Celtra Dataset Visualizations - Project 2 IDS

Semir Salkić[1]

[1] *ss9343@student.uni-lj.si, 63190409*

Introduction

This report provides insight and detailed description of two datasets from Celtra. First dataset is platform usage data, which recorded user activity over period of 3 months. It contains 622,078 data points (rows). Every row contains accounts (ID), users (ID), sessions (ID), activity, activity location, and a timestamp. Second dataset is Celtra sessions data, which recorded advertisement traffic and relevant data on display devices during period of two years(2018 and 2019). It contains 4,823,186 data points. Every row contains accounts (ID), campaigns (ID), creative (ID), platform, SDK, Sessions (requested, loaded, rendered, interacted) creative load attempts and viewable time.

Platform usage data

Data contains 11 unique account IDs. Users have 265 distinctive IDs. It is important mentioning that users can be anonymous. Each user(employee) belongs to a defined account (i.e. company). While using platform, data is recorded in form of activity. Activity location is detailed categorization of activities. Activity is categorized in five major categories:

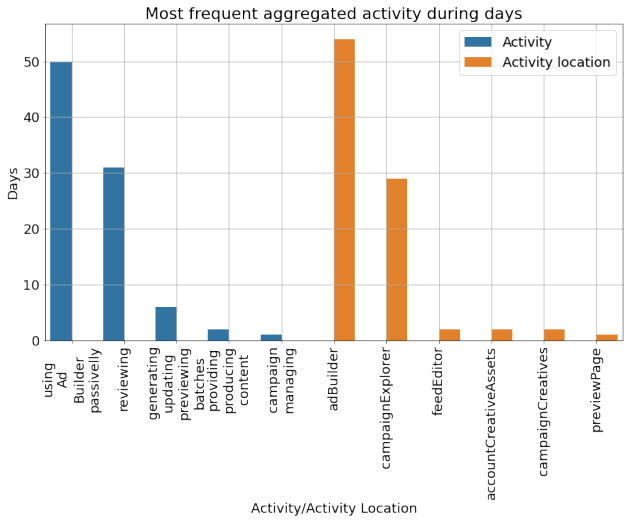


Figure 1. Activity bar plot. Activity is counted in a day if it was a most frequent activity.

- using Ad Builder passively
- Reviewing
- Generating/updating/previewing batches
- Providing producing content
- Campaign managing

In Figure 1 we can see two bar plots showing most frequent activities and their location during data recording time frame (1^{st} June - 29^{th} September 2020). From this, we can see that Ad Builder is dominantly used part of the platform. Having in

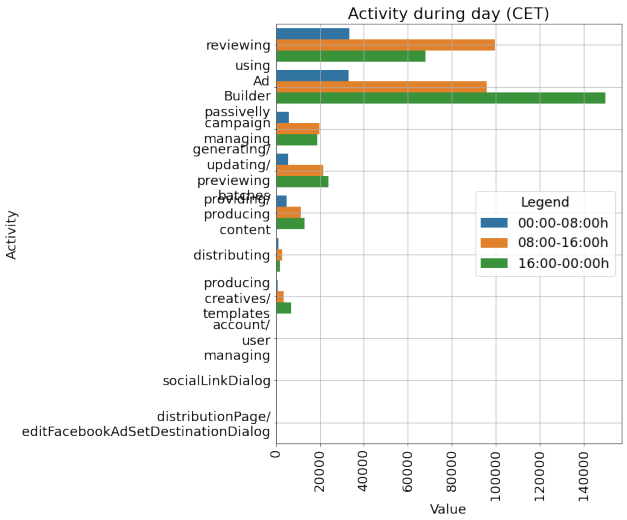


Figure 2. Activity during day. Each category is shown in day with 3 intervals

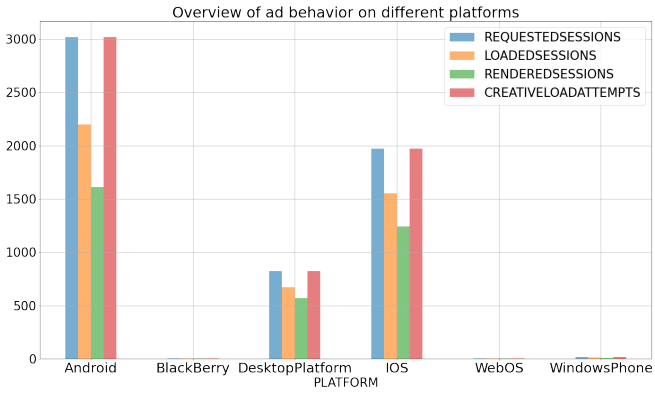
mind that platform is used 24 h due to geographical diversity of users and work teams, this results in different time zones (practice common in marketing companies). In the Figure 2, data is labeled in three intervals. Night (00 : 00 - 08 : 00), work (08 : 00 - 16 : 00) and evening (16 : 00 - 00 : 00) intervals. It can be seen diverse behavior for different activities. For example, most frequent category - "Ad Builder" is mostly used during evening, which can support claim that recorded users are from different time zones. Other categories are showing also unexpected behavior where we can see that majority of categories are used during work and evening hours.

Sessions data

This dataset is collected over period of two full years. Further grouping can divide attributes of dataset to couple of groups.First group is general data, such as: UTC date, account/campaign ID, account/campaign creation date, creative ID, creative creation date. Other groups describes user interaction with the published advertisement. Each ad is shown on a device which can be defined by platform and SDK[1], which can be defined as second (device) group. Third group is creative group (i.e. group of ads for certain product are united and same creative). This group is described by creative ID and creative load attempts (number of ads that were attempted to be loaded on the devices) in a event of user interaction. Last group is sessions. They are defined as a simple request from user device to view the ads. Each data point is described with requested sessions (number of requested sessions from users device), loaded sessions (successful sessions loaded on the device), rendered sessions (shown on the device) and sessions with interactions. This information is probably most important for marketing companies since the main goal is to maximize usage of their ads. Also, viewable time is available as a metric for each data point.

In Figure 3 overview of different platforms is presented. It is possible to conclude that most popular platform for Celtra users is Android. It is interesting to mention that platform such as: WebOS, BlackBerry are sparse. Also, we can see that smartphone (mobile) devices are preferable for the users much more then Desktop platform (PC). As can be seen in Figure 5

Figure 3. Overview of ad behavior on different platforms It is shown that requested sessions closely follow load attempts. It is noticeable that in each case more than 50% of requested ads are rendered.



sessions with interaction is presented. Sessions with interactions are successful session where ad is interacted by user on their device. Here data from 2018 and 2019 is presented separately to support the claim that this data contains distinctive patterns. Besides that, smoothed is included where general nature of whole dataset is presented in simpler manner. Figure 5 shows two time intervals. Yellow time interval represents beginning summer holidays, where expected marketing traffic is going up due to couple of facts. Generally, third Sunday of June is Father's Day, which is holiday adopted as international day by more and more countries around the world. Also, June in the USA (the biggest general consumer) is declared international camping month [1]. Besides that, June is traditionally start of summer vacations for most people. Green time interval(29^th November - 25^th December) is marked as winter holidays. In this interval we can notice biggest growth of interactions. Reason for this is couple of international holidays (e.g. Black Friday, Cyber Monday, Catholic Christmas). This mixture of commercial and religious holidays forms a time interval where increasingly growing population of consumers are interacting with various ads on wanted product.

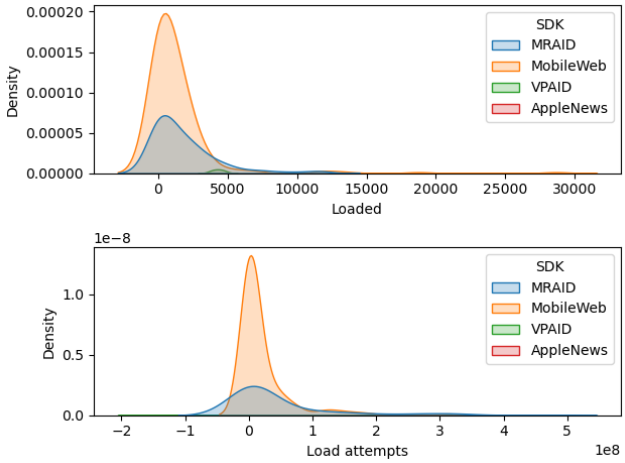


Figure 4. KDE density distributions for different SDK (grouped by Accounts). We can see that SDK of choice for most accounts is Mobile Web.

Figure 4 shows KDE distributions by different SDK. Data beforehand is grouped by same accounts to form distributions on loaded sessions and load attempts. We can see that distribution is "normal-like" and centered around zero. Load attempts distributions is distributed with small quartiles and noticeably skewed to the right. It is shown that as expected, mobile web is most preferred SDK (also in Fig. 3).

Conclusion

In this report we have presented Celtra platform usage and sessions dataset. W have shown activity behavior of users in the first dataset and interesting behavior during 24h, which can be beneficial insight for adaptive maintenance and load balancing of platform resources and traffic. In the second dataset, discovered patterns were presented. By sessions interaction visualization we proved that we can connect unreasonable growth in some parts of the year with user events and make sense out of it. With this we can say that acquired inference can be used to further analyse data and provide meaningful insight for construction of predictive models.

[1] *Software development kit*

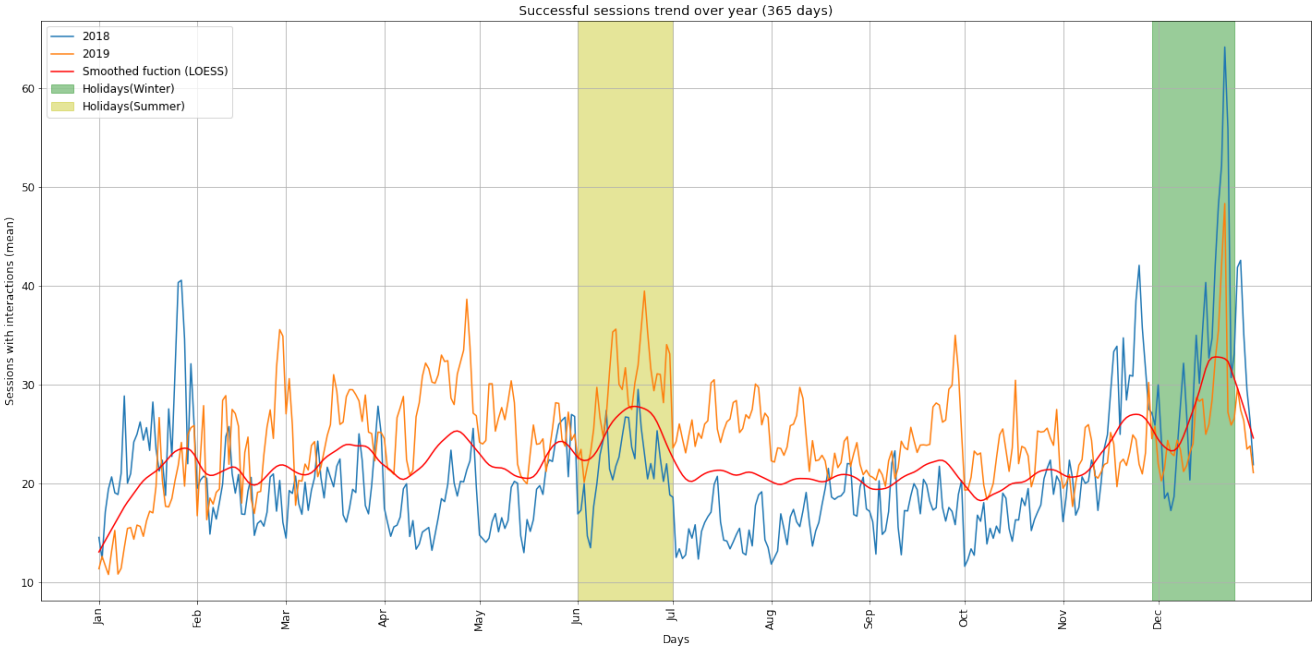


Figure 5. Sessions with interactions during year. Highlighted holidays confirm significant growth in comparison to the rest of the year.

References

[1] National Day Calendar 2020. National camping month – june, 2020.