

«Золотая жила»

Аналитика аудитории веб-платформы иностранного языка

Васюнина София

Вводная информация



Цель

Привлечение 1000
новых клиентов



Проблема

Убыточное привлечение
новых клиентов



Идея решения

Привлекать новых клиентов
таргетированно по полу и возрасту.
Определить категории, в которых расходы
на обслуживание клиента низкие, а доходы
от клиента высокие.

Какие данные имеем?

	Пол	Возраст	Категория возраста	Платежи в приложении	Себестоимость оказанных услуг	Прибыль
Клиенты						

При **первичной оценке** данных было 8715 строк с клиентами. Было отмечено, что в датасете присутствует отрицательные значения прибыли(убыточные клиенты). Их доля среди общих составила 8,2%. Наибольшие потери прибыли по первичным данным составили 9 643 950 рублей(аномально большая цифра при условии, что это 1 клиент). Стандартное отклонение составило 127100, что говорит о **сильной неоднородности** данных.

Первостепенная задача: очистка данных от выбросов



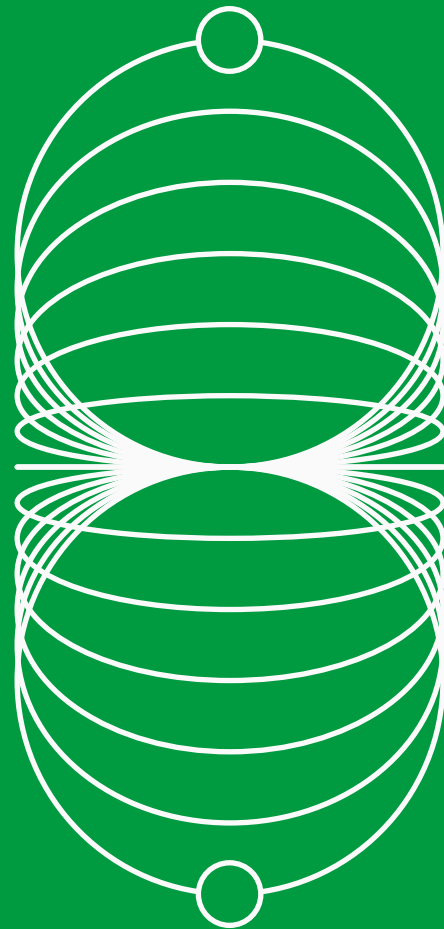
Была найдена аномальная строка с отрицательной прибылью = -9643950. Для более точной аналитики приняли эту строку за выброс и удалили.



Пропусков значений не было найдено. Строки в столбце с полом приведены к одинаковому виду.



Столбцы возраста заполнены корректными данными(от 18 до 49 лет). Проверили также, что прибыль действительно равна оплата минус затраты. Все строчки также заполнены корректно.



Текущая ситуация и базовые метрики

01 Количество клиентов
8714

02 Среднее значение прибыли
с одного клиента
1651

03 Общая прибыль
143.918.100

04 Количество клиентов, которых
планируется привлечь
1000

05 Количество клиентов, которых
планируется привлечь
до 15000

Анализ категории клиентов 18-24 лет



Мужчины 18-24 лет

- Количество: 3679
- Средняя прибыль: 13156
- Медианная прибыль: 2500
- Общая прибыль: 48400050



Женщины 18-24 лет

- Количество: 2712
- Средняя прибыль: 26278
- Медианная прибыль: 26100
- Общая прибыль: 71266950

Прибыль от **женщин** почти в **1,5 раза** больше прибыли от мужчин, поэтому лучше выбрать первую категорию. Однако стоит отметить, что медианное и среднее значение в категории мужчин довольно сильно отличаются. Это значит, что **есть категория мужчин, от которых прибыли было значительно больше, чем от большинства.**

Анализ половозрастных групп

Категория	Кол-во клиентов	Средняя прибыль	Медианная прибыль	Общая прибыль
Женщины 18-24 лет	2712	26278	26100	71266950
Мужчины 18-24 лет	3679	13156	2500	48400050
Женщины 25-39 лет	1212	7879	7900	9549950
Мужчины 25-39 лет	1020	14434	8200	14722400
Женщины 40+ лет	72	-344(убыток)	-500	-24750
Мужчины 40+ лет	19	184	0	3500

Анализ закономерностей

Возрастные тренды:

- 18-24 года: максимальная прибыльность
- 25-39 лет: средняя прибыль
- 40+ лет: низкая прибыльность

Самая прибыльная категория:

Женщины 18-24 лет, общая прибыль составила почти 71,3 млн рублей

Золотая жила

Перегруппировав данные на более узкие групповые категории, было выявлено, что самой прибыльной категорией являются **мужчины 20-22 лет**. Прибыль с одного клиента в среднем составила **79 тыс рублей**.

Почему “золотая жила” не была видна изначально?



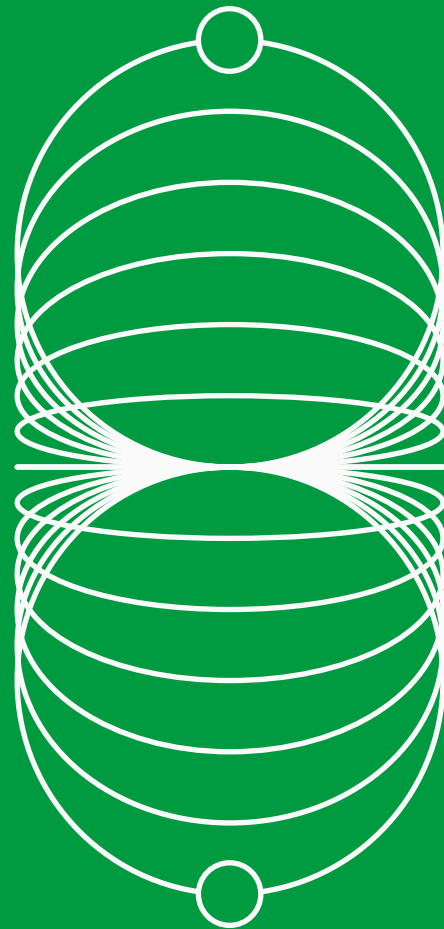
Исходная группировка по категориям была слишком “грубой”. В категории 18-24 было 6391 клиент.



Прибыльные клиенты составляют меньшинство, а значит их вклад в среднее значение прибыли был невелик на общем фоне клиентов.




Вывод: широкие возрастные группы могут скрывать прибыльные сегменты клиентов



Другие прибыльные категории

1. Мужчины 22-25 лет(53257 руб - средняя прибыль с клиента)
2. Женщины 20-22 лет(29705 руб - средняя прибыль с клиента)
3. Мужчины 30-34 лет(24959 руб - средняя прибыль с клиента)



**Спасибо за
внимание!
Готова ответить на
ваши вопросы**