

1. Общая информация

Мобильная игра зарабатывает на показе рекламы. Есть несколько форматов: **int** – полноэкранная реклама; **banner** – горизонтальный баннер, висит внизу экрана пока юзер играет; **rv** – видео, которое юзер смотрит по желанию и за просмотр получает монетки или др. ресурсы. Информация о каждом показе отправляется в аналитику и содержит доход от показа, формат показанной рекламы, рекламную сеть от которой был показ и др.

eCPM – доход за 1000 показов рекламы.

2. Бизнес задача

Провести исследование показов рекламы с целью понять на сколько отличается eCPM в зависимости от географии (города) юзера, версии ОС и других параметров. Интересны любые инсайты и находки по eCPM-ам.

Этот csv файл содержит информацию о показах рекламы для юзеров США:
https://drive.google.com/drive/folders/1bjWlzyXV8ALcoxZYrdNQH6_MOm3nS3PI?usp=sharing

3. Что требуется

- Для каждого формата рекламы:
 - построить распределение средних eCPM-ов юзеров
 - построить распределение eCPM-ов по городам, версиям ОС
 - построить распределение доходов по городам, версиям ОС
- Учесть статистическую значимость (отобразить, отфильтровать незначимое или хотя бы прокомментировать). Объяснить решение.
- Описать наблюдения и вывод по исследованию.

4. Дополнительные (по желанию)

Исследовать:

- Какой процент доходов содержит информацию о городе юзера
- Зависит ли eCPM от времени дня

5. Вспомогательная информация

Столбец **profile_id** – id юзера

Столбец **event_json** содержит данные о событии показа рекламы:

ad_type – формат рекламы (int, banner или rv)

placement – когда/где был показ рекламы в игре

view_index – номер показа конкретного формата рекламы для юзера (не поддерживается для баннеров)

revenue – доход от конкретного показа рекламы ($eCPM = revenue * 1000$)

ads_ltv – накапливающийся доход от рекламы, равен сумме всех revenue юзера

network_name – рекламная сеть, от которой был показ

Остальные столбцы должны быть интуитивно понятны

6. После выполнения

По готовности, опубликуйте Jupyter Notebook с выполненным заданием на github и пришлите ссылку на репозиторий.

При оценке выполненного задания будет оцениваться:

- на сколько понятно визуализированы данные, хорошо ли выбраны средства для презентации данных (типы диаграмм, таблиц или др.)
- чистота и понятность кода
- описания наблюдений в процессе исследования и сделанные выводы