



# CODEFLIX: Measuring Churn Rate

Learn SQL from Scratch

Earl Puzon

# Questions and Topics Covered

1. Getting familiar with Codeflix
2. What is the overall churn rate by month?
3. Compare the churn rates between segments

# 1. Getting Familiar with Codeflix

## How many months has the company been operating?

- Per the first query, the lowest valued date is selected as the company's start date and the highest valued date is selected as the company's latest operating date. This shows that the company has been running from December 2016 – March 2017.
- A total of 3 months can be utilized to calculate the churn rate.

## What segments of users exists?

- On the second query, it selects the distinct segments and classifies them as "users". Segments 87 and 30 exists in the given table.

--1) Select the first and latest subscription start dates.

```
SELECT MIN(subscription_start) AS 'company_subscription_start',  
MAX(subscription_start) AS 'latest_subscription_start'  
FROM subscriptions;  
|
```

--2) Select the types of segments on the table.

```
SELECT DISTINCT segment AS 'users'  
FROM subscriptions  
LIMIT 10;
```

Users
87
30

Company Start Date	Latest Company End Date
12-01-16	03-30-17

# 2. Calculating Overall Churn Rate by Month (Part 1)

## Defining Churn Rate:

- It is the rate in which customers stops or ceases subscription to the service. Each month, there will be a number of customers that will cancel their subscription with Codeflix or any subscription-based services due to various reasons.
- Monthly Churn Rate is calculated by determining the number of customers that cancelled subscriptions and divide it by the number of active subscriptions during the month.
- To calculate the churn rate for each month, we need to determine the cancelled and active subscription data by creating a temporary table that combines all of that information.

--3) Create a new table showing the months can be utilized to calculate the 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,  
           '2017-03-30' AS last_day  
  
)
```

--Join the subscriptions table and months table into a temporary table called "cross\_join"

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

# 2. Calculating the Overall Churn Rate by Month (Part 2)

## Obtaining the Active Subscription Data:

- In order to obtain the subscription data, a temporary table needs to be created that selects the active and cancelled subscriptions during the month.
- As listed on the following query, the table is created by selecting the identification number for each subscriber, the first day of each month that is utilized to classify the month, and an “IF THEN” statement in which each user is classified as “1” if their subscription existed in the beginning of the month or if their subscription has not ended as of yet. This statement creates a column called “is\_active” which identifies the active subscriptions for the month.
- The query shows how the active subscribers data is selected for segments 87 and segment 30.

```
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',  
  --Statement above states that IF value of subscription date is less than  
  --the beginning of the month (subscription started before the given month)  
  --and subscription end date value is greater than the beginning of the month  
  --(subscription has not been cancelled) for Segment 87, then the  
  --subscription is active for Segment 87.  
  CASE WHEN (subscription_start < first_day)  
         AND (subscription_end > first_day OR subscription_end IS NULL)  
         AND (segment = 30)  
       THEN 1  
       ELSE 0  
       END AS 'is_active_30',  
  --Statement above states that IF value of subscription date is less than  
  --the beginning of the month (subscription started before the given month)  
  --and subscription end date value is greater than the beginning of the month  
  --(subscription has not been cancelled) for Segment 30, then the  
  --subscription is active for Segment 30.
```

# 2. Calculating the Overall Churn Rate by Month (Part 3)

## Obtaining the Cancelled Subscription Data:

- To grab the data for subscribers who have cancelled during the month, another “IF THEN” statement needs to be created for each segment.
- On the query, the statement selects the users that have cancelled by stating that if their subscription end date lands on the middle of the month, then they are labeled as 1 on a column called “is\_canceled”.
- The query shows these “is\_canceled” columns created for both segments 30 and 87.
- The purpose of these columns is to identify the number of users that have cancelled during the month.

```
CASE WHEN (subscription_end BETWEEN first_day AND last_day)
AND (segment = 30)
THEN 1
ELSE 0
END AS 'is_canceled_30',
```

--Statement above states that if the subscription final date for segment 30 falls between the beginning and the end of the month, then the subscription is cancelled for that segment.

```
CASE WHEN (subscription_end BETWEEN first_day AND last_day)
AND (segment = 87)
THEN 1
ELSE 0
END AS 'is_canceled_87'
```

--Statement above states that if the subscription final date for segment 87 falls between the beginning and the end of the month, then the subscription is cancelled for that segment.

```
FROM cross_join),
```

# 2. Calculating the Overall Churn Rate by Month (Part 4)

## Combining the Active and Cancelled Subscriptions Data:

- Now that we have identified the columns that selects the number of active and cancelled subscriptions, we can combine them into a temporary table.
- On this query, each month is selected with its associated aggregate number of active subscriptions and aggregate number of cancelled subscriptions for each segment.
- The results for each segment is seen on the table below by selecting everything from the temporary table.

--Create status\_aggregate table below:

```
status_aggregate AS (  
SELECT month, SUM(is_active_30) AS 'sum_active_30', SUM(is_active_87) AS  
'sum_active_87', SUM(is_cancelled_87) AS 'sum_cancelled_87',  
SUM(is_cancelled_30) AS 'sum_cancelled_30'  
FROM status  
GROUP BY 1)
```

--The statement above creates a temporary table summarizing the active and cancelled subscriptions for Segments 30 and 87.

Month	Active Users (30)	Cancelled Users (30)	Active Users (87)	Cancelled Users (87)
January	291	22	278	70
February	518	38	462	148
March	716	81	531	247

# 2. Calculating the Overall Churn Rate by Month (Part 5)

## Calculating the Churn Rate:

- Now that the total active and cancelled subscription numbers for each month is available to us by creating the “status\_aggregate” table, we can calculate the churn rate for both segments.
- The query calculates the churn rate for each segment by taking the total number of cancelled subscriptions and dividing it by the aggregate active subscription number for each month.
- The results of the monthly churn rate each segment can be seen on the table below. A more detailed comparison of the churn rates for both segments can be seen on the next slide.

```
SELECT month, ROUND(1.0 * sum_canceled_87 / sum_active_87, 2) AS  
'churn_rate_87', ROUND(1.0 * sum_canceled_30 / sum_active_30, 2) AS  
'churn_rate_30'  
FROM status_aggregate;  
  
--Calculate the churn rate above by taking the total canceled  
subscriptions for that month divided by the total active subscription.
```

Month	Churn Rate (30)	Churn Rate (87)
January	8 %	25 %
February	7 %	32 %
March	11 %	47 %



# 3. Comparing Churn Rates for Segments 87 and 30

Segment	Average Churn Rate (3-Month Period)
30	9 %
87	35 %

## Notes for Segments 87 and 30:

- Based on the available data, it is discovered that segment 30 has the lower churn rate compared to segment 87.
- On a 3-month period, segment 87 has a higher average monthly churn rate than segment 30.
- As the company grows, they should focus on expanding segment 30 as data shows that users are more likely to be satisfied that this segment provides.
- Since the company started, it seems that segment 30 provides a better product that retains customer each month. While churn rate is currently going up for segments, segment 30's increase in rates is less drastic in comparison to segment 87.

## Segment 30 vs. Segment 87 Churn Rate

