User segmentation based on mobile app behavior

October 7 to November 3, 2019

Table of contents

Objectives

1. Identify the most common events performed by users who view contacts.
2. Perform an analysis of the events that are connected to the target event.
3. Analyze user groups based on different metrics: retention rate, time spent in the app, frequency of events, conversion to the target event - 'contacts_show'.
4. Test statistical hypotheses:
   - Hypothesis 1: Conversions to contact views between users who installed the application via a link from yandex and users who installed the application via a link from google have statistically significant differences.
   - Hypothesis 2: Site search (search) to contact views and recommender (tips_show) to contact views conversions are different.

Final conclusion

- Data is available from October 7 to November 3, 2019. There is almost a 28-day time difference.
- Total events 4293, average events per user 17.3, median 9, maximum events per user 478.
- There is a maximum time of 0 days 01:38:05 between the first and target events, an average of 0 days 00:09:21, and a median of 0 days 00:04:49. Based on the histogram, the most values fall on the ~20 minutes.
- Who completed the target event: they like photos_show more - 20.2% versus 13.9% in the second group; less like advert_open - 6.5% versus 8.9% in the second group; they like search more - 11.3% versus 10.5% in the second group; they like map less - 4.4% versus 5.8% in the second group; they like favorites_add more - 2.4% versus 1.9% in the second group; differences in tips_click and tips_show are not significant.
- Users who were attracted on October 14 showed the highest retention rates, as indicated by the light spots on the chart.

Final conclusion

- Four groups were formed after clustering. There are 64 users in the smallest group and 3294 users in the largest group.
- It takes 9:02 minutes for group 2 to reach the target event, while group 3 takes 12:38 minutes. Group leader 3 spends the most time since it performs more actions.
- The highest conversion rate was in group 2 (62%).
- Group 2 has the highest retention rate, perhaps because users like to search for things on their own.
- At a given critical level of statistical significance, there are no statistically significant differences between the conversions of users who installed the application via a link from yandex and users who installed the application via a link from google. The source has no effect on the target event's views.
- At a given critical level of statistical significance, there are statistically significant differences between conversion rates of site search (search) to contact views and recommendation system (tips_show) to contact views.

Analyzing user activity for anomalies
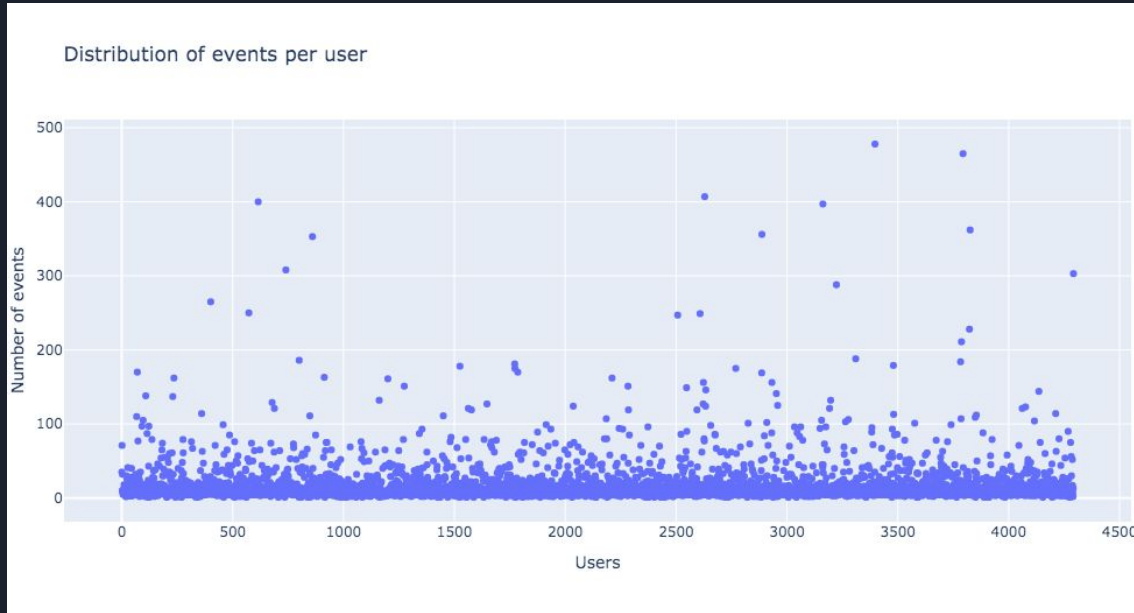
Distribution of events per user

- Total events 4293;
- Average events per user 17.3;
- Median 9;
- Maximum events per user 478.

Data provided by Yandex (IT company))

Analyze the distribution of events by day

The distribution of events by day
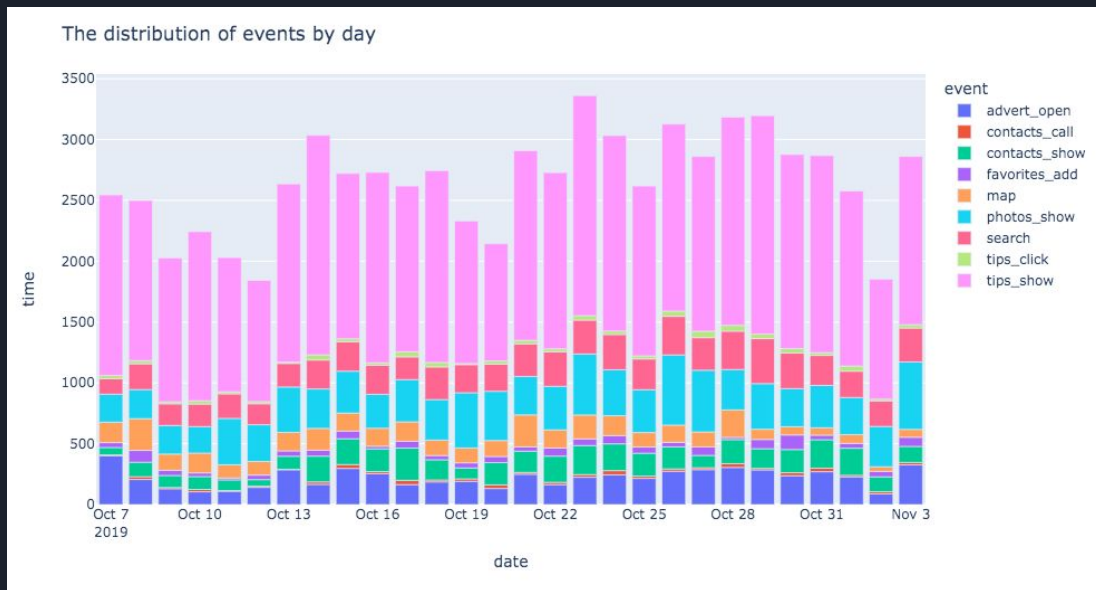
Since we see very few tips_click events, the vast majority of tips_show events (the user saw recommended ads) are automatic and do not indicate interaction with the mobile application.

Data provided by Yandex (IT company)

Analyze the distribution of events by day without tips_show

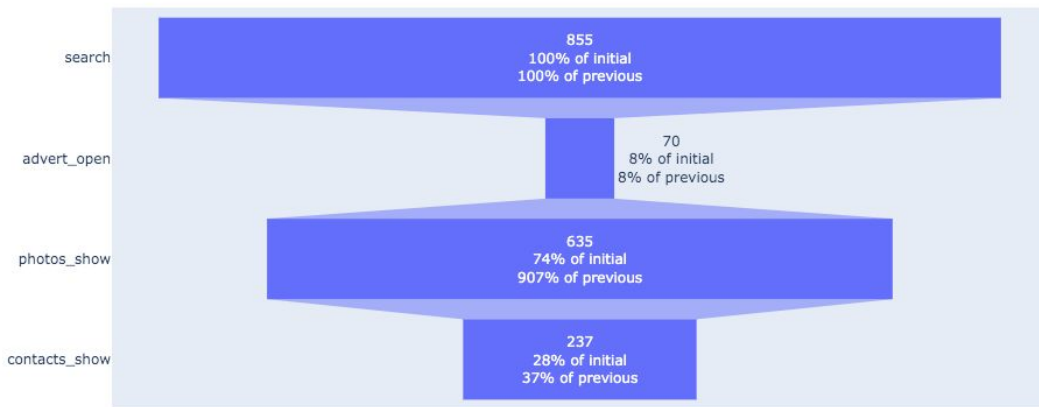The distribution of events by day without tips_show

- Users often use independent search for ads on the site;
- Users do not often add ads to favorites;
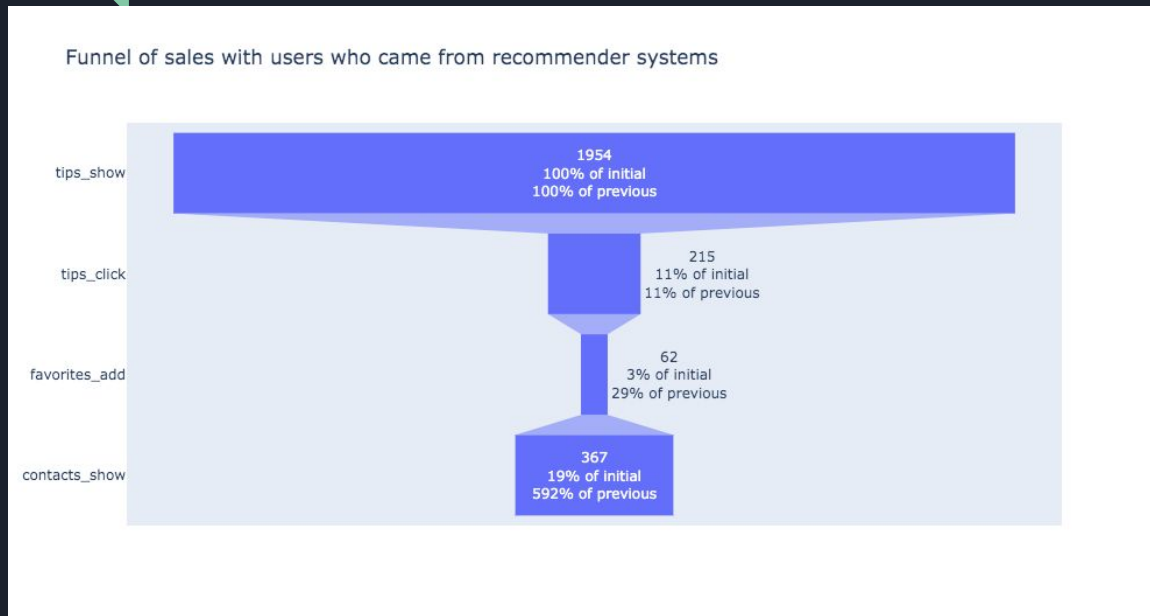- Users often browse contacts on the site.

Funnel of sales: search, advert_open, photos_show, contacts_show

Funnel of sales with users who came from search engine

|	
search	855 100% of initial 100% of previous
advert_open	70 8% of initial 8% of previous
photos_show	635 74% of initial 907% of previous
contacts_show	237 28% of initial 37% of previous

It takes a lot of time to search and only 8% of users open ads. Users probably have the choice to open the ad or the photos right away. 74% of users open photos immediately + 8% of users open ads = 82%. 27% of initial users view contacts.

Funnel of sales: tips_show, tips_click, favorites_add, contacts_show

Funnel of sales with users who came from recommender systems

tips_show
1954
100% of initial
100% of previous

tips_click
215
11% of initial
11% of previous

favorites_add
62
3% of initial
29% of previous

contacts_show
367
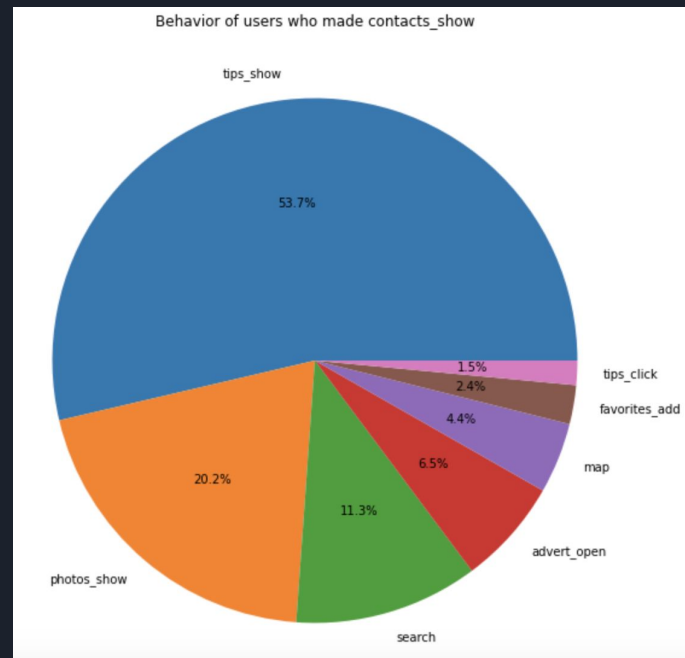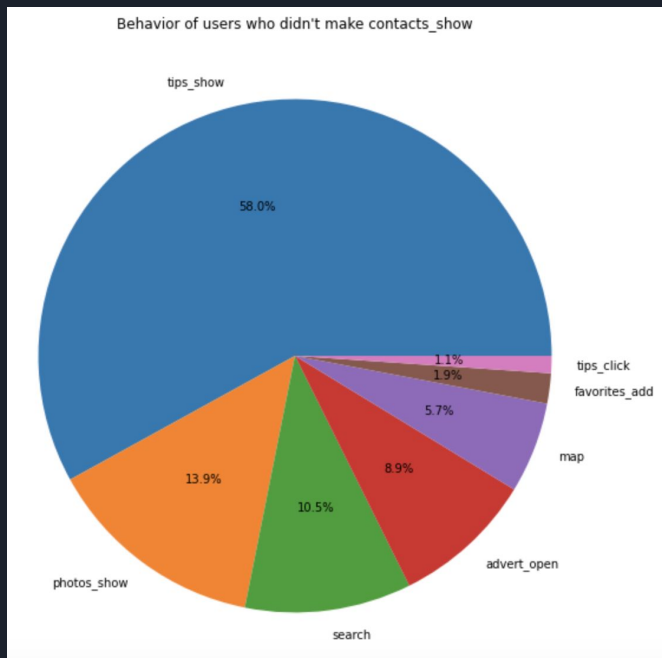19% of initial
592% of previous

The distorted shape of the funnel tells us that there are problems in the processes. Many users enter the funnel, but only a few make it to the next stage. This indicates a marketing error - showing the wrong ads to the wrong audience. There is a need to improve the ad serving system, since only 18% of contacts are viewed.

Data provided by Yandex (IT company)

The time between the first action and the target
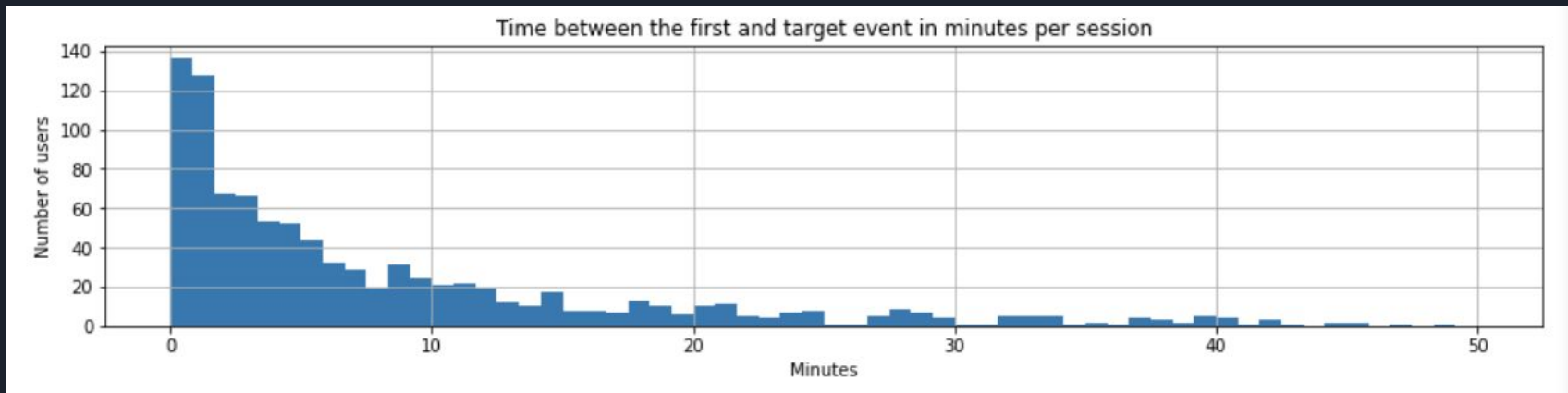
Time between the first and target event in minutes per session

There is a maximum time of 0 days 01:38:05 between the first and target events, an average of 0 days 00:09:21, and a median of 0 days 00:04:49. Based on the histogram, the most values fall on the ~20 minutes.

Data provided by Yandex (IT company)

11

Compare users based on the frequency of their events

Behavior of users who didn't make contacts_show

tips_show 58.0%
tips_click 1.1%
favorites_add 1.9%
map 5.7%
advert_open 8.9%
search 10.5%
photos_show 13.9%

Behavior of users who made contacts_show

tips_show 53.7%
tips_click 1.5%
favorites_add 2.4%
map 4.4%
advert_open 6.5%
search 11.3%
photos_show 20.2%

Compare users based on the frequency of their events

Who completed the target event:

- they like photos_show more - 20.2% versus 13.9% in the second group;
- less like advert_open - 6.5% versus 8.9% in the second group;
- they like search more - 11.3% versus 10.5% in the second group;
- they like map less - 4.4% versus 5.8% in the second group;
- they like favorites_add more - 2.4% versus 1.9% in the second group;
- differences in tips_click and tips_show are not significant.

Compare users based on retention rate
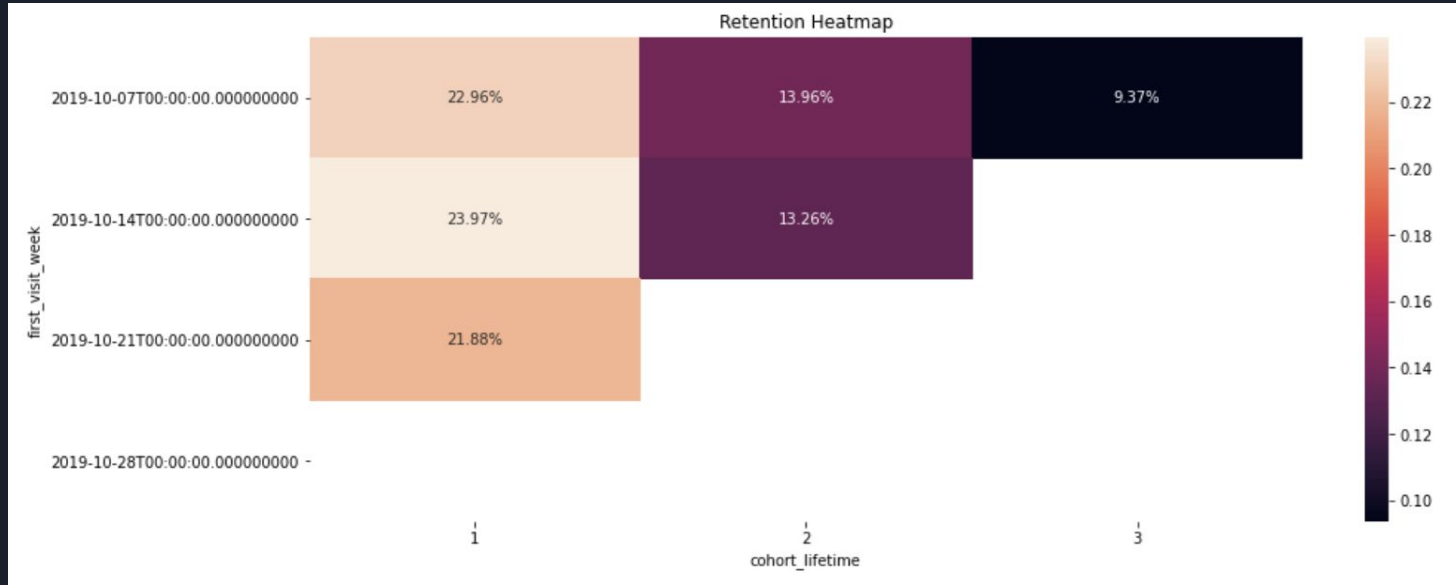
Retention Heatmap

Users who were attracted on October 14 showed the highest retention rates, as indicated by the light spots on the chart.

Data provided by Yandex (IT company)

User segmentation

Four groups were formed after clustering. There are 64 users in the smallest group and 3294 users in the largest group. Clusters have the following characteristics

Group 0

Despite being the second largest group (590 users), this is not the most active group. Users in this group make a small number of events (18) and sessions (4), active for 12 days. The vast majority of users see ads in Yandex recommendations, but seldom open them and do not view photos or contacts. There is a better situation than in group 1, but it is much worse than in groups 2 and 3. Advertisements may be saved to favorites and purchases delayed.

Group 1

This group has the most participants (3294) and is the most passive. There are 10 events in total, 1 session and 1 day of activity. A majority of users come from the recommendation system of all three search engines and rarely search for something themselves. Users open ads occasionally, look at photos a little bit, and rarely make the target event - viewing contacts - themselves.

User segmentation

Group 2

The smallest group of users (64). Users take part in 65 events and 12 sessions during a 12 day period. A lot of users search for ads on their own, look at photos and view contacts. Google and Yandex are the leading sources of traffic. These users clearly understand the product / service they are looking for and therefore most often only use search.

Group 3

It is the third largest (236) and most active group. Users perform more events in this groups (67), and quite a few sessions (5) are active within the past week. Most users open ads in Yandex recommendations, but rarely see photos. They use contacts as a target event. Number of contacts_shows per user is the winner.

User segmentation

Analyzing the time difference between user events from different clusters

- It takes 9:02 minutes for group 2 to reach the target event, while group 3 takes 12:38 minutes. Group leader 3 spends the most time since it performs more actions.
- The time between the most inactive group 1 and the slightly more active group 0 is practically the same. Group 1 spends 9:34 minutes and group 0 spends 8:54 minutes.

Identifying clusters that convert to target events more frequently (conversion to a target event)

- The highest conversion rate was in group 2 (62%).

User segmentation

Identifying clusters that return to the mobile app frequently (Retention rate)

- Group 2 has the highest retention rate, perhaps because users like to search for things on their own.

Testing statistical hypotheses

Hypothesis 1: Conversions to contact views between users who installed the application via a link from yandex and users who installed the application via a link from google have statistically significant differences.

- Total unique google users = 1107
- Total unique google users who made contacts_show = 264
- Conversion in contacts_show for google users = 23.85%
- Total unique yandex users = 1882
- Total unique yandex users who made contacts_show = 467
- Conversion in contacts_show for yandex users = 24.81%

p-value: 0.55

At a given critical level of statistical significance, there are no statistically significant differences between the conversions of users who installed the application via a link from yandex and users who installed the application via a link from google. The source has no effect on the target event's views.

Testing statistical hypotheses

Hypothesis 2: Site search (search) to contact views and recommender (tips_show) to contact views conversions are different

- Total unique 'tips_show' users = 2761
- Total unique 'tips_show' users who made 'contacts_show' = 492
- Conversion to 'contacts_show' for 'tips_show' users = 18.02%
- Total unique 'search' users = 1640
- Total unique 'search' users who made 'contacts_show' = 362
- Conversion to 'contacts_show' for 'search' users = 22.18%

p-value: 0.0008

At a given critical level of statistical significance, there are statistically significant differences between conversion rates of site search (search) to contact views and recommendation system (tips_show) to contact views.

Recommendations

The ad serving system needs to be improved. Users see ads in all search engines, but are not interested in looking at photos or even contacts. Users are not shown what is relevant them. It is necessary to find out what interests them and build algorithms based on this information.

Source

1. Data provided by Yandex (IT company)