User Churn Rate for Codeflix

Analyzing Data with SQL

Michelle Caler

July 9, 2020

codecademy Pro guided project

Presentation Outline

1. Aims of the Project
2. The Data Set
3. Analysis and Results
4. Conclusions and Recommendations

Background

Codeflix is a fictional video streaming start-up company. After four months of offering its streaming service, the company is interested in calculating its churn rate.

In addition to the overall churn rate, the company is interested in knowing how the churn rate compares between its two user segments.

Project Aims

This project aims to answer the following questions:

1. What is the overall churn trend since the company started?

2. Based on their churn rates, which segments of its subscriber base should Codeflix focus on expanding?

The Data Set

Data for Codeflix were generated by Codecademy from December 1, 2016 through March 31, 2017. Using this data, monthly churn rates for January, February, and March of 2017 could be calculated[1].

2000 fictional subscribers were generated. Users were divided into two segments: “87” and “30.” The significance of these segments was not provided.

1: Churn rates for December 2016 could not be calculated, as the fictional data was generated with a minimum subscription length of 31 days.

The Data Set: Context

Since data for this exercise was generated by Codecademy and is therefore proprietary to that company, the data table used in this analysis cannot be provided with this presentation.

To give some context to the SQL searches provided on GitHub, a table named “subscriptions” was provided for analysis. Columns in that table included ID (an integer) subscription_start and subscription_end (both strings), and segment (an integer).

Monthly Retention and Cancellation Numbers

Before churn rates could be calculated, the number of users retained per month was determined, as well as the number of cancellations during that month. These numbers are provided below. The SQL search used to generate them is provided on GitHub. The total number of subscribers at the start of a month can be found by adding the “Retained” column to the “Cancelled” column.

Month	Overall		“87” segment		“30” segment	
	Retained	Cancelled	Retained	Cancelled	Retained	Cancelled
January	478	92	209	70	269	22
February	799	186	319	148	480	38
March	917	342	283	258	634	84

Overall Churn Rate by Month

From these values, churn rates could be calculated as:

churn rate = cancelled / (retained + cancelled)

For the case of the “87” and “30” segments, SQL was used to calculate these rates directly. The search is included on GitHub. For the overall churn rate, data from the previous table was inserted into the above expression and calculated via Microsoft Excel.

Month	Overall Churn Rate	“87” segment Churn Rate	“30” segment Churn Rate
January	0.16	0.25	0.08
February	0.19	0.32	0.07
March	0.27	0.49	0.12

Overall Trends

- Codeflix's overall number of total subscribers has increased over each of the three months considered, but so has its churn rate.

- March saw a jump in the overall churn rate to over 25%.

- The churn rate for the "87" segment is much greater than that of the "30" segment.

Trends by segment

- The churn rate for the “87” segment is at least triple that of the “30” segment.
- For each month considered, the “30” segment had a larger total number of subscribers than the “87” segment.
- Both segments saw a significant increase in churn rate in March.

Conclusions

- Codeflix's churn rate increased every month from January 2017 through March 2017, with an uptick in churn rate in March.

- The "30" segment has a much lower churn rate than the "87" segment, despite having a larger total number of subscribers.

- Nearly half of subscribers in the "87" segment chose to end their subscriptions in March 2017.

Recommendations

Clearly, **emphasis should be put on expanding the “30” segment**. This segment is less likely to cancel their subscription.

Since the precise distinction between “87” and “30” is unknown, more specific recommendations are difficult to put forth.

1. Does the “30” segment have a differently-structured subscription plan? If so, that should be marketed instead of the “87” plan.
2. Is the “30” segment overall more financially stable than the “87” segment? If so, the subscription service should be marketed in places that segment is more likely to see.