***Задание 1:***

*Предложить методологию исследования успешности релиза сервиса. У PFM есть сервис Анализ личных финансов (АЛФ). Сервис можно посмотреть в Сбербанк Онлайн на главном экране справа от ленты сторис. Сервис представляет из себя агрегацию расходов и доходов клиента помесячно по категориям. АЛФ помогает клиентам контролировать свои расходы и понять, куда были потрачены деньги.  
Основные метрики, которые важны для сервиса:  
• MAU – количество клиентов, которые хотя бы раз в месяц заходят в сервис  
• DAU / MAU – отношение средней дневной аудитории сервиса к MAU  
• Retention – кривая жизни клиента в сервисе. За 0 когорту берутся пользователи, которые зашли в АЛФ в отчетном месяце и не заходили в АЛФ в 6 прошлых месяцев. За Retention 1-ого месяца считается доля клиентов из 0 когорты, которые зашли в АЛФ в следующий за отчетным месяц.*

*Релиз:  
 В новой версии у клиентов появилась возможность переключаться с круговой диаграммы на столбчатую диаграмму (раньше в АЛФ была только круговая диаграмма). Также добавилась возможность перейти к следующему месяцу и посмотреть прогноз своих расходов и доходов (который посчитал с помощью модели машинного обучения)  
Клиенты обновляются на новую версию постепенно. Вся информация по версиям приложения и кликам в приложении доступна. Доступна статистка по заходу в АЛФ, переключению с бублика на столбики, переход к прошлому и будущему месяцу, проваливание внутрь той или иной категории расходов.*

*Гипотеза:  
 Релиз привел к росту MAU, DAU/MAU, Retention и активности внутри АЛФ (переходы на прошлый и будущий месяц и проваливание внутрь категорий). Как проверить успешность релиза и подтвердить или опровергнуть гипотезы? Какие еще гипотезы можно предложить?*

Ответ:

1. Формулируем гипотезу: *Добавление функции смены круговой диаграммы на столбчатую, а также возможность просмотра расходов за прошлые и будущие периоды приведет к росту MAU, DAU/MAU, Retention и активности внутри АЛФ.*
2. Поскольку обновление уже идет, мы понимаем, что у нас уже есть пользователи, которые перешли на новую версию. Принимаем решение провести AB тестирование.
3. Определяем метрики: *MAU, DAU/MAU, Retention, Количество активных действий (смена круговой диаграммы на столбчатую, движение по периодам, проваливание в категорию, свайпы до конца изображения).*

Добавляю метрики: *процент роста/снижения количества транзакций по карте, процент роста/снижения суммы покупок по карте, процент снижения/роста снятия наличности, процент снижения/роста перевода средств в другие банки.*

1. Пользователей– разделяем на 2 когорты (0 – те, кто не заходил в последние 6 месяцев, но зашел в текущем(отчетном), и 1 когорту – тех, кто в текущем месяце зашел и заходил ранее в течении полугода).

Данных пользователей мы сегментируем по таким признакам, как: устройство (Android/IOS), пол, возраст (категории: до 18, 18-30, 30-60, 60+), регион проживания (или терр банк).

**Очень важный момент.** Старт тестирования начать проводить на группе А/А. То есть на двух равнозначных группах, которые НЕ ОБНОВИЛИСЬ до новой версии.

Далее, спустя месяц исследования, мы обновляем одну группу до новой версии, если никаких статистически значимых отличий в поведении пользователей не наблюдаем. И только теперь наблюдаем за изменением метрик, анализируя с периодом до обновления.

Выборка должна выглядеть (примерно) следующим образом:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Когорта | Устройство пользователя | Пол | Возраст | Город (ТБ) | Количество пользователей |
| 0 когорта (не заходили 6 месяцев) | Android | М | 18-30 | Москва | 1000 |
| СПБ | 1000 |
| Калининград | 1000 |
| 30-60 | Москва | 1000 |
| СПБ | 1000 |
| Калининград | 1000 |
| Ж | 18-30 | Москва | 1000 |
| СПБ | 1000 |
| Калининград | 1000 |
| 30-60 | Москва | 1000 |
| СПБ | 1000 |
| Калининград | 1000 |
| IOS | М | 18-30 | Москва | 1000 |
| СПБ | 1000 |
| Калининград | 1000 |
| 30-60 | Москва | 1000 |
| СПБ | 1000 |
| Калининград | 1000 |
| Ж | 18-30 | Москва | 1000 |
| СПБ | 1000 |
| Калининград | 1000 |
| 30-60 | Москва | 1000 |
| СПБ | 1000 |
| Калининград | 1000 |
| 1 когорта (заходили последние 6 месяцев) | Android | М | 18-30 | Москва | 1000 |
| СПБ | 1000 |
| Калининград | 1000 |
| 30-60 | Москва | 1000 |
| СПБ | 1000 |
| Калининград | 1000 |
| Ж | 18-30 | Москва | 1000 |
| СПБ | 1000 |
| Калининград | 1000 |
| 30-60 | Москва | 1000 |
| СПБ | 1000 |
| Калининград | 1000 |
| IOS | М | 18-30 | Москва | 1000 |
| СПБ | 1000 |
| Калининград | 1000 |
| 30-60 | Москва | 1000 |
| СПБ | 1000 |
| Калининград | 1000 |
| Ж | 18-30 | Москва | 1000 |
| СПБ | 1000 |
| Калининград | 1000 |
| 30-60 | Москва | 1000 |
| СПБ | 1000 |
| Калининград | 1000 |

Причем такая выборка должна быть как для группы A(которая не обновлена до новой версии), так и для группы B(обновленная, которая изначально является аналогичной группе A и по показателям ведет себя аналогично).

1. Определяем ключевую метрику: возьмем *Retention*, поскольку изначально важно, чтобы клиенты, которые длительно не пользуются приложением – начали им пользоваться. Именно поэтому изначально бралась когорта из тех клиентов, кто пол года не пользуется приложением.
2. Ставим цели: что плохо, а что хорошо.

Для принятия или отклонения гипотезы будем использовать тест-t критерия Стьюдента, который рассчитывается с помощью функции библиотеки scipy

stats.ttest\_1samp(), в результате чего получим показатель p-value, который будем сравнивать с пороговым значением alpha=0.05.

На вход будем передавать значения полученных метрик и средние показатели текущего(отчетного) месяца. Проводим, как двусторонний тест, так и односторонний (чтобы понять, привели ли изменения к росту метрик).

1. Таким образом проверяем показатели MAU, DAU, DAU/MAU, Retention. Делаем вывод о статистической значимости. Пока делать вывод об успешности релиза рано.
2. Теперь проверяем остальные метрики: имеется ли рост (или отсутствует снижение или это снижение статистически незначимо) покупок (по количеству или сумме), а также отсутствует ли рост снятия наличных и переводов средств в сторонние банки в среднем по выборке, относительно контрольной группы и предыдущего периода.
3. Если метрики показали статистически значимый рост (в частности MAU, DAU/MAU, Retention) и нет статистически значимых негативных оттоков, мы признаем релиз успешным.

***Задание 2: Предложить контрольную группу для исследования*** *У PFM есть сервис Цели, в котором клиенты открывают цели и копят на них. В сервисе цели можно добавить фотографию своей цели и видеть прогрессбар своих накоплений. Сервис доступен на главном экране Сбербанк Онлайн. При открытии цели у клиента открывается один из видов счетов: сберегательный счет (можно пополнять и снимать), вклад Управляй (можно пополнять и снимать + есть неснижаемый остаток) или вклад Пополняй (можно только пополнять). При этом клиент видит свою цель именно в разделе целей, а не в разделе вкладов и счетов. Клиент может в разделе вклады и счета открыть один из перечисленных видов счетов/вкладов, но тогда у него не будет красивой карточки цели и прогрессбара.*

*Гипотеза: клиенты, которые открывают Цель копят больше денег, чем те, кто открывают сберегательный счет/управляй/пополняй не в рамках цели. Как подобрать контрольную группу (по каким критериям) и провести это исследование? В критерии может входить любая информация по банковским продуктам клиента и его  
операциям.*

Клиенты:

1. Разделяем клиентов на тех, кто вклад оформил и нет. Нам нужны те, кто оформил.

2. Разделение по возрастным категориям:

- до 18

- 18-30

- 30-60

- более 60

3. Разделение по территориальному признаку, например, городу проживания или территориальному банку. Взять несколько крупных городов и несколько небольших.

3.2. Можно разделить по наличию супругов/детей, но данная информация не всегда корректна в системах банка.

3.3. Добавим уровень заработной платы или доходу по карте (поскольку информация по заработной плате не всегда явная, если у клиента нет зарплатного проекта в Сбере).

4. Полученные сегменты разделяем по типу вклада:

- СберСчет

- Управляй

- Пополняй

5. И уже данные сегменты делим на тех, кто ставил себе цель и не ставил.

6. Уравниваем группы. Должно быть одинаковое количество людей в каждой категории.

7. Категории целей также должны быть разделены. Например, до 500 000 рублей, от 500 000 до 1 000 000 рублей, от 1 000 000 до 10 000 000 рублей. Предварительно посмотреть средние, медианные, квартильные показатели. (данный показатель нужен для анализа данных после отработки изначальной гипотезы).

8. Требуется разделение на то, с какой суммы, клиенты начали накопление: от 1000 рублей, от 50 000 рублей, от 200 000 рублей. Это примерно. Хорошо бы посмотреть на средние, квартильные и медианные значения вкладов и двигаться уже от данных цифр.

Для выборки, которая оформила вклад в качестве цели данные будут выглядеть так:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Количество клиентов | Возраст | Город | Тип вклада | Сумма первоначального взноса | Средняя ЗП |
| 1000 | 18-30 | Москва | СберСчет | 1000 — 10 000 | 50 000 — 100 000 |
| 1000 | 30-60 | Москва | СберСчет | 1000 — 10 000 | 50 000 — 100 000 |
| 1000 | 18-30 | Москва | Управляй | 1000 — 10 000 | 50 000 — 100 000 |
| 1000 | 30-60 | Москва | Управляй | 1000 — 10 000 | 50 000 — 100 000 |
| 1000 | 18-30 | Москва | Пополняй | 1000 — 10 000 | 50 000 — 100 000 |
| 1000 | 30-60 | Москва | Пополняй | 1000 — 10 000 | 50 000 — 100 000 |
| 1000 | 18-30 | Москва | СберСчет | 10 000 — 100 000 | 50 000 — 100 000 |
| 1000 | 30-60 | Москва | СберСчет | 10 000 — 100 000 | 50 000 — 100 000 |
| 1000 | 18-30 | Москва | Управляй | 10 000 — 100 000 | 50 000 — 100 000 |
| 1000 | 30-60 | Москва | Управляй | 10 000 — 100 000 | 50 000 — 100 000 |
| 1000 | 18-30 | Москва | Пополняй | 10 000 — 100 000 | 50 000 — 100 000 |
| 1000 | 30-60 | Москва | Пополняй | 10 000 — 100 000 | 50 000 — 100 000 |
| 1000 | 18-30 | Москва | СберСчет | 1000 — 10 000 | 100 000 — 300 000 |
| 1000 | 30-60 | Москва | СберСчет | 1000 — 10 000 | 100 000 — 300 000 |
| 1000 | 18-30 | Москва | Управляй | 1000 — 10 000 | 100 000 — 300 000 |
| 1000 | 30-60 | Москва | Управляй | 1000 — 10 000 | 100 000 — 300 000 |
| 1000 | 18-30 | Москва | Пополняй | 1000 — 10 000 | 100 000 — 300 000 |
| 1000 | 30-60 | Москва | Пополняй | 1000 — 10 000 | 100 000 — 300 000 |
| 1000 | 18-30 | Москва | СберСчет | 10 000 — 100 000 | 100 000 — 300 000 |
| 1000 | 30-60 | Москва | СберСчет | 10 000 — 100 000 | 100 000 — 300 000 |
| 1000 | 18-30 | Москва | Управляй | 10 000 — 100 000 | 100 000 — 300 000 |
| 1000 | 30-60 | Москва | Управляй | 10 000 — 100 000 | 100 000 — 300 000 |
| 1000 | 18-30 | Москва | Пополняй | 10 000 — 100 000 | 100 000 — 300 000 |
| 1000 | 30-60 | Москва | Пополняй | 10 000 — 100 000 | 100 000 — 300 000 |

Так выглядит выборка клиентов в рамках одного города среди тех, кто поставил себе цель. Для сравнения потребуется такая же выборка клиентов, которые цель не ставили. Такие итерации необходимо будет проделать по разным городам (Например: Москва, СПБ, Воронеж, Липецк, Калининград, Владивосток).

Примечание: возрастные группы до 18 лет и более 60 лет – опустила. Их можно добавить, в зависимости от широты исследования и доступности ресурсов.

Таким образом мы сможем сравнить в среднем, копят ли клиенты больше, используя цели. А далее мы можем посмотреть в разрезе разных категорий: возрастов, городов, средней зарплаты и стартового вклада: при каких показателях клиенты копят больше. Уже поняв ситуацию, можно будет делать сервис более персонализированным и предлагать оформить цели только тем, кому это будет действительно полезно.

Дополнительные параметры, которые можно использовать в сегментации клиентской базы и анализе данных:

- является ли клиент новым;

- совершает ли покупки на еженедельной/ежемесячной основе с применением карты;

- какой процент поступлений тратит;

- имеются ли кредитные/ипотечные продукты;

- имеются доп вклады.

**Теперь перейду к конкретным метрикам и методике их расчета.**

**Задание 3.**

Во второй таблице имеется две строки с одинаковым client\_id, при этом данное поле имеет указатель PRIMARY KEY, что говорит о том, что это должен быть уникальный идентификатор. Поэтому мы либо удаляем метку PK, либо удаляем строку. Но странным кажется, что клиент в какой-то момент меняет пол. Поэтому необходима перепроверка данных.

SQL-запрос:

SELECT

a.tb\_name AS ТБ,

d.gender AS Пол,

a.report\_dt AS Отчетная\_дата,

CASE

WHEN d.age <= 18 THEN 'Младше 18'

WHEN d.age <= 30 THEN 'От 18 до 30'

WHEN d.age <= 60 THEN 'От 30 до 60'

ELSE 'Старше 60'

END AS Группа\_возраста,

COUNT(DISTINCT a.client\_dk) AS Количество\_клиентов,

AVG(a.salary) AS Средний\_размер\_заработной\_платы

FROM

clnt\_aggr a

JOIN

clnt\_data d ON a.client\_dk = d.client\_dk

AND

a.report\_dt >= d.actual\_from\_dt AND a.report\_dt <= d.actual\_to\_dt

GROUP BY

ТБ,

Пол,

Группа\_возраста,

Отчетная\_дата;