codecademy

Churn Rates at Codeflix

Analyze Data with SQL

Ahmad Ammash

Thursday, September 14, 2023

Overview of the Project

The team at Codeflix, a streaming video startup that launched four months ago, wants to evaluate the progress their business has made since its launch. They want to analyze the churn rates across two of their customer segments to measure the attractiveness of their service to their customers.

After determining the months whose churn rates can be calculated with the data available and investigating the customer segments of interest, this project answered the questions about the overall trend in the churn rate over the studied period and compared the rates between the customer segments.

Ultimately, insights generated from this analysis can be used by the management team at Codeflix to gauge the company’s performance and potentially focus on the customer segments with higher churn rates.

Table of Contents

1. Choosing the Month(s) & Exploring the Customer Segments

The subscriptions dataset provides information on the unique user ID, their subscription start and end dates, as well as the customer segment they belong to.

The range of dates covers the months of December 2016 through March 2017. However, we can only determine the churn rate for the first three months of 2017 since no users were active before the first day of December.

User ID	Subscription Start	Subscription End	Segment Number
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

```
SELECT *
FROM subscriptions
LIMIT 5;


SELECT
 MIN(subscription_start) AS min_date,
 MAX(subscription_start) AS max_date
 FROM subscriptions;

SELECT
DISTINCT segment
FROM subscriptions;
```

Minimum Subscription Start Date	Maximum Subscription Start Date
2016-12-01	2017-03-30

The results also show that Codeflix is interested in two specific segments: numbers 87 and 30.

Segments
87
30

2. Methodology for Calculating Churn Rates

The methodology followed to calculate churn rates is as follows:

1. Create a temporary table formed of the months we want to calculate the churn rate for. These are January through March 2017 (inclusive).
2. Apply a CROSS JOIN on the subscriptions and months tables (another temporary table).
3. Create the temporary status table which is formed of four Boolean (1 or 0 values) columns indicating whether a user from a specific segment (number 30 or 87) is classified as ‘active’ or ‘canceled’. A user is classified as ‘active’ at the start of a specific month when their subscription start date is before the first day of that month and their cancellation date is either after the first day of the month or non-existent (they did not cancel). On the contrary, a user is counted as canceled when their cancellation date is between the first and last days of the months we are interested in. CASE WHEN statements are used to create columns for the active/canceled users for each of the two user segments.
4. Create an aggregated version of the temporary status table that sums each of its columns and groups by the month. This table shows the total number of active and canceled users in each of the months under study and in each customer segments.
5. Calculate the churn rates for each month and user segment combination using the aggregated table created in the previous step.

2. Methodology for Calculating Churn Rates (continued)

1. Creating the Temporary Months Table

```
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
),
```

2. Creating the CROSS JOIN Table

```
 cross_join AS
 (
  SELECT *
  FROM subscriptions CROSS
JOIN months
 ),
```

4. Creating the Temporary Status Aggregate Table

```
 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
 )
```

3. Creating the Temporary Status Table

```
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
 ),
```

3. Overall Churn Trend since the Company Started

The results show that the overall churn rate has increased over the first three months of 2017 and since Codeflix started. This is not a good sign for the company and management as it shows that less users are sticking with the service with time. Moreover, the churn rate experienced a sharp increase between February (around 19%) and March (around 27%).

```
 SELECT
 month,
 100.0*(sum_canceled_87 +
sum_canceled_30)/(sum_active_87 +
sum_active_30) AS 'Overall_Churn'
 FROM status_aggregate
```

Month	Overall Churn Rate (%)
January	16.16%
February	18.97%
March	27.42%

4. Comparing the Churn Rates between User Segments

The results show that the user segment 87 has a significantly higher churn rate than the user segment 30. Besides, user segment 87 saw a considerable increase in the churn rate between February and March. Relating this finding to the overall churn rate (on the previous slide), we can say that the user segment 87 is a major driver of the high overall churn rate at Codeflix.

Such insights can help management to focus their attention on customer segments, such as number 87, which have a higher churn rate. The management needs to make changes at the company to make the service more attractive for that specific group of users.

```
SELECT
 month,
 100.0*sum_canceled_87/sum_active_8
7 AS 'Churn_87',
 100.0*sum_canceled_30/sum_active_3
0 AS 'Churn_30'
 FROM status_aggregate
```

Month	Churn Rate (Segment 87)	Churn Rate (Segment 30)
January	25.17%	7.56%
February	32.03%	7.33%
March	48.58%	11.73%

Key Points

1. The overall churn rate at Codeflix has been on the rise between January and March 2017 indicating that many users are not satisfied with the service and are possibly choosing to switch to a competitor.
2. There is a significant difference in churn rates between different customer segments. Customer segment number 87 is a major contributor to the overall high churn rate at Codeflix. On the contrary, user segment number 30 has a much lower churn rate.