CodeCademy Capstone Churn Project

Ben Guerrieri

Getting Familiar

- **Q** - How many segments do you see?
- **A** - 2, Segments 30 and 87 using the following query:

```
SELECT *
FROM subscriptions
LIMIT 100;
```

11	2016-12-01	2017-01-17	87
12	2016-12-01	2017-02-07	87
13	2016-12-01	Ø	30
14	2016-12-01	2017-03-07	30
15	2016-12-01	2017-02-22	30

- **Q** - How many months has the company been operating?
- **A** - The first subscription began on December 1, 2016 and the latest subscription began on March 30, 2017. Therefore the company has been operating for at lease a total of 4 months.

```
SELECT MIN(subscription_start),
   MAX(subscription_start)
FROM subscriptions;
```

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

Getting Familiar Cont'd

- **Q** - Which months do you have enough info to calculate a churn rate?
- **A** - The only months that we have enough information to calculate churn are the first three months of 2017 as the earliest subscription end date is January 1, 2017 and the latest is March 31, 2017.

```
1  SELECT MIN(subscription_end),
2     MAX(subscription_end)
3  FROM subscriptions;
```

Query Results

MIN(subscription_end)	MAX(subscription_end)
2017-01-01	2017-03-31

Overall Churn Rate

- Q - What is the overall churn rate for the company over the first three months across all segments?
- A - The overall churn rate from January 1, 2017 through March 31, 2017 is 22.03%.
    - To create a churn, I had to utilize multiple temporary tables to establish counts of subscriptions versus cancellations. These tables included:
        - A temporary table (months) to define the parameters for the months
        - A cross-joined table of “months” with the “subscription” table
        - A status table which adds the number 1 or 0 depending on whether or not the subscription beginning and ending date fits inside each individual month, and
        - An aggregate table which sums up the subscriptions and cancellations.

```
52  SELECT 1.0 * sum_cancelled / sum_active AS 'churn_rate'
53  FROM status_aggregate;
54
```

Query Results

churn_rate
0.220326936744847

Overall Churn Rate Cont'd

```sql
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_cancelled
  FROM cross_join
),

status_aggregate AS
(
  SELECT month,
    SUM(is_active) AS sum_active,
    SUM(is_cancelled) AS sum_cancelled
  FROM status
)
```

Overall Churn Trend

- Q - What is the overall churn trend since the company started?
- A - The overall churn trend by month is increasing by an average of 5.5% over the first three months. (January - 16.14%, February 18.88%, March 27.16%). This would mean that the number of cancellations is increasing faster than the rate of new subscriptions.
    - To get a rate by month, simply modify the overall churn trend by adding:
        - The months column to the status_aggregate temporary table
        - A group by function to the end of the status_aggregate temporary table, and
        - The months column to the final churn rate table to see all three months churn rates.

```
44  status_aggregate AS
45  (
46    SELECT month,
47      SUM(is_active) AS sum_active,
48      SUM(is_cancelled) AS sum_cancelled
49    FROM status
50    GROUP BY month
51  )
52
53  SELECT month,
54      1.0 * sum_cancelled / sum_active AS 'churn_rate'
55  FROM status_aggregate;
56
```

Query Results

month	churn_rate
2017-01-01	0.16140350877193
2017-02-01	0.188832487309645
2017-03-01	0.27164416203336

Segment Churn Rate Analysis

- Q - What is the overall churn rate for each segment since the company started?
- A - The churn rate is different for each segment of subscriptions. For segment 87 the church rate is 36.98% over the first 3 months of 2017. For segment 30 the church rate is 9.43% over the first 3 months of 2017.
    - To create individual segment churn rates, I had to utilize multiple temporary tables to establish criteria for each. These tables included:
        - A temporary table (months) to define the parameters for the months
        - A cross-joined table of “months” with the “subscription” table
        - A status table which adds the number 1 or 0 depending on whether or not the subscription beginning and ending date fits inside each individual month, and
        - An aggregate table which sums up the subscriptions and cancellations.
- These aggregates allow me to complete the churn calculation. (see SQL code on next slide)

```
71  SELECT 1.0 * sum_cancelled_87 / sum_active_87 AS 'churn_rate_87',
72    1.0 * sum_cancelled_30 / sum_active_30 AS 'churn_rate_30'
73  FROM status_aggregate;
74
```

Query Results

churn_rate_87	churn_rate_30
0.36985236985237	0.0943025540275049

Segment Churn Rate Analysis Cont’d

```sql
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 segment = 87
      AND subscription_start < first_day
      AND (
        (subscription_end >= first_day)
        OR (subscription_end IS NULL))
      THEN 1
      ELSE 0
  END AS is_active_87,
    CASE
      WHEN segment = 30
      AND subscription_start < first_day
      AND (
        (subscription_end >= first_day)
        OR (subscription_end IS NULL))
      THEN 1
      ELSE 0
  END AS is_active_30,
    CASE
      WHEN segment = 87
      AND subscription_end
          BETWEEN first_day AND last_day
      THEN 1
      ELSE 0
  END AS is_cancelled_87,
    CASE
      WHEN segment = 30
      AND subscription_end
          BETWEEN first_day AND last_day
      THEN 1
      ELSE 0
  END AS is_cancelled_30
  FROM cross_join
),

status_aggregate AS
(
  SELECT SUM(is_active_87) AS sum_active_87,
    SUM(is_active_30) AS sum_active_30,
    SUM(is_cancelled_87) AS sum_cancelled_87,
    SUM(is_cancelled_30) AS sum_cancelled_30
  FROM status
)
```

Segment Churn Rate Analysis Cont’d

- Q - Which user segment should they focus on expanding?
- A - Based upon the total number of subscriptions for each (1000 subs per segment) and solely on individual churn rates, the segment they should focus on is segment 30 subscriptions. This segment of users has 27% less churn than segment 87.