

# Churn Rates of Codeflix

Hirofumi Fukuzawa

# The aims of whole analysis

## 1. Get familiar with the company.

- How many months has the company been operating?
- Which months do you have enough information to calculate a churn rate?
- What segments of users exist?

## 2. What is the overall churn trend since the company started?

## 3. Compare the churn rates between user segments.

- Which segment of users should the company focus on expanding?

# 1. Get familiar with Codeflix

- **How many months has the company been operating?**
  - The company has operated 4 months.
  - The oldest data recorded was at 2016-12-01, while the newest one was 2017-03-31 as a result.

MIN(subscription_start)	MAX(subscription_start)	MAX(subscription_end)
2016-12-01	2017-03-30	2017-03-31

```
--How many months has the company been operating?  
  
SELECT MIN(subscription_start),  
       MAX(subscription_start),  
       MAX(subscription_end)  
FROM subscriptions;
```

# 1. Get familiar with Codeflix

- **Which months do you have enough information to calculate a churn rate?**
  - 2017.1~2017.3 are relevant to calculate churn rates. The reason of it is that the oldest data of subscription\_end was 2017-01-01.
  - Churn rate is the number of users who cancel during the month divided by the number of users who exist before the beginning of the month.

MIN(subscription_end)
2017-01-01

- **What segments of users exist?**
  - There are two segment : 87 and 30.

segment
87
30

```
--Which months do you have enough information to  
calculate a churn rate?
```

```
SELECT MIN(subscription_end)  
FROM subscriptions;
```

```
--What segments of users exist?
```

```
SELECT DISTINCT segment  
FROM subscriptions;
```

## 2. What is the overall churn trend since the company started?

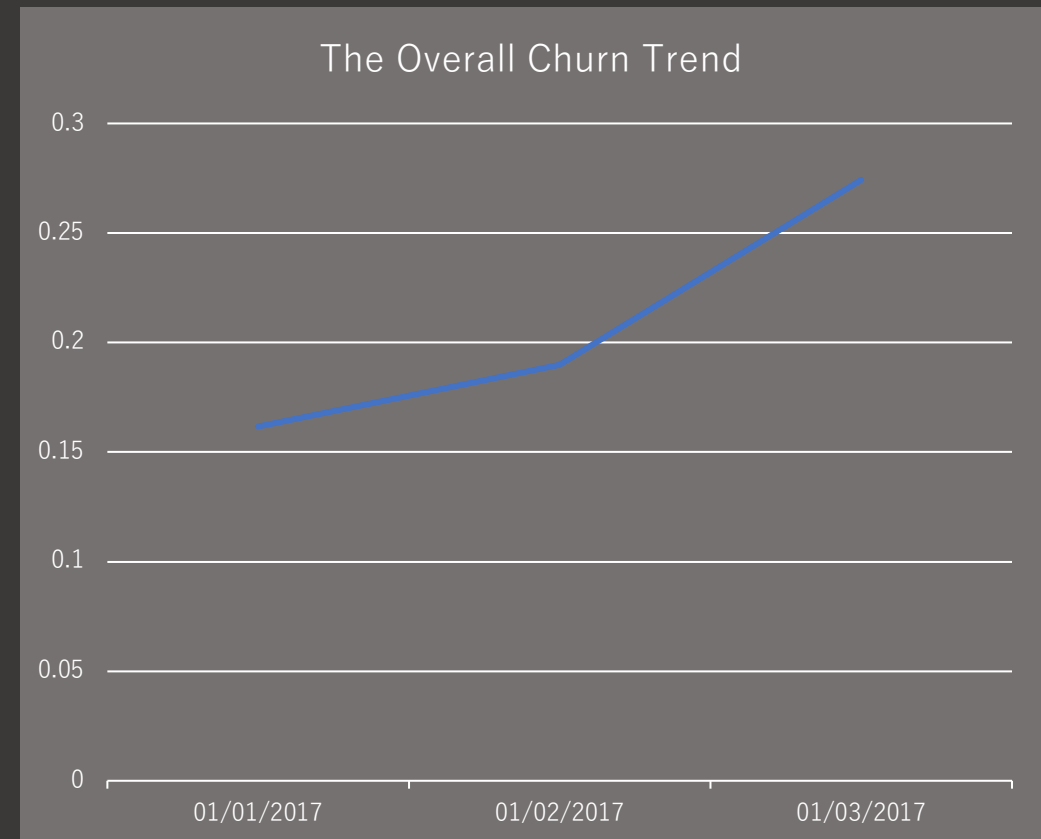
The overall churn rate increased over 3 months.

Especially, there was a dramatic rise between February and March.

This trend is not good for Codeflix.

But the reason of increase the rate might be just that it was the beginning of the service running.

month	churn_rate
2017-01-01	0.161687170474517
2017-02-01	0.189795918367347
2017-03-01	0.274258219727346



## 2. What is the overall churn trend since the company started?

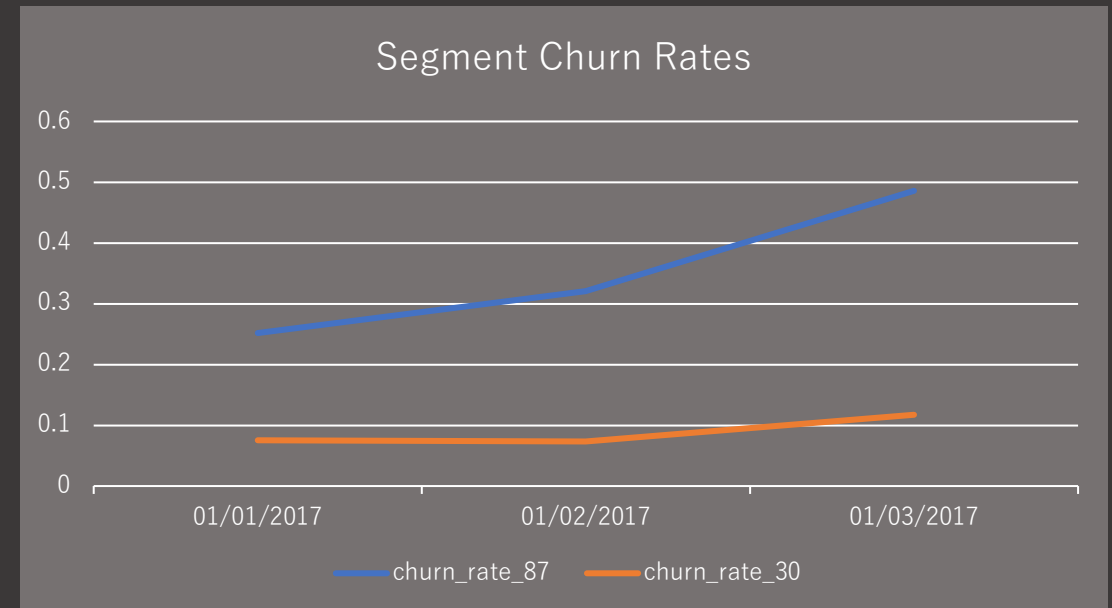
```
WITH months AS(
    SELECT
        '2017-01-01' AS 'first_day',
        '2017-01-31' AS 'last_day'
    UNION
    SELECT
        '2017-02-01' AS 'first_day',
        '2017-02-28' AS 'last_day'
    UNION
    SELECT
        '2017-03-01' AS 'first_day',
        '2017-03-31' AS 'last_day'),
    cross_join 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 active,
            SUM(is_canceled) AS canceled
        FROM status
        GROUP BY month)
SELECT
    month,
    1.0 * canceled / active AS churn_rate
FROM status_aggregate;
```

### 3. Compare the churn rates between user segments.

Both churn rates rose in the 3 months, but segment 87 had more or less 4 times as high churn rate as segment 30.

While there was a small dip of rate in segment 30 in February, the rate of segment 87 had increased throughout 3 months.

We have to know why the dip can be seen in the February on the purpose to find somewhat small good point, also why 87 had higher churn rate in order to find some causes of quit.



month	churn_rate_87	churn_rate_30
2017-01-01	0.251798561151079	0.0756013745704467
2017-02-01	0.32034632034632	0.0733590733590734
2017-03-01	0.485875706214689	0.11731843575419

# 3. Compare the churn rates between user segments.

--What is the overall churn trend since the company started?

```
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(
        (segment = 87)
        AND((subscription_start < first_day)
            AND (subscription_end > first_day
                OR subscription_end IS NULL))
        ) THEN 1
    ELSE 0
    END AS is_active_87,
```

```

CASE
WHEN(
    (segment = 30)
    AND((subscription_start < first_day)
        AND(subscription_end > first_day
            OR subscription_end IS NULL))
    ) THEN 1
ELSE 0
END AS is_active_30,
CASE
WHEN(
    (segment = 87)
    AND(subscription_end BETWEEN
        first_day AND last_day)
    ) THEN 1
ELSE 0
END AS is_canceled_87,
CASE
WHEN(
    (segment = 30)
    AND(subscription_end BETWEEN
        first_day AND last_day)
    ) 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
GROUP BY month;
```