

# Подписка Tinkoff Pro

Она такая одна



Команда №1

# С чем мы имеем дело?

## ПРОБЛЕМА:

**Как привлечь больше пользователей подписки и/или увеличить срок ее использования?**

## DANO:

- Клиентская база в 20000 человек разных возрастов
- 28 переменных, 26 из которых — количественные

## Предварительный анализ

X

- Обработка пропущенных значений: доход, пол, количество детей
- Обработка выбросов: отрицательное количество дней, аномальные цифры, расхождение данных
- Выделим факторы, коррелирующие с выгодой
- Нахождение средних значений и медиан: активность и выгода
- Построение кривой распределения данных

# Численные характеристики

## Корреляция

	выгода
доход	0.118
возраст в банке	0.201
активных дней	0.181
выгода	1.000
кол-во опций	0.272
НС	0.160
флаг оплаты подписки	0.144
оборот	0.346
кол-во транзакций	0.114
транзакций фастфуд	0.108
транзакций рестораны	0.123

## Медианы и среднее

Характеристика	Выгода	Активные дни
Медиана	222.430	120.000
Среднее	361.027	145.839

## Гипотезы

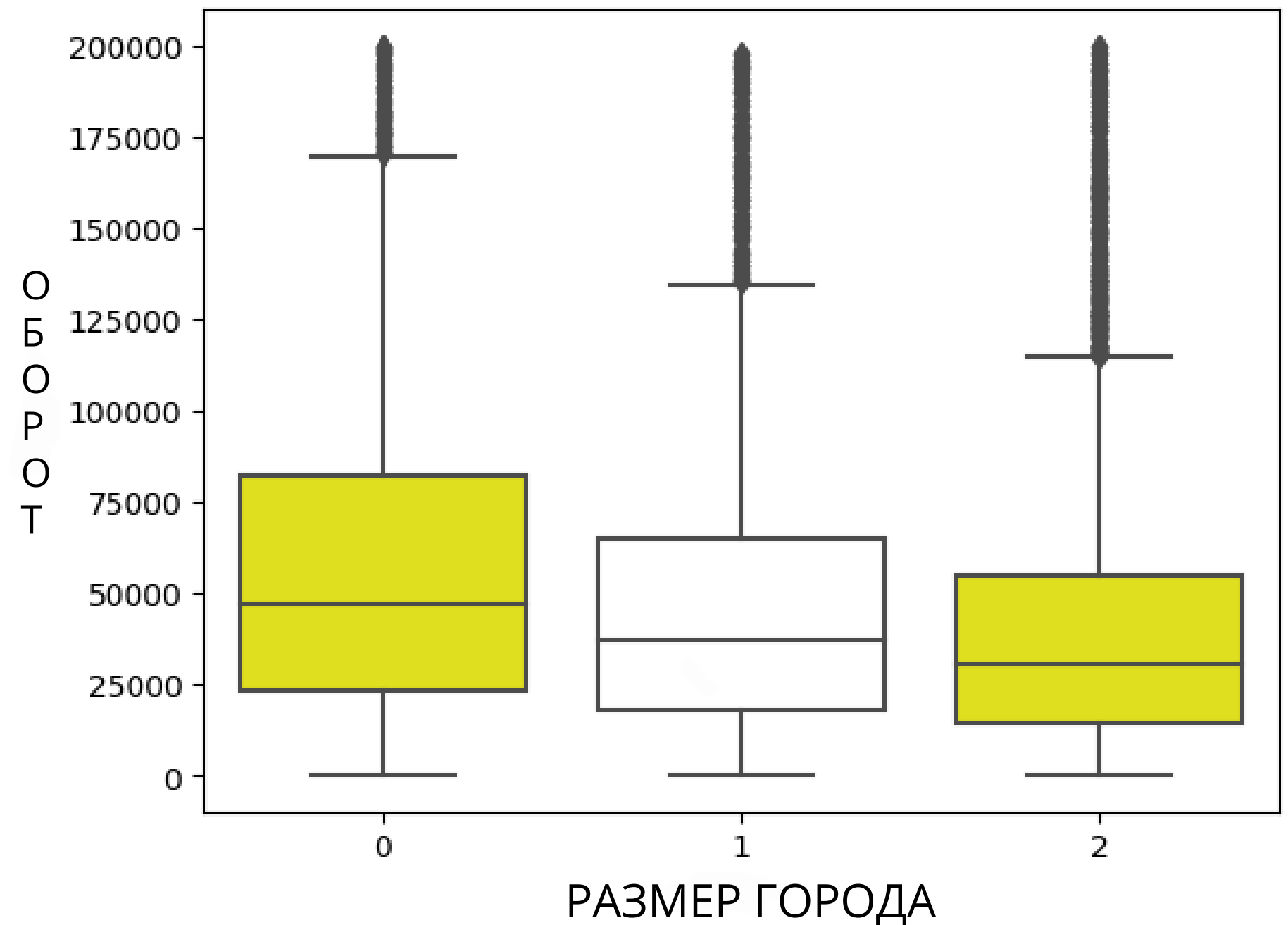
**X**

- Для увеличения количества пользователей подписки можно улучшить ее условия для жителей маленьких городов
- Разные возрастные группы привлекают разные элементы подписки

## Работа с данными



- Выделение групп среди городов по упоминаниям
- Медиана и среднее по оборотам и транзакциям в городах
- График распределения оборотов по городам



# Проверка гипотез

## Медианы и среднее

гео	медиана по обороту	среднее по обороту
0	49102.045	66974.770
1	38102.230	52212.299
2	30649.265	44423.131

## Медианы и среднее

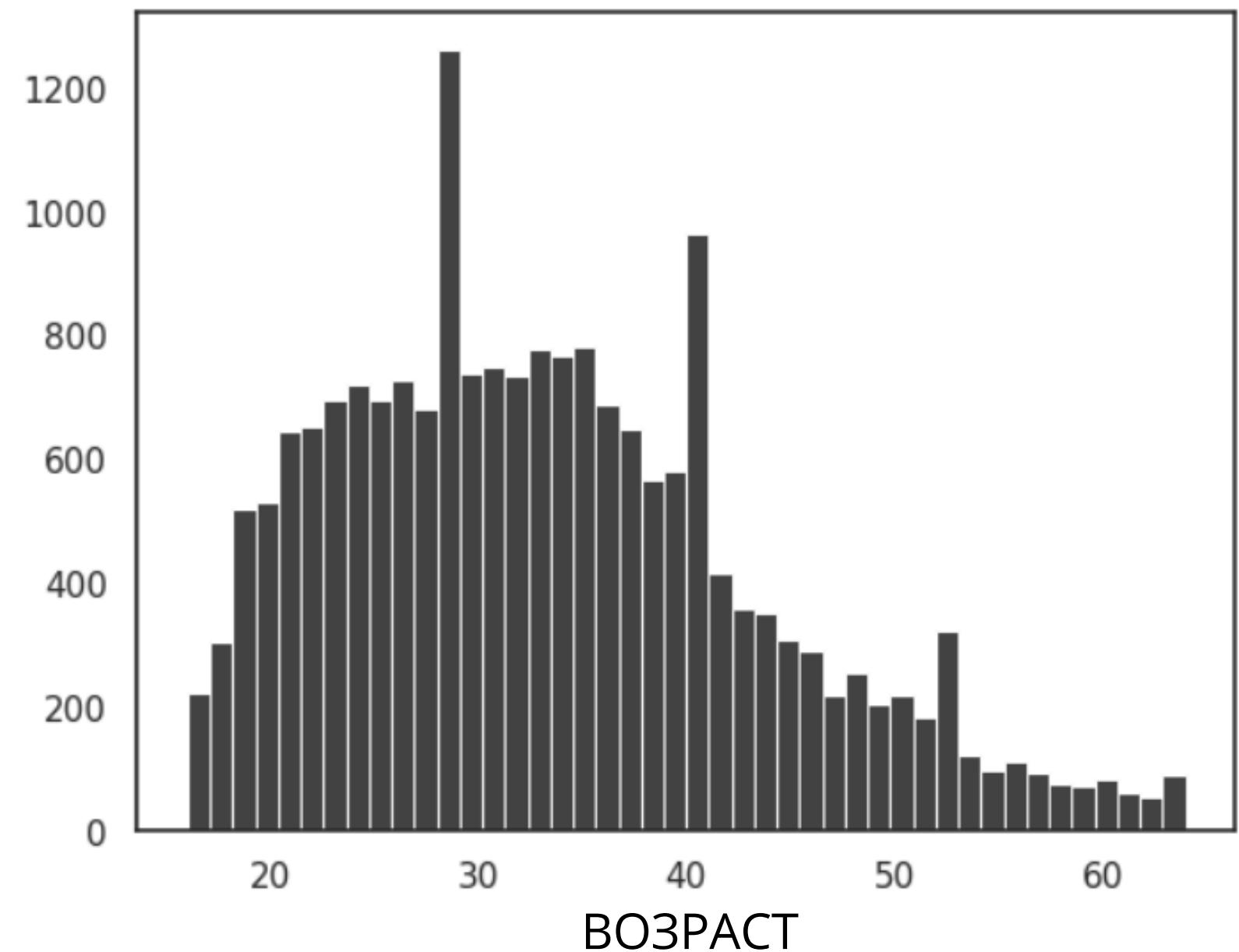
X

гео	медиана по транзакциям	среднее по транзакциям
0	71.0	79.451
1	73.0	81.905
2	62.0	72.041

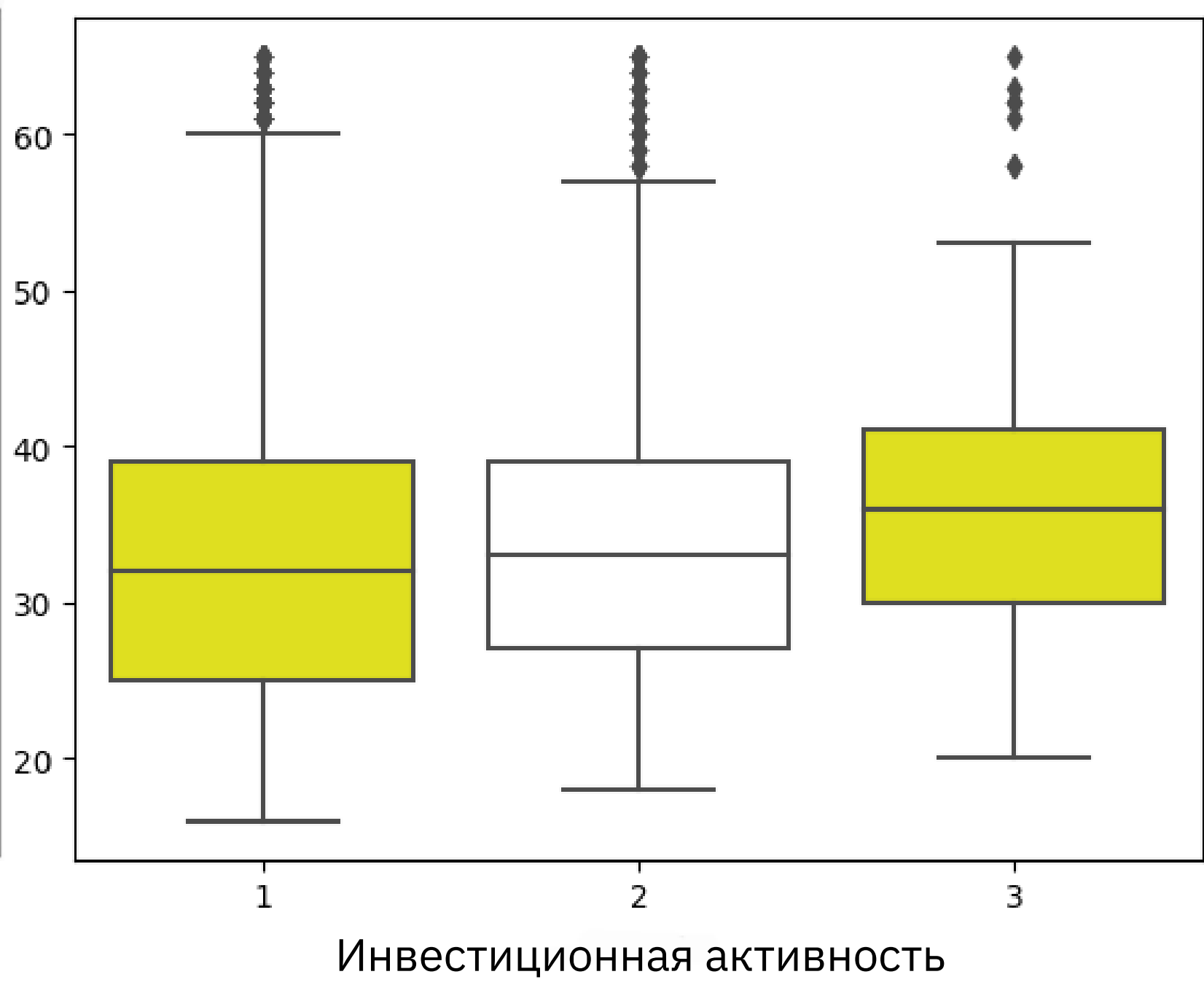
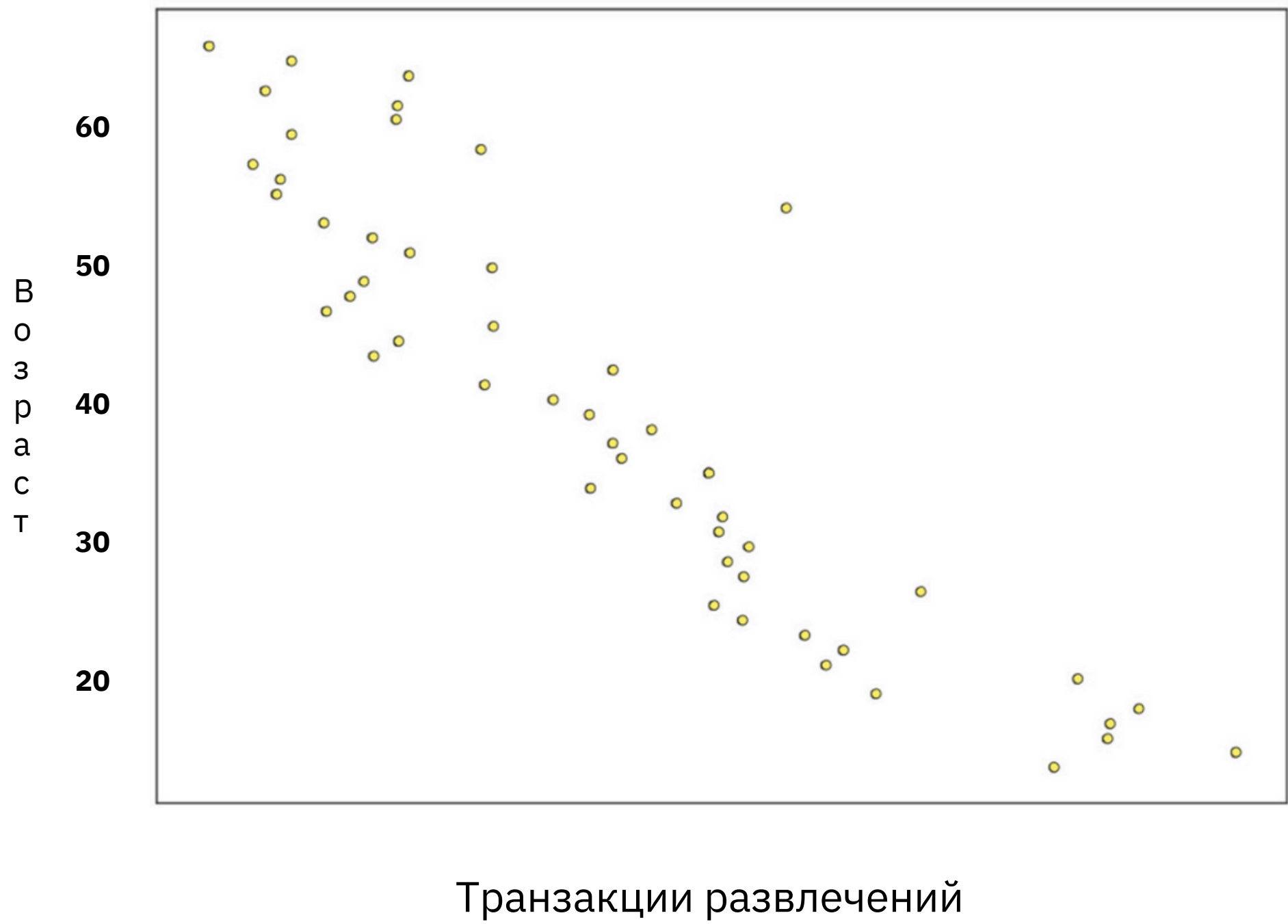
## Работа с данными



- **Выделение групп людей по возрасту из характера распределения**
- **Формируем предположения для каждой из групп**
- **Строим графики из предположений**







# Выводы и комментарии

## Гипотеза про города



- **Повышение кэшбэка в маленьких городах будет эффективно.**



## Гипотеза про возраст



- **Адаптация условий подписки под возраст будет выгодна для младшей и средней возрастных групп**



## Ограничения и перспективы

X

- Ограничения: не учли вид счета(бизнес счет)
- Некоторые представители возрастных групп могут не разделять моду и быть недовольны.

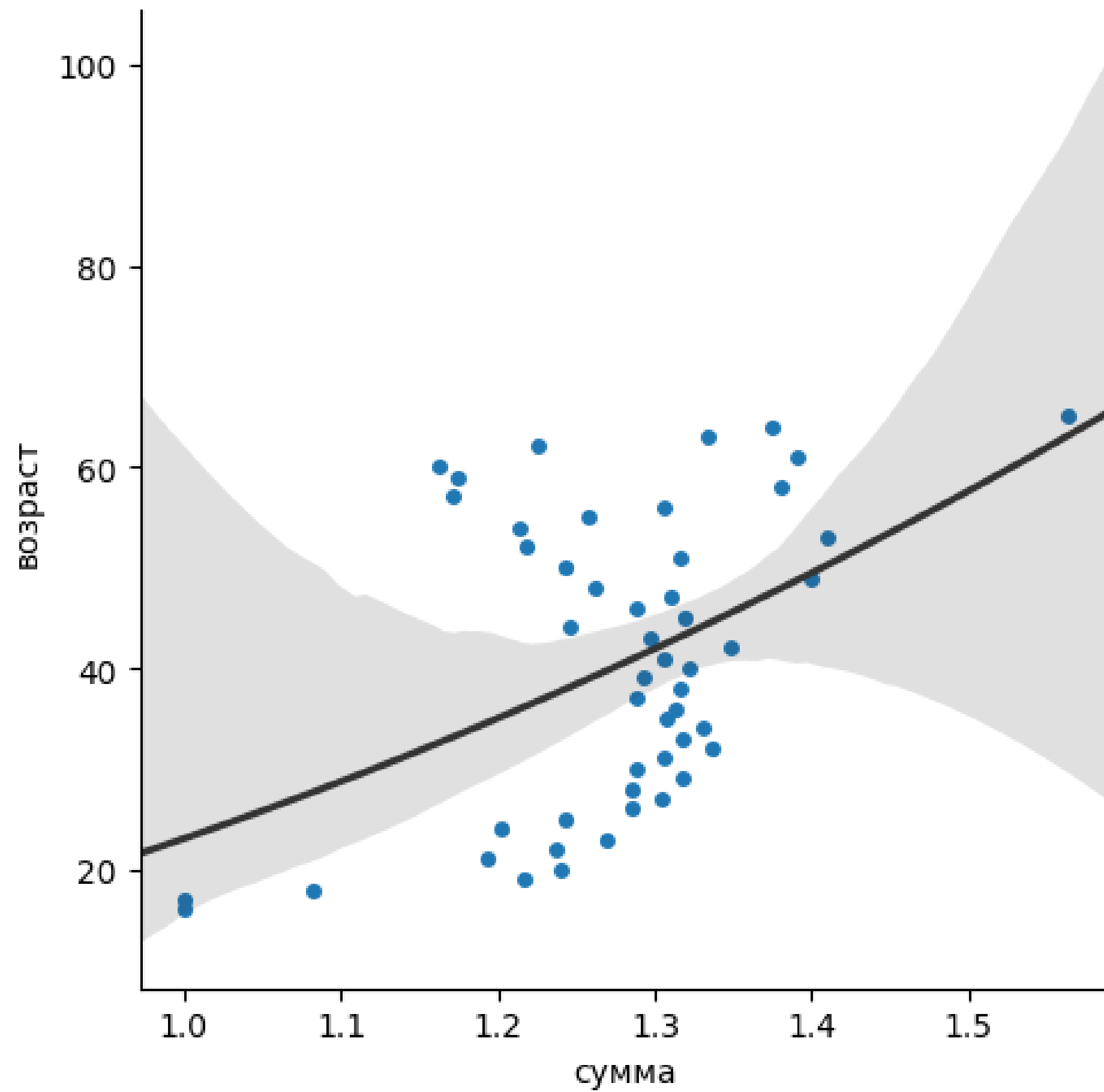
# Будем рады ответить на ваши вопросы

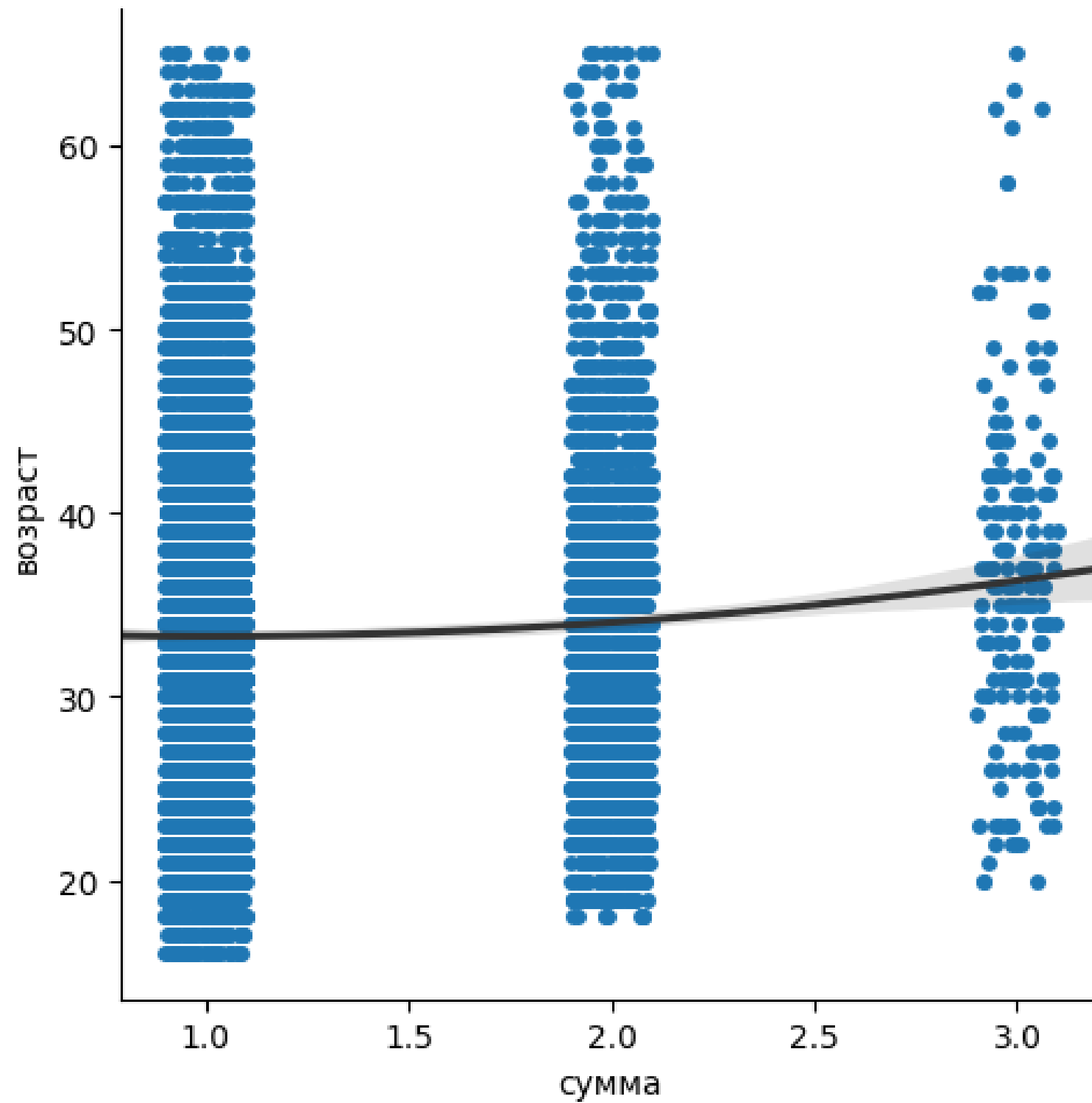
Исследование спонсировано:

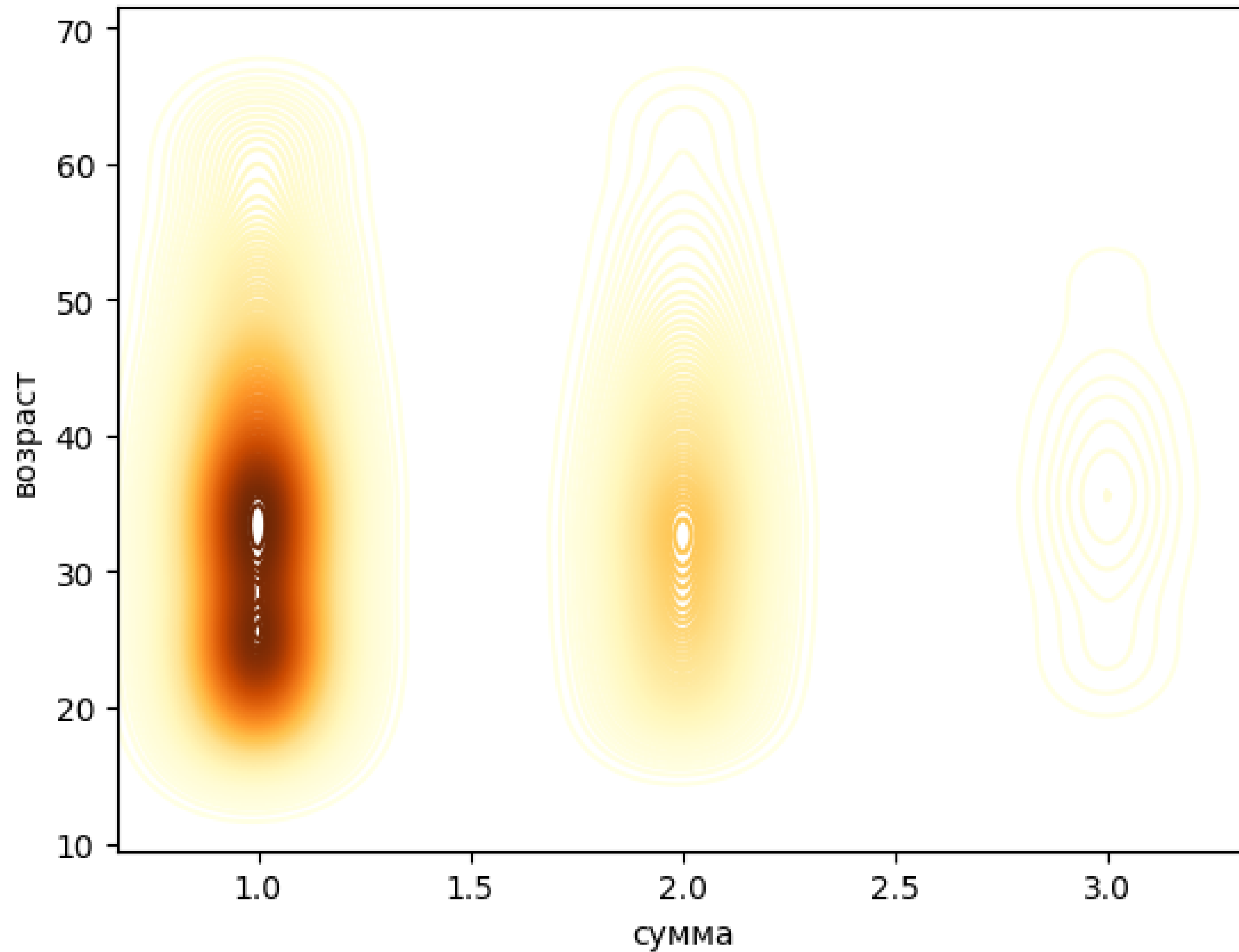


D A  
N O









*#Работа с датафреймом*



```
data_origin = pd.read_csv(r"C:\Users\dimas\Downloads\tinkoff_pro_new.csv")  
data = data_origin.copy(deep = True)
```

```
data['доход'] = data['доход'].fillna(data['оборот'])#заменяем пустые ячейки в доходе на оборот
```

```
data = data[data['доход'] > 0] #доход, равный нулю, не учитываем  
#пропуски в доходе заменили на оборот
```

```
data = data[~data['пол'].isna()]#удаляем пропуски в колонке 'пол'
```

```
data['кол-во детей'] = data['кол-во детей'].fillna(0)#раз люди пропустили графу 'кол-во детей', то и детей у них нет
```

```
data['возраст в банке'] = data['возраст в банке'].fillna(0) # заполняем единственный пропуск  
data = data[data['возраст в банке'] > 0] #В датафрейме встречались отрицательные значения данной колонки, что невозможно + чистим нолики
```

```
data['дата отключения'] = data['дата отключения'].fillna('2022-12-1') #если человек не отключил подписку, считаем, что она действовала до конца года
```

```
data.to_csv(r"C:\Users\dimas\OneDrive\Рабочий стол\ML\table_for_pro")
```