

Введение в статистику

Проект 10. Обязательный

Золотая жила 2

Кейс, проблема, идея



Кейс

- Веб-платформа для изучения иностранного языка
- Клиенты потребляют разный контент и платят по разным тарифам
- Доходы от некоторых клиентов даже меньше, чем расходы платформы на обслуживание этих клиентов
- Планируется привлечь 1000 новых клиентов
- Средняя стоимость привлечения одного клиента — 15000 ₽
- При тестировании гипотез нужно использовать уровень значимости 5%, а при определении минимально детектируемого эффекта использовать мощность 80%



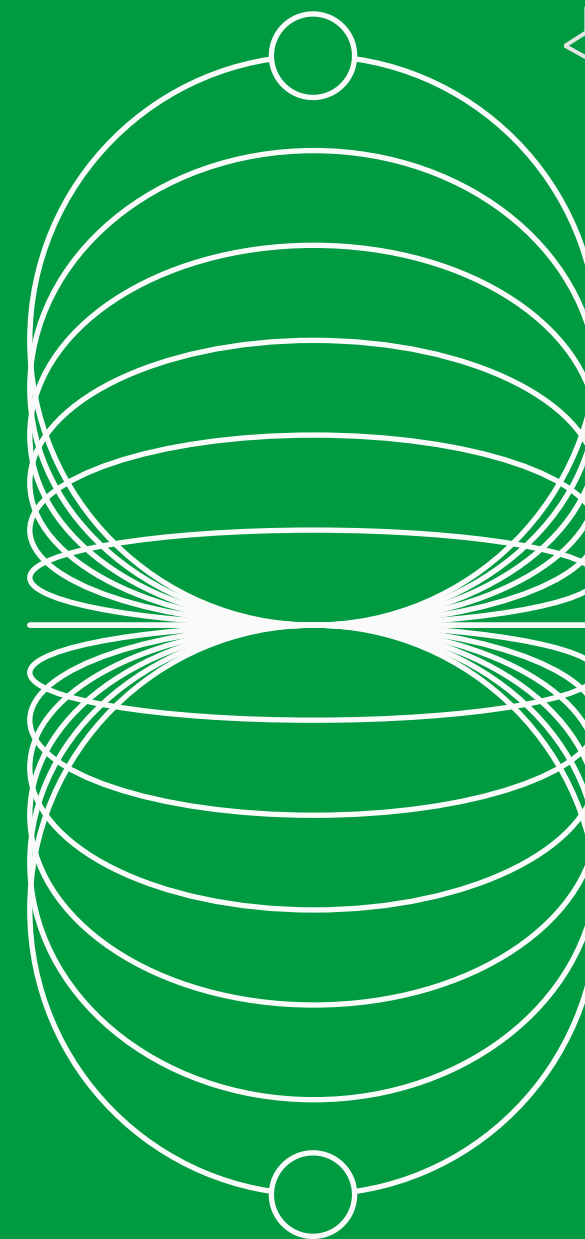
Проблема

Привлечение новых клиентов может повлечь убытки



Идея решения

Привлекать новых клиентов в заранее выбранной категории пола и возраста: возможно, удастся найти категорию, где клиенты платят много, а их обслуживание стоит мало



Роль, задача, результат



Роль

Аналитик отдела рекламы и маркетинга



Задача

- Таргетировать рекламу — найти категорию клиентов, в которой платформа получает максимум прибыли в расчёте на клиента
- Подготовить слайды с презентацией результата



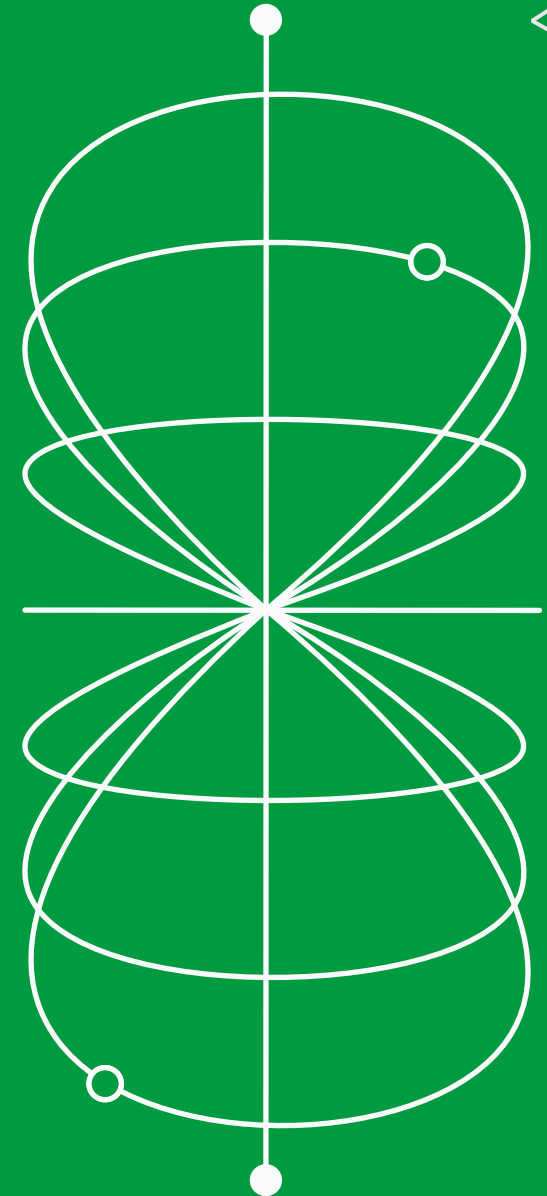
Доступные данные

Датасет [golden_mine.csv](#): прибыль по текущим клиентам, пол, возраст



Ожидаемый результат

- Презентация на слайдах в формате pdf
- Jupyter Notebook в Google Colab с расчётами



Требования к слайдам



Если слайды или Jupyter Notebook не приложен, решение кейса **оценивается в 0 баллов**

Понятность и внешний вид

- Внешний вид презентации не мешает воспринимать информацию
- Понятно, на какие вопросы отвечает каждый слайд
- Содержимое таблиц, графиков понятно из слайда без необходимости открывать исходный датасет
- Выводы явно сформулированы

МАКСИМУМ 5 БАЛЛОВ

Обоснованность

- Выводы основаны на таблицах, графиках, показателях, полученных из данных
- Таблицы и графики получены скриншотом или картинкой из Jupyter Notebook, поэтому их можно перепроверить
- Выводы явно сформулированы

МАКСИМУМ 5 БАЛЛОВ

Реакция заказчика

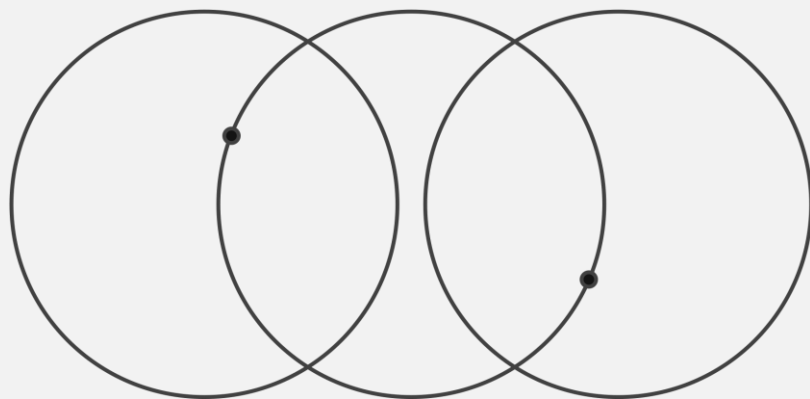
- **0 баллов**
Не принимает, ищет другого исполнителя
- **1 балл**
Частично принимает, считает необходимым отдать на доработку текущему исполнителю
- **2 балла**
Принимает, готов пересылать слайды от своего имени, под свою ответственность

МАКСИМУМ 2 БАЛЛА

Максимальная оценка — **10 баллов**

Бонусные баллы могут поднять основную оценку, но не выше **10 баллов**, даже если в сумме получится 11 или 12 баллов

Подход к решению



01

Показать на слайде текущую ситуацию:

- Сколько клиентов
- Сколько новых клиентов планируется привлечь, и сколько это стоит в пересчете на клиента
- Какова прибыль от клиента
- Результаты теста статистической значимости прибыльности привлечения клиентов без таргетирования
- Вывод о привлечении клиентов без таргетирования пола или возраста, учитывая статистическую и экономическую значимость



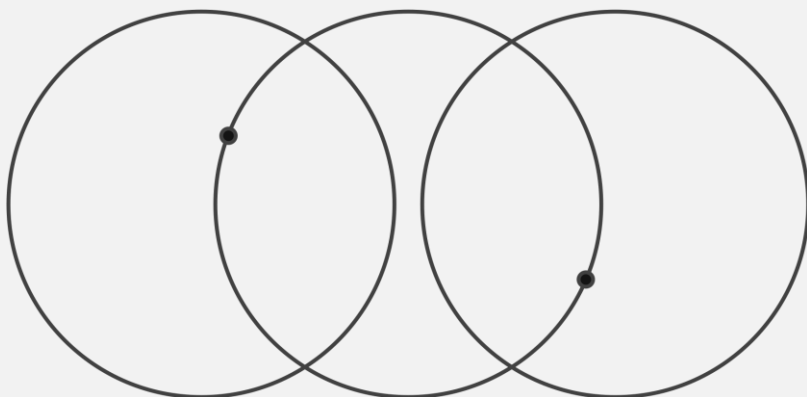
02

Показать на слайде, из какой половозрастной категории стоит привлекать клиентов

Для двух категорий клиентов: «Мужчины 18-24 лет» и «Женщины 18-24 лет»:

- Проверить статистическую значимость разницы между ними в прибыльности
- Прокомментировать экономическую значимость разницы между ними
- Провести тесты значимости прибыльности привлечения клиентов в каждой из категорий
- Сделать вывод: выбрать одну категорию

Подход к решению



01

Подготовить слайды синего уровня

Для двух подкатегорий клиентов: «Женщины 18-21 лет» и «Женщины 22-24 лет»:

- Проверить статистическую значимость разницы между ними в прибыльности
- Для каждой подкатегории проверить, что прибыль статистически значимо выше расходов на привлечение
- Объяснить выбор уровня значимости в тестах

02

Прокомментировать идею сравнивать подкатегории «Женщины 18 лет», «Женщины 19 лет», «Женщины 20 лет», ... «Женщины 24 лет», попарно тестируя значимость разницы в прибыльности

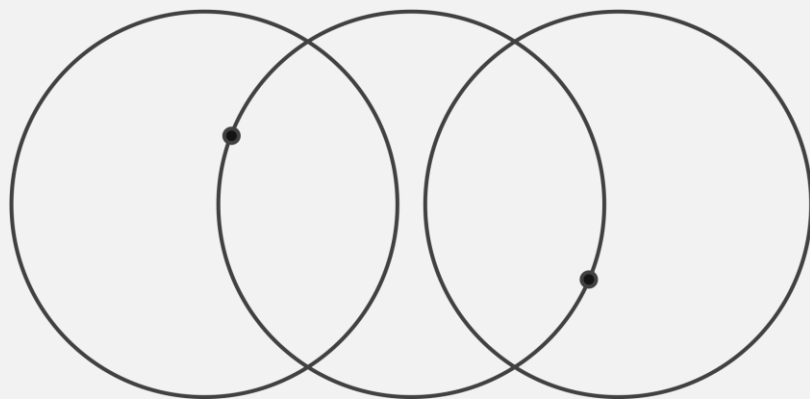
- Как выбрать уровень значимости?
- Что случится с минимальным детектируемым эффектом? Качественного ответа достаточно

03

Прокомментировать идею сравнивать подкатегории «Мужчины 18 лет», «Мужчины 19 лет», «Мужчины 20 лет», ... «Мужчины 24 лет», попарно тестируя значимость разницы в прибыльности

Чёрный уровень

Подход к решению



01

Подготовить слайды синего и красного уровня



02

Проверить значимость разницы в прибыльности категорий «Золотая жила» (найдена в проекте 1), и категории «Женщины 18-24 лет»

03

Определить минимальный детектируемый эффект в тесте, проверяющем, что прибыль в категории «Мужчины 18 лет» выше стоимости привлечения

04

Определить минимальный детектируемый эффект в серии тестов, проверяющих, что прибыль хотя бы в одной из категорий в категории «Мужчины 18 лет», «Мужчины 19 лет», ... «Мужчины 24 лет» выше стоимости привлечения. Не забывать про корректировку уровня значимости

Эффект называется детектируемым, если его присутствие ровно в одной категории обнаруживается соответствующим тестом в серии с требуемой вероятностью 80%

Что и когда нужно сдать



Что сдавать?

- Презентация в слайдах в формате pdf
- Jupyter Notebook в Google Colab с расчётами



Когда сдавать?

Сроки сдачи указаны в информационной системе

Сдача кейса — **необходимое** условие прохождения курса

