

گزارش تحلیل حل مسائل تجاری

کافه بازار

پاسخ چالش ۱

مهدی ناصری

۰۹۱۲۷۱۶۲۶۲۹

mahdi.nasseri@gmail.com

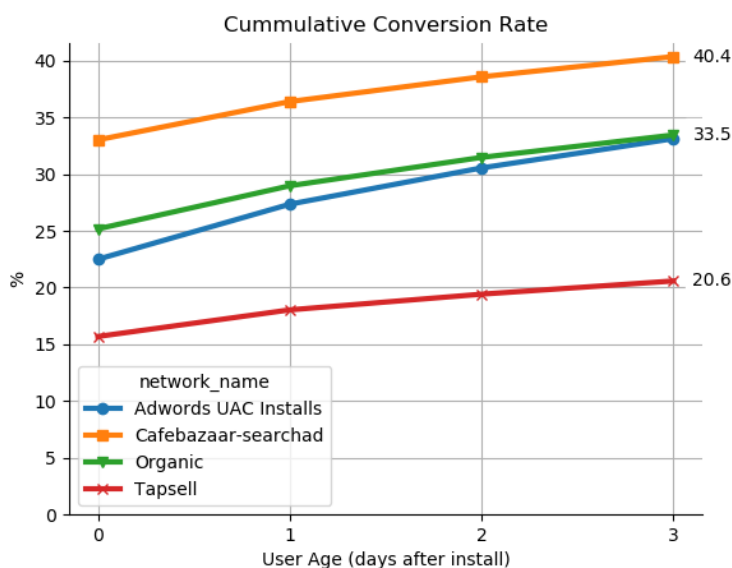
مهر ۱۳۹۸

پاکسازی داده‌ها

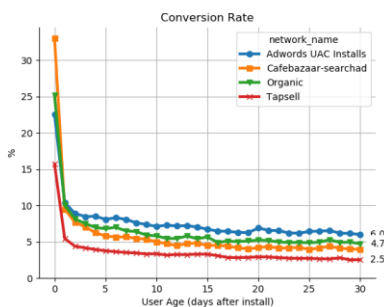
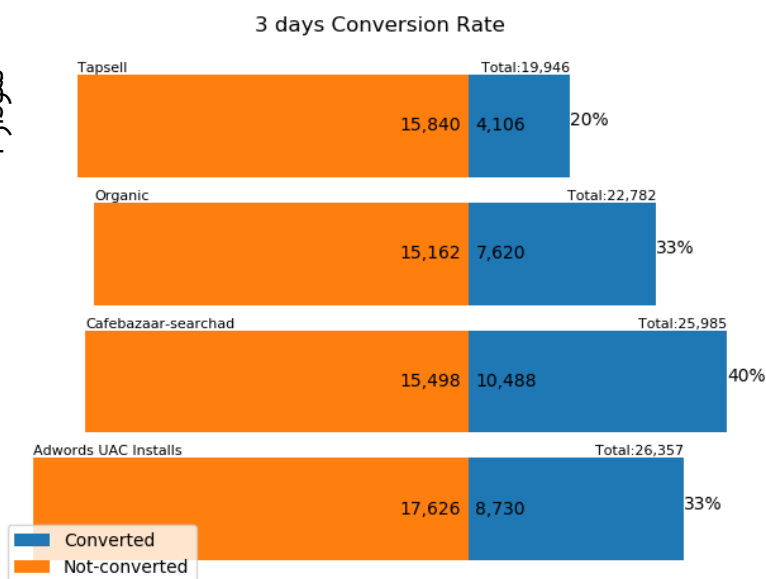
برای پاکسازی داده‌ها این موارد بررسی و در موارد موجود رفع شد

- **تکراری بودن رکوردها:** تنهای یک مورد رکورد تکراری برای نصب یکی از کاربران ثبت شده بود.
- **عدم یکپارچگی فرمت ستون‌ها:** ناهمخوانی در ستون‌ها وجود نداشت به جز ستون `created_at` که فرمت زمان‌ها یکسان شد.
- **عدم یکپارچگی منطقی فیلدها:** بررسی شد آیا ممکن است رکوردی اشتباه منطقی داشته باشد یا نه. برای این مورد توالی زمانی تاریخ نصب و رویدادها بررسی شد که موردی وجود نداشت.
- **ناموجودی رکوردها:**
 - هیچ فیلدی از هیچ رکوردی NULL نبود
 - تعداد بیش از ۵۰۰۰ کاربر وجود داشت که event برای اونها ثبت شده بود ولی رکورد install نداشتند. رکورد install این کاربران با توجه به رکوردهای event تولید شد.

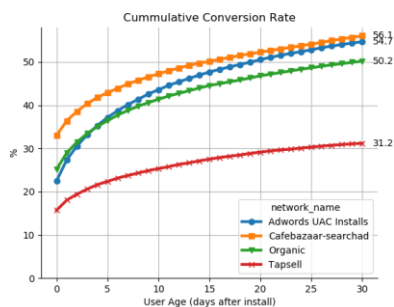
نمودار ۱



نمودار ۲



نمودار ۳



نمودار ۴

نرخ تبدیل

سه روزه کاربر به کاربر فعال

- یافته کلیدی

کاربران cafebazaar-searchad نرخ تبدیل ۳ روزه بالاتری دارند. کاربران adword و organic در مقام دوم بوده و کاربران tapsell با اختلاف در مقام چهارم هستند.

- تحلیل

به طور کلی کاربران cafebazaar-searchad نرخ تبدیل بالاتری را نسبت به سایر شبکه‌ها دارند. این نرخ در سه روز اول ۴۰٪ و در ۳۰ روز اول به حدود ۵۶٪ رسیده است. بیشترین تبدیل در همان روز اول اتفاق افتاده است. کاربران adwords رشد مدام بیشتری دارند به طوری که در روز ۳ در مقام ۳م بوده ولی بعد از ۳۰ روز به نزدیکی ۵۴٪ می‌رسند. این کاربران بیشترین نرخ تبدیل را در روزهای بعد از نصب دارند (نمودار ۳).

کاربران تپسل کمترین نرخ تبدیل را در کل دوره دارند.

- شرح روش

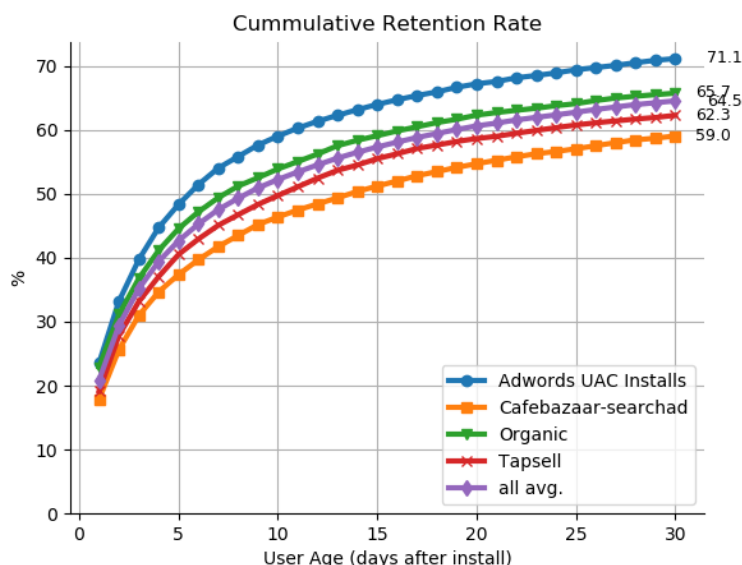
نمودار ۱: نرخ تجمیعی تبدیل کاربران در ۳ روز اول را به تفکیک شبکه جذب کاربر نشان می‌دهد. منظور از نرخ تجمیعی این است که تا مثلاً روز ۳م به طور کلی از اول چه نرخ از کاربران فعال شده‌اند.

نمودار ۲: نرخ تبدیل سه روزه را به تفکیک هر شبکه نشان می‌دهد.

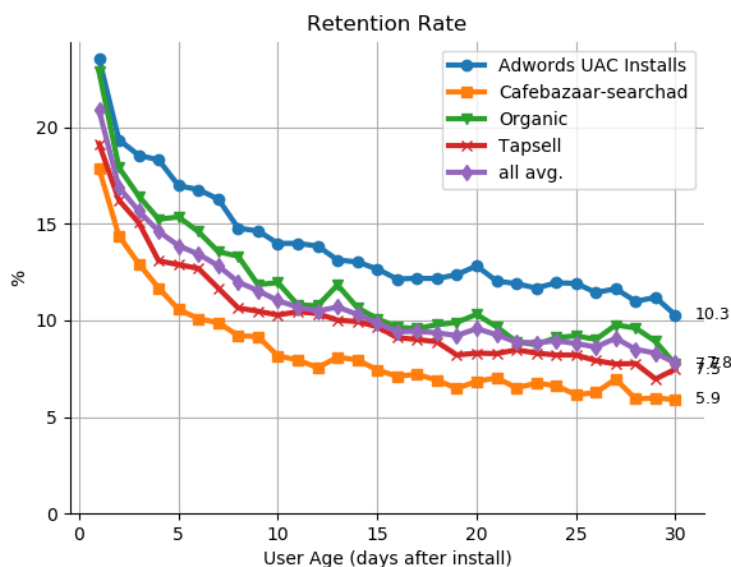
نمودار ۳: نرخ تبدیل را در هر روز به طور مستقل نشان می‌دهد.

نمودار ۴: نرخ تجمیعی را تا روز ۳۰م نمایش می‌دهد.

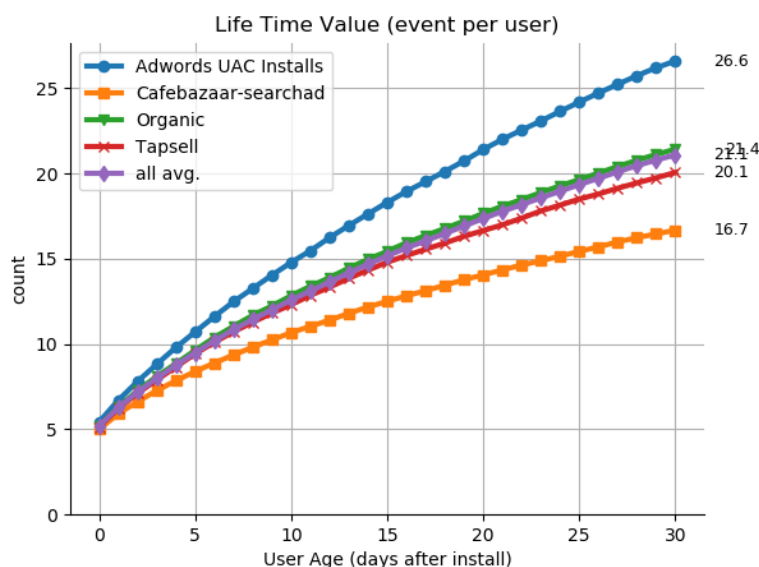
نمودار ۱



نمودار ۲



نمودار ۳



نرخ نگهداشت

۳۰ روزه کاربر فعال

- یافته کلیدی

کاربران adwords بیشترین نرخ نگهداشت ۳۰ روزه (حدوداً ۷۱٪) دارند. کاربران organic در مقام دوم و tapsell و cafebazaar در مقام سوم و چهارم هستند. (نمودار ۱)

در کل ارزش طول عمر کاربران فعال adwords بیشتر از همه است. این کاربران به طور میانگین در ۳۰ روز اول بعد از فعال شدن حدود ۲۶ فعالیت (event) انجام داده‌اند. (نمودار ۳). کاربران cafebazaar-searchad کمترین ارزش طول عمر را دارند (۱۶ رویداد به ازای هر کاربر)

- تحلیل

به طور کلی کاربران فعال با نرخ مناسبی در روزهای بعد فعال نگهداشته می‌شوند. به طور میانگین ۶۴٪ کاربرانی که فعال می‌شوند تا روز ۳۰م حداقل یک فعالیت دیگر انجام داده‌اند.

این نرخ به طور خاص برای روز ۳۰م برابر با ۷/۵٪ است. یعنی ۷٪ کاربرانی که فعال می‌شوند ۳۰ روز بعد هم حداقل یک فعالیت انجام می‌دهند. (نمودار ۲).

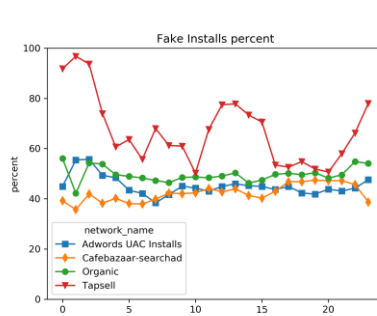
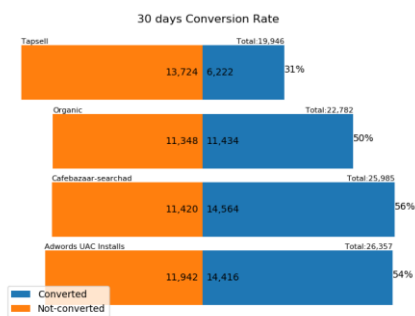
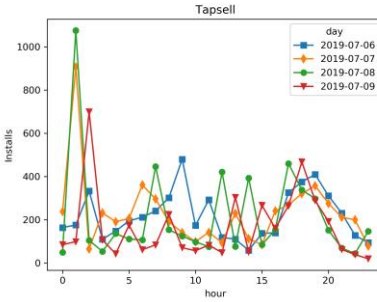
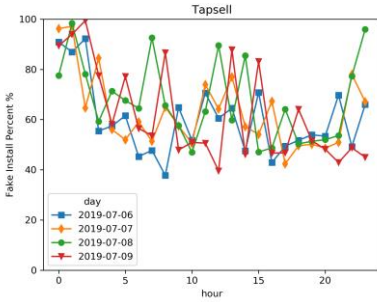
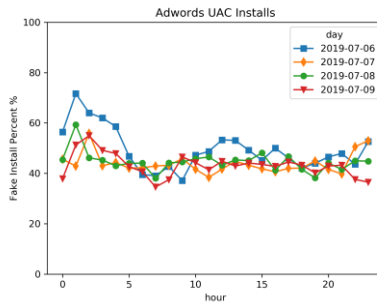
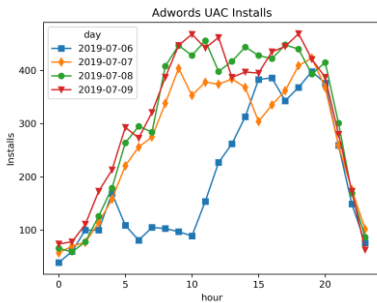
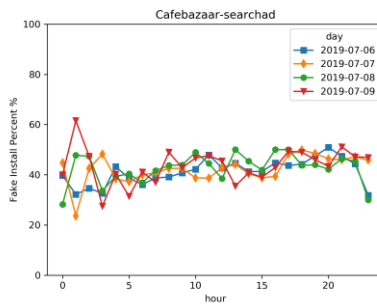
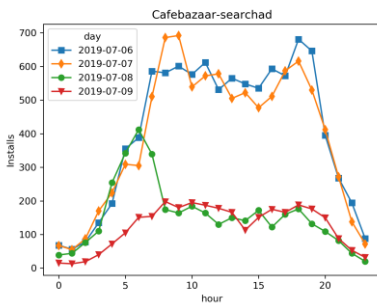
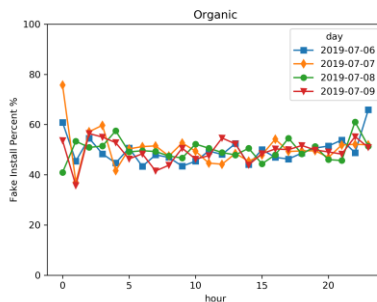
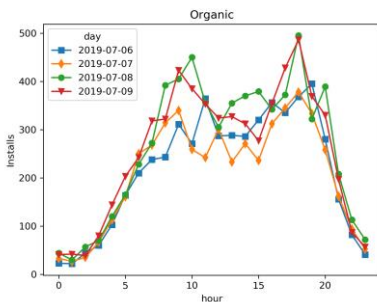
نمودار ۳ نشان می‌دهد که ارزش طول عمر کاربران با توجه به شبکه جذب آنها ممکن است تفاوت معناداری داشته باشد. با این تحلیل کاربران adwords کیفیت بالاتری نسبت به بقیه شبکه‌ها دارند.

- شرح روش

نمودار ۱: نرخ تجمیعی نگهداشت (بازگشت) کاربران فعال را در ۳۰ روز اول بعد از فعال شدن نشان می‌دهد.

نمودار ۲: نرخ بازگشت روزانه را تا روز ۳۰م نمایش می‌دهد.

نمودار ۳: ارزش طول عمر کاربران فعال را به صورت تجمیعی تا روز ۳۰م نمایش می‌دهد. ارزش طول عمر کاربر (LTV) به صورت میانگین تعداد کل event به ازای هر کاربر محاسبه شده است.



نصب‌های Fake

نصب‌هایی که فعال نشده‌اند

یافته کلیدی

شبکه tapsell به طرز محسوسی نصب‌های fake می‌گیرد.

تحلیل

همانطور که در نمودار ۴ مشاهده می‌شود. تعداد نصب گرفته شده در ساعات شبانه روز توسط tapsell اصلا با نرم رفتاری بقیه شبکه‌ها و مردم متناسب نیست. برای مثال در ساعت بامداد نرخ نصب همه شبکه‌ها کم است و این طبیعی‌ست در حالی که در tapsell هیچ قاعده‌ای وجود ندارد. بنابراین احتمال بالا این نصب‌ها fake هستند. همانطور که نمودار ۴ سمت راست نشان داده شده نرخ fake بودن هم نامنظم و هم بالاست. در حالی که بقیه شبکه‌ها رفتار ثابتی از نظر تعداد نصب، روند نصب و نرخ fake بودن دارند.

در نمودار ۵-چپ نرخ کلی نصب‌های fake به تفکیک شبکه‌ها نمایش داده شده است. این نمودار هم به طور کلی یافته اصلی این تحلیل را تایید می‌کند.

نمودار ۵-راست نرخ fake بودن نصب‌ها را به تکیه هر شبکه در ساعات شبانه‌روز نمایش می‌دهد. همانطور که مشخص است به جز tapsell بقیه شبکه‌ها نرخ تقریباً ثابتی را در طول روز دارند.

رفتار مشکوک

نسبت بالای نصب fake از طریق شبکه‌ها. نصب fake به نصب‌های گفته شده است که در عرض ۳۰ روز هیچ فعالیتی نداشتند. تعداد نصب و نرخ fake بودن به تفکیک هر شبکه در هر روز و هر ساعت محاسبه شده است.

روش

برای هر شبکه به طور مجزا تعداد نصب‌ها در نمودارهای سمت چپ نشان داده شده است. در این نمودارها تعداد نصب به تفکیک هر روز در ساعات شبانه‌روز نمایش داده شده است. در نمودارهای سمت چپ نرخ fake بودن به تفکیک هر روز و هر ساعت نمایش داده شده است.