

# Открытый хакатон School of Analytics 2023

## Введение

AdTech агентство «TechDesign», предоставляющее решения в области рекламных технологий, разработало новую модель расчета ставки (т.е. цены за переход) по объявлениям для рекламной кампании. В качестве альтернативы существующей модели ставок «Максимальная Ставка» компания представила новое рекламное решение – модель ставок «Средняя Ставка».

После внедрения новой опции компания «eStore», являющаяся одним из клиентов агентства и проводящая ретаргетинговую рекламную кампанию с использованием модели «Максимальная Ставка», решила протестировать альтернативную модель расчета.

## Условия задачи

Для того, чтобы провести сравнение между моделями расчета «Средняя Ставка» и «Максимальная Ставка», в течение следующего месяца компания «eStore» проводила A/B тестирование. Как аналитику данных «eStore» Вам предстоит проанализировать результаты проведения A/B тестирования и представить аналитический отчет менеджменту компании.

Вы получили данные с результатами проведения A/B тестирования и хотите убедиться, что эксперимент проводился правильно и полученный результат оказался статистически достоверным.

В рамках выполнения данной задачи Вы определили для себя следующие шаги:

**Задание 1.** Проверка разделения наблюдений на экспериментальную и контрольную группы на основе данных, полученных после проведения эксперимента

**Задание 2.** Подсчет метрик на основе данных, полученных после проведения эксперимента. Проверка нормальности распределения

**Задание 3.** Формулирование гипотезы. Подбор статистического критерия для тестирования гипотезы

**Задание 4.** Определение параметров для A/B групп. Проведение A/A теста для оценки параметров. Проверка репрезентативности и однородности выборок

**Задание 5.** Подсчет ключевых метрик. Оценка статистической значимости полученных результатов. Оценка ошибки первого и второго рода

## Данные

Данные, собранные в течение 30 дней с момента начала проведения эксперимента, содержатся в файле «**test\_group.csv**». Файл «**control\_group.csv**» содержит данные пользователей, включенных в контрольную группу.

Описание и тип данных переменных приведен в таблице ниже:

Название	Тип данных	Описание
Campaign Name	object	# разделение на экспериментальную и контрольную группы
Date	object	# дата проведения рекламной кампании
Total Advertising Spend (USD)	int64	# затраты на рекламу
Total Impressions Count	int64	# просмотр пользователем рекламного объявления
Total Reach Count	int64	# уникальные пользователи, просмотревшие рекламное объявление
Total Clicks on Website	int64	# пользователи, перешедшие на сайт по ссылке в рекламном объявлении
Total Searches Initiated	int64	# выполнение пользователем поиска на сайте
Total View Content Actions	int64	# просмотр пользователем деталей товара
Total Add to Cart Actions	int64	# добавление пользователем товара в корзину
Total Purchases Count	int64	# приобретение пользователем товара

## Инструкция по выполнению

Ваш итоговый отчет должен содержать следующие части:

### 1. Дизайн эксперимента

Перед проведением эксперимента аналитики компании разработали дизайн эксперимента, который включал в себя детальное описание следующих компонентов:

- Цель эксперимента
- Нулевая и альтернативная гипотезы
- Определение изменения и целевые метрики
- Экспериментальная и контрольная группы
- Уровень статистической значимости
- Параметры эксперимента
  - Размер выборки
  - Факторы воздействия
  - Длительность эксперимента

Поскольку файл с описанием дизайна эксперимента был утерян, Вам необходимо восстановить приведенные выше компоненты на основе вводной информации по условиям задачи и полученных данных.

### 2. Проведение и анализ эксперимента

Аналитики компании дополнительно составили детальное описание этапов проведения эксперимента, используемых метрик и методологии оценки результатов.

К сожалению, Вы не смогли получить доступ к файлу с деталями проведения A/B тестирования, что затрудняет поставленную перед Вами задачу.

Чтобы убедиться в правильности проведенного эксперимента и достоверности полученных результатов, Вам необходимо составить детальное описание процесса выполнения каждого из пяти шагов, которые Вы обозначили для себя ранее.

### 3. Выводы и рекомендации

На основе полученных выводов Вам необходимо сформулировать рекомендации для компании «eStore» с целью выбора оптимальной модели расчета ставки по рекламным объявлениям. Необходимо объяснить, как предложенные Вами рекомендации соотносятся с результатами A/B тестирования.

## Формат сдачи

Каждая команда должна назначить одного участника, ответственного за коммуникацию в рамках хакатона и загрузку итогового решения.

Выполненные задания необходимо загрузить по ссылке в Google-форму: <https://forms.gle/PKZvvQyuF8xqAWff8>

В поле «Участники команды» необходимо через запятую указать Фамилию Имя всех участников команды.

В поле «Загрузка решения» необходимо прикрепить 2 файла:

- Решение заданий в формате .ipynb
- Итоговый отчет в формате .pdf

Просьба дополнительно направить решение на почту: [hackathon@sanalytics.education](mailto:hackathon@sanalytics.education)

Ответы на задания хакатона принимаются **до 23:55 12 сентября**.

## Оценивание

Полученные от участников решения заданий оцениваются по 10-балльной системе. Решение заданий в формате кода на Python и аналитический отчет имеют одинаковый вес в итоговой оценке.

Для участников, направивших решение заданий раньше установленного дедлайна, предусмотрена система дополнительных баллов:

Ответ отправлен <b>до 23:55 2 сентября</b>	<b>+ 2 балла</b> к итоговой оценке
Ответ отправлен <b>до 23:55 8 сентября</b>	<b>+ 1 балл</b> к итоговой оценке
Ответ отправлен <b>до 23:55 12 сентября</b>	<b>+ 0 баллов</b> к итоговой оценке

Обратите внимание, что Ваше решение должно содержать минимальный процент правильно выполненных заданий для того, чтобы иметь возможность получить дополнительные баллы за раннюю сдачу.

Решения, направленные раньше установленного дедлайна, но не прошедшие порог оценивания в 40% (без учета дополнительных баллов), не будут допущены до финального этапа.