# **Проектная работа**

## **Что такое проект**

Поздравляем! Вы прошли курс в тренажёре. Самое время проверить полученные знания на практике и решить аналитический кейс. Выполнять работу вы будете самостоятельно.

Как закончите работу над проектом, отправьте его на проверку ревьюеру. В течение суток вы получите комментарии. Их нужно учесть: доработать проект и вернуть ревьюеру обновлённый вариант.

Скорее всего, вы снова получите комментарии по кейсу. Это нормально — доработка может проходить в несколько этапов.

Проект завершён, когда засчитаны все исправления.

## **Описание проекта**

Вы — маркетинговый аналитик развлекательного приложения Procrastinate Pro+. Несмотря на огромные вложения в рекламу, последние несколько месяцев компания терпит убытки. Ваша задача — разобраться в причинах и помочь компании выйти в плюс.

Есть данные о пользователях, привлечённых с 1 мая по 27 октября 2019 года:

* лог сервера с данными об их посещениях,
* выгрузка их покупок за этот период,
* рекламные расходы.

Вам предстоит изучить:

* откуда приходят пользователи и какими устройствами они пользуются,
* сколько стоит привлечение пользователей из различных рекламных каналов;
* сколько денег приносит каждый клиент,
* когда расходы на привлечение клиента окупаются,
* какие факторы мешают привлечению клиентов.

## **Описание данных**

В вашем распоряжении три датасета. Файл visits\_info\_short.csv хранит лог сервера с информацией о посещениях сайта, orders\_info\_short.csv — информацию о заказах, а costs\_info\_short.csv — информацию о расходах на рекламу.

Структура visits\_info\_short.csv:

* User Id — уникальный идентификатор пользователя,
* Region — страна пользователя,
* Device — тип устройства пользователя,
* Channel — идентификатор источника перехода,
* Session Start — дата и время начала сессии,
* Session End — дата и время окончания сессии.

Структура orders\_info\_short.csv:

* User Id — уникальный идентификатор пользователя,
* Event Dt — дата и время покупки,
* Revenue — сумма заказа.

Структура costs\_info\_short.csv:

* dt — дата проведения рекламной кампании,
* Channel — идентификатор рекламного источника,
* costs — расходы на эту кампанию.

## **Инструкция по выполнению проекта**

Рекомендуем ещё раз просмотреть урок «Разбор кейса» из темы «Юнит-экономика» перед началом работы.

### **Шаг 1. Загрузите данные и подготовьте их к анализу**

Загрузите данные о визитах, заказах и рекламных расходах из CSV-файлов в переменные.

**Пути к файлам**

* /datasets/visits\_info\_short.csv. [Скачать датасет](https://code.s3.yandex.net/datasets/visits_info_short.csv);
* /datasets/orders\_info\_short.csv. [Скачать датасет](https://code.s3.yandex.net/datasets/orders_info_short.csv);
* /datasets/costs\_info\_short.csv. [Скачать датасет](https://code.s3.yandex.net/datasets/costs_info_short.csv).

Изучите данные и выполните предобработку. Есть ли в данных пропуски и дубликаты? Убедитесь, что типы данных во всех колонках соответствуют сохранённым в них значениям. Обратите внимание на столбцы с датой и временем.

### **Шаг 2. Задайте функции для расчёта и анализа LTV, ROI, удержания и конверсии.**

Разрешается использовать функции, с которыми вы познакомились в теоретических уроках.

Это функции для вычисления значений метрик:

* get\_profiles() — для создания профилей пользователей,
* get\_retention() — для подсчёта Retention Rate,
* get\_conversion() — для подсчёта конверсии,
* get\_ltv() — для подсчёта LTV.

А также функции для построения графиков:

* filter\_data() — для сглаживания данных,
* plot\_retention() — для построения графика Retention Rate,
* plot\_conversion() — для построения графика конверсии,
* plot\_ltv\_roi — для визуализации LTV и ROI.

### **Шаг 3. Исследовательский анализ данных**

* Составьте профили пользователей. Определите минимальную и максимальную даты привлечения пользователей.
* Выясните, из каких стран пользователи приходят в приложение и на какую страну приходится больше всего платящих пользователей. Постройте таблицу, отражающую количество пользователей и долю платящих из каждой страны.
* Узнайте, какими устройствами пользуются клиенты и какие устройства предпочитают платящие пользователи. Постройте таблицу, отражающую количество пользователей и долю платящих для каждого устройства.
* Изучите рекламные источники привлечения и определите каналы, из которых пришло больше всего платящих пользователей. Постройте таблицу, отражающую количество пользователей и долю платящих для каждого канала привлечения.

После каждого пункта сформулируйте выводы.

### **Шаг 4. Маркетинг**

* Посчитайте общую сумму расходов на маркетинг.
* Выясните, как траты распределены по рекламным источникам, то есть сколько денег потратили на каждый источник.
* Постройте визуализацию динамики изменения расходов во времени (по неделям и месяцам) по каждому источнику. Постарайтесь отразить это на одном графике.
* Узнайте, сколько в среднем стоило привлечение одного пользователя (CAC) из каждого источника. Используйте профили пользователей.

Напишите промежуточные выводы.

### **Шаг 5. Оцените окупаемость рекламы**

Используя графики LTV, ROI и CAC, проанализируйте окупаемость рекламы. Считайте, что на календаре 1 ноября 2019 года, а в бизнес-плане заложено, что пользователи должны окупаться не позднее чем через две недели после привлечения. Необходимость включения в анализ органических пользователей определите самостоятельно.

* Проанализируйте окупаемость рекламы c помощью графиков LTV и ROI, а также графики динамики LTV, CAC и ROI.
* Проверьте конверсию пользователей и динамику её изменения. То же самое сделайте с удержанием пользователей. Постройте и изучите графики конверсии и удержания.
* Проанализируйте окупаемость рекламы с разбивкой по устройствам. Постройте графики LTV и ROI, а также графики динамики LTV, CAC и ROI.
* Проанализируйте окупаемость рекламы с разбивкой по странам. Постройте графики LTV и ROI, а также графики динамики LTV, CAC и ROI.
* Проанализируйте окупаемость рекламы с разбивкой по рекламным каналам. Постройте графики LTV и ROI, а также графики динамики LTV, CAC и ROI.

Ответьте на такие вопросы:

* + Окупается ли реклама, направленная на привлечение пользователей в целом?
  + Какие устройства, страны и рекламные каналы могут оказывать негативное влияние на окупаемость рекламы?
  + Чем могут быть вызваны проблемы окупаемости?

Напишите вывод, опишите возможные причины обнаруженных проблем и промежуточные рекомендации для рекламного отдела.

### **Шаг 6. Напишите выводы**

* Выделите причины неэффективности привлечения пользователей.
* Сформулируйте рекомендации для отдела маркетинга.

## **Оформление**

Задание выполните в Jupyter Notebook. Код пишите в ячейках типа code, текстовые пояснения — в ячейках типа markdown. Применяйте форматирование и заголовки.

## **Критерии проверки**

Ревьюеры ожидают, что в этом проекте вы:

* подготовили данные к анализу,
* выбрали подходящие типы визуализации,
* верно интерпретировали графики,
* правильно рассчитали и проанализировали метрики,
* аргументировали метриками рекомендации маркетологам,
* сформировали логичную структуру проекта,
* написали «чистый» и эффективный код,
* оставили комментарии,
* сделали корректные выводы.