codecademy

Codeflix - Calculating Churn Rates

Learn SQL from Scratch
Matt Kerr
Cohort 2018/12/03

Table of Contents

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 THE COMPANY.

We have been provided with customer subscription data from the company Codeflix.

Our goal is to evaluate their monthly churn rates and search for any trends that might inform their growth strategy.

First, let's just take a look at the subscriptions table they've given us to get a general sense of how it's structured.

```sql
1  SELECT *
2  FROM subscriptions
3  LIMIT 15;
```

Query Results			
id	subscription_start	subscription_end	segment
1	2016-12-01	2017-02-01	87
2	2016-12-01	2017-01-24	87
3	2016-12-01	2017-03-07	87
4	2016-12-01	2017-02-12	87
5	2016-12-01	2017-03-09	87
6	2016-12-01	2017-01-19	87
7	2016-12-01	2017-02-03	87
8	2016-12-01	2017-03-02	87
9	2016-12-01	2017-02-17	87
10	2016-12-01	2017-01-01	87
11	2016-12-01	2017-01-17	87
12	2016-12-01	2017-02-07	87
13	2016-12-01	∅	30
14	2016-12-01	2017-03-07	30
15	2016-12-01	2017-02-22	30

1. GET FAMILIAR WITH THE COMPANY. - CONT'D

How many months has the
company been operating?

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

1. GET FAMILIAR WITH THE COMPANY. - CONT'D

How many months has the company been operating?

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

```sql
SELECT MIN(subscription_start) AS 'sub_start_min',
    MAX(subscription_start) AS 'sub_start_max',
    MIN(subscription_end) AS 'sub_end_min',
    MAX(subscription_end) AS 'sub_end_max'
FROM subscriptions;
```

Query Results			
sub_start_min	sub_start_max	sub_end_min	sub_end_max
2016-12-01	2017-03-30	2017-01-01	2017-03-31

The earliest subscription_start date is 2016-12-01 and the latest subscription_end date is 2017-03-31. Based on the data we have been provided, we can see that Codeflix has been operating from at least December 1, 2016 until March 31, 2017.

1. GET FAMILIAR WITH THE COMPANY. - CONT'D

How many months has the company been operating?

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

```sql
1   SELECT MIN(subscription_start) AS 'sub_start_min',
2       MAX(subscription_start) AS 'sub_start_max',
3       MIN(subscription_end) AS 'sub_end_min',
4       MAX(subscription_end) AS 'sub_end_max'
5   FROM subscriptions;
```

Query Results

sub_start_min	sub_start_max	sub_end_min	sub_end_max
2016-12-01	2017-03-30	2017-01-01	2017-03-31

Monthly churn rate is equal to:

(the number of customers who cancel their subscription in a given month)
÷
(the number of subscribed customers at the start of that month)

1. GET FAMILIAR WITH THE COMPANY. - CONT'D

How many months has the
company been operating?

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

```
1   SELECT MIN(subscription_start) AS 'sub_start_min',
2       MAX(subscription_start) AS 'sub_start_max',
3       MIN(subscription_end) AS 'sub_end_min',
4       MAX(subscription_end) AS 'sub_end_max'
5   FROM subscriptions;
```

Query Results			
sub_start_min	**sub_start_max**	**sub_end_min**	**sub_end_max**
2016-12-01	2017-03-30	2017-01-01	2017-03-31

The earliest subscription_start date is 2016-12-01. This means that the denominator of the monthly churn rate equation is zero for all months before January 2017, and so the first month for which monthly churn can be calculated is January 2017.

The latest subscription_start and subscription_end dates are 2017-03-30 and 2017-03-31, respectively. Given the scope of this project, it seems reasonable to assume that data has only provided until 2017-03-31. And so, the last month for which one can calculate churn is March 2017.

There is enough information to calculate monthly churn rates for January 2017 through March 2017.

1. GET FAMILIAR WITH THE COMPANY. - CONT'D

What segments of users exist?

```
1   SELECT DISTINCT segment
2   FROM subscriptions;
```

Query Results
segment
87
30

There are two different user segments, designated as 87 and 30, included within the subscriptions table data.

2. WHAT IS THE OVERALL CHURN TREND SINCE THE COMPANY STARTED?

```
1   WITH months AS (
2       SELECT '2017-01-01' AS first_day,
3           '2017-01-31' AS last_day
4       UNION
5       SELECT '2017-02-01',
6           '2017-02-28'
7       UNION
8       SELECT '2017-03-01',
9           '2017-03-31'),
10  cross_join AS (
11      SELECT *
12      FROM subscriptions
13      CROSS JOIN months),
14
15  status AS (
16      SELECT id,
17          first_day AS month,
18          CASE
19              WHEN segment == 87
20                  AND subscription_start < first_day
21                  AND (subscription_end >= first_day
22                  OR subscription_end IS NULL)
23                  THEN 1
24              ELSE 0
25          END AS is_active_87,
26          CASE
27              WHEN segment == 30
28                  AND subscription_start < first_day
29                  AND (subscription_end >= first_day
30                  OR subscription_end IS NULL)
31                  THEN 1
32              ELSE 0
33          END AS is_active_30,
```

```
34          CASE
35              WHEN segment == 87
36                  AND subscription_end >= first_day
37                  AND subscription_end <= last_day
38                  THEN 1
39              ELSE 0
40          END AS is_canceled_87,
41          CASE
42              WHEN segment == 30
43                  AND subscription_end >= first_day
44                  AND subscription_end <= last_day
45                  THEN 1
46              ELSE 0
47          END AS is_canceled_30
48      FROM cross_join),
49
50  status_aggregate AS (
51      SELECT month,
52          SUM(is_active_87) AS sum_active_87,
53          SUM(is_active_30) AS sum_active_30,
54          SUM(is_canceled_87) AS sum_canceled_87,
55          SUM(is_canceled_30) AS sum_canceled_30
56      FROM status
57      GROUP BY 1)
58
59  SELECT month,
60      (1.0 * sum_canceled_87 / sum_active_87) AS churn_87,
61      (1.0 * sum_canceled_30 / sum_active_30) AS churn_30
62  FROM status_aggregate
63  GROUP BY 1;
64
65
66
```

month	churn_87	churn_30
2017-01-01	0.25089605734767	0.0756013745704467
2017-02-01	0.316916488222698	0.0733590733590734
2017-03-01	0.476894639556377	0.116991643454039

The churn rates for user segments 87 and 30 can be calculated nicely using the code that the capstone project guide helps us to create (shown to the left).

But our churn rate analysis can be made more concise and flexible (for an unknown number of different segments) by altering this code, as suggested by the bonus section of the capstone project guide.

2. WHAT IS THE OVERALL CHURN TREND SINCE THE COMPANY STARTED? - CONT'D

```
1    WITH months AS (
2        SELECT '2017-01-01' AS first_day,
3            '2017-01-31' AS last_day
4        UNION
5        SELECT '2017-02-01',
6            '2017-02-28'
7        UNION
8        SELECT '2017-03-01',
9            '2017-03-31'),
10   cross_join AS (
11       SELECT *
12       FROM subscriptions
13       CROSS JOIN months),
14
15   status AS (
16       SELECT id,
17           first_day AS month,
18           segment,
19           CASE
20               WHEN subscription_start < first_day
21                   AND (subscription_end >= first_day
22                   OR subscription_end IS NULL)
23                   THEN 1
24               ELSE 0
25           END AS is_active,
26           CASE
27               WHEN subscription_end >= first_day
28                   AND subscription_end <= last_day
29                   THEN 1
30               ELSE 0
31           END AS is_canceled
32       FROM cross_join)
33
```

```
34   SELECT segment,
35       month,
36       (1.0 * SUM(is_canceled) / SUM(is_active)) AS churn
37   FROM status
38   GROUP BY 1, 2;
39
```

segment	month	churn
30	2017-01-01	0.0756013745704467
30	2017-02-01	0.0733590733590734
30	2017-03-01	0.116991643454039
87	2017-01-01	0.25089605734767
87	2017-02-01	0.316916488222698
87	2017-03-01	0.476894639556377

Query Results

Using this improved code, we can easily calculate the monthly churn rates for any number of different segments. Similar to how we were previously grouping our results by month, we are now grouping our results by both month **and** segment.

2. WHAT IS THE OVERALL CHURN TREND SINCE THE COMPANY STARTED? - CONT'D

```sql
1   WITH months AS (
2       SELECT '2017-01-01' AS first_day,
3           '2017-01-31' AS last_day
4       UNION
5       SELECT '2017-02-01',
6           '2017-02-28'
7       UNION
8       SELECT '2017-03-01',
9           '2017-03-31'),
10  cross_join AS (
11      SELECT *
12      FROM subscriptions
13      CROSS JOIN months),
14
15  status AS (
16      SELECT id,
17          first_day AS month,
18          segment,
19          CASE
20              WHEN subscription_start < first_day
21                  AND (subscription_end >= first_day
22                  OR subscription_end IS NULL)
23                  THEN 1
24              ELSE 0
25          END AS is_active,
26          CASE
27              WHEN subscription_end >= first_day
28                  AND subscription_end <= last_day
29                  THEN 1
30              ELSE 0
31          END AS is_canceled
32      FROM cross_join)
33
```

```sql
34  SELECT month,
35      (1.0 * SUM(is_canceled) / SUM(is_active)) AS churn
36  FROM status
37  GROUP BY 1;
```

| Query Results	
month	churn
2017-01-01	0.16140350877193
2017-02-01	0.188832487309645
2017-03-01	0.27164416203336

By removing the segment column from the final SELECT statement, and then grouping by only the month column, we can also look at Codeflix's monthly churn rates across all user segments. This would be useful for getting an overall sense of how well the subscription service is retaining users.

2. WHAT IS THE OVERALL CHURN TREND SINCE THE COMPANY STARTED? - CONT'D
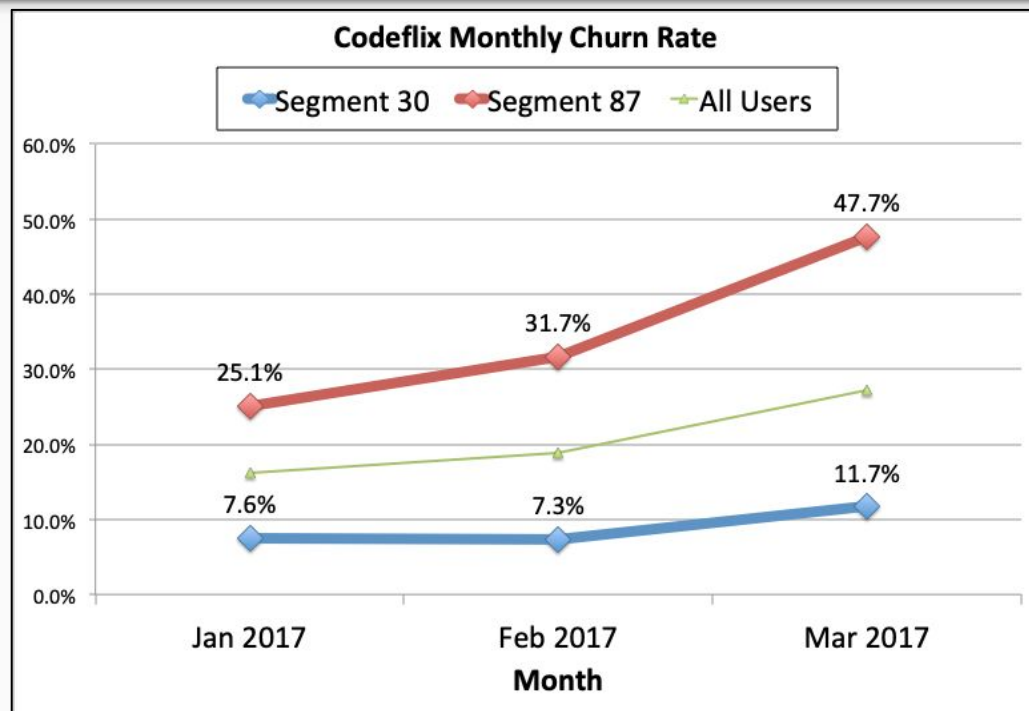

Codeflix Monthly Churn Rate
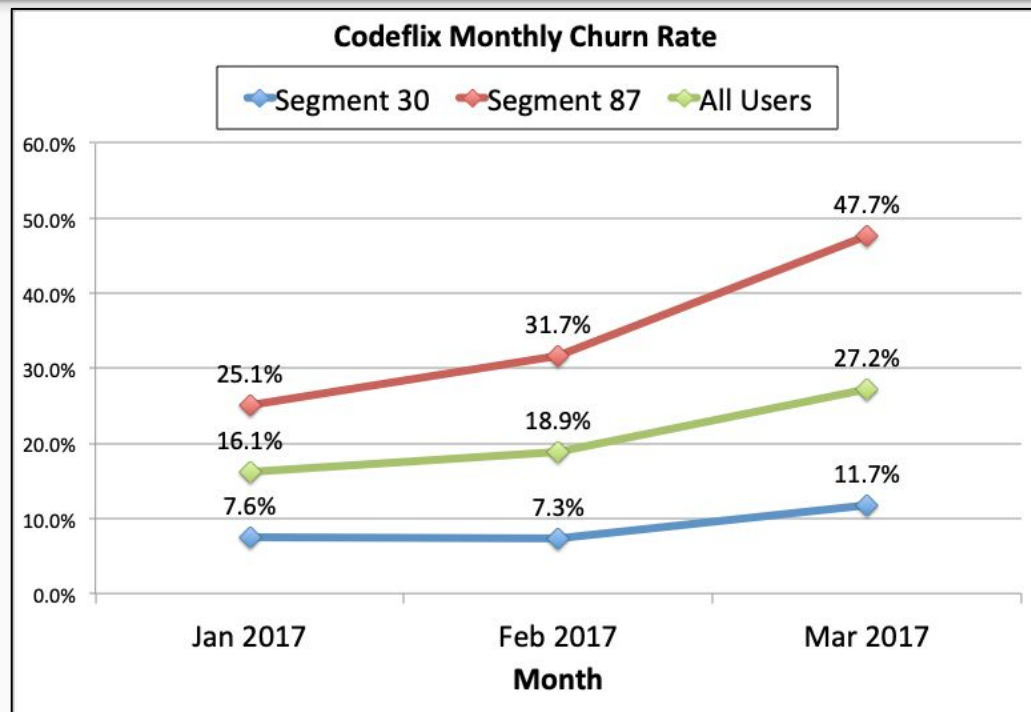
By plotting the results of the previous two queries, we can see that there is a general upward trend in the monthly churn rates, for all users overall, as well as for each segment (30 and 87) individually.

3. COMPARE THE CHURN RATES BETWEEN USER SEGMENTS.

Codeflix Monthly Churn Rate
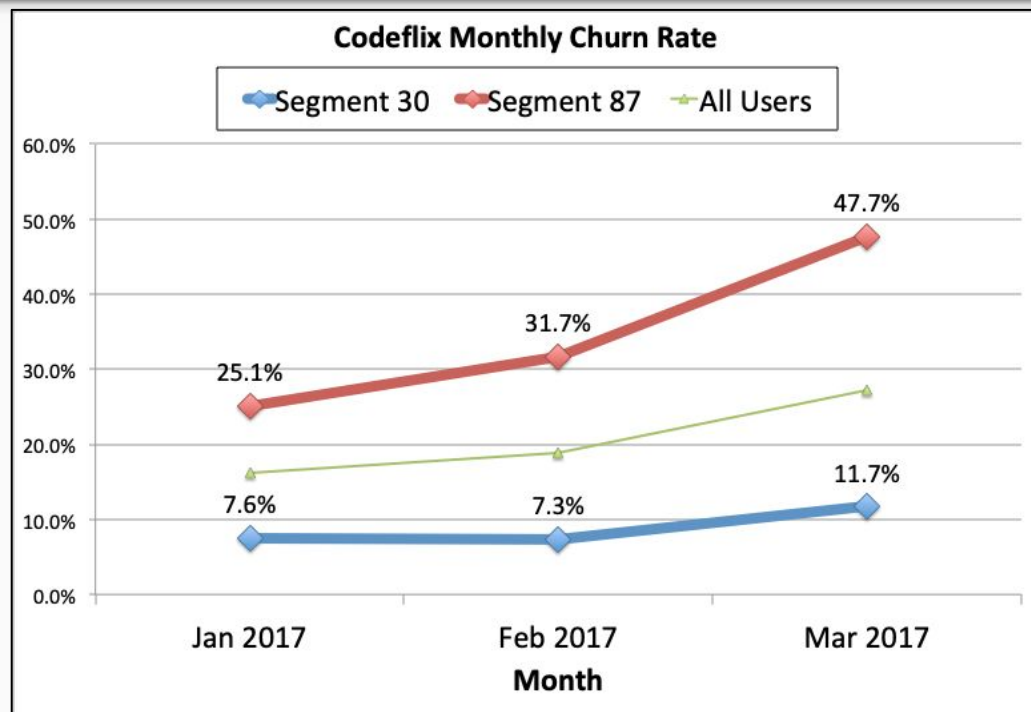
We can also see that the churn rate for segment 87 is more than 3 times the churn rate for segment 30, for each of the three months evaluated.

3. COMPARE THE CHURN RATES BETWEEN USER SEGMENTS. - CONT'D

Codeflix Monthly Churn Rate

Segment 30 — Segment 87 — All Users

	Jan 2017	Feb 2017	Mar 2017
Segment 87	25.1%	31.7%	47.7%
All Users		19.0%	27.0%
Segment 30	7.6%	7.3%	11.7%

Which segment of users should the company focus on expanding?

Codeflix appears better able to retain users of segment 30, compared to users of segment 87, as indicated by segment 30's considerably lower monthly churn rates.

Therefore, Codeflix should focus on expanding the number of users in segment 30. Of course, that is just the simple answer. It's the conclusion, I assume, that the creators of this capstone project had hoped we could reach by the end of this course.

BUT…

3. COMPARE THE CHURN RATES BETWEEN USER SEGMENTS. - CONT'D

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

The truth is, we don't know anything about what differentiates user segments 30 and 87. Was this an A/B test? What sort of variable was tested? One possibility is that these two segments represent two different customer demographics, in which case the focused expansion of segment 30 seems like reasonable advice, at first.

But in the real world, **additional metrics** (such as user acquisition cost, segment market size, and average revenue per user) and **potential remedies** (like demographic/family discounts) would need to be considered before making such consequential recommendations.

3. COMPARE THE CHURN RATES BETWEEN USER SEGMENTS. - CONT'D

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

For instance, what if the cost to acquire users for segment 87 is 1/10th the cost to acquire users for segment 30?

And what if segment 87 users are primarily young college students with minimal income? Perhaps by lowering the subscription cost slightly for students (if they sign up with .edu email addresses) for a 4-year term, Codeflix can better retain and garner loyalty from this segment.

And what if the segment 87 market size is 100 times greater than the segment 30 market size?

If it no longer seems as obvious that Codeflix should focus on expanding segment 30, then I have made my point. But to drive it home, here it is in fewer words...

3. COMPARE THE CHURN RATES BETWEEN USER SEGMENTS. - CONT'D

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

Churn rate is just one piece of the puzzle and should be used in conjunction with other available user and market metrics to develop a robust and successful growth strategy.