Formal Problem Statement (FPS)

Author: Rajat Jain

Last Updated: 2018-04-10

Contents

1. Goal

Adobe sells its Creative Cloud products as various single apps and bundled subscription plans. One of such bundle is Photography plan which bundles Adobe Photoshop with Adobe Lightroom and a few other services. It is one of Adobe's most popular packages and was originally tailored to bring in more professional as well as amateur photographers into the family by offering discounted prices on the software of their need. From the customers' perspective, it is also a good way to get cheaper Photoshop than buying it as a single app subscription. Pricing page has more details. Hence it attracts a lot of non-photographers as well.

Adobe currently uses the information from the plan purchased to run targeted marketing campaigns and to send learning tutorial recommendations to the customers to increase their engagement in-product; and to eventually increase customer retention. An example campaign could be, identified photographers may receive a marketing communication reminding them that they also have cloud storage associated with their Adobe account and should use it to reliably save their photos online. However, these campaigns are often found to be less effective on photography plan customers. Inacurate targeting due to a mix of photographers and other customers is often understood to be the reason behind it.

To acurately target photographers, we need to classify the customers buying this bundle into "true" photographers v/s "non" photographers.

2. Key Stakeholders

a. Business Owners

Adobe's Creative Cloud Product & Marketing team led by Juanjo will sign-off on the requirements of this project and will refine the output requirements, if needed.

b. SMEs

Photoshop and Lightroom Product Management and Product Analytics team are the subject matter experts on the features of Photoshop and Lightroom. They also are the domain expert on what kind of features are

used by different classes of their users. Damon Lapoint will help us understand that domain knowledge and incorporate it into data cleansing and model building.

c. Customers

The output of this project will primarily be consumed by the Photography Product Marketing team at Adobe lead by Sr. Director Jim. Additionally, Johnson's Creative Cloud Learning Content team will be responsible for running the campaigns to target "true" photographers with content encouraging them to engage with Lightroom CC product.

3. Impact

Being able to identify "true" photographers will help the business will be able to,

a. increase the effectiveness of the campaigns targeted towards photographers.
b. steer other customers towards the right products or plans for them and increase overall engagement.
c. measure accurately and increase customer's engagement with products and eventually increasing retention in longer term.

4. Execution Plan

We will build models based on various customer attributes and their usage of the produts and product features, to segment paid photography plan customers into "true" Photographer and "non" Photographer segments. The output of this project will be a table on Hadoop consisting of all qualifying members with their corresponding segment at a given snapshot date/time.

a. Data Sources

The following major data sources are available to us:

- Customer profile consisting various attributes like geo location, plan purchased etc. This data resides in Adobe's financial systems and is synched over to a centralized Hadoop cluster.
- Customer's product and product feature usage data collected by various Adobe apps running on their machines.
- Results from a survey where customers have self-classified themselves into various job functions.

All the above data sources have been classified as "Company Confidential" or "Customer Confidential" data and may not be exported out of Adobe's enterprise network.

5. Deployment Plan

Model(s) will live on-prem on Adobe owned enterprise wide data lake. They will be re-trained every week to adjust for the changes to the market, offers and customer behaviors.

6. Resources

a. Human Resources

We will need one data scientist to work full time on this project for the estimated period of time.

b. Machine Resources

No special hardware is required at this time based on the estimated data volume. For future iterations, specialized hardware or data processing platforms may be required depending on the growth in size of data and desired frequency of model update.

7. Major Milestones

The first version of the project will take seven weeks to complete.

Estimated Start Date	ETA
04/11/2018	05/30/2018

During this period, the following major milestones will be tracked.

a. Ingest:

Data exploration, clean-up, ingestion and feature generation will take three weeks and is estimated to be completed by 05/02/2018.

b. Model:

Model(s) will be built on the ingested data and will take three weeks. They will be available for validation(if required) by 05/23/2018.

c. Validate:

No validation testing on production data is planned yet.

d. Deploy:

If accepted, the model(s) will be deployed to production. Deployment plan is as discussed earlier. Deployment process will take seven days. ETA: 06/30/2018.

e. Iterate:

Plans to iterate for next version, if any. TBD.

*All dates are tentatively estimated based on the expected start date. Exact milestone dates will be known when the project starts.

8. Success Criteria

The ultimate desired business impact from this project is the lift in customer's engagement with Adobe products and eventual lift in retention rates. However, both of these are long term measure and may not be tangible in short term.

The success of this project will be measured by the accuracy of classifying "true" photographers among the paid photography plan customer base. We expect to achieve 70% or more accuracy to be successful.

9. Risks

The possible risks to this project are as below:

- Product Marketing team may decide to not use the classification in running their campaign if the accuracy of the model(s) is less than acceptable.
- Adobe products like Photoshop tend to have thousands of major and minor instrumented features, localized into many languages. Using the data as available may not help much and may even be sparse. To avoid this, we would need to build a taxonomy of features with help from the product SMEs.
- Since the primary data for this project is customer's product usage, we may not be able to classify the users with no or low usage.