Portfolio Projects.

CODEFLIX USER CHURN

Analyse data with SQL

Ameen Azmi

DATA ANALYST

Table of Content

Portfolio Projects.

1. Get familiar with data

DATA ANALYST

1.1 Codeflix and the data

Codeflix, a promising startup in the streaming video industry, is keen on understanding and improving its user retention. One of the key metrics to gauge user loyalty and business health is the churn rate, which indicates the percentage of subscribers who discontinue their service over a given period.

To effectively measure and analyze the churn rate, we will utilize the data from the company's subscriptions table. This table include four columns in it:

- id - belong to subscriber
- subscription_start - the start of subscription date
- subscription_end - the end of subscription date
- segment - which segment belong to customer subscription

subscriptions	
name	**type**
id	INTEGER
subscription_start	TEXT
subscription_end	TEXT
segment	INTEGER
Rows: 2000	

1.2 How many months has the company been operating?

We are using first date subscription and last date subscription to see how many months the company has been operating.

Codeflix requires a minimum subscription length of 31 days, so a user can never start and end their subscription in the same month.

```
SELECT  MIN(subscription_start),
MAX(subscription_start)
FROM subscriptions;
```

Query Results	
MIN(subscription_start)	MAX(subscription_start)
2016-12-01	2017-03-30

1.3 What segments of user exists?

We are using group by to see what segments are ready for customers.

There are two group of segment exists in Codeflix. Later we will be using this segments to calculate Churn rate.

Query Results	
segment	subscribers
30	1000
87	1000

```
SELECT segment, COUNT(segment)
as subscribers
FROM subscriptions
GROUP BY segment;
```

Portfolio Projects.

2. What is the overall Churn trend since company started?

DATA ANALYST

2.1 Churn Rate

Our final goal is to compare this two group of segments Churn Rate.

If the churn rate is low, meaning fewer subscribers are leaving the service, Codeflix will likely see subscriber growth. Conversely, if the churn rate is high, indicating many subscribers are discontinuing their service, Codeflix may struggle to grow its subscriber base.

To calculate the churn rate:

Churn rate = cancellations/ active subscribers

2.2 Calculate Churn Rate

```
with months as
(select
'2017-01-01' as first_day,
 '2017-01-31' as last_day
 UNION
 select
'2017-02-01' as first_day,
 '2017-02-28' as last_day
UNION
select
'2017-03-01' as first_day,
 '2017-03-31' as last_day),
cross_join as
 (select *
  from subscriptions
 cross join months),
status as (
 select id,
 first_day as month,
 case
   when   (subscription_start   <
first_day) and
  (subscription_end  >  first_day
or subscription_end is null)
 then 1 else 0
 end as is_active,
  case  when  (subscription_end
between first_day and last_day)
 then 1 else 0
 end as is_canceled
from cross_join),
```

```
status_aggregate as (
 select month,
 sum(is_active) as sum_active,
       sum(is_canceled)       as
sum_canceled
 from status
 group by month)
 select month,
  1.0*sum_canceled/sum_active  as
churn_rate
 from status_aggregate;
```

Within this 3 months, amount of cancellation from Codeflix subscribers keep increase.

Query Results	
month	**churn_rate**
2017-01-01	0.161687170474517
2017-02-01	0.189795918367347
2017-03-01	0.274258219727346

Portfolio Projects.

3. Compare the Churn rate between user segments

DATA ANALYST

3.1 Calculate Churn Rate Per Segment

```
with months as
(select
'2017-01-01' as first_day,
 '2017-01-31' as last_day
 UNION
 select
'2017-02-01' as first_day,
 '2017-02-28' as last_day
UNION
select
'2017-03-01' as first_day,
 '2017-03-31' as last_day
 ),
cross_join as (select *
  from subscriptions
  cross join months),
```

```
status as (
 select id,
 first_day as month,
 case
 when (subscription_start <
first_day) and
 (subscription_end > first_day or
subscription_end is null) and
(segment=87)
 then 1 else 0
 end as is_active_87,
 case
 when (subscription_start <
first_day) and
 (subscription_end > first_day or
subscription_end is null)
```

```
 and (segment=87) then 1 else 0
 end as is_active_30,
 case when (subscription_end
between first_day and last_day)
and (segment=87)
 then 1 else 0
 end as is_canceled_87,
 case when (subscription_end
between first_day and last_day)
and (segment=30)
 then 1 else 0
 end as is_canceled_30
from cross_join),
status_aggregate as (
 select
 month,
```

```
sum(is_active_87)                as
sum_active_87,
       sum(is_active_30)         as
sum_active_30,
      sum(is_canceled_87)        as
sum_canceled_87,
      sum(is_canceled_30)        as
sum_canceled_30
 from status
 group by month)
 select month,

1.0*sum_canceled_87/sum_active_
87 as churn_rate_87,

1.0*sum_canceled_30/sum_active_
30 as churn_rate_30
 from status_aggregate;
```

3.2 Compare Churn Rate by Segment

As we can see Churn Rate for segment 87 is higher than Churn Rate for segment 30. Subscribers for segment 87 tend to leave with Churn Rate keep increase every month.

However, subscribers in segment 30 rather to stay and enjoy the subscription than leaving. Even though their Churn Rate keep increase every month.

Query Results		
month	churn_rate_87	churn_rate_30
2017-01-01	0.251798561151079	0.079136690647482
2017-02-01	0.32034632034632	0.0822510822510823
2017-03-01	0.485875706214689	0.15819209039548