codecademy

Churn Rates

Learn SQL from Scratch

Project: Churn Rates

Codefix marketing department is interested in how the churn compares between two segments of users. They provide us with a dataset containing subscription data for users who were acquired through two distinct channels.

Codeflix marketing department want to know:

1. WHAT IS THE OVERALL CHURN RATE BY MONTH?
2. COMPARE THE CHURN RATES BETWEEN SEGMENTS

1. Codeflix dataset

Project: Churn Rates

1.1 SQL Table ‘subscriptions’

The marketing department provide us with a dataset, table ***‘subscriptions’***, containing subscription data for users who were acquired through two distinct channels. Dataset contains 2.000 files or registers.

```
select * from
subscriptions
LIMIT 4;
```

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

- **id** (INTEGER) : the subscription id
- **subscription_start** (TEXT) : the start date of the subscription
- **subscription_end** (TEXT : the end date of the subscription
- **segment** (INTEGER) : identifies segment the subscription owner belongs to

Project: Churn Rates

1.2 TABLE ‘subscriptions’

We detect two segment the owner belongs to.

- *Segment 30*
- *Segment 87*

The range of months of data provided is between 2016-12-01 and 2017-03-30.

```
select segment, count(*) from subscriptions
group by segment;
```

Query Results

segment	count(*)
30	1000
87	1000

```
select MIN(subscription_start),
MAX(subscription_start) from subscriptions;
```

Query Results

MIN(subscription_start)	MAX(subscription_start)
2016-12-01	2017-03-30

2. Churn rate

Project: Churn Rates

2.1 Churn rate

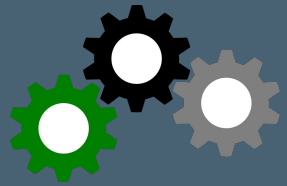
Churn rate is the percent of subscribers that have canceled within a certain period.

To calculate the **churn rate**, we only will be considering users who are subscribed at the beginning of the month. The churn rate is the number of these users who cancel during the month divided by the total number.

Project: Churn Rates

2.2 What is the overall churn rate by month?

```
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
),
cross_join AS
  (SELECT * FROM subscriptions
   CROSS JOIN months),
status AS
  (SELECT id, first_day AS month,
    CASE
      WHEN (subscription_start < first_day)
        AND (
        subscription_end > first_day
        OR subscription_end IS NULL
        ) THEN 1
      ELSE 0
    END as is_active,
    CASE
      WHEN (subscription_end
        BETWEEN first_day and last_day)
        THEN 1
      ELSE 0
    END AS is_canceled
    FROM cross_join) ,
 status_aggregate AS
  (SELECT month,
   SUM(is_active) AS sum_active,
   SUM(is_canceled) AS sum_canceled
   from status
   GROUP BY month)
 SELECT month,
   round((1.0*sum_canceled/sum_active),2) AS
 churn_rate
 from status_aggregate;
```



When code run it calculates the churn rate by month

Query Results

month	churn_rate	active	canceled
2017-01-01	0.16	569	92
2017-02-01	0.19	980	186
2017-03-01	0.27	1247	342

The churn rate is higher in the last month (March). There is a monthly increase in the churn rate. Although the number of subscriptions increases also increases the number of cancellations.

Project: Churn Rates

2.3 Compare the churn rates between segments

Query Results						
month	churn_rate_87	active_87	canceled_87	churn_rate_30	active_30	canceled_30
2017-01-01	0.25	278	70	0.08	291	22
2017-02-01	0.32	462	148	0.07	518	38
2017-03-01	0.49	531	258	0.12	716	84

There is a clear increase in segment 30 and a large decrease in segment 87. We could say that there is a movement between segments. It could be that users prefer segment 30 than segment 87 once they have tested segment 87.