Data Analysis and SQL test

‘Test By Groove Collaborative
‘Presenter - Nupur Bandre

Agenda

Goal: To understand outstanding customer retention trend and perform retention analysis.

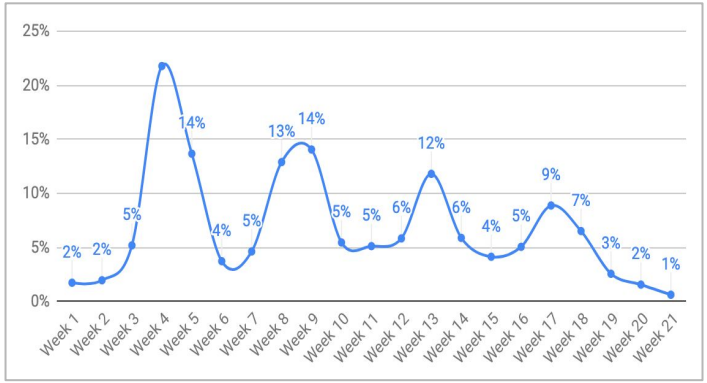
Scope: Provide understanding of following trends

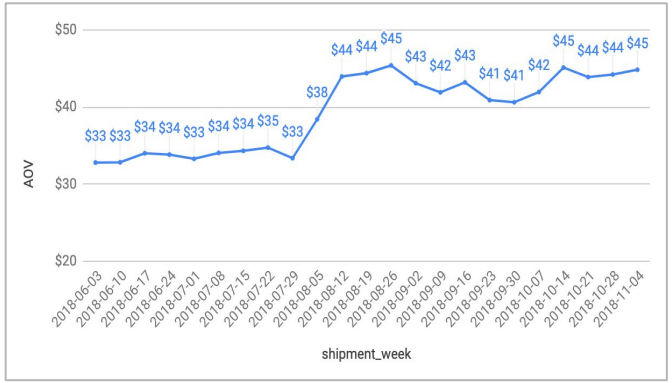
1. Weekly Retention Analysis Chart & Average order Value (AOV)
2. Retention Rate (Addressing gaps in analysis as next steps)
    a. Overall
    b. Channel level (Top 4 Channels)
3. Key Learnings
4. Question 2 (Big lever of impact AOV & CRC)
5. Question 3 (Data requirement & Data Mining)
6. Scripts

Retention Analysis

Retention based on weekly cohorts and understanding weekly trend with average order value (AOV)

Ship Week	Week 0	Week 1	Week 3	Week 5	Week 7	Week 9	Week 11	Week 13	Week 15	Week 17	Week 19	Week 21
2018-06-03	2488	3.90%	9.85%	33.56%	10.01%	26.05%	9.85%	18.65%	8.28%	16.68%	8.48%	5.31%
2018-06-10	6704	3.77%	28.74%	15.74%	17.08%	20.24%	11.62%	18.85%	8.74%	17.05%	8.92%	1.31%
2018-06-17	8675	3.57%	30.18%	14.90%	18.47%	19.23%	13.03%	15.93%	10.79%	16.37%	5.83%	
2018-06-24	7696	3.78%	28.30%	16.41%	19.48%	18.10%	10.14%	17.35%	9.04%	16.19%		
2018-07-01	8505	3.79%	26.16%	17.51%	17.44%	18.68%	9.51%	18.69%	8.45%	12.11%		





Highlights:

1. Customer retention spikes every 4 weeks
   a. Indication of customer preference between 1,2,3,4 month subscription model for products.
   b. ~7% average retention rate
   c. WoW increase in AOV by $12 since the beginning of the time

Next Steps:

1. Identify the products that most attract customers to comeback. [No data available for this]
2. Additional efforts to attract customers for Subscription based products. [Need more data to gain more insights]

Retention Rate (By Weekly Cohorts, weekly trend)

Tracking retention rate for weekly cohorts and week trends ***(Please Note: End of the months is not adjusted for the subscriptions)***

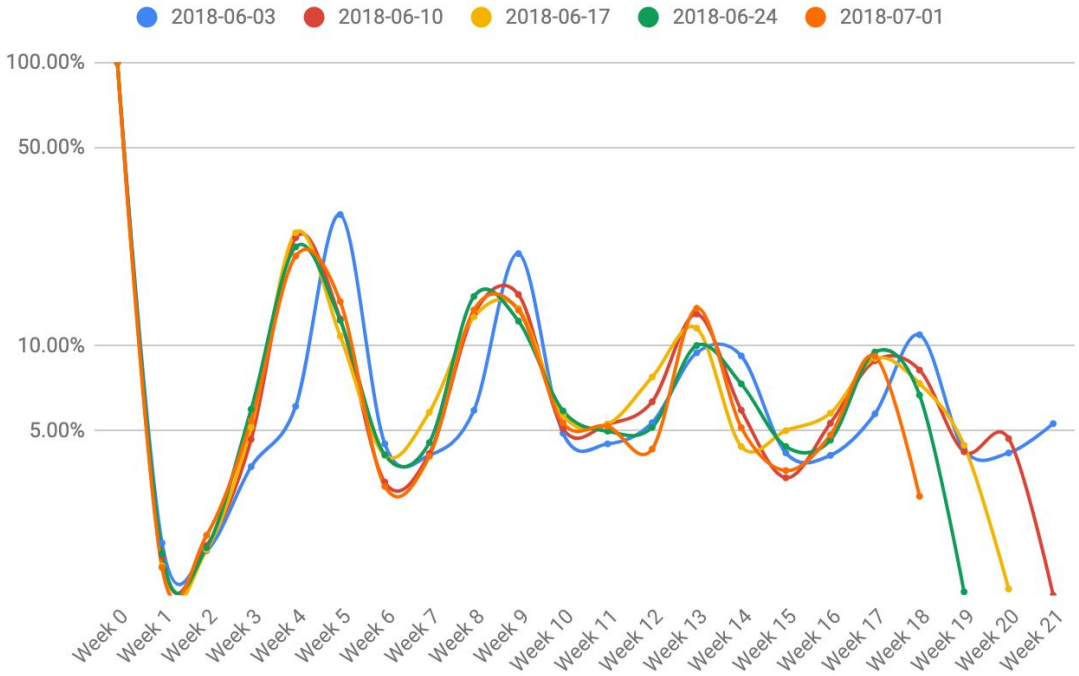


Chart for Increase in retention

Months (M)	Increase in Retention	Week of 2018-07-01	Week of 2018-06-03
Week 4 (M1)	14.62%	20.73%	6.11%
Week 8 (M1)	7.46%	13.37%	5.91%
Month 13 (M1)	4.11%	13.56%	9.45%
Month 17(M1)	3.42%	9.17%	5.75%

Highlights: ***[Data for 6/3 skewed due to end of month]***

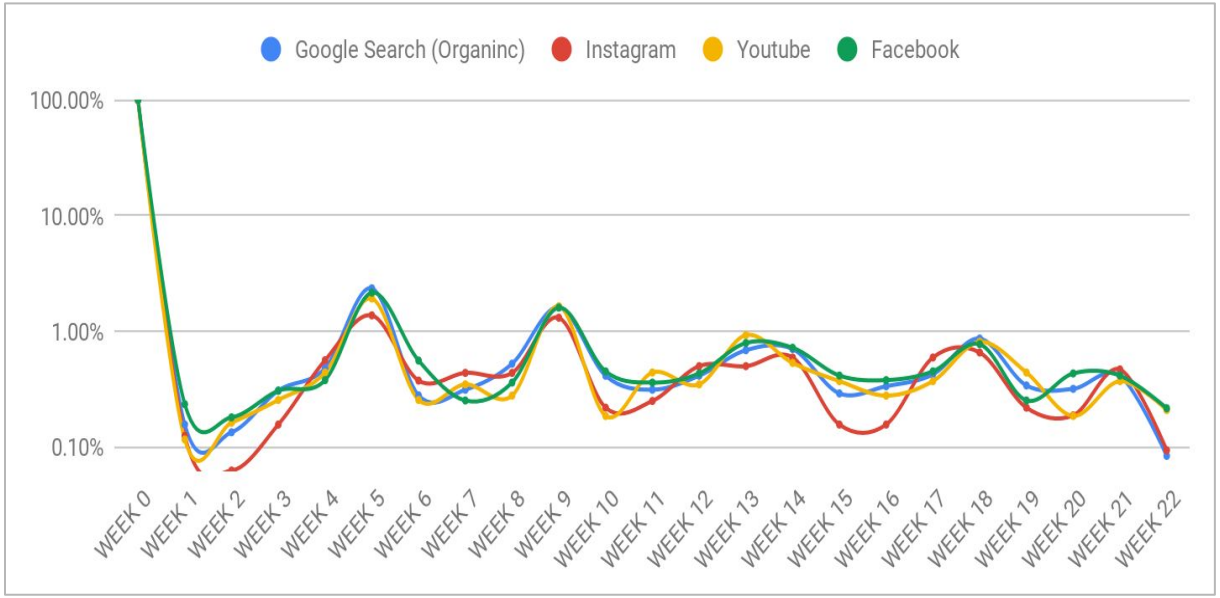
1. Increase in retention between the weekly cohorts (shown as per the above chart)

Gaps in the Analysis [Next steps]:

1. Data is skewed due to end of the month
   a. Consider adjusting the weeks to consider end of month trends.
   b. Make any activity after 25 of month as a part of next month.

Retention Rate (By Weekly Cohorts & Acq. Channel)

Considering top 4 channels by % of new users acquired



Highlights:

1. Channel ID 2 and 7 are top 2 acquisition channels
2. Customers acquired by Instagram has the worst retention rate and is constantly down compared to other channels
3. Customers acquired by Facebook and Youtube are more likely to come back

Next Steps:

1. Identify the gaps in the retention rate is low for Instagram [Like quality of service, customer satisfaction and products]
   a. What is causing the drop?
   b. How to fix it?

Assumed Channel	Acq channel ID	Customers
Google Search (Organinc)	2	17964
Instagram	4	3210
Youtube	5	4326
Facebook	7	5564

Question 2

Which is the bigger lever for revenue impact: an increase in 3+ order retention of 500 basis points or an increase in 3+ order average order value of $3? Drawing from past experience, which do you think will be easier to achieve and why?

- According to my experiences, the bigger impact depends on how the revenue model is established, customer retention is definitely the biggest lever for revenue in longer run but in AOV can impact revenue significantly in a shorter term.

1. If the business model is subscription based then I would want more customers to stay with me and in that terms increasing order retention will be more profitable since you get 2-fold advantage
    a. Recurring gains
    b. Lower acquisition cost (Lesser marketing cost)

2. If the business model is a non-subscription based (eg. marketplace), then more I sell will always be profitable for the business, AOV is a simple money maker which should add extreme impact to the business revenue

3. In a blend of both the models, both metrics will be equally beneficial.
    a. Improving **customer retention of 500 bips** would be a fantastic idea and will be a no brainer since it brings recurring profits to the business, but at the same time it would be very hard to achieve and may require additional marketing costs.
    b. On the other hand, **average order value (AOV)** will bring immediate profits and compared to acquiring more retainable customers, this is an easier choice in terms of cost.

4. According to my understanding **AOV** can be easily improved by adding more products or merchants, decreasing the price of the products and selling them in bundles / as accessories to persuade customers to add more items to the cart, where as **customer retention** requires a whole channel of marketing (acquisition, customer relations ships, 1 to 1 interactions etc.) which can be very challenging.,

5. In conclusion, I would have AOV increase as a short / medium term goal to boost my profits, and simultaneously have a long term goal of improving customer retention

Question 3

Suggest two other data sets you wish you had access to and which analytic techniques you would suggest for mining the overall data set for further insights.

1. **User Engagement:** Another very crucial data source would be user behavior websites, like pageviews, clicks etc, which give a better idea about which type of products are generally people looking at and are very popular in the market, to focus the marketing efforts towards those specific products which will help in optimizing and directing the growth and marketing. Also user engagement goes hand in hand with user retention to identify customers needs and provide them the best user experience.

2. **Customer experience databases:** This includes databases like **Operations, Fulfillment and customer support** databases which would be very beneficial, and should give a fair understanding of how the orders are impacting the GMV, importantly how many of the orders placed are just canceled by customers or we don't have inventory to fulfill the orders etc. Moreover customer feedbacks and operational gaps would be a good to have in order to identify the loopholes in the business and improve upon it to provide the best experience

Data Mining techniques:

1. The best way for data mining is to **track daily, weekly, monthly, yearly trends** and perform **gap analysis** for identifying the issue that needs fixing.

2. In terms of customer related data, I think **clustering** a very useful techniques, which helps in tie knot to each and every behavior of the customer and is a very powerful way to identify different trends and potential opportunities to improve user experience as well as increase revenue

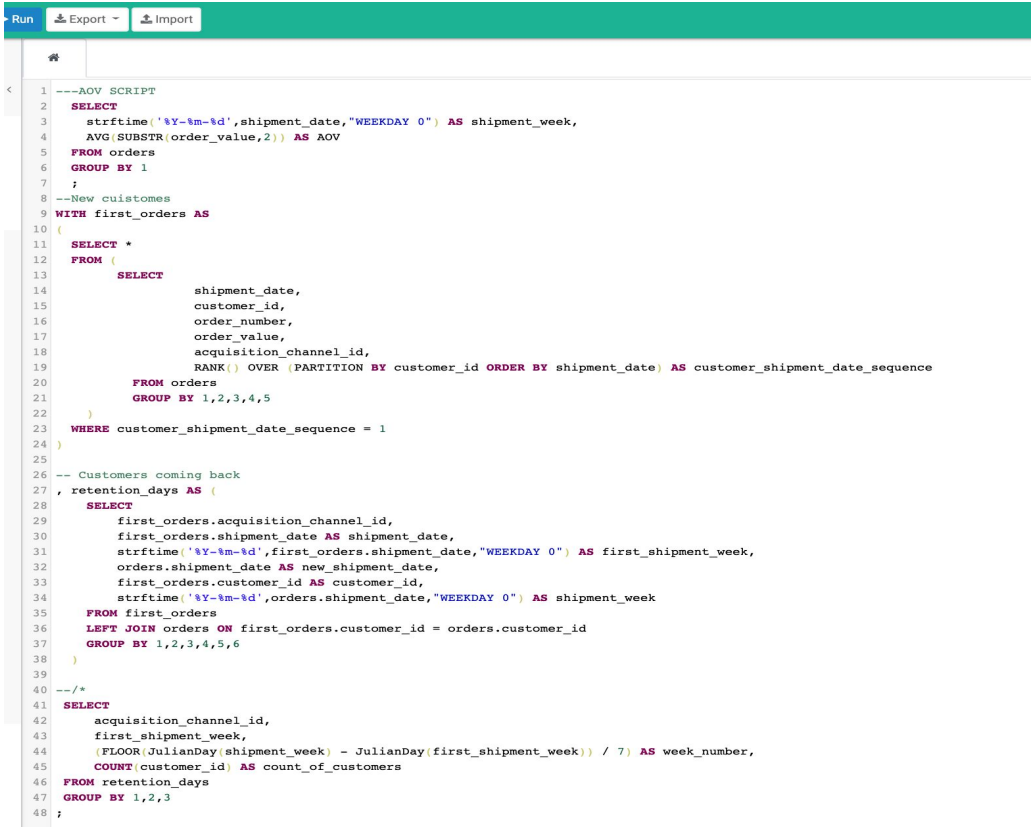
3. Prediction and forecasting is again a very good technique for growth and marketing, which helps in identifying the potential in the market and provides a guideline on how to churn your business.

4. Lastly one of the techniques which I have been using a lot recently on one of my projects is outlier detection, data can be skewed in many ways and sometimes the skewed data can cause a huge anomaly which can directly affect customer experience, and revenue.

Scripts

Please find the scripts in the following sheet:

https://docs.google.com/spreadsheets/d/1PXyqyYIecv5RVkzsajqojqhSH79ak7K1TX8GXgq7qSo/edit#gid=1137483304




```
---AOV SCRIPT
  SELECT
    strftime('%Y-%m-%d',shipment_date,"WEEKDAY 0") AS shipment_week,
    AVG(SUBSTR(order_value,2)) AS AOV
  FROM orders
  GROUP BY 1
  ;
--New cuistomes
WITH first_orders AS
(
  SELECT *
  FROM (
        SELECT
                  shipment_date,
                  customer_id,
                  order_number,
                  order_value,
                  acquisition_channel_id,
                  RANK() OVER (PARTITION BY customer_id ORDER BY shipment_date) AS customer_shipment_date_sequence
          FROM orders
          GROUP BY 1,2,3,4,5
    )
  WHERE customer_shipment_date_sequence = 1
)

-- Customers coming back
, retention_days AS (
    SELECT
        first_orders.acquisition_channel_id,
        first_orders.shipment_date AS shipment_date,
        strftime('%Y-%m-%d',first_orders.shipment_date,"WEEKDAY 0") AS first_shipment_week,
        orders.shipment_date AS new_shipment_date,
        first_orders.customer_id AS customer_id,
        strftime('%Y-%m-%d',orders.shipment_date,"WEEKDAY 0") AS shipment_week
    FROM first_orders
    LEFT JOIN orders ON first_orders.customer_id = orders.customer_id
    GROUP BY 1,2,3,4,5,6
  )

--/*
 SELECT
     acquisition_channel_id,
     first_shipment_week,
     (FLOOR(JulianDay(shipment_week) - JulianDay(first_shipment_week)) / 7) AS week_number,
     COUNT(customer_id) AS count_of_customers
 FROM retention_days
 GROUP BY 1,2,3
 ;
```


Thank You