Journey App Team 2 Deliverable 2

Andrew Whittum, Giacomo Ieronutti, Jaewoo Kang

For this deliverable we continued to build on the work that we did for deliverable 1. To summarize our work for deliverable 1, we added several new columns to the dataset. Two of these were “Activation Day” and “Activity Day” which represent the day of the week the content was sent to the user and the day of the week when the user opened it. We also added a column for “Activity Delay” which represents the amount of time elapsed between when the user received a piece of content and when they opened it. We also added “Activation Time” and “Activity Time” to isolate the time of day when a user received and opened a piece of content (isolating this from the date).

Some further preprocessing we did for the dataset since last time was to remove rows where the user looked at the same piece of content multiple times. There were some pieces of content which the user had looked at many times, we wanted each content to only appear once in the data set, so we only considered the first instance when a user opened a piece of content.

Something we looked at for this deliverable was the possibility of breaking down the dataset by customer id instead of user id, but we found that there were 78 total companies in the dataset and some of them had hundreds of thousands of entries, so this made it very difficult to find one time each day, the client agreed and we decided to abandon this line of inquiry.

Another technique we looked into was running an NLP analysis on the “Name” column. For each user we ran a vectorizer on the “Name” column and then we one-hot-encoded all of the unique words that the user saw. Then we calculated the average completion for content when the user saw each word. The client had expressed some interest in this earlier in the semester, but he did not seem as interested when we showed him our results. We will not be continuing with this line of inquiry although we may include this code in the final deliverable.

For the algorithm that we are using for each user, we are first looking at rows with 100% completion. Then we are looking at the Channel column, specifically we are looking first for times when the user accessed the content via a mobile notification. Then we are looking to see if any of these instances occurred within one hour of receiving the notification. If there are any instances, we will return them If not, we look for times when the user accessed the notification with a delay of more than one hour. Then we will look at the direct notifications, and similarly break it down by delay time. For all of the leaf nodes in our algorithm, we will return up to 3 times. Once we have exhausted all rows with 100% completion, if we still haven’t returned any times, we will sort the rest of the rows by completion, only considering 85% to 100% completion. Then we will return the time with the highest completion.

At any point in the algorithm if we find a time that fulfils all the requirements, we will return that time and then move on to the next day. We will run this algorithm for each day of the work week. In addition to the time, we will return the Device, Channel, Delay, Completion and Content Type corresponding the time that we are returning. A sample output is included on the next page, and a visual representation of the algorithm is including on the following page.

Monday

Time	Device	Channel	Delay	Action	Content Type
18:07:54	Android	Direct Connection	0.13	100	FLIPBOOK
08:46:13	Android	Direct Connection	0.79	100	LIVE_EVENT
11:22:33	Android	Direct Connection	0.87	100	EXTERNAL_VIDEO

Tuesday

Time	Device	Channel	Delay	Action	Content Type
11:18:27	Android	Direct Connection	0.31	100	INTERNAL_VIDEO

Wednesday

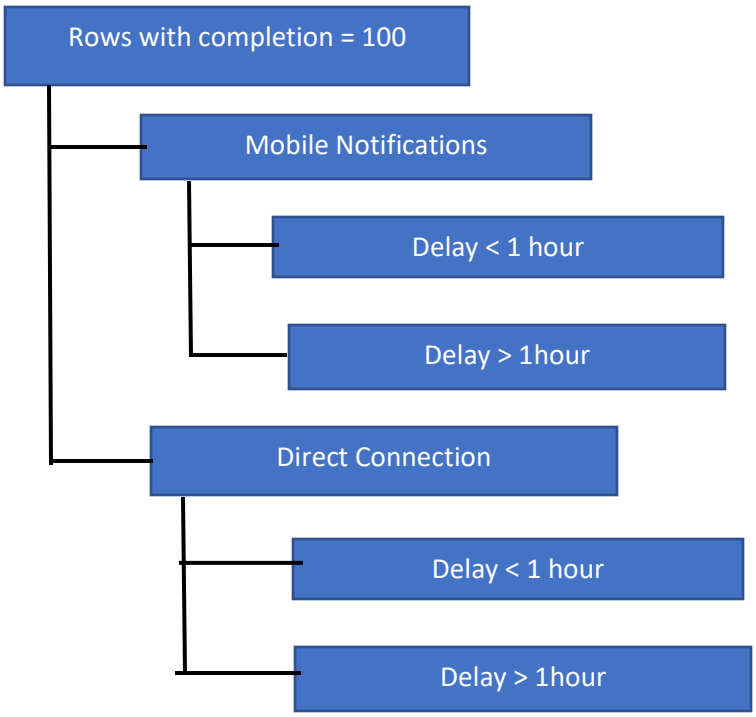
Time	Device	Channel	Delay	Action	Content Type
15:21:44	Android	Direct Connection	0.36	100	INTERNAL_VIDEO

Thursday

Time	Device	Channel	Delay	Action	Content Type
13:08:43	iOS	Mobile Notification	0.77	100	INTERNAL_VIDEO

Friday

Time	Device	Channel	Delay	Action	Content Type
19:32:50	Android	Direct Connection	1.27	100	INTERNAL_VIDEO
18:14:47	Android	Direct Connection	2.21	100	EXTERNAL_VIDEO
20:11:01	iOS	Direct Connection	2.69	100	INTERNAL_VIDEO



Sort the rest of the rows by completion, return the time for the highest completion