

Отчет

Задача

Определить критерии конвертации пользователей, выявить зависимости конвертируемости.

Данные

В качестве исходных данных имелись данные по действиям пользователей за 2 месяца разбитые по дням (01.05-30.06) и все регистрации пользователей за первый месяц (01.05-31.05). Объемы данных 34ГБ и 74МБ соответственно. Примеры данных:

Actions Data Frame						
USER_ID	TIMESTAMP	TYPE	DOCUMENT_TYPE	UPDATE	OTHER_USER_FRIEND	OTHER_USER_ID
683294001	2015-05-25 02:10:00	DOCUMENT_DELETED	PHOTO_ALBUM	null	null	812038158
1392727142	2015-05-25 19:23:20	GUEST	null	null	false	260427154
715014160	2015-05-25 11:53:20	CREATE_DOCUMENT	USER_MESSAGE	false	false	1718441850

Registration Data Frame	
USED_ID	TIMESTAMP
1270248661	2015-05-17 11:40:00
1173602184	2015-05-13 15:43:20
62481419	2015-05-23 20:10:00

Подготовка данных

Данные о регистрации пользователей были разбиты по разным файлам (на каждый день один файл). Далее, для того чтобы определить порог конвертируемости (порог количества действий пользователя в день/неделю, который показывает его конвертацию), понадобилось сгенерировать следующие данные: на каждый день регистрации пользователей смотрится их активность через месяц на протяжении недели (например, для пользователей зарегистрировавшихся 01.05 смотрится активность спустя 24 дня на протяжении недели: 25.05-31.05). В результате получилось 217 файлов, на каждый день регстрации (31) по 7 дней активности. Пример выходного файла:

Выходной файл для 1 дня регистрации (01.05) спустя 24 дня (т.е. 1 дня активности из 7, 25.05)

USER_ID	TIMESTAMP	TYPE	DOCUMENT_TYPE	UPDATE	OTHER_USER_FRIEND	OTHER_USER_ID
1881587348	2015-05-25 13:00:00	CREATE_DOCUMENT	USER_MESSAGE	false	true	1722734546
1881587348	2015-05-25 00:13:20	MARK_PHOTO	null	null	true	1148639005
430661138	2015-05-25 16:36:40	CREATE_DOCUMENT	APP_INVITATION	false	null	812038158

После этого, была создана резульативная таблица, в которой содержалась следующая информация: для каждого сочетания дня регистрации и дня активности были вычислены количество совершенных действий и уникальных пользователей. Полученная таблица имеет следующий вид:

Result Data Frame

reg_day	act_day	uniq_users	events
01-05	25-05	12477	427926
01-05	26-05	12235	423028
01-05	27-05	12109	429909

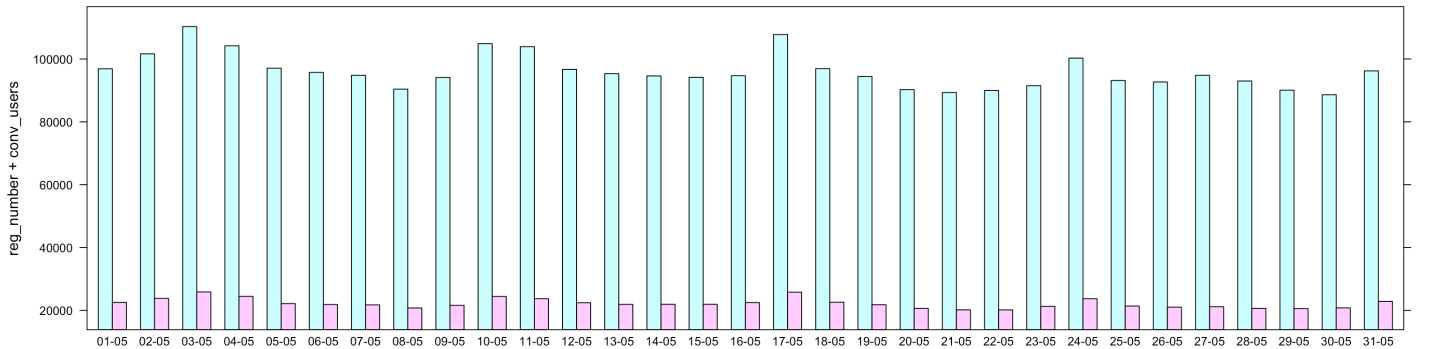
Первичный анализ

Первым шагом был анализ последней недели первого месяца активности пользователей. Были отсечены пользователи, которые никак не сконвертировались, т.е. ни разу не были активны на протяжении последней недели первого месяца. Таким образом, мы посчитали conversion rate как отношение хоть как-то “сконвертированных” пользователей ко всем зарегистрированным для каждого дня регистрации. Так же, были добавлены показатели среднего количества действий пользователя за неделю и медиана этой величины. В результате получилась следующая таблица:

Таблица со сводкой данных за последнюю неделю первого месяца активности

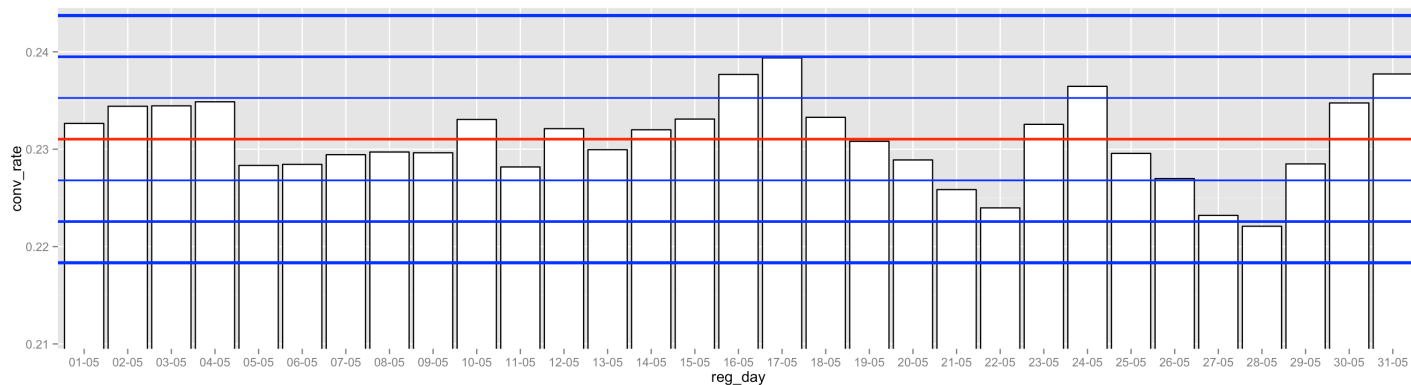
reg_day	reg_number	conv_users	mean_actions	actions_median	conv_rate
01-05	99614	23056	128.6734	25	0.2314534
02-05	104573	24343	120.5533	25	0.2327848
03-05	115553	26735	114.3660	26	0.2313657
04-05	107516	25006	116.0208	24	0.2325793
05-05	100640	22707	123.7068	25	0.2256260

Гистограмма, описывающая разницу между зарегистрировавшимися пользователями и оставшимися через месяц



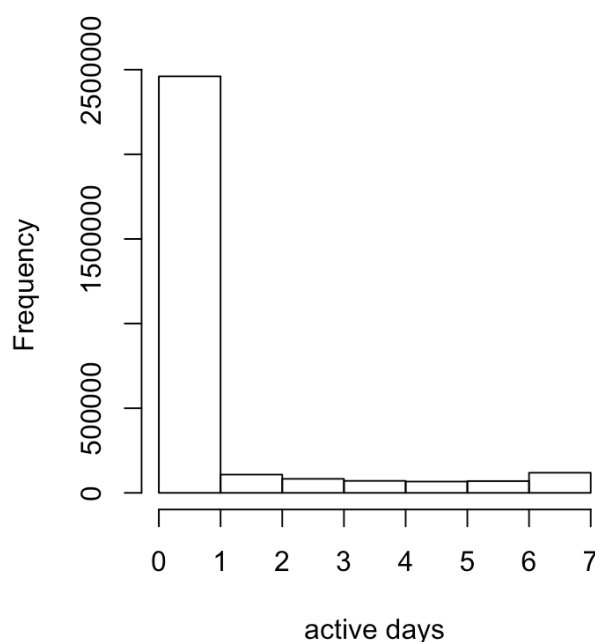
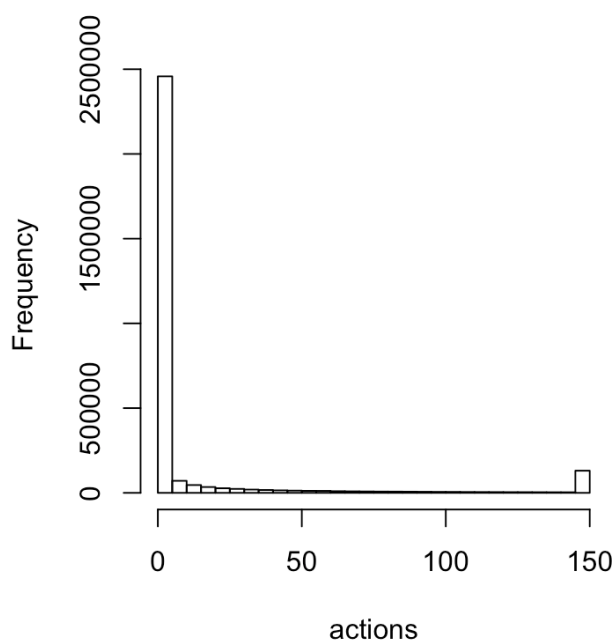
Чтобы выявить аномальные дни, был проведен статистический анализ отклонений conversion rate. Было рассчитано среднее значение и стандартное отклонение conversion rate. Гипотеза состоит в том, что если conversion rate одного из дней лежит вне пределах трех сигм ($3 \cdot \sigma$), такой день считается аномальным и подлежит последующему анализу. Аномальных дней выявлено не было. Для анализа был построен график с этими показателями (красная линия — среднее значение, синие линии - стандартное отклонение, сигмы):

Гистограмма, показывающая отклонения от среднего показателя conversion rate



Далее была произведена агрегация данных. Были собраны данные активности на протяжении последней недели первого месяца для каждого пользователя. На основе этих данных была составлена гистограмма распределения количества действий по количеству пользователей и аналогичная гистограмма для количества активных дней (150 трактовать как 150 и больше):

Гистограмма распределения количества действий и активных дней по количеству пользователей



Было принято решение выделить 5 групп конвертации: unconverted, inactive, rare, average и active. Для каждой группы конвертации нужно было выбрать порог. Для этого анализировались основные описательные статистики данных, такие как: медиана, первый и третий квантили и тд. Поскольку распределение по дням активности дало мало конкретной информации, пороги выбирались на основе данных количества действий (150 трактовать как 150 и больше):

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	1.0	6.0	25.0	121.4	99.0	40400.0

После анализа было принято окончательное решение выбора значений порогов: для группы rare был выбран порог который соответствует первому квантилю выборки — 6. Для группы average был выбран порог соответствующий медиане выборки — 26. И для группы active был выбран порог соответствующий третьему квантилю выборки — 100 действий в неделю. Естественным образом для группы inactive был выбран порог до 6 действий. Результатом первичного анализа является следующая таблица:

Результативная таблица описывающая группы конвертации пользователей и их активность

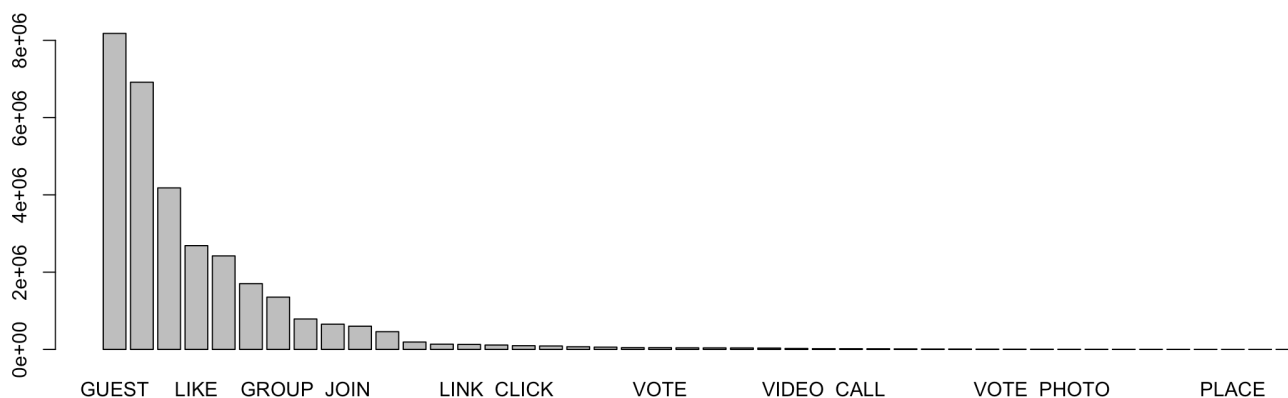
id	day_1	day_2	day_3	day_4	day_5	day_6	day_7	active_day	all group
----	-------	-------	-------	-------	-------	-------	-------	------------	-----------

964322704	1	1	5	7	41	13	4	7	72	average
599236160	0	0	0	0	1	0	0	1	1	inactive
1313261972	1	0	0	0	0	0	0	1	1	inactive
1179320426	0	5	2	13	0	20	6	5	46	average
1150663505	4	42	36	9	26	7	20	7	144	active
1686861514	0	6	20	5	6	0	32	5	69	average

actions	group
0	unconverted
1-6	inactive
7-26	rare
27-100	average
100+	active

Первой попыткой было предсказание конвертации пользователя по первым двум часам активности после регистрации. Нужно было выделить факторы (типы действий пользователя), по которым будет прогнозироваться конвертация. Для этого было взято в рассмотрение распределение факторов по количеству пользователей. Выглядело оно следующим образом:

Распределение факторов (типов действий) по количеству пользователей



В рассмотрение были взяты 12 наиболее встречающихся факторов, выглядит этот список так:

```
## [1] "GUEST"          "PHOTO_TITLE"    "USER_MESSAGE"
## [4] "LIKE"           "FRIEND_INVITE"  "PHOTO_ALBUM"
## [7] "MARK_PHOTO"     "GROUP_JOIN"     "DOCUMENT_DELETED"
## [10] "USER_STATUS"    "USER_PERSONAL_DATA" "GROUP_PHOTO_TITLE"
```

Далее все действия за первые два часа после регистрации были агрегированы для каждого пользователя по выделенным факторам. Каждый пользователь был помечен группой конвертации в соответствии с полученными ранее данными. В результате был получен следующий набор данных пригодный для построения модели:

Результативный набор данных для построения модели

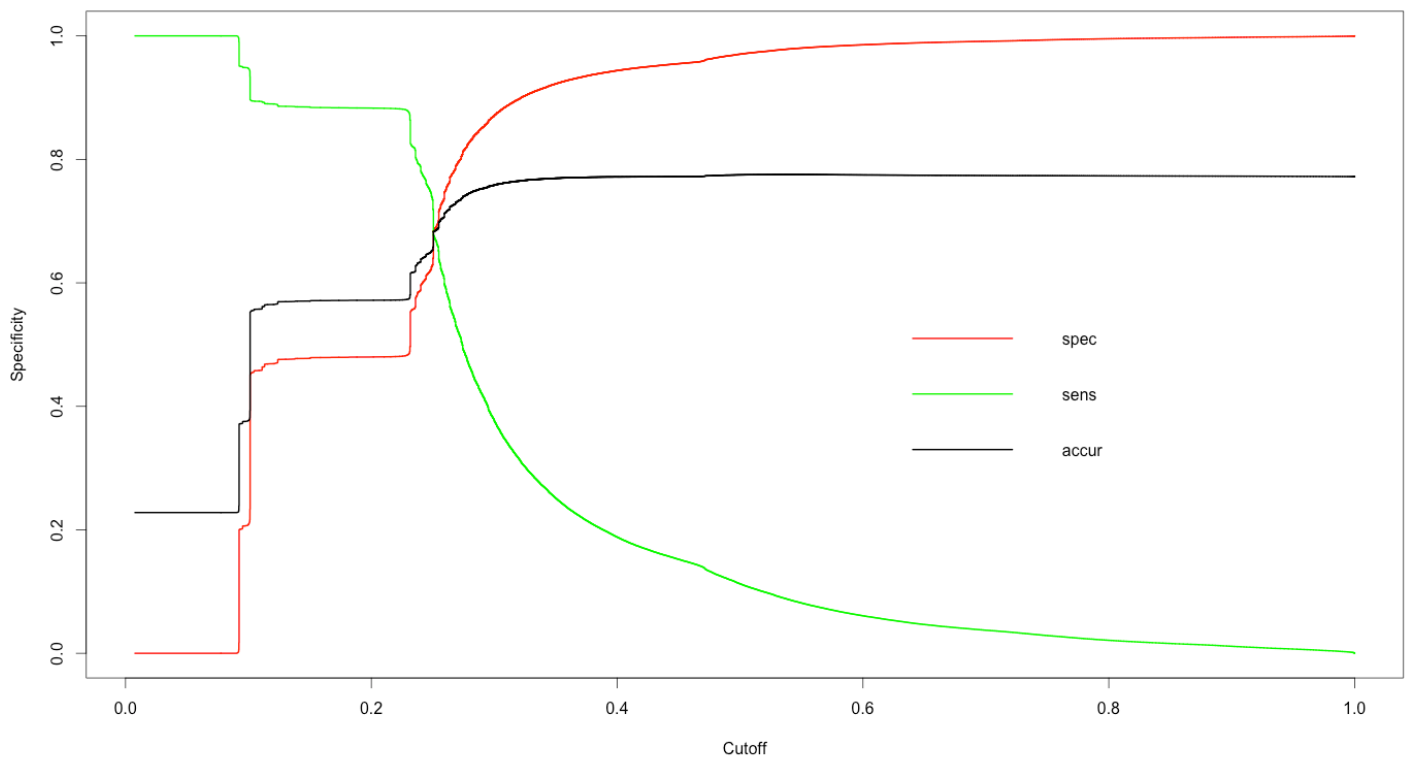
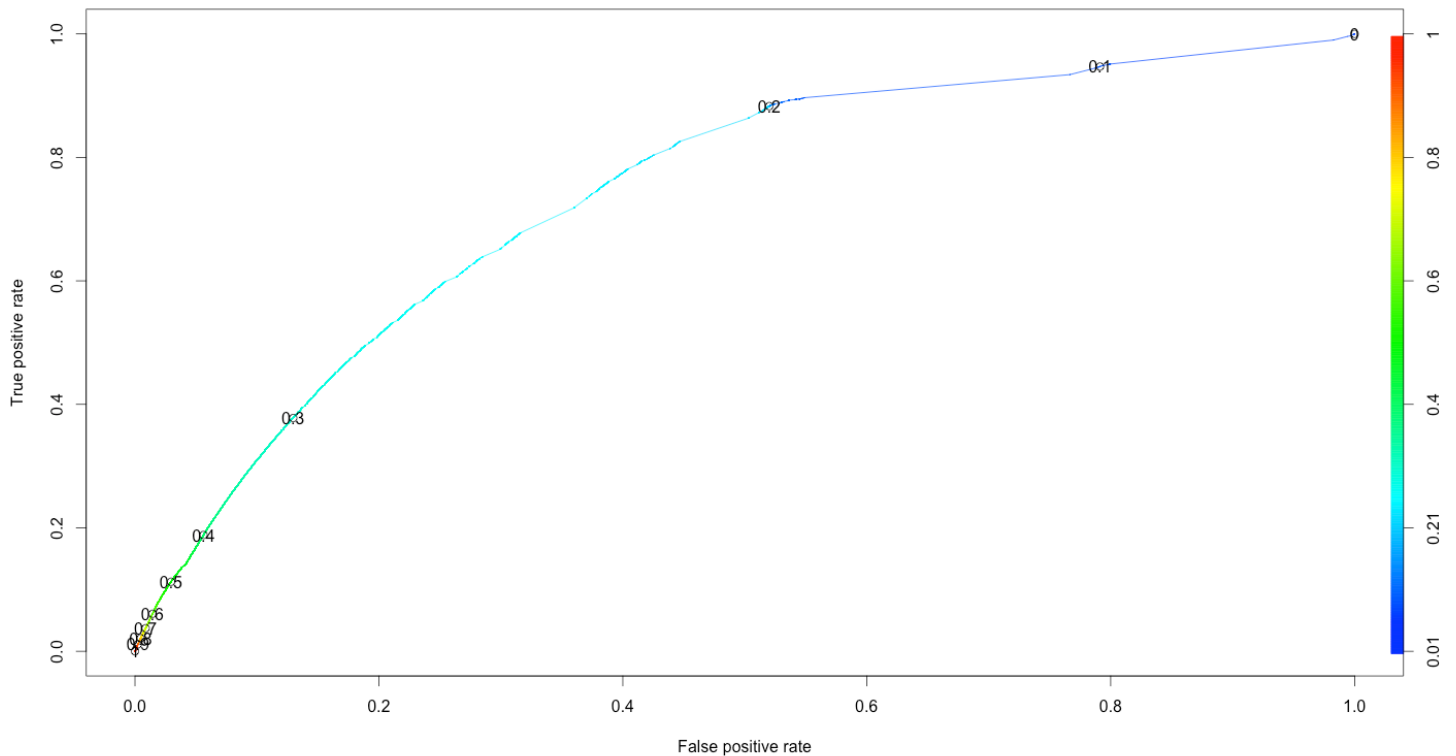
USER_ID	GUEST	LINK_CLICK	PHOTO_ALBUM	PHOTO_TITLE		USER_MESSAGE	GROUP
394580146	0	0	1	0		0	inactive

1754004182	1	1	1	4		1	unconverted
1045482474	1	0	1	0		0	average
1047107302	0	1	1	0		0	unconverted

Модель

Для анализа была выбрана линейная статистическая модель — логистическая регрессия. Данные были перемешаны и разделены на две части, 2/3 и 1/3, обучающий и тестовый набор данных соответственно. Модель была построена на обучающем наборе данных по всем факторам без учета их взаимодействия. С помощью модели была сделана попытка предсказания группы конвертации для тестового набора данных. Показатели результата применения модели:

AUC, specificity, sensitivity, accuracy



Порогом вероятности положительного срабатывания классификатора была выбрана точка пересечения кривых specificity, sensitivity и accuracy (0.25) как самая оптимальная. Поскольку показатели AUC и Accuracy не дают нам исчерпывающий ответ о качестве модели (в виду особенностей данных), в качестве целевой функции были выбраны показатели precision, recall и F1-мера:

```

test/train
precision — 0.387/0.386
recall — 0.677/0.678
f1 — 0.492/0.492

```

Далее модель была применена на обучающем наборе данных. Графики AUC, specificity, sensitivity и accuracy как и показатели precision, recall и F1-мера не имеют значимых отличий от результатов применения на тестовом наборе данных, что говорит о возможном недообучении модели. Следующим шагом было усложнение модели, были собраны, сагрегированны и добавлены данные о действиях пользователей за следующие два часа (2-4 час) после регистрации. На обновленных данных точно так же была применена логистическая регрессия. Результаты применения:

```

test/train
precision — 0.403/0.402
recall — 0.660/0.660
f1 — 0.500/0.500

```

Модель дала пренебрежимо малый прирост производительности по сравнению с предыдущей. Для дальнейшего анализа было решено воспользоваться случайным лесом (random forest). Были построены разные вариации моделей: 150000 семплов с 12 факторами (первые 2 часа после регистрации), 300000 семплов с 12 и 24 факторами (первые 2 и последующие 2 часа после регистрации). Результаты применения случайного леса:

	150k, 12 test/train	300k, 12 test/train	300k, 12+12 test/train
precision —	0.541/0.529	0.540/0.531	0.570/0.565
recall —	0.176/0.172	0.186/0.185	0.223/0.223
f1 —	0.266/0.260	0.277/0.274	0.321/0.320

Можно заметить, что показатели F1-меры заметно ниже, чем у линейной модели. Далее, для улучшения результатов было решено к случайному лесу применить регуляризацию. Были так же построены разнообразные модели с различными показателями коэффициентов регуляризации и количества факторов. Результаты применения регуляризации:

	150k, 12, c = 1 test/train	150k, 12+12, c = 1 test/train	150k, 12+12, c = 0.0001 test/train
precision —	0.484/0.473	0.501/0.497	0.501/0.496
recall —	0.246/0.241	0.291/0.294	0.291/0.293
f1 —	0.326/0.319	0.368/0.369	0.368/0.368

Регуляризация дала заметный прирост производительности, но, тем не менее, показатель F1-меры остается низким. Показатель precision остается на более менее приемлемом уровне, а recall, как с регуляризацией, так и без нее, находится на очень низком уровне. Это говорит нам о том, что модель имеет высокий false negative rate. В связи с этим, было принято решение вручную сдвинуть порог принятия решения на более низкое значение для улучшения показателя recall, и, соответственно F1-меры. Эмпирическим путем было подобрано оптимальное значение порога — 0.001. После введения нового порога показатели recall и F1-мера заметно улучшились, а precision упал в допустимых пределах:

```

regularized random forest
300k, 12+12, c = 1, p = 0.001
test/train
precision — 0.425/0.425
recall — 0.665/0.657
f1 — 0.519/0.516

```

Основные показатели выше, чем у линейной модели, что позволяет нам проводить дальнейший анализ. Далее, после длительных экспериментов, было принято решение выделить 14 наиболее важных факторов, на основе которых будет проводиться анализ. По этим факторам была взята информация за первые два и последующие два часа пребывания пользователей на портале. Также в качестве дополнительного фактора был добавлен пол пользователя. По этим факторам была построена линейная и нелинейная модели (логистическая регрессия и случайный лес деревьев соответственно). Итоговый набор факторов:

```
## [1] "PHOTO_ALBUM"      "GUEST"             "USER_PERSONAL_DATA"
## [4] "GENDER"           "FRIEND_INVITE"     "APP_INVITATION"
## [7] "GROUP_JOIN"       "USER_STATUS"       "USER_MESSAGE"
## [10] "LIKE"             "SEND_PRESENT"      "MARK_PHOTO"
## [13] "PHOTO_TITLE"      "DOCUMENT_DELETED"  "RE_SHARE"
```

Итоговая таблица моделей

	regularized random forest	
	300k train set, 15 features	300k train set, 29 features
	test/train	test/train
precision	0.433/0.441	0.433/0.441
recall	0.611/0.620	0.671/0.680
f1	0.507/0.515	0.526/0.535
	logistic regression	
	300k train set, 15 features	300k train set, 29 features
	test/train	test/train
precision	0.373/0.374	0.385/0.382
recall	0.719/0.723	0.707/0.718
f1	0.491/0.493	0.498/0.499

После построения моделей можно приступить к изучению значимости переменных (факторов). Будем опираться на показатель z-value линейной модели и на variable importance случайного леса. Рассмотрим модель с 29 факторами (14 факторов за первые 2 часа активности + 14 за последующие 2 часа + пол пользователя):

```
## [1] "random forest variables importance"
```

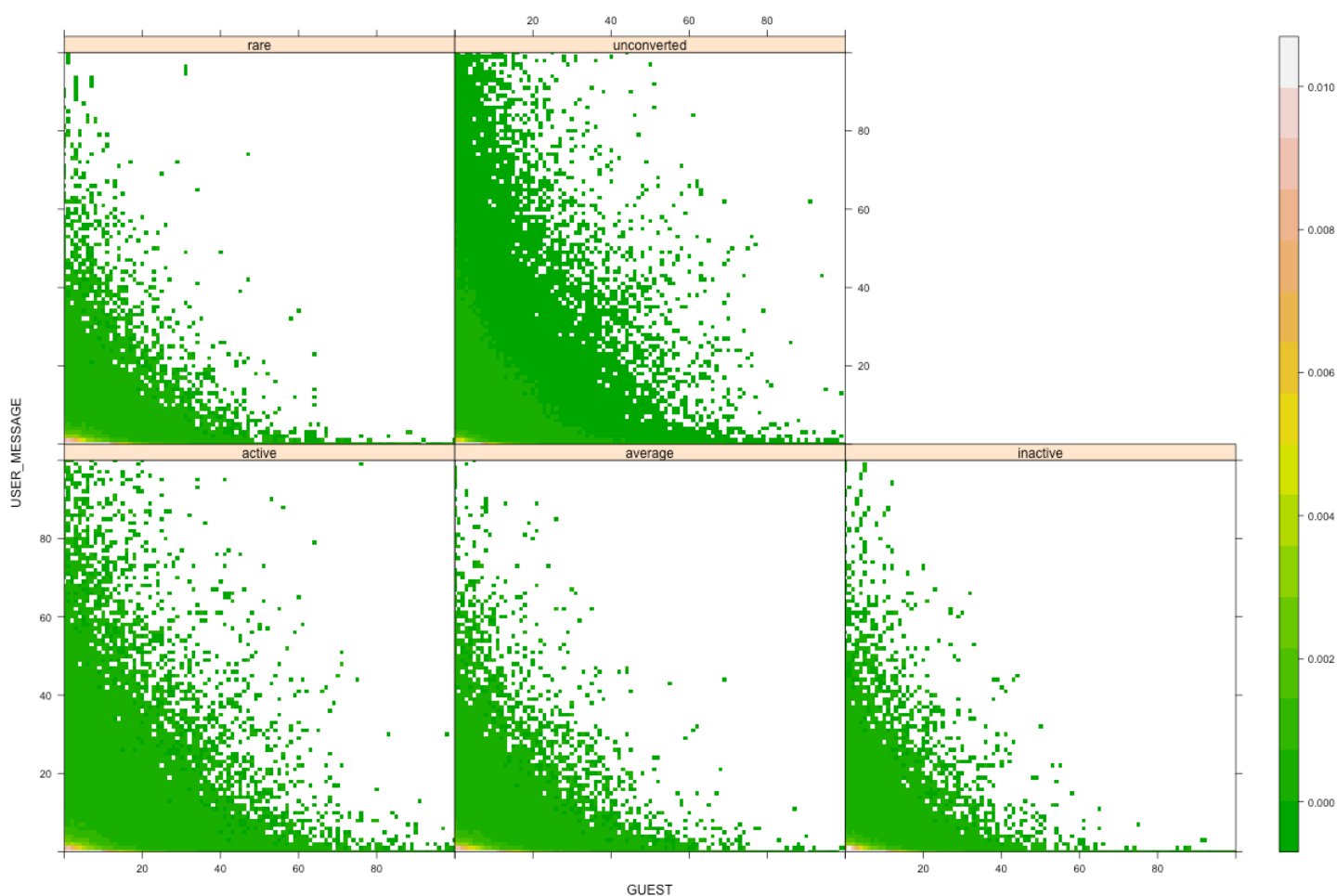
```
##          PHOTO_TITLE          GUEST          PHOTO_ALBUM
##          11684             8528             5430
##          USER_MESSAGE      FRIEND_INVITE          LIKE
##          4127             3548             3337
##          GUEST_2          USER_STATUS  USER_PERSONAL_DATA
##          2995             2269             2030
##          MARK_PHOTO      GROUP_JOIN      DOCUMENT_DELETED
##          2015             1975             1906
##          USER_MESSAGE_2          GENDER          LIKE_2
##          1802             1255             1139
##          PHOTO_TITLE_2      FRIEND_INVITE_2          USER_STATUS_2
##          1106             730             660
##          PHOTO_ALBUM_2      GROUP_JOIN_2          MARK_PHOTO_2
##          623             567             483
##          DOCUMENT_DELETED_2  USER_PERSONAL_DATA_2          SEND_PRESENT
##          468             455             429
##          APP_INVITATION          RE_SHARE      APP_INVITATION_2
##          399             313             161
##          SEND_PRESENT_2          RE_SHARE_2
##          150             60
```

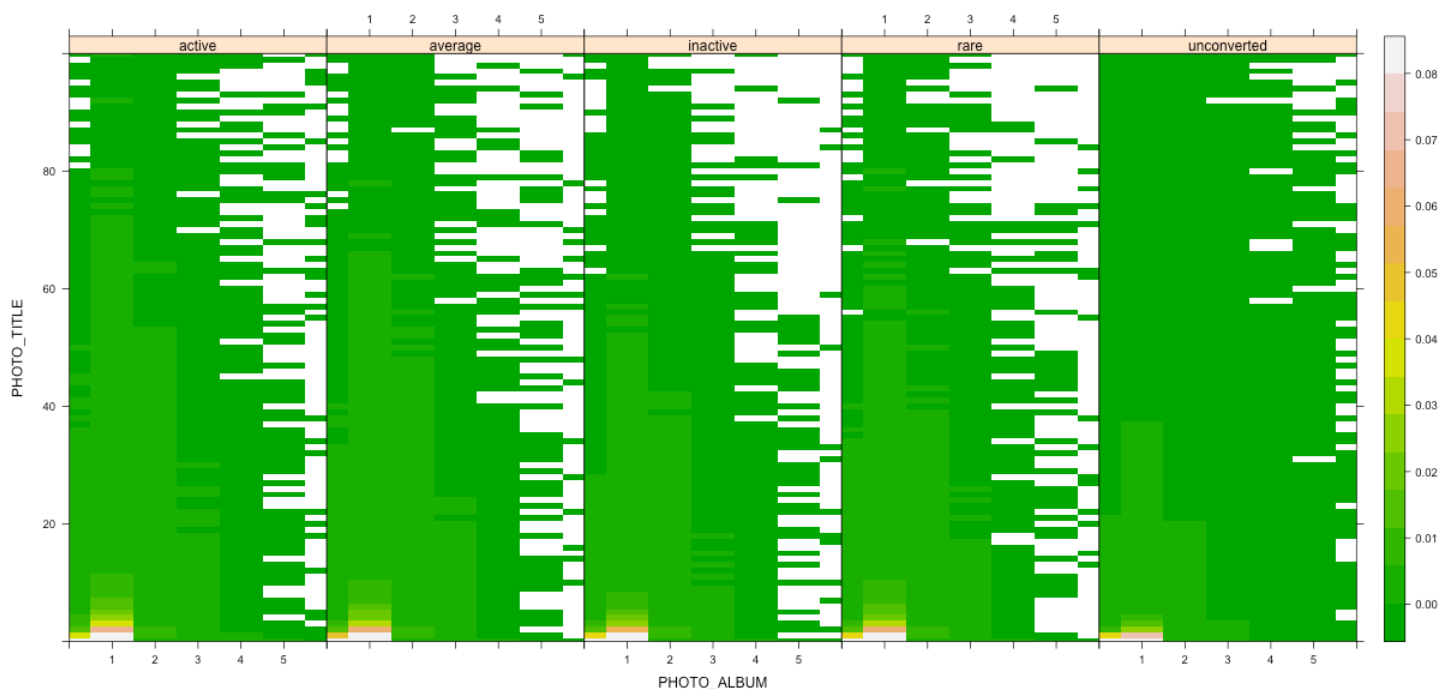
```
## [1] "logistic regression variables importance"
```

##	PHOTO_ALBUM	GUEST	USER_PERSONAL_DATA
##	100	39	29
##	GENDER2	FRIEND_INVITE	PHOTO_ALBUM_2
##	29	24	23
##	GUEST_2	GROUP_JOIN	USER_STATUS
##	17	14	12
##	APP_INVITATION	USER_PERSONAL_DATA_2	LIKE_2
##	11	11	10
##	FRIEND_INVITE_2	LIKE	GROUP_JOIN_2
##	9	8	8
##	USER_MESSAGE	APP_INVITATION_2	USER_STATUS_2
##	8	7	6
##	SEND_PRESENT	PHOTO_TITLE	SEND_PRESENT_2
##	5	5	4
##	DOCUMENT_DELETED	MARK_PHOTO	DOCUMENT_DELETED_2
##	3	3	1
##	USER_MESSAGE_2	MARK_PHOTO_2	PHOTO_TITLE_2
##	0	0	-2
##	RE_SHARE_2	RE_SHARE	
##	-5	-7	

Логистическая регрессия дает представление об общей значимости факторов, линейной, она не учитывает их взаимодействие. Случайный лес дает представление о значимости с учетом взаимодействий факторов между собой. Сравнивая значимость факторов линейной и нелинейной модели, можно заключить, что факторы PHOTO_TITLE, USER_MESSAGE, LIKE и MARK_PHOTO не имеют особой значимости по отдельности, но нелинейная модель говорит о том, что эти факторы становятся значимыми вступая во взаимодействие с другими. После анализа взаимодействий этих факторов с другими, были выявлены следующие зависимости, представленные на графиках ниже:

Выявленные зависимости переменных





Анализируя эти зависимости были сформированы две новых переменных (фактора):

$PHOTO_PER_ALBUM = PHOTO_TITLE / PHOTO_ALBUM$ (количество фото на альбом)

$USR_MSG_AND_GUEST = USER_MESSAGE / (4 * GUEST + 70)$ (сочетание сообщений пользователя и его походов в гости)

Также была выдвинута гипотеза о том, что фактор $USER_MESSAGE$ станет значимым, если смотреть среднее количество сообщений уникальному пользователю. Для этого были собраны данные и создан фактор $USR_MSG_PER_UNIQ$.

После добавления данных факторов к общему списку, значимость переменных в линейной модели выглядела следующим образом:

##	PHOTO_ALBUM	GUEST	GENDER2
##	97	40	29
##	USER_PERSONAL_DATA	FRIEND_INVITE	PHOTO_ALBUM_2
##	28	23	23
##	GUEST_2	GROUP_JOIN	PHOTO_PER_ALBUM
##	17	14	14
##	USR_MSG_PER_UNIQ	USER_STATUS	APP_INVITATION
##	12	12	11
##	USR_MSG_AND_GUEST	USER_PERSONAL_DATA_2	LIKE_2
##	11	11	9
##	FRIEND_INVITE_2	LIKE	GROUP_JOIN_2
##	9	8	8
##	APP_INVITATION_2	MSG_UNIQ_USERS	USER_STATUS_2
##	7	7	6
##	SEND_PRESENT	SEND_PRESENT_2	MARK_PHOTO
##	4	4	3
##	DOCUMENT_DELETED	USER_MESSAGE_2	DOCUMENT_DELETED_2
##	1	1	0
##	PHOTO_TITLE_2	MARK_PHOTO_2	RE_SHARE_2
##	0	0	-5
##	RE_SHARE	PHOTO_TITLE	USER_MESSAGE
##	-7	-9	-14

Можно сделать вывод, что более сложные факторы (их взаимодействия) работают лучше, чем их аналоги без взаимодействия. Допустим, фактор $PHOTO_TITLE$ стал, как мы видим, отрицательным, а новый фактор $PHOTO_PER_ALBUM$ стал значимее, чем был старый, что говорит нам о том, что добавленный фактор работает лучше.

Это значит, что пользователя, который загружает много фотографий, но при этом не создавал ни одного фотоальбома, модель будет воспринимать как пользователя с негативным поведением. Если же пользователь будет загружать фотографии непосредственно в альбомы, которых больше одного (дефолтного), то такое поведение будет распознаваться моделью, как поведение ведущее к конверсии. Аналогичная ситуация с фактором USER_MESSAGE. Добавление трех более сложных факторов, включающих в себя этот фактор работает лучше, чем сам фактор отдельно. Мы видим, что значимость фактора USER_MESSAGE отрицательна, в то время как факторы USR_MSG_PER_UNIQ, USR_MSG_AND_GUEST и MSG_UNIQ_USERS имеют довольно высокую значимость. Из этого следует, что пользователь, пишущий много сообщений, но при этом мало ходящий в гости вероятнее всего не сконвертируется, и то, что пользователь, пишущий много сообщений, но при этом разным пользователям конвертируется намного лучше, чем пользователь, который пишет только одному-двум другим пользователям. Данные преобразования модели подняли ее значение F1 не значимо, меньше, чем на пол процента, и довели это значение до 0.5.

Далее, была выдвинута гипотеза о том, что фидбэк пользователя (действия по отношению к нему) значимо влияет на конвертацию этого пользователя. Для проверки этой гипотезы были собраны данные по этим действиям от 2 часов после регистрации до 4 дней. Список новых факторов:

```
## [1] "INC.FRIEND_INVITE"          "INC.GUEST"
## [3] "INC.LIKE.USER_PHOTO"       "INC.MARK_PHOTO"
## [5] "INC.SEND_PRESENT"          "INC.USER_MESSAGE"
## [7] "INC.LIKE.PHOTO_COMMENT"    "INC.PHOTO_COMMENT"
## [9] "INC.LIKE.USER_STATUS"      "INC.RE_SHARE"
## [11] "INC.LIKE.USER_STATUS_COMMENT" "INC.USER_STATUS_COMMENT"
## [13] "INC.LIKE.PHOTO_ALBUM"      "INC.USER_PHOTOPIN_INFO"
## [15] "INC.LIKE.GROUP_PHOTO"      "INC.LIKE.GROUP_PHOTO_COMMENT"
## [17] "INC.PHOTO_ALBUM_COMMENT"
```

На этих факторах была построена сначала линейная, затем нелинейная (случайный лес) модель. Результаты значимости факторов были следующими:

```
## [1] "logistic regression variables importance"
```

```
##          INC.SEND_PRESENT          INC.LIKE.USER_PHOTO
##                53                35
##          INC.GUEST          INC.USER_MESSAGE
##                35                14
##    INC.LIKE.PHOTO_ALBUM INC.LIKE.USER_STATUS_COMMENT
##                4                3
##    INC.PHOTO_COMMENT INC.LIKE.GROUP_PHOTO_COMMENT
##                3                1
##    INC.LIKE.PHOTO_COMMENT          INC.MARK_PHOTO
##                1                1
##    INC.USER_PHOTOPIN_INFO    INC.PHOTO_ALBUM_COMMENT
##                1                1
##    INC.LIKE.GROUP_PHOTO          INC.RE_SHARE
##                1                0
##    INC.LIKE.USER_STATUS          INC.FRIEND_INVITE
##                0                -3
##    INC.USER_STATUS_COMMENT
##                -6
```

```
## [1] "random forest variables importance"
```

```

##          INC.GUEST          INC.USER_MESSAGE
##          27008              7257
##          INC.MARK_PHOTO      INC.LIKE.USER_PHOTO
##          4255                3313
##          INC.FRIEND_INVITE    INC.PHOTO_COMMENT
##          1780                1214
##          INC.SEND_PRESENT     INC.LIKE.USER_STATUS
##          1199                1083
##          INC.USER_STATUS_COMMENT INC.USER_PHOTOPIN_INFO
##          597                 312
##          INC.LIKE.PHOTO_COMMENT INC.RE_SHARE
##          302                 256
##          INC.LIKE.USER_STATUS_COMMENT INC.LIKE.PHOTO_ALBUM
##          241                 124
##          INC.LIKE.GROUP_PHOTO INC.LIKE.GROUP_PHOTO_COMMENT
##          99                  28
##          INC.PHOTO_ALBUM_COMMENT
##          13

```

Можно заметить, что значимость у моделей выглядит по-разному. Линейная модель отображает прямое влияние фактора на конвертацию пользователя, а случайный лес отображает влияние с учетом разнообразных взаимодействий факторов. Можно выделить самые значимые переменные по этим данным: INC.SEND_PRESENT, INC.LIKE.USER_PHOTO, INC.GUEST, INC.USER_MESSAGE, INC.MARK_PHOTO, INC.FRIEND_INVITE.

Далее, все факторы фидбэка были добавлены к остальным факторам основных данных. На новых, объединенных данных снова были построены линейная модель и случайный лес. Показатели моделей:

Показатели моделей

	logistic regression	random forest
	300k train set, 50 features	300k train set, 50 features
	test/train	test/train
precision	0.393/0.394	0.433/0.441
recall	0.737/0.738	0.821/0.832
f1	0.513/0.514	0.567/0.574

Сразу можно заметить, что показатели нелинейной модели заметно выше, чем у линейной. Дальше, можно посмотреть на общую картину значимости факторов обеих моделей. Выглядела она так:

```
## [1] "logistic regression variables importance"
```

```

##          PHOTO_ALBUM          INC.SEND_PRESENT
##          92                  40
##          GUEST              INC.LIKE.USER_PHOTO
##          34                  27
##          USER_PERSONAL_DATA    PHOTO_ALBUM_2
##          24                  20
##          INC.USER_MESSAGE      GENDER2
##          18                  18
##          FRIEND_INVITE        GUEST_2
##          17                  13
##          USR_MSG_PER_UNIQ      APP_INVITATION
##          13                  11
##          GROUP_JOIN           USER_STATUS
##          11                  11
##          PHOTO_PER_ALBUM      INC.GUEST
##          11                  11
##          MSG_UNIQ_USERS       USR_MSG_AND_GUEST
##          9                    8
##          USER_PERSONAL_DATA_2  APP_INVITATION_2
##          8                    6
##          FRIEND_INVITE_2      GROUP_JOIN_2
##          6                    6
##          LIKE_2              LIKE
##          5                    5
##          USER_STATUS_2       MARK_PHOTO
##          5                    2
## INC.LIKE.USER_STATUS_COMMENT  SEND_PRESENT
##          2                    1
##          INC.LIKE.GROUP_PHOTO INC.LIKE.GROUP_PHOTO_COMMENT
##          1                    1
##          DOCUMENT_DELETED     INC.LIKE.PHOTO_COMMENT
##          1                    0
##          INC.PHOTO_COMMENT     INC.USER_PHOTOPIN_INFO
##          0                    0
##          DOCUMENT_DELETED_2   INC.PHOTO_ALBUM_COMMENT
##          0                    0
##          INC.RE_SHARE         INC.LIKE.USER_STATUS
##          0                    0
##          MARK_PHOTO_2        PHOTO_TITLE_2
##          -1                  -1
##          INC.LIKE.PHOTO_ALBUM  INC.MARK_PHOTO
##          -1                  -1
##          SEND_PRESENT_2      RE_SHARE_2
##          -2                  -4
##          INC.USER_STATUS_COMMENT INC.FRIEND_INVITE
##          -5                  -6
##          RE_SHARE            PHOTO_TITLE
##          -7                  -9
##          USER_MESSAGE_2      USER_MESSAGE
##          -11                 -14

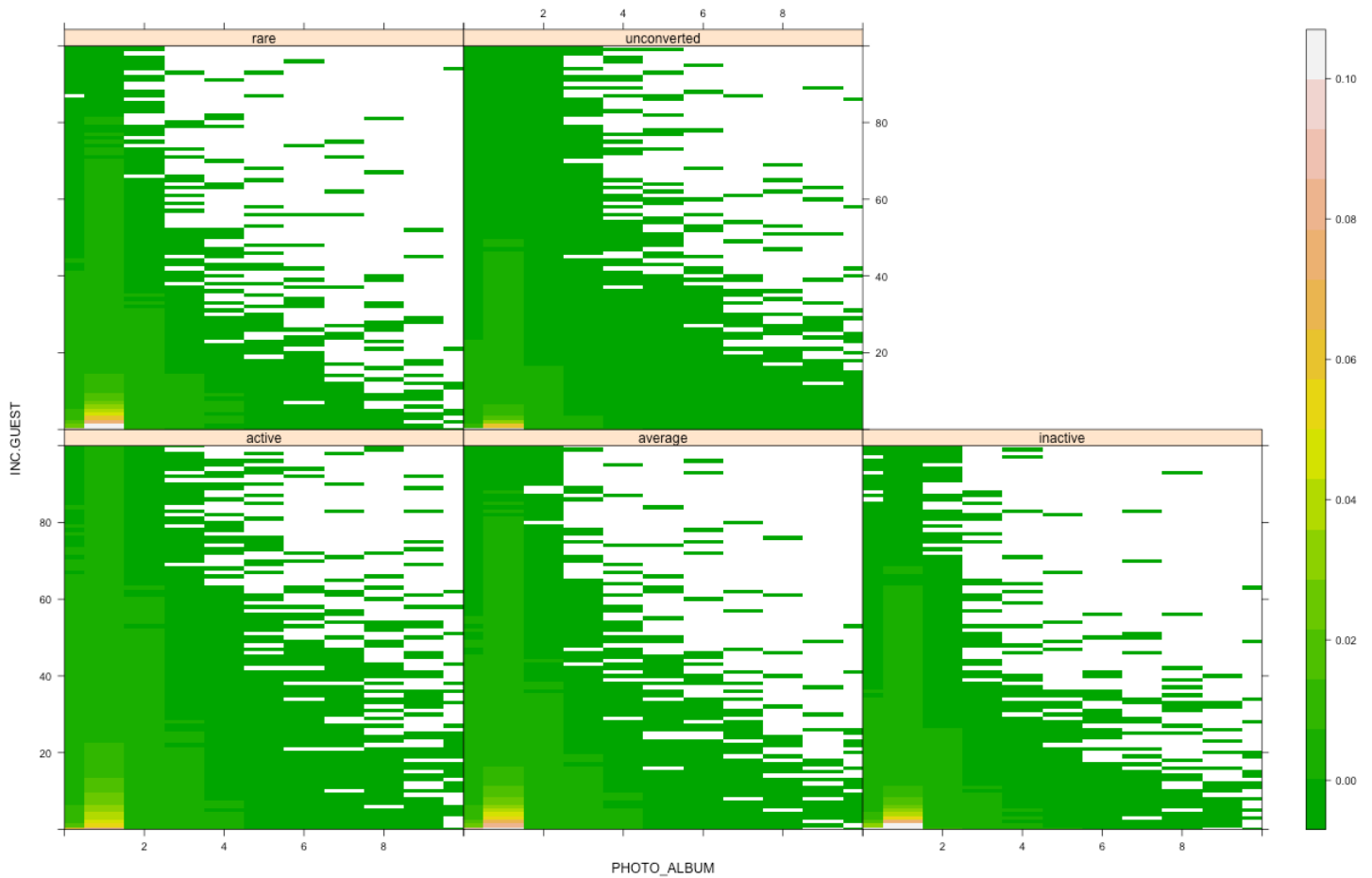
```

```
## [1] "random forest variables importance"
```

##	INC.GUEST	GUEST
##	25305	6175
##	INC.USER_MESSAGE	FRIEND_INVITE
##	4812	2900
##	PHOTO_TITLE	INC.MARK_PHOTO
##	2850	2656
##	PHOTO_ALBUM	LIKE
##	2478	2424
##	PHOTO_PER_ALBUM	INC.LIKE.USER_PHOTO
##	2309	2296
##	USR_MSG_AND_GUEST	GUEST_2
##	2097	2088
##	USER_PERSONAL_DATA	USER_STATUS
##	1709	1650
##	GENDER	GROUP_JOIN
##	1621	1468
##	MARK_PHOTO	DOCUMENT_DELETED
##	1355	1281
##	INC.FRIEND_INVITE	USR_MSG_PER_UNIQ
##	1215	1172
##	USER_MESSAGE_2	INC.SEND_PRESENT
##	1017	966
##	USER_MESSAGE	LIKE_2
##	878	709
##	MSG_UNIQ_USERS	PHOTO_TITLE_2
##	702	657
##	INC.PHOTO_COMMENT	INC.LIKE.USER_STATUS
##	617	556
##	FRIEND_INVITE_2	USER_STATUS_2
##	471	462
##	GROUP_JOIN_2	APP_INVITATION
##	359	353
##	INC.USER_STATUS_COMMENT	USER_PERSONAL_DATA_2
##	351	320
##	MARK_PHOTO_2	DOCUMENT_DELETED_2
##	316	292
##	SEND_PRESENT	PHOTO_ALBUM_2
##	282	244
##	RE_SHARE	INC.USER_PHOTOPIN_INFO
##	211	172
##	INC.LIKE.PHOTO_COMMENT	INC.RE_SHARE
##	144	135
##	APP_INVITATION_2	INC.LIKE.USER_STATUS_COMMENT
##	123	115
##	SEND_PRESENT_2	INC.LIKE.PHOTO_ALBUM
##	89	52
##	INC.LIKE.GROUP_PHOTO	RE_SHARE_2
##	46	39
##	INC.LIKE.GROUP_PHOTO_COMMENT	INC.PHOTO_ALBUM_COMMENT
##	12	6

Можно сделать вывод, что фидбэк пользователя значимо влияет на его конвертацию. По данным нелинейной модели был проведен анализ по выявлению нелинейных связей (взаимодействий факторов). Была выявлена единственная значимая связь:

Выявленные зависимости переменных



Был создан новый фактор, который описывается следующей формулой: $INC.GUEST_PER_ALBUM = INC.GUEST / PHOTO_ALBUM$, и трактуется как количество пришедших гостей на каждый фотоальбом. Можно заметить, что с каждой следующей группой конвертации количество гостей на каждый фотоальбом возрастает. После добавление этой переменной в линейную модель значимость факторов стала выглядеть следующим образом:

```

##          PHOTO_ALBUM          INC.SEND_PRESENT
##          93                  40
##          GUEST              INC.LIKE.USER_PHOTO
##          33                  27
##          USER_PERSONAL_DATA  PHOTO_ALBUM_2
##          24                  19
##          INC.GUEST          GENDER2
##          18                  18
##          INC.USER_MESSAGE    FRIEND_INVITE
##          18                  17
##          PHOTO_PER_ALBUM     GUEST_2
##          13                  13
##          USR_MSG_PER_UNIQ     APP_INVITATION
##          13                  11
##          USER_STATUS        GROUP_JOIN
##          11                  10
##          MSG_UNIQ_USERS      USR_MSG_AND_GUEST
##          9                   8
##          USER_PERSONAL_DATA_2 APP_INVITATION_2
##          7                   6
##          FRIEND_INVITE_2     GROUP_JOIN_2
##          6                   6
##          LIKE_2              LIKE
##          5                   5
##          USER_STATUS_2       MARK_PHOTO
##          5                   2
## INC.LIKE.USER_STATUS_COMMENT SEND_PRESENT
##          2                   1
## INC.LIKE.GROUP_PHOTO_COMMENT INC.LIKE.GROUP_PHOTO
##          1                   1
##          DOCUMENT_DELETED    INC.LIKE.PHOTO_COMMENT
##          0                   0
##          INC.PHOTO_ALBUM_COMMENT INC.PHOTO_COMMENT
##          0                   0
##          DOCUMENT_DELETED_2  INC.USER_PHOTOPIN_INFO
##          0                   0
##          INC.RE_SHARE        INC.LIKE.USER_STATUS
##          0                   0
##          PHOTO_TITLE_2       MARK_PHOTO_2
##          0                   -1
##          INC.LIKE.PHOTO_ALBUM SEND_PRESENT_2
##          -2                  -2
##          INC.MARK_PHOTO      RE_SHARE_2
##          -3                  -4
##          INC.FRIEND_INVITE    INC.USER_STATUS_COMMENT
##          -5                  -5
##          RE_SHARE            USER_MESSAGE_2
##          -6                  -11
##          PHOTO_TITLE         USER_MESSAGE
##          -12                 -14
##          INC.GUEST_PER_ALBUM
##          -16

```

//TODO: написать, как именно трактовать результат новой фичи, снова запутался и не смог разобраться.

После этого, набор данных был очищен от малозначимых факторов. Это привело к пренебрежимо малому падению значения F1, что является знаком того, что эти факторы были действительно малозначимыми и сильно не влияют на модель. Далее, для получения более полной картины, линейная модель была обучена на всей обучающей выборке размеров в 2 млн. экземпляров. Значение показателя F1 поднялось менее, чем на пол пункта, а значимость факторов стала выглядеть следующим образом:

##	PHOTO_ALBUM	INC.SEND_PRESENT	GUEST
##	239	109	85
##	USER_PERSONAL_DATA	INC.LIKE.USER_PHOTO	PHOTO_ALBUM_2
##	68	66	52
##	FRIEND_INVITE	GENDER2	INC.GUEST
##	50	47	47
##	GUEST_2	PHOTO_PER_ALBUM	INC.USER_MESSAGE
##	44	40	40
##	APP_INVITATION	USR_MSG_PER_UNIQ	USER_STATUS
##	35	32	32
##	GROUP_JOIN	USR_MSG_AND_GUEST	LIKE
##	30	23	22
##	MSG_UNIQ_USERS	INC.FRIEND_INVITE	INC.USER_STATUS_COMMENT
##	21	-8	-9
##	INC.MARK_PHOTO	USER_MESSAGE_2	RE_SHARE
##	-9	-17	-18
##	INC.GUEST_PER_ALBUM	PHOTO_TITLE	USER_MESSAGE
##	-34	-37	-39

Статистический анализ

Альбомы

Из всех несконверченных имеют 0 альбомов и 0 фотографий 47%

Из всех несконверченных имеют 0 альбомов и хотя бы 1 фотографию 7%
— 57% из них имеют ровно 1 фотографию

— из всех несконверченных имеют 1 альбом и 0 фотографий 27%

Из всех несконверченных имеют 1 альбом и хотя бы 1 фотографию 15%
— 46% из них имеют ровно 1 фотографию

Из всех сконверченных имеют 0 альбомов и 0 фотографий 12%

Из всех сконверченных имеют 0 альбомов и хотя бы 1 фотографию 9%
— 50% из них имеют ровно 1 фотографию

Из всех сконверченных имеют хотя бы 1 альбом и 0 фотографий 32%
— 96% из них имеют ровно 1 альбом

Из всех сконверченных имеют 1 альбом и хотя бы 1 фотографию 38%
— 42% из них имеют ровно 1 фотографию

Из всех сконверченных имеют больше 1 альбома и хотя бы 1 фотографию 6%

(хорошо сконверченный = active + average)

Из всех хорошо сконверченных имеют 0 альбомов и 0 фотографий 11%

Из всех хорошо сконверченных имеют 0 альбомов и хотя бы 1 фотографию 10%
— 40% из них имеют ровно 1 фотографию

Из всех хорошо сконверченных имеют хотя бы 1 альбом и 0 фотографий 28%

— 96% из них имеют ровно 1 альбом

Из всех хорошо сконверченных имеют хотя бы 1 альбом и хотя бы 1 фотографию 49%

— 34% из них имеют ровно 1 фотографию

— 83% из них имеют ровно 1 фотоальбом

— 83% из них имеют ровно 1 фотоальбом и хотя бы 1 фотографию (больше 1 фотографии 51%)

— 32% из них имеют ровно 1 фотоальбом и ровно 1 фотографию

Из всех хорошо сконверченных имеют больше 1 альбома и хотя бы 1 фотографию 7%

если смотреть отдельно по каждой из этих групп статистика ровно такая же, все одинаково

Фотографии

Из всех несконверченных имеют 0 фотографий 75%

Из всех несконверченных имеют хотя бы 1 фотографию 24%

— 45% из них имеют ровно 1 фотографию

Из всех сконверченных имеют 0 фотографий 45%

— 53% из них хотя бы 1 раз сходили в гости

— 16% из них хотя бы 1 раз кинули инвайт

— 17% из них хотя бы 1 раз написали сообщение

— — 38% из них получили хотя бы 3 гостей потом

— — 14% из них получили хотя бы 1 лайк на фото потом

— — 27% из них получили хотя бы 4 сообщения потом

— — 8% из них получили хотя бы 1 инвайт потом

— — 9% из них получили хотя бы 1 подарок потом

— — 35% из них в итоге получили фидбэк

Из всех сконверченных имеют хотя бы 1 фотографию 55%

— 39% из них ровно 1 фотографию имеют

— 14% из них ровно 2 фотографии имеют

— 72% из них хотя бы 1 раз сходили в гости

— 29% из них хотя бы 1 раз кинули инвайт

— 32% из них хотя бы 1 раз написали сообщение

— — 65% из них получили хотя бы 3 гостей потом

— — 45% из них получили хотя бы 1 лайк на фото потом

— — 43% из них получили хотя бы 4 сообщения потом

— — 16% из них получили хотя бы 1 инвайт потом

— — 20% из них получили хотя бы 1 подарок потом

— — 66% из них в итоге получили фидбэк

Из всех сконверченных имеют больше 1 фотографии 33%

Из всех сконверченных имеют больше 2 фотографий 25%

Из всех сконверченных имеют больше 4 фотографии 18%

Из всех супер-сконверченных имеют 0 фотографий 38%

— 52% из них хотя бы 1 раз сходили в гости

— 15% из них хотя бы 1 раз кинули инвайт

- 23% из них хотя бы 1 раз написали сообщение
- 58% из них получили хотя бы 3 гостей потом
- 23% из них получили хотя бы 1 лайк на фото потом
- 48% из них получили хотя бы 4 сообщения потом
- 12% из них получили хотя бы 1 инвайт потом
- 14% из них получили хотя бы 1 подарок потом
- 59% из них в итоге получили фидбэк

Из всех супер-сконверченных имеют хотя бы 1 фотографию 62%

- 33% из них ровно 1 фотографию имеют
- 13% из них ровно 2 фотографии имеют
- 74% из них хотя бы 1 раз сходили в гости
- 29% из них хотя бы 1 раз кинули инвайт
- 43% из них хотя бы 1 раз написали сообщение
- 80% из них получили хотя бы 3 гостей потом
- 57% из них получили хотя бы 1 лайк на фото потом
- 64% из них получили хотя бы 4 сообщения потом
- 20% из них получили хотя бы 1 инвайт потом
- 28% из них получили хотя бы 1 подарок потом
- 82% из них в итоге получили фидбэк

Из всех супер-сконверченных имеют больше 1 фотографии 41%

Из всех супер-сконверченных имеют больше 2 фотографий 32%

Из всех супер-сконверченных имеют больше 4 фотографий 24%

Остальное

Из всех несконверченных сходили в гости 0 раз 68%

Из всех сконверченных сходили в гости 0 раз 36%

- 42% из них хотя бы 1 раз загрузили фото
- 10% из них хотя бы 1 раз кинули инвайт
- 7% из них хотя бы 1 раз написали сообщение
- 39% из них получили хотя бы 3 гостей потом
- 23% из них получили хотя бы 1 лайк на фото потом
- 27% из них получили хотя бы 4 сообщения потом
- 9% из них получили хотя бы 1 инвайт потом
- 10% из них получили хотя бы 1 подарок потом
- 40% из них в итоге получили фидбэк

Из всех сконверченных сходили в гости хотя бы 1 раз 63%

- 17% из них ровно 1 раз
- 51% из них больше 4 раз
- 27% из них больше 9 раз
- 62% из них хотя бы 1 раз загрузили фото
- 31% из них хотя бы 1 раз кинули инвайт
- 35% из них хотя бы 1 раз написали сообщение
- 61% из них получили хотя бы 3 гостей потом
- 36% из них получили хотя бы 1 лайк на фото потом
- 41% из них получили хотя бы 4 сообщения потом
- 15% из них получили хотя бы 1 инвайт потом

- 18% из них получили хотя бы 1 подарок потом
 - 60% из них в итоге получили фидбэк
-

Из всех несконверченных инвайтили 0 раз 90%

Из всех сконверченных инвайтили 0 раз 76%

- 50% из них хотя бы 1 раз загрузили фото
- 57% из них хотя бы 1 раз сходили в гости
- 21% из них хотя бы 1 раз написали сообщение
- 49% из них получили хотя бы 3 гостей потом
- 28% из них получили хотя бы 1 лайк на фото потом
- 34% из них получили хотя бы 4 сообщения потом
- 4% из них получили хотя бы 1 инвайт потом
- 12% из них получили хотя бы 1 подарок потом
- 49% из них в итоге получили фидбэк

Из всех сконверченных инвайтили хотя бы 1 раз 24%

- 69% из них хотя бы 1 раз загрузили фото
 - 83% из них хотя бы 1 раз сходили в гости
 - 36% из них хотя бы 1 раз написали сообщение
 - 65% из них получили хотя бы 3 гостей потом
 - 42% из них получили хотя бы 1 лайк на фото потом
 - 43% из них получили хотя бы 4 сообщения потом
 - 39% из них получили хотя бы 1 инвайт потом
 - 24% из них получили хотя бы 1 подарок потом
 - 65% из них в итоге получили фидбэк
-

Из всех несконверченных написали 0 сообщений 90%

Из всех сконверченных написали 0 сообщений 74%

- 50% из них хотя бы 1 раз загрузили фото
- 55% из них хотя бы 1 раз сходили в гости
- 20% из них хотя бы 1 раз отправили инвайт
- 46% из них получили хотя бы 3 гостей потом
- 27% из них получили хотя бы 1 лайк на фото потом
- 26% из них получили хотя бы 4 сообщения потом
- 11% из них получили хотя бы 1 инвайт потом
- 13% из них получили хотя бы 1 подарок потом
- 44% из них в итоге получили фидбэк

Из всех сконверченных написали хотя бы 1 сообщение 25%

- 27% из них ровно 1 сообщение
- 51% из них больше 3 сообщения
- 29% из них больше 9 сообщений
- 70% из них хотя бы 1 раз загрузили фото
- 88% из них хотя бы 1 раз сходили в гости
- 33% из них хотя бы 1 раз отправили инвайт
- 72% из них получили хотя бы 3 гостей потом
- 43% из них получили хотя бы 1 лайк на фото потом
- 64% из них получили хотя бы 4 сообщения потом
- 17% из них получили хотя бы 1 инвайт потом

- 21% из них получили хотя бы 1 подарок потом
- 77% из них в итоге получили фидбэк

Дополнительно

Из всех сконверченных, кто загрузил фото — сходили в гости 72%
Из всех сконверченных, кто сходил в гости — загрузили фото 63%
То есть из тех, кто загрузил фото — 72% которые и в гости сходили, и фото загрузили.
А из тех, кто ходил в гости — 63% которые и в гости сходили, и фото загрузили.
Следовательно, нам более интересны и важны люди, которые загружают фото? Походу нет, первых просто меньше.

Из всех сконверченных загрузили хотя бы 1 фото и сходили в гости 0 раз 15%
— 51% из них получили хотя бы 3 гостей потом
— 35% из них получили хотя бы 1 лайк на фото потом
— 34% из них получили хотя бы 4 сообщения потом
— 13% из них получили хотя бы 1 инвайт потом
— 14% из них получили хотя бы 1 подарок потом
— 53% из них в итоге получили фидбэк

Из всех сконверченных загрузили 0 фото и сходили в гости хотя бы 1 раз 23%
— из них получили хотя бы 3 гостей потом 44%
— из них получили хотя бы 1 лайк на фото потом 14%
— из них получили хотя бы 4 сообщения потом 31%
— из них получили хотя бы 1 инвайт потом 9%
— из них получили хотя бы 1 подарок потом 10%
— из них в итоге получили фидбэк 40%

Из всех сконверченных загрузили хотя бы 1 фото и сходили в гости хотя бы 1 раз 39%
— из них получили хотя бы 3 гостей потом 71%
— из них получили хотя бы 1 лайк на фото потом 50%
— из них получили хотя бы 4 сообщения потом 47%
— из них получили хотя бы 1 инвайт потом 18%
— из них получили хотя бы 1 подарок потом 23%
— из них в итоге получили фидбэк 71%

Из всех сконверченных загрузили 0 фото и сходили в гости 0 раз 20%
— из них получили хотя бы 3 гостей потом 30%
— из них получили хотя бы 1 лайк на фото потом 14%
— из них получили хотя бы 4 сообщения потом 22%
— из них получили хотя бы 1 инвайт потом 8%
— из них получили хотя бы 1 подарок потом 7%
— из них в итоге получили фидбэк 30%

Фидбэк: первые 2 часа

(получили фидбэк = 2 или больше входящих действий из списка основных)

Всего фидбека получили 18%

Всего фидбека получили (>=5) 9%

Из тех, кто получил фидбэк:

- сконвертились 46%
- не сконвертились 54%

Из тех, кто не получил фидбэк:

- сконвертились 17%
- не сконвертились 83%

Из всех сконверченных получили фидбэк 37%

- из них загрузили: 0 фотографий — 20%, хотя бы 1 фотографию 80%
- из них сходили в гости: 0 раз — 16%, хотя бы 1 раз 84% (>3 — 59%)
- из них кинули инвайт: 0 раз — 65%, хотя бы 1 раз 35%
- из них написали сообщений: 0 раз — 47%, хотя бы 1 раз 53% (>4 — 29%)

Из всех несконверченных не получили фидбэк 88%

- из них загрузили: 0 фотографий — 80%, хотя бы 1 фотографию 20%
- из них сходили в гости: 0 раз — 72%, хотя бы 1 раз 28%
- из них кинули инвайт: 0 раз — 93%, хотя бы 1 раз 7%
- из них написали сообщений: 0 раз — 95%, хотя бы 1 раз 5%

Из всех супер-сконверченных получили фидбэк 47%

Из всех хорошо-сконверченных получили фидбэк 38%

Из всех средне-сконверченных получили фидбэк 33%

Из всех плохо-сконверченных получили фидбэк 28%

Из всех несконверченных получили подарок 1%

Из всех сконверченных получили подарок 3%

Из супер-сконверченных получили подарок 4%

Из всех, кто получил подарок:

- сконвертились 62%
- не сконвертились 38%

Из всех, кто не получил подарка:

- сконвертились 23%
- не сконвертились 77%

Из всех несконверченных получили 2 или больше гостей 13%

Из всех сконверченных получили 2 или больше гостей 36%

Из супер-сконверченных получили 2 или больше гостей 45%

Из всех, кто не получил гостей:

- сконвертились 18%
- не сконвертились 82%

Из всех, кто получил гостей:

- сконвертились 45%
- не сконвертились 55%

Из всех несконверченных получили лайк к фотке 2%

Из всех сконверченных получили лайк к фотке 9%

Из супер-сконверченных получили лайк к фотке 12%

Из всех, кто не получил лайк к фотке:

- сконвертились 22%
- не сконвертились 78%

Из всех, кто получил лайк к фотке:

- сконвертились 57%
- не сконвертились 43%

Из всех несконверченных получили 2 и больше сообщений 6%

Из всех сконверченных получили 2 и больше сообщений 17%

Из супер-сконверченных получили 2 и больше сообщений 26%

Из всех, кто не получил сообщений:

- сконвертились 20%
- не сконвертились 80%

Из всех, кто получил хотя бы 1 сообщение:

- сконвертились 45%
- не сконвертились 55%

Из всех, кто не получил инвайт:

- сконвертились 22%
- не сконвертились 78%

Из всех, кто получил инвайт:

- сконвертились 47%
- не сконвертились 53%

Фидбэк: первый день

(получили фидбэк = 5 или больше входящих действий из списка основных)

Всего фидбека получили 22%

Из тех, кто получил фидбэк:

- сконвертились 52%
- не сконвертились 48%

Из тех, кто не получил фидбэк:

- сконвертились 14%
- не сконвертились 86%

Из всех сконверченных получили фидбэк 50%

- из них загрузили: 0 фотографий — 15%, хотя бы 1 фотографию 85%
- из них сходили в гости: 0 раз — 10%, хотя бы 1 раз 90% (>3 — 77%)
- из них кинули инвайт: 0 раз — 63%, хотя бы 1 раз 37%
- из них написали сообщений: 0 раз — 27%, хотя бы 1 раз 73% (>4 — 50%)

Из всех несконверченных не получили фидбэк 88%

- из них загрузили: 0 фотографий — 80%, хотя бы 1 фотографию 20%
- из них сходили в гости: 0 раз — 72%, хотя бы 1 раз 28%
- из них кинули инвайт: 0 раз — 93%, хотя бы 1 раз 7%
- из них написали сообщений: 0 раз — 95%, хотя бы 1 раз 5%

Из всех супер-сконверченных получили фидбэк 67%

Из всех хорошо-сконверченных получили фидбэк 53%

Из всех средне-сконверченных получили фидбэк 45%

Из всех плохо-сконверченных получили фидбэк 36%

Из всех несконверченных получили подарок 2%
Из всех сконверченных получили подарок 10%
Из супер-сконверченных получили подарок 15%

Из всех, кто не получил подарка:

- сконвертились 22%
- не сконвертились 78%

Из всех, кто получил подарок:

- сконвертились 65%
 - не сконвертились 35%
-

Из всех несконверченных получили 2 или больше гостей 15%
Из всех сконверченных получили 2 или больше гостей 51%
Из супер-сконверченных получили 2 или больше гостей 67%

Из всех, кто не получил гостей:

- не сконвертились 88%
- сконвертились 12%

Из всех, кто получил гостей:

- сконвертились 45%
 - не сконвертились 55%
-

Из всех несконверченных не получили лайк к фотке 95%
Из всех сконверченных получили лайк к фотке 25%
Из супер-сконверченных получили лайк к фотке 34%

Из всех, кто не получил лайк к фотке:

- не сконвертились 80%
- сконвертились 20%

Из всех, кто получил лайк к фотке:

- сконвертились 60%
 - не сконвертились 40%
-

Из всех несконверченных получили 2 и больше сообщений 10%
Из всех сконверченных получили 2 и больше сообщений 36%
Из супер-сконверченных получили 2 и больше сообщений 54%

Из всех, кто не получил сообщений:

- не сконвертились 83%
- сконвертились 17%

Из всех, кто получил хотя бы 1 сообщение:

- сконвертились 48%
 - не сконвертились 52%
-

Изолированное тестирование: первые 2 часа

Выделим кластер пользователей с одинаковыми основными показателями:

- загрузили фото: 1-4 раз
- ходили в гости: 1-4 раз
- инвайтили: 1-4 раз
- писали сообщения: 1-4 раз
- заполнили пользовательские данные

- не дарили подарков
- не приглашали в приложения
- не репостили

Из тех, кто не загрузил фотографию:

- сконвертились 40% / не сконвертились 60%
- получили инвайт 9% / не получили 91%
- получили гостей 45% / не получили 55%
- получили сообщения 42% / не получили 58%
- получили фидбэк 41% / не получили 59%

Из тех, кто загрузил фотографию:

- сконвертились 50% / не сконвертились 50%
- получили инвайт 15% / не получили 85%
- получили гостей 57% / не получили 43%
- получили сообщения 50% / не получили 50%
- получили фидбэк 62% / не получили 38%

Из тех, кто не отправлял инвайт:

- сконвертились 45% / не сконвертились 55%
- получили фидбэк 57% / не получили 43%

Из тех, кто отправил инвайт:

- сконвертились 48% / не сконвертились 52%
- получили фидбэк 61% / не получили 39%

Из тех, кто не сходил в гости:

- сконвертились 44% / не сконвертились 56%
- получили фидбэк 57% / не получили 43%

Из тех, кто сходил в гости:

- сконвертились 48% / не сконвертились 52%
- получили фидбэк 60% / не получили 40%

Из тех, кто не писал сообщений:

- сконвертились 44% / не сконвертились 56%
- получили фидбэк 37% / не получили 63%

Из тех, кто написал сообщение:

- сконвертились 50% / не сконвертились 50%
- получили фидбэк 74% / не получили 26%

Из тех, кто не заполнил пользовательские данные:

- сконвертились 40% / не сконвертились 60%
- получили инвайт 10% / не получили 85%
- получили гостей 43% / не получили 43%
- получили сообщения 44% / не получили 50%
- получили фидбэк 52% / не получили 48%

Из тех, кто заполнил пользовательские данные:

- сконвертились 47% / не сконвертились 53%
- получили инвайт 14% / не получили 85%
- получили гостей 54% / не получили 43%
- получили сообщения 49% / не получили 50%
- получили фидбэк 58% / не получили 42%

Из тех, кто получил подарок (3%):

- сконвертились 68%
- не сконвертились 32%

Из тех, кто не получил подарок (97%):

- сконвертились 46%
- не сконвертились 54%

Из тех, кто получил фидбэк:

- сконвертились 50%
- не сконвертились 50%

Из тех, кто не получил фидбэк:

- сконвертились 45%
- не сконвертились 65%

Выделим другой кластер пользователей с одинковыми основными показателями:

- загрузили фото: 0 раз
- ходили в гости: 0 раз
- инвайтили: 0 раз
- писали сообщения: 0 раз
- не заполнили пользовательские данные
- не дарили подарков
- не приглашали в приложения
- не репостили

Из тех, кто не загрузил фотографию:

- сконвертились 9% / не сконвертились 91%
- получили фидбэк 1% / не получили 99%

Из тех, кто загрузил фотографию:

- сконвертились 24% / не сконвертились 76%
- получили фидбэк 14% / не получили 86%

Из тех, кто не отправлял инвайт:

- сконвертились 9% / не сконвертились 91%
- получили фидбэк 1% / не получили 99%

Из тех, кто отправлял инвайт:

- сконвертились 30% / не сконвертились 70%
- получили фидбэк 8% / не получили 92%

Из тех, кто не сходил в гости:

- сконвертились 9% / не сконвертились 56%
- получили фидбэк 1% / не получили 43%

Из тех, кто сходил в гости:

- сконвертились 23% / не сконвертились 52%
- получили фидбэк 7% / не получили 40%

Из тех, кто не писал сообщений:

- сконвертились 9% / не сконвертились 56%
- получили фидбэк 1% / не получили 63%

Из тех, кто написал сообщение:

- сконвертились 29% / не сконвертились 50%
- получили фидбэк 48% / не получили 26%

Из тех, кто не заполнил пользовательские данные:

- сконвертились 9% / не сконвертились 60%
- получили фидбэк 1% / не получили 48%

Из тех, кто заполнил пользовательские данные:

- сконвертились 21% / не сконвертились 53%
- получили фидбэк 4% / не получили 42%

Из тех, кто получил фидбэк (2%):

- сконвертились 21%
- не сконвертились 79%

Из тех, кто не получил фидбэк (98%):

- сконвертились 9%
 - не сконвертились 91%
-

Выделим кластер пользователей с одинаковыми основными показателями:

- получили лайк к фото: 0 раз
- получили сообщение: 0 раз
- получили оценку: 0 раз
- получили инвайт: 0 раз
- получили подарок: 0 раз

Из тех, кто получил подарок (1%):

- сконвертились 53% / не сконвертились 47%

Из тех, кто не получил подарок (99%):

- сконвертились 17% / не сконвертились 83%

Из тех, кто получил инвайт (1%):

- сконвертились 40% / не сконвертились 60%

Из тех, кто не получил инвайт (99%):

- сконвертились 17% / не сконвертились 83%

Из тех, кто получил лайк к фото (1%):

- сконвертились 54% / не сконвертились 46%

Из тех, кто не получил лайк к фото (99%):

- сконвертились 17% / не сконвертились 83%

Из тех, кто получил сообщение (7%):

- сконвертились 38% / не сконвертились 62%

Из тех, кто не получил сообщение (93%):

- сконвертились 17% / не сконвертились 83%

Из тех, кто получил оценку (10%):

- сконвертились 36% / не сконвертились 64%

Из тех, кто не получил оценку (90%):

- сконвертились 17% / не сконвертились 83%
-

Выделим кластер пользователей с одинаковыми основными показателями:

- получили лайк к фото: 1-5 раз
- получили сообщение: 1-5 раз
- получили оценку: 1-5 раз
- получили инвайт: 1-5 раз
- получили подарок: 1-5 раз

Из тех, кто получил подарок (18%):

- сконвертились 65% / не сконвертились 35%

Из тех, кто не получил подарок (82%):

- сконвертились 58% / не сконвертились 42%

Из тех, кто получил инвайт (12%):

— сконвертились 59% / не сконвертились 41%

Из тех, кто не получил инвайт (88%):

— сконвертились 59% / не сконвертились 41%

Из тех, кто получил лайк к фото (42%):

— сконвертились 61% / не сконвертились 39%

Из тех, кто не получил лайк к фото (58%):

— сконвертились 52% / не сконвертились 48%

Из тех, кто получил сообщение (60%):

— сконвертились 59% / не сконвертились 41%

Из тех, кто не получил сообщение (40%):

— сконвертились 57% / не сконвертились 43%

Из тех, кто получил оценку (67%):

— сконвертились 56% / не сконвертились 44%

Из тех, кто не получил оценку (90%):

— сконвертились 58% / не сконвертились 42%

Выделим кластер пользователей с одинокими основными показателями:

— получили лайк к фото: 0 раз

— получили сообщение: 0 раз

— получили оценку: 0 раз

— получили инвайт: 0 раз

— получили подарок: 0 раз

— загрузили фото: 0 раз

— ходили в гости: 0 раз

— инвайтили: 0 раз

— писали сообщения: 0 раз

— не заполнили пользовательские данные

— не дарили подарков

— не приглашали в приложения

— не репостили

Из тех, кто получил подарок (<1%):

— сконвертились 47% / не сконвертились 35%

Из тех, кто не получил подарок (>99%):

— сконвертились 9% / не сконвертились 91%

Из тех, кто получил инвайт (1%):

— сконвертились 23% / не сконвертились 41%

Из тех, кто не получил инвайт (99%):

— сконвертились 9% / не сконвертились 91%

Из тех, кто получил лайк к фото (<1%):

— сконвертились 49% / не сконвертились 51%

Из тех, кто не получил лайк к фото (99%):

— сконвертились 9% / не сконвертились 91%

Из тех, кто получил сообщение (1%):

— сконвертились 16% / не сконвертились 84%

Из тех, кто не получил сообщение (40%):

— сконвертились 9% / не сконвертились 91%

Из тех, кто получил оценку (6%):

— сконвертились 24% / не сконвертились 76%

Из тех, кто не получил оценку (90%):

— сконвертились 9% / не сконвертились 91%

Изолированное тестирование: первый день

Выделим кластер пользователей с одинаковыми основными показателями:

- загрузили фото: 1-4 раз
- ходили в гости: 1-4 раз
- инвайтили: 1-4 раз
- писали сообщения: 1-4 раз
- заполнили пользовательские данные
- не дарили подарков
- не приглашали в приложения
- не репостили

Из тех, кто не загрузил фотографию:

- сконвертились 30% / не сконвертились 70%
- получили инвайт 14% / не получили 86%
- получили гостей 61% / не получили 39%
- получили сообщения 57% / не получили 43%
- получили фидбэк 29% / не получили 71%

Из тех, кто загрузил фотографию:

- сконвертились 43% / не сконвертились 57%
- получили инвайт 23% / не получили 77%
- получили гостей 76% / не получили 24%
- получили сообщения 65% / не получили 35%
- получили фидбэк 54% / не получили 46%

Из тех, кто не отправлял инвайт:

- сконвертились 40% / не сконвертились 60%
- получили фидбэк 42% / не получили 58%

Из тех, кто отправил инвайт:

- сконвертились 42% / не сконвертились 58%
- получили фидбэк 51% / не получили 49%

Из тех, кто не сходил в гости:

- сконвертились 45% / не сконвертились 55%
- получили фидбэк 44% / не получили 56%

Из тех, кто сходил в гости:

- сконвертились 43% / не сконвертились 57%
- получили фидбэк 51% / не получили 49%

Из тех, кто не писал сообщений:

- сконвертились 38% / не сконвертились 62%
- получили фидбэк 20% / не получили 80%

Из тех, кто написал сообщение:

- сконвертились 44% / не сконвертились 56%
- получили фидбэк 69% / не получили 31%

Из тех, кто не заполнил пользовательские данные:

- сконвертились 35% / не сконвертились 65%

— получили фидбэк 37% / не получили 63%

Из тех, кто заполнил пользовательские данные:

— сконвертились 40% / не сконвертились 60%

— получили фидбэк 47% / не получили 53%

Из тех, кто получил подарок (6%):

— сконвертились 62%

— не сконвертились 38%

Из тех, кто не получил подарок (9%):

— сконвертились 40%

— не сконвертились 60%

Из тех, кто получил фидбэк:

— сконвертились 45%

— не сконвертились 55%

Из тех, кто не получил фидбэк:

— сконвертились 35%

— не сконвертились 65%

Выделим другой кластер пользователей с одинокими основными показателями:

— загрузили фото: 0 раз

— ходили в гости: 0 раз

— инвайтили: 0 раз

— писали сообщения: 0 раз

— не заполнили пользовательские данные

— не дарили подарков

— не приглашали в приложения

— не репостили

Из тех, кто не загрузил фотографию:

— сконвертились 5% / не сконвертились 95%

— получили фидбэк 0% / не получили 100%

Из тех, кто загрузил фотографию:

— сконвертились 17% / не сконвертились 83%

— получили фидбэк 7% / не получили 93%

Из тех, кто не отправлял инвайт:

— сконвертились 5% / не сконвертились 95%

— получили фидбэк 0% / не получили 100%

Из тех, кто отправлял инвайт:

— сконвертились 22% / не сконвертились 78%

— получили фидбэк 4% / не получили 96%

Из тех, кто не сходил в гости:

— сконвертились 5% / не сконвертились 95%

— получили фидбэк 0% / не получили 100%

Из тех, кто сходил в гости:

— сконвертились 19% / не сконвертились 81%

— получили фидбэк 4% / не получили 96%

Из тех, кто не писал сообщений:

— сконвертились 5% / не сконвертились 95%

— получили фидбэк 0% / не получили 100%

Из тех, кто написал сообщение:

- сконвертились 27% / не сконвертились 73%
- получили фидбэк 40% / не получили 60%

Из тех, кто не заполнил пользовательские данные:

- сконвертились 5% / не сконвертились 95%
- получили фидбэк 0% / не получили 100%

Из тех, кто заполнил пользовательские данные:

- сконвертились 14% / не сконвертились 86%
- получили фидбэк 1% / не получили 60%

Из тех, кто получил фидбэк:

- сконвертились 21%
- не сконвертились 79%

Из тех, кто не получил фидбэк:

- сконвертились 5%
- не сконвертились 95%

Выделим кластер пользователей с одинаковыми основными показателями:

- получили лайк к фото: 0 раз
- получили сообщение: 0 раз
- получили оценку: 0 раз
- получили инвайт: 0 раз
- получили подарок: 0 раз

Из тех, кто получил подарок:

- сконвертились 48% / не сконвертились 52%

Из тех, кто не получил подарок:

- сконвертились 12% / не сконвертились 88%

Из тех, кто получил инвайт:

- сконвертились 35% / не сконвертились 65%

Из тех, кто не получил инвайт:

- сконвертились 12% / не сконвертились 88%

Из тех, кто получил лайк к фото:

- сконвертились 48% / не сконвертились 52%

Из тех, кто не получил лайк к фото:

- сконвертились 12% / не сконвертились 88%

Из тех, кто получил сообщение:

- сконвертились 37% / не сконвертились 63%

Из тех, кто не получил сообщение:

- сконвертились 12% / не сконвертились 88%

Из тех, кто получил оценку:

- сконвертились 30% / не сконвертились 70%

Из тех, кто не получил сообщение:

- сконвертились 12% / не сконвертились 88%

Выделим кластер пользователей с одинаковыми основными показателями:

- получили лайк к фото: 1-5 раз
- получили сообщение: 1-5 раз
- получили оценку: 1-5 раз

- получили инвайт: 1-5 раз
- получили подарок: 1-5 раз

Из тех, кто получил подарок (20%):

- сконвертились 65% / не сконвертились 35%

Из тех, кто не получил подарок (80%):

- сконвертились 54% / не сконвертились 46%

Из тех, кто получил инвайт (14%):

- сконвертились 57% / не сконвертились 43%

Из тех, кто не получил инвайт (86%):

- сконвертились 55% / не сконвертились 45%

Из тех, кто получил лайк к фото (45%):

- сконвертились 58% / не сконвертились 42%

Из тех, кто не получил лайк к фото (55%):

- сконвертились 50% / не сконвертились 50%

Из тех, кто получил сообщение (60%):

- сконвертились 51% / не сконвертились 63%

Из тех, кто не получил сообщение:

- сконвертились 58% / не сконвертились 88%

Из тех, кто получил оценку (70%):

- сконвертились 53% / не сконвертились 47%

Из тех, кто не получил сообщение (30%):

- сконвертились 57% / не сконвертились 43%

Выделим кластер пользователей с одинаковыми основными показателями:

- получили лайк к фото: 0 раз
- получили сообщение: 0 раз
- получили оценку: 0 раз
- получили инвайт: 0 раз
- получили подарок: 0 раз
- загрузили фото: 0 раз
- ходили в гости: 0 раз
- инвайтили: 0 раз
- писали сообщения: 0 раз
- не заполнили пользовательские данные
- не дарили подарков
- не приглашали в приложения
- не репостили

Из тех, кто получил подарок (<1%):

- сконвертились 38% / не сконвертились 62%

Из тех, кто не получил подарок (>99%):

- сконвертились 5% / не сконвертились 95%

Из тех, кто получил инвайт (<1%):

- сконвертились 17% / не сконвертились 83%

Из тех, кто не получил инвайт (99%):

- сконвертились 5% / не сконвертились 95%

Из тех, кто получил лайк к фото (<1%):

— сконвертились 38% / не сконвертились 62%

Из тех, кто не получил лайк к фото (99%):

— сконвертились 5% / не сконвертились 95%

Из тех, кто получил сообщение (1%):

— сконвертились 10% / не сконвертились 90%

Из тех, кто не получил сообщение (40%):

— сконвертились 5% / не сконвертились 95%

Из тех, кто получил оценку (7%):

— сконвертились 16% / не сконвертились 84%

Из тех, кто не получил оценку (90%):

— сконвертились 5% / не сконвертились 95%

Проверка фидбэка

Был выбран кластер пользователей размером 1000 человек, из пользователей, которые:
(для каждого пользователя были подсчитаны все связи с другими пользователями)

Сконвертировались

1. Из всего фидбека
 - 1.1) 30% реального фидбэка (в ответ на действие конкретному пользователю)
 - 1.1) 70% просто входящее действие (пришло в ответ на любую активность или просто так)
2. 72% первыми сделали действие
 - 2.1) 12% получили фидбэк от этого пользователя
 - 2.2) 60% не получили фидбэк от этого пользователя
3. 28% первыми получили действие
 - 3.1) 5% дали фидбэк
 - 3.2) 23% не дали фидбэк

Не сконвертировались

1. Из всего фидбека
 - 1.1) 31% реального фидбэка (в ответ на действие конкретному пользователю)
 - 1.1) 69% просто входящее действие (пришло в ответ на любую активность или просто так)
2. 76% первыми сделали действие
 - 2.1) 11% получили фидбэк от этого пользователя
 - 2.2) 65% не получили фидбэк от этого пользователя
3. 24% первыми получили действие
 - 3.1) 4% дали фидбэк
 - 3.2) 20% не дали фидбэк

Сконвертировались и получили фидбэк

1. Из всего фидбека
 - 1.1) 28% реального фидбэка (в ответ на действие конкретному пользователю)
 - 1.1) 72% просто входящее действие (пришло в ответ на любую активность или просто так)
2. 67% первыми сделали действие
 - 2.1) 13% получили фидбэк от этого пользователя
 - 2.2) 54% не получили фидбэк от этого пользователя
3. 33% первыми получили действие

- 3.1) 6% дали фидбэк
- 3.2) 27% не дали фидбэк

Сконвертировались и не получили фидбэк

1. Из всего фидбека
 - 1.1) 42% реального фидбэка (в ответ на действие конкретному пользователю)
 - 1.1) 58% просто входящее действие (пришло в ответ на любую активность или просто так)
2. 90% первыми сделали действие
 - 2.1) 7% получили фидбэк от этого пользователя
 - 2.2) 83% не получили фидбэк от этого пользователя
3. 10% первыми получили действие
 - 3.1) 2% дали фидбэк
 - 3.2) 8% не дали фидбэк

Не сконвертировались и получили фидбэк

1. Из всего фидбека
 - 1.1) 26% реального фидбэка (в ответ на действие конкретному пользователю)
 - 1.1) 74% просто входящее действие (пришло в ответ на любую активность или просто так)
2. 66% первыми сделали действие
 - 2.1) 12% получили фидбэк от этого пользователя
 - 2.2) 54% не получили фидбэк от этого пользователя
3. 34% первыми получили действие
 - 3.1) 29% дали фидбэк
 - 3.2) 5% не дали фидбэк

Не сконвертировались и не получили фидбэк

1. Из всего фидбека
 - 1.1) 40% реального фидбэка (в ответ на действие конкретному пользователю)
 - 1.1) 60% просто входящее действие (пришло в ответ на любую активность или просто так)
2. 88% первыми сделали действие
 - 2.1) 8% получили фидбэк от этого пользователя
 - 2.2) 80% не получили фидбэк от этого пользователя
3. 12% первыми получили действие
 - 3.1) 0% дали фидбэк
 - 3.2) 11% не дали фидбэк

Проверка фото: первые 2 часа

Выделим кластер пользователей с одинаковыми основными показателями:

- загрузили фото: 1 или больше раз
- ходили в гости: 0 раз
- писали сообщения: 0 раз
- не заполнили пользовательские данные
- не дарили подарков
- не приглашали в приложения
- не репостили

Без инвайтов

Из тех, кто загрузили фотки и не получили лайка:

— 24% сконвертились

Из тех, кто загрузили фотки и получили лайк (1%):

— 50% сконвертились

С инвайтами

Из тех, кто загрузили фотки и не получили лайка:

— 42% сконвертились

Из тех, кто загрузили фотки и получили лайк (8%):

— 56% сконвертились

Выделим кластер пользователей с одинаковыми основными показателями:

— загрузили фото: 1 или больше раз

— ходили в гости: 1-5 раз

— писали сообщения: 1-5 раз

— заполнили пользовательские данные

— не дарили подарков

— не приглашали в приложения

— не репостили

Без инвайтов

Из тех, кто загрузили фотки и не получили лайка:

— 47% сконвертились

Из тех, кто загрузили фотки и получили лайк (13%):

— 58% сконвертились

С инвайтами

Из тех, кто загрузили фотки и не получили лайка:

— 50% сконвертились

Из тех, кто загрузили фотки и получили лайк (17%):

— 57% сконвертились

Проверка фото: первый день

Выделим кластер пользователей с одинаковыми основными показателями:

— загрузили фото: 1 или больше раз

— ходили в гости: 0 раз

— писали сообщения: 0 раз

— не заполнили пользовательские данные

— не дарили подарков

— не приглашали в приложения

— не репостили

Без инвайтов

Из тех, кто загрузили фотки и не получили лайка:

— 16% сконвертились

Из тех, кто загрузили фотки и получили лайк (2%):

— 45% сконвертились

С инвайтами

Из тех, кто загрузили фотки и не получили лайка:

— 33% сконвертились

Из тех, кто загрузили фотки и получили лайк (15%):

— 49% сконвертились

Выделим кластер пользователей с одинаковыми основными показателями:

— загрузили фото: 1 или больше раз

— ходили в гости: 1-5 раз

— писали сообщения: 1-5 раз

— заполнили пользовательские данные

— не дарили подарков

— не приглашали в приложения

— не репостили

Без инвайтов

Из тех, кто загрузили фотки и не получили лайка:

— 42% сконвертились

Из тех, кто загрузили фотки и получили лайк (21 %):

— 51% сконвертились

С инвайтами

Из тех, кто загрузили фотки и не получили лайка:

— 44% сконвертились

Из тех, кто загрузили фотки и получили лайк (26%):

— 54% сконвертились

По полу

Из тех, кто загрузил фотки, получил к ним лайки и сконвертился: — 63% женщины — 37% мужчины

Итог

Как итог, можно сделать следующие заключения: