

Тестовые задания

Задание CTR & EvPM Analysis & Visualisation

На основе данных из файлов [interview.X.csv](#) и [interview.y.csv](#) разработать клиент-серверное WEB-приложение для визуализации и анализа последовательностей событий.

Требования к frontend-у:

- Применение любых js/css-фреймворков (предпочтительный вариант); для backend developer допустимо использовать статические страницы, отчеты, pdf/doc/xls документы, упрощенный UI (альтернативный вариант);
- Аккуратный и современный внешний вид;
- Визуализация графиков изменения показателей кампании во времени по [CTR](#) и [EvPM](#) с возможностью выбора/переключения между типами событий (см. описание полей [interview.y.csv](#));
- Визуализация агрегационных таблиц для следующих полей: (см. описание полей [interview.X.csv](#))
 - mm_dma;
 - site_id.

С возможностью выбора/переключения между типами событий.

Агрегация - это вычисление метрик для, например, различных значений mm_dma.

Агрегационные таблицы должны включать следующие столбцы:

- Число показов (impressions);
- [CTR](#);
- [EvPM](#).

Требования к backend-у:

- Применение Python, Java EE, JavaScript/Node.js или C#;
- Оптимизация приветствуется;
- Обработка запросов для графиков изменения показателей кампании во времени по [CTR](#) и [EvPM](#);
- Обработка запросов агрегационных таблиц показателей кампании.

Задание DS, ML, AI

Data Science & Machine Learning на основе данных из файлов [interview.X.csv](#) и [interview.y.csv](#).

1. Провести разведочный анализ данных, например в Jupyter.
Творческая задача на усмотрение исполнителя.
2. Провести корреляционный анализ между показами (характеристиками показа) и событиями, например - registration, content и т.д.
Визуализация на усмотрение исполнителя.
3. Обучить модель, способную прогнозировать вероятность возникновения события fclick или другого события для данного показа (характеристик показа).
Обосновать выбор модели и оценить ее качество.

Метрики CTR и EvPM

CTR

$CTR (click\ rate) = 100 * click_count / impression_count \%$.

[Ссылка на википедию](#)

EvPM

$EvPM (event\ rate) = 1000 * event_count / impression_count \%$ для запрошенного типа события. Признак view/click-through учитывать не требуется (т.е. при вычислении EvPM для (v)registration в числителе имеем сумму числа registration и vregistration).

Данные тестового задания

Загрузка данных

Файлы данных interview.X.csv и interview.y.csv [доступны к загрузке по этой ссылке](#).

Описание данных

У файлов interview.X.csv и interview.y.csv есть общий ключ - идентификатор показа (uid).

interview.X.csv

Лог показов (просмотров).

1. reg_time - timestamp;
2. uid - идентификатор показа;
3. fc_imp_chk - число предшествующих показов:

Значение	Описание
-1	N/A
0	1 impression
1	2-5 impressions
2	6-10 impressions
3	11-20 impressions
4	21+ impressions

4. fc_time_chk - время с момента последнего показа:

Значение	Описание
-1	N/A
0	Less than a minute
1	1-10 minutes
2	11-30 minutes
3	31-60 minutes
4	1-3 hours
5	4-24 hours
6	One or more days
7	Never before

5. utmtr - время просмотра в часов поясе пользователя:

Значение	Описание
-1	N/A
0	00:00 - 03:00
1	03:00 - 06:00
2	06:00 - 09:00
3	09:00 - 12:00

4	12:00 - 15:00
5	15:00 - 18:00
6	18:00 - 21:00
7	21:00 - 00:00

6. mm_dma - DMA;
7. osName - OS;
8. model - модель устройства;
9. hardware - тип устройства;
10. site_id - сайт, где был просмотр.

interview.y.csv

Лог событий.

1. uid - идентификатор показа, в результате которого произошло событие;
2. tag - тип события:
 - буква <v> в начале некоторых событий означает, что событие произошло без клика = view-through. В противном случае имеем click-through событие;
 - событие fclick - первый клик.