

# Исследование эффективности публикаций вакансий категории “Менеджер проектов”

Март 2024

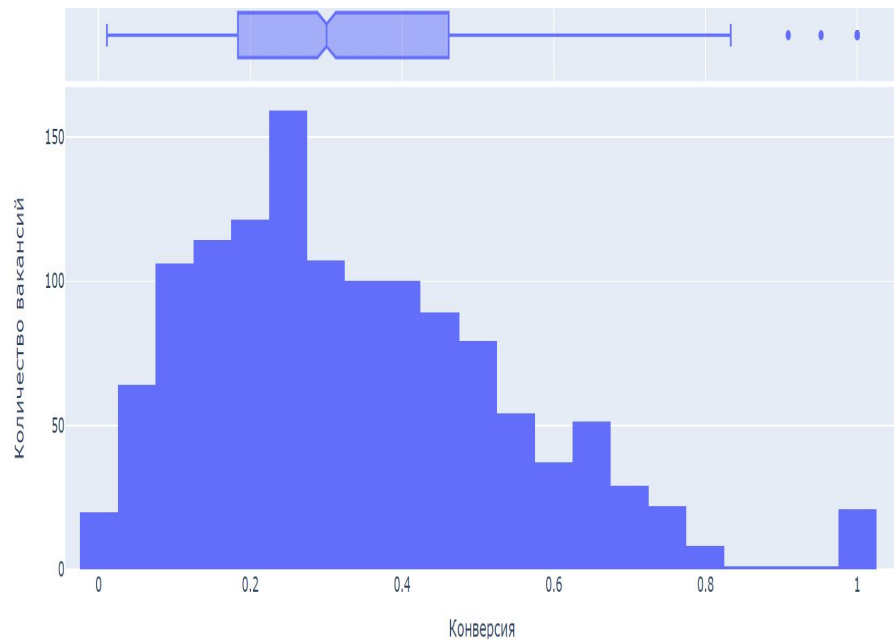
# Описание проекта

- \* Исследование проводилось для кадрового агентства, предоставившего данные по закрытию вакансий за период 2023 - начало 2024 года.
- \* Предмет исследования: группа публикаций вакансий **Project Manager/Менеджер проектов**.
- \* Провести анализ опубликованных вакансий за 2023 год с целью подтверждения или опровержения гипотез относительно влияния определенных условий на вакансии (перечень гипотез в приложении к ТЗ).
- \* В случае обнаружения дополнительных условий, влияющих на отклик кандидатов, которые не были ранее учтены в исходных гипотезах, включить в проект их изложение.

# Основная метрика

Рассчитана метрика конверсии отклики/просмотры:

1 - максимально возможная конверсия, 0 - минимально возможная конверсия



Анализ рассчитанного признака показывает нормальное распределение, это значит он будет хорошо работать как статистический критерий

Средние показатели конверсии по всем вакансиям примерно **0.3**

# Данные разделены по категориям вакансий



# Уровни конверсии в категории “Project manager”

очень низкий уровень: менее 0.20

низкий уровень: от 0.20 до 0.22

ниже среднего: от 0.22 до 0.29

средний уровень: от 0.29 до 0.32

выше среднего: от 0.32 до 0.35

высокий: от 0.35 до 0.41

очень высокий: более 0.41



Низкий уровень: до 22%

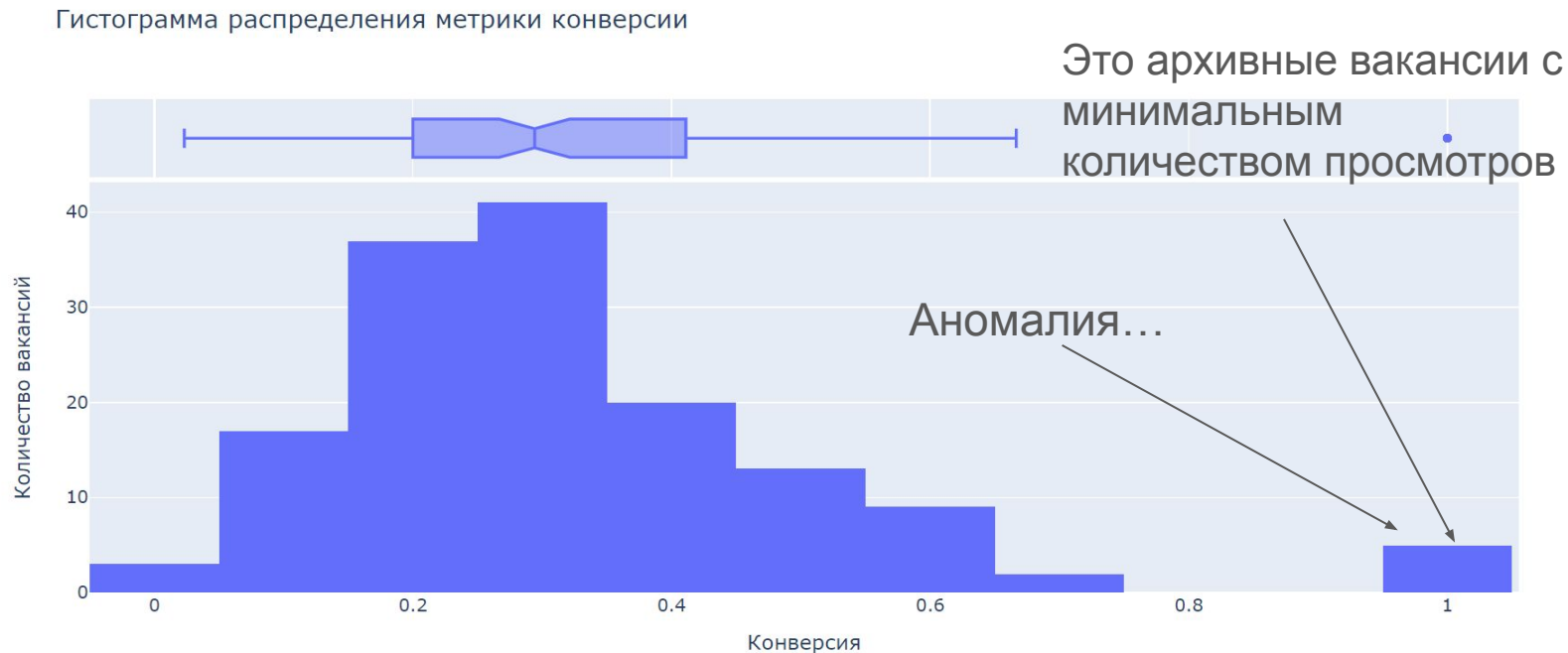


Средний уровень: более 22% и до 35%



Высокий уровень: более 35 %

# Анализ распределения конверсии и поиск аномалий



# Количество просмотров в категории “Project manager”

По особенностям распределения признака конверсия становится очевидно, что также **нужно учитывать и количество просмотров**

очень низкий уровень: менее 23

низкий уровень: от 23 до 27

ниже среднего: от 27 до 37

средний уровень: от 37 до 42

выше среднего: от 42 до 49

высокий: от 49 до 53

очень высокий: более 53

Низкий уровень: до 27 просмотров

Средний уровень: более 27 и до 49 просмотров

Высокий уровень: более 49 просмотров

# Соотношение вакансий по уровням конверсии и просмотрам

Много случаев, когда высокая конверсия у вакансий с низким количеством просмотров. Там может быть сильное влияние случайных факторов. Выберем самые надежные примеры.

Просмотры \ Конверсия			
	Высокий	Средний	Низкий
Высокий	13	14	21
Средний	15	16	18
Низкий	21	19	10

Дальше в возьмем для этой группы название “лучшие вакансии”



# Гипотезы 1 названия (Junior, Junior+ итд.) на всей выборке

Матрица корреляции

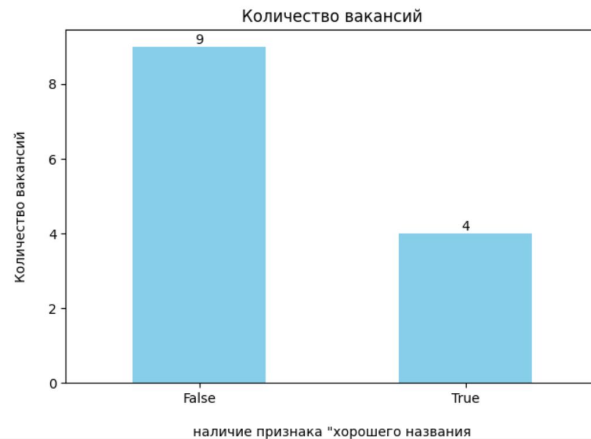
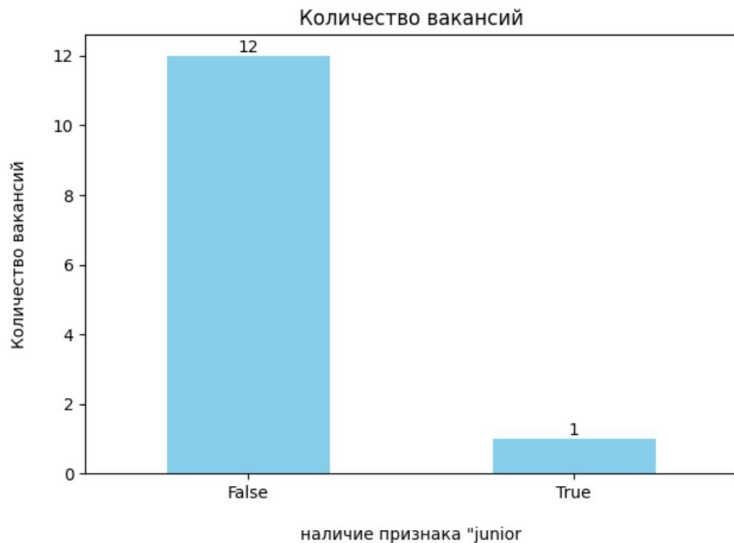
good_title	1	0	0	0.06	0	0	0
junior_tag	0	1	0.95	0	0	0.066	0.12
junior_list	0	0.95	1	0.015	0	0	0.26
attractive	0.06	0	0.015	1	0.92	0.31	0.4
conv	0	0	0	0.92	1	0.28	0.45
views	0	0.066	0	0.31	0.28	1	0.72
responses	0	0.12	0.26	0.4	0.45	0.72	1
	good_title	junior_tag	junior_list	attractive	conv	views	responses

Влияния на показатели успешности не наблюдается

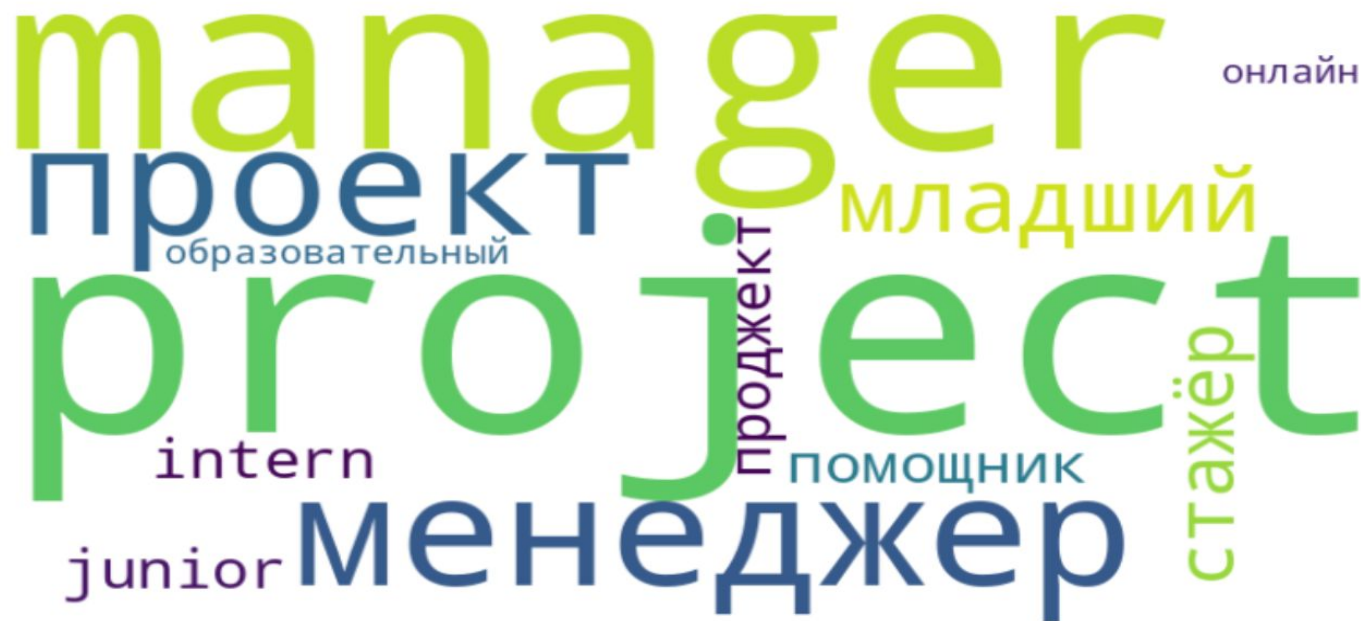
- `good_title` : {Помощник руководителя проекта  
Руководитель проекта, Менеджер проекта, Координатор проекта,  
Менеджер IT проектов}
- `junior_tag`: признак в названии должности
- `junior_list`: признак в любом месте текста

# Гипотезы 1 названия (Junior, Junior+ итд.) на лучших вакансиях

Именно эти признаки в лучших вакансиях встречаются редко:  
предпочтительными оказываются названия “project” латиницей или транслит



Гипотезы 1: облако слов названий лучших вакансий



A word cloud visualization of job titles. The words are arranged in a cluster, with 'manager' and 'project' being the largest. Other visible words include 'online', 'junior', 'intern', 'assistant', 'stage', 'educational', 'youngest', 'project', and 'junior'.

manager  
онлайн  
проект  
образовательный  
project  
младший  
intern  
проект  
помощник  
junior  
менеджер  
стажёр

Примечания по ряду гипотез, для которых не было примеров в данных

Технически невозможной оказалась проверка некоторых гипотез и некоторых “отрицательных признаках”, поскольку не было таких данных в датасете

- Признаков “Middle, Senior”
- Списка стека с ['битрикс24', 'bitrix24', 'activecollab', 'sql', 'powerbi', 'illustrator']
- Признаков для проверки гипотезы про прототипирование и дизайн

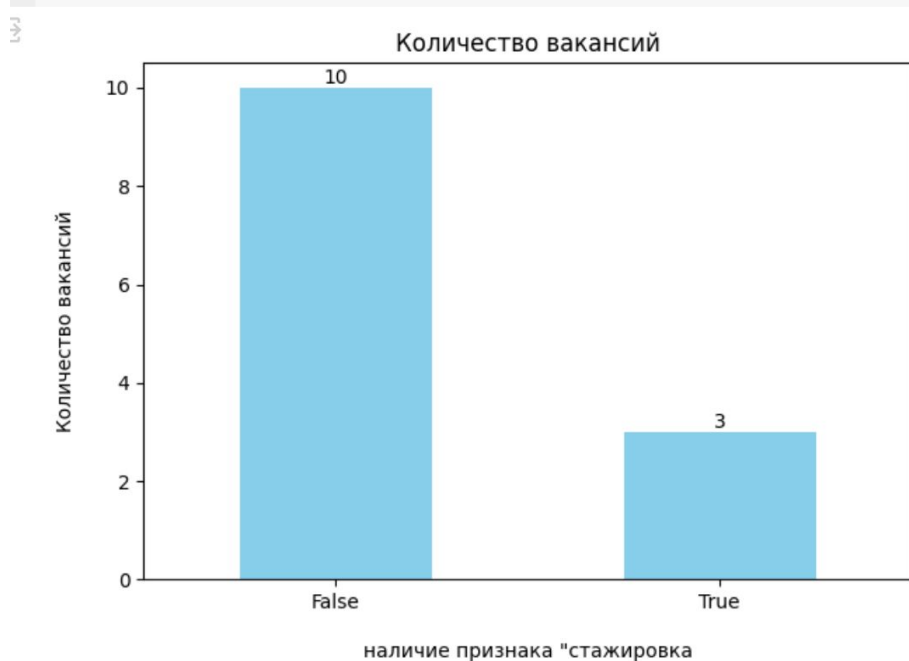
## Гипотезы 2: признак “стажировка” на всей выборке



Не наблюдается  
существенного влияния  
этого признака:

- trainee\_list - признак во всех текстах
- trainee\_tag - признак в названии вакансии

## Гипотезы 2: признак “стажировка” у лучших вакансий



Это признак встречается у лучших вакансий, но все-таки в меньшинстве наблюдаемых случаев

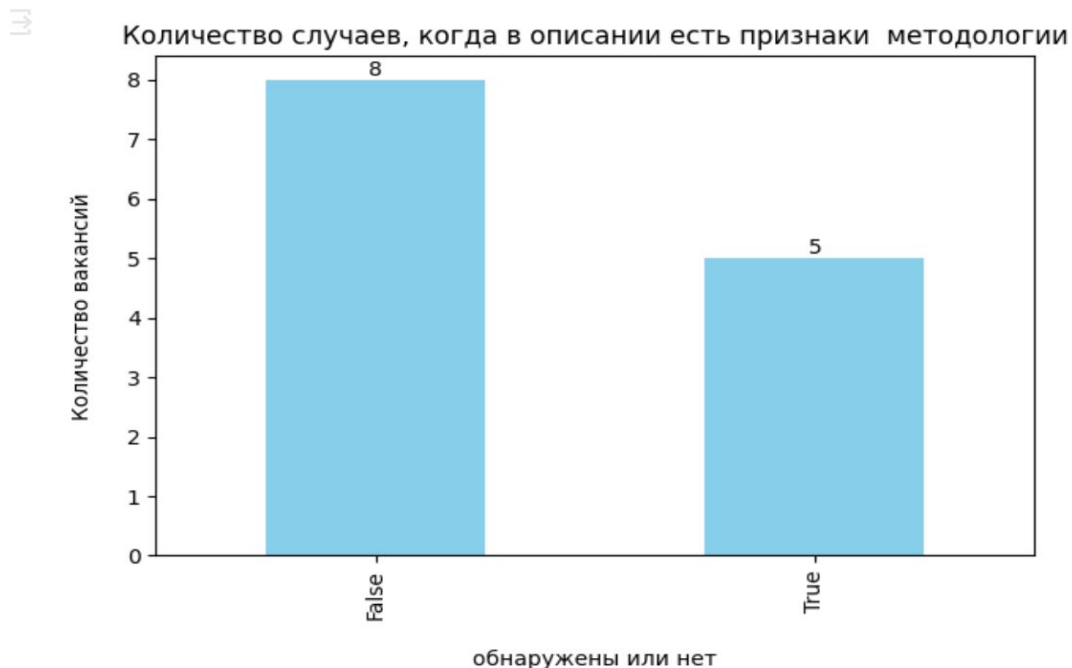
### Гипотезы 3: признаки методологии управления проектами на всей выборке

**Матрица корреляции**

agile_list	1	0.97	0	0	0	0.24
method_list	0.97	1	0	0	0	0.3
attractive	0	0	1	0.92	0.31	0.4
conv	0	0	0.92	1	0.28	0.45
views	0	0	0.31	0.28	1	0.72
responses	0.24	0.3	0.4	0.45	0.72	1
	agile_list	method_list	attractive	conv	views	responses

Вектор влияния не противоречит гипотезе, но связь слишком слабая.

### Гипотезы 3: признаки методологии управления проектами у лучших вакансий



Не наблюдается связи этого признака с успешными вакансиями, возможны оба варианта



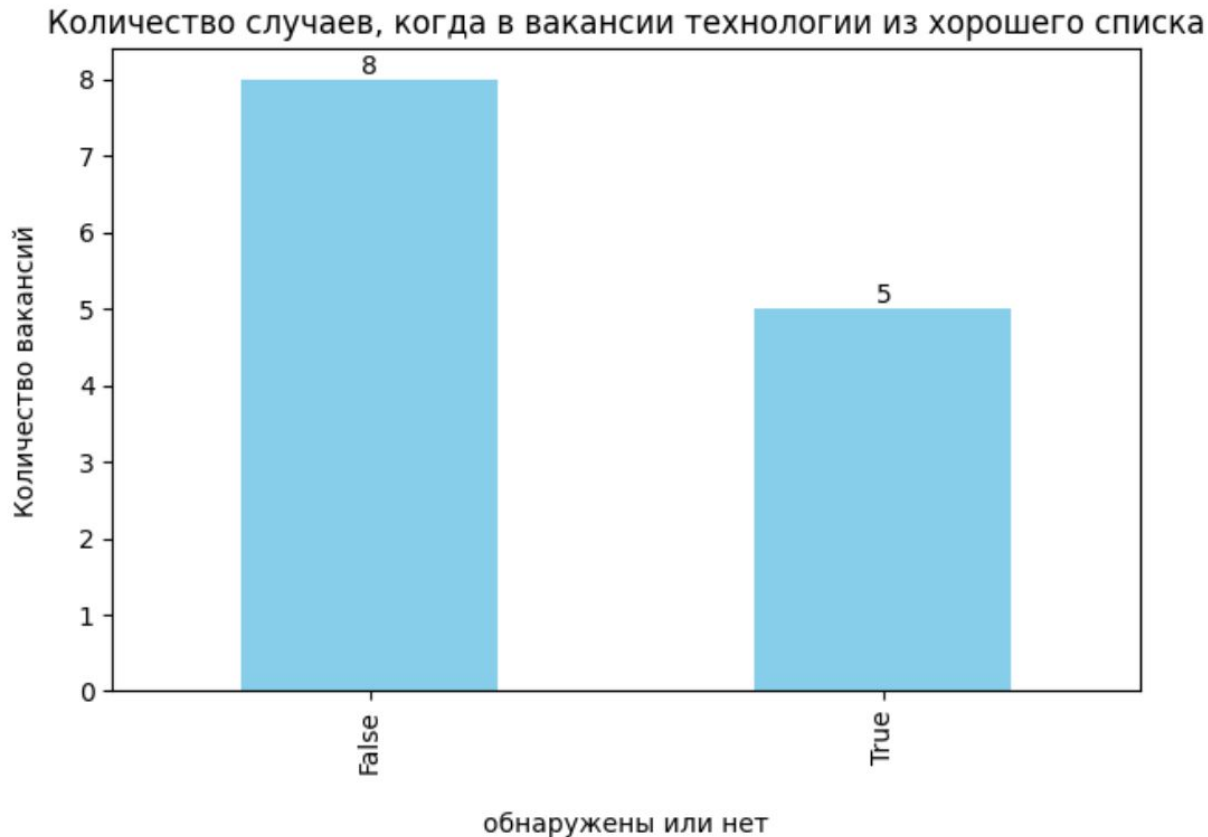
# Гипотезы 4: Технологии, ПО на всем датасете

Матрица корреляции

crm_list	1	0.4	0.4	0.92	0.072	0	0.7	0
good_list_all	0.4	1	1	0.4	0	0	0	0.29
good_list_short	0.4	1	1	0.4	0	0	0	0.29
bad_list_all	0.92	0.4	0.4	1	0.072	0	0.7	0
attractive	0.072	0	0	0.072	1	0.92	0.31	0.4
conv	0	0	0	0	0.92	1	0.28	0.45
views	0.7	0	0	0.7	0.31	0.28	1	0.72
responses	0	0.29	0.29	0	0.4	0.45	0.72	1
	crm_list	good_list_all	good_list_short	bad_list_all	attractive	conv	views	responses

- Влияние “хорошего списка технологий” (good\_list\_all) и его краткой версии (без Excel и другого софта общего назначения): по направлению - как в гипотезе, по силе связей - незначительное.
- “Плохой список” - пустой, за исключением CRM, который больше оснований включить в “хороший список”

## Гипотезы 4: Технологии, ПО у лучших вакансий



Картина распределения похожа на ту же, что и в методологии проектов, и это понятно, т.к. близкие параметры.

Не наблюдается связи этого признака с успешными вакансиями

## Гипотезы 5: английский язык на всей выборке

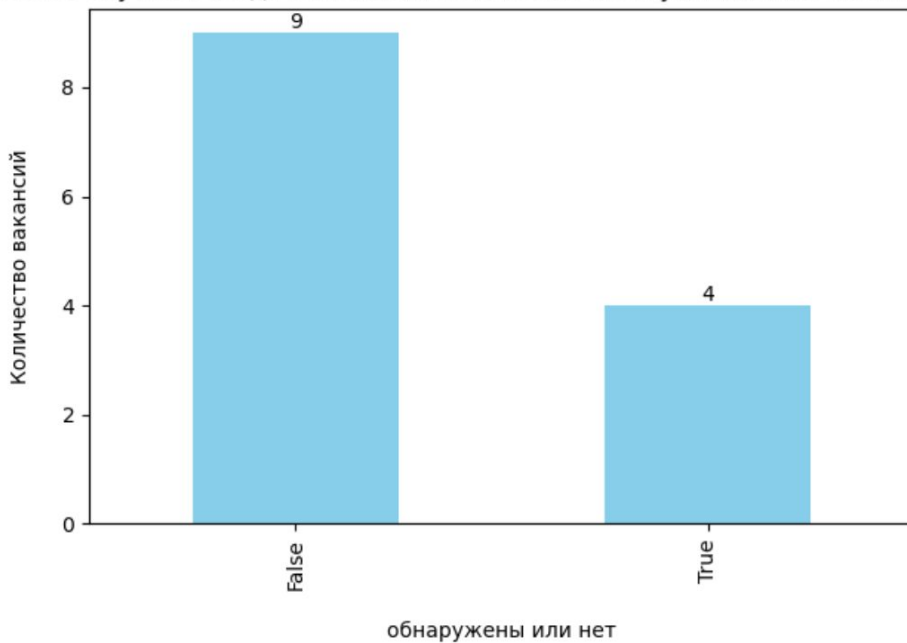
**Матрица корреляции**

english_any	1	0.26	0.52	0	0	0.13	0
english_low	0.26	1	0.42	0.053	0	0	0
english_high	0.52	0.42	1	0.044	0	0	0
attractive	0	0.053	0.044	1	0.92	0.31	0.4
conv	0	0	0	0.92	1	0.28	0.45
views	0.13	0	0	0.31	0.28	1	0.72
responses	0	0	0	0.4	0.45	0.72	1

Связи признаков  
английского языка и  
признаков эффективности  
вакансии не наблюдается.

## Гипотезы 5: английский язык на лучших вакансиях

Количество случаев, когда в вакансии есть какие-либо упоминания английского языка



Случаев, когда есть  
какое-либо  
упоминание  
английского языка -  
меньшинство

# Гипотезы 6, формат работы на всей выборке

Матрица корреляции

format	1	1	1	1	0	0	0.4	0.35
удаленка	1	1	0.94	0.69	0.049	0.2	0.61	0.43
офис	1	0.94	1	0.77	0.042	0.27	0.53	0.32
гибрид	1	0.69	0.77	1	0	0	0.38	0.2
attractive	0	0.049	0.042	0	1	0.92	0.31	0.4
conv	0	0.2	0.27	0	0.92	1	0.28	0.45
views	0.4	0.61	0.53	0.38	0.31	0.28	1	0.72
responses	0.35	0.43	0.32	0.2	0.4	0.45	0.72	1
	format	удаленка	офис	гибрид	attractive	conv	views	responses

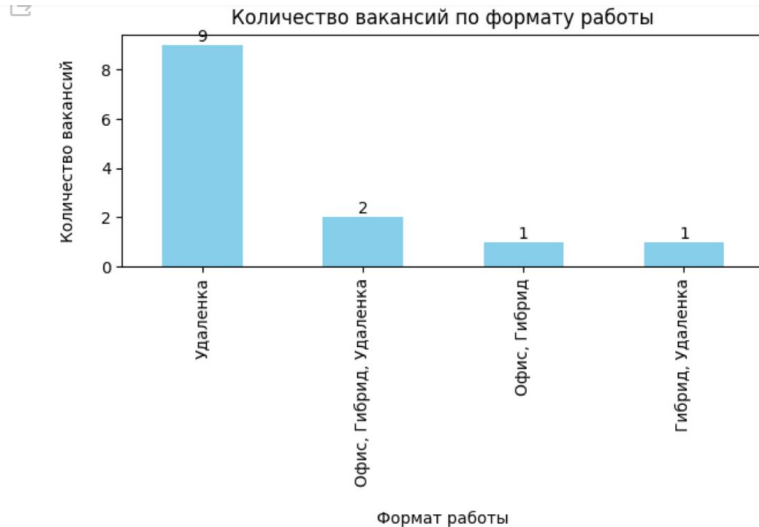
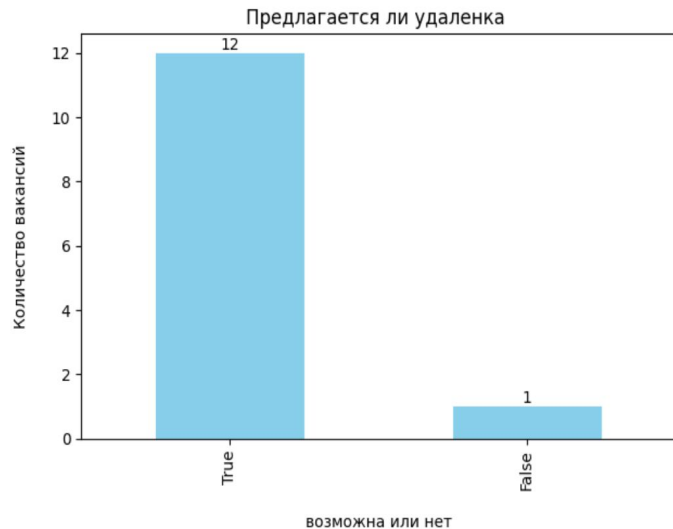
На всей выборке вакансий:

Можно наблюдать тенденции влияния формата работы, но не сильные.

“Удаленка” чуть больше связана с успешными показателями, чем другие форматы.

# Гипотезы 6, формат работы на лучших вакансиях

На лучших вакансиях влияние признака “удаленка” более заметно



# Гипотезы 7, город на всей выборке

Матрица корреляции

москва	1	0.65	0.26	0	0.36	0	0	0.11	1	0	0	0	0.11	0.4	0.32	0.2
петербург	0.65	1	0	0	0	0	0	0	1	0	0	0	0.096	0	0.28	0
екатеринбург	0.26	0	1	0.27	0	0.06	0.27	0	1	0.19	0.16	0.2	0.053	0.43	0.3	0
нижний	0	0	0.27	1	0.22	0.41	0.7	0.35	1	0	0	0	0.0059	0	0	0
казань	0.36	0	0	0.22	1	0	0.22	0	1	0	0	0	0	0.02	0.058	0
воронеж	0	0	0.06	0.41	0	1	0.41	0.14	1	0	0	0	0.01	0	0.44	0
волгоград	0	0	0.27	0.7	0.22	0.41	1	0.35	1	0	0	0	0.0059	0	0	0
тюмень	0.11	0	0	0.35	0	0.14	0.35	1	1	0	0.1	0	0.063	0	0	0
city	1	1	1	1	1	1	1	1	1	0.06	0.24	0.053	0.23	0.51	0	0
удаленка	0	0	0.19	0	0	0	0	0.06	1	0.94	0.69	0.049	0.2	0.61	0.43	
офис	0	0	0.16	0	0	0	0	0.1	0.24	0.94	1	0.77	0.042	0.27	0.53	0.32
гибрид	0	0	0.2	0	0	0	0	0	0.053	0.69	0.77	1	0	0	0.38	0.2
attractive	0.11	0.09	0.05	0.0059	0	0.01	0.0059	0.063	0.23	0.049	0.042	0	1	0.92	0.31	0.4
conv	0.4	0	0.43	0	0.02	0	0	0	0.51	0.2	0.27	0	0.92	1	0.28	0.45
views	0.32	0.28	0.3	0	0.058	0.44	0	0	0	0.61	0.53	0.38	0.31	0.28	1	0.72
responses	0.2	0	0	0	0	0	0	0	0	0.43	0.32	0.2	0.4	0.45	0.72	1
москва																
петербург																
екатеринбург																
нижний																
казань																
воронеж																
волгоград																
тюмень																
city																
удаленка																
офис																
гибрид																
attractive																
conv																
views																
responses																

Из списка городов в гипотезе есть слабо выраженная связь успешных показателей с городами:

- Москва
- Екатеринбург

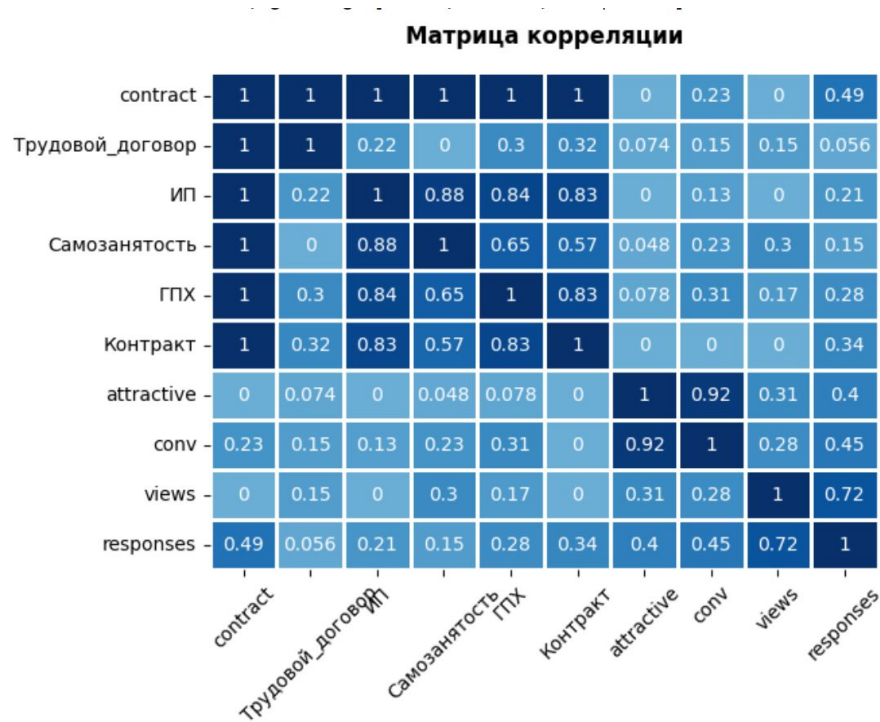
# Гипотезы 7, город на лучших вакансиях



- **Москва и Санкт-Петербург** совпадают с гипотезой.
- **Ростов-на-Дону и Тверь** - новые города (но это единичные случаи, нет данных для вывода о закономерности)

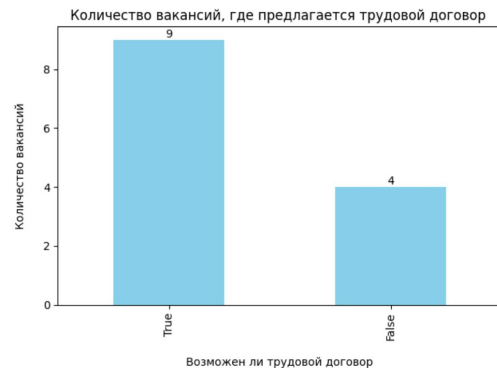
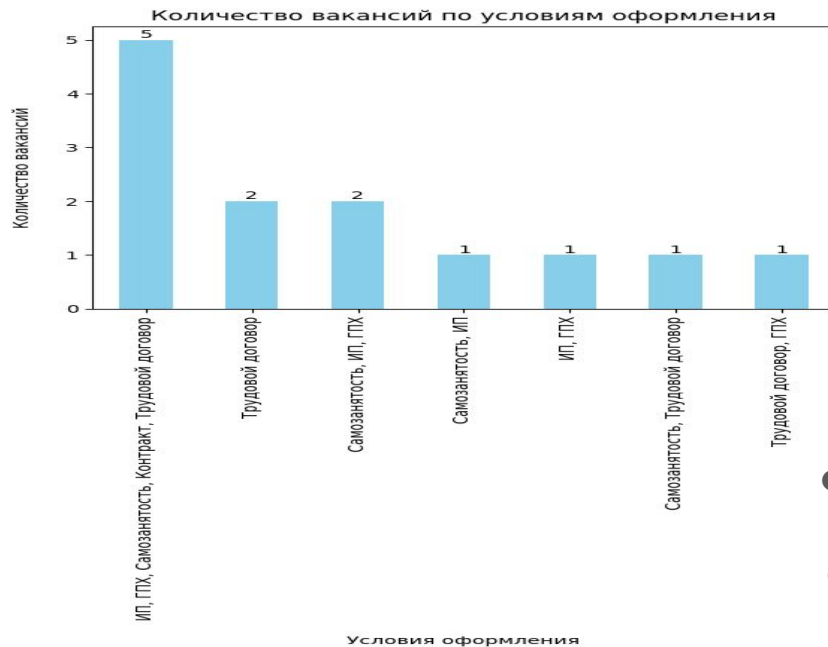


## Гипотеза 8, формат оформления на всем датасете



Не наблюдается  
существенных  
преимуществ у какого  
либо формата  
оформления

# Гипотеза 8, формат оформления на лучших вакансиях



- Трудовой договор предлагается чаще, но не обязательно
- Второй фактор (предположительно): количество вариантов в списке (ассортимент возможностей)

# Гипотезы 9: зарплата, на всем датасете

Матрица корреляции

salary_min	1	0.7	0	0.12	0	0
salary_max	0.7	1	0.74	0.3	0	0
attractive	0	0.74	1	0.92	0.31	0.4
conv	0.12	0.3	0.92	1	0.28	0.45
views	0	0	0.31	0.28	1	0.72
responses	0	0	0.4	0.45	0.72	1
	salary_min	salary_max	attractive	conv	views	responses

- В датасете более половины записей без этого параметра, может быть искажение
- Наблюдается взаимосвязь верхнего уровня з/п и категории привлекательности, посчитаем какие это значения:

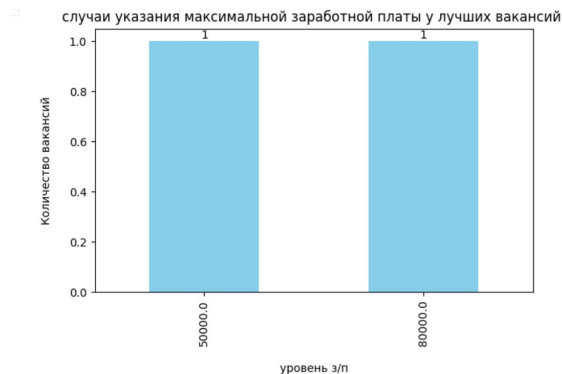
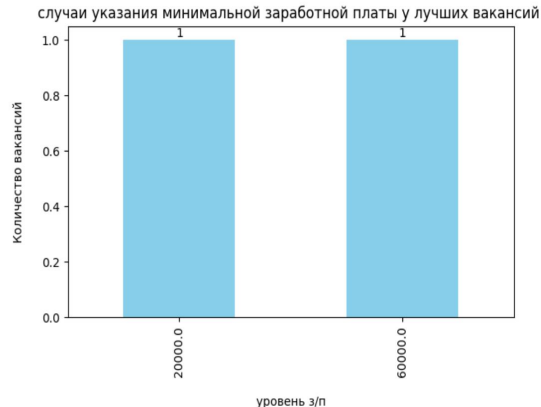
```
↳ attractive
high      80000.0
low       50000.0
middle    60000.0
Name: salary_max, dtype: float64
```

# Гипотезы 9: зарплата, на примерах лучших вакансий

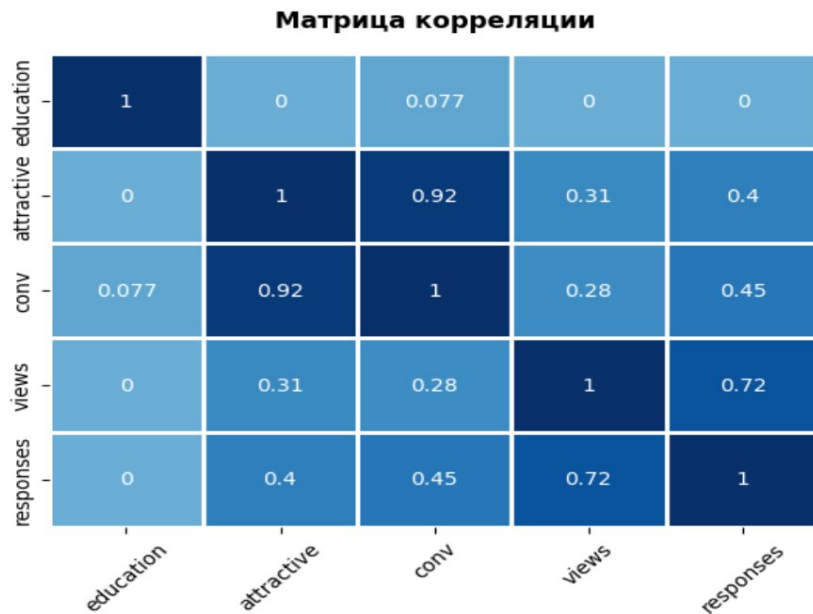
Проверке этих гипотез мешает существенное количество вакансий без указания зарплаты.

По данному датасету можно посчитать, что зарплата не влияет на успешность вакансии.

Но это не вывод, а результат дефицита информации. Вакансии без указания з\п более распространены.



## Гипотеза 10, образование, на всем датасете



Связи признака “Образование” в требованиях вакансии с параметрами успешности вакансии не наблюдается.

- Признак в датасете 147: в пропорции 143/4

# Гипотеза 10, образование, на лучших вакансиях

Наблюдается интересная закономерность: у всех лучших вакансий в категории образования - "Другое.." Но если присмотреться - этот признак в датасете в большинстве именно такой.



Другое, Около-релевантное (например, техническое образование для разработчика или QA, общее бизнес образование для маркетолога), Профильное (например, ИТ образование для разработчика) 143

Около-релевантное (например, техническое образование для разработчика или QA, общее бизнес образование для маркетолога), Профильное (например, ИТ образование для разработчика) 4

Name: education, dtype: int64

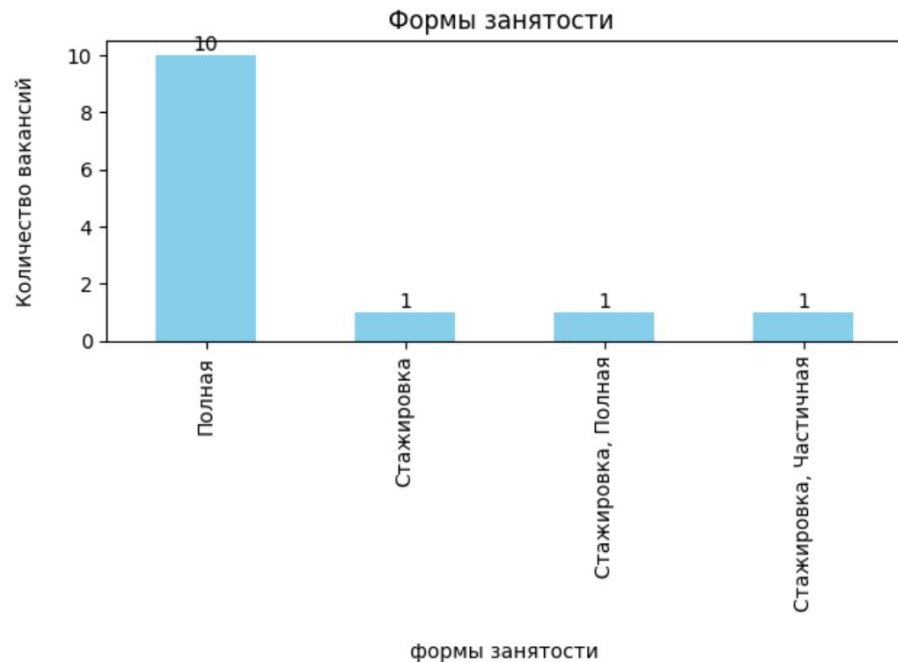
# Гипотеза 11: форма занятости на общей выборке

Матрица корреляции

employment	1	1	1	1	1	1	0	0	0.17	0.62
Полная	1	1	0.64	0.76	0.52	0.12	0.083	0	0.13	0.093
Проектная	1	0.64	1	0.9	0.58	0.7	0	0	0	0.32
Частичная	1	0.76	0.9	1	0.49	0.62	0	0	0	0.26
Стажировка	1	0.52	0.58	0.49	1	0.86	0.071	0	0.15	0.24
Волонтерство	1	0.12	0.7	0.62	0.86	1	0.037	0.095	0	0
attractive	0	0.083	0	0	0.071	0.037	1	0.92	0.31	0.4
conv	0	0	0	0	0	0.095	0.92	1	0.28	0.45
views	0.17	0.13	0	0	0.15	0	0.31	0.28	1	0.72
responses	0.62	0.093	0.32	0.26	0.24	0	0.4	0.45	0.72	1

Примечательно, что сама по себе форма занятости показывает связь с метриками откликов, но какая именно - не очевидно

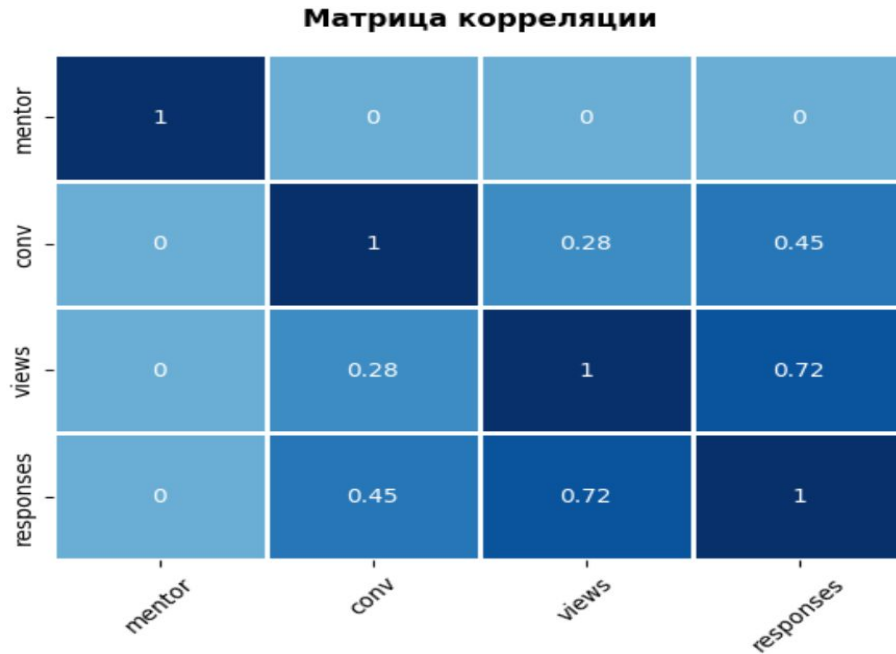
# Гипотеза 11: форма занятости на лучших вакансиях



- Преобладают предложения с полной занятостью
- Есть 3 предложения со стажировкой
- Предложение, где только стажировка - это определенно не оплачиваемая стажировка

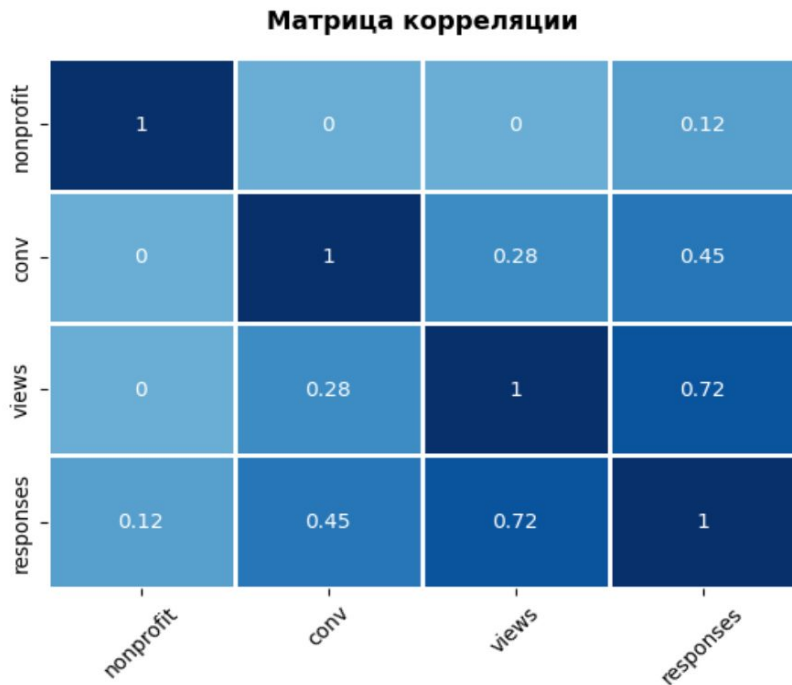


# Гипотезы 13: менторство, наставник



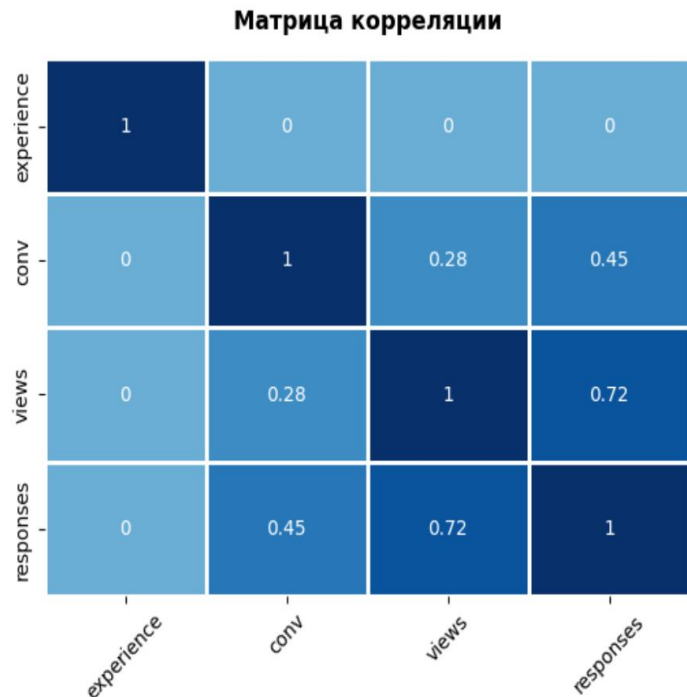
Влияния на  
обнаруживается

# Гипотеза 14: некоммерческий опыт



- В данных всего 5 примеров с таким случаем
- Влияния на статистику не обнаруживается

# Гипотеза 15: опыт категории



- Связи с показателями не наблюдается
- В данных диспропорция по этому признаку: 147 : 137/12/4
- Во всех примерах лучших вакансий: все 3 категории

# Условия работы у лучших вакансий (облако слов)





# Итоги (кратко по списку)

- **название должности/позиции** : вероятнее латиницей без признаков Junior
- **формат работы**: вероятнее удаленка
- **опыт** : вероятнее перечень из всех категорий (но мало других примеров в данных)
- **образование**: “Другое....” (но мало других примеров в данных)
- **город** : вероятнее Москва
- **з/п**: можно не указывать, но если есть верхняя граница, желательно от 80000 рублей
- **формат оформления**: вероятнее по ТК, но лучше “широкий ассортимент”
- **занятость**: вероятнее полная. Стажировки не исключаются, но больше данных за полную занятость.
- **обязанности**: вероятнее “управление разработкой...”
- **условия работы**: вероятнее часовой пояс МСК, ТК РФ, удаленка, ‘интересный проект...’, ‘карьерный рост...’
- **требования**: не подтвердилось влияние требования по языку, agile и некоторые другие