

Дозвон, успешно

Бегемотики

Малышкин Алексей Андреевич
Мерзликин Матвей Александрович
Тестов Максим Евгеньевич
Мицкевич Елизавета Анатольевна
Паламарчук София Игоревна



Датасет

№3 Тинькофф: голосовые роботы и звонки

Количественные данные

возраст, доход, время ожидания,
длительность звонка, длительность
разговора, время ожидания ответа

Качественные характеристики

- пол, регион проживания, образование,
доход, возраст
- группа и вид звонка, результат звонка,
уточнение результата звонка, оператор,
точное время соединения

78 435 клиентов

33% женщин

41% мужчин

26% пол не
указан



Введение

Предвар. анализ

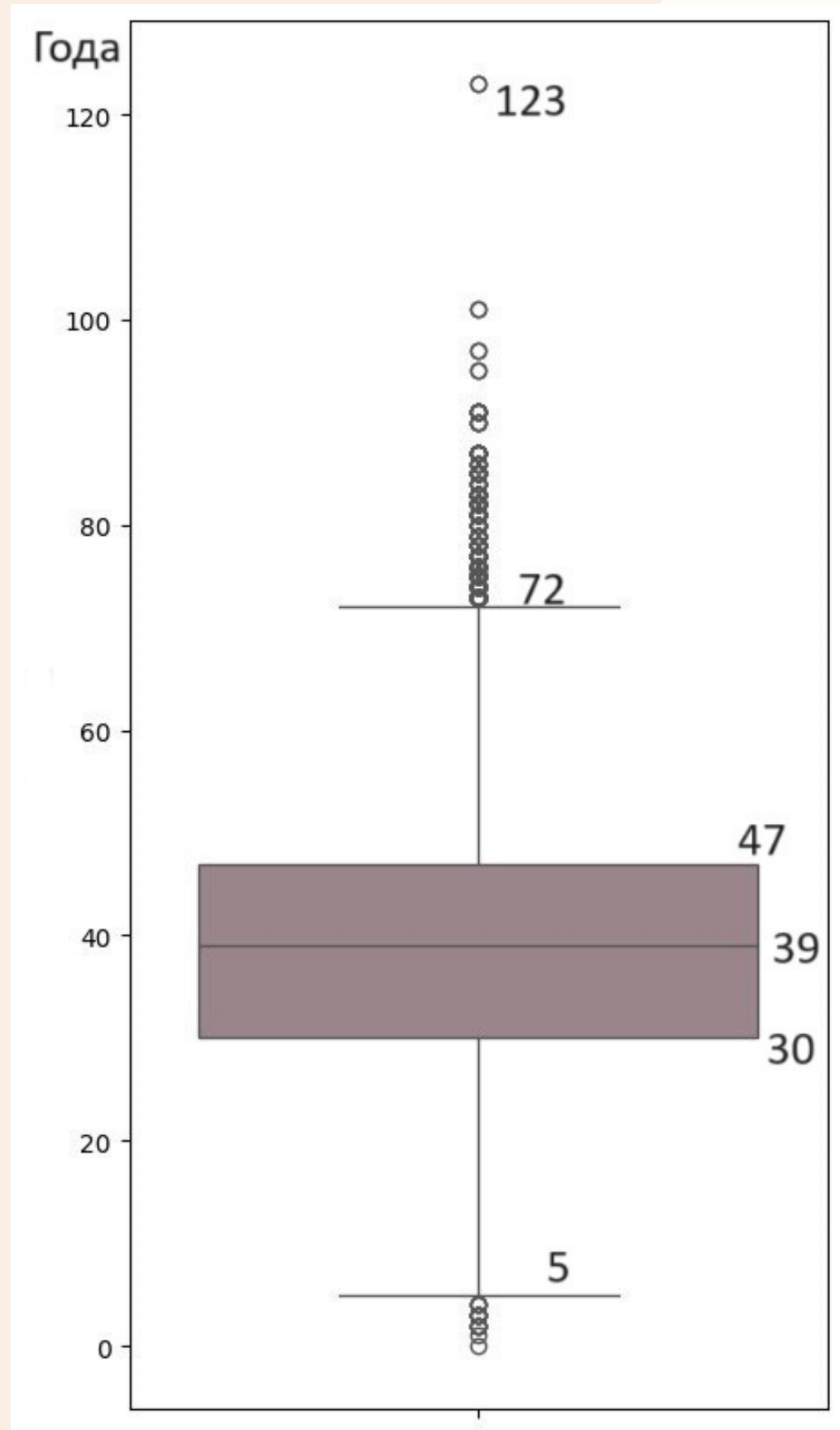
Гипотеза

Мат. модель

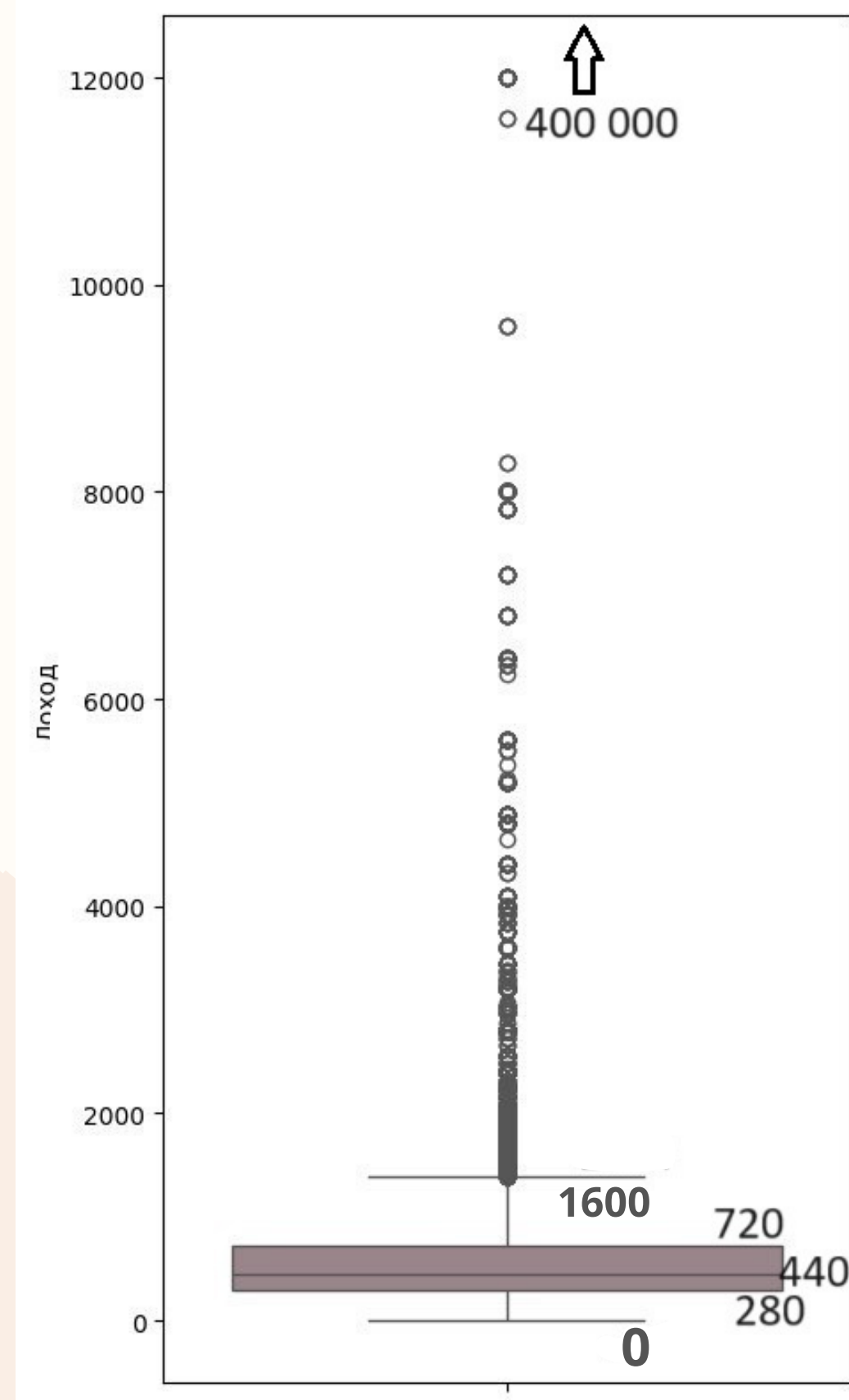
Регресс. модель

Итоги

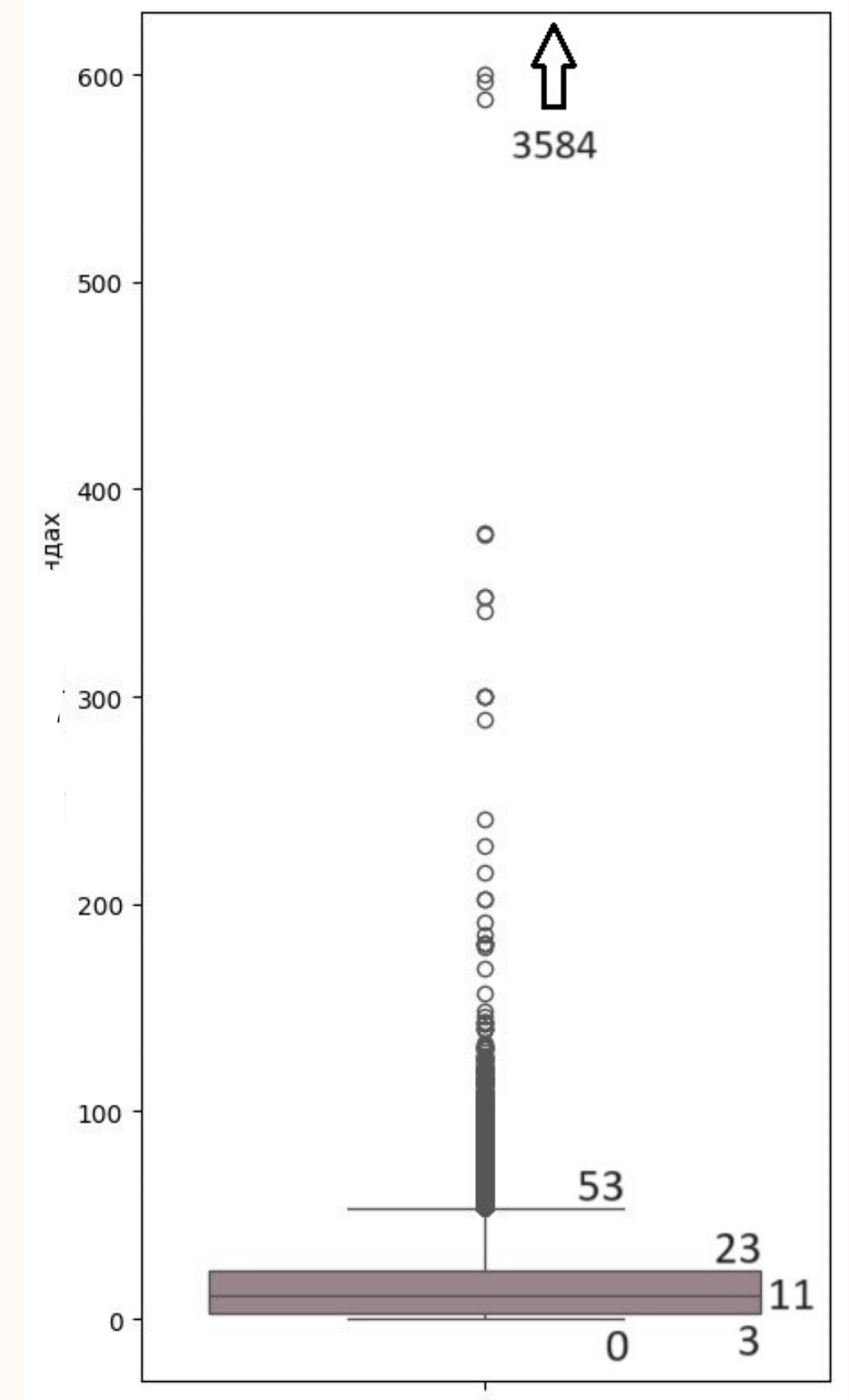
Обработка выбросов



Возраст



Доход



Время разговора в секундах



Подготовка данных

После очистки базы от выбросов осталось

Введение новых переменных/понятий

IS_SUCCESSFUL

переменная принимает значение "1" если дозвон успешный и значение "0" в ином случае

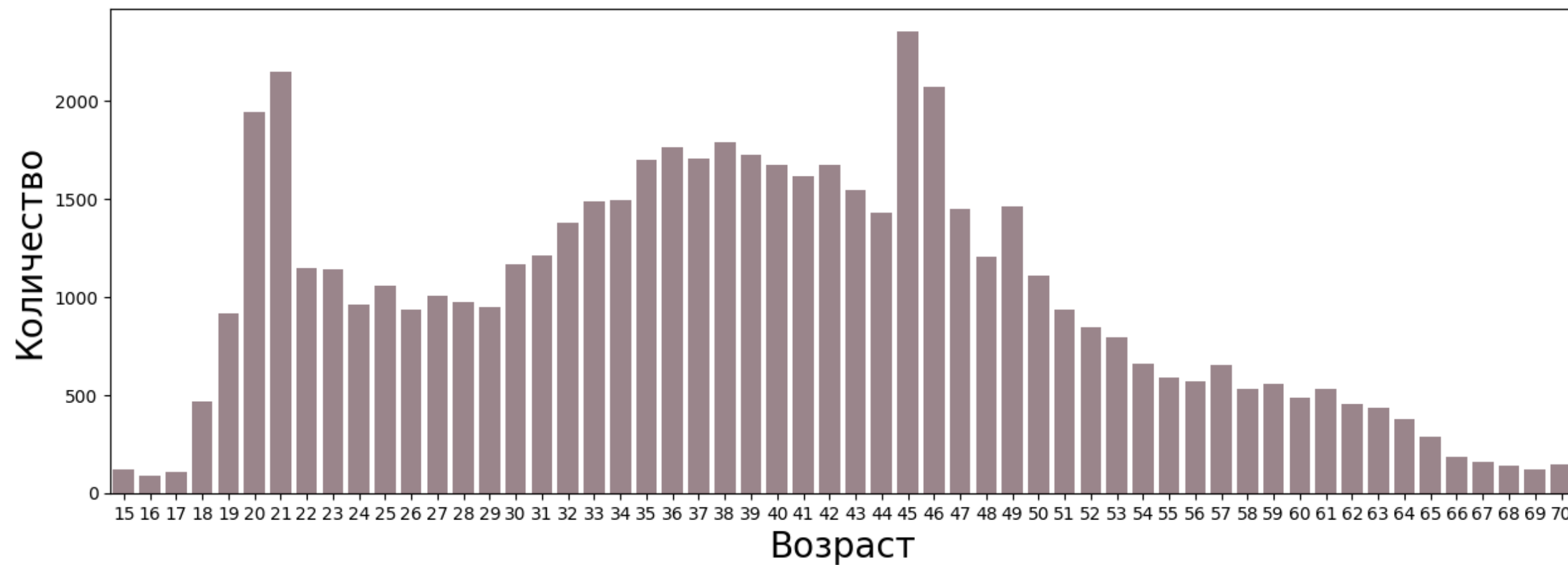
Конверсия

отношение количества успешных звонков к общему количеству звонков

- 57 100 клиентов
- 27211 мужчин
- 2110 женщин

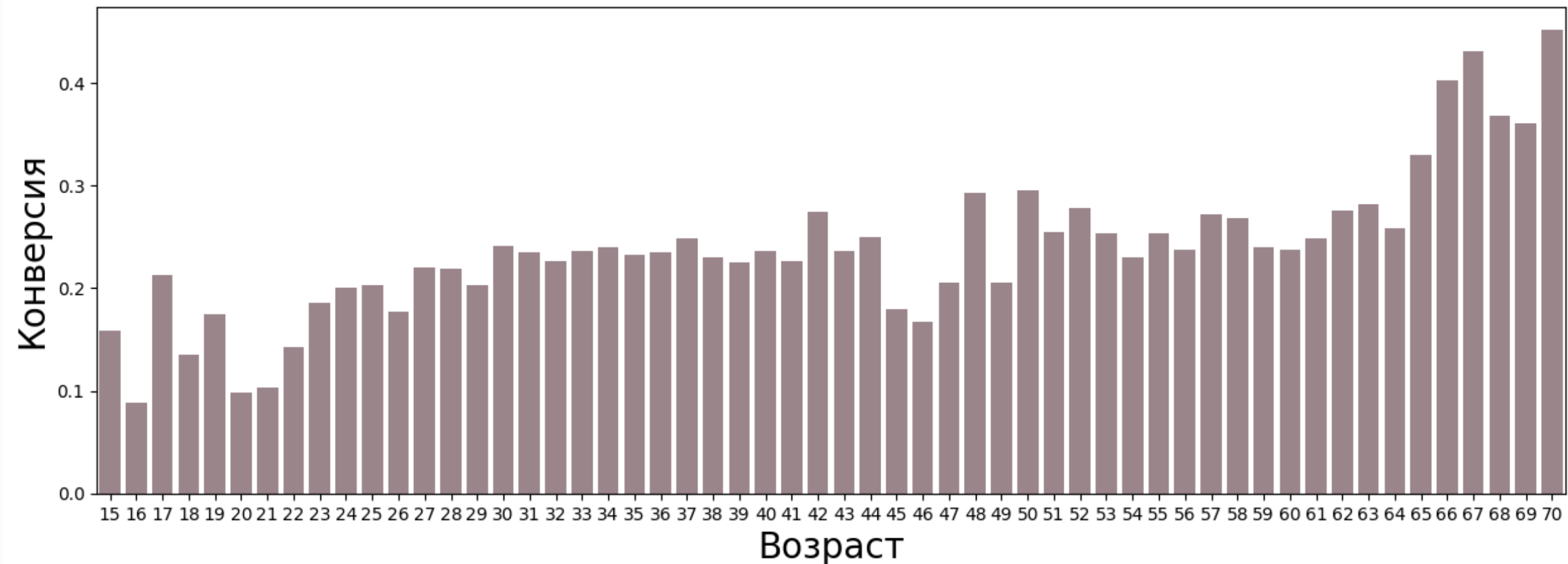


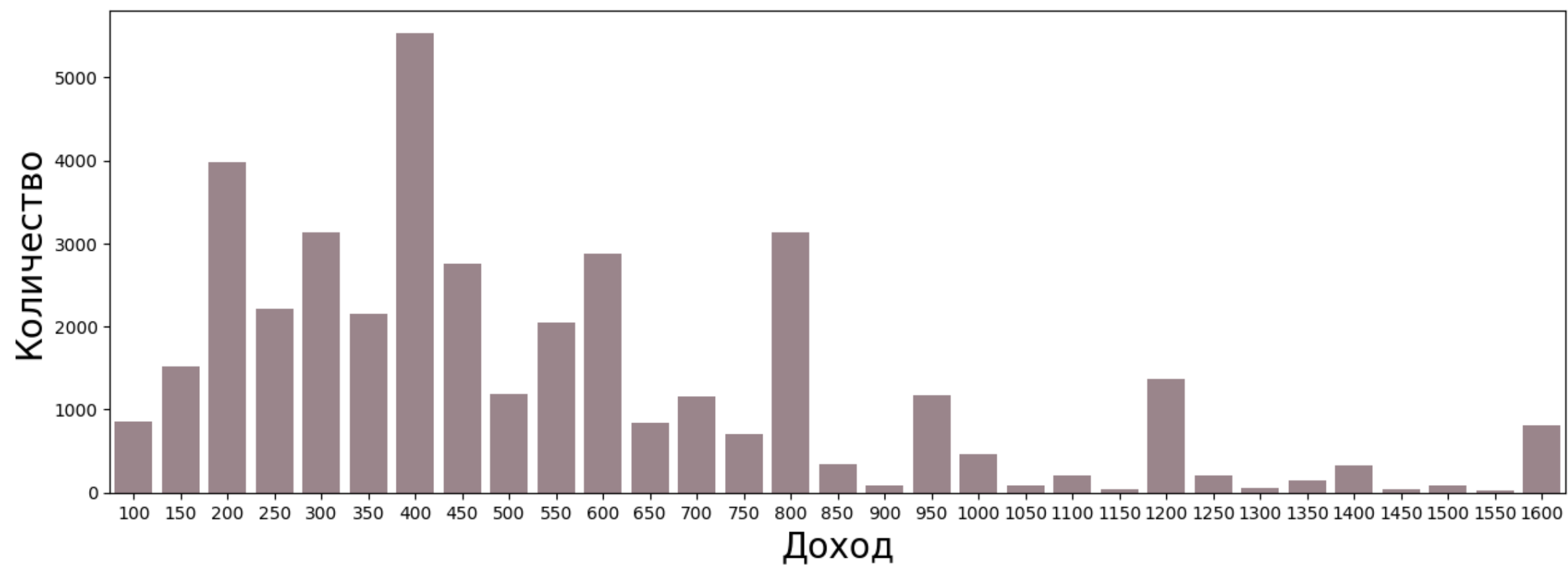
Графики распределения/конверсии



Большая часть клиентской базы – люди 28–42 лет

Корреляция **0.78**
Самая высокая конверсия – люди **66–70 лет**



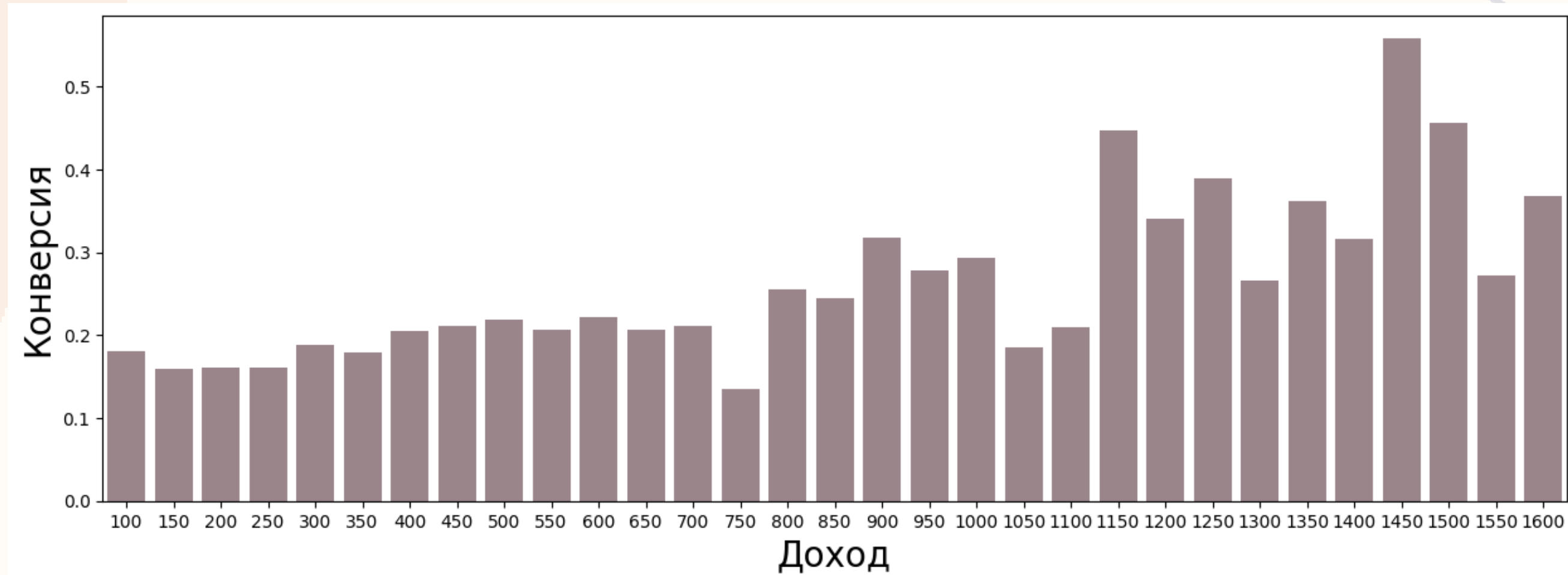


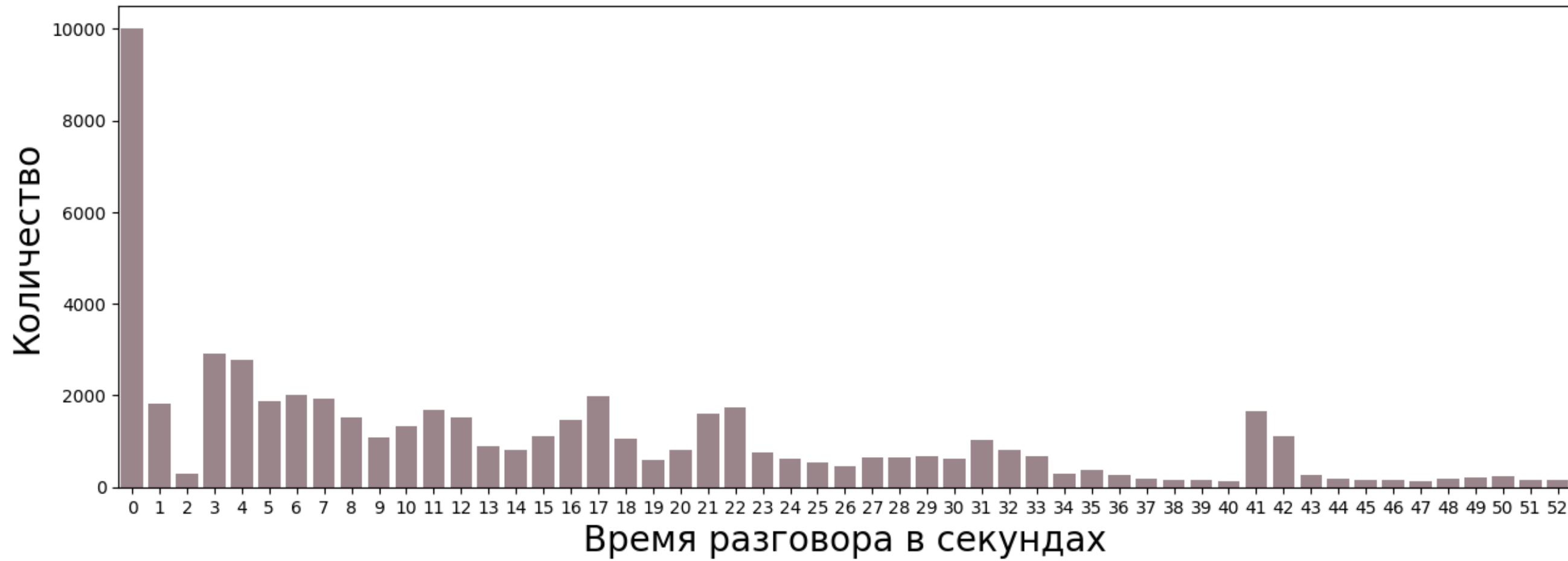
Мода 400 единиц

Медиана 850 единиц

Среднее значение 550 единиц

Корреляция 0.76

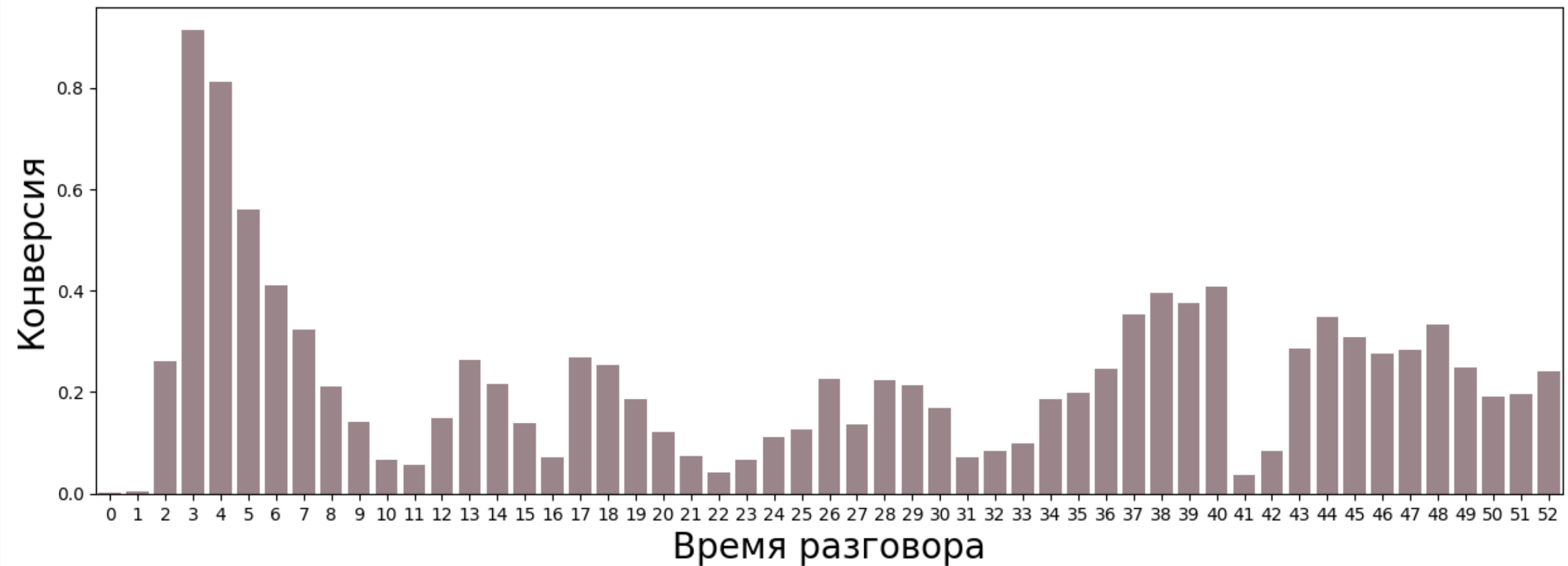




Мода 0 секунд

Медиана 27 секунд

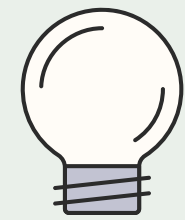
Наибольшая конверсия в момент времени 3 секунды





Исследовательский вопрос

Звонки какой возрастной группе людей реже всего оказываются успешными?

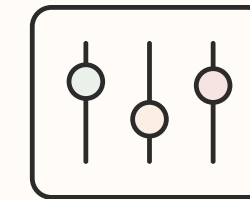


Гипотеза

Звонки представителям поколения Z реже всего оказываются успешными

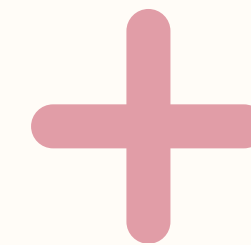


**поколение Z рождалось в 1997-2012гг



Поколение Z

Нет лишнего времени



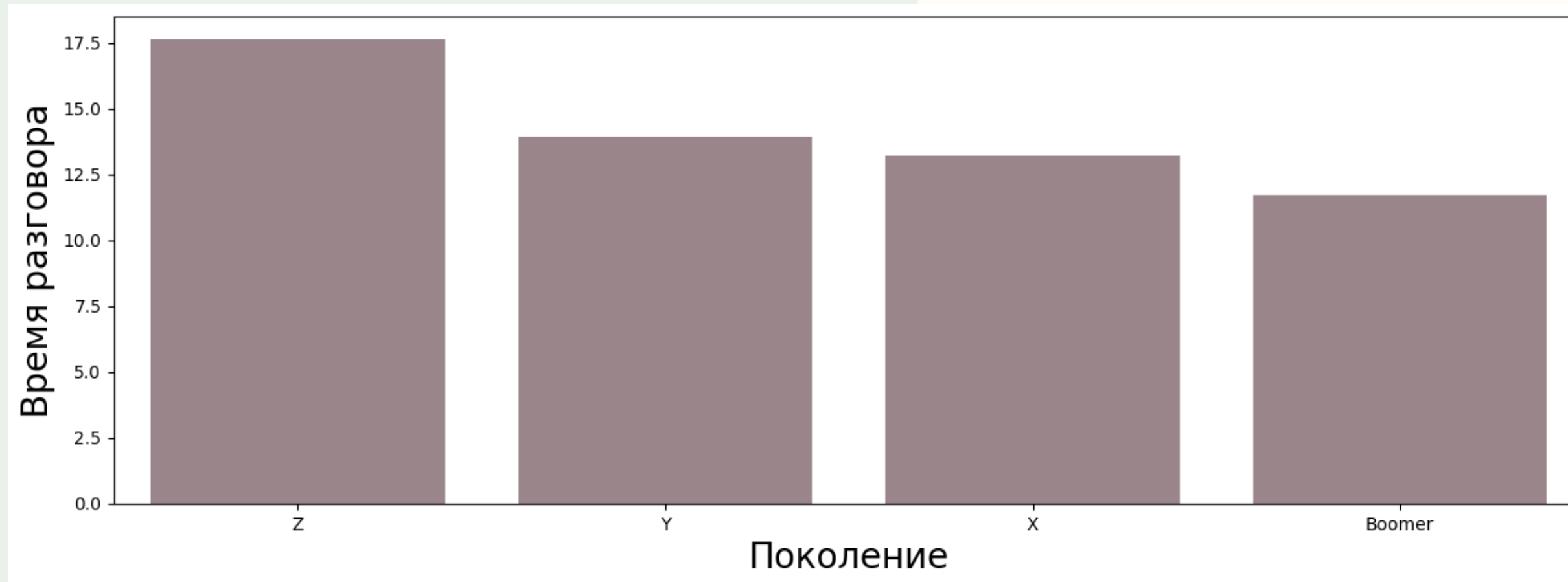
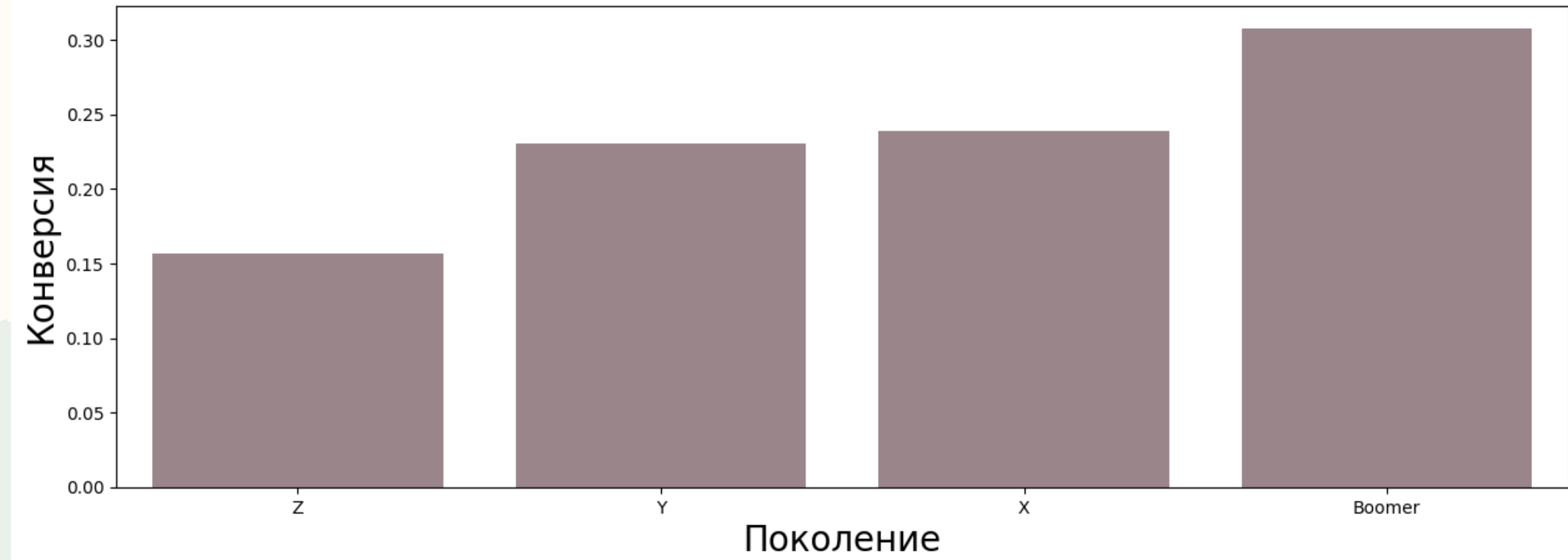
Наслышаны о телефонных мошенниках

“Телефобия”



НЕ примут телефонный звонок

Графики зависимости конверсии/времени разговора от поколения



NSUC

SUC

Z

NZ

9475

34989

1666

10969

Хи-квадрат Пирсона

Почему выбрали хи-квадрат Пирсона?

- 1) Сопоставляемые показатели измерены в номинальной шкале (0, 1)
- 2) Сопоставляемые группы независимы
- 3) Все значения в клетках матрицы больше 10

Мат.модель

Проверка гипотезы

Проверка устойчивости гипотезы

Регресс.модель



Математическое подтверждение

5%

Входной уровень значимости

H0

В генеральной совокупности **нет статистических различий** конверсии между поколением Z и остальными поколениями.

H1

В генеральной совокупности **есть статистические различия** конверсии между поколением Z и остальными поколениями.



Математическое подтверждение

5%

Входной уровень значимости

statistic 412.95

dof 1

p-value 8.3×10^{-92}

p-value < 0.05, нулевая гипотеза
отвергается, конверсия зависит от
принадлежности к поколению Z



Проверка на подвыборках

Мат.модель

Проверка гипотезы

Проверка устойчивости гипотезы

Регресс.модель

```
res = pd.DataFrame ( {"income": [], доход некой группы клиентов
```

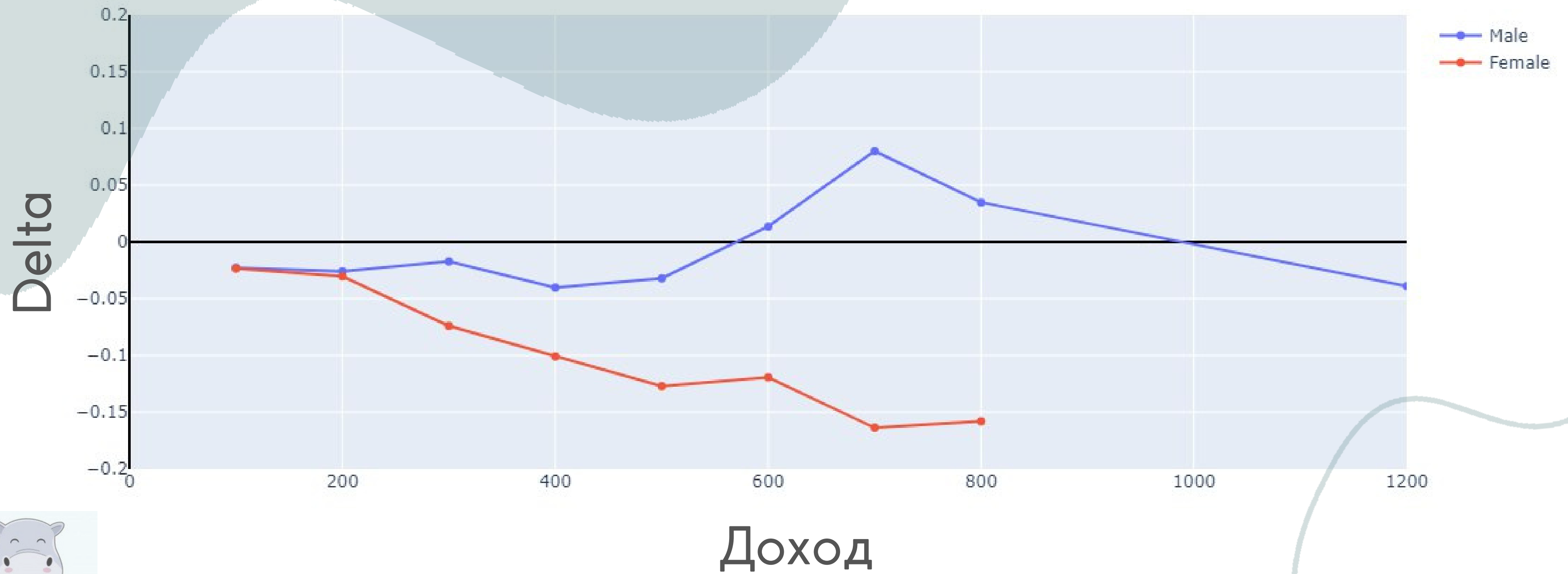
```
"gender" : [], пол некой группы клиентов
```

```
"delta" : [], разница конверсий старших поколений и поколения Z при заданных значениях пола и дохода
```

```
"pvalue": [] } ) результат применения хи-квадрата Пирсона из предыдущего слайда к части генеральной совокупности, ограниченной заданным полом и доходом
```



Delta в зависимости от дохода



“Неправильные зуммеры”

income	gender	delta	pvalue
600.0	M	0.013587	0.726998
700.0	M	0.080000	0.065219
800.0	M	0.034843	0.383979

H0

Между результатом звонка и принадлежностью к поколению Z статистическая взаимосвязь отсутствует*

H1

Связь между результатом звонка и принадлежностью к поколению Z статистически значима*



Промежуточные выводы

p-value > 0.05, значит нулевая гипотеза принимается.

Более высокая конверсия у поколения Z в некоторых случаях не имеет для нас статистической значимости

Звонки представителям поколения Z действительно менее успешны, чем звонки людям старших поколений



Построение линейной регрессии

$$\text{conversion} = k \times \text{age} + b$$

- вероятность совершить успешный звонок человеку заданного возраста

k - коэффициент

b - свободный член

age - возраст клиента



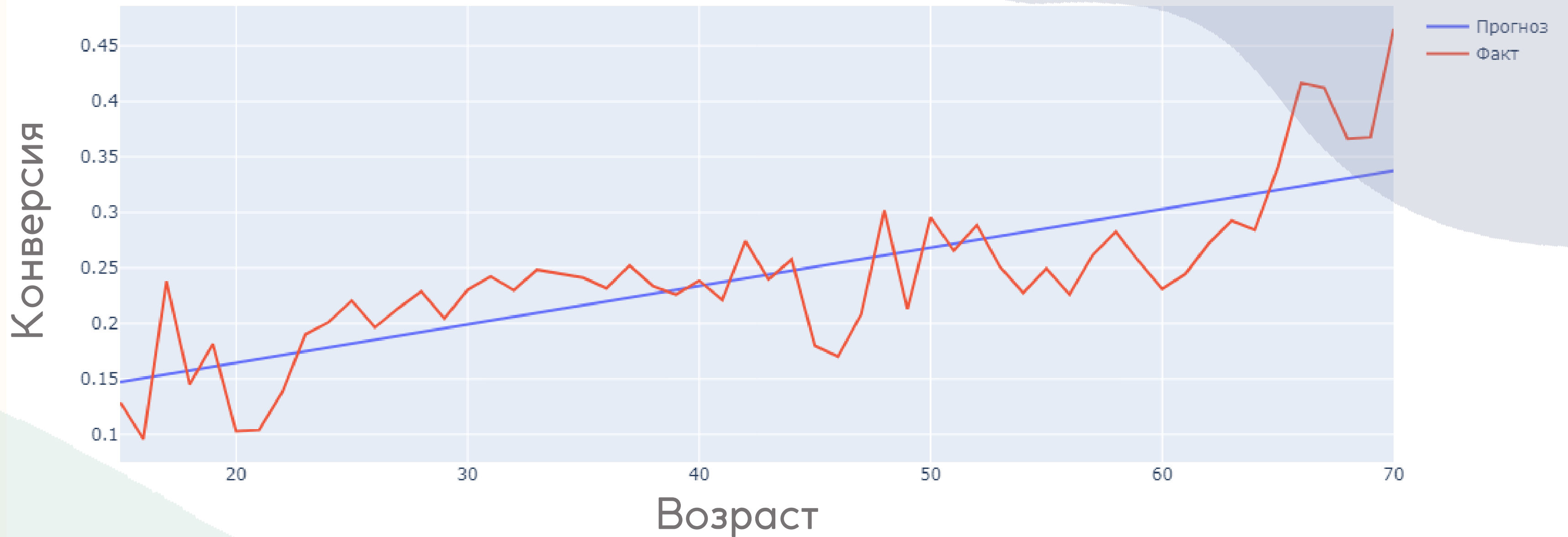
Мат.модель

Проверка
гипотезы

Проверка
устойчивости
гипотезы

Регресс.модель

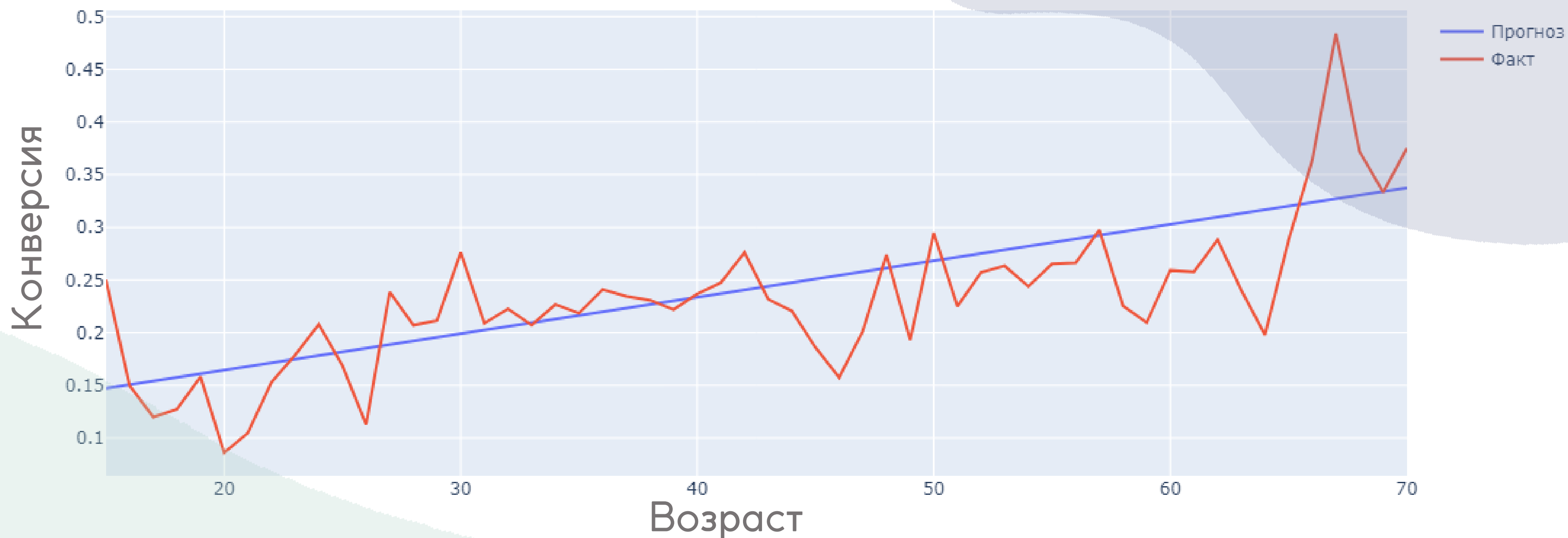
Обучающая выборка



Коэффициент детерминации (R^2) = 0.6



Тестовая выборка



Коэффициент детерминации (R^2) = 0.52



Полученные коэффициенты детерминации **выше 0.5**, значит модель считается **приемлемой** и может достаточно точно предсказывать вероятность успешности звонка клиенту с заданным значением **возраста**

Коэффициент	Значения	p-value
k	0.004	$1,9 \cdot 10^{-6}$
b	0.09	$8.1 \cdot 10^{-13}$

$$\text{conversion} = 0.004 * \text{age} + 0.09$$



Интерпретация



Есть статистически значимые различия конверсии у поколения Z и у людей старших поколений.



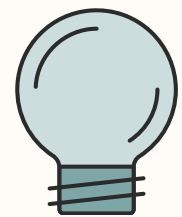
Несмотря на большую продолжительность звонков, звонки поколению Z реже оказываются успешными



Вывод

DA

NO



Гипотеза подтвердилась,
звонки представителям
поколения Z действительно
реже всего оказываются
успешными



Гипотеза не подтвердилась
для мужчин поколения Z с
доходом 600 - 900 единиц



Policy implication



Полученная при помощи линейной регрессии формула может помочь банку в принятии решения о форме общения с клиентом



Перспективы исследования

Углубить: собрать данные о доходах и расходах на один звонок

Смотреть в разрезе: регионов

Долгосрочная динамика: собирать данные в течение нескольких лет



Использованные источники

+1 Люди – Исследование предпочтительного вида общения

Исследование отпусков в России

Исследование компании Oracle

