Тестовое задание для продуктового аналитика мобильного и интернет-банка

[Задание 1 (SQL)](#_xfxq4lzhlwh)

[SQL1](#_ppqanjq0cix6)

[SQL2](#_nneds6qz0h31)

[Задание 2 (когортный анализ)](#_745pvirvsf62)

[Задание 3 (разметка)](#_3kprumsyc5gg)

[Задание 4 (портфолио)\*](#_pjadw4y8u8pi)

# Задание 1 (SQL)

Приведите в качестве решения скриншот с результатом, скрипт (или ссылку на запрос в среде, в которой решали задания, например <http://sqlfiddle.com/>), а также свои комментарии относительно предложенного решения.

## SQL1

Отобрать клиентов по г. Москва с суммарными остатками по клиенту от 20 000 на последнюю дату.

create table #Клиенты (client\_id varchar(1), FIO varchar (255) , Region varchar(50), account\_num int)

insert into #Клиенты values

('A','Иванов','Москва',111),

('A','Иванов','Москва',222),

('B','Петров','Иваново',333),

('C','Сидоров','Москва',444)

create table #Cчета ([Date] date, Summa\_USD money, Account\_num int)

insert into #Cчета values

('2012-01-01',15000,111),

('2012-02-01',10000,111),

('2012-02-01',5000,222),

('2012-03-01',30000,333),

('2012-04-01',20000,444)

**Решение**

Автоматически созданный замещающий текст:
SQL 
' ' Петров ' , 'Лваново' 
into Клиенты va1ues ( 'С' , 'сидоров' , 'Москва ' 
MYSQL - 
Vtew Sample Ft661e 
id 
С Clear 
(255) 
Text to OOL 
Region vapchap(se) 
1 cpeate 
З insept 
4 insept 
S insept 
б insept 
8 cpeate 
18 insept 
11 insept 
12 insert 
13 insept 
14 insept 
tab1e Клиенты (c1ient 
into Клиенты va1ues ( 'д' 
into Клиенты va1ues ( , ' Иванов ' 
into Клиенты va1ues ( ' 
vapchap(1), FIO varchar 
account 
пит 
' Иванов ' 
' Москва ' 
' Москва ' 
111); 
222); 
333); 
444) ; 
tab1e счета (0ate date, 
Summa 
USD int, kcount пит 
int) 
int), 
Regton 
Москва 
11 
SELECT 
c.c1ient id, 
C.FIO, 
. Region, 
SUM(s . Summa_USD) AS tota1 Ьа1апсе 
FROM Клиенты 
# объединяем таблицы Клиенты и счета 
INNER 301N счета s USIMG (kcount_num) 
#фильтруем только записи для клиентов из Москвы 
into 
into 
into 
into 
into 
счета 
счета 
счета 
счета 
счета 
va1ues( 
va1ues( 
va1ues( 
va1ues( 
va1ues( 
' 2912-91 
' 2912-92 
' 2912-92 
' 2912-ез 
' 2812-94 
-81 ' 
-81 ' 
-81 ' 
-91 ' , 2€еее,444) 
3rowser •k 
выбираем только записи последней датой из таблицы счета 
с . Region 
' Москва ' ДНО s . Date 
(SELECT rux(oate) FROM счета) 
13 WHERE 
14 GROLP ву с. id, 
. FIO, . Region 
16 # фильтруем клиентов, у которых суммарный баланс на последнюю дату составляет не менее 2€ вев 
17 SUM(s. summa_uso) 2€еее] 
But[6 Sc6ema 
cltent td 
с 
E6tt ” 
аил SQL » 
E6tt ” 
total Balance 
сидоров 

Этот запрос объединяет таблицы #Клиенты и #Cчета по полю account\_num, фильтрует только записи для клиентов из Москвы, выбирает только записи с последней датой из таблицы #Cчета и группирует данные по клиентам. Затем запрос отфильтровывает только тех клиентов, у которых суммарный баланс на последнюю дату составляет не менее 20 000.

SELECT

c.client\_id,

c.FIO,

c.Region,

SUM(s.Summa\_USD) AS total\_balance

FROM Клиенты c

# объединяем таблицы Клиенты и Cчета

INNER JOIN Cчета s ON c.account\_num = s.Account\_num

#фильтруем только записи для клиентов из Москвы,

# выбирает только записи с последней датой из таблицы Cчета

WHERE c.Region = 'Москва' AND s.Date = (SELECT MAX(Date) FROM Cчета)

GROUP BY c.client\_id, c.FIO, c.Region

# фильтруем клиентов, у которых суммарный баланс на последнюю дату составляет не менее 20 000

HAVING SUM(s.Summa\_USD) >= 20000

 select

Region,Date,

max(Date) over (partition by FIO order by Date) as max\_date,

client\_id,account\_num,FIO,Summa\_USD

 from

(

SELECT \*

FROM Клиенты a

JOIN Cчета b

USING (Account\_num)

) AS account\_base

 where         Region = 'Москва' and Summa\_USD >= 20000

order by Date

## SQL2

Таблица #oper хранит информацию о количестве произведенных операций на каждую календарную дату.

Вывести на каждую дату количества операций, совершенное с начала месяца по указанную дату включительно накопительным итогом.

create table #oper (date date, cnt int)

insert into #oper values

('2019-06-02', 1985),

('2019-06-03', 1577),

('2019-06-04', 1597),

('2019-06-05', 1468),

('2019-07-06', 82),

('2019-07-08', 1689),

('2019-07-09', 1556),

('2019-07-10', 1480),

('2019-07-11', 1405),

('2019-07-12', 1502)

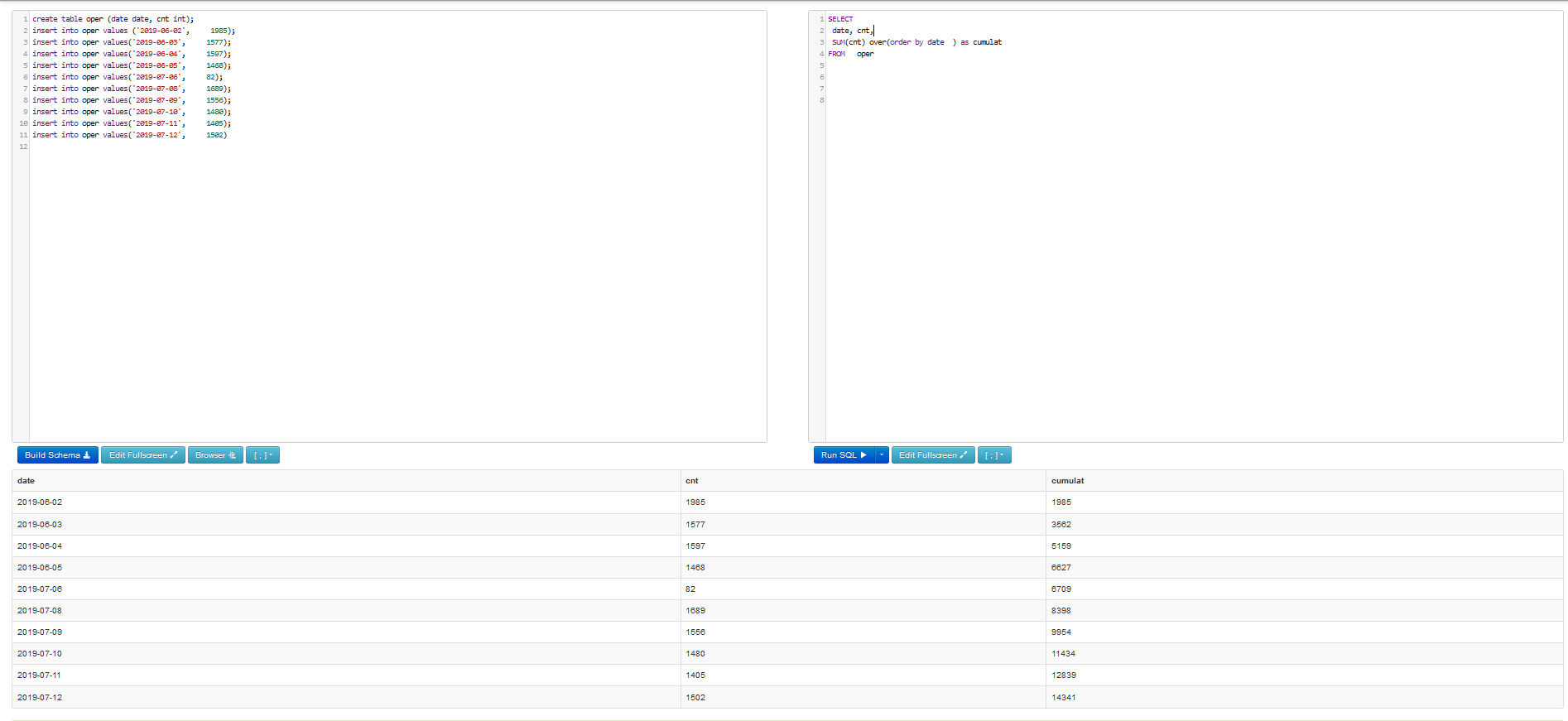
**Решение**

SELECT

date, cnt,

SUM(cnt) over(order by date ) as cumulat

FROM oper



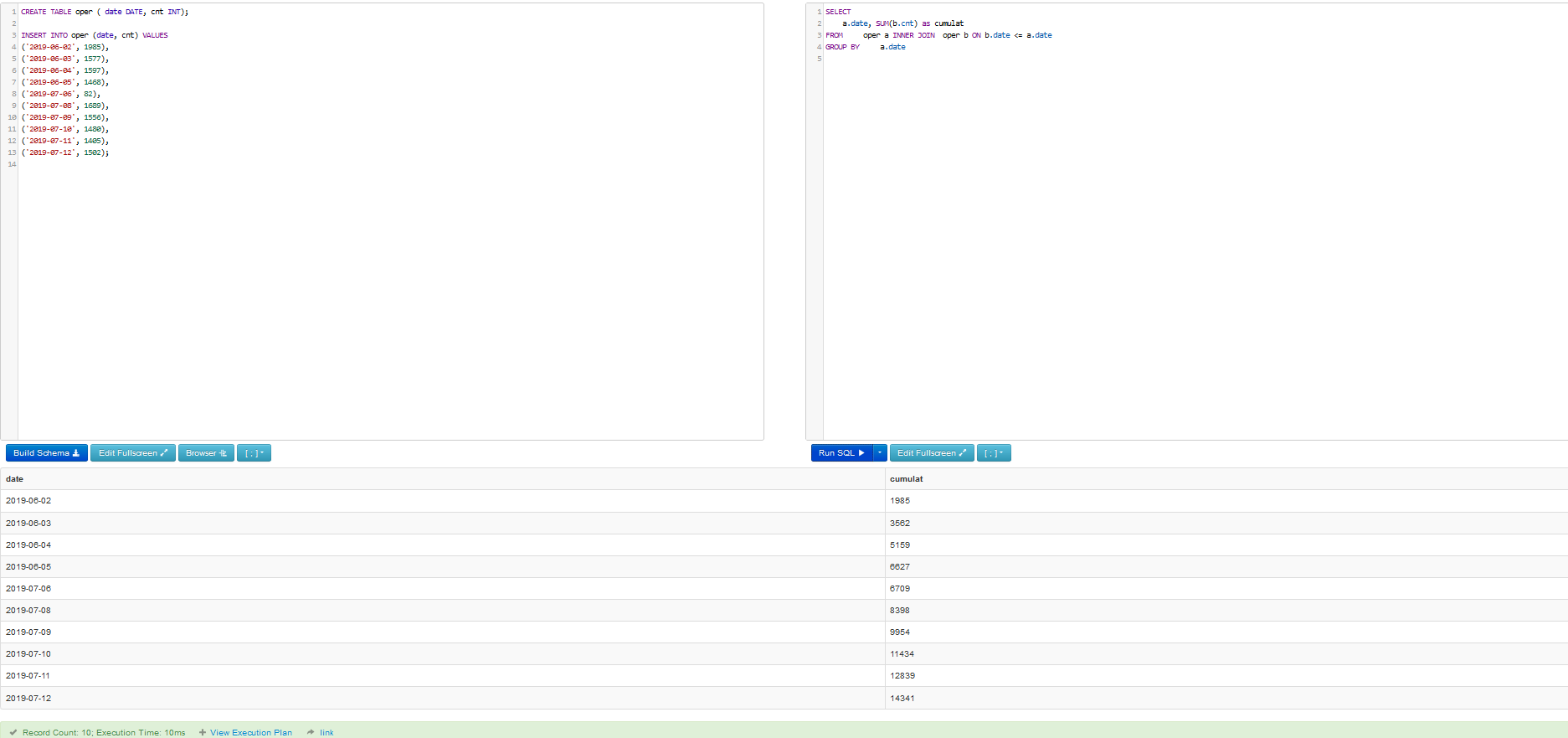
Вариант 2

SELECT

a.date, SUM(b.cnt) as cumulat

FROM oper a INNER JOIN oper b ON b.date <= a.date

GROUP BY a.date



Этот запрос объединяет таблицы oper a и oper b по полю date, где происходит суммирование кол-ва операций с начала месяца по указанную дату включительно накопительным итогом, по условию b.date <= a.date . Затем группировка по дате.

## SQL3

Даны месячные срезы сегментов клиентов, нужно получить по каждому клиенту периоды действия каждого сегмента.

create table #segment ([date] date, ClientID varchar(6),SegmentID int)

insert into #segment values

('2018-01-31' ,'A11111', 2),

('2018-02-28' ,'A11111', 2),

('2018-03-31' ,'A11111', 1),

('2018-04-30' ,'A11111', 1),

('2017-11-30' ,'B22222', 1),

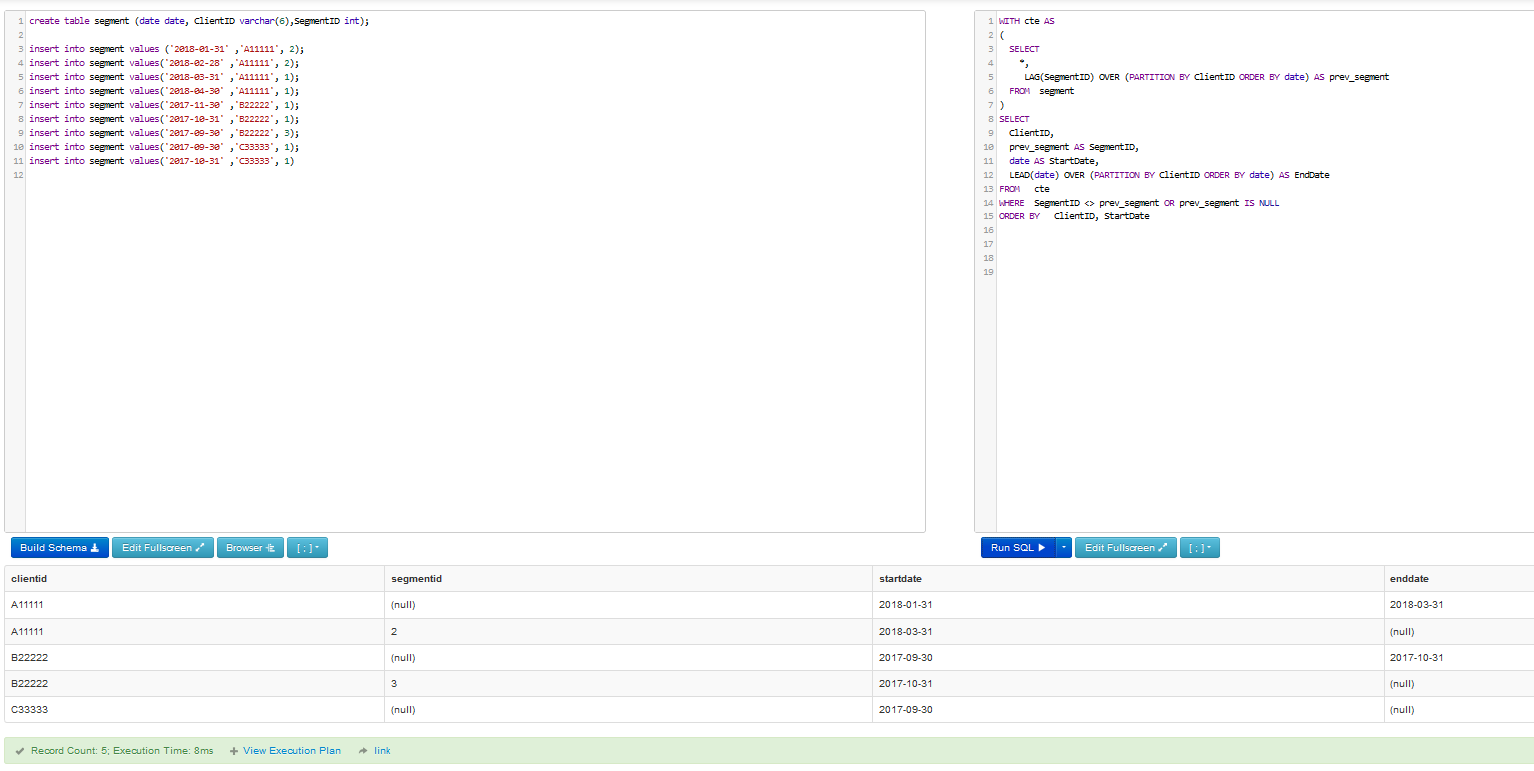
('2017-10-31' ,'B22222', 1),

('2017-09-30' ,'B22222', 3),

('2017-09-30' ,'C33333', 1),

('2017-10-31' ,'C33333', 1)

**Решение**



В этом запросе мы сначала создаем общую таблицу выражений (CTE), которая содержит все данные из таблицы #segment, а также столбец prev\_segment, который содержит значение предыдущего сегмента для каждой записи, отсортированных по дате. Затем мы выбираем из CTE только те строки, где произошло изменение сегмента, используя условие SegmentID <> prev\_segment OR prev\_segment IS NULL. Наконец, мы вычисляем даты начала и конца каждого периода, используя функцию LEAD() для получения даты следующей записи в таблице. Результат сортируется по ClientID и StartDate.

Запрос вернет следующие столбцы: ClientID - идентификатор клиента, SegmentID - идентификатор сегмента, StartDate - дата начала периода, и EndDate - дата окончания периода. Если сегмент для клиента не изменился, то значение SegmentID и EndDate будут равны соответственно предыдущему сегменту и следующей записи в таблице. Если сегмент продолжается до конца исходных данных, то EndDate будет равно NULL.

WITH cte AS

(

SELECT

\*,

LAG(SegmentID) OVER (PARTITION BY ClientID ORDER BY date) AS prev\_segment

FROM segment

)

SELECT

ClientID,

prev\_segment AS SegmentID,

date AS StartDate,

LEAD(date) OVER (PARTITION BY ClientID ORDER BY date) AS EndDate

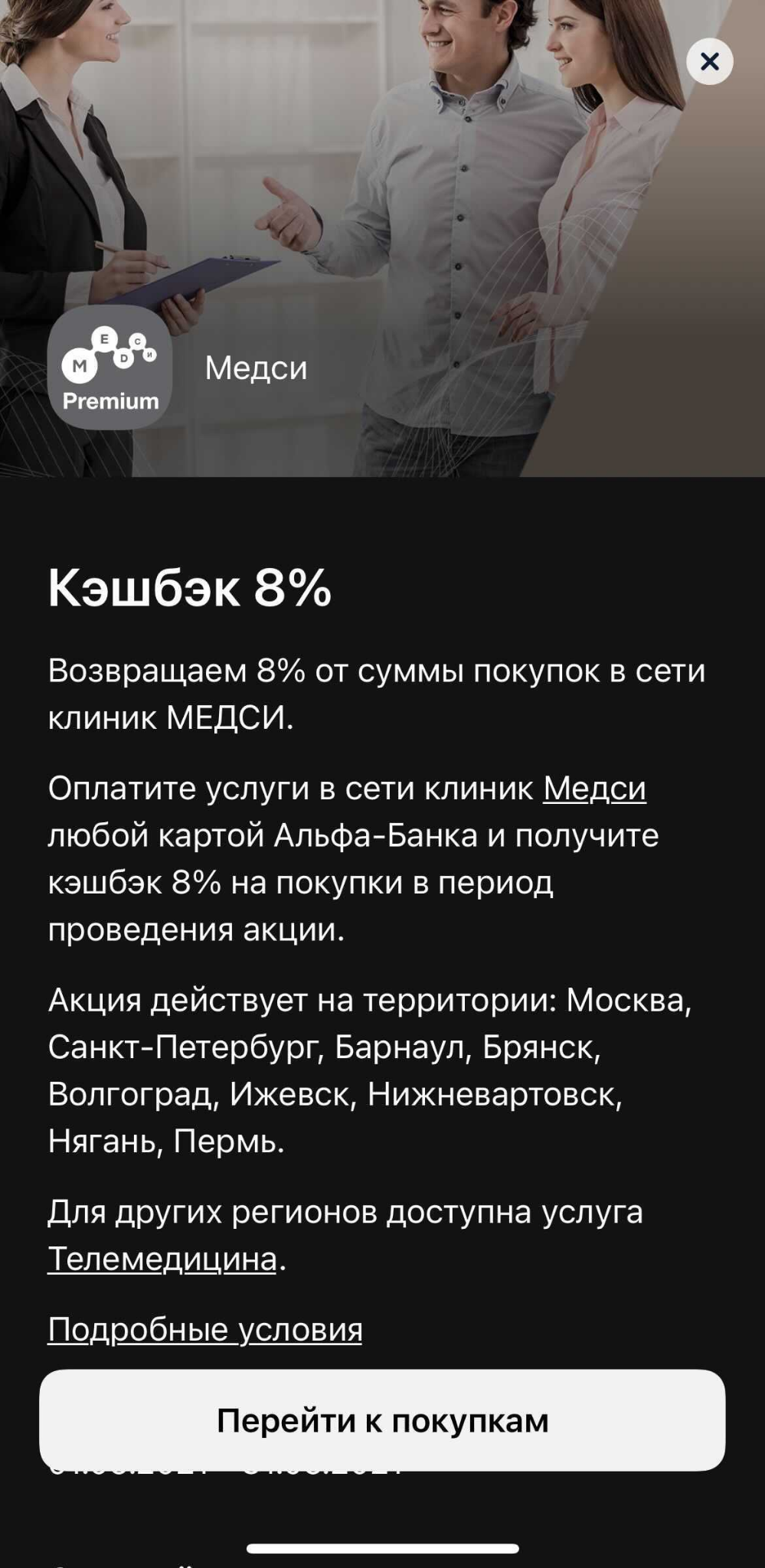
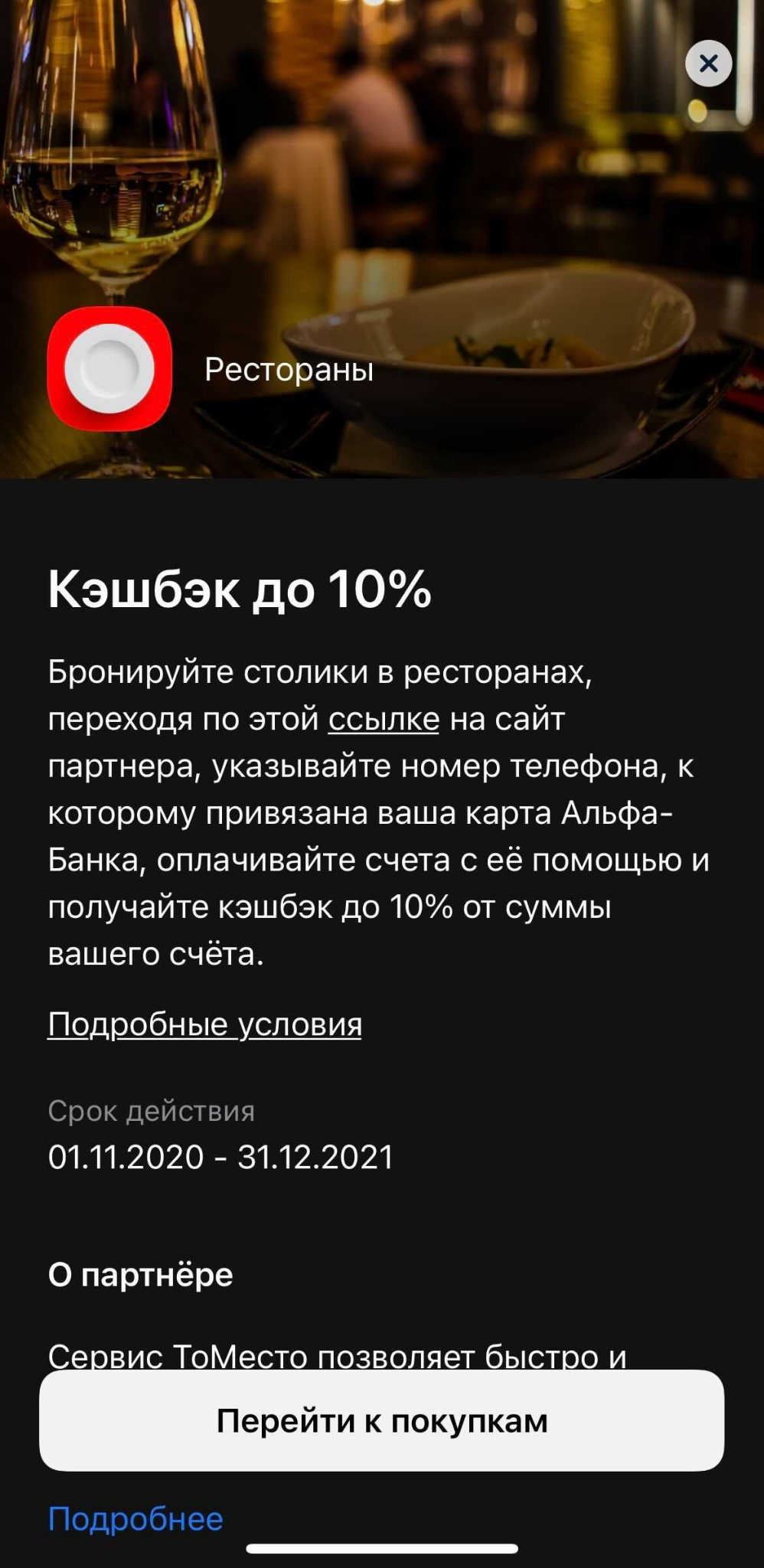
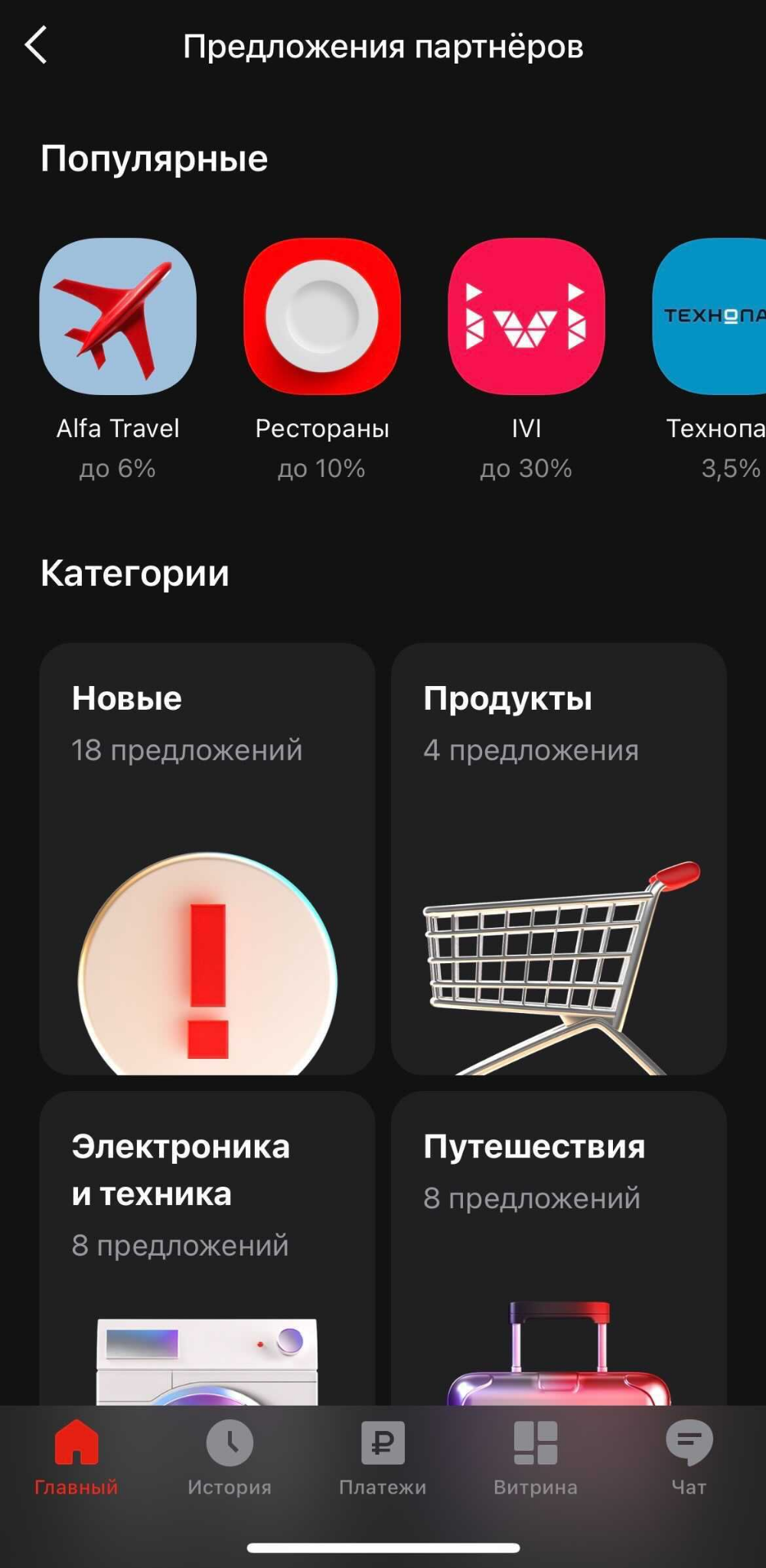
FROM cte

WHERE SegmentID <> prev\_segment OR prev\_segment IS NULL

ORDER BY ClientID, StartDate

# Задание 2 (разметка и метрики)

Партнерская витрина - это набор предложений партнеров банка, которые предоставляют различного рода скидки нашим клиентам.



На витрину можно перейти разными способами: через внешние и внутренние коммуникации (SMS, email, push, баннеры/уведомления внутри приложения).

Партнеры могут предоставлять данные о совершенных нашими клиентами конверсиях (транзакциях).

* Какие наборы метрик для оценки эффективности партнерской витрины вы можете предложить?

Что потребуется сделать?

* Определиться с системой трекинга данных.
* Подготовить ТЗ на установку SDK аналитической системы для разработчиков.
* Подготовить карту событий для разметки в приложении.
* Внедрить аналитику в приложение.
* Протестировать сбор данных.

Набор метрик для оценки эффективности партнерской витрины:

1. Активность пользователей.

- Количество посетителей партнерской витрины (VP)

- Количество конверсий (транзакций) на партнерской витрине (CP)

- Количество повторных посещений партнерской витрины (RPV)

- Количество просмотров товаров/услуг на партнерской витрине (PV)

- Соотношение количества посетителей и конверсий на партнерской витрине (CR)

2.Доходность

- Средний чек на партнерской витрине (AOV)

- Срок жизни клиентов, привлеченных через партнерскую витрину (CLV)

- Расходы на привлечение клиентов через партнерскую витрину (CAC) и их окупаемость (ROI)

3. Аудитория и паттерны поведения.

Какие пользователи приходят в приложение и каковы их паттерны взаимодействия с продуктом.

* Опишите возможный набор систем и способы их интеграции, которые позволят осуществлять трекинг полного пути пользователя и его конверсионных действий.

Каждое приложение — это новый проект для анализа данных.

Сегодня ни один сервис не может закрыть все задачи сквозной аналитики по анализу данных в мобильном приложении — приложения слишком сложны.

1. Выбираем систему трекинга данных в мобильных приложениях, такие как Google Analytics, Яндекс.Метрика, Google Firebase , Amplitude , AppsFlyer

и другие, которые позволяют отслеживать поведение пользователей на сайте и на витрине партнеров.

2. [Google Tag Manager](https://www.youtube.com/watch?v=C5uI4aOlLOg)  упрощает процесс установки кодов отслеживания и позволяет быстро вносить изменения в коды.

3. CRM-системы, в которых хранятся данные о клиентах и их воронки взаимодействия с компанией.

4. API (SDK) для интеграции с системами партнеров и получения данных о конверсиях и взаимодействии клиентов с их витринами.

* Какие сырые данные нужны и в каком виде для полноты картины?

1. Информация о внешних и внутренних коммуникациях с клиентами (SMS, email, push-уведомления, баннеры/уведомления внутри приложения).

2. Данные о визитах клиентов на сайт компании и на партнерские витрины.

3. Информация о конверсиях (транзакциях) на партнерских витринах.

4. Данные о товарах/услугах, которые были просмотрены на сайте и на партнерских витринах.

5. Информация о поведении клиентов на сайте и на партнерских витринах (например, время, проведенное на странице, просмотренные страницы, взаимодействие с элементами интерфейса, модель тлф в сессии и т.п. ).

6. расходы из рекламных кабинетов.

* Какая разметка коммуникаций и действий пользователя вам понадобится для отслеживания всего пути клиента?

Для отслеживания всего пути клиента нам понадобится разметка, которая позволит идентифицировать и отслеживать каждое действие пользователя, начиная с того момента, когда он взаимодействует с внешними коммуникациями (SMS, email, push, баннеры/уведомления внутри приложения) и до того момента, когда он совершает конверсионное действие (эвент - перейти к покупкам ) на партнерской витрине.

Этапы :

* Определяем персонажи пользователей,
* описываем пользовательские истории (User Story),
* составляем карту событий пользователя (Customer Journey Map)
* формируем технические требования к приложению (каким должно быть приложение, что оно должно уметь и как это будет работать с аналитикой)

Разметка должна содержать информацию о:

* типе коммуникации (SMS, email, push, баннеры/уведомления внутри приложения),
* времени и дате отправки,
* идентификаторе пользователя, который получил коммуникацию,
* а также о том, какой партнерский материал был показан на витрине,
* какие конверсионные действия были совершены и когда.

Таким образом, для каждого пользователя нам понадобится записывать все события, связанные с его путем от взаимодействия с коммуникацией до совершения конверсионного действия на партнерской витрине.

Это позволит нам :

* отслеживать все этапы пути пользователя,
* выявлять успешные комбинации коммуникаций и материалов на витрине,
* а также определять, какие коммуникации и материалы на витрине могут быть оптимизированы для улучшения конверсии (UI/UX).

# Задание 3 (когортный анализ)

Приведите в качестве решения скриншот с результатом, а также свои комментарии относительно предложенного решения (с указанием среды, в которой решали задания).   
в приложенном файле хранятся данные об активности клиентов с сентября 2021 года по февраль 2022  
Необходимо на основе этих данных провести когортный анализ  
Данные: [data\_test.csv](https://raw.githubusercontent.com/Propsih/test/main/data_test.csv)  
Описание полей:

CLIENT\_ID - идентификатор клиента;

VALUE\_DAY - отчетная дата;

DIGITAL\_30\_CNT - кол-во входов в мобильное приложение за последние 30 дней на отчетную дату;

TRAN\_ACTIVE\_30\_CNT - кол-во транзакций в мобильном приложении за последние 30 дней на отчетную дату;

OPER\_ACTIVE\_30\_CNT - кол-во операций (транзакции+нефинансовые операции) в мобильном приложении за последние 30 дней на отчетную дату

* постройте визуальное представление удержания клиентов в статусе:
  + активных клиентов банка (наличие строки для клиента на отчетную дату);
  + диджитал активных (активных в мобильном приложении: наличие хотя бы одного входа);
  + операционно активных;
  + транзакционно активных;
* какие выводы можно сделать на основе этих данных?;
* если построить проникновение транзакционно активных клиентов в когорты диджитал активных клиентов, то мы увидим, что существенная часть клиентов использует мобильное приложение, но не совершает транзакции (платежи и переводы) в нем. Какие гипотезы такого поведения можно предположить? Как их можно проверить ?;
* как можно исправить эту ситуацию и увеличить проникновение транзакционно активных клиентов в диджитал активных?

**Решение На Гитхабе - https://github.com/Bear2232/Alfa\_b**

# Задание 4 (портфолио)

Сделайте мини-портфолио по 2-3 сделанным вами ранее отчетам/исследованиям (чувствительную информацию можете упустить, если прикладываете скрины отчетов - “заблюрить” значения):

* какая цель и задачи стояли, какие пользователи результата,
* для исследований - какие методы и инструменты были выбраны,
* опишите источники данных, требовались ли доработки в источниках для обеспечения полноты данных и для матчинга,
* какие инструменты использовались для выгрузки данных из источников и для их объединения, для автоматизации,
* каким образом производился матчинг данных разных систем-источников,
* какие средства визуализации использовались и каким образом выглядел конечный результат.

|  |  |
| --- | --- |
| Исследование 1 |  |
| какая цель и задачи стояли, какие пользователи результата, | ЦА – Кафе\бары ,  Цель : отраслевая витрина по всем платформам телесмотрения;  - анализ распределения пользователей по количеству просмотренного контента.  - теле смотрение у пользователей, которые смотрят много от пользователей, которые смотрят мало контента. |
| для исследований - какие методы и инструменты были выбраны, |  |
| опишите источники данных, требовались ли доработки в источниках для обеспечения полноты данных и для матчинга, | PowerBI ,csv файл |
| какие инструменты использовались для выгрузки данных из источников и для их объединения, для автоматизации, | PowerBI |
| каким образом производился матчинг данных разных систем-источников, | 1 источник данных |
| какие средства визуализации использовались и каким образом выглядел конечный результат. | YDL ,  Витрины : годовых и месячных показателей ,  Тайминг телесмотрения по платформам ,временным интервалам  результат представлен ниже |

Автоматически созданный замещающий текст:
В2В TV 2021-2022 смотрибельность контента 
Годовые показатели 
1 755 
Теле 
чно 
Месячные показатели 
1 755 
тд писок 
113301 
тд лисок 
113301 
- копия 
длит-ть ф«ьютр. кт«гента, 
142 743 706 
по плате.“ ам (сек) 
длит-ть ф«ьютр. кт«гента, 
142 743 706 
AVG 'тсьютра 
1 259,86 
абм 
AVG 'тсьютра 
1 259,86 
Сравнение теле смотрения пользователей (много/мало) 
АВ11 
AB1cLy8'esq.:nJ2sEOZ 
Авлендтх,ыниютањ.чн 
АВ EWLgZ3cUrFNM7WZYDn 
АВ FtANJfFvgxtBnXEUZp 
good user 
10367' 
11scm 
164 572 
114839 
417363 
431S2 
142698 
и 108 
18185 
176888 
124680 
12372 

Автоматически созданный замещающий текст:
Теле 
Платф'* ма_смот рение 
rttv•h 
по дням ( шт,'день) 
2001 по 
сьштрения те дыми недели 
77 ssu 
9 S16S83 
673080 
73S82jg 
143609 
9810133 
и 83800 
1438ss 
149223 
ь дззм8 
918 дм 
333“ т 
778 
виз 
33 
170220 
25.05.20 

|  |  |
| --- | --- |
| Исследование 2 |  |
| какая цель и задачи стояли, какие пользователи результата, | Анализ поведения пользователей в моб. Приложении-видеоаналитика.  Цель :  -DAU  показателей эвентов событий по дням |
| для исследований - какие методы и инструменты были выбраны, | Маркировка эвентов событий по дням |
| опишите источники данных, требовались ли доработки в источниках для обеспечения полноты данных и для матчинга, |  |
| какие инструменты использовались для выгрузки данных из источников и для их объединения, для автоматизации, |  |
| каким образом производился матчинг данных разных систем-источников, |  |
| какие средства визуализации использовались и каким образом выглядел конечный результат. | Tableau:   * Диаграмма распределения количества событий по типу события , возможность среза - со скроллингом по дням * Количество взаимодействий по дням * пользователи по источникам |

