

.руХеLeS

Платонов Константин
Томулец Станислав
Коротчин Андрей
Алдаг Алсу
Безрук Александра

Попробовали.
А что дальше?



Датасет

Данные о заявках на
подписку Tinkoff
Premium за март 2023 г

73711

заявок

Основные переменные:

- флаги использования услуг
- сумма и количество транзакций клиента за последние 1, 3, 5 месяцев
- длительность подписки
- длительность пробного периода
- возрастной сегмент клиента

Много бинарных переменных

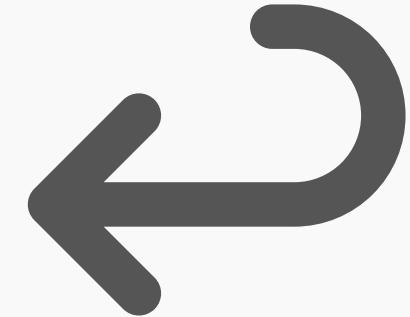
Чистка данных

Были убраны:

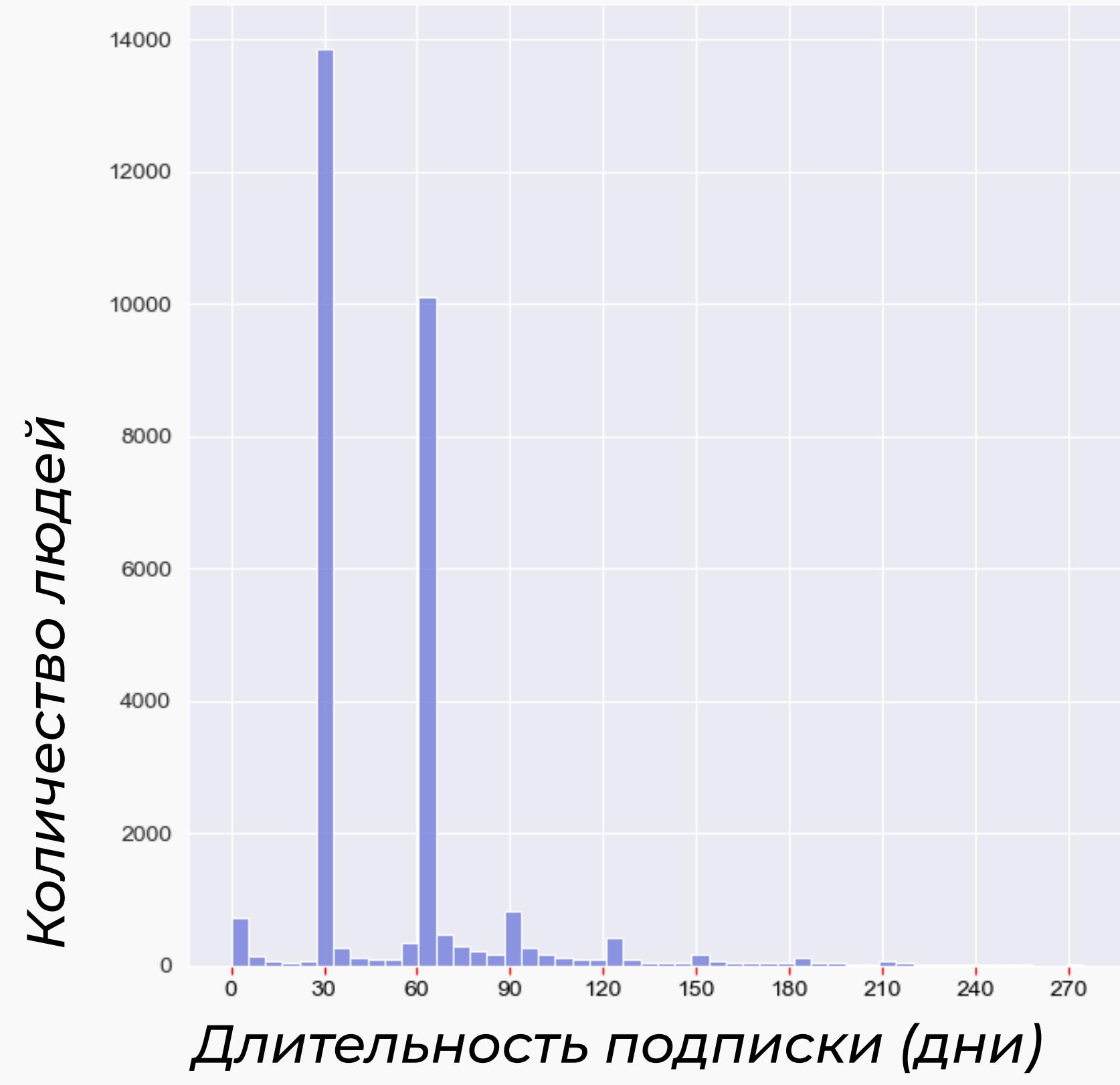
- пустые значения по age_segment
- пустые значения по subscription_time
- заявки, у которых нет пробного периода или он больше 1 месяца
- дубликаты клиентов

Окончательная выборка: 29 868 чел.

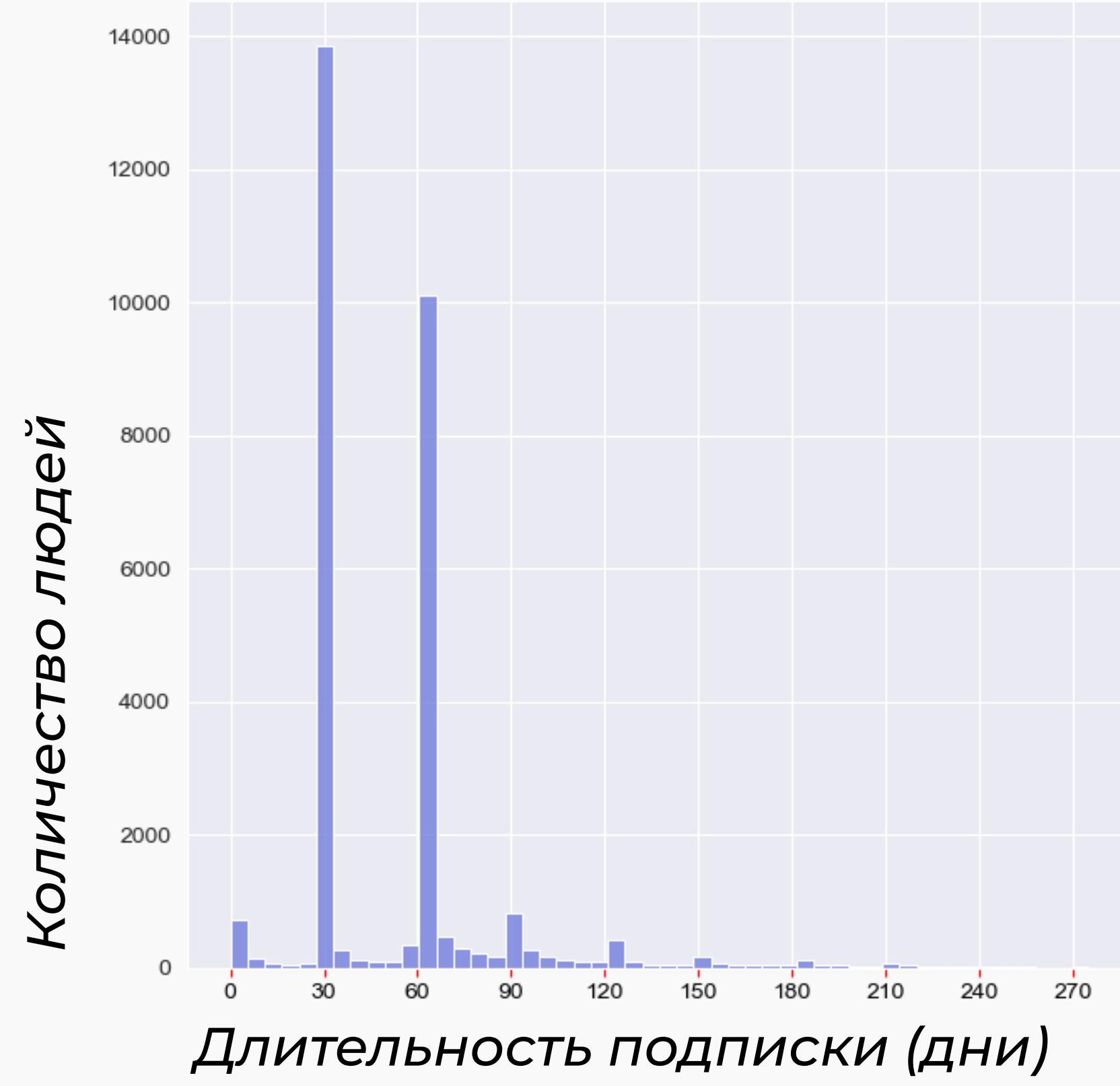
Люди, которые оформили подписку и у которых был пробный период 1 месяц



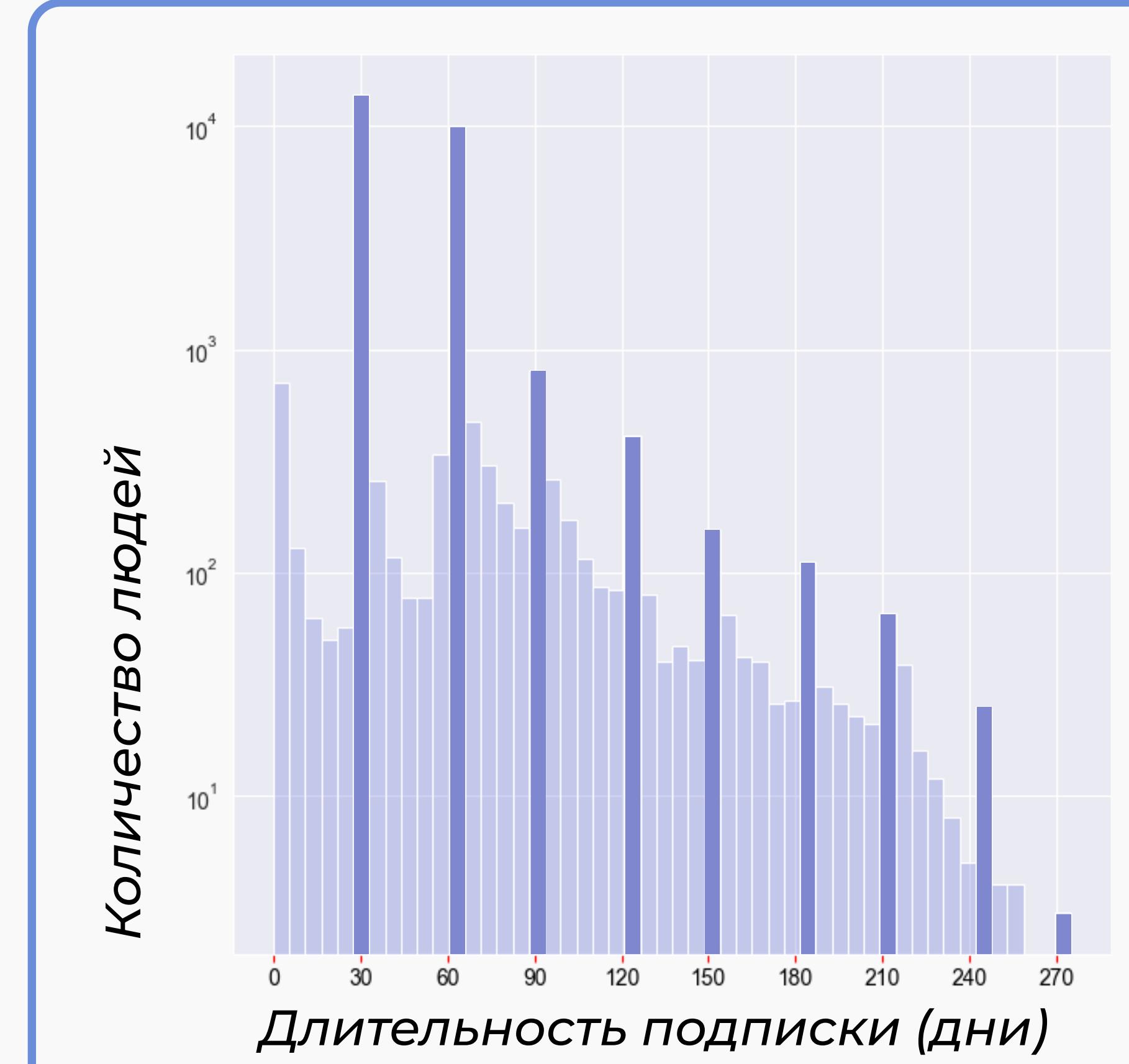
Предварительный анализ



Предварительный анализ



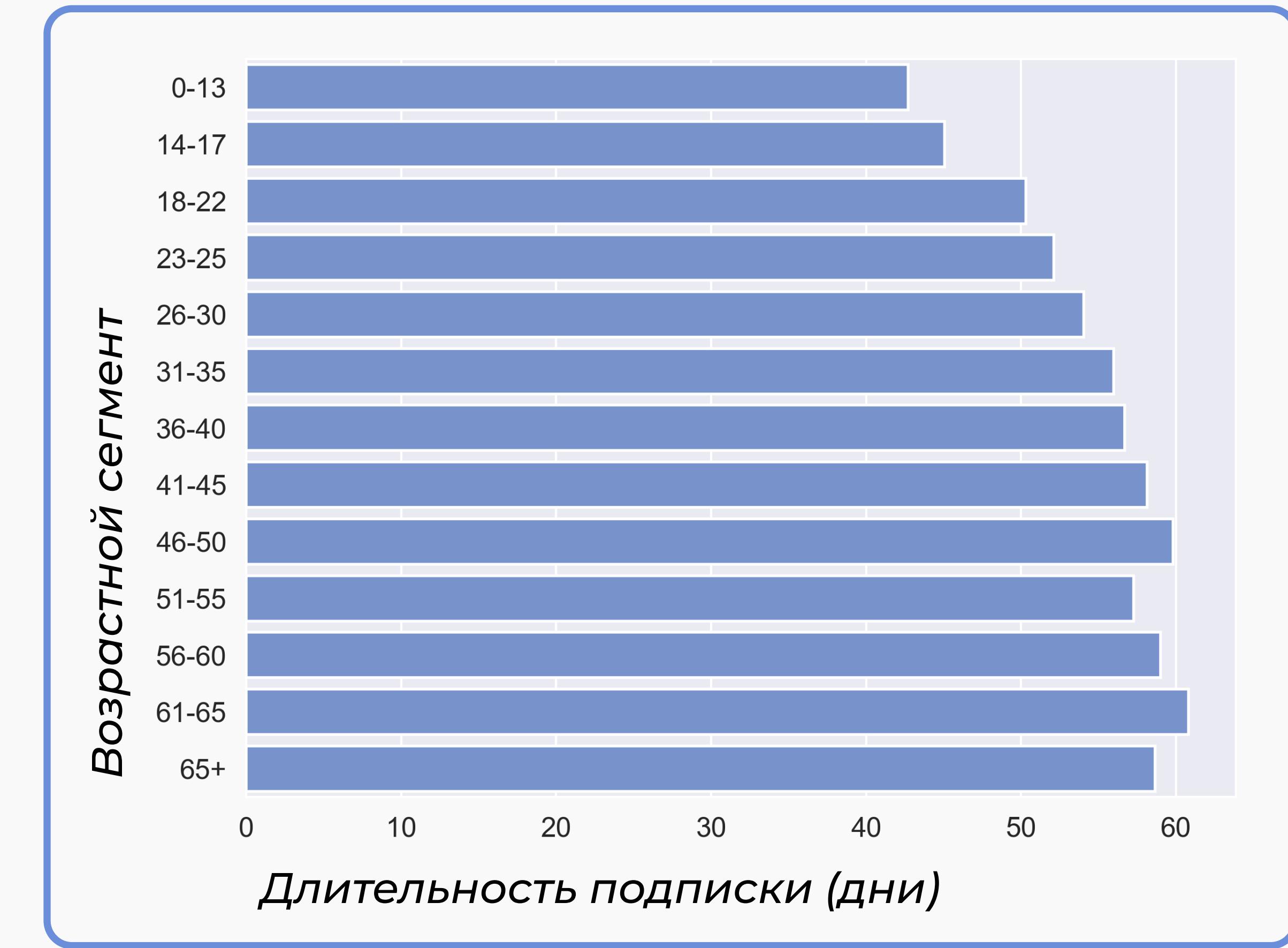
Линейная шкала



Логарифмическая шкала

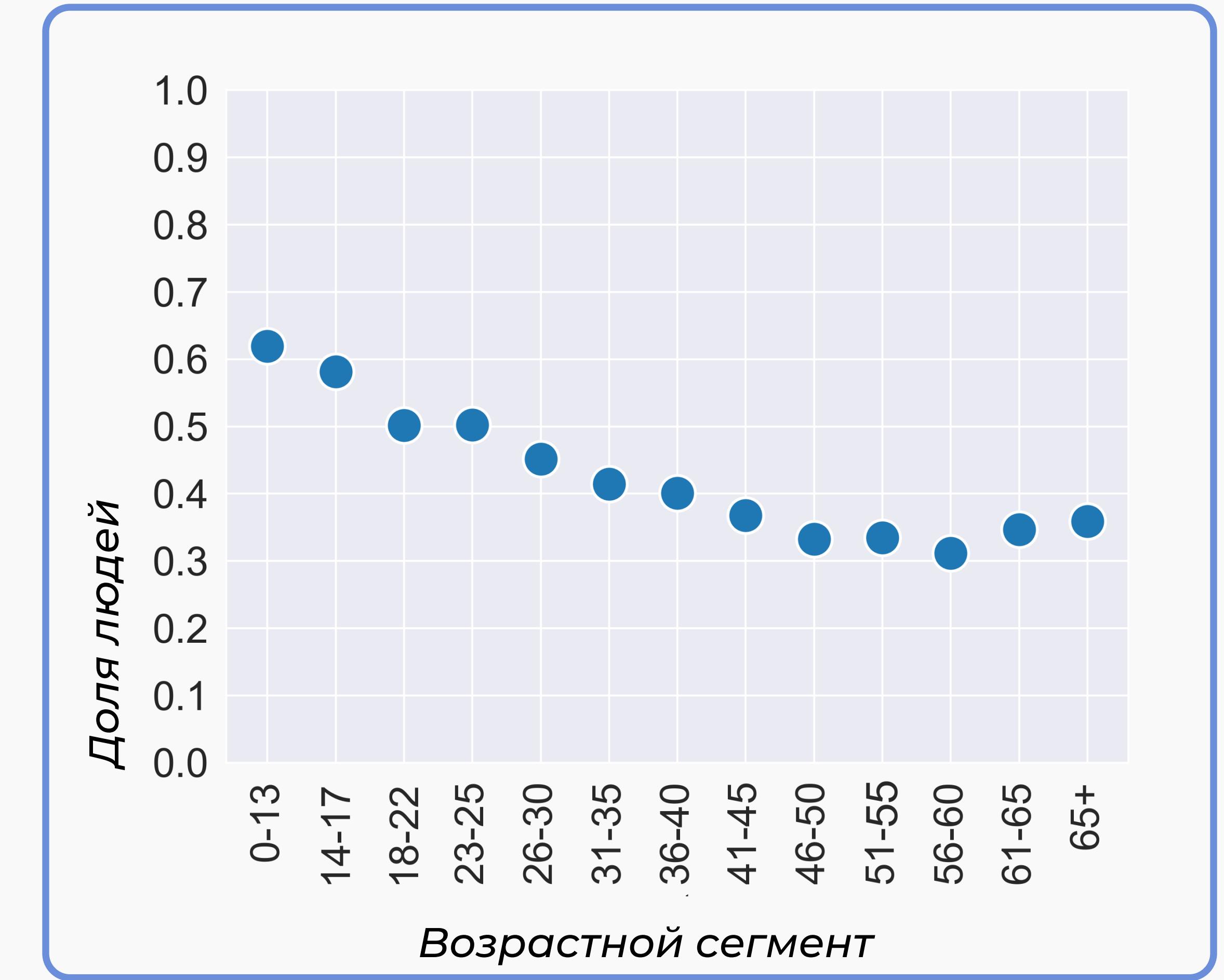
Предварительный анализ

Среднее значение
времени подписки в
каждом возрастном
сегменте



Предварительный анализ

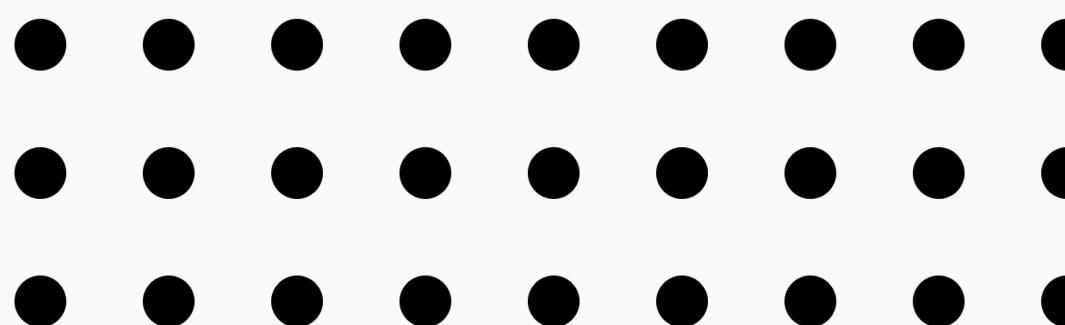
Доля людей,
отказавшихся платить
за подписку после
пробного периода





Исследовательский вопрос

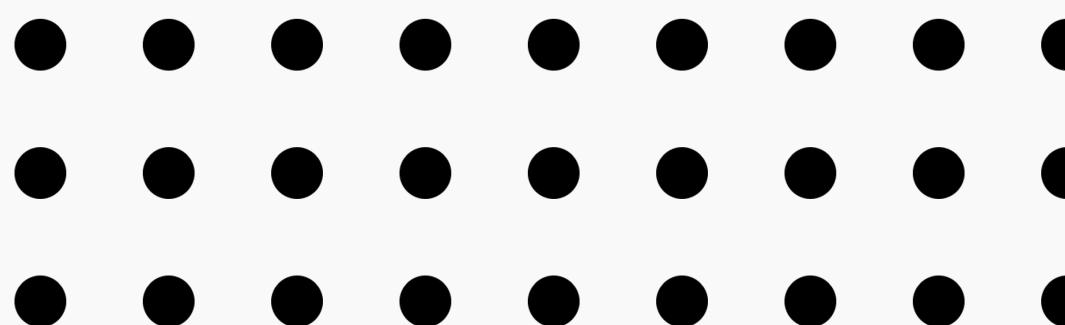
Как возраст влияет на продление
подписки после окончания пробного
периода?



Гипотеза



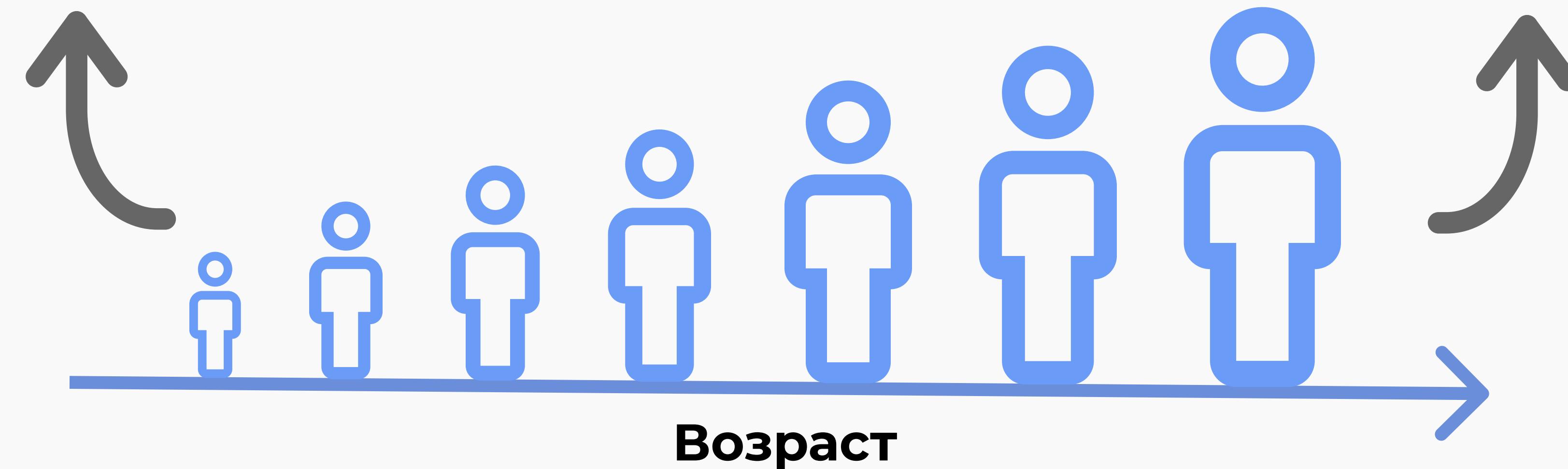
*С увеличением возраста уменьшается
доля людей, отказывающихся от
продления подписки после окончания
пробного периода*



Механизм

**Стимулы оформлять
подписку лишь ради
бесплатного пробного
периода**

**Стимулы оформлять
платную подписку
именно ради её
бонусов**



Подготовка данных

Замена текстовой переменной
`age_segment` на категориальную
`age_category`

Обработка выборок малого размера:

- Удаление данных о пользователях младше 14 и старше 60
- Объединение возрастных сегментов 51-55 и 56-60

Создание новой бинарной переменной `cancelled_trial`



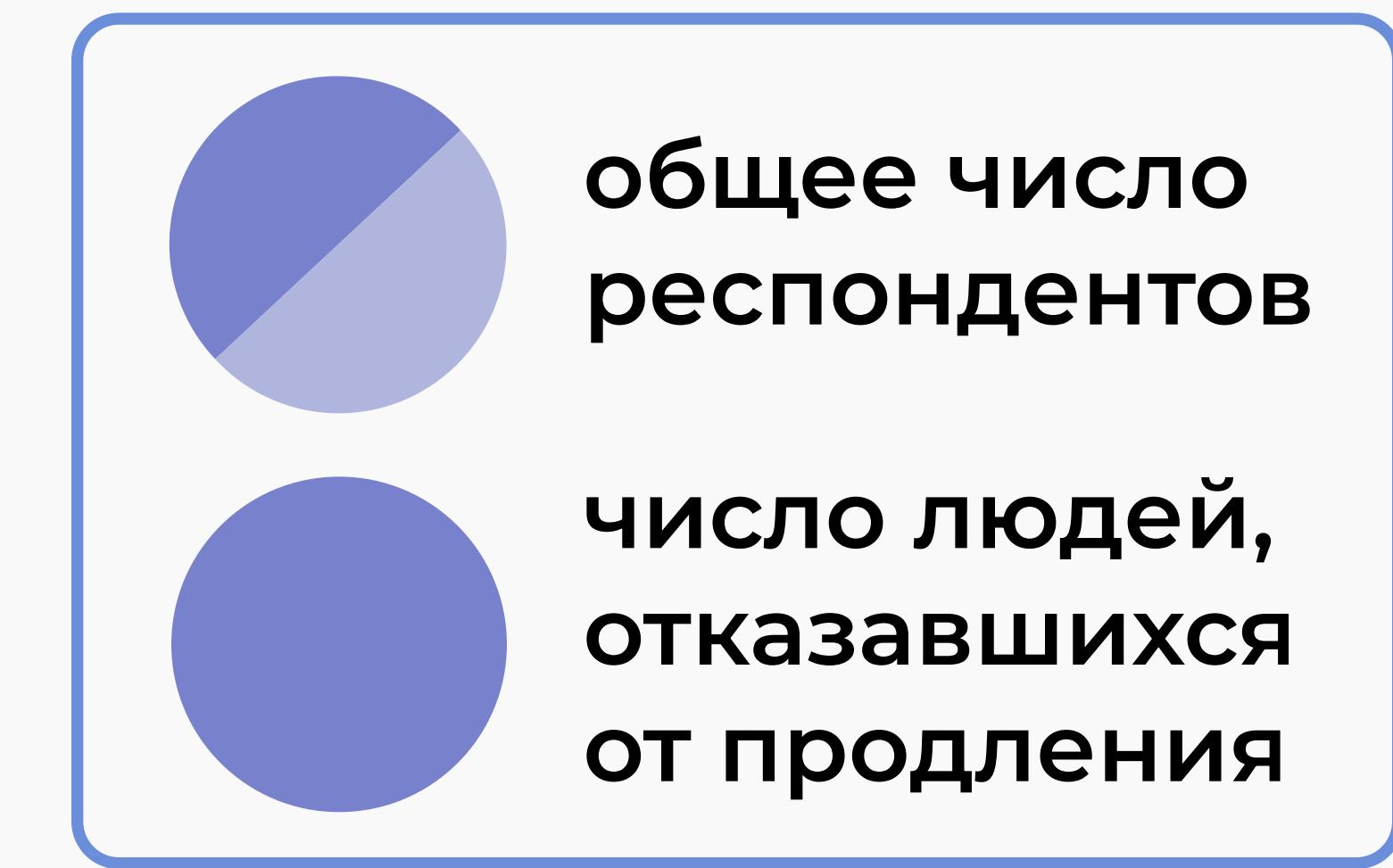
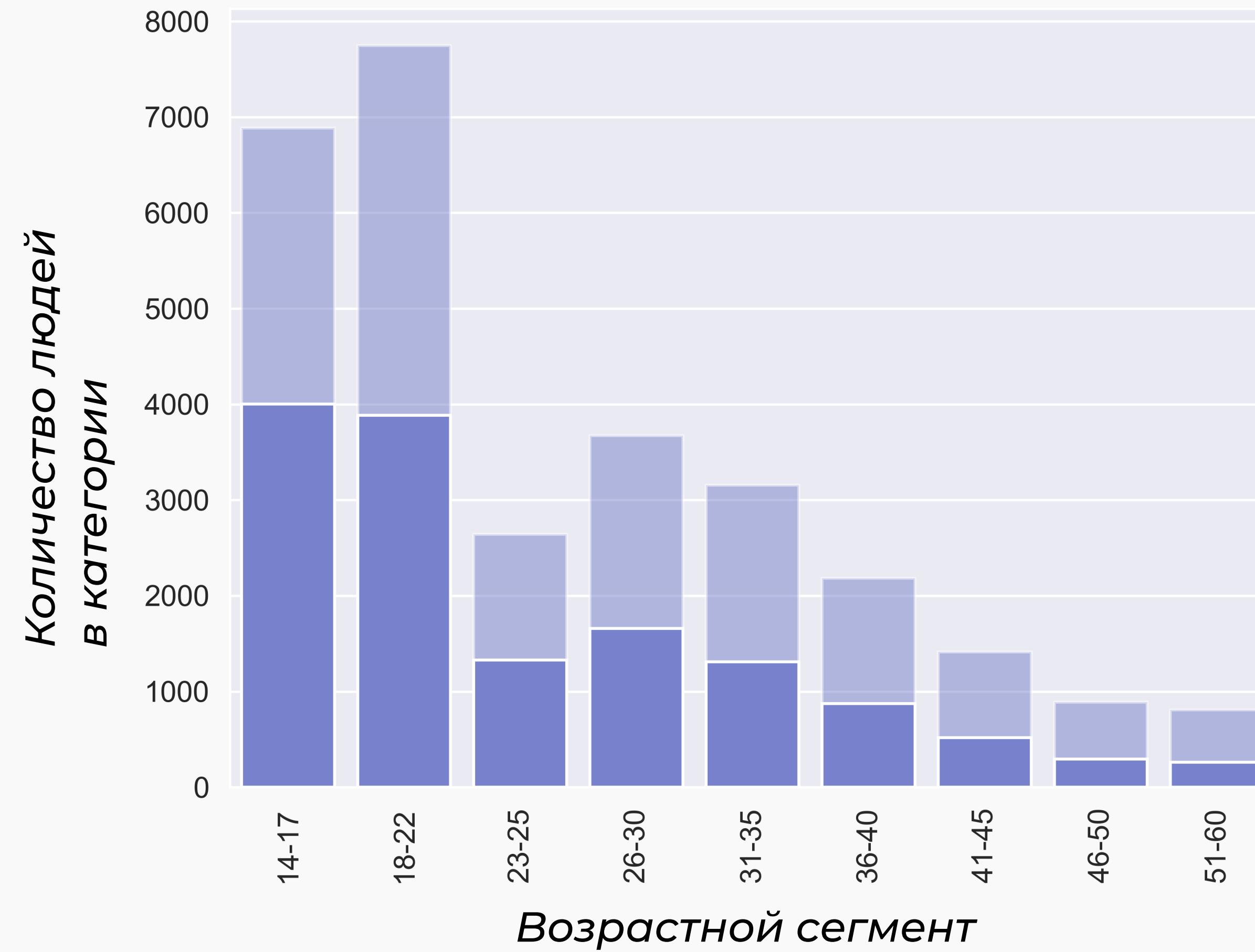
Исследуемые данные

Флаг отказа после пробного периода
(cancelled_trial*)

Возрастная категория
(age_category)

*В каждой возрастной категории считаем долю людей, которые не продлили подписку после окончания пробного периода

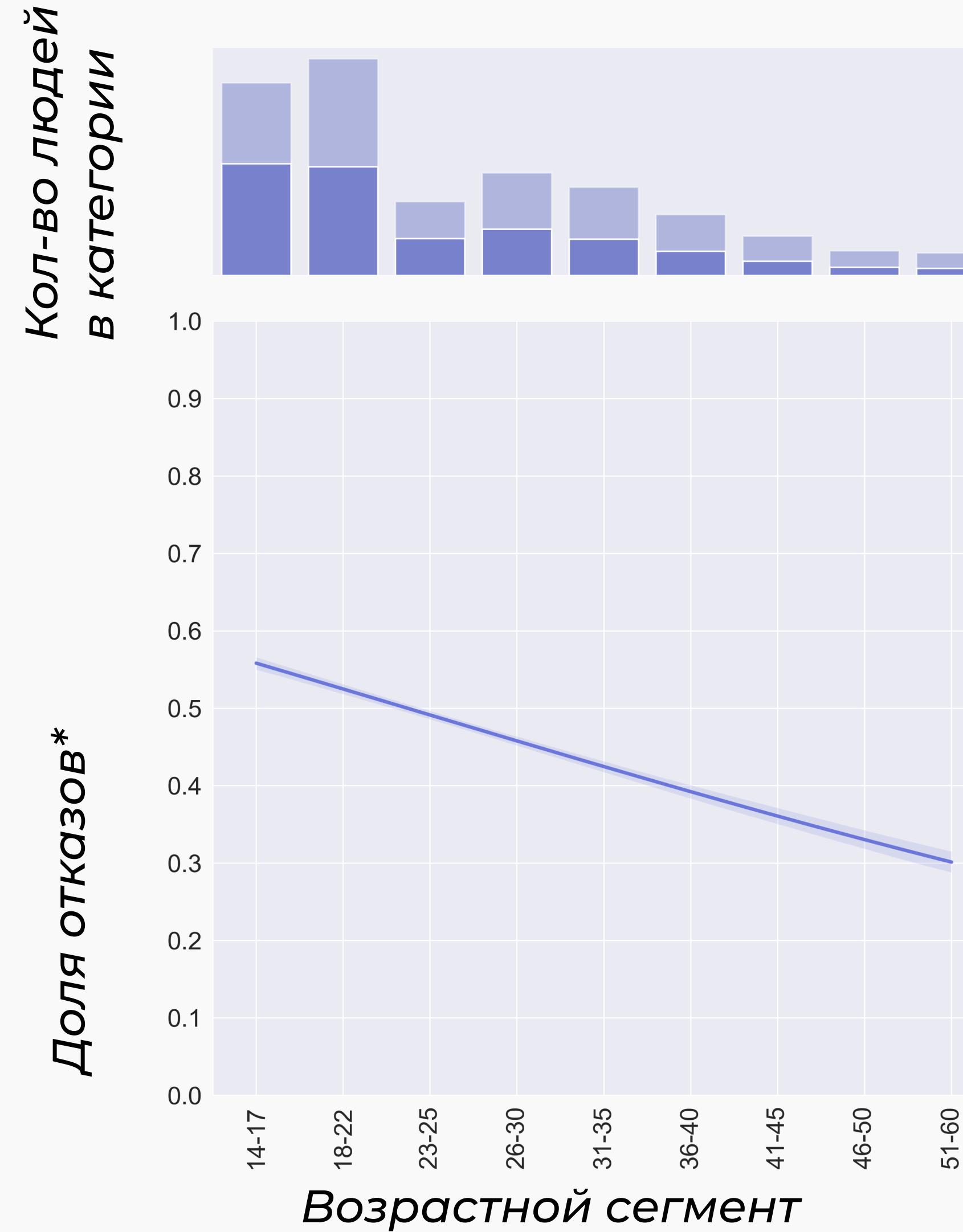
Исследуемые данные



Распределение людей из выборки по возрастному сегменту

Математическая модель

Пробит регрессия



1 – отказ от подписки
0 – продление подписки

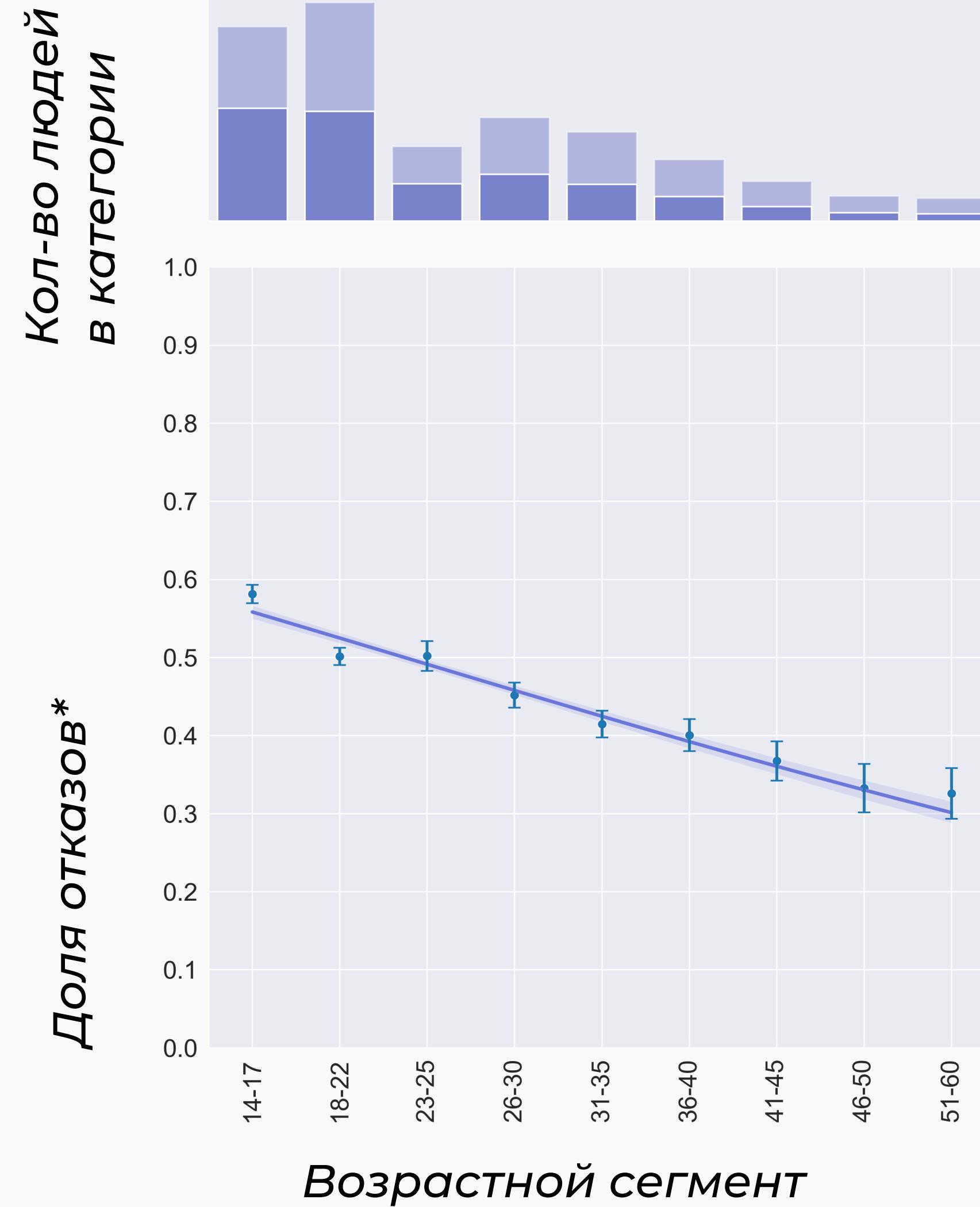
$$\hat{y} = \frac{1}{1 + e^{-(\text{const} + a_c * x)}}$$

	coef	P > z	0.025	0.975
const	0.2298	0.000	0.203	0.256
age_category (a_c)	-0.0837	0.000	-0.090	-0.077

*Доля отказов от подписки сразу после окончания пробного периода

Математическая модель

Пробит регрессия



1 – отказ от подписки
0 – продление подписки

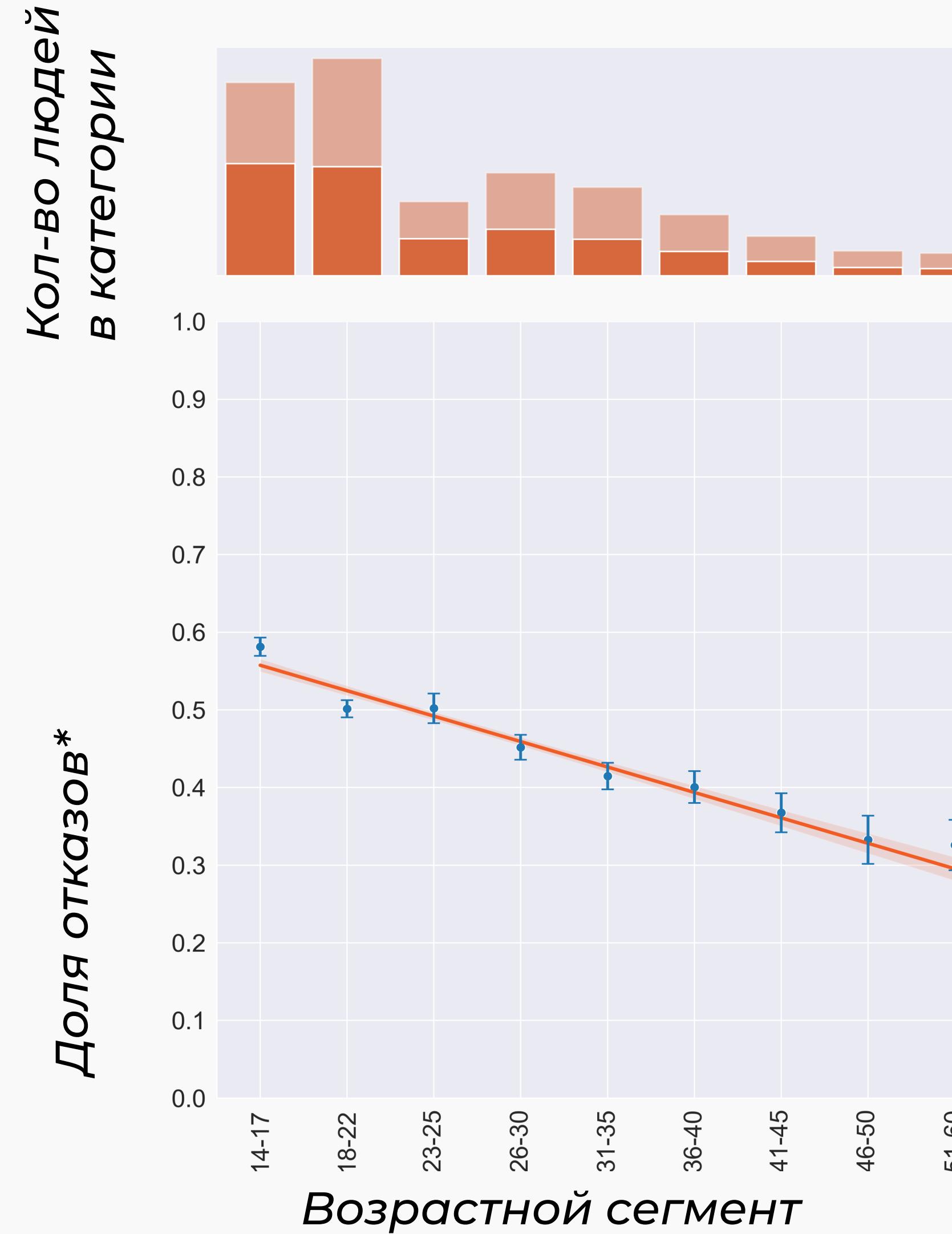
$$\hat{y} = \frac{1}{1 + e^{-(\text{const} + a_c * x)}}$$

	coef	P > z	0.025	0.975
const	0.2298	0.000	0.203	0.256
age_category (a_c)	-0.0837	0.000	-0.090	-0.077

*Доля отказов от подписки сразу после окончания пробного периода

Математическая модель

Линейная регрессия



1 – отказ от подписки
0 – продление подписки

$$\hat{y} = \text{const} + a_c * x$$

	coef	P > z	0.025	0.975
const	0.5902	0.000	0.580	0.600
age_category (a_c)	-0.0327	0.000	-0.035	-0.030

*Доля отказов от подписки сразу после окончания пробного периода

Добавление в регрессию новой переменной

От двух переменных

	coef	P > z	0.025	0.975
const	0.2283	0.000	0.191	0.265
age_category	-0.1012	0.000	-0.109	-0.094
income*	7.559e-07	0.000	7e-07	8.13e-07

От одной переменной

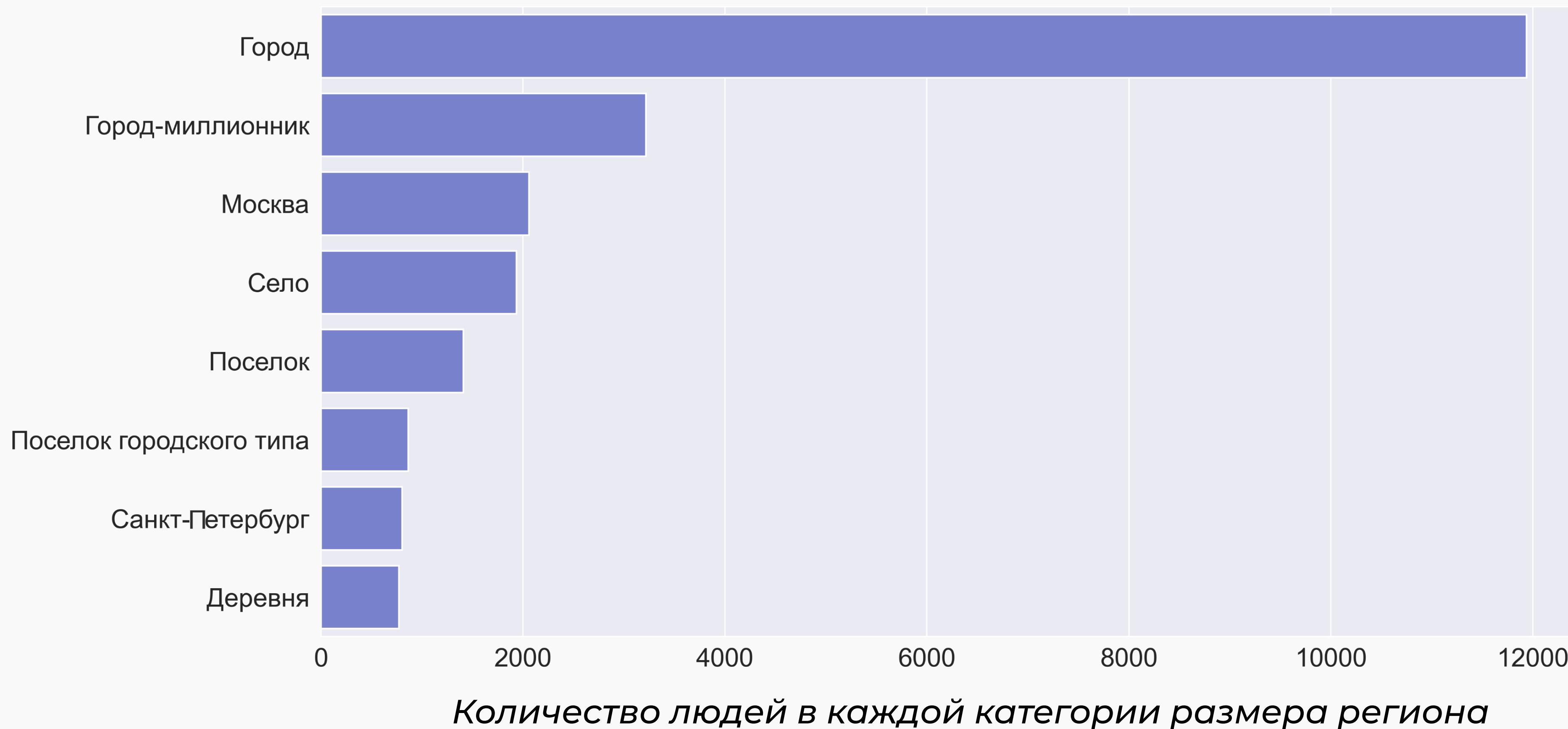
	coef	P > z
const	0.2298	0.000
age_category	-0.0837	0.000

Статистическая значимость возрастной
переменной сохраняется

*income – переменная дохода респондента, поступающего на карты/счета Тинькофф банка

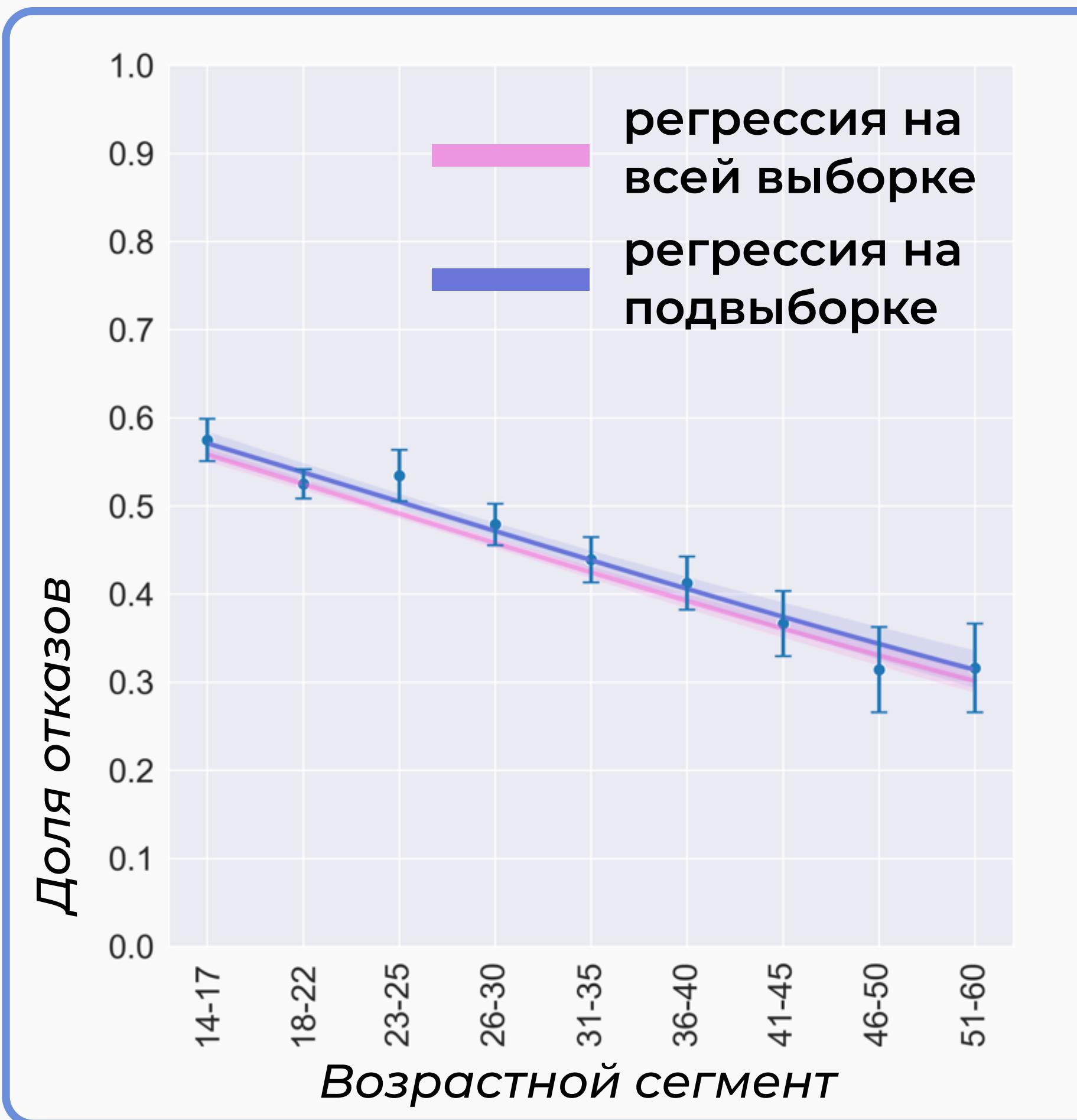
Анализ по подвыборкам

Разделение на подвыборки по размеру региона респондента

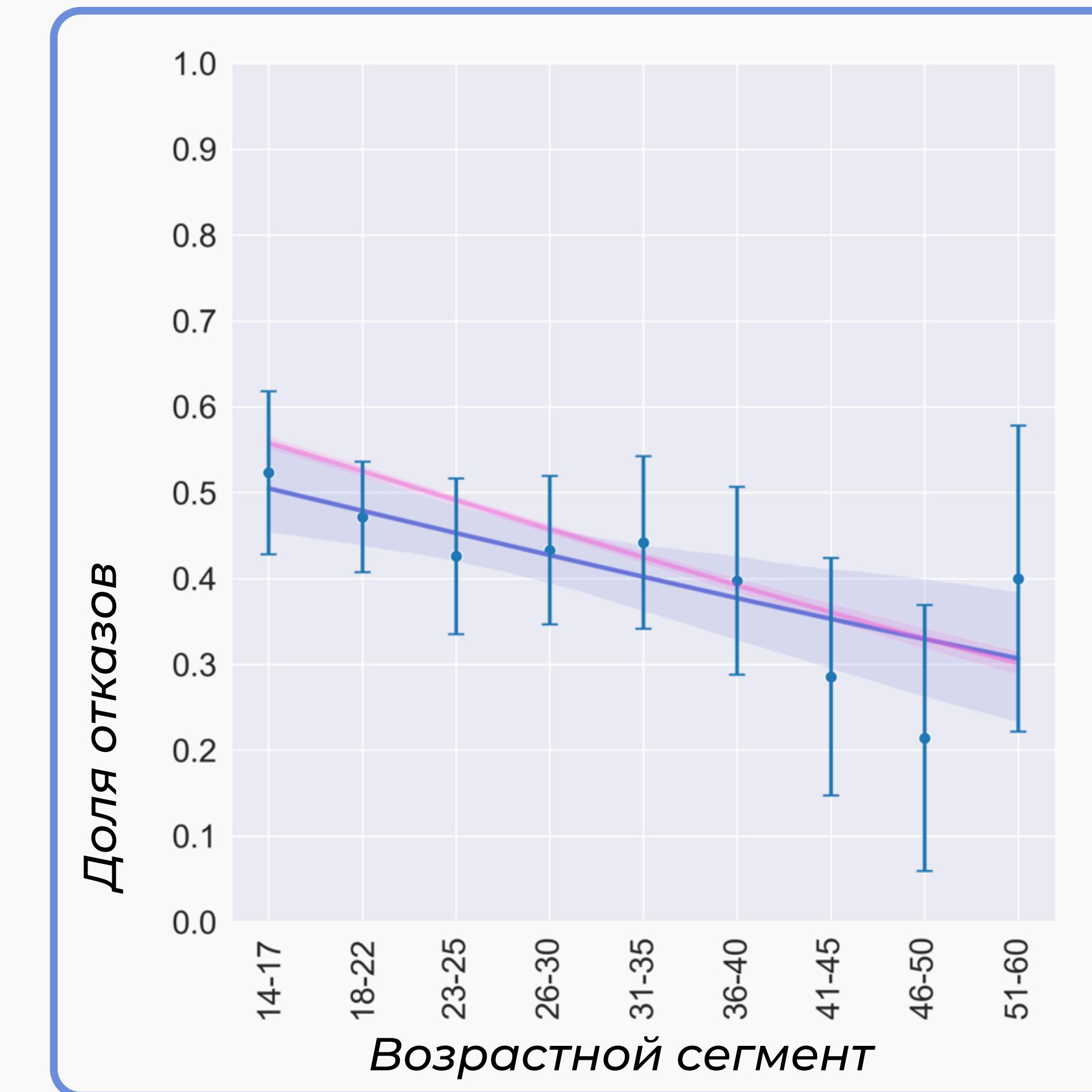


Анализ по подвыборкам

Пробит регрессия



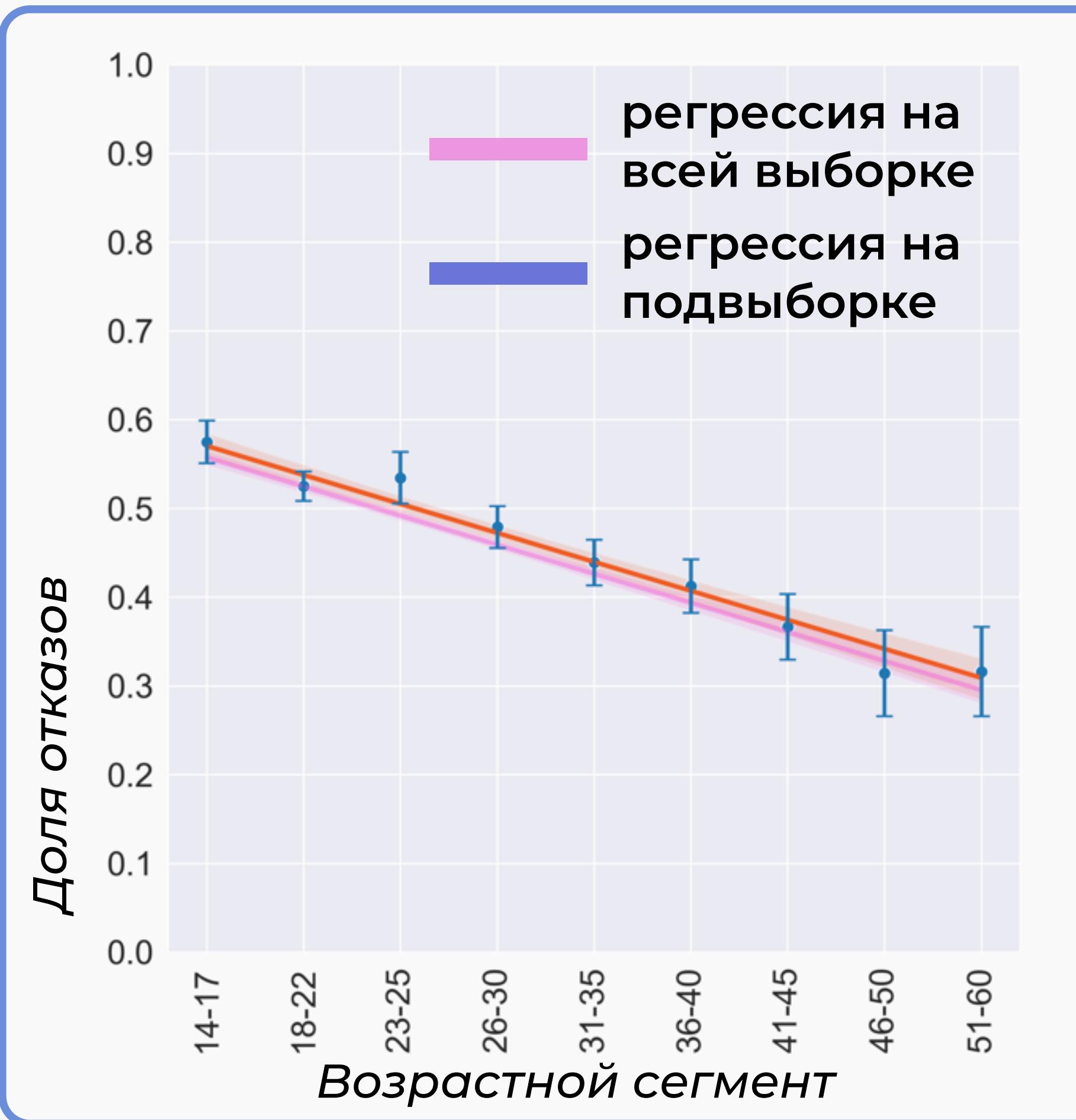
Город



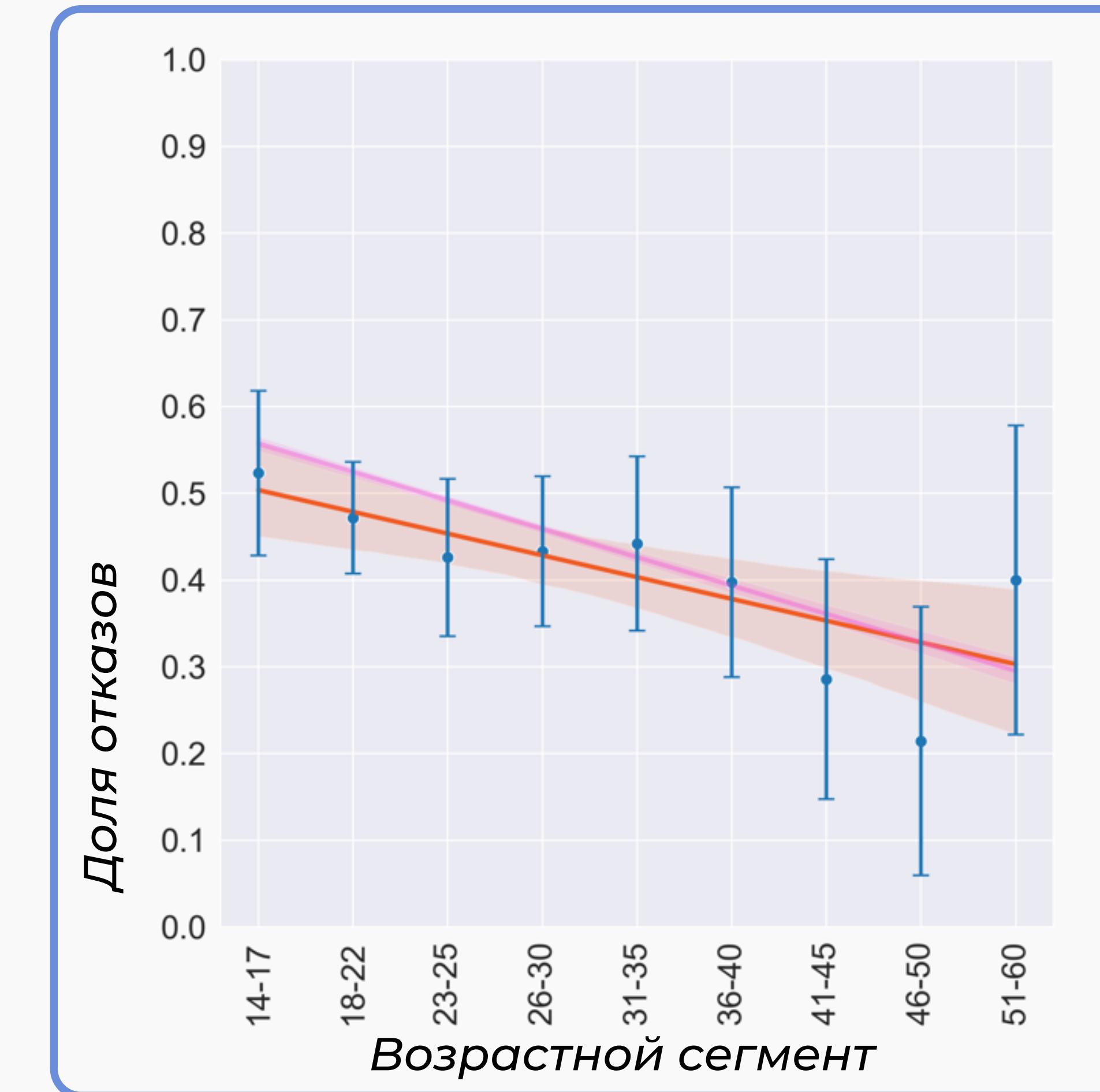
Поселок городского типа

Анализ по подвыборкам

Линейная регрессия



Город



Поселок городского типа

Интерпретация

Между переменными возраста и доли отказов от подписки существует статистически значимая отрицательная взаимосвязь

Модель устойчива, т.к. знак и р-значение коэффициента при независимой переменной сохраняется при добавлении новой переменной

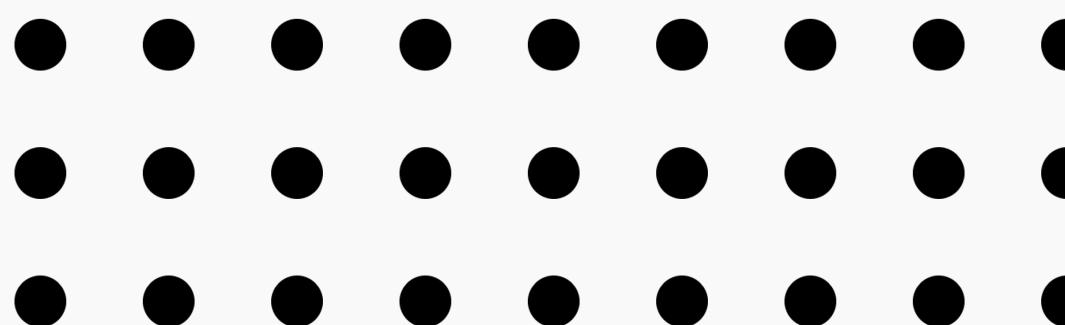
Вывод



Для максимизации LTV следует:

Привлечь больше
людей среднего и
старшего возраста
31-60 лет

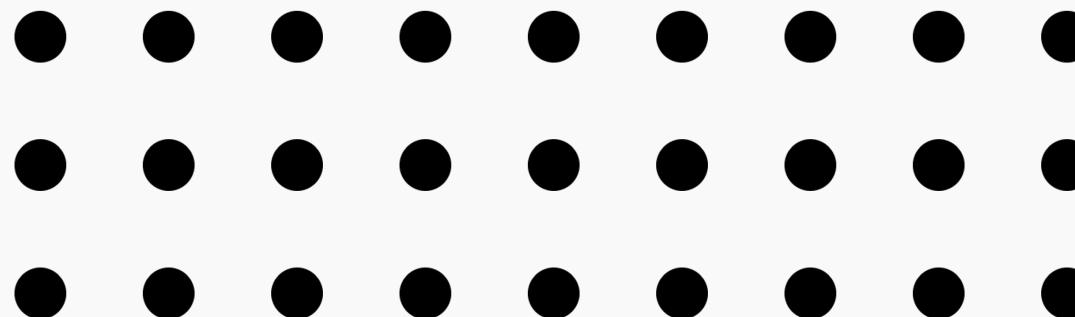
Увеличить
конверсию в
платную подписку
среди молодых
людей



Практическая польза



Tinkoff сможет учитывать поведение каждого возрастного сегмента за время использования пробного периода при внедрении новых фич в подписку



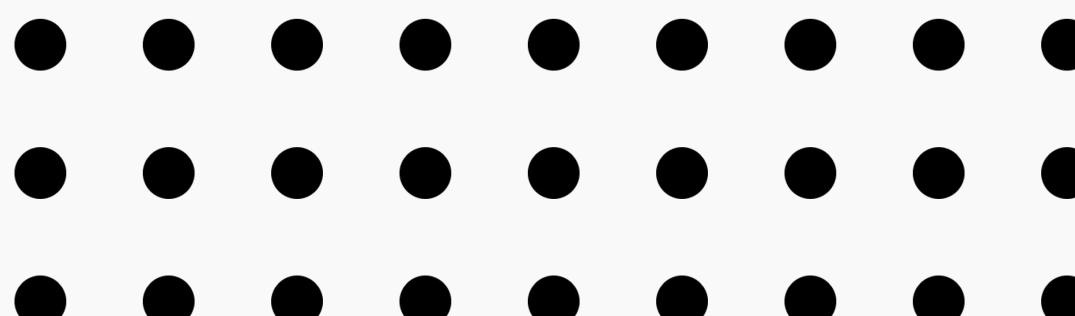


Policy Implication

Для привлечения людей постарше:

Партнерство с:

- частными клиниками
(«Академика Ройтберга»)
- частными детскими садами
(«MAGIC CASTLE Education»)

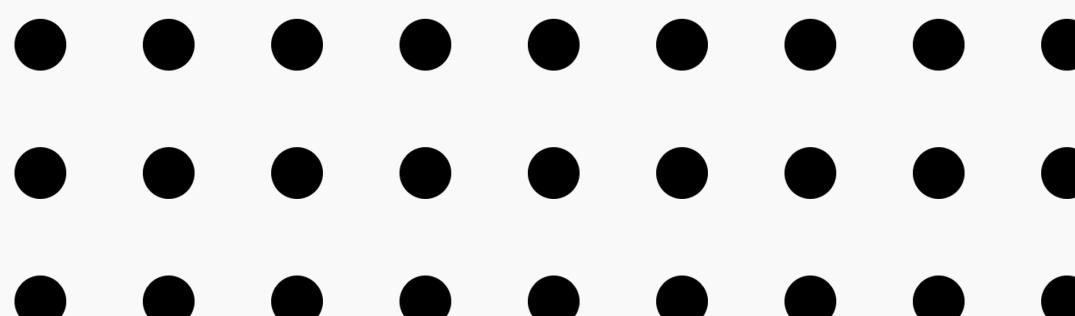


Policy Implication

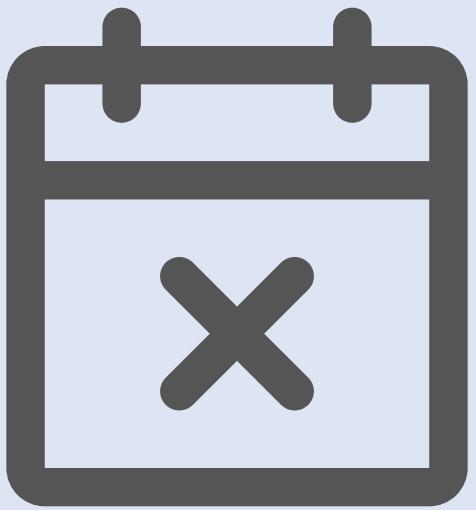


Для увеличения конверсии среди молодых:

Предоставление льгот группе людей в
возрасте 18-22 года
(скидки 50% на оформление подписки)



Ограничения



Вывод нашего исследования нельзя распространить на:

- 1** клиентов НЕ Tinkoff Банка
- 2** людей, обладающих подпиской Tinkoff Pro
- 3** респондентов НЕ из СНГ

Перспективы



Собрать данные, которых нет в изначальном датасете:

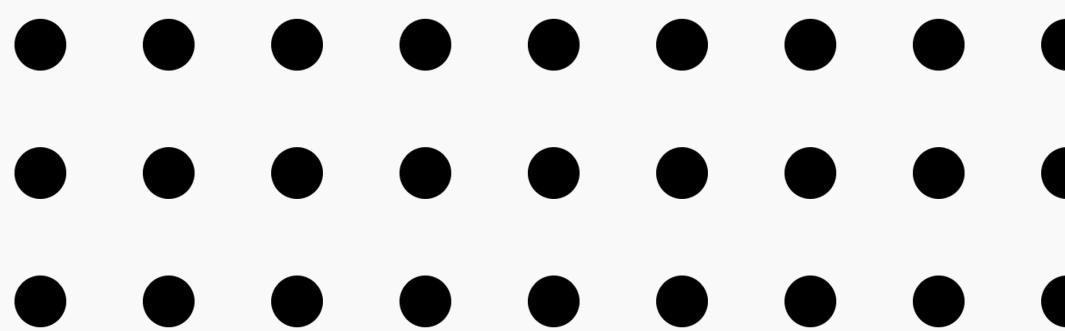
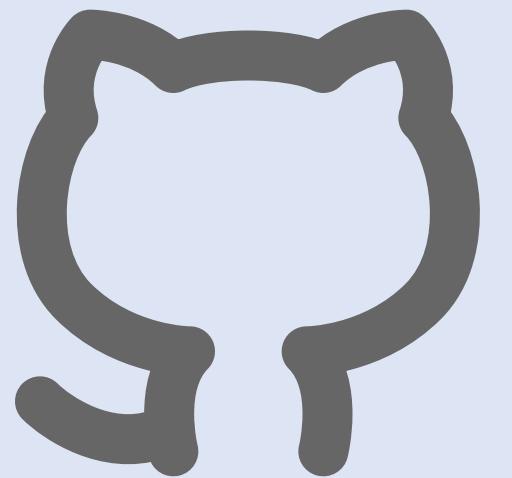
- конкретный возраст клиента
- конкретный регион клиента
- доход клиента вне Tinkoff банка
- гендер клиента



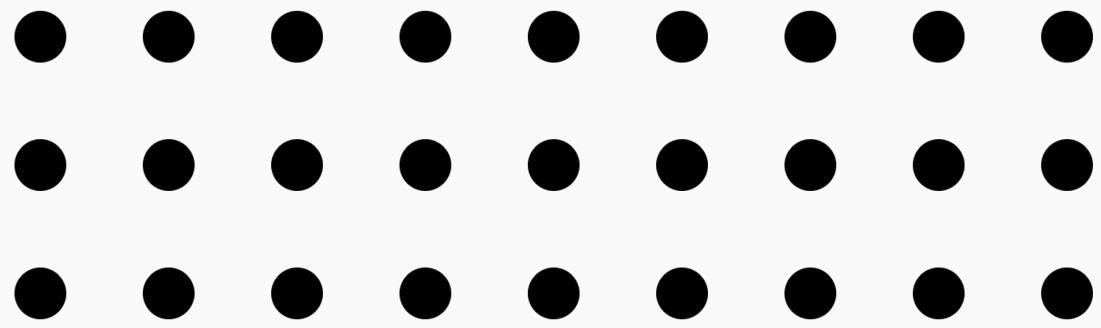
Исследовать зависимости на данных подписки Tinkoff Pro



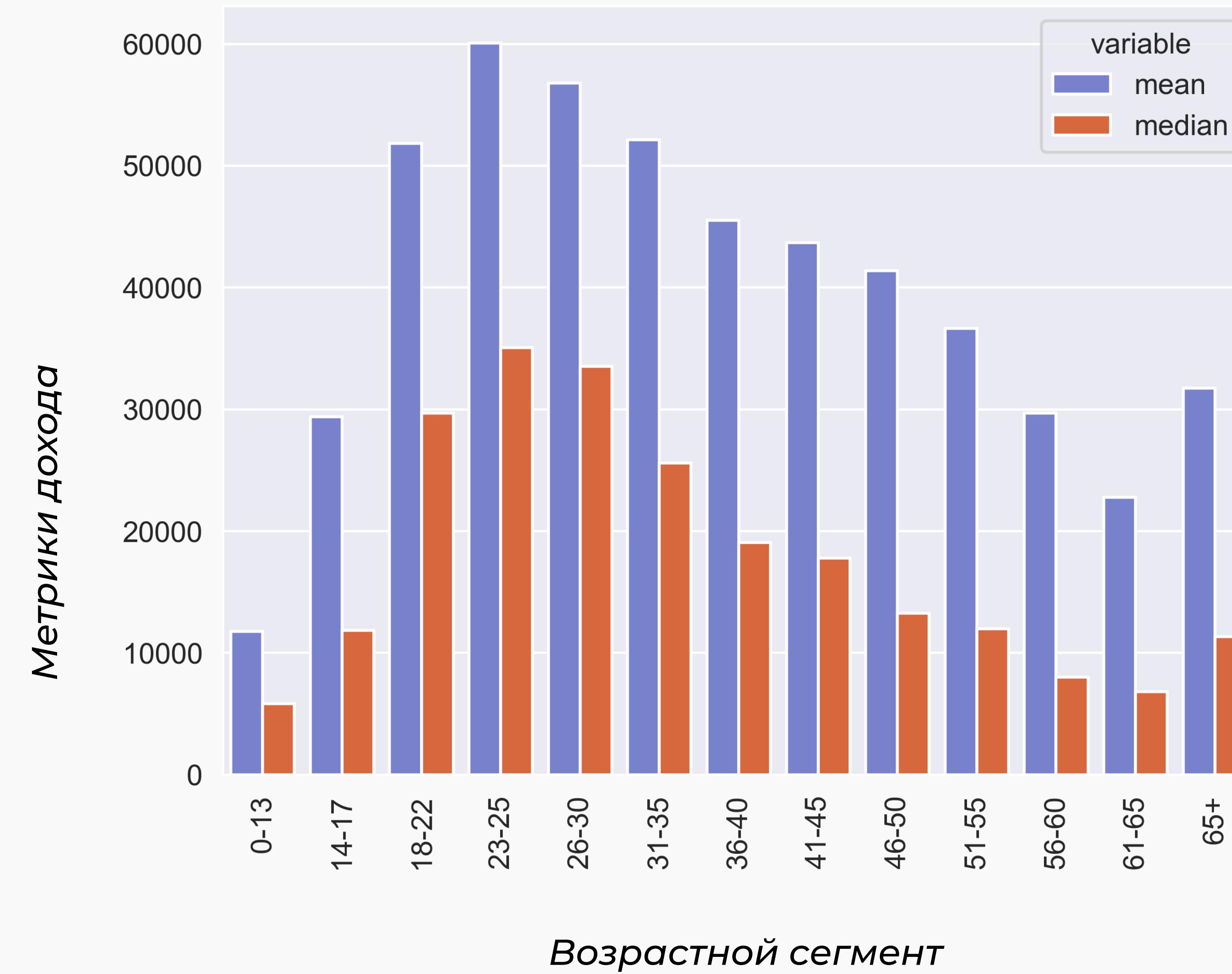
Наш Github Репозиторий



ПРИЛОЖЕНИЕ



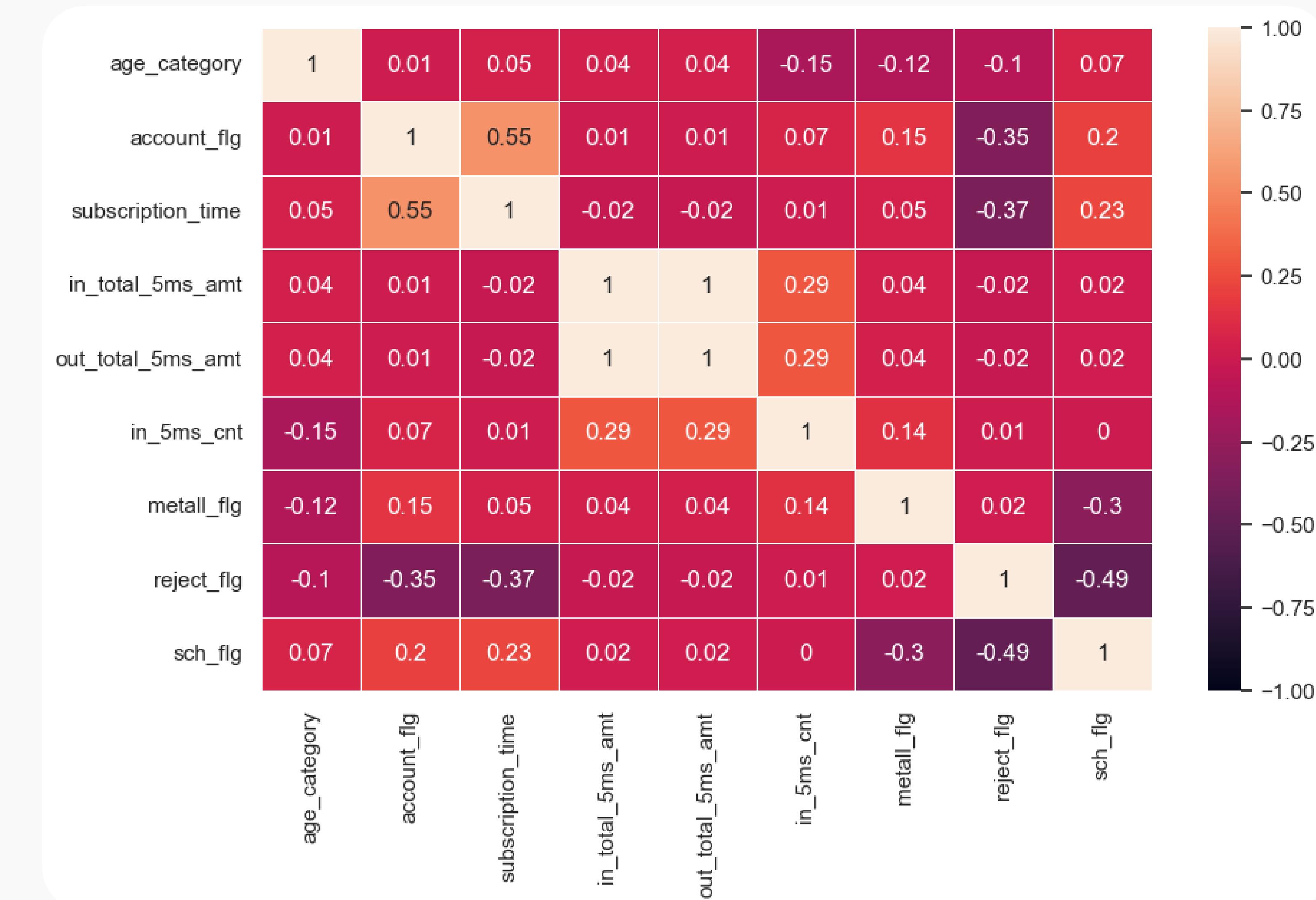
Распределение дохода по возрастным категориям



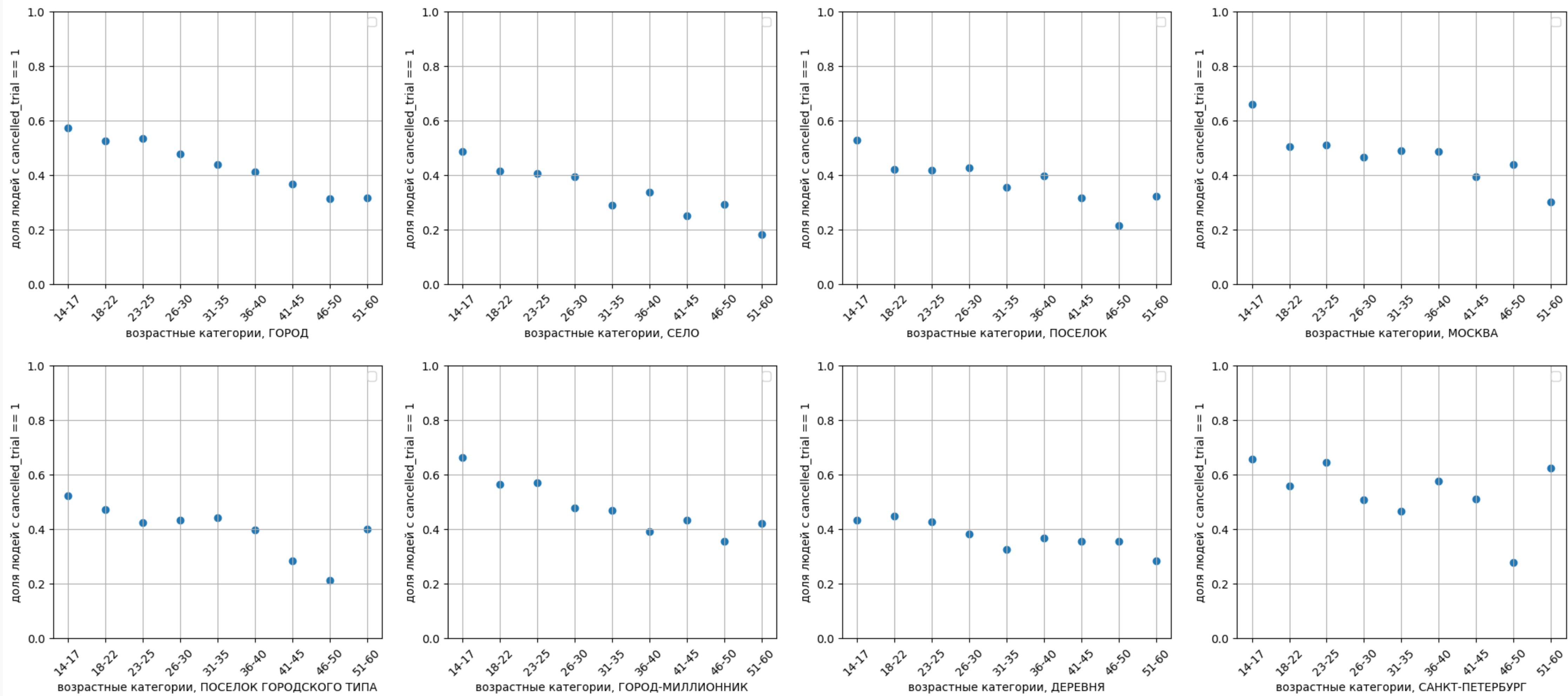
Распределение сумм транзакций пополнения за 5 месяцев до заявки



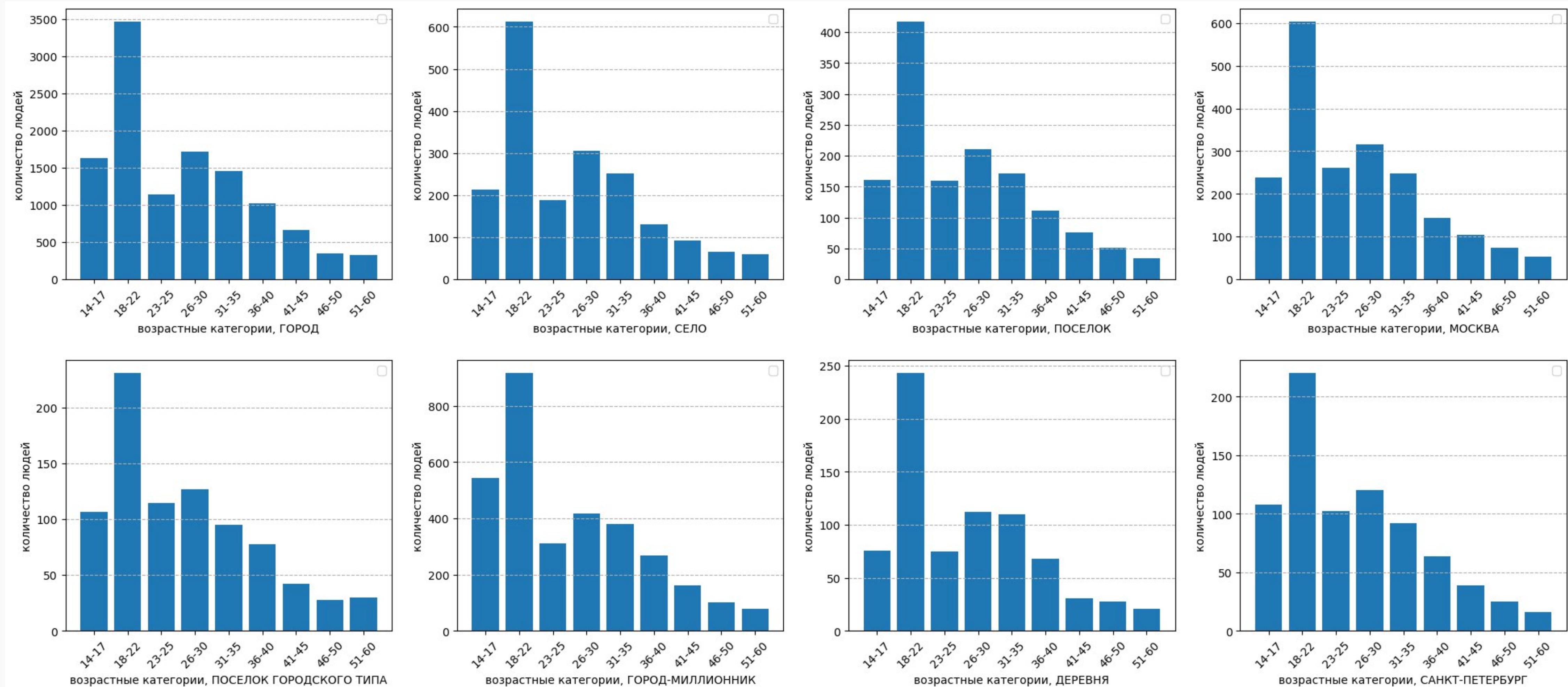
Матрица корреляций



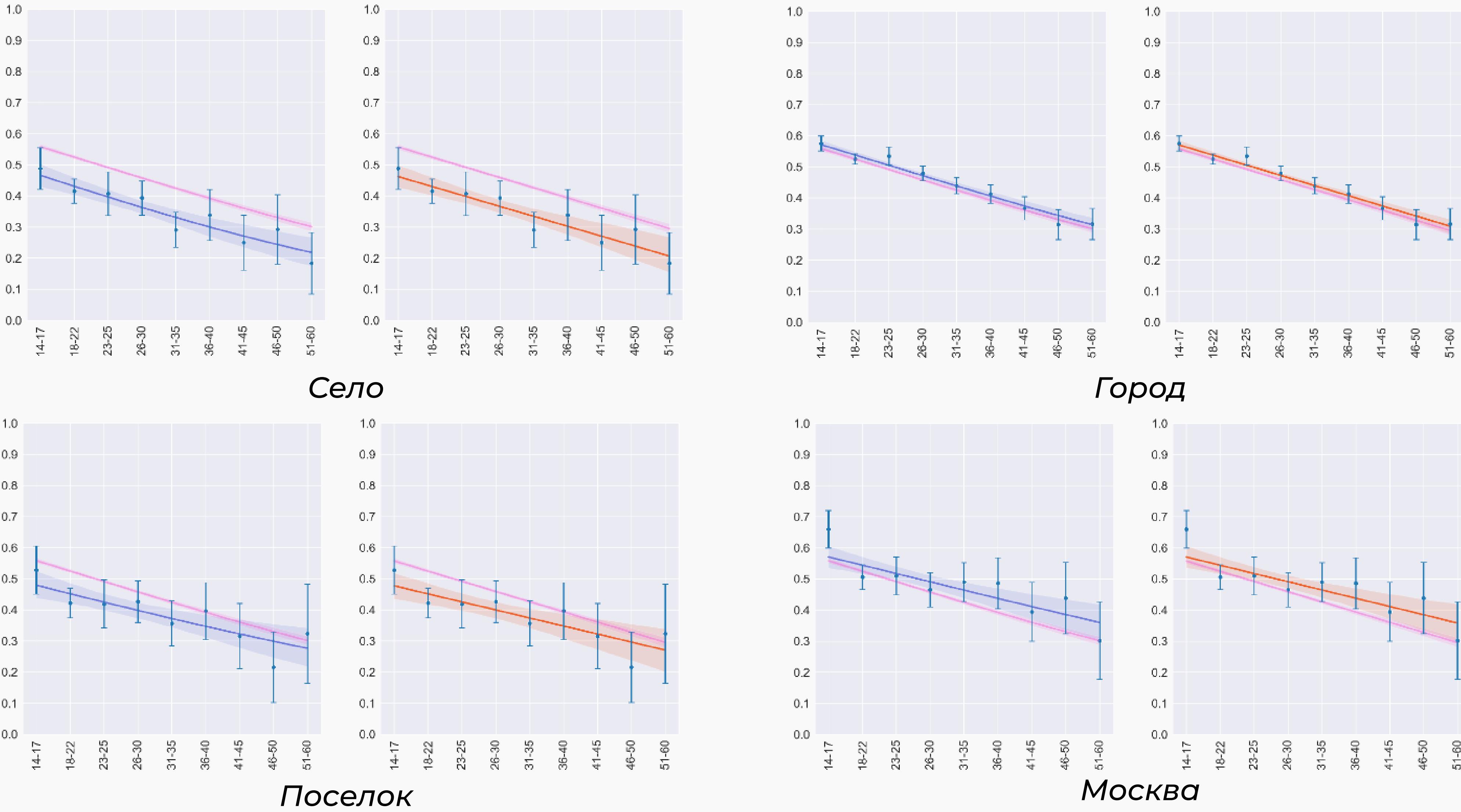
Анализ по подвыборкам - scatter plot



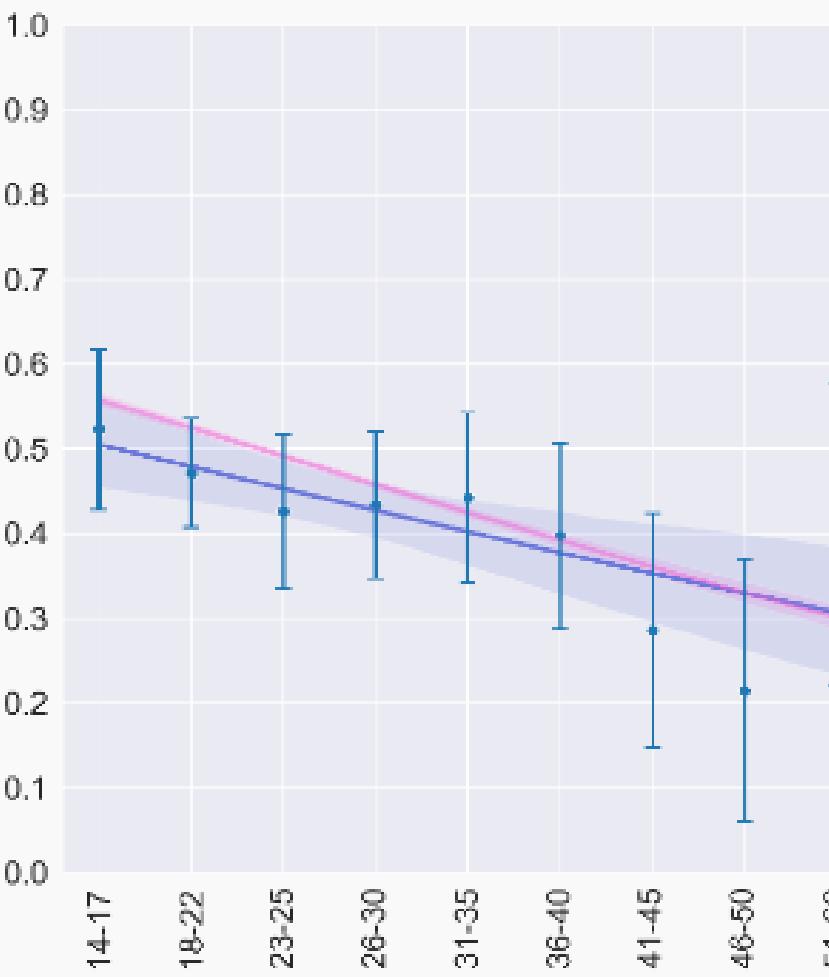
Анализ по подвыборкам - распределение людей в возрастных сегментах



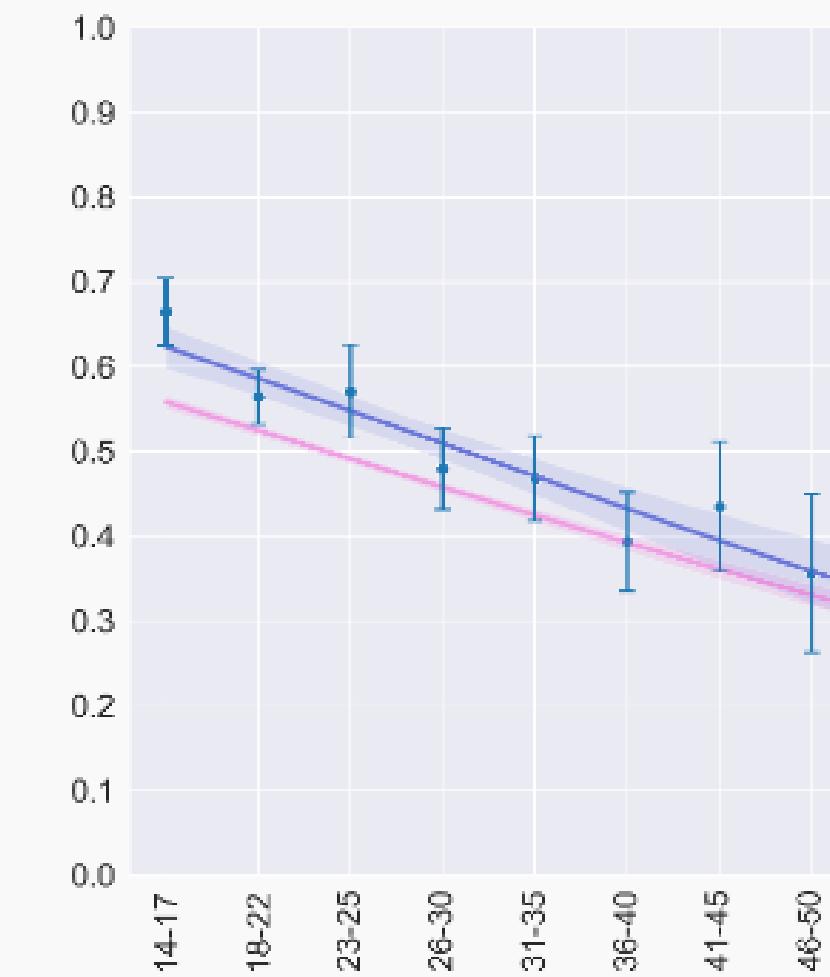
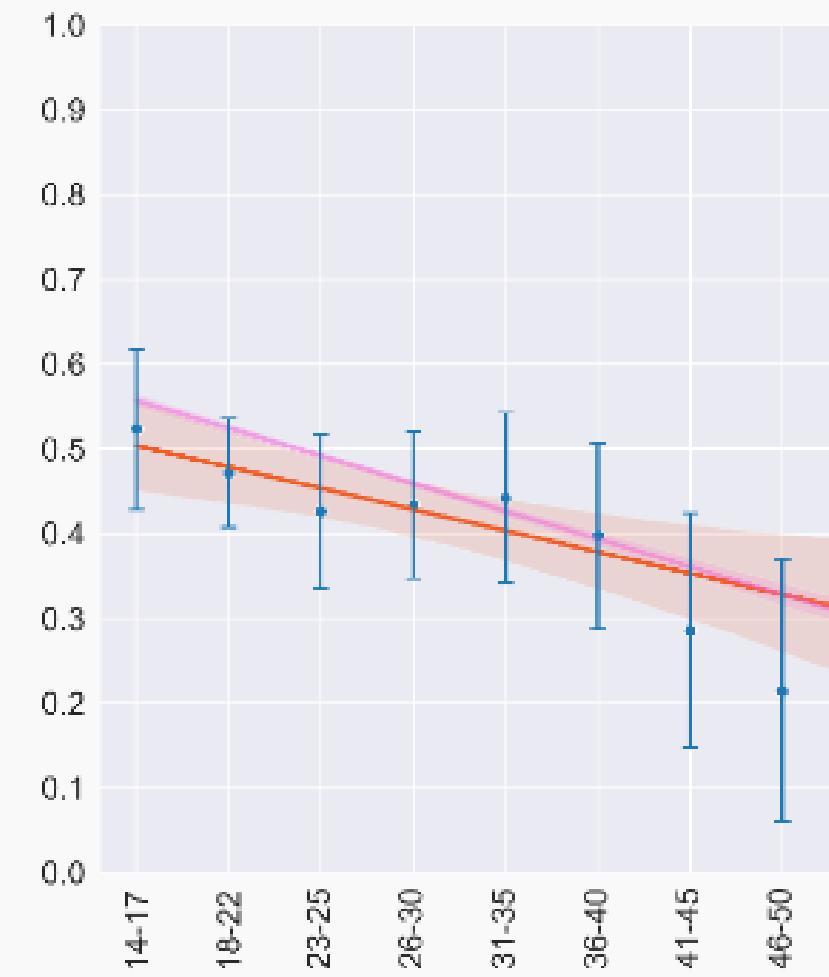
Анализ по подвыборкам - регрессии



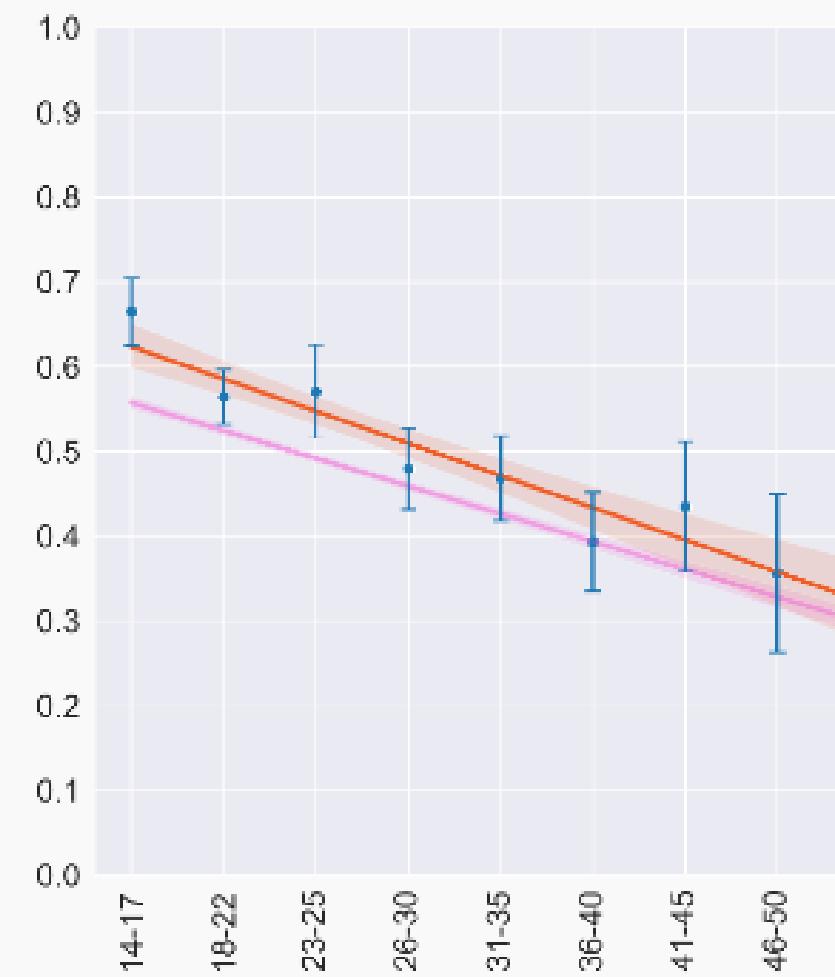
Анализ по подвыборкам - регрессии



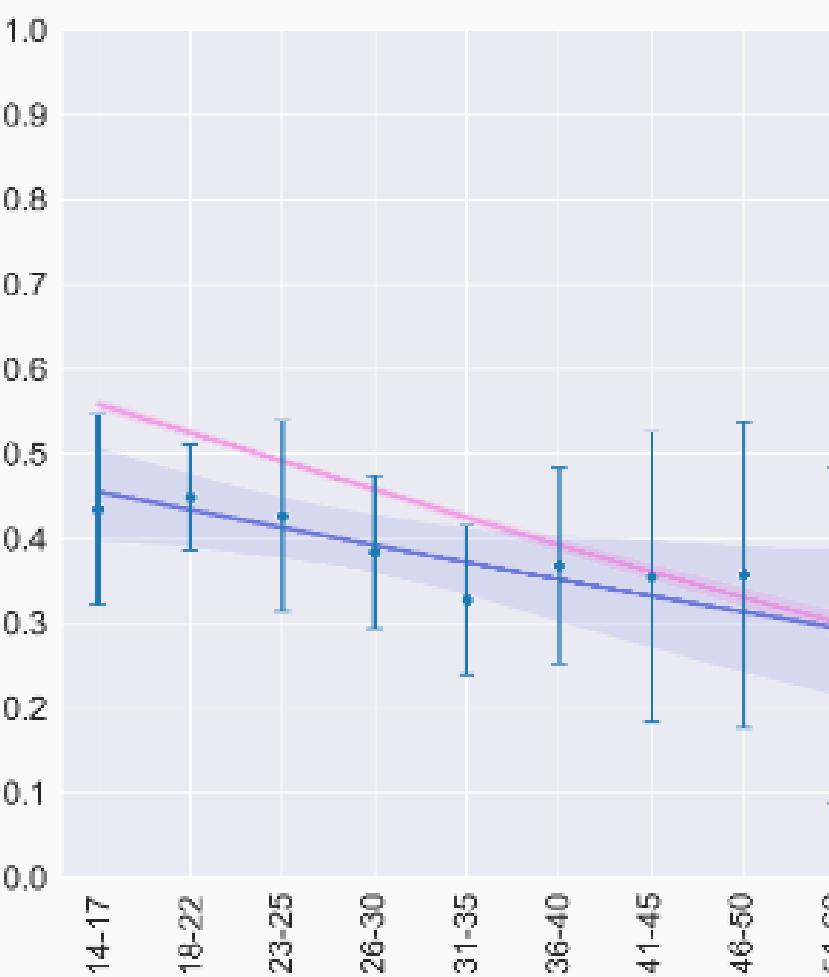
Поселок городского типа



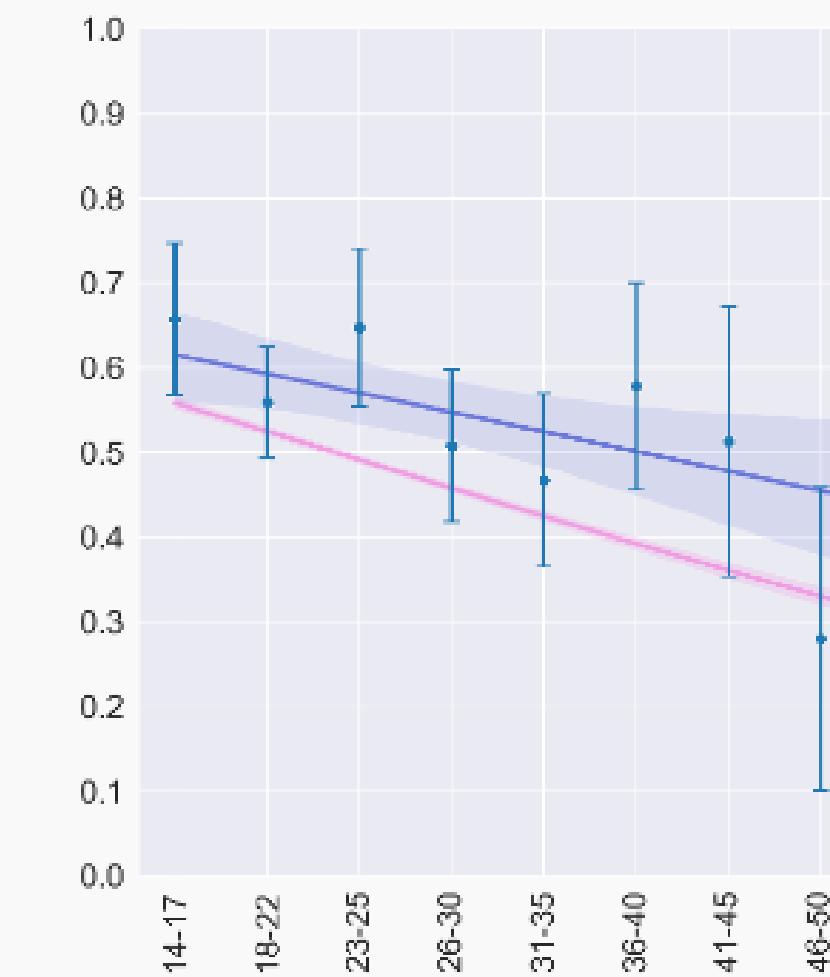
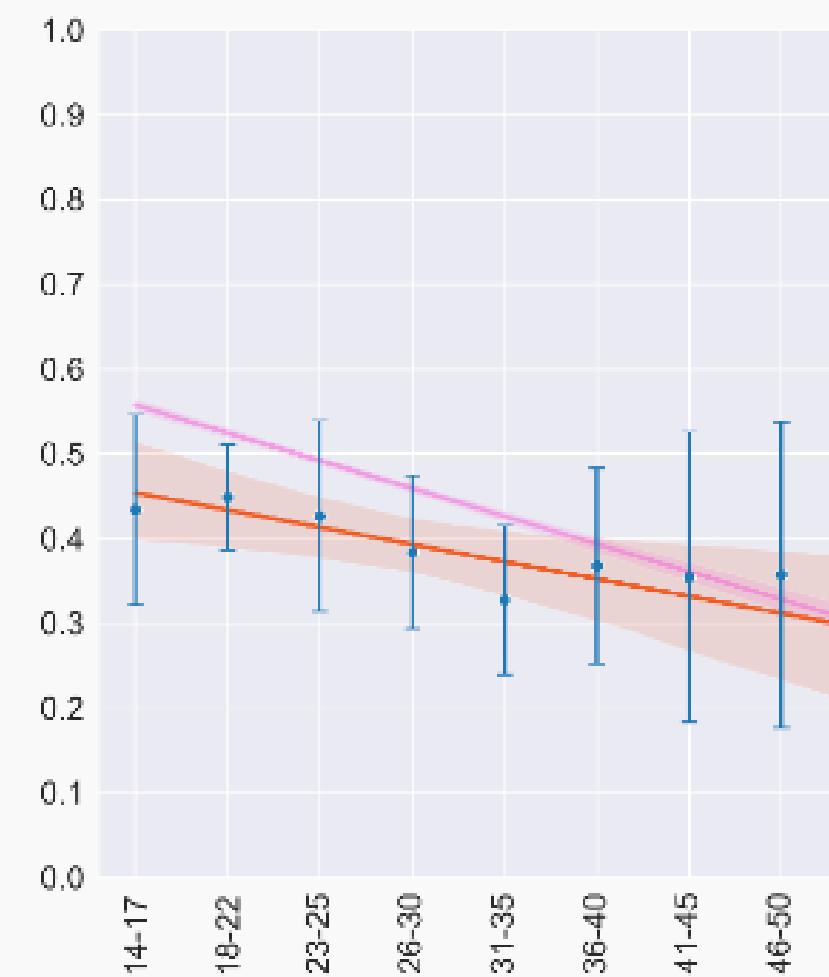
Город-миллионник



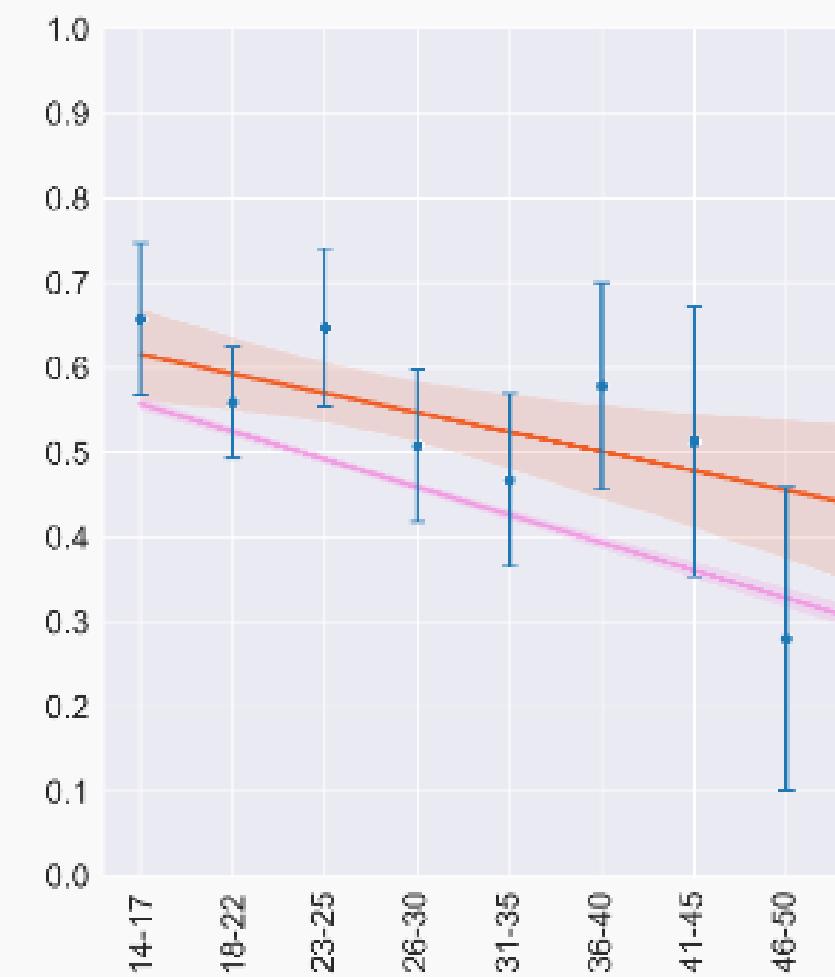
Поселок городского типа



Деревня



Москва



Анализ по подвыборкам (лин. регрессия)

ГОРОД

```
OLS Regression Results
=====
Dep. Variable: y R-squared: 0.019
Model: OLS Adj. R-squared: 0.019
Method: Least Squares F-statistic: 232.4
Date: Tue, 19 Dec 2023 Prob (F-statistic): 5.68e-52
Time: 11:38:34 Log-Likelihood: -8425.6
No. Observations: 11775 AIC: 1.686e+04
Df Residuals: 11773 BIC: 1.687e+04
Df Model: 1
Covariance Type: nonrobust
=====

            coef  std err      t  P>|t|  [0.025  0.975]
-----
x1     -0.0327  0.002   -15.244  0.000   -0.037  -0.028
const   0.6358  0.011    58.293  0.000    0.614   0.657
=====
Omnibus: 42440.200 Durbin-Watson: 1.977
Prob(Omnibus): 0.000 Jarque-Bera (JB): 1814.972
Skew: 0.055 Prob(JB): 0.00
Kurtosis: 1.080 Cond. No. 12.6
=====

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
inverted.
```

СЕЛО

```
OLS Regression Results
=====
Dep. Variable: y R-squared: 0.019
Model: OLS Adj. R-squared: 0.019
Method: Least Squares F-statistic: 37.26
Date: Tue, 19 Dec 2023 Prob (F-statistic): 1.25e-09
Time: 11:38:34 Log-Likelihood: -1314.2
No. Observations: 1917 AIC: 2632.
Df Residuals: 1915 BIC: 2644.
Df Model: 1
Covariance Type: nonrobust
=====

            coef  std err      t  P>|t|  [0.025  0.975]
-----
x1     -0.0319  0.005   -6.104  0.000   -0.042  -0.022
const   0.5261  0.027    19.774  0.000    0.474   0.578
=====
Omnibus: 8955.990 Durbin-Watson: 2.067
Prob(Omnibus): 0.000 Jarque-Bera (JB): 302.349
Skew: 0.484 Prob(JB): 2.22e-66
Kurtosis: 1.312 Cond. No. 12.7
=====
```

ПОСЕЛОК

```
OLS Regression Results
=====
Dep. Variable: y R-squared: 0.012
Model: OLS Adj. R-squared: 0.011
Method: Least Squares F-statistic: 16.96
Date: Tue, 19 Dec 2023 Prob (F-statistic): 4.03e-05
Time: 11:38:34 Log-Likelihood: -978.29
No. Observations: 1392 AIC: 1961.
Df Residuals: 1390 BIC: 1971.
Df Model: 1
Covariance Type: nonrobust
=====

            coef  std err      t  P>|t|  [0.025  0.975]
-----
x1     -0.0258  0.006   -4.119  0.000   -0.038  -0.014
const   0.5290  0.032    16.530  0.000    0.466   0.592
=====
Omnibus: 5809.183 Durbin-Watson: 2.029
Prob(Omnibus): 0.000 Jarque-Bera (JB): 222.556
Skew: 0.363 Prob(JB): 4.70e-49
Kurtosis: 1.180 Cond. No. 12.9
=====

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
inverted.
```

МОЛКА

```
OLS Regression Results
=====
Dep. Variable: y R-squared: 0.012
Model: OLS Adj. R-squared: 0.012
Method: Least Squares F-statistic: 24.94
Date: Tue, 19 Dec 2023 Prob (F-statistic): 6.42e-07
Time: 11:38:34 Log-Likelihood: -1467.5
No. Observations: 2039 AIC: 2939.
Df Residuals: 2037 BIC: 2950.
Df Model: 1
Covariance Type: nonrobust
=====

            coef  std err      t  P>|t|  [0.025  0.975]
-----
x1     -0.0265  0.005   -4.994  0.000   -0.037  -0.016
const   0.6237  0.027    23.202  0.000    0.571   0.676
=====
Omnibus: 7509.686 Durbin-Watson: 2.043
Prob(Omnibus): 0.000 Jarque-Bera (JB): 323.737
Skew: -0.006 Prob(JB): 5.03e-71
Kurtosis: 1.048 Cond. No. 12.8
=====
```

Анализ по подвыборкам (лин. регрессия)

ПОСЕЛОК ГОРОДСКОГО ТИПА												
OLS Regression Results												
Dep. Variable:	y	R-squared:	0.012									
Model:	OLS	Adj. R-squared:	0.011									
Method:	Least Squares	F-statistic:	10.10									
Date:	Tue, 19 Dec 2023	Prob (F-statistic):	0.00153									
Time:	11:38:34	Log-Likelihood:	-607.05									
No. Observations:	853	AIC:	1218.									
Df Residuals:	851	BIC:	1228.									
Df Model:	1											
Covariance Type:	nonrobust											
<hr/>												
coef	std err	t	P> t	[0.025	0.975]							
x1	-0.0251	0.008	-3.178	0.002	-0.041	-0.010						
const	0.5540	0.041	13.593	0.000	0.474	0.634						
<hr/>												
Omnibus:	3522.368	Durbin-Watson:		1.924								
Prob(Omnibus):	0.000	Jarque-Bera (JB):		135.903								
Skew:	0.251	Prob(JB):		3.08e-30								
Kurtosis:	1.110	Cond. No.		12.8								
<hr/>												
Notes:												
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified												
<hr/>												
ГОРОД-МИЛЛИОННИК												
OLS Regression Results												
Dep. Variable:	y	R-squared:	0.026									
Model:	OLS	Adj. R-squared:	0.026									
Method:	Least Squares	F-statistic:	86.32									
Date:	Tue, 19 Dec 2023	Prob (F-statistic):	2.77e-20									
Time:	11:38:34	Log-Likelihood:	-2256.6									
No. Observations:	3175	AIC:	4517.									
Df Residuals:	3173	BIC:	4529.									
Df Model:	1											
Covariance Type:	nonrobust											
<hr/>												
coef	std err	t	P> t	[0.025	0.975]							
x1	-0.0380	0.004	-9.291	0.000	-0.046	-0.030						
const	0.6994	0.020	34.339	0.000	0.659	0.739						
<hr/>												
Omnibus:	12057.787	Durbin-Watson:		1.960								
Prob(Omnibus):	0.000	Jarque-Bera (JB):		474.118								
Skew:	-0.112	Prob(JB):		1.11e-103								
Kurtosis:	1.120	Cond. No.		12.0								
<hr/>												

ДЕРЕВНЯ												
OLS Regression Results												
Dep. Variable:	y	R-squared:	0.007									
Model:	OLS	Adj. R-squared:	0.006									
Method:	Least Squares	F-statistic:	5.707									
Date:	Tue, 19 Dec 2023	Prob (F-statistic):	0.0171									
Time:	11:38:34	Log-Likelihood:	-535.81									
No. Observations:	764	AIC:	1076.									
Df Residuals:	762	BIC:	1085.									
Df Model:	1											
Covariance Type:	nonrobust											
<hr/>												
coef	std err	t	P> t	[0.025	0.975]							
x1	-0.0203	0.009	-2.389	0.017	-0.037	-0.004						
const	0.4947	0.044	11.320	0.000	0.409	0.580						
<hr/>												
Omnibus:	3446.126	Durbin-Watson:		2.067								
Prob(Omnibus):	0.000	Jarque-Bera (JB):		124.654								
Skew:	0.406	Prob(JB):		8.54e-28								
Kurtosis:	1.196	Cond. No.		13.1								
<hr/>												
Notes:												
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified												
<hr/>												
САНКТ-ПЕТЕРБУРГ												
OLS Regression Results												
Dep. Variable:	y	R-squared:	0.009									
Model:	OLS	Adj. R-squared:	0.008									
Method:	Least Squares	F-statistic:	7.054									
Date:	Tue, 19 Dec 2023	Prob (F-statistic):	0.00807									
Time:	11:38:34	Log-Likelihood:	-561.77									
No. Observations:	786	AIC:	1128.									
Df Residuals:	784	BIC:	1137.									
Df Model:	1											
Covariance Type:	nonrobust											
<hr/>												
coef	std err	t	P> t	[0.025	0.975]							
x1	-0.0229	0.009	-2.656	0.008	-0.040	-0.006						
const	0.6616	0.043	15.362	0.000	0.577	0.746						
<hr/>												
Omnibus:	3225.086	Durbin-Watson:		1.871								
Prob(Omnibus):	0.000	Jarque-Bera (JB):		126.474								
Skew:	-0.228	Prob(JB):		3.44e-28								
Kurtosis:	1.088	Cond. No.		12.6								
<hr/>												

Анализ по подвыборкам с разделением на группы в зависимости от канала подачи заявки

```
только те, у кого channel_desc == "МБ"
OLS Regression Results
=====
Dep. Variable: y R-squared: 0.018
Model: OLS Adj. R-squared: 0.017
Method: Least Squares F-statistic: 454.8
Date: Tue, 19 Dec 2023 Prob (F-statistic): 4.76e-100
Time: 11:09:23 Log-Likelihood: -18279.
No. Observations: 25510 AIC: 3.656e+04
Df Residuals: 25508 BIC: 3.658e+04
Df Model: 1
Covariance Type: nonrobust
=====
            coef    std err      t   P>|t|    [0.025    0.975]
-----
x1       -0.0308   0.001   -21.327   0.000   -0.034   -0.028
const     0.6448   0.007    93.978   0.000     0.631     0.658
=====
Omnibus: 91133.321 Durbin-Watson: 2.023
Prob(Omnibus): 0.000 Jarque-Bera (JB): 3959.957
Skew: -0.059 Prob(JB): 0.00
Kurtosis: 1.073 Cond. No. 10.9
=====
```

```
только те, у кого channel_desc == "Web"
OLS Regression Results
=====
Dep. Variable: y R-squared: 0.010
Model: OLS Adj. R-squared: 0.010
Method: Least Squares F-statistic: 39.37
Date: Tue, 19 Dec 2023 Prob (F-statistic): 3.89e-10
Time: 11:09:20 Log-Likelihood: -2297.8
No. Observations: 3886 AIC: 4600.
Df Residuals: 3884 BIC: 4612.
Df Model: 1
Covariance Type: nonrobust
=====
            coef    std err      t   P>|t|    [0.025    0.975]
-----
x1       -0.0185   0.003   -6.274   0.000   -0.024   -0.013
const     0.3529   0.016    21.771   0.000     0.321     0.385
=====
Omnibus: 839.222 Durbin-Watson: 1.977
Prob(Omnibus): 0.000 Jarque-Bera (JB): 847.661
Skew: 1.072 Prob(JB): 8.57e-185
Kurtosis: 2.200 Cond. No. 13.0
=====
```

Регрессия по сэмплам

