

مرحله اول چالش تحلیل داده برای حل مسائل تجاری کافهبازار - مهر ۱۳۹۸

فایل پیوست حاوی اطلاعات مربوط به کاربران اپ اندروید دیوار است.

<https://drive.google.com/open?id=1rbXEGszvK1NpdPM0ULf1Q4yN8Q6dKtN>

توضیحات ستون های این جدول به شرح زیر است:

1- adid

شناسه یکتای هر کاربر

2- network_name

شبکه تبلیغاتی که کاربر با دیدن تبلیغات این شبکه اپ دیوار را نصب کرده است. مقادیر این ستون این موارد هستند:

Organic- Tapsell- Adword UAC Installs - Cafebazaar searchad

نصب های organic، نصب هایی هستند که بدون دیدن تبلیغات اپ را نصب کرده اند.

3- installed_at

زمان نصب کاربر (با دقت ثانیه)

4- activity kind

نوع فعالیت انجام شده که شامل مقادیر install و event است. منظور از event، فعالیت هایی است که کاربر داخل اپ انجام می دهد.

5- event_name

نام event انجام شده در اپ، شامل مقادیر get_contact و new_post_submit. منظور از get_contact، زدن دکمه «اطلاعات تماس» در صفحه آگهی است. منظور از new_post_submit، زدن دکمه «تایید» در صفحه سوم فرایند ثبت آگهی در اپ است، که به معنای پایان ثبت آگهی است. این ستون فقط برای سطرهایی که activity_kind = event باشد مقدار دارد و در غیر اینصورت null است.

6- created_at

زمان به وقوع پیوستن activity (با دقت ثانیه)

ابتدا تمیز بودن و صحت دیتای جدول را بررسی کنید. پس از آن به سوالات زیر پاسخ دهید:

- ۱- برای مقایسه کیفیت کاربران نصب های network های تبلیغاتی، چند متریک تعریف می شود که دو مورد آنها موارد زیر هستند:
 - درصد کاربرانی که نهایتاً تا سه روز بعد از نصب اپ، حداقل یک event مهم در اپ انجام داده اند (get_contact یا new_post_submit)
 - نرخ بازگشت (retention rate) کاربران فوق الذکر (یعنی کسانی که سه روز بعد از نصب اپ حداقل یک event مهم انجام داده اند) در بازه زمانی 30 روز بعد از اولین event بعد از نصببرای network های مختلف، این دو متریک را محاسبه کنید و با نمودارهای مناسب نمایش دهید.

۲- اگر در بررسی هایتان به رفتارهای مشکوکی برخورد کردید، آنها را با نمودارهای مناسب توضیح دهید.