6. Ограничения и перспективы
7. Применение и Policy implication



Предварительный анализ данных

Данные - несбалансированные пространственные, в срезе за июнь, июль и август 2023 года.

Переменные

- **определяемые соискателем:** персональные данные, навыки, предпочтения по работе;
- **определяемые работодателем:** статусы взаимодействия, данные о вакансии;
- **создаваемые сервисом:** идентификаторы взаимодействия, даты.

- **189402** уникальных resume id
- **319481** уникальных vacancy id

Предварительный анализ данных

initial state final state	response	invitation	discard	NaN
response	242739	102976	72006	0
invitation	0	80523	846	0
discard	0	0	76	0
NaN	0	0	0	834

Статусы взаимодействия

Предобработка данных: обработка выбросов

Качественные переменные

дубликаты уникального topic id	51
--------------------------------	----

первоначальный статус взаимодействия	834
финальный статус взаимодействия	834
уровень образования	102
предпочитаемый график работы	1773

наличие докторской степени у соискателя	60
---	----

соответствие возраста и уровня образования	532
--	-----

Количественные переменные

возраст ниже 1-ого перцентиля	3954
-------------------------------	------

аномальный стаж работы	6240
------------------------	------

соотношение минимальной и максимальной предлагаемых зарплат	29380
---	-------

ожидаемые зарплаты меньше 3-его и больше 97-ого перцентилей	25631
---	-------

Предобработка данных: обработка выбросов



Предобработка данных: замена переменных

Нет высшего образования

Есть высшее образование

1. **education_level :**
 - secondary: 0
 - special_secondary: 0
 - unfinished_higher: 0 / 1
 - bachelor: 1
 - master: 1
 - higher: 1
 - candidate: 1
 - doctor: -

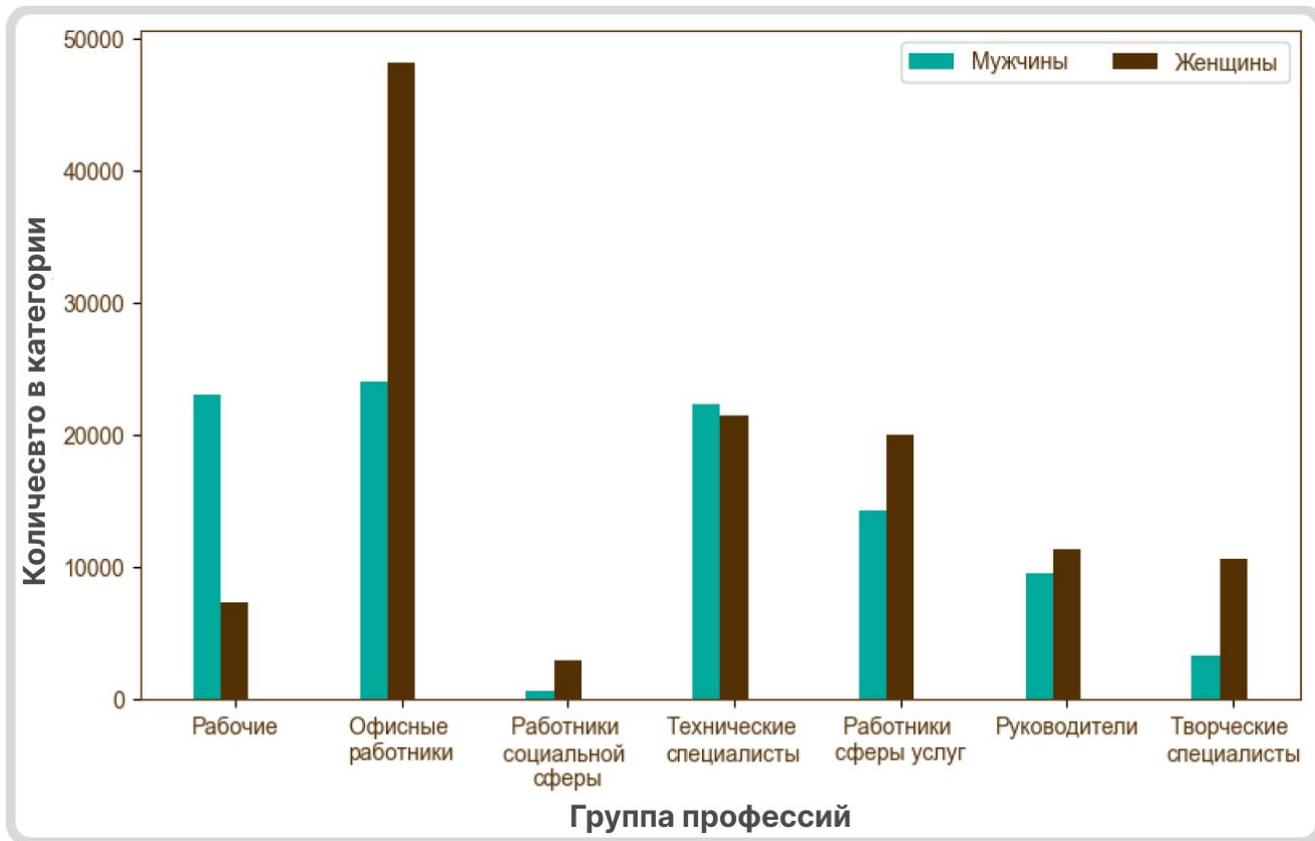
2. **gender:**
 - female: 0
 - male: 1

3. **final_state:**
 - discard: 0
 - invitation: 1

Предобработка данных: классификация профессий

Тип воротничка	Тип профессии	Примеры	Количество
Синий	Рабочие	Кладовщик, сварщик	30477
Белый	Офисные работники	Бухгалтер, экономист	72185
Серый	Социальные работники	Учитель, воспитатель	3640
Фиолетовый	Технические специалисты	Программист, инженер	43863
Розовый	Работники сферы услуг	Официант, кассир	34227
Золотой	Руководители	СЕО, начальник производства	20850
Воротничок отсутствует	Творческие специалисты	Дизайнер, контент-менеджер	13905

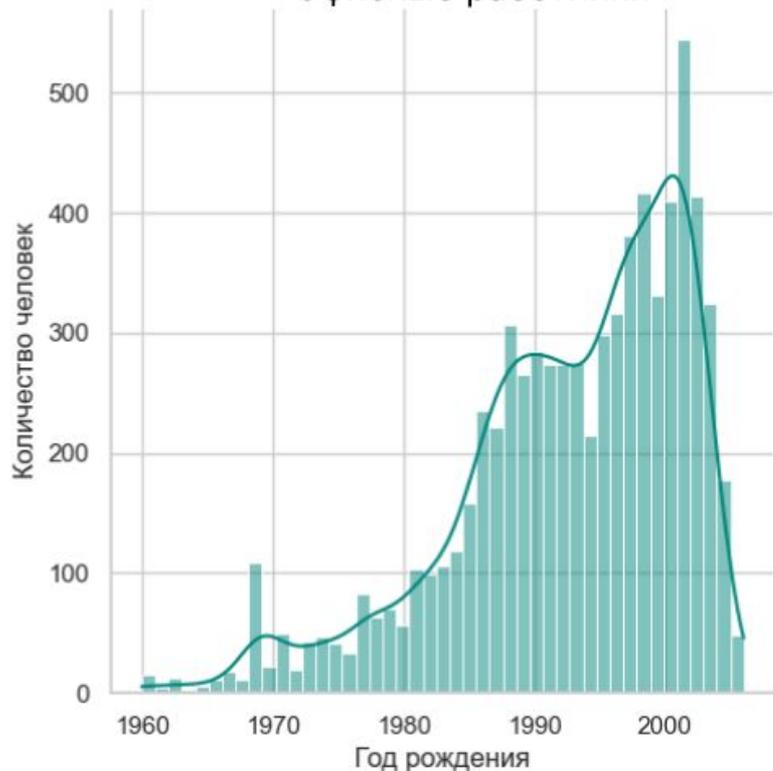
Предобработка данных: классификация профессий



Распределение полов по профессиям

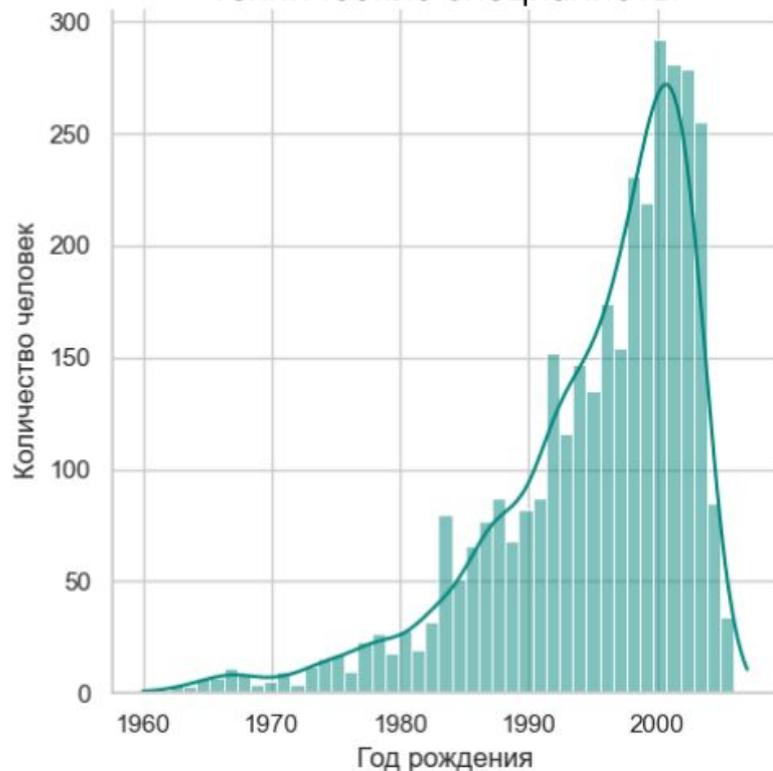
Предобработка данных: классификация по образованию

Офисные работники*



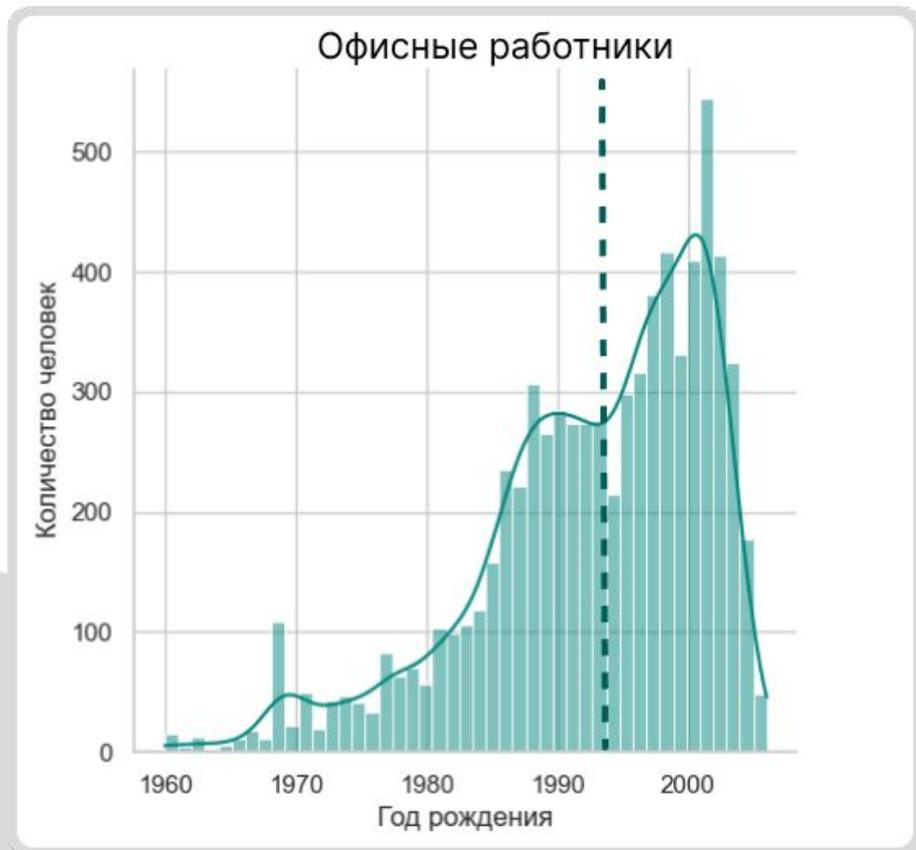
Распределение неоконченного высшего образования

Технические специалисты*



*оставшиеся типы профессий в приложении

Предобработка данных: классификация по образованию



Замена неоконченного высшего образования на:

Среднее образование

рабочие

работники сферы услуг

офисные работники
родившиеся до 1994

Высшее образование

работники социальной сферы

технические специалисты

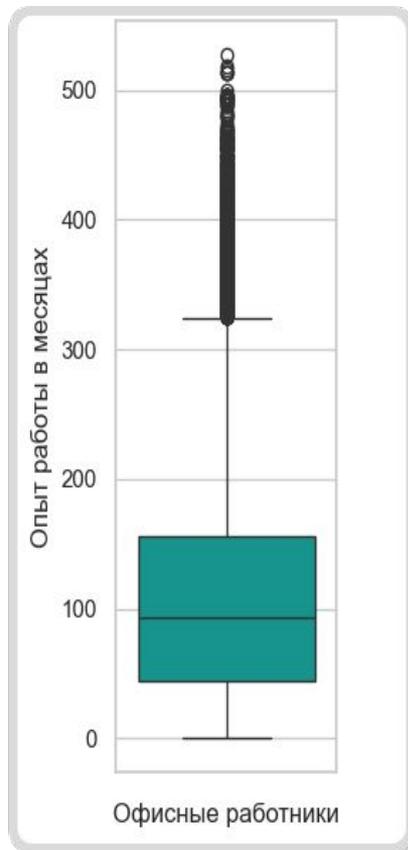
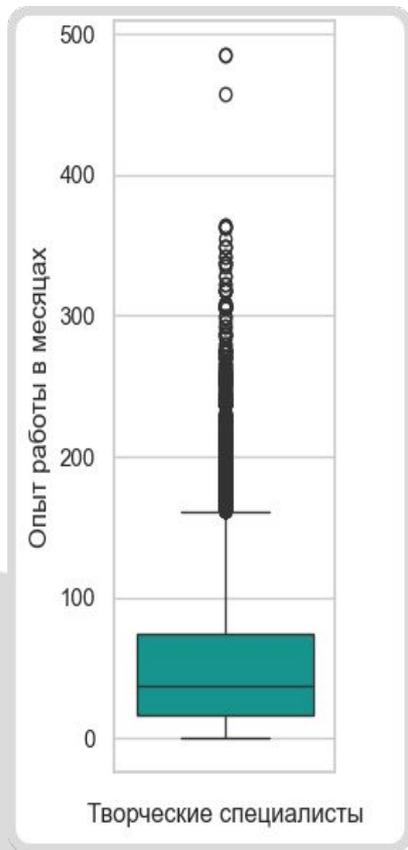
творческие специалисты

руководители

офисные работники
родившиеся после 1994

Распределение неоконченного высшего образования

Предобработка данных: классификация по опыту работы



Опыт работы	Количество офисных работников
Нет опыта	2744
До 25-ого перцентиля	17566
От 25-ого до 50-ого перцентилей	17337
От 50-ого до 75-ого перцентилей	17302
От 75-ого перцентиля	17236

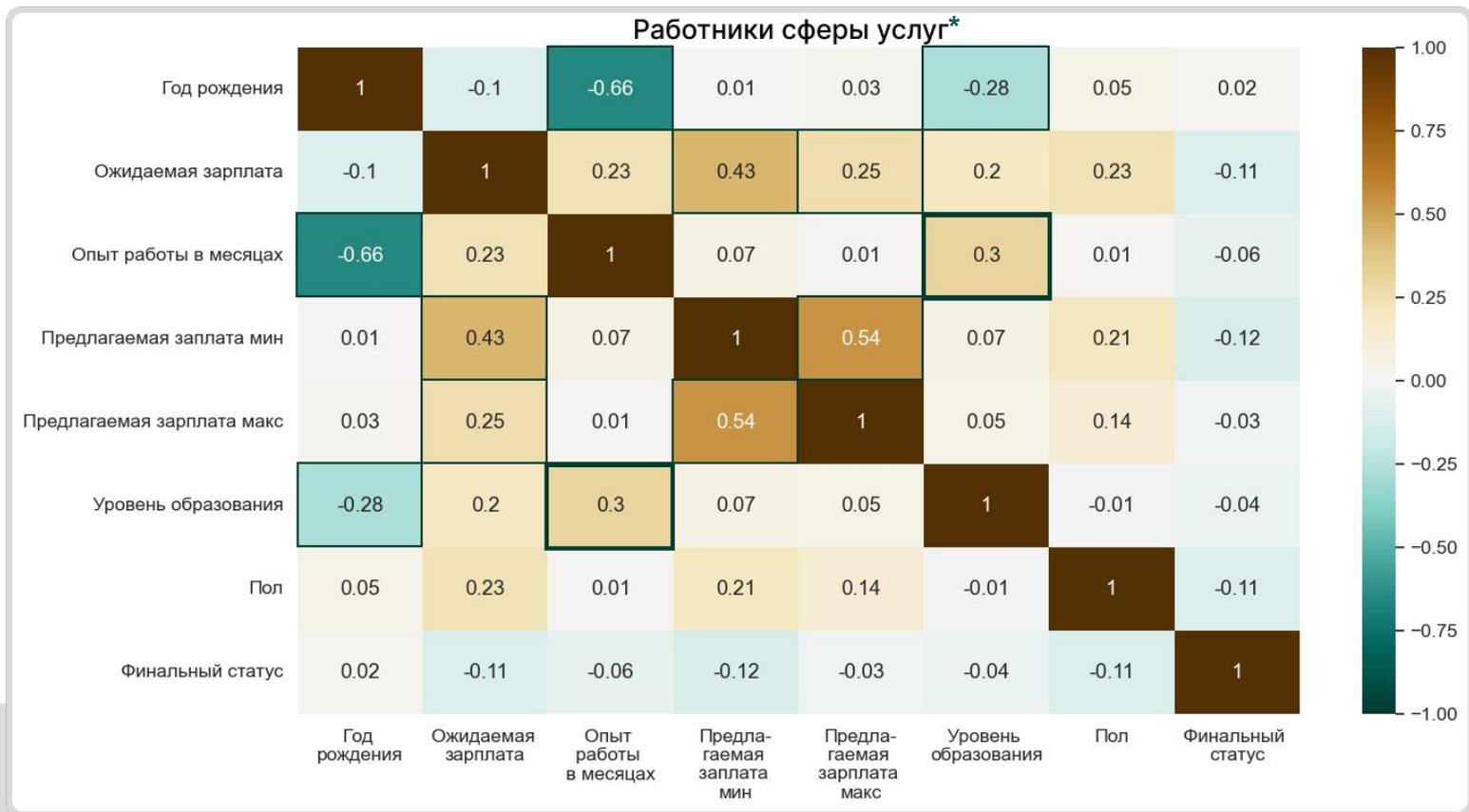
Распределение опыта работы

Предобработка данных: поиск зависимых метрик



*оставшиеся типы профессий в приложении

Предобработка данных: поиск зависимых метрик



*оставшиеся типы профессий в приложении

Исследовательский вопрос и гипотеза

Что для работодателя важнее: трудовой стаж или уровень образования?

На финальный статус взаимодействия между работодателем и соискателем опыт работы влияет сильнее, чем уровень образования.

Механизм

У соискателя есть опыт работы



Приобретены практические навыки



Работодателю не придется тратить дополнительные ресурсы на обучение сотрудника



Человека с большей вероятностью пригласят на собеседование

Соискатель получил образование



Есть теоретические знания



Не являются для работодателя сигналом, что сотрудник успешно справится с предлагаемой работой



На принятие решения повлияют другие факторы

Математическое подтверждение: рассматриваемые переменные

Объясняемая переменная	Принимаемые значения
final status	0 или 1

- Отдельно рассматриваем мужчин и женщин

Объясняющая переменная	Принимаемые значения
work experience	целое неотрицательное число месяцев
education level	0 или 1

Математическое подтверждение: проверка значимости переменных

Нулевая гипотеза

- рассматриваемая переменная **не влияет** на final state

Альтернативная гипотеза

- рассматриваемая переменная **влияет** на final state

Математическое подтверждение: проверка значимости переменных

Значения χ^2 -квadrата для мужчин	Офисные работники
Уровень образования	460.79
Опыт работы	1291.36

Значения χ^2 -квadrата для женщин	Офисные работники
Уровень образования	401.38
Опыт работы	1253.85

*оставшиеся типы профессий и формула в приложении

Уровень значимости $\alpha = 0.05$

Критическое значения критерия χ^2 -квadrат для данного α

- по уровню образования: 3.8
- по опыту работы: 9.5

Математическое подтверждение: стандартизация для сравнения коэффициентов

Объясняющая переменная	Тип переменной
work experience	непрерывная
education level	булевая

Не можем сравнивать
нестандартизованные
переменные **разных** типов

Стандартизируем с помощью
z-value

$$z = (x - \mu) / \sigma$$

- x - значение исследуемого показателя
- μ - среднее значение показателя
- σ - стандартное отклонение

Математическое подтверждение: построение логистической регрессии

Коэффициенты регрессий для мужчин	Уровень образования	Опыт работы	Выполняется ли гипотеза?
Рабочие	-0.423	-0.395	Нет
Офисные работники	-0.238	-0.488	Да
Социальные работники	-	-	-
Технические специалисты	-0.151	-0.109	Нет
Руководители	-0.204	-0.345	Да
Творческие специалисты	-	-0.257	Да

Коэффициенты регрессий для женщин	Уровень образования	Опыт работы	Выполняется ли гипотеза?
Рабочие	-0.268	-0.368	Да
Офисные работники	-0.168	-0.393	Да
Социальные работники	-0.130	-0.207	Да
Технические специалисты	-0.165	-0.098	Нет
Руководители	-0.198	-0.256	Да
Творческие специалисты	-	-0.135	Да

Стандартизированные коэффициенты

Математическое подтверждение: проверка устойчивости модели

Значения χ -квадрата для мужчин	Офисные работники из Москвы	Рабочие из регионов	Случайные 85% от всех программистов
Уровень образования	87.18	6.18	106.97
Опыт работы	1291.36	7.98	185.67

Значения χ -квадрата для женщин	Офисные работники из Москвы	Рабочие из регионов	Случайные 85% от всех программистов
Уровень образования	116.01	3.23	84.99
Опыт работы	552.59	15.84	111.69

Выделены незначимые χ -квадраты

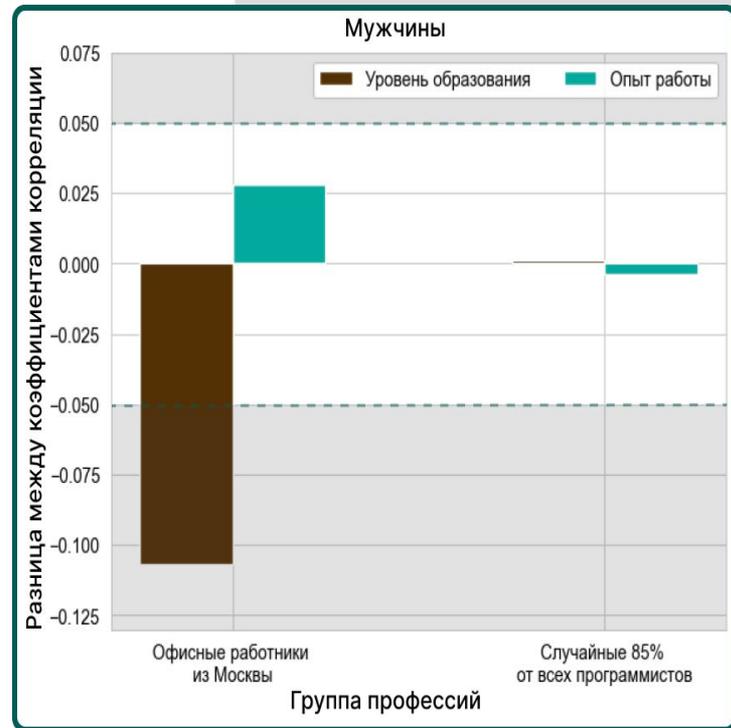
Математическое подтверждение: проверка устойчивости модели

Коэффициенты регрессий для мужчин	Уровень образования	Значимо ли различие по уровню образования?
Офисные работники из Москвы	-0.131	Да
Случайные 85% от всех программистов	-0.152	Нет

Коэффициенты регрессий для мужчин	Опыт работы	Значимо ли различие по опыту работы?
Офисные работники из Москвы	-0.516	Нет
Случайные 85% от всех программистов	-0.105	Нет

Выделены неустойчивые показатели

|Значимая разница| = 0.05



Разница между коэффициентами корреляции

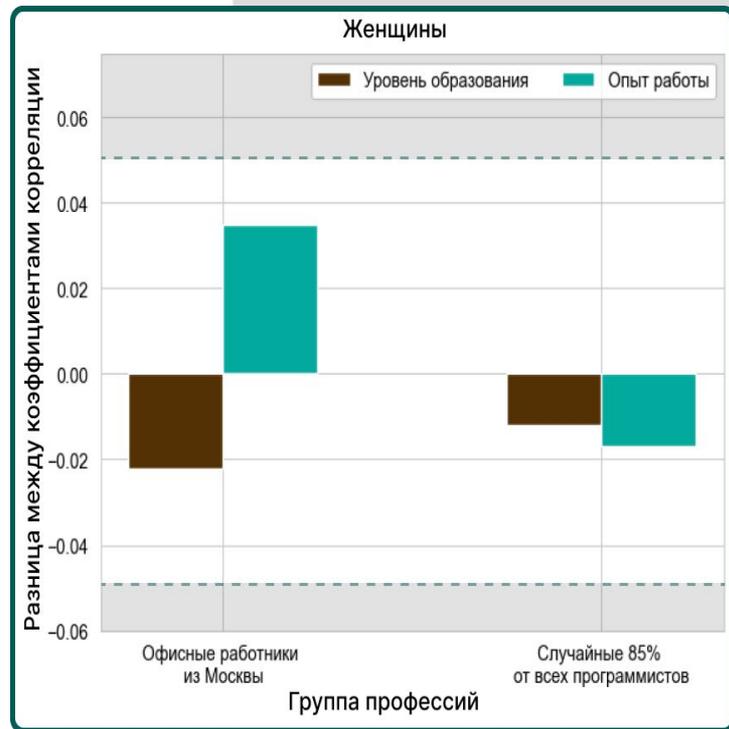
Математическое подтверждение: проверка устойчивости модели

Коэффициенты регрессий для женщин	Уровень образования	Значимо ли различие по уровню образования?
Офисные работники из Москвы	-0.146	Нет
Случайные 85% от всех программистов	-0.153	Нет

Коэффициенты регрессий для женщин	Опыт работы	Значимо ли различие по опыту работы?
Офисные работники из Москвы	-0.428	Нет
Случайные 85% от всех программистов	-0.081	Нет

Нет неустойчивых показателей

|Значимая разница| = 0.05



Разница между
коэффициентами корреляции

Интерпретация результата

- Модель устойчива **не для всех** рассматриваемых групп

- Гипотеза **верна** для **некоторых** комбинаций типов профессий и полов

- Значения найденных показателей **контринтуитивны**

Интерпретация результата

На **final state** и знаки коэффициентов в регрессии может влиять **третий фактор**

Пример: завышенные требования работодателя к квалифицированным сотрудникам

Альтернативный механизм

Большой опыт работы и высокий уровень образования повышают квалификацию соискателя

Работодатель имеет больше требований к квалифицированным сотрудникам

На этап собеседования приглашают меньшее количество соискателей

Ограничения

**Рассмотрен только
летний сезон**

**Маленькое количество
соискателей старше 50 лет**

**Внешние факторы, не
отражённые в переменных**

**Неполный список
профессий**

Перспективы

**Расширить
классификацию
профессий**

**Собрать данные о
результатах собеседований**

**Собрать данные о важных для
каждого работодателя
характеристиках соискателя**

**Собрать данные о
вузах, в которых
соискатели получали
образование**

Применение

- Программа, которая будет предсказывать финальный статус взаимодействия по вводным данным соискателя и работодателя

- Улучшение механизмов рекомендаций вакансий для работодателей и соискателей на hh.ru

Policy implications

Заинтересованные группы

- Государство
- Центры занятости населения
- hh.ru
- Ищущие работу

Суть

- Улучшение качества психологической помощи безработным, которых реже приглашают на собеседования, в центрах занятости населения
- Создание нового сервиса на hh.ru с психологами и экспертами, оказывающими психологическую поддержку контрольной группе

Команда



Ковальчук Артём Игоревич

Программист



Малакшанидзе Анна
Давидовна

Аналитик



Мирошникова Александра
Владимировна

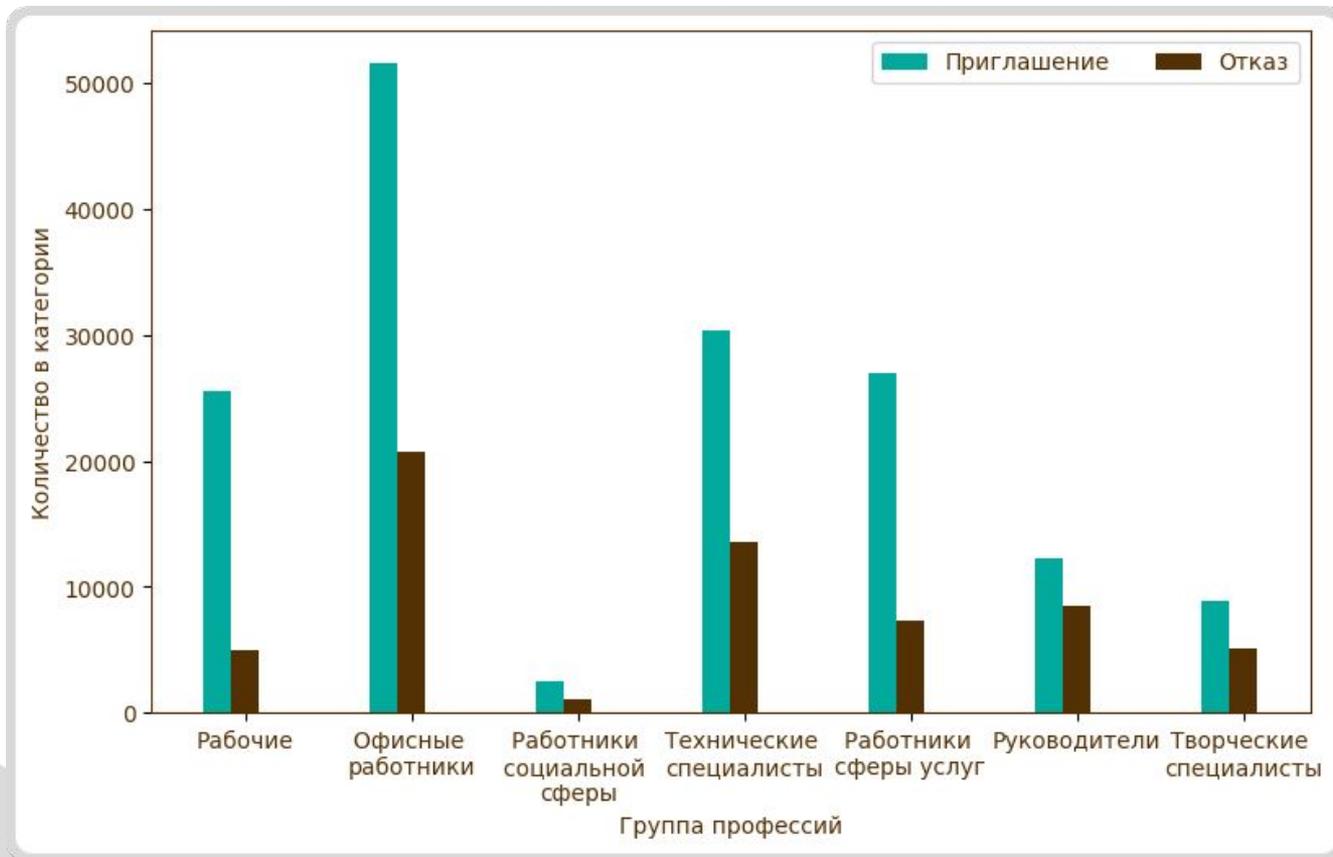
Аналитик



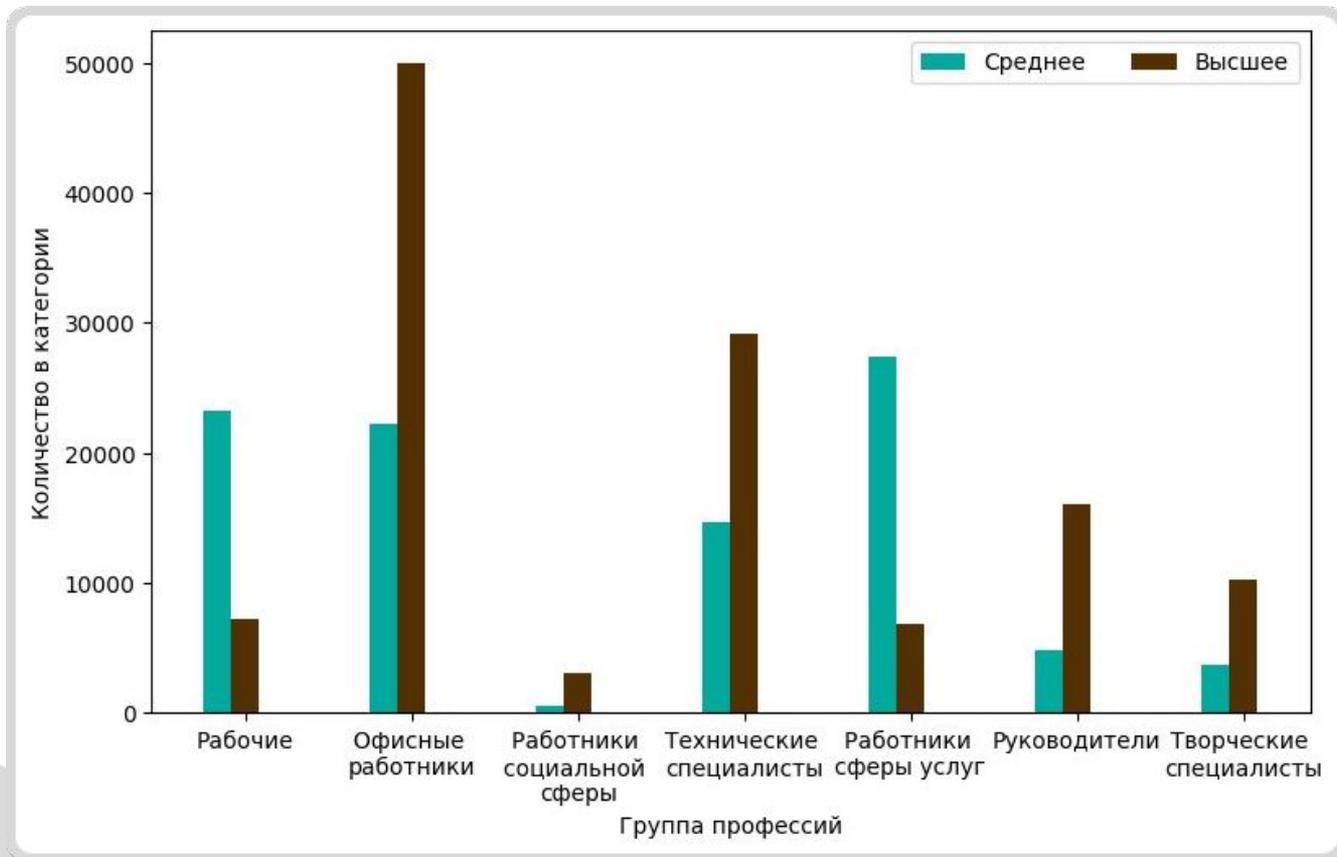
Казакова Евгения Яковлевна

Программист

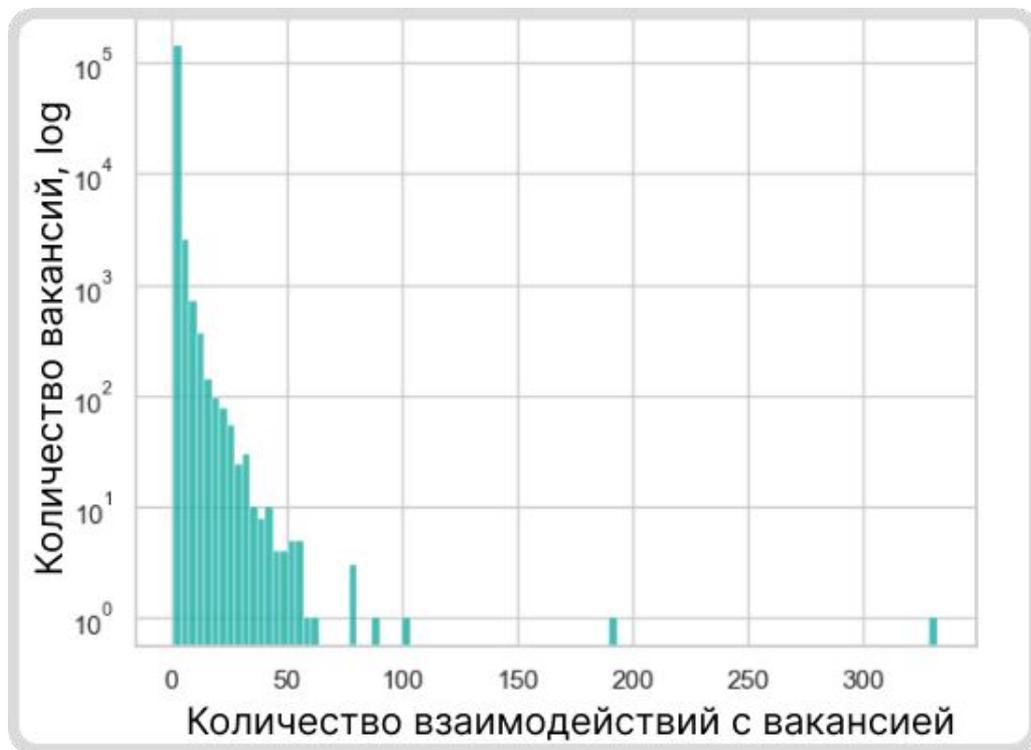
Приложение: распределение финальных статусов



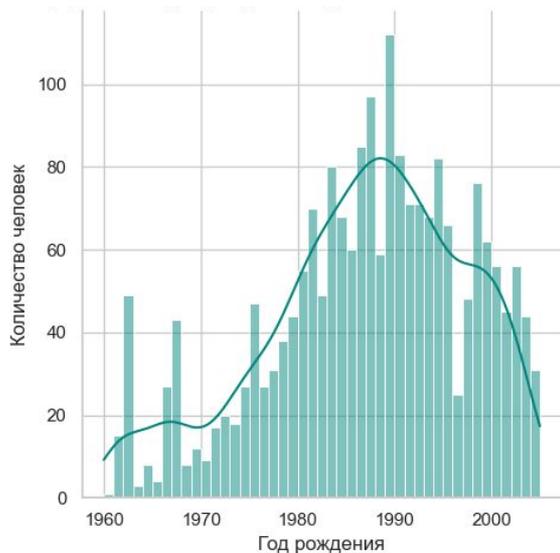
Приложение: распределение уровней образования



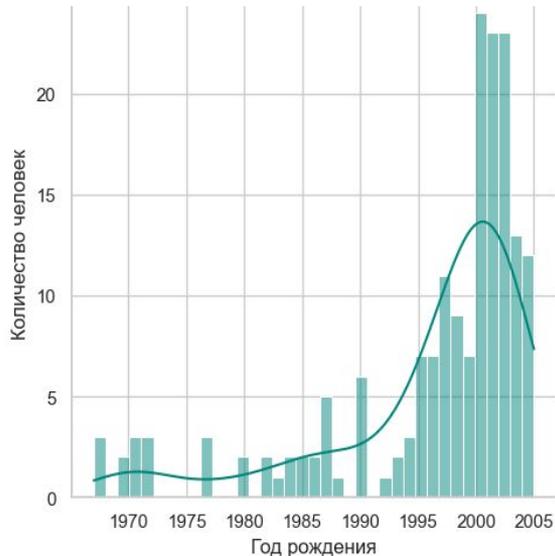
Приложение: распределение вакансий по количеству уникальных взаимодействий



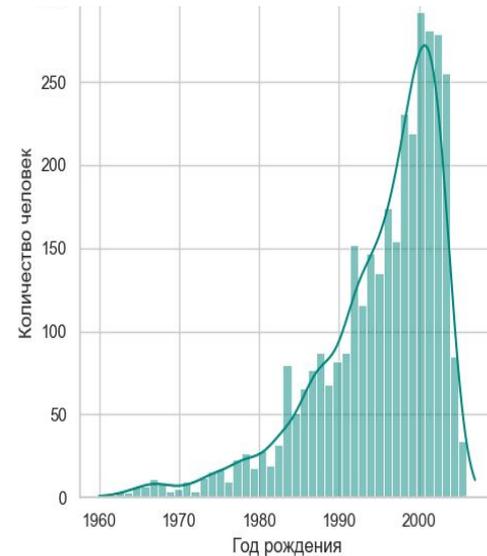
Приложение: распределение неоконченного высшего образования



Рабочие

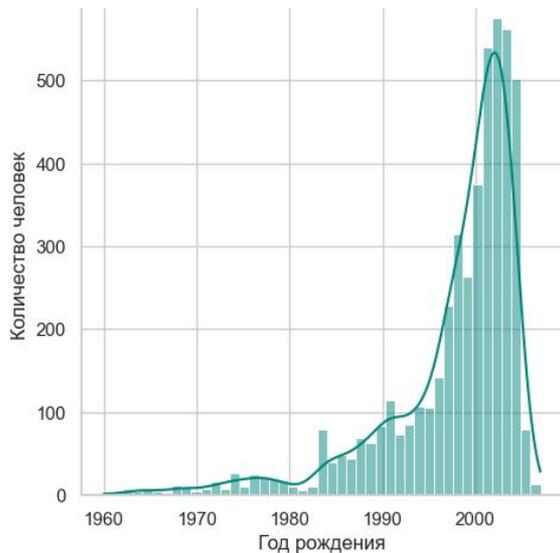


Работники социальной сферы

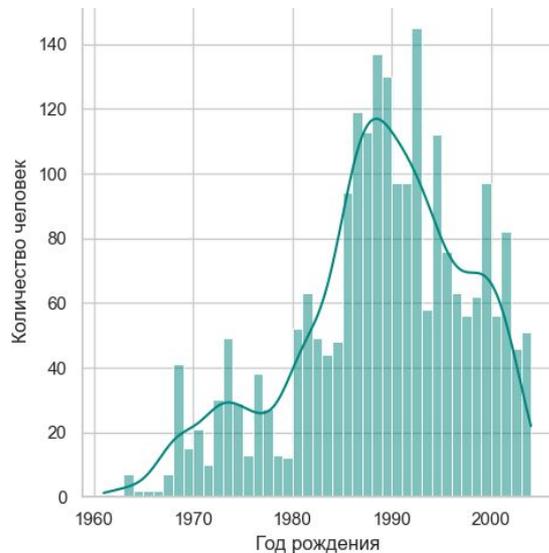


Технические специалисты

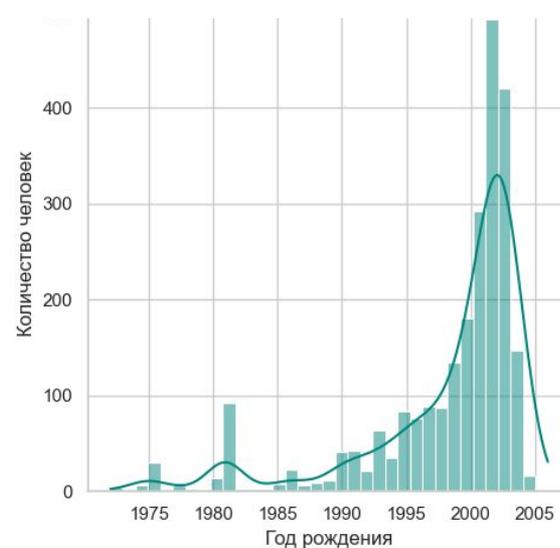
Приложение: распределение неоконченного высшего образования



Работники сферы услуг

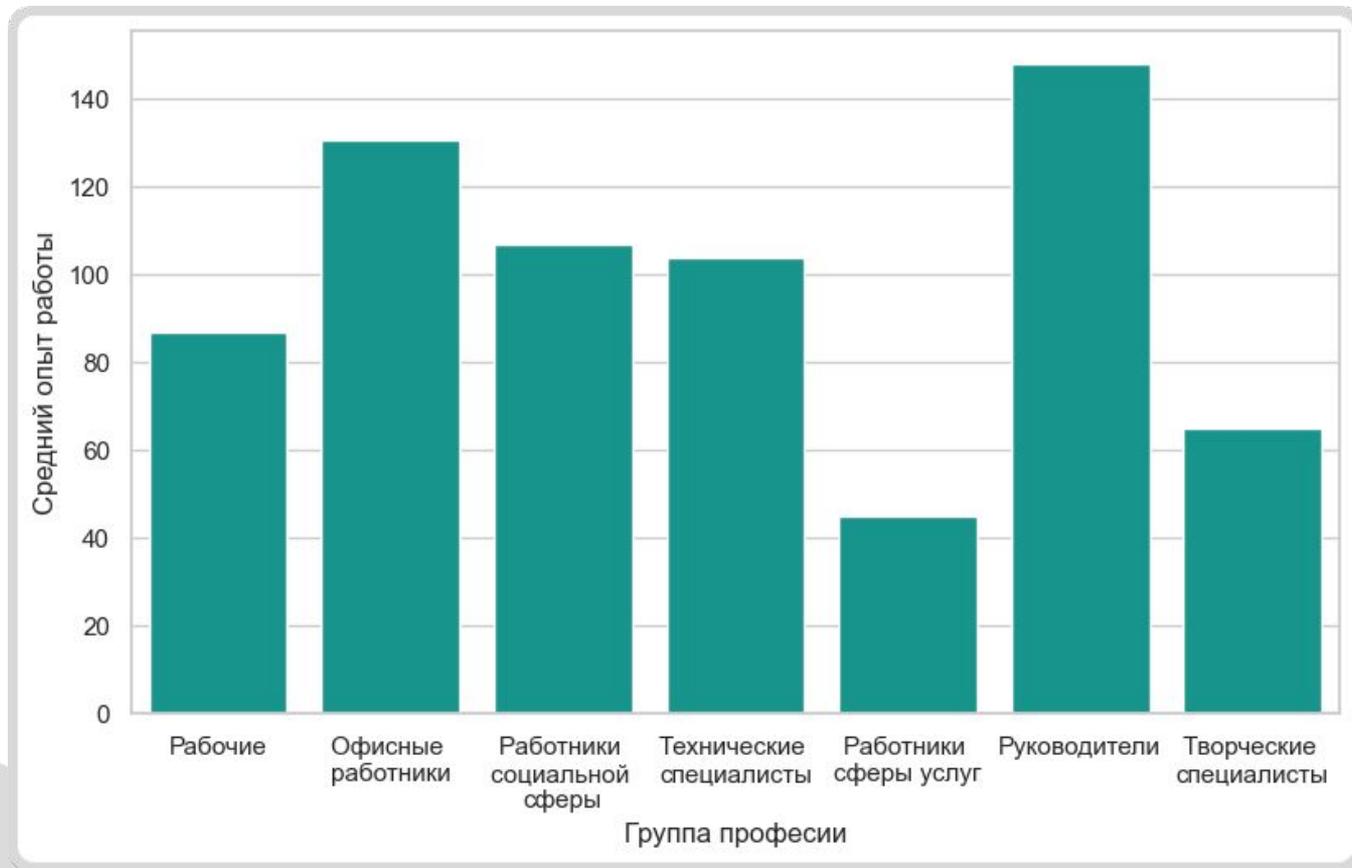


Руководители

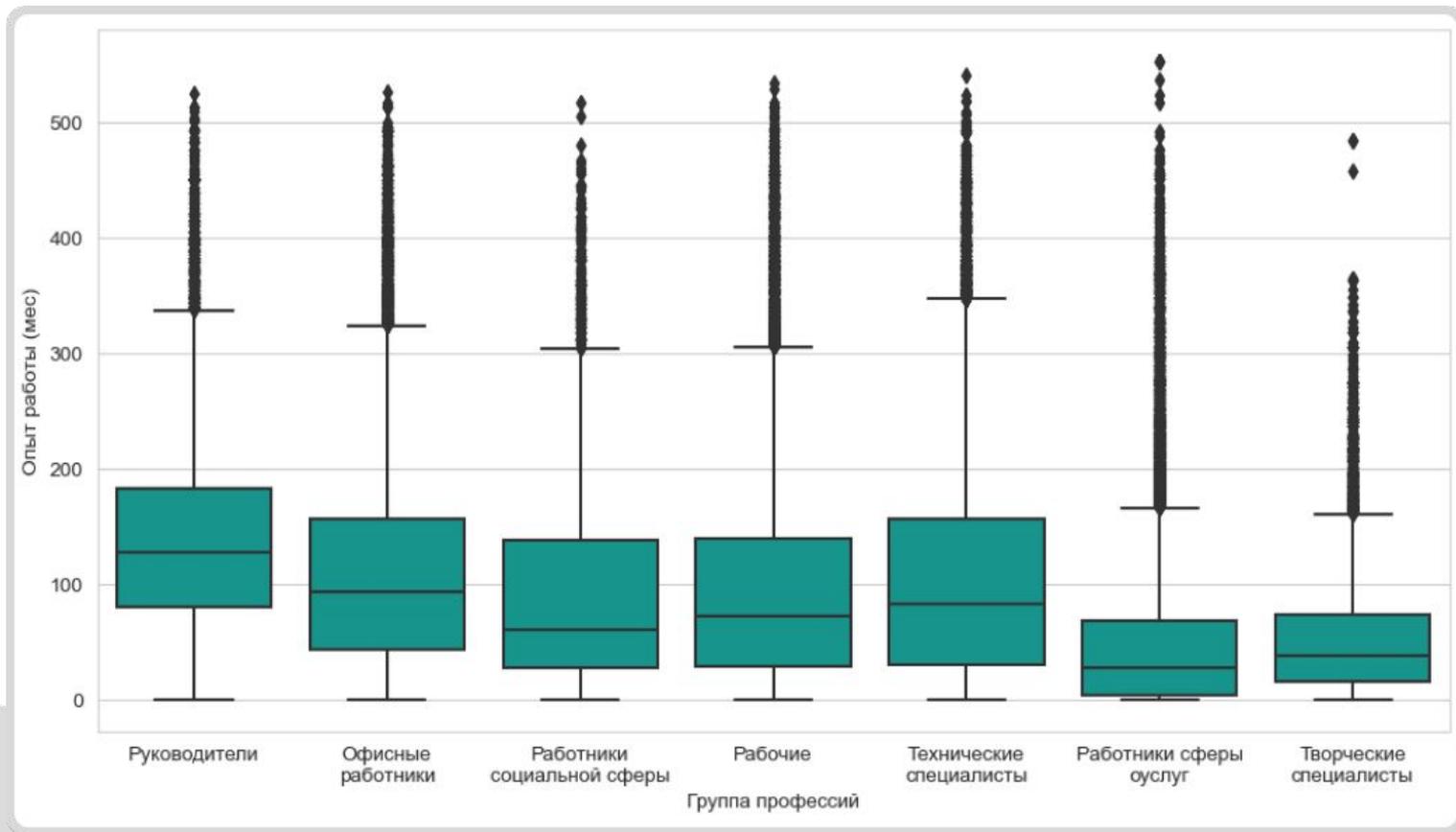


Творческие специалисты

Приложение: средний опыт работы по профессиям



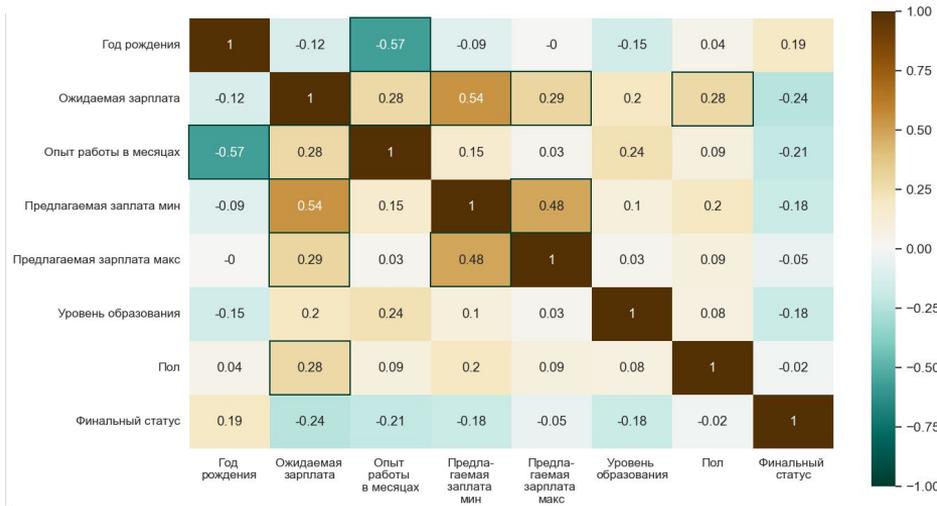
Приложение: анализ распределения опыта работы



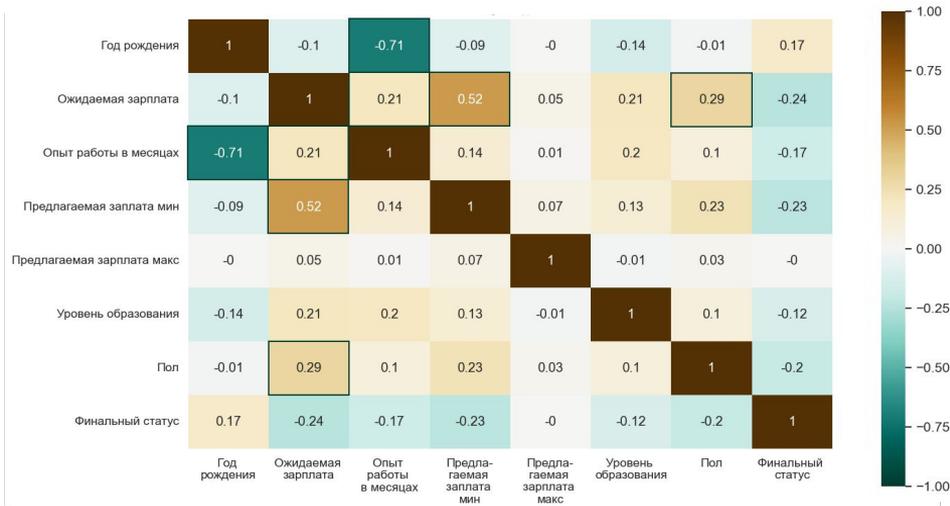
Приложение: классификация по опыту работы

Опыт работы	Офисные работники	Рабочие	Работники социальной сферы	Технические специалисты	Работники сферы услуг	Руководители	Творческие специалисты
Нет опыта	2744	3063	250	2900	7097	130	1257
До 25-ого перцентиля	17566	6951	854	10274	6867	5275	3269
От 25-ого до 50-ого перцентилей	17337	6766	852	10285	6796	5178	3154
От 50-ого до 75-ого перцентилей	17302	6897	838	10244	6726	5117	3122
От 75-ого перцентиля	17236	6800	846	10160	6741	5150	3103

Приложение: поиск зависимых метрик

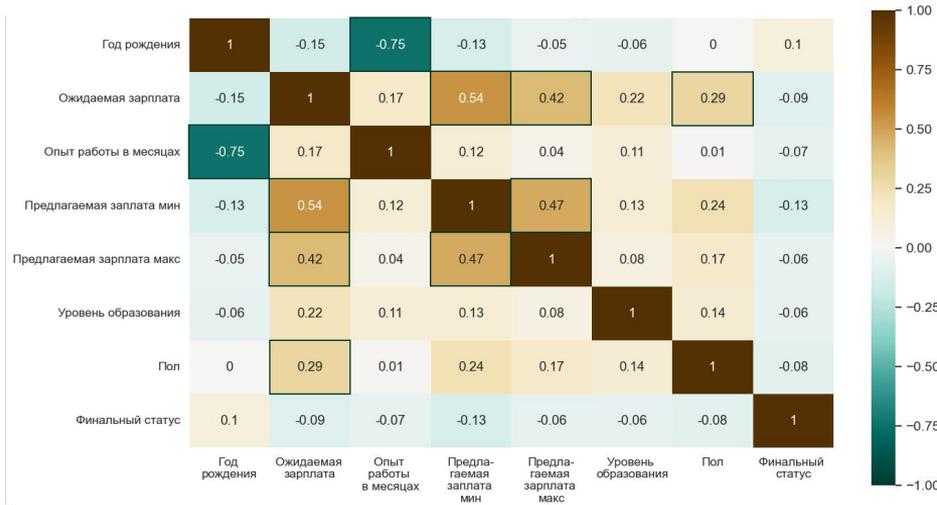


Рабочие

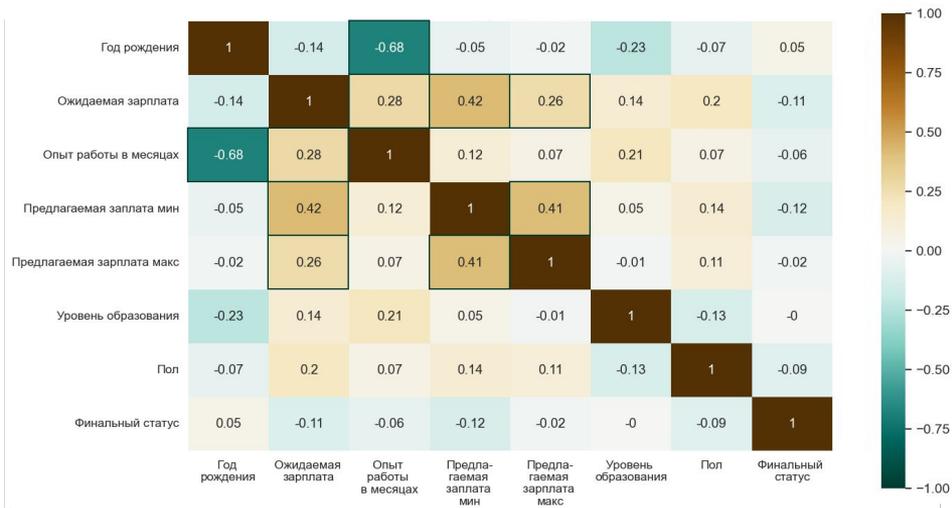


Руководители

Приложение: поиск зависимых метрик

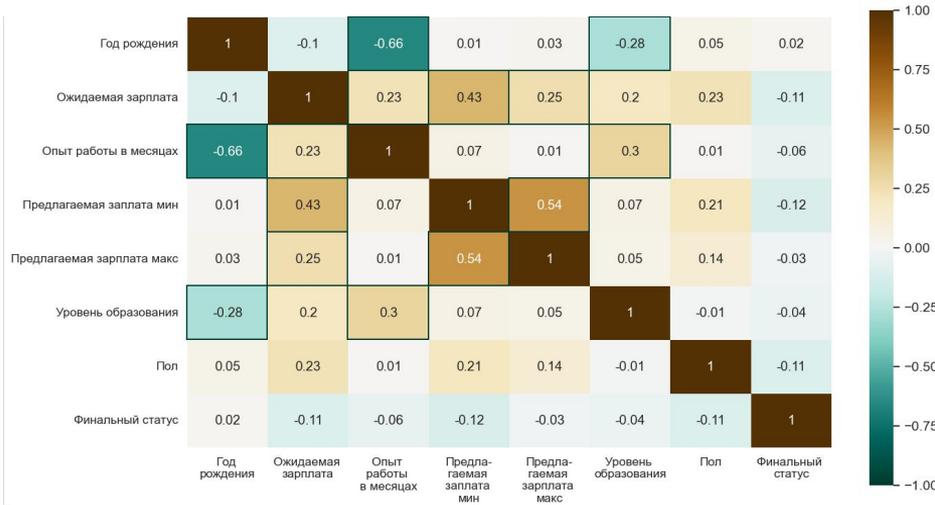


Работники социальной сферы

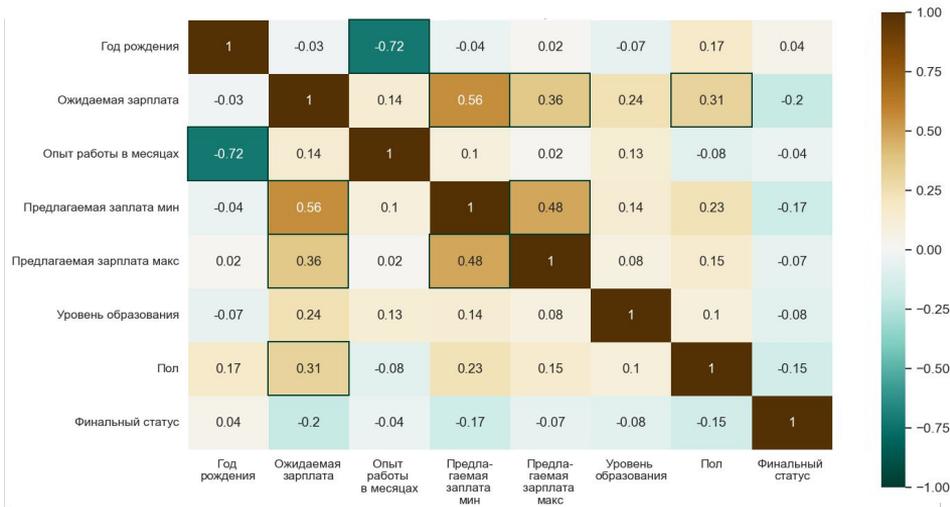


Творческие специалисты

Приложение: поиск зависимых метрик



Работники сферы услуг



Технические специалисты

Приложение: поиск незначимых переменных по критерию χ -квадрат

Мужчины	Офисные работники	Рабочие	Работники социальной сферы	Технические специалисты	Руководители	Творческие специалисты
Уровень образования	460.79	873.07	0.0046	126.48	121.96	1.00
Опыт работы	1291.36	896.56	8.66	212.15	260.27	33.18

Женщины	Офисные работники	Рабочие	Работники социальной сферы	Технические специалисты	Руководители	Творческие специалисты
Уровень образования	401.38	99.86	8.93	93.41	98.53	1.36
Опыт работы	1253.85	303.27	36.58	133.46	187.48	33.35

$$\chi\text{-квадрат} = \sum (O_i - E_i)^2 / E_i$$

- O_i - наблюдаемое значение показателя
- E_i - ожидаемое значение показателя

Выделены незначимые χ -квадраты

Ссылка на Google Диск с файлами

- <https://drive.google.com/drive/folders/1LipOan5qk5ZAfKzaywqMVYYxjgW73K5q?usp=sharing>

