

Goals:

Find out these numbers in each month from April 2020 to April 2021.

- Total # of drop-off users
- Total # of returning users
- Total # of new users

เนื่องจากระบบของเราเป็น transaction based จึงจำเป็นต้องกำหนด churn period เพื่อใช้นิยามของ drop off users และ returning users โดยเราได้กำหนดนิยามต่างๆ ดังนี้

- Total # of drop-off users in month M:
 - จำนวน users ทั้งหมดที่ไม่ได้ใช้งานระบบนานกว่า **churn period** (Days) ใน month M
- Total # of returning users in month M::
 - drop-off users ที่กลับมาใช้งานใน month M
 - หาก user มีการ drop-off และกลับมาใช้งานในเดือนเดียวกัน จะถือว่าในเดือนนั้นมี # of returning users และ # of drop-off users เพิ่มขึ้นมา 1 user
- Total # of new users in month M:
 - users ของเราที่เพิ่งมี transaction แรกในระบบเราใน month M

***churn periods** เนื่องจากเราค้นพบว่า user type แต่ละประเภทนั้นมีลักษณะการใช้งานระบบเราแตกต่างกันอย่างชัดเจนเราจึงให้กำหนดให้แต่ละ users type จะมี churn period ดังนี้

- A = 360 วัน
- B = 360 วัน
- C = 120 วัน
- D = 260 วัน

Input:

1. daily_transactions.csv:

โดยไฟล์ที่ได้รับจะมีทั้งหมด 3 columns ดังนี้

- _id: users id
- event_date: วันที่ users คนนั้นมีการสร้าง transaction ในระบบ
- usertype: ประเภทของ users ประกอบด้วย 4 ค่าได้แก่ A,B,C,D ซึ่ง 1 users id จะมีเพียงแค่ 1 usertype เท่านั้นหากพบ users ที่มีมากกว่า 1 usertype ให้ตัด user คนนั้นออกจากการคำนวณ

Expected Output:

1. File CSV ที่ประกอบด้วย 4 columns

month	dropoff_users	returning_users	new_users
April 2020	23127	10293	3810
May 2020	39494	12752	3278
...			
April 2021	81237	19804	8921

****** ตัวเลขที่เห็นเป็นเพียงค่า mock เท่านั้นไม่ใช่ผลเฉลยที่มีไว้ cross-check แต่อย่างใด

2. source code ที่ใช้ในการคำนวณ เพื่อให้ได้มาซึ่งไฟล์ CSV ในข้อ 1
3. วิธีการรันของ source code ข้อที่ 2 เพื่อให้ได้มาซึ่งไฟล์ CSV ในข้อ 1

In addition, please prepare a visualization for an audience of mixed technical abilities on your results and thinking process.