

И в царстве, и в сказке

Аверков Фёдор
Адаменко Матей
Кислицин Иван
Нефедов Никита
Некрасова Дарья
Валуева Дарья



TINKOFF



**УНИВЕРСИТЕТ
ИННОПОЛИС**

Датасет

Данные о заработной плате разных регионов и сфер деятельности России за 2017 год

Основные переменные:

- Psu- название населённого пункта
- Status-тип населённого пункта
- Age- возраст
- Industry- отрасль занятости
- Invage- логарифм заработной платы
- Public- частное или государственное мероприятие

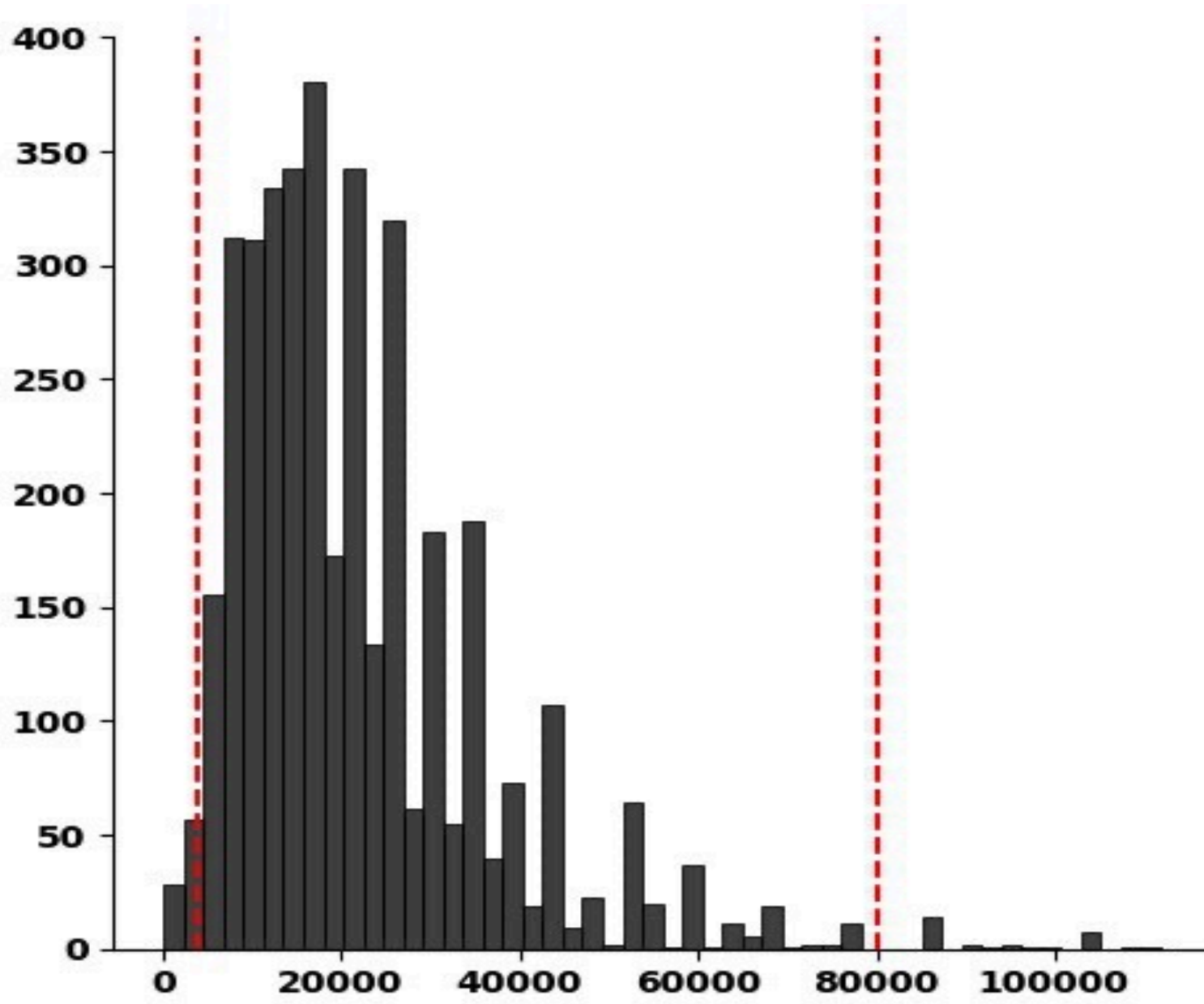
Чистка данных

Убрали:

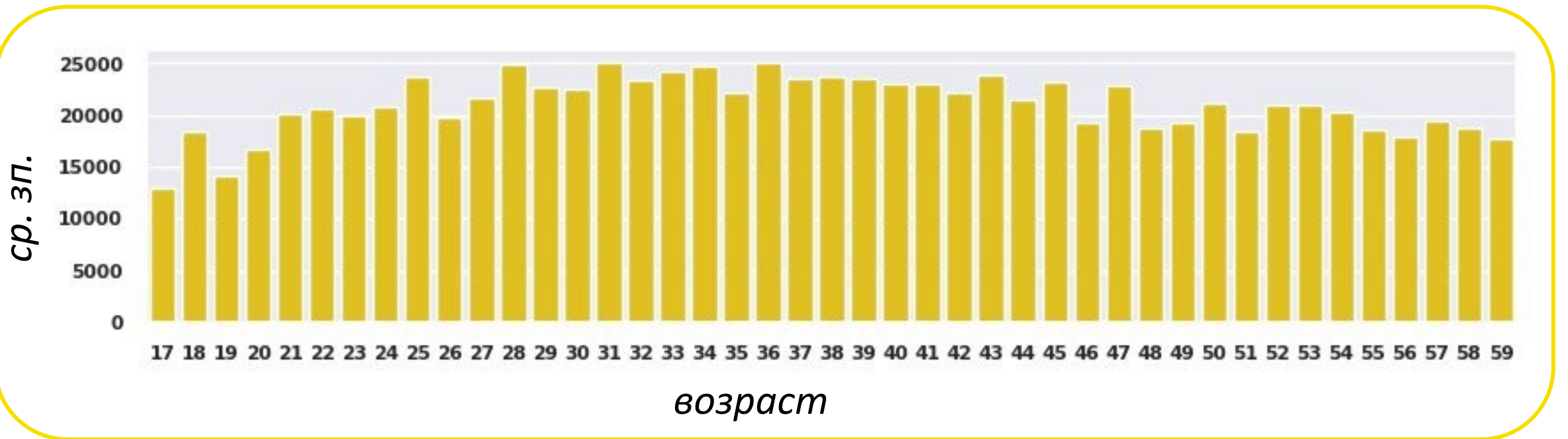
- Нижнее значение по Inwage (-59 строчек)
- Верхнее значение по Inwage (-46 строчек)

Окончательная выборка составляет: 3766 строчек

Чистка данных

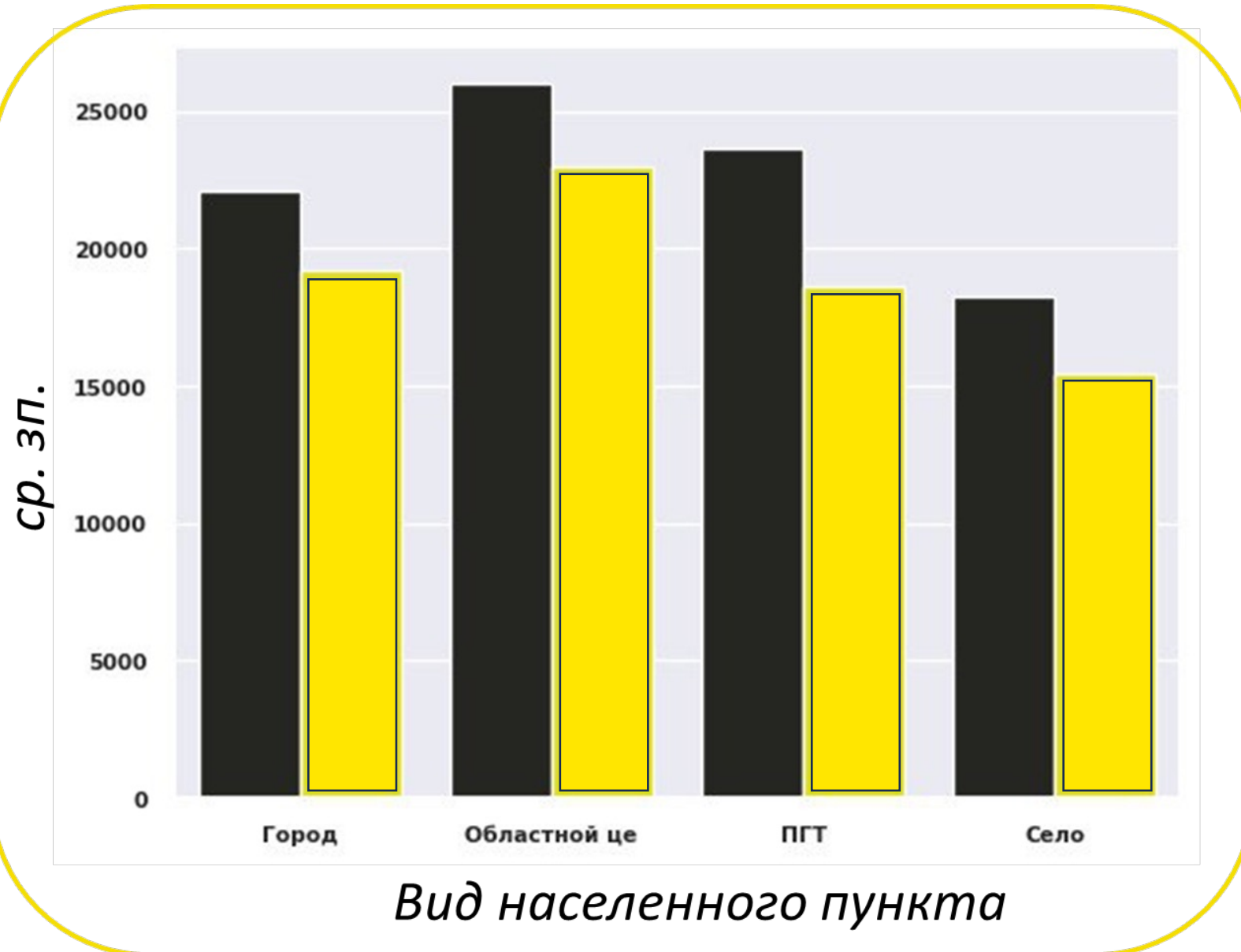


Предварительный анализ



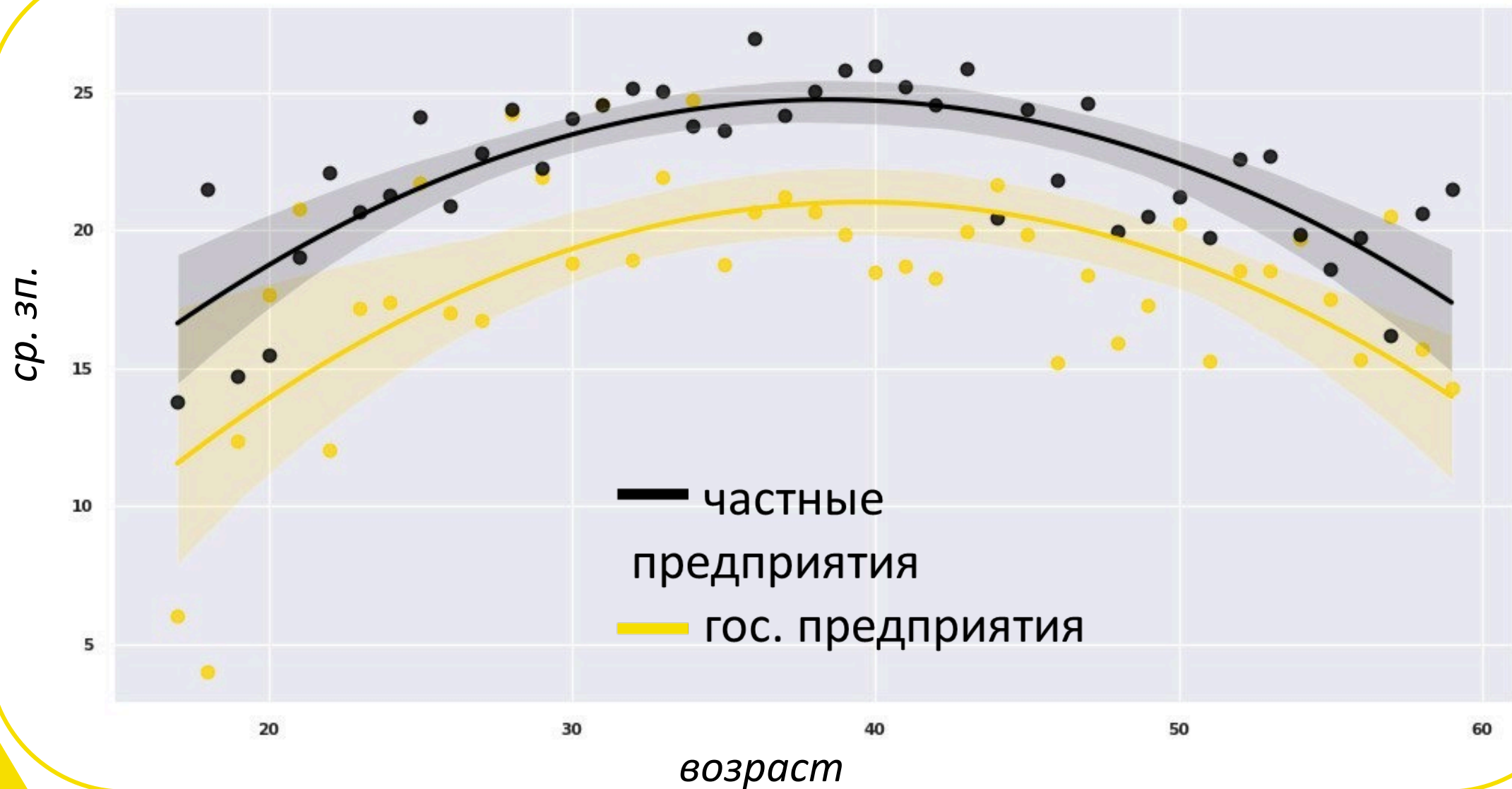
Линейная шкала

Предварительный анализ

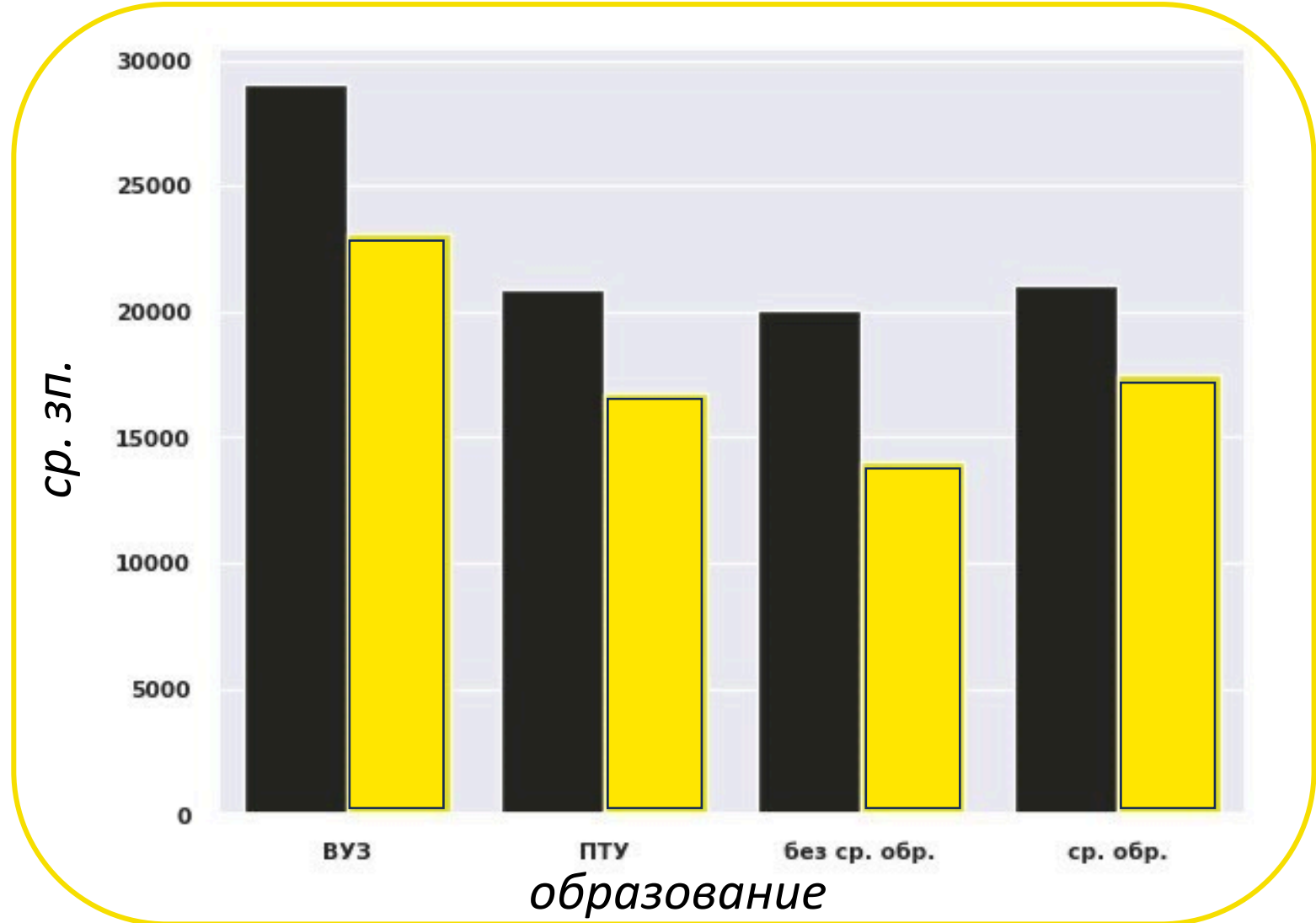
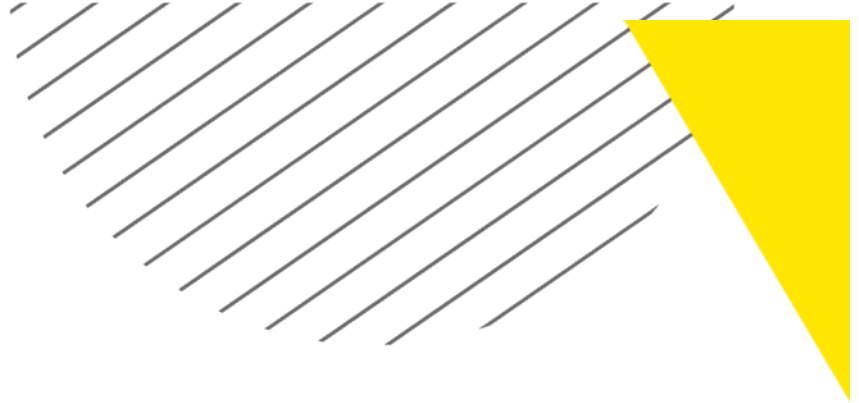


— частные
предприятия
— гос. предприятия

Предварительный анализ

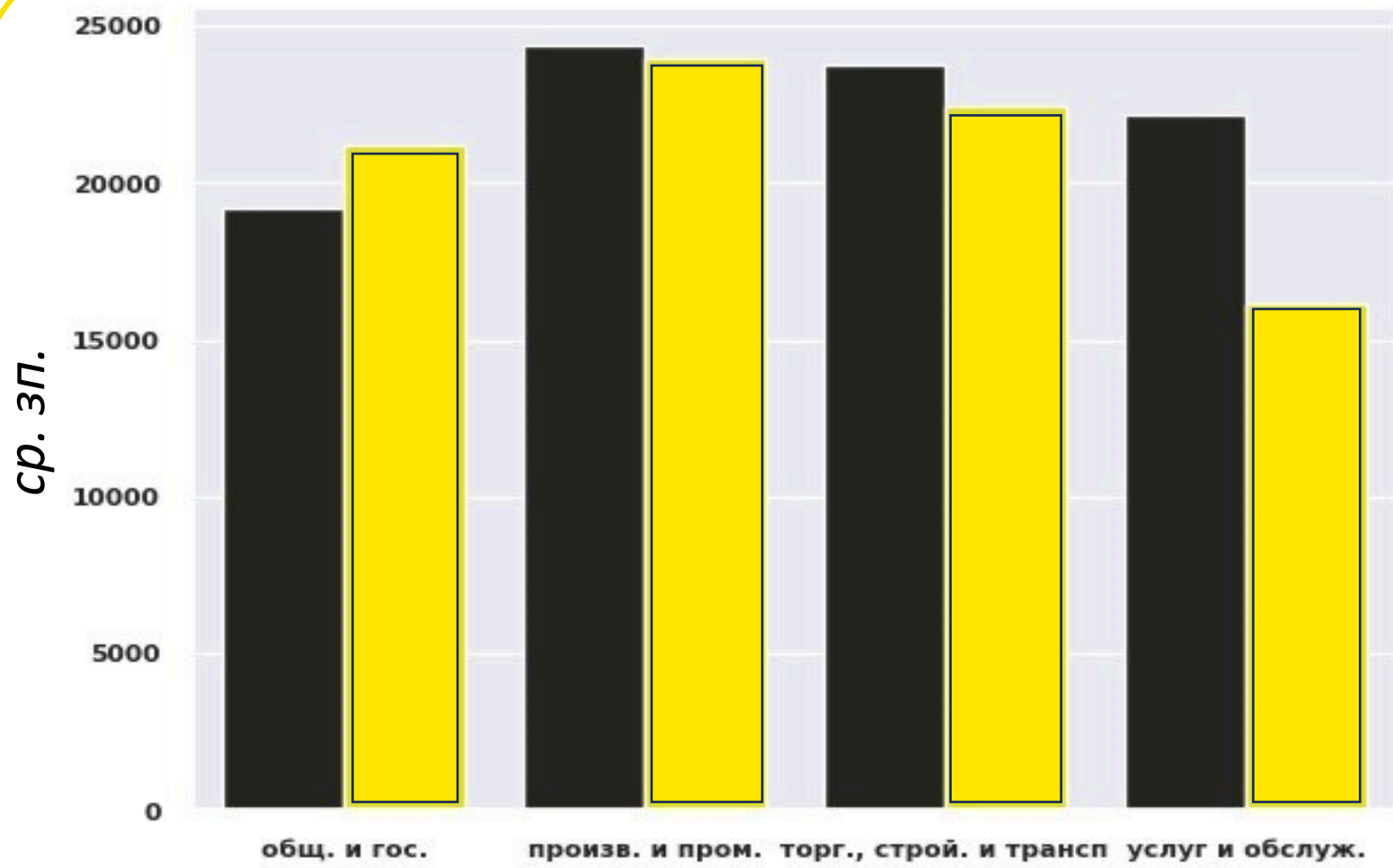


Предварительный анализ



— частные предприятия
— гос. предприятия

Предварительный анализ



Сфера деятельности

— частные предприятия
— гос. предприятия

Исследовательский вопрос

От чего зависит заработная плата?

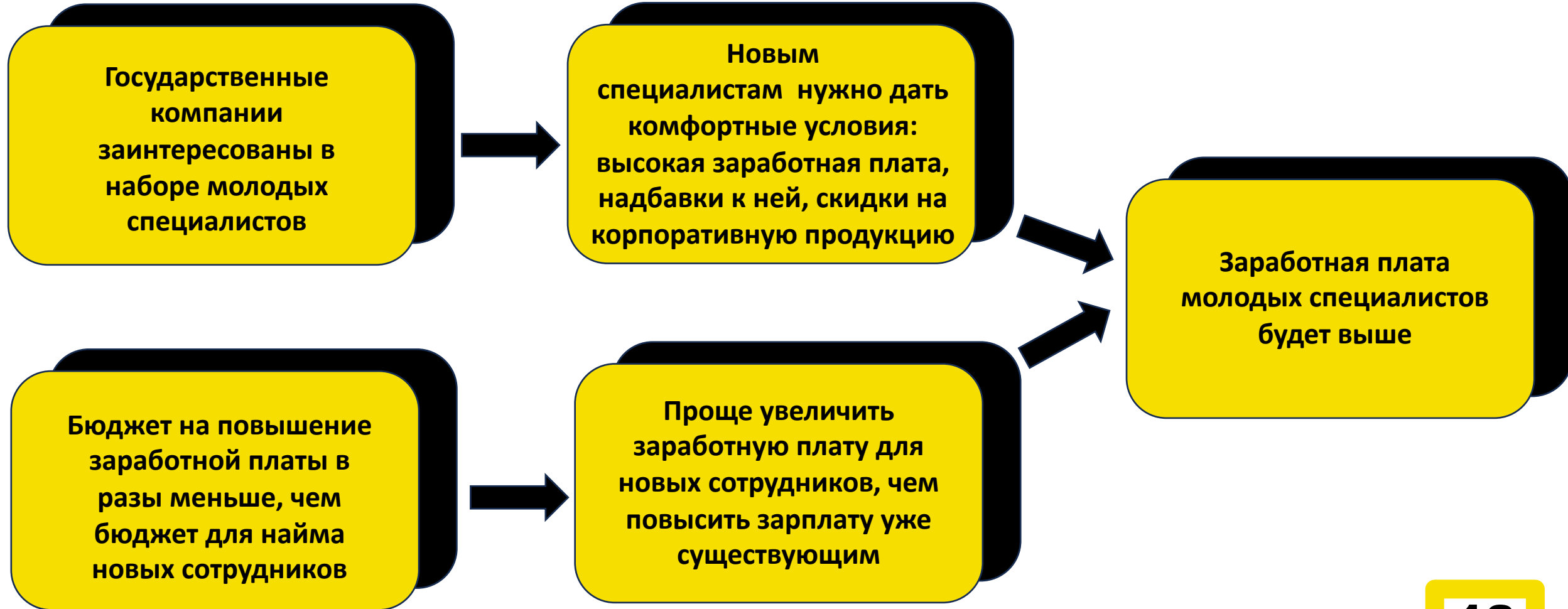
Гипотеза №1

Заработная плата будет зависеть от типа предприятия(частное/государственное) и в частных будет выше.

Гипотеза №2

Заработная плата будет зависеть от возраста и у более взрослых в государственных компаниях зп будет выше.

Механизм



Подготовка данных

Замена переменной `age` на бинарную `the_young`

Замена текстовой переменной `Inwade` на `exp_Inwade`

Объединили данные из `Industry` в категориальную переменную `Industry_group`

Математическая модель

$$H_0: D_1 = D_2 = D_3 = D_4$$

H_1 : Дисперсии не равны

Критерий Левене

№	Значение статистики	P-значение	Число элементов выборки	Итог
1	11.884	$9.688 * 10^{-8}$	3766	Отклоняем H_0

Математическая МОДЕЛЬ

Поправка Боннферони: $p < 0.05/6$

№№	Значение статистики	P-значение	Число элементов выборок	Итог
M1 = M2	27.422	$1.636 * 10^{-7}$	958+521	Медианы не равны
M2 = M3	6.238	0.0125	521+1061	Медианы равны
M3 = M4	73.825	$8.538 * 10^{-18}$	1061+1226	Медианы не равны
M1 = M3	92.026	$8.554 * 10^{-22}$	958+1061	Медианы не равны
M2 = M4	18.167	$2.025 * 10^{-5}$	521+1226	Медианы не равны
M1 = M4	1.847	0.174	958+1226	Медианы равны

Критерий Краскела — Уоллиса

Интерпретация

- Имеется статистически значимая взаимосвязь между средним значением переменной `lnwage` и бинарной переменной `public`.
- Имеется статистически значимое совместное влияние переменных `public` и `is_younger` на количественную переменную `lnwage`.

Вывод:

Зарплата зависит от:

Заработная плата зависит от возраста и вида компании: молодым специалистам лучше идти работать на государственные предприятия

Практическая польза

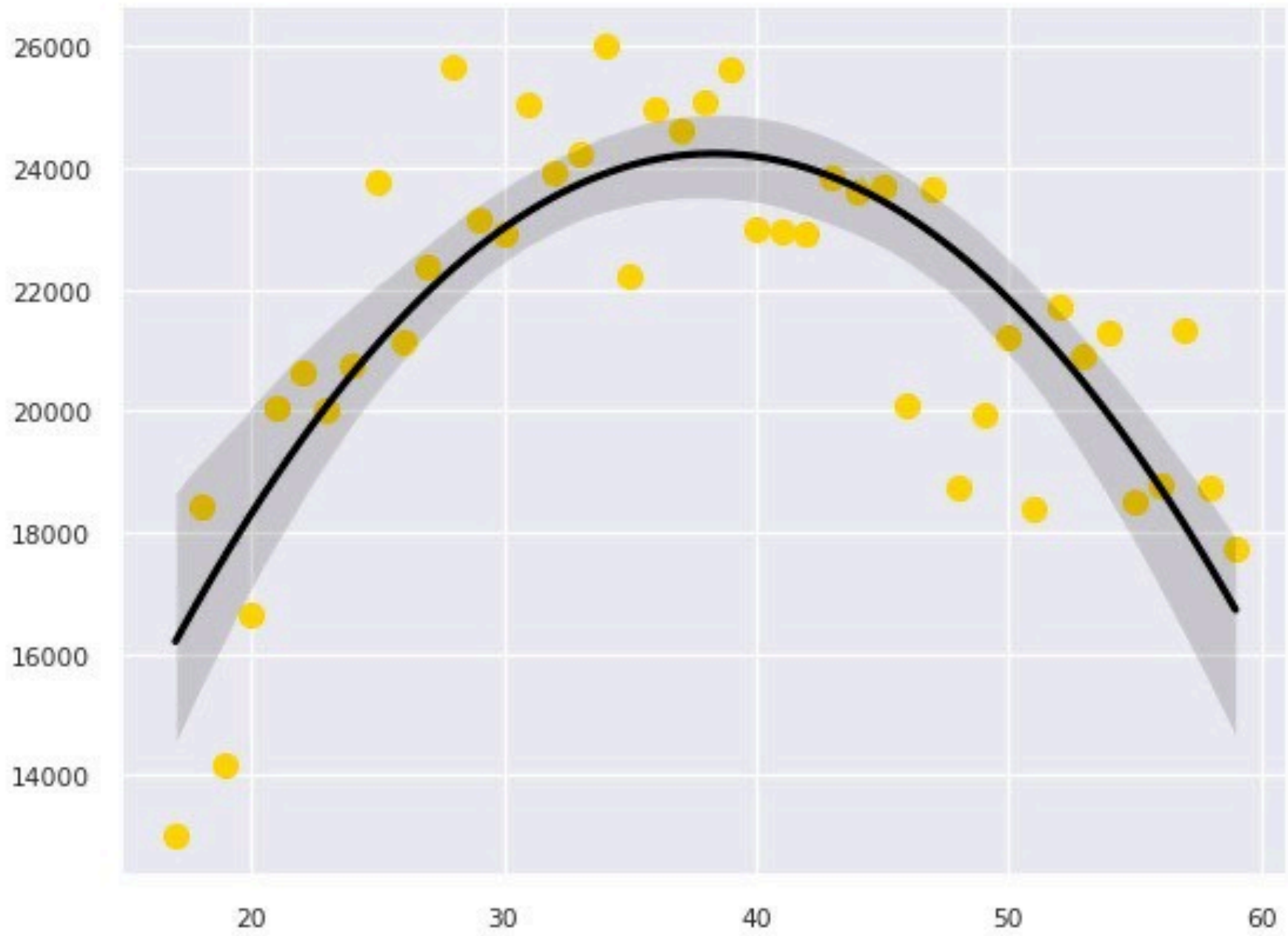
Помощь людям с тем, чтобы определиться с выбором вида компании, в которой им выгоднее работать.

Policy Implication

Платформы для профориентации и поиска работы

Policy Implication

Реклама государственных компаний и привлечение туда специалистов





Questions and answers

Приложение

