Objective

Four months into launching, video streaming service CodeFlix is interested in measuring their user churn rate .
The marketing department is particularly interested in how the churn compares between two segments of users.

In this project I'll help them figure that out by answering following questions:
1. What segments of users exists?
2. Which months do you have enough information to calculate churn rate?
3. What is the overall churn trend since the company started?
4. Compare the churn rates between user segments.
5. Which segment of users should the company focus on expanding?

The queries used to find data are included in a separate .sql file

Method

Churn rate is the fraction of subscribers that have canceled within a certain period, usually a month. For a user base to grow, the churn rate must be less than the new subscriber rate for the same period.

CodeFlix has the minimum subscription policy of 31 days , meaning that user can't cancel in the month of subscription.

To calculate the churn rate for the month we divide the number of users who canceled during that month by the total number of subscribers on the beginning of that month

$$\frac{cancellations}{total\ subscribers}$$

Getting familiar with our data

Database Schema
subscriptions

name	type
id	INTEGER
subscription_start	TEXT
subscription_end	TEXT
segment	INTEGER

Rows: 2000

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

Inspecting data

There are 2 segments of subscribers that are acquired through two distinct channels and are equally distributed in given dataset

By calculating churn rates for them separately we can get a sense of how acquiring method influence subscription duration

Data provided contains data for first 4 months of service operating.
That means that we can analyze data for 3 of them: January, February and March of 2017

segment	count
30	1000
87	1000

first_date	last_date
2016-12-01	2017-03-31

This table adds up the statuses of active and cancelled users for each segment to provide an overall total for each month

Ultimately this data will be used for final calculation, but for now it gives us information about fluctuation of users among segments

We can see that both segments have roughly the same number of active users throughout months, but the segment 87 have noticeably more cancellations

Let's see if churn rates reflect this

sum_active_87	sum_active_30	sum_canceled_87	sum_canceled_30
278	291	70	22
462	518	148	38
531	716	258	84

Churn rates

month	overall_churn
01	0.16
02	0.19
03	0.27

month	chrune_87	chrune_30
01	0.25	0.08
02	0.32	0.07
03	0.49	0.12

Conclusion

- For first four months of running CodeFlix managed to contain relatively steady churn rates by obtaining more subscribers than it was losing
- In January it has 16% churn that increased to 27% in March which is somewhat expected and still indicates positive fluctuation of subscribers
- It seems that neither segment has trouble obtaining new customers
- Users from segment 87 have significantly higher churn rates so that may be something worth further analysis, CodeFlix could inspect reasons for users in segment 87 unsubscribing and find ways to prevent that