



Уговор дорожке денег

DA, NO есть нюанс

Теребов Михаил
Турицына Ульяна
Сафонов Арсений

Чепелюк Илья
Новолаев Никита





01

Разведочный анализ



319 481


Уникальных вакансий

189 402

Уникальных резюме

499 949

Уникальных взаимодействий

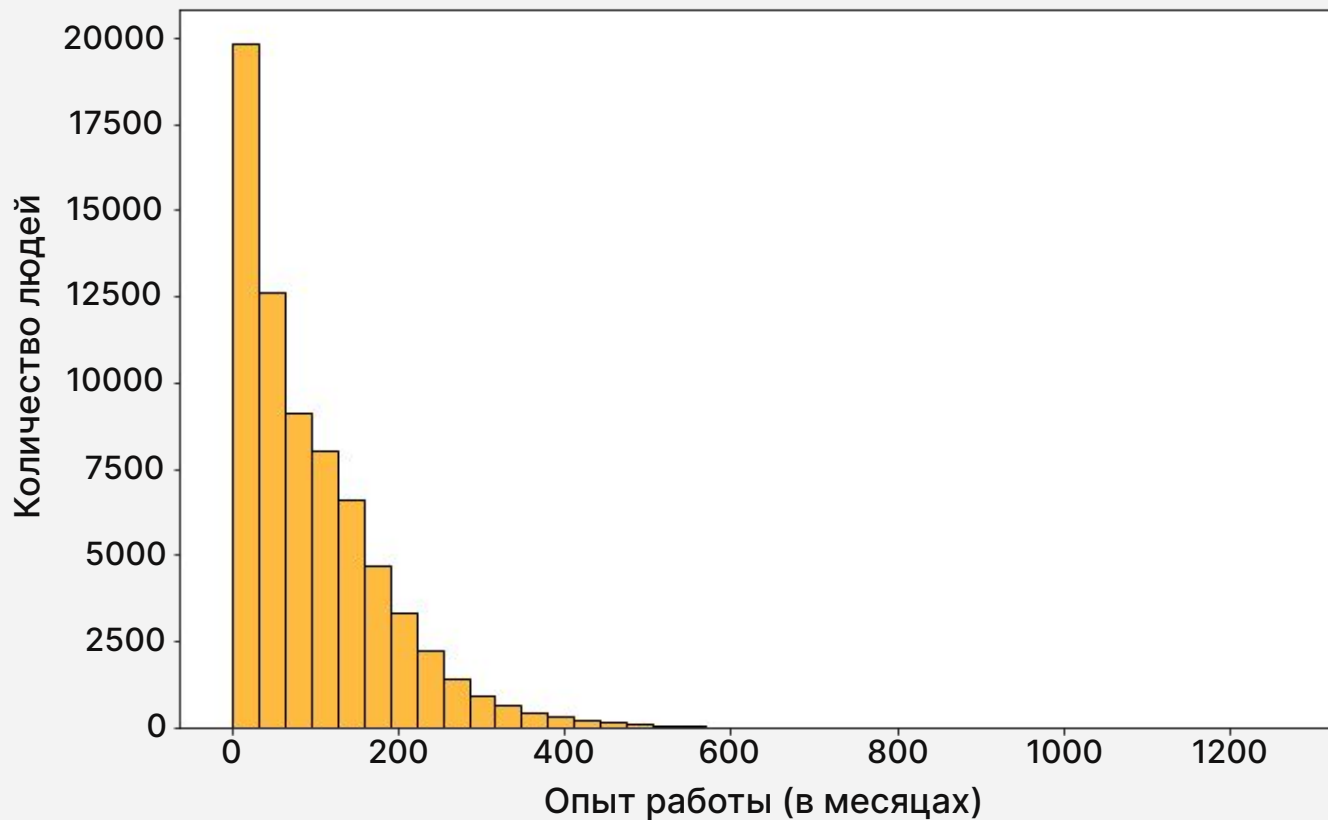


Структура базы данных

Качественные	Количественные
первоначальный статус взаимодействия	желаемый уровень зарплаты соискателя
финальный статус взаимодействия	год рождения соискателя
профессия	кол-во месяцев рабочего опыта
уровень образования соискателя	предлагаемая зарплата вакансии (от-до)
пол	

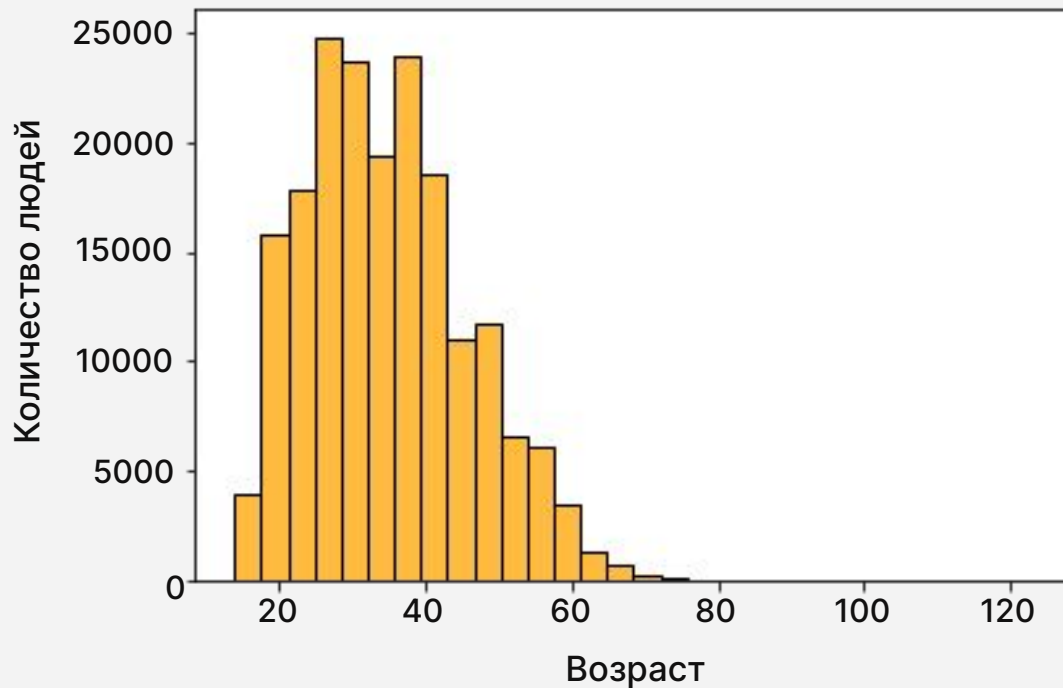
 Указаны только данные, использующиеся в исследовании

Распределение по опыту работы

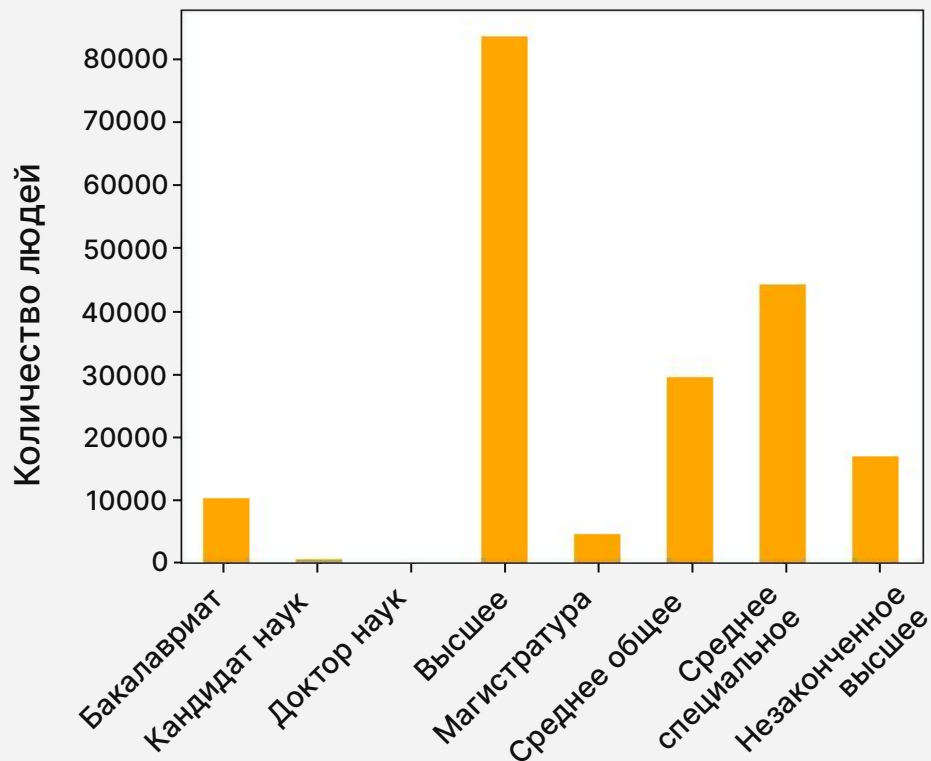




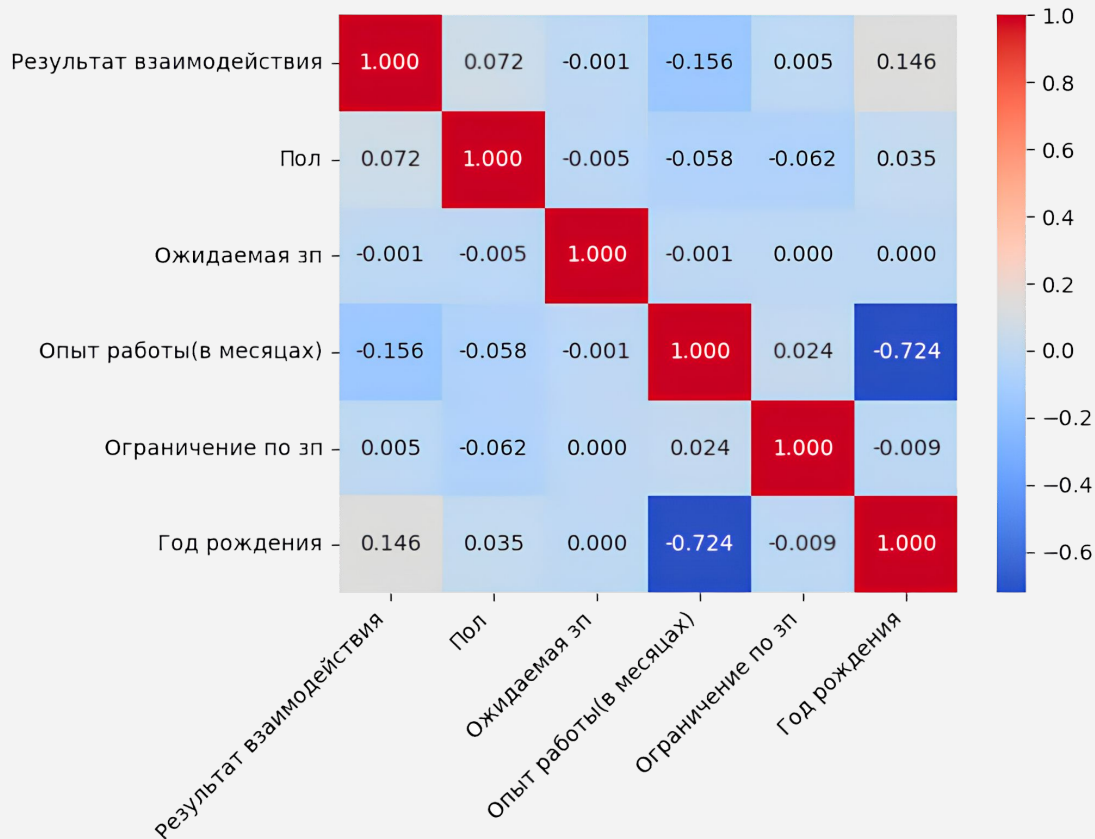
Распределение по возрасту



Распределение по уровню образования




Матрица корреляций





02

Постановка ГИПОТЕЗЫ



Исследовательский вопрос и гипотеза

Исследовательский вопрос

Как **отклонение** ожидаемой заработной платы от среднерыночной **влияет на приглашение** на собеседование?

Гипотеза

Отклонение ожидаемой заработной платы **негативно влияет** на приглашение на собеседование **не во всех сферах деятельности**



Механизм



В некоторых отраслях
наблюдается дефицит
квалифицированных
кадров

Работодатель
стремится пригласить
на собеседование как
можно больше людей

На собеседование
проходят даже кандидаты
с завышенными
зарплатными ожиданиями



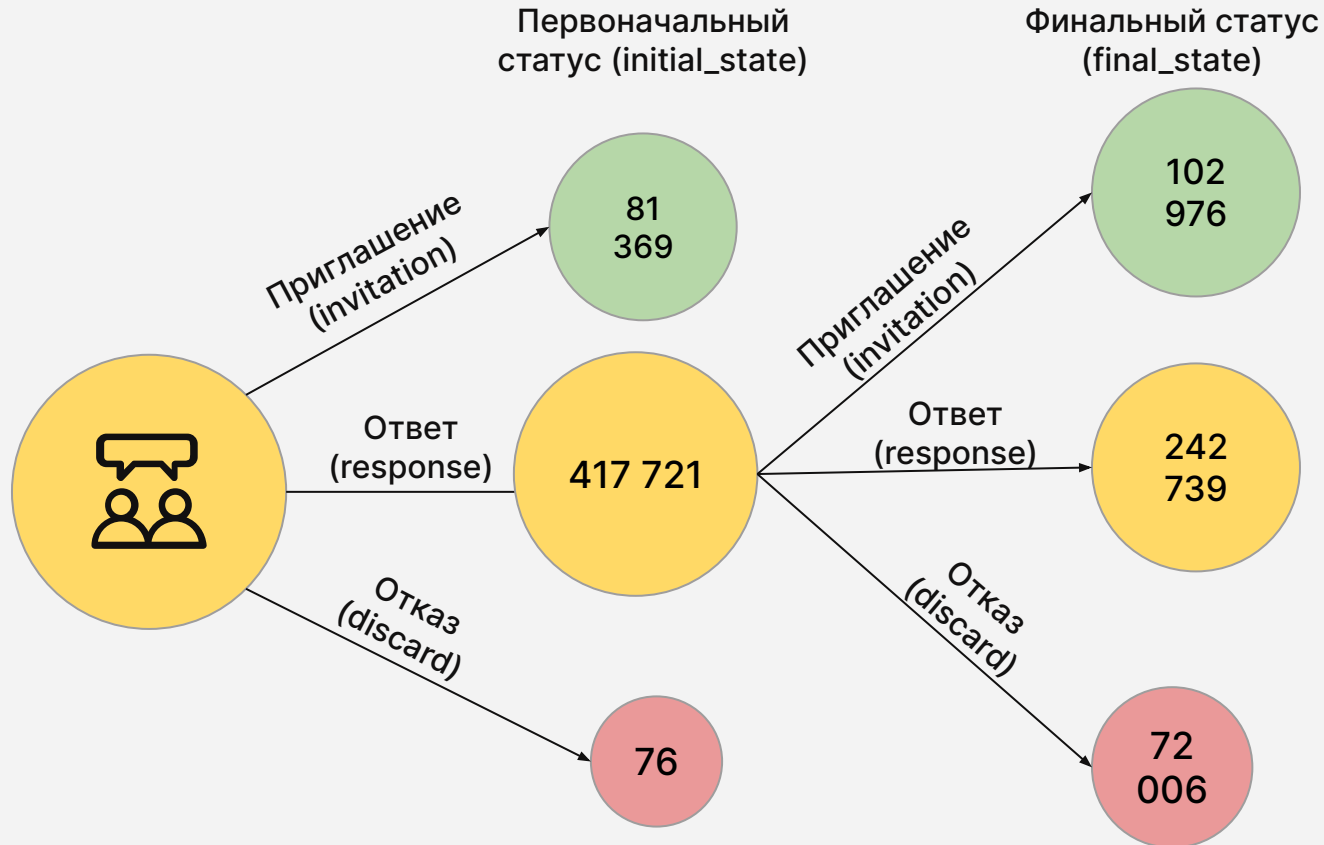


03

Подготовка данных



Статусы взаимодействия



Итого
184 345
242 739
72 082



Очистка от выбросов

Убираем взаимодействия со статусом response-response

Убираем резюме с ожидаемой зарплатой выше quantile 0.95

Убираем вакансии с верхней границей зарплаты выше quantile 0.95 (или где она вообще отсутствует)

Убираем резюме, если опыт меньше возраста менее чем на 14 лет

Очистка не повлияла на общую структуру данных (распределение полов, рабочего опыта, возрастов, уровня образования)*



73 913

(-245 568 от исходного кол-ва)

Уникальных вакансий

70 638

(-118 764 от исходного кол-ва)

Уникальных резюме

113 276

(-386 673 от исходного кол-ва)

Уникальных взаимодействий

Вводим новые переменные

**Уровень
образования**

Значение	Соответствующие уровни образования
0	Среднее общее
1	Среднее специальное
2	Незаконченное высшее
3	Высшее, Бакалавриат
4	Магистратура, Кандидат наук, Доктор наук



Вводим новые переменные

Медианная зарплата

1. Для каждой вакансии находим **среднее арифметическое** между нижней и верхней границей зарплаты
2. Полученные значения **разделяем по профессиям**
3. Берем **медиану** из средних в каждой профессии



Вводим новые переменные

**Относительное
отклонение**



(Ожидаемая зарплата - Медианная зарплата)

Медианная зарплата



04

Проверка ГИПОТЕЗЫ



Проверка гипотезы

Нулевая гипотеза: во всех сферах работы отклонение от средней зарплаты негативно влияет на приглашение на собеседование, с учетом опыта и образования

Альтернативная гипотеза: не во всех сферах работы отклонение от средней зарплаты негативно влияет на приглашение на собеседование, с учетом опыта и образования





Проверка гипотезы

Т.к целевая переменная бинарная, где **1** - **успех**(приглашение на собеседование), а **0** - **неудача**(отказ)

Для определения уровня значимости относительного отклонения ожидаемой зарплаты от медианной по сферам используем модель логистической регрессии из библиотеки statsmodels.

При первом просмотре строили модель для каждой из сфер деятельности.

Передадим в модель следующие переменные:

у	х
Финальный статус взаимодействия	Относительное отклонение , опыт работы, уровень образования






Проверка гипотезы



Причины выбора логистической регрессии в качестве модели:

- Бинарная критериальная переменная
 - Возможность оценивать влияние отклонения от зп на результат независимо от опыта работы и уровня образования
 - Возможность получить подробную информацию о влиянии свободных переменных
- 

Проверка гипотезы

Сфера деятельности	Z-value	p-value	Количество наблюдений	Значимость отклонения
IT	-22.968	0.0	14 708	Yes
Услуги	-10.584	0.0	39 321	Yes
Медицина	-3.506	0.0	1 141	Yes
Правовая деятельность	-17.996	0.0	8 291	Yes
Дизайн	-4.629	0.0	858	Yes
Образование	-5.344	0.0	981	Yes
Промышленность	-2.23	0.026	12 226	Yes
Менеджмент	-22.626	0.0	35 750	Yes

* p-уровень значимости - 0.05



Устойчивость матмодели

При разделении данных на группы по различным параметрам датасета, **различий между ними найдено не было**, кроме столбца “Тип трудоустройства вакансии” :

Для вакансий с полным трудоустройством есть сильная зависимость результата взаимодействия от отклонения от зарплаты, для остальных - **зависимости нет**

Матмодель устойчива





Гипотеза опровергается



Гипотеза опровергается

Мы поняли, что во всех сферах деятельности отклонение зарплатных ожиданий влияет на вероятность пройти на собеседование



Уточним гипотезу

Существуют профессии, в которых отклонение зарплатных ожиданий не влияет на приглашение на собеседование

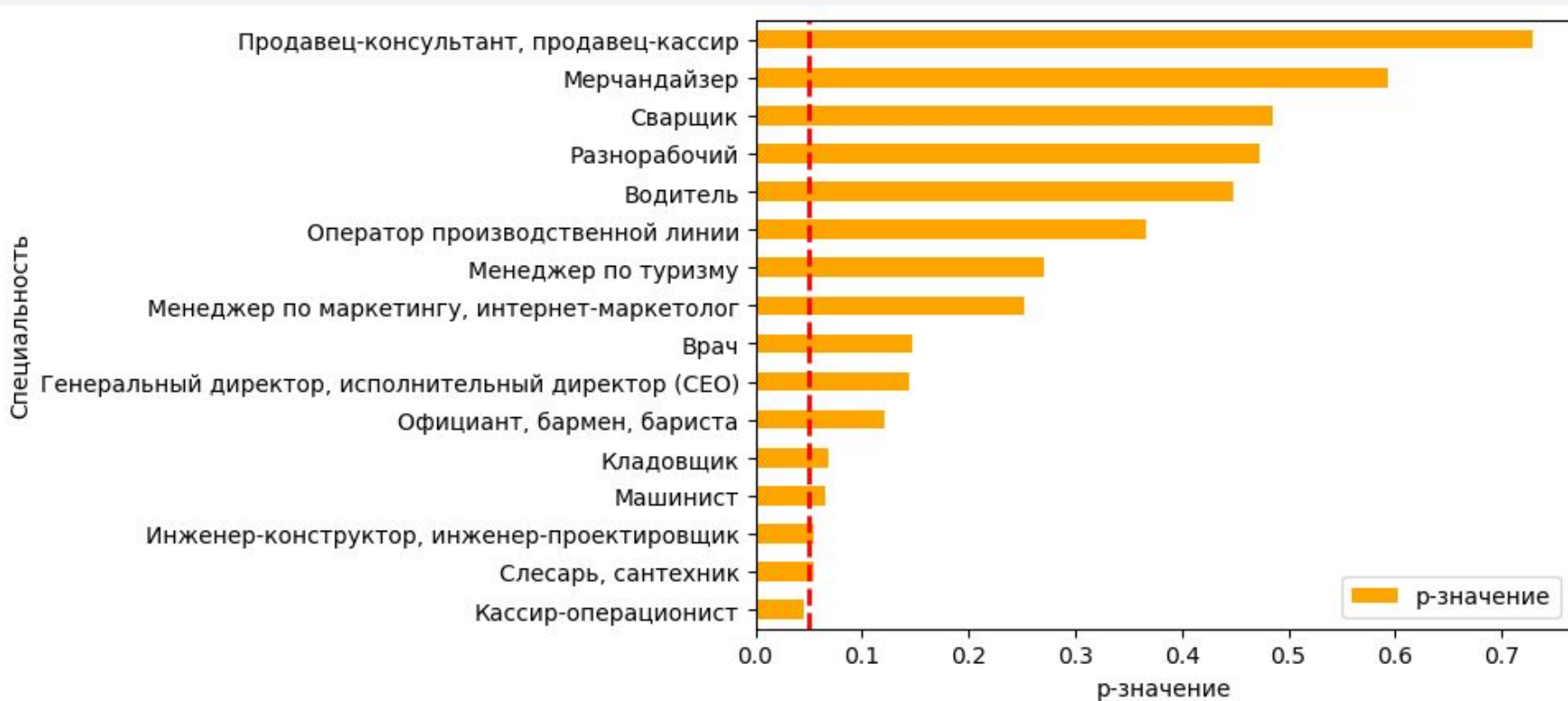


Проверка гипотезы для профессий

Специальность	p-значение	Средняя зп для специальности	Количество наблюдений	Значимость отклонения
Менеджер по маркетингу, интернет-маркетолог	0.253	72500.0	1232	No
Официант, бармен, бариста	0.121	64000.0	9917	No
Водитель	0.449	85000.0	382	No
Сварщик	0.485	90000.0	1008	No
Инженер-конструктор, инженер-проектировщик	0.055	85000.0	3016	No
Слесарь, сантехник	0.054	67500.0	916	No
Врач	0.147	100000.0	572	No
Мерчандайзер	0.593	49575.0	752	No
Кладовщик	0.068	61000.0	443	No
Продавец-консультант, продавец-кассир	0.729	52725.0	559	No
Генеральный директор, исполнительный директор ...	0.145	140000.0	250	No
Менеджер по туризму	0.271	64500.0	574	No
Оператор производственной линии	0.367	60000.0	1035	No
Машинист	0.065	120000.0	839	No
Разнорабочий	0.473	64712.5	755	No

* p-уровень значимости - 0.05

Проверка гипотезы для профессий



Интерпретация результатов

1. В сферах деятельности **наблюдается зависимость** между относительным **отклонением** ожидаемой зарплаты и **вероятностью приглашения** на собеседование

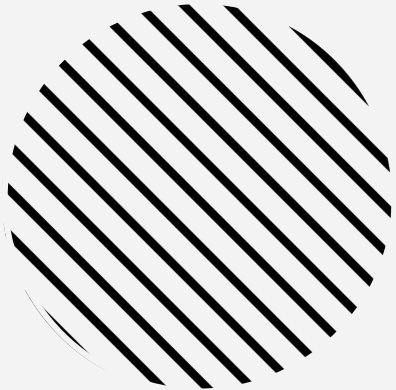
2. При подробном рассмотрении профессий обнаружено, что такая зависимость отсутствует для:

- Профессий с дефицитом квалифицированных кадров
- Профессий со особенностями оплаты труда
- Профессий с высокой текучестью кадров



05

**Ограничения,
польза, перспективы**



Ограничения

Образование

1. “Высшее” без уточнений
2. Нет информации о ВУЗе

Опыт работы

1. Релевантен ли опыт
2. Как давно был получен

Временной промежуток

Невозможно оценить влияние сезонности

Результаты собеседования

Нет информации был ли соискатель принят и на каких условиях



Польза и перспективы



Польза

Модель, предсказывающая вероятность приглашения на собеседование относительно завышения ожидаемой зарплаты с учетом опыта и образования



Перспективы

Собрать данные:

1. о результатах собеседований
2. о регионе и ВУЗе, в котором получено образование
3. об опыте соискателя

Учесть другие показатели: навыки, регион. Рассмотреть данные за большой временной промежуток.



Policy implication

Рекомендательная система для сервисов подбора персонала:

01

Помогает
соискателям в
составлении резюме



02

Сервис становится
удобнее



03

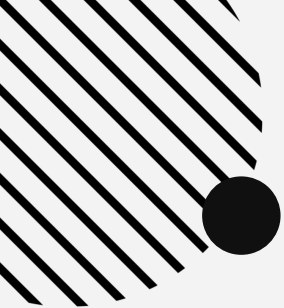
Им пользуется
больше людей



04



Это приносит
большую прибыль
компании





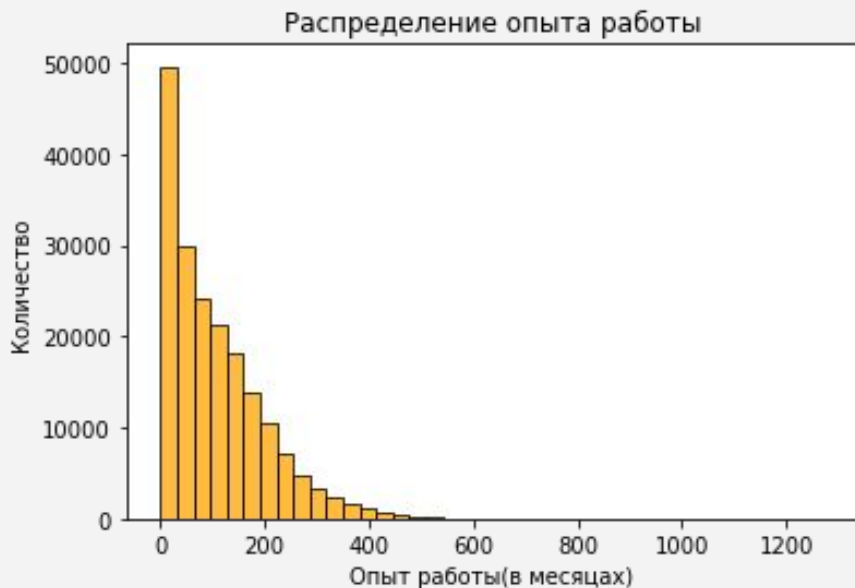
Приложение

Проверка на изменения: распределение полов

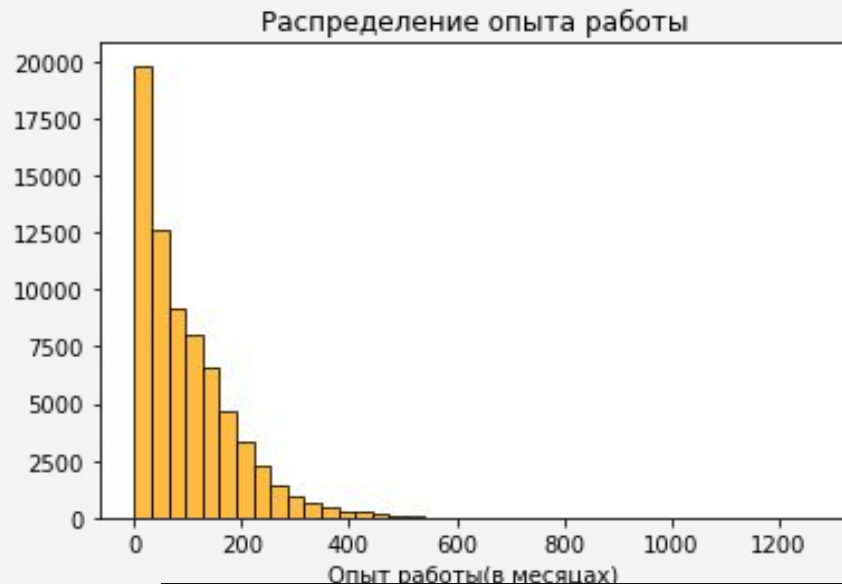
	ДО обработки	ПОСЛЕ обработки
 Мужчин	47%	44%
 Женщин	53%	56%

Проверка на изменения: распределение рабочего опыта

ДО обработки



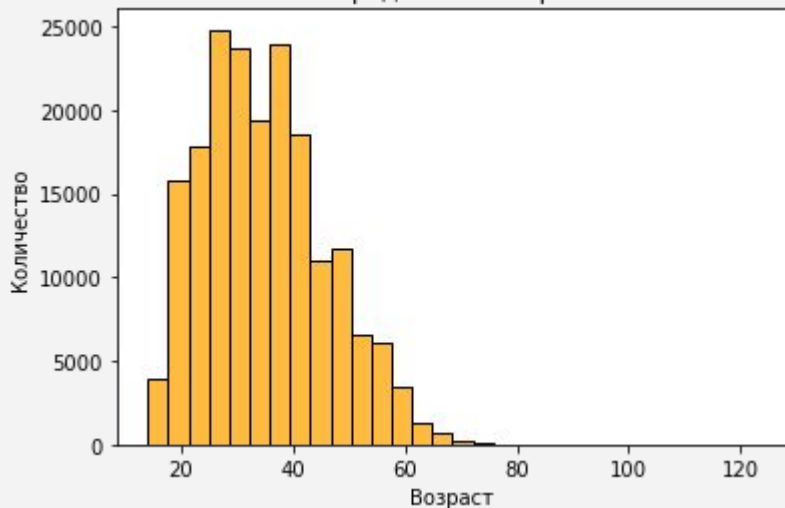
ПОСЛЕ обработки



Проверка на изменения: распределение возраста

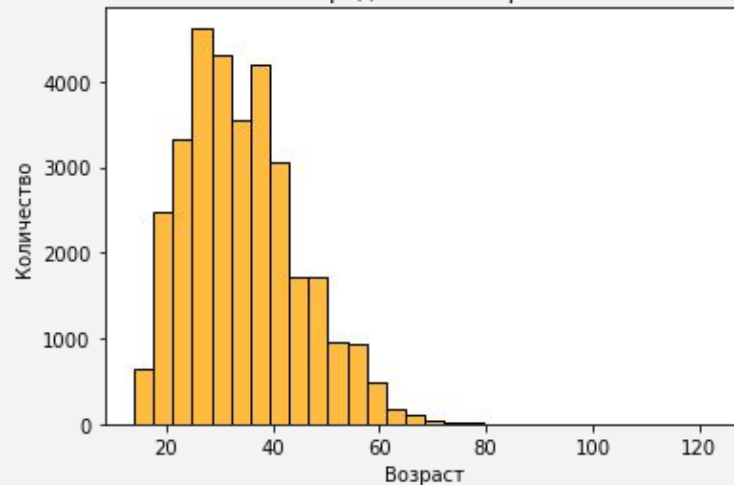
ДО обработки

Распределение возраста



ПОСЛЕ обработки

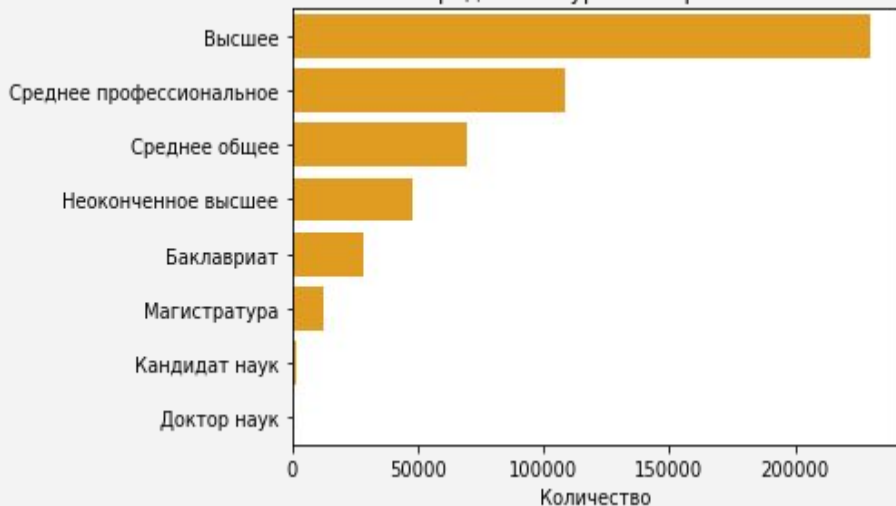
Распределение возраста



Проверка на изменения: распределение уровня образования

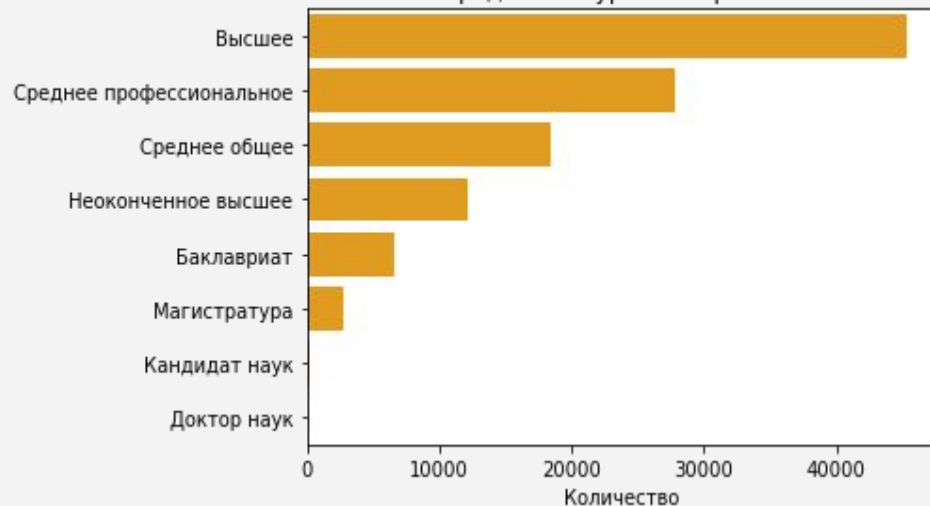
ДО обработки

Распределение уровня образования

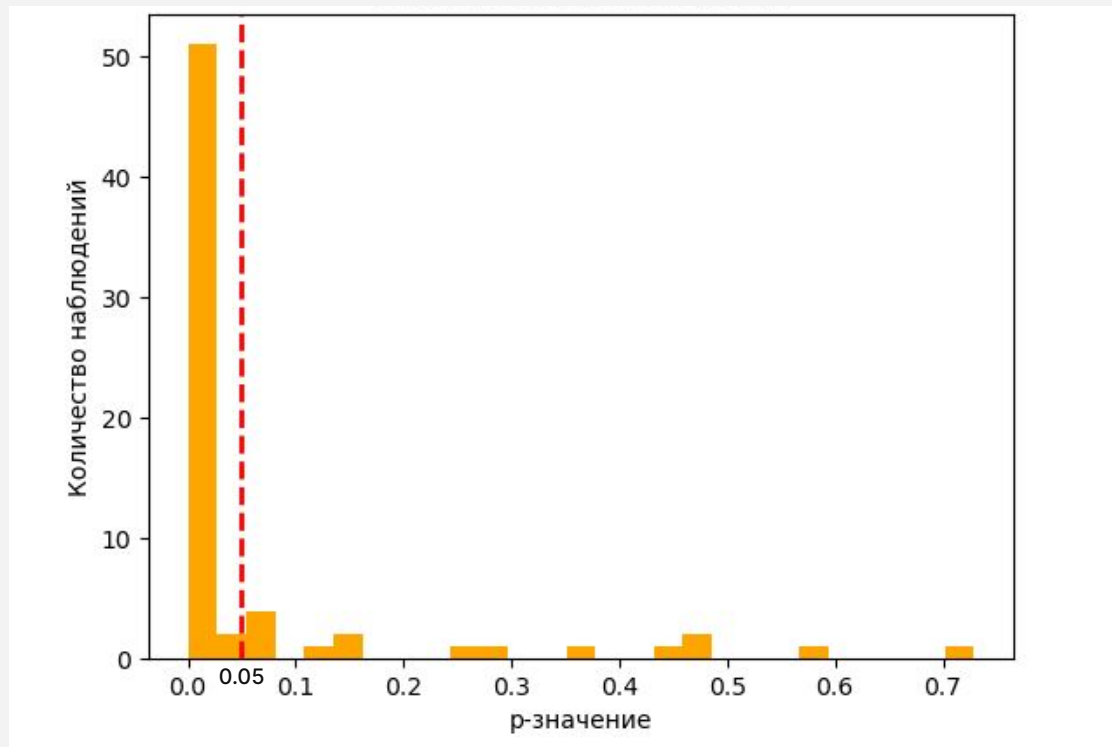


ПОСЛЕ обработки

Распределение уровня образования



Распределение Р-значений



Github



<https://github.com/DA,NOestnuans>



1	Дизайнер, художник	0.000	60000.0	858	YES
2	Программист, разработчик	0.040	90000.0	1089	YES
3	Бухгалтер	0.000	58875.0	2183	YES
4	Менеджер по маркетингу, интернет-маркетолог	0.253	72500.0	1232	No
5	Официант, бармен, бариста	0.121	64000.0	9917	No
6	Менеджер по персоналу	0.008	65000.0	889	YES
7	Менеджер по логистике, менеджер по ВЭД	0.000	75000.0	3430	YES
8	Оператор ПК, оператор базы данных	0.000	47500.0	3007	YES
9	Специалист по подбору персонала	0.001	64250.0	1201	YES
10	Менеджер по закупкам	0.000	75000.0	1914	YES
11	Директор магазина, директор сети магазинов	0.000	80000.0	5162	YES
12	Делопроизводитель, архивариус	0.000	53750.0	2417	YES
13	Юрисконсульт	0.000	65000.0	4481	YES
14	Инженер-энергетик, инженер-электрик	0.000	90000.0	807	YES
15	Руководитель отдела продаж	0.000	135000.0	1295	YES
16	Специалист технической поддержки	0.000	50000.0	2860	YES
17	Водитель	0.449	85000.0	382	No
18	Учитель, преподаватель, педагог	0.002	52500.0	458	YES
19	Офис-менеджер	0.004	55000.0	491	YES
20	Охранник	0.011	59050.0	440	YES

20	Охранник	0.011	59050.0	440	YES
21	Воспитатель, няня	0.000	55000.0	523	YES
22	Сварщик	0.485	90000.0	1008	No
23	Экономист	0.000	62500.0	513	YES
24	Психолог	0.000	57500.0	319	YES
25	Системный администратор	0.002	61500.0	902	YES
26	Курьер	0.000	80000.0	2466	YES
27	Секретарь, помощник руководителя, ассистент	0.000	65000.0	408	YES
28	SMM-менеджер, контент-менеджер	0.000	55000.0	5725	YES
29	Тестировщик	0.000	59750.0	3146	YES
30	Руководитель проектов	0.000	100000.0	1527	YES
31	Инженер-конструктор, инженер-проектировщик	0.055	85000.0	3016	No
32	Супервайзер	0.000	90125.0	745	YES
33	Финансовый менеджер	0.000	80000.0	712	YES
34	Кассир-операционист	0.046	50000.0	389	YES
35	Слесарь, сантехник	0.054	67500.0	916	No
36	Диспетчер	0.000	57500.0	4625	YES
37	Торговый представитель	0.003	82700.0	1666	YES
38	Упаковщик, комплектовщик	0.013	60000.0	4183	YES
39	Врач	0.147	100000.0	572	No

	Специальность	p-значение	Средняя зп для специальности	Количество наблюдений	Значимость отклонения
39	Врач	0.147	100000.0	572	No
40	Повар, пекарь, кондитер	0.000	65000.0	1968	YES
41	Администратор	0.003	57500.0	419	YES
42	Мерчандайзер	0.593	49575.0	752	No
43	Аналитик	0.000	70000.0	839	YES
44	Кладовщик	0.068	61000.0	443	No
45	Начальник производства	0.000	110000.0	756	YES
46	Юрист	0.000	70000.0	232	YES
47	Начальник смены, мастер участка	0.000	80000.0	1270	YES
48	Продавец-консультант, продавец-кассир	0.729	52725.0	559	No
49	Генеральный директор, исполнительный директор ...	0.145	140000.0	250	No
50	Менеджер по туризму	0.271	64500.0	574	No
51	Координатор отдела продаж	0.000	65000.0	2531	YES
52	Оператор производственной линии	0.367	60000.0	1035	No
53	Технолог	0.000	75000.0	1588	YES

	Специальность	p-значение	Средняя зп для специальности	Количество наблюдений	Значимость отклонения
53	Технолог	0.000	75000.0	1588	YES
54	Машинист	0.065	120000.0	839	No
55	Менеджер ресторана	0.000	81500.0	825	YES
56	Начальник склада	0.000	76900.0	2022	YES
57	Фитнес-тренер, инструктор тренажерного зала	0.003	75000.0	1539	YES
58	Товаровед	0.006	55000.0	1088	YES
59	Оператор call-центра, специалист контактного ц...	0.003	50000.0	2037	YES
60	Механик	0.002	85000.0	269	YES
61	Сервисный инженер, инженер-механик	0.000	80000.0	532	YES
62	Уборщица, уборщик	0.000	42500.0	1044	YES
63	Разнорабочий	0.473	64712.5	755	No
64	Кредитный специалист	0.000	60000.0	648	YES
65	Инженер по охране труда и технике безопасности...	0.025	72500.0	190	YES
66	Специалист по кадрам	0.000	57500.0	270	YES
67	Менеджер/руководитель АХО	0.011	75000.0	204	YES