code|cademy

CodeFlix

Analyze User Churn with SQL
Landon Woollard
08/16/23

Table of Contents

1. Getting familiar with the company.
   a. How many months has the company been operating?
   b. Which months do you have enough information to calculate a churn rate?
   c. What segments of users exist?

2. Analyzing user churn.
   a. What is the overall churn trend since the company started?
   b. How do the churn rates compare between user segments.
   c. Which segment of users should the company focus on expanding?

1. Getting familiar with the company.

1.1 How many months has the company been operating for?

- The company has been operating for four months.
  - The first subscriptions were in December, 2016
  - The data is current through the end of March, 2017

start_month	COUNT(*)
12-2016	570
01-2017	507
02-2017	460
03-2017	463

end_month	COUNT(*)
∅	570
01-2017	507
02-2017	460
03-2017	463

```
--identify unique values in each column, and generate
counts to get a better feel for what is in the
database prior to calculations.

SELECT STRFTIME('%m-%Y',subscription_start)
          AS start_month,
     COUNT(*)
FROM subscriptions
GROUP BY STRFTIME('%m-%Y',subscription_start)
ORDER BY STRFTIME('%Y',subscription_start);


SELECT STRFTIME('%m-%Y',subscription_end)
          AS end_month,
     COUNT(*)
FROM subscriptions
GROUP BY STRFTIME('%m-%Y',subscription_end);
```

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

- Codeflix requires a minimum subscription length of 31 days.
  - A user can never start and end their subscription in the same month.
- Churn can be calculated for January, February, and March, but NOT December.
  - December was the first month users could sign-up, so there were no cancellations yet.

start_month	end_month
12-2016	∅
01-2017	01-2017
02-2017	02-2017
03-2017	03-2017

$$\text{Churn} = \frac{\text{Cancellations during month}}{\text{Active Subscriptions at start of month}}$$

1.3 What segments of users exist?

- There are two segments of users.
  - Segment 30
  - Segment 87

segment	COUNT(*)
30	1000
87	1000

```
--identify unique values in each column, and generate
counts to get a better feel for what is in the
database prior to calculations.

SELECT segment, COUNT(*)
FROM subscriptions
GROUP BY segment;
```

2. Analyzing User Churn

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

- Overall we see that the user churn for CodeFlix over the first three months is increasing.
  - Based on this data, churn can be projected using a quadratic regression with an R² statistic of 1

month	churn_perc
2017-01-01	16.140350877193
2017-02-02	18.8832487309645
2017-03-03	27.164416203336
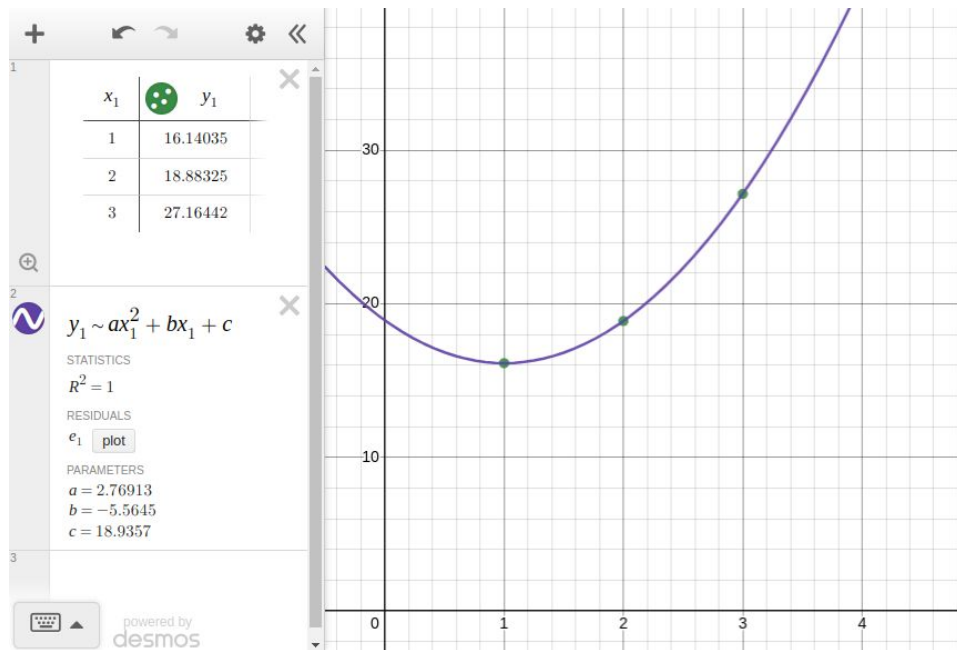
2.2 How do the churn rates compare between user segments.

- Segment 87 has a churn rate approximately 4 times higher than segment 30.

- Segment 87 has consistently increased churn month-to-month.

- During February, segment 30 showed a slight decrease in churn.

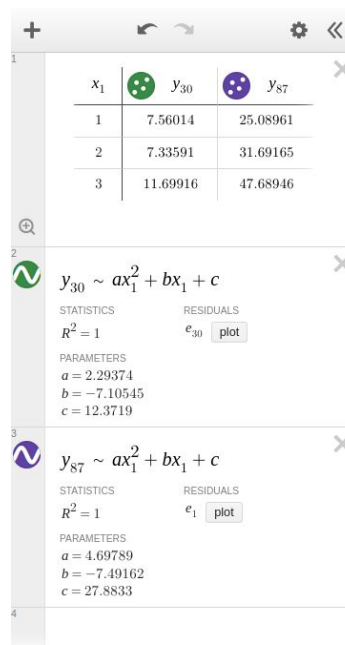- Segment 30 shows smaller fluctuations in churn than segment 87

SEGMENT 30	
month	churn_perc
2017-01-01	7.56013745704467
2017-02-02	7.33590733590734
2017-03-03	11.6991643454039

SEGMENT 87	
month	churn_perc
2017-01-01	25.089605734767
2017-02-02	31.6916488222698
2017-03-03	47.6894639556377

2.2 How do the churn rates compare between user segments.

SEGMENT 30	
month	churn_perc
2017-01-01	7.56013745704467
2017-02-02	7.33590733590734
2017-03-03	11.6991643454039

SEGMENT 87	
month	churn_perc
2017-01-01	25.089605734767
2017-02-02	31.6916488222698
2017-03-03	47.6894639556377

x_1	y_{30}	y_{87}
1	7.56014	25.08961
2	7.33591	31.69165
3	11.69916	47.68946

$y_{30} \sim ax_1^2 + bx_1 + c$

STATISTICS   RESIDUALS

$R^2 = 1$ e_{30} plot

PARAMETERS
$a = 2.29374$
$b = -7.10545$
$c = 12.3719$

$y_{87} \sim ax_1^2 + bx_1 + c$

STATISTICS   RESIDUALS

$R^2 = 1$ e_1 plot

PARAMETERS
$a = 4.69789$
$b = -7.49162$
$c = 27.8833$

Both data sets fit to a quadratic regression model with an R² statistic of 1
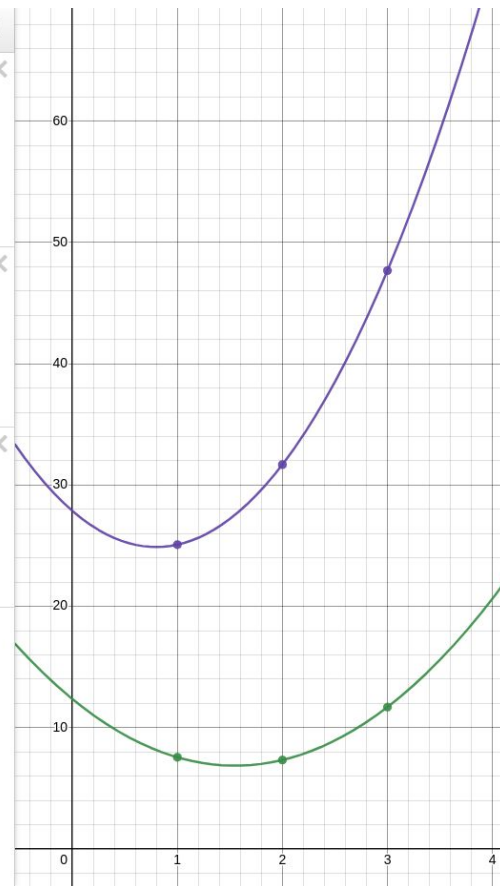
powered by desmos

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

- CodeFlix should focus on expanding Segment 30
    - Lower churn overall when compared to segment 87.
        - Lower churn ⟶ less money spent acquiring new customers.
        - Lower churn ⟶ greater reliability of profit.
    - Churn is increasing at a lesser rate when compared to segment 87.