Sample

“Mevod”

Towards Growth Hacking

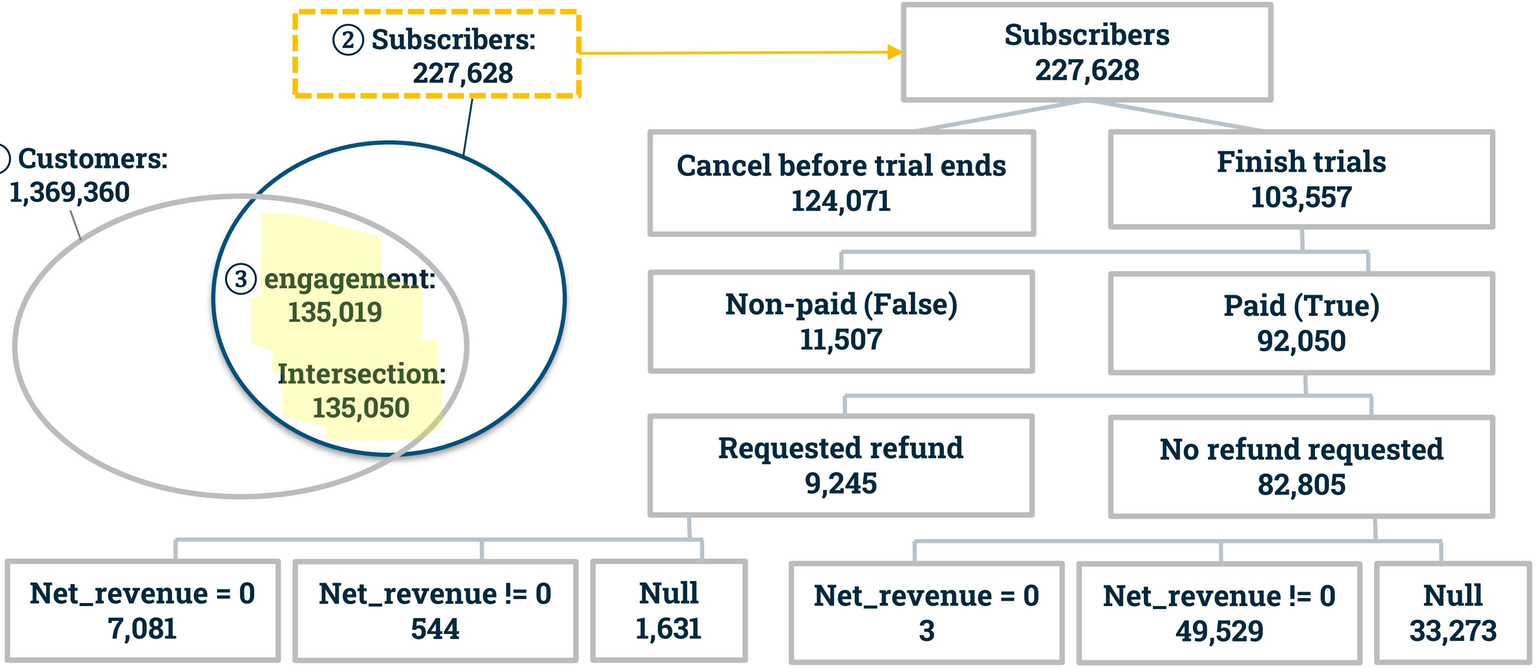
Jiyoung Kim
Spring 2020

Table of contents

Mevod: Basic information

- Towards the dominant regional market player
  - Focusing on local customers is key
  - Acquiring or producing native content (mostly generated in Egypt and Turkey now)
  - Offering superior customer service to establish their committed presence
- Business model
  - On mobile, tablet from iTunes or Google Play stores
  - On smart TV by mirroring these apps or through downloading an app in the TV app store operated by the parent Telco (cable and internet bills from Telco)
- Pricing Schemes
  - No trial fee | Discounted trial fee | 14-day trial period | 7-day trial period
  - Currently, all are committed to 4-month contracts

Data structure



Note: Figures indicate the number of unique customers

Trial Churn modelling

Part 1

1. Prediction model
2. Revenue Analysis
3. Distribution of LTV

1. Churn Prediction – Summary of Results

- Narrowed the focus down to trials period and built a prediction model that classified churners
    - Decided to focus on trial churn as we noticed greater churn rate at this phase as a result of funnel analysis
    - ‘not churn’ is defined as conversion to paid subscription from a trial period with no refund made

- The final random forest model shows 86% accuracy and 0.92 AUC, determining variables being engagement activities i.e. number of videos watched per duration

- Business Interpretation
    - (Revenue analysis) With our prediction model, we simulated offering discounts to predicted churners, which resulted in revenue increase by 3.1% with 50% decision threshold
    - (CLV) A distribution of expected CLV is right skewed and mean CLV is 4.7 with 10% discount rate, but discount rate should be discussed with financial department to better reflect the time value

In which stage we lose our users most?

We lose lots of users in two phases: ‘sign up → completion of trial’, ‘paid subscription → renewal’

- Critical Path (Customer Journey)

Acquisition (Sign up) → Activation (Finish trials) → Revenue (Subscribe paid plan) → Retention (Renewal)

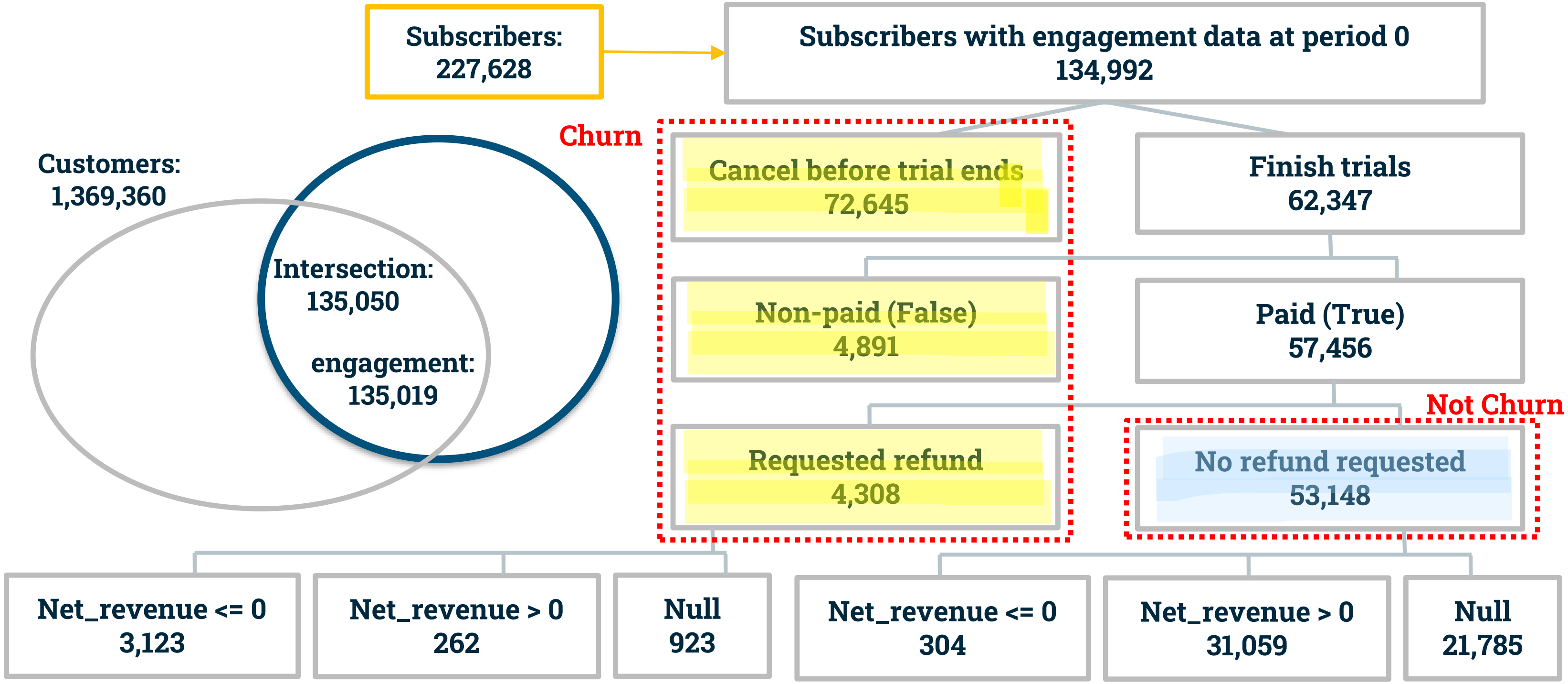
- Funnel Analysis on ‘customer dataset’ (where we have most number of users)

	Sign up	Finish Trials	Subscribe paid plan	Renewal
Number of users	1,369,360	708,433	543,852	127,193
Ratio	100%	51.73%	76.77%	23.39%

- Interpretation

1) After signing up, 52% of users finish the trial period → 48% of the users left before finishing trials
2) 77% of users who finished trial converted to paid subscription → most of them converted to paid subscriptions
3) 24% of users who converted to paid subscription made at least one renewal → 76% of users left before renewal

Modelling Procedure - Data structure for churn modelling



Prediction Models

Built up prediction models and selected Random Forest that predicted churners with 86% accuracy

- Model Performance Comparison (Metrics)

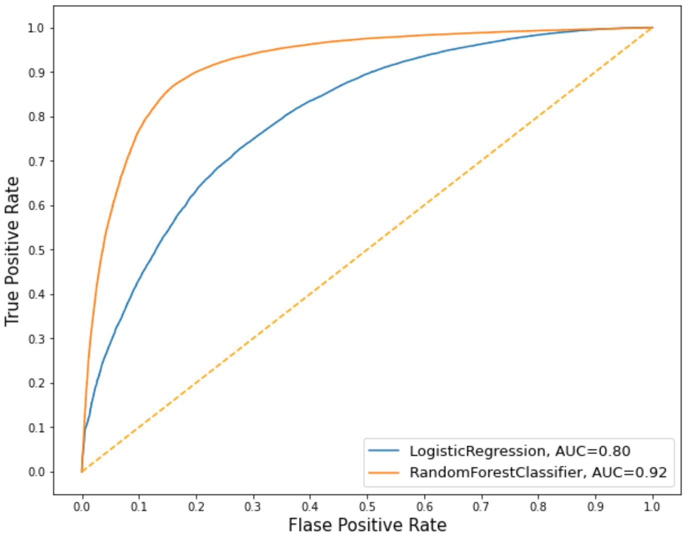
	Accuracy	F1-score	AUC
Logistic Regression	0.74	0.80	0.80
Random Forest	0.86	0.89	0.92

- Accuracy: A proportion of model outputs that tally with test data
- F1-Score: A measure of balance between Precision and Recall
- AUC: How capable in distinguishing two output classes

- ROC curve shows the tradeoff for different threshold

1) **Increasing decision threshold**: Fewer predicted churners
   - Good to use when direct cost of action matters
2) **Decreasing decision threshold**: More predicted churners
   - Good to use when opportunity cost of passing out one user matters

- ROC Curve



Random Forest (orange line) showed better performance than Logistic Regression (blue line) in almost all range of x-axis

Nudging customers could add 1~3% revenue growth

Pro forma churn impact illustration

Decision threshold: 50%

base price: AED 4.5, offer price: AED 2.3

Count of id	Column Labels				*Lift/(loss):*	*3.1%*
Row Labels	0	1	Grand Total		Base revenue	Proforma rev.
base	21,063		21,063			$94,784
churn		31,229	31,229			$0
offer	194	1,511	1,705			$3,836
Grand Total	21,257	32,740	53,997		$95,657	$98,620

Decision threshold: 66.7%

Count of id	Column Labels				*Lift/(loss):*	*2.8%*
Row Labels	0	1	Grand Total		Base revenue	Proforma rev.
base	21,082		21,082			$94,869
churn		31,358	31,358			$0
offer	175	1,382	1,557			$3,503
Grand Total	21,257	32,740	53,997		$95,657	$98,372

Decision threshold: 90%

Count of id	Column Labels				*Lift/(loss):*	*1.9%*
Row Labels	0	1	Grand Total		Base revenue	Proforma rev.
base	21,221		21,221			$95,495
churn		31,903	31,903			$0
offer	36	837	873			$1,964
Grand Total	21,257	32,740	53,997		$95,657	$97,459

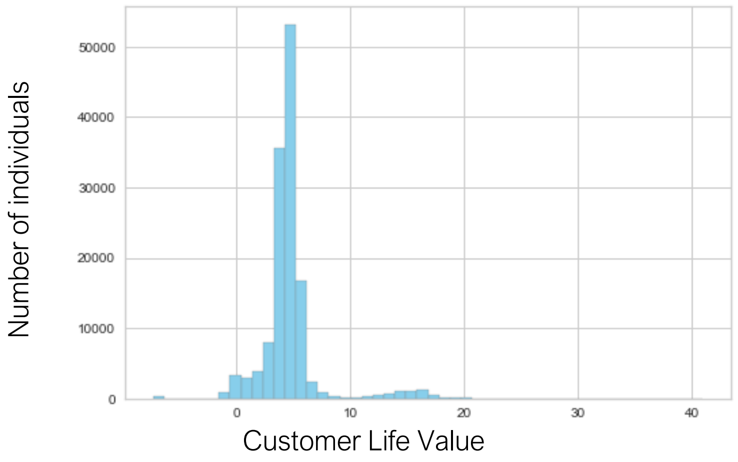
Key Takeaway

- (Scope) Conducted on test set - 53,997 users
- (Revenue modelling) Simulated offering 50% discounts to the predicted churners with 5% acceptance rate
  - (Base) The expected revenue was AED 95,657 without offering discounts
  - (Results) The revenue increased by 3.1% with 50% threshold
- (Further discussion) Contingent on offer acceptance rate, decision threshold, and discount rates

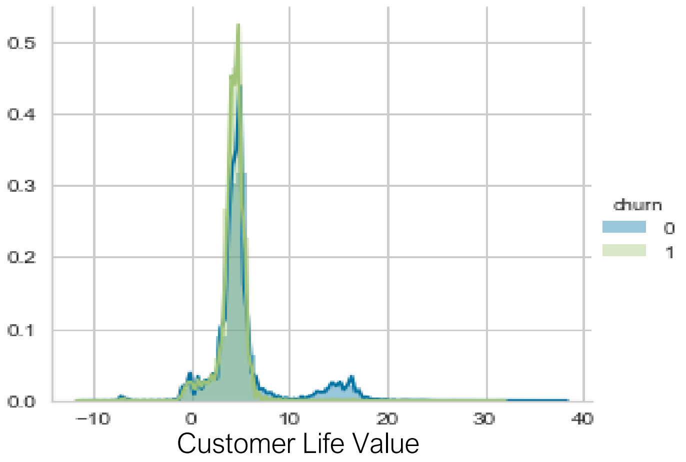
Distribution of Customer Lifetime Value

A distribution of expected CLV is right skewed and mean CLV is AED 4.7

Distribution of expected CLV



Distribution of expected CLV on 'churn' and 'not churn' group



Interpretation

- (Scope) This analysis is based on churn prediction - 134,992 users
- (Results) Mean CLV: AED 4.7, Maximum CLV : AED 37.95, Minimum CLV : AED -11.48

Advertising channel efficiency

Part 2

1. CAC comparison
2. CAC vs LTV
3. Advertising strategy

2. Advertising channel efficiency– Summary of Results

- Advertising channels effectiveness: Facebook > Search > Pinterest > Google > Referral > Affiliate > Email > Email blast
    - Channel effectiveness seemed robust regardless that a customer is converted or not
    - Mean CAC was 0.91 on Facebook, 1.20 on search, 1.52 on google, 3.90 on email, and 5.41 on email blast

- ‘Facebook, Search, Pinterest, Google, Referral’ are our trusted marketing channels
    - Lower CAC across time, 87% of customers are acquired from these channels
    - Higher churn rate of 63% as compared to 53% of the remaining channels

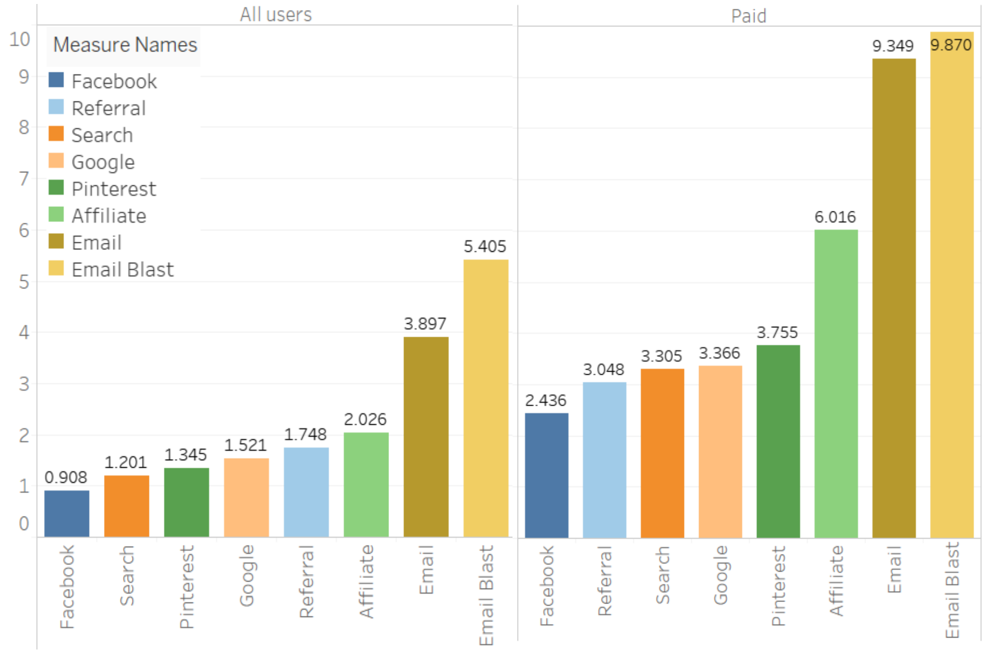
- ‘Affiliate, Email, Email blast’ channels needs further investigation
    - Higher CAC in general but with fluctuation, meaning it decreased significantly for the last 2 months when budgets were cut down, and went up rapidly during tier 4~7
    - Further experiments to test marginal CAC is desired

- LTV was less than ‘3 * CAC’ for all channels, indicating that our business is not in a good shape
    - Optimization of paid ads channels and discovery of non-paid channels are encouraged to reach an ideal state

Effective channels are consistent in both all vs paid users

Paid users seem to prefer similar channels with population users

CAC by Channel for "all vs paid"

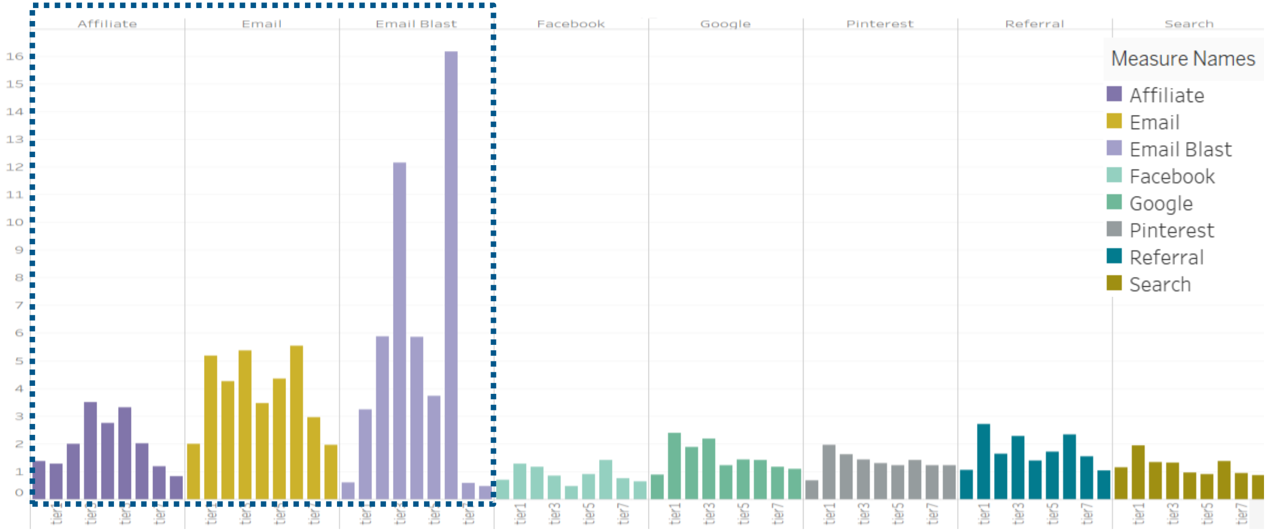


- We looked CAC of two groups for robustness check
- (Results) Channel effectiveness seemed robust regardless that a customer is converted (paid) or not
- (All) Mean CAC was 0.91 on Facebook, 1.20 on search, 1.52 on google
- (Paid) To gain a customer who would convert to paid subscription, we spent 2.44 on Facebook, 3.05 on referral, 3.67 on google

Does CAC per channel vary as time passes?

CAC of Email blast, Email, Affiliate channels fluctuates

CAC over time by channel

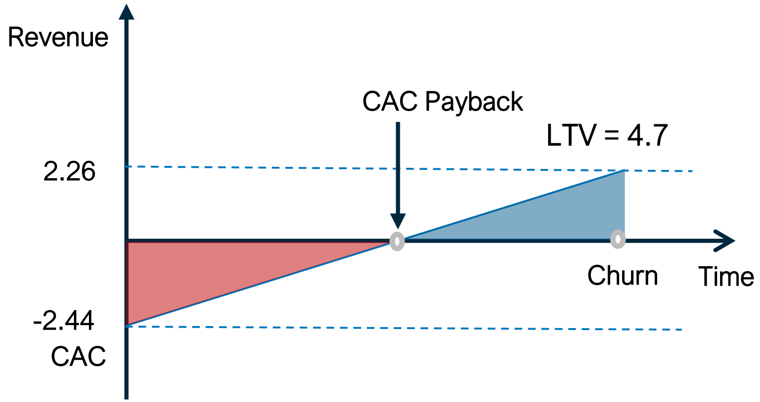


- CAC went up rapidly during tier 4~7 for channels of Email blast, Email, Affiliate (subject of future analysis)
- Other channels e.g. Facebook, search, referral show relatively stable CAC across time frame

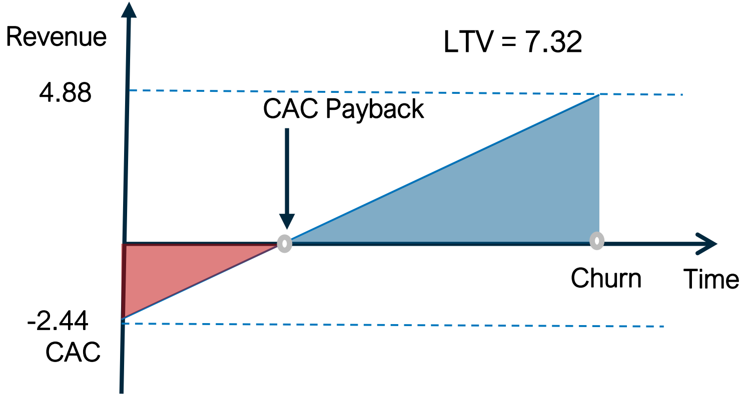
Are we earning more than spending? - Facebook

We pay 2.44 to get a new paid user, but that user spends 4.7 and churns

- Current: LTV < 3 * CAC



- Ideal: LTV >= 3 * CAC



Key Takeaway

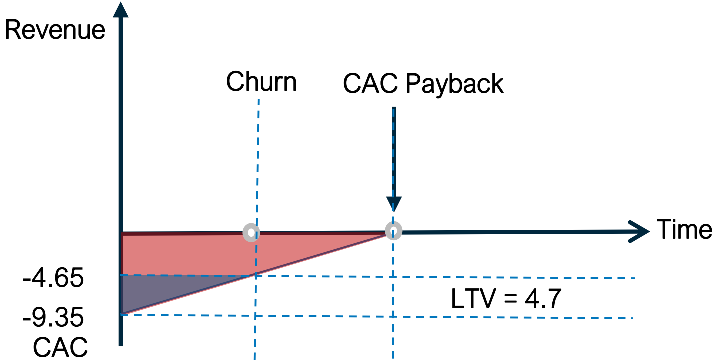
- Facebook is the most effective channel and LTV > CAC (earning more than acquisition cost)
- LTV < 3 *CAC, not good enough to maintain business

Note: Mean LTV was used, which is calculated based on trial churn modelling

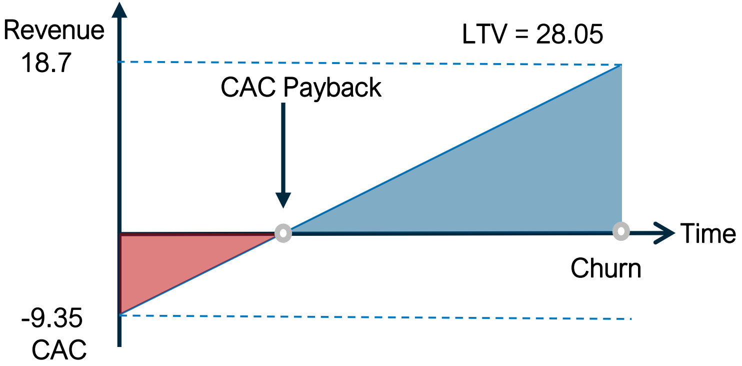
Are we earning more than spending? - Email

We pay 9.35 to get a new paid user, but that user spends 4.7 and churns

- Current: LTV < CAC



- Ideal: LTV >= 3 * CAC



Key Takeaway

- Email is the second least effective channel and LTV < CAC, we do not reach payback point
- To satisfy LTV > 3 *CAC in this channel, LTV should be greater than 28.05

Note: Mean LTV was used, which is calculated based on trial churn modelling

Advertising strategy recommendations

Advertising channel optimization

- Budget allocation on upcoming quarter
  - Facebook, Search, Pinterest, Google, Referral have lower CAC than other channels
  - Email, Email blast, Affiliate have lower churn rate; there is a value in keeping these channels
- Effectiveness of media channel spend
  - Design an experiment to test Marginal CAC specifically for Email, Email blast, Affiliate
- Digital media targeting
  - Incorporating segmentation information into media buys

Recommended future analysis

- Campaign ROI
- Sentiment Analysis

A/B Testing

Part 3

1. Testing structure
2. Interpretation of results
3. Repeated trials
4. Business impact

3. A/B Testing – Summary of Results

- We ran an A/B test to compare whether the removal of join fees leads to an increase in conversion* rate (*conversion to paid subscription after trial ended)
- The removal of join fees led to a statistically significant lift in conversion of 3.37% as compared to base (paying join fees)
  - The optimal sample size was 2,946, and sequential test required 745 observations on average, future A/B tests could benefit from reduced sample size (current sample size 33,495)
- Business Interpretation
  - Total revenue is likely to increase despite of the loss from join fee due to the gain on increased conversion; however, the result will differ depending on the number of users, subscription fees, and join fees
  - Further research on how join fees disincentivize user experience might be required before waiving join fees

Testing Procedure

- Hypothesis

  Users of ‘Mevod’ are more likely to convert if there are no join fees.
  Hence, waiving join fees to variant B group will show an increase in conversion to paid subscriptions

- Experimental design

  I. Variant A: Users need to pay join fees
  II. Variant B: Users do not need to pay join fees

- Sampling method

  Variant A had 159,229 users (82.6%) and Variant B had 33,482 users (17.4%)

Data Preparation

- Scope
  - Total 227,628 customers in ‘subscribers' dataset as join fees are known for this dataset only
  - Out of 226,628 users, 34,904 users who had null value were excluded in this analysis
  - There are negative values for join fees which accounted for 0.0067%; considered as ‘no join fee’

- Conversion variable
  - ‘paid_TF’ (True/False): if user has made a successful payment

Does Variant B affect conversion rates?

Waiving join fees improved the conversion rate from 29.1% to 32.4% at 95% significance level

- Hypotheses

  H_o: Conversion A = Conversion B
  H_A: Conversion A ≠ Conversion B

- Test statistic

  Calculate z-score $z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$ where $\hat{p}$ = Conversion B, p = Conversion A, n= 33,495

  Z-Score = 13.577 > $Z_{\alpha/2}$ = 1.96

H_A: **Conversion_B ≠ Conversion_A**	H_A: **Conversion_B > Conversion_A**
True	**True**

- Hence, we reject the null hypothesis, Variant B resulted in a 3.37% lift

Assumptions

1. Variant A represents the population; therefore, we use a one-sample z-test
2. Use 2-tailed test to consider both scenarios: lift and hurt
3. Sampling is done in a representative manner

What is the optimal sample size?

The optimal sample size is 2,946

- Calculate size using this formula

$$n^* = \left(t_{\alpha/2}\sqrt{2\bar{p}(1-\bar{p})} + t_\beta\sqrt{p_0(1-p_0)+p_1(1-p_1)}\right)^2 \frac{1}{\delta^2}$$

where p-bar = {(Conversion A) + (Conversion B)} / 2
p0 = Conversion A
p1 = Conversion B
δ = Conversion B – Conversion A (min detectable effect)
tα/2 = 1.96
tβ = 0.842 (t-value of 0.8)

Confidence Level = .95
Power = .80

Assumptions

1. Data is large enough to use normal approximation to binomial distribution
2. Sampling is done in a representative manner

Will the results be consistent for repeated trials?

Results are stable for 10 repeated trials with the optimal sample size, with a two-sided z-score test

- For each trial, we used the optimal sample size 2,946
  - Minimum detectable effect of 3.37%
  - Power = 80%
  - Alpha = 5%
  - Variance of observations for Variant A and B are equal

- Results

Trial	1	2	3	4	5	6	7	8	9	10
Conversion rate B	32.1%	32.1%	32.3%	32.1%	32.6%	33.3%	32.8%	31.4%	32.1%	33.9%
Z-score	3.68	3.64	3.88	3.64	4.25	5.10	4.45	2.83	3.64	5.79
Reject Null	True	True	True	True	True	True	True	True	True	True

Assumptions

1. Population proportion is known and equal to Variant A proportion; therefore, we use a one-sample z-test
2. Trials are independent and identically distributed (IID)
3. Sampling is done in a representative manner

Sequential Testing reduces the sample size by 74%

The average number of observations was 745 iterations for 10 repeated trials

- For each trial,

  Minimum detectable effect of 3.37%
  Power = 80%
  Alpha = 5%
  Population mean (proportion) = 29.06%

- Results

Trial	1	2	3	4	5	6	7	8	9	10
Number of observations	971	911	513	207	582	926	754	636	252	1916
Reject Null	True	True	True	True	True	True	True	True	True	True

- Steps

1. Randomly sample 2,946 users from Variant B, without replacement
2. P(xi=1) = 0.2906 under H0; and P(xi =1) = 0.3243 under H1
3. From Type I error = 5% and Type II error = 20%, calculate upper and lower bounds
   a. Upper = ln(1/α) = 2.99
   b. Lower = ln(β) = -1.6
4. Observe first outcome, X1 (the first observation of the sampled)
5. Update ln(λn) = ln(λn -1) + ln(λ(xi))
   a. If Xi =True, ln(λ(xi)) = ln (p1/p0)
   b. If Xi = False, ln (λ(xi)) = ln((1-p1)/(1-p0))
6. If ln(λn) > Upper, stop test and reject Null Hypothesis
   If ln(λn) < Lower, stop test and do not reject Null Hypothesis
7. Else, continue test (repeat step 4-7)

Business impact

Believability of Results

- We may need more qualitative research on why join fees are contributing to an increased conversion rate and how join fees disincentivize user experience

Assessing Business Value

- The average join fee is 0.1151 whereas average monthly subscription fee is 4.7320. Assuming waiving join fees increase conversion rate by 3.37% on 100 users, total revenue increased as following:

Trial	Revenue from monthly fee	Revenue from join fee	Total Revenue
Before	100 * 29.04% * 4.7320 = 137.54	100* 0.1151 = 11.51	141.97
After	100 * 32.43%* 4.7320 = 153.48	-	**153.48**

- This calculation is based at the point of conversion (no renewal was counted) with only 100 users, hence, the revenue change will differ depending on the number of users, monthly fees and join fees
- Scale of implementation and reproducibility of results may not be the same in different cities across geographical boundaries

Appendix

A) Cohort Analysis
B) Churn model in details
C) CAC Calculation

Appendix A. Cohort Analysis

Retention Levels by Cohort

Cohort Table by period – number of users alive at each period

	period 0	period 1	period 2	period 3	period 4	period 5	period 6	period 7	period 8	period 9	period 10	period 11	period 12
tier 1	237,436	104,955	42,352	13,892	4,464	3,212	1,388	1,088	890	766	503	1	1
tier 2	103,644	44,022	18,307	5,052	3,104	1,633	1,140	896	725	341	3	3	2
tier 3	114,977	50,752	21,228	5,619	3,717	1,758	1,320	1,024	452	5	2		
tier 4	113,522	46,415	16,210	5,270	2,578	1,677	1,276	477	6				
tier 5	191,968	80,539	25,084	8,046	3,164	2,290	747	3					
tier 6	133,429	55,228	13,593	5,135	3,054	1,221	8	1					
tier 7	104,005	45,377	11,012	5,317	1,624	23							
tier 8	168,230	71,045	8,410	2,313	42	3							
tier 9	202,149	83,070	4,337	62	9								

Interpretation

- Tier1 means users who created their account between '2019-06-30' and '2019-07-30' and Tier2 means users who created their account during the following month
  Tier1 had 237,436 users at payment period_0 and 104,955 moved to period_1
  Among them, 42,352 users moved to period_2 and we had only one user at period 12

- We are losing lots of customers in the initial stage and have less than 10,000 users from period 4

Appendix A. Cohort Analysis

Retention Rates by Cohort

Cohort Table by period – percentage of users alive at each period

	period 0	period 1	period 2	period 3	period 4	period 5	period 6	period 7	period 8	period 9	period 10	period 11	period 12
tier 1	100.0%	44.2%	17.8%	5.9%	1.9%	1.4%	0.6%	0.5%	0.4%	0.3%	0.2%	0.0%	0.0%
tier 2	100.0%	42.5%	17.7%	4.9%	3.0%	1.6%	1.1%	0.9%	0.7%	0.3%	0.0%	0.0%	0.0%
tier 3	100.0%	44.1%	18.5%	4.9%	3.2%	1.5%	1.1%	0.9%	0.4%	0.0%	0.0%		
tier 4	100.0%	40.9%	14.3%	4.6%	2.3%	1.5%	1.1%	0.4%	0.0%				
tier 5	100.0%	42.0%	13.1%	4.2%	1.6%	1.2%	0.4%	0.0%					
tier 6	100.0%	41.4%	10.2%	3.8%	2.3%	0.9%	0.0%	0.0%					
tier 7	100.0%	43.6%	10.6%	5.1%	1.6%	0.0%							
tier 8	100.0%	42.2%	5.0%	1.4%	0.0%	0.0%							
tier 9	100.0%	41.1%	2.1%	0.0%	0.0%								

Interpretation

- The trend is clearer with percentages of the remaining users at each period
- By looking at period 2, we can see that the retention rate keeps going down from 17.8% to 2.1%, meaning not as much as customers continue to stay as before

  Recently acquired users are not satisfied with our service and left the platform early, resulting in lower retention rate in earlier stages of customer life cycle
- At period 4, only 1~3% of users remained out of 100%

Appendix B. Churn model in details

Logistic Regression vs Random Forest

Logistic Regression

- Performance Metrics
  Accuracy: 0.74, F1-score: 0.80, AUC: 0.80

  Predicted potential churners with 74% accuracy and AUC of 0.80

- Confusion Matrix

	Predicted not churn	Predicted churn	Total
Actual not churn	12,949	8,308	21,257
Actual churn	5,626	27,114	32,740

Random Forest

- Performance Metrics
  Accuracy: 0.86, F1-score: 0.89, AUC: 0.92

  Predicted potential churners with 86% accuracy and AUC of 0.92

- Confusion Matrix

	Predicted not churn	Predicted churn	Total
Actual not churn	16,591	4,666	21,257
Actual churn	2,934	29,806	32,740

We used 'Random Forest ' as our final prediction to conduct revenue modelling

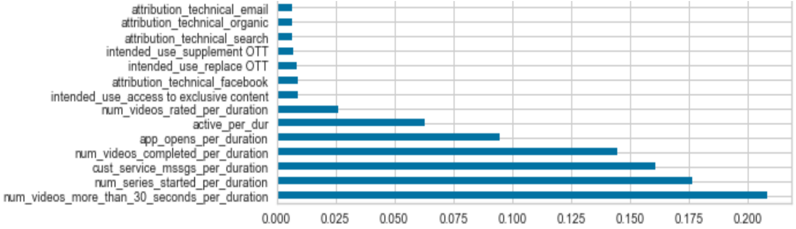
Appendix B. Churn model in details

Significant Features in predicting churn

Triggers and behaviors that equate to churn or retention

- Important features are consistent across logistic regression and random forest models
- Features have strong correlations with ‘churn’ or ‘retention’ are mostly related to engagement activities:

	Coef.	Std.Err.	z	P>|z|
active_per_dur	-0.0141	0.0482	-0.2923	0.7700
app_opens_per_duration	-1.8036	0.2709	-6.6579	0.0000
cust_service_mssgs_per_duration	21.5364	0.3470	62.0561	0.0000
num_videos_completed_per_duration	-1.8888	0.1196	-15.7984	0.0000
num_videos_more_than_30_seconds_per_duration	3.4344	0.1353	25.3803	0.0000
num_videos_rated_per_duration	-34.8361	1.0472	-33.2667	0.0000
num_series_started_per_duration	-4.1104	0.0804	-51.1345	0.0000
attribution_technical_pinterest_organic	-0.7460	0.7420	-1.0054	0.3147
op_sys_Android	0.1692	0.0529	3.1975	0.0014
op_sys_iOS	-0.1191	0.0507	-2.3467	0.0189



Positive to churn

Number of messages sent to customer representatives are likely to be associated with negative experience
Higher number of videos watched over 30 seconds may indicate that users are not happy with contents

Negative to churn

Giving a rating, watching videos through 95%, and using apps often are behaviors related to retention

Appendix C. CAC Calculation

How much we spent to acquire one user?

Facebook, Search, Pinterest, Google are best channels in terms of marketing efficiency

CAC by tier by channel

	date	Facebook	Email	Search	google	affiliate	email_blast	pinterest	referral
Tier 1	2019-06-30	0.71	1.99	1.15	0.89	1.37	0.60	0.68	1.06
Tier 2	2019-07-31	1.27	5.17	1.94	2.38	1.28	3.23	1.97	2.70
Tier 3	2019-08-31	1.16	4.27	1.34	1.89	1.99	5.87	1.62	1.65
Tier 4	2019-09-30	0.86	5.37	1.32	2.19	3.49	12.15	1.44	2.29
Tier 5	2019-10-31	0.46	3.46	0.97	1.23	2.75	5.85	1.31	1.40
Tier 6	2019-11-30	0.91	4.34	0.91	1.44	3.31	3.73	1.22	1.72
Tier 7	2019-12-31	1.41	5.55	1.37	1.41	2.02	16.16	1.42	2.34
Tier 8	2020-01-31	0.76	2.96	0.94	1.16	1.18	0.58	1.23	1.55
Tier 9	2020-02-29	0.63	1.97	0.86	1.10	0.84	0.47	1.22	1.03
mean	-	0.91	3.90	1.20	1.52	2.03	5.41	1.34	1.75

- Advertising channels effectiveness: Facebook > Search > Pinterest > Google > Referral > Affiliate > Email > Email blast
  Facebook, Search, Pinterest, Google shows lower CAC consistently; these are our trusted marketing channels
- Further investigation is needed for affiliate, email, email blast channels
  Despite the high average CAC, there channels showed much lower CAC for the last 2 months when the budget was cut down

Appendix C. CAC Calculation

How much we spent to acquire one paid user?

Marketing Channel Efficiency: Facebook > referral > Search > Google

CAC by tier by channel for paid customers only

	date	Facebook	email	Search	google	affiliate	email_blast	pinterest	referral
Tier 1	2019-06-30	1.89	4.46	3.54	2.14	2.89	1.15	2.35	2.04
Tier 2	2019-07-31	3.53	12.24	5.36	5.17	5.94	6.70	5.38	4.60
Tier 3	2019-08-31	3.25	10.22	3.79	4.06	4.80	11.43	4.67	2.92
Tier 4	2019-09-30	2.46	12.74	3.82	4.62	11.58	23.60	4.35	4.03
Tier 5	2019-10-31	1.22	8.11	2.55	2.65	9.65	10.73	3.48	2.41
Tier 6	2019-11-30	2.39	10.67	2.47	3.32	8.17	7.19	3.45	3.03
Tier 7	2019-12-31	3.72	12.99	3.47	3.14	5.32	26.00	3.85	3.86
Tier 8	2020-01-31	1.92	7.52	2.48	2.65	3.42	1.03	3.10	2.80
Tier 9	2020-02-29	1.54	5.18	2.26	2.54	2.37	1.00	3.17	1.74
mean	-	2.44	9.35	3.31	3.37	6.02	9.87	3.76	3.05

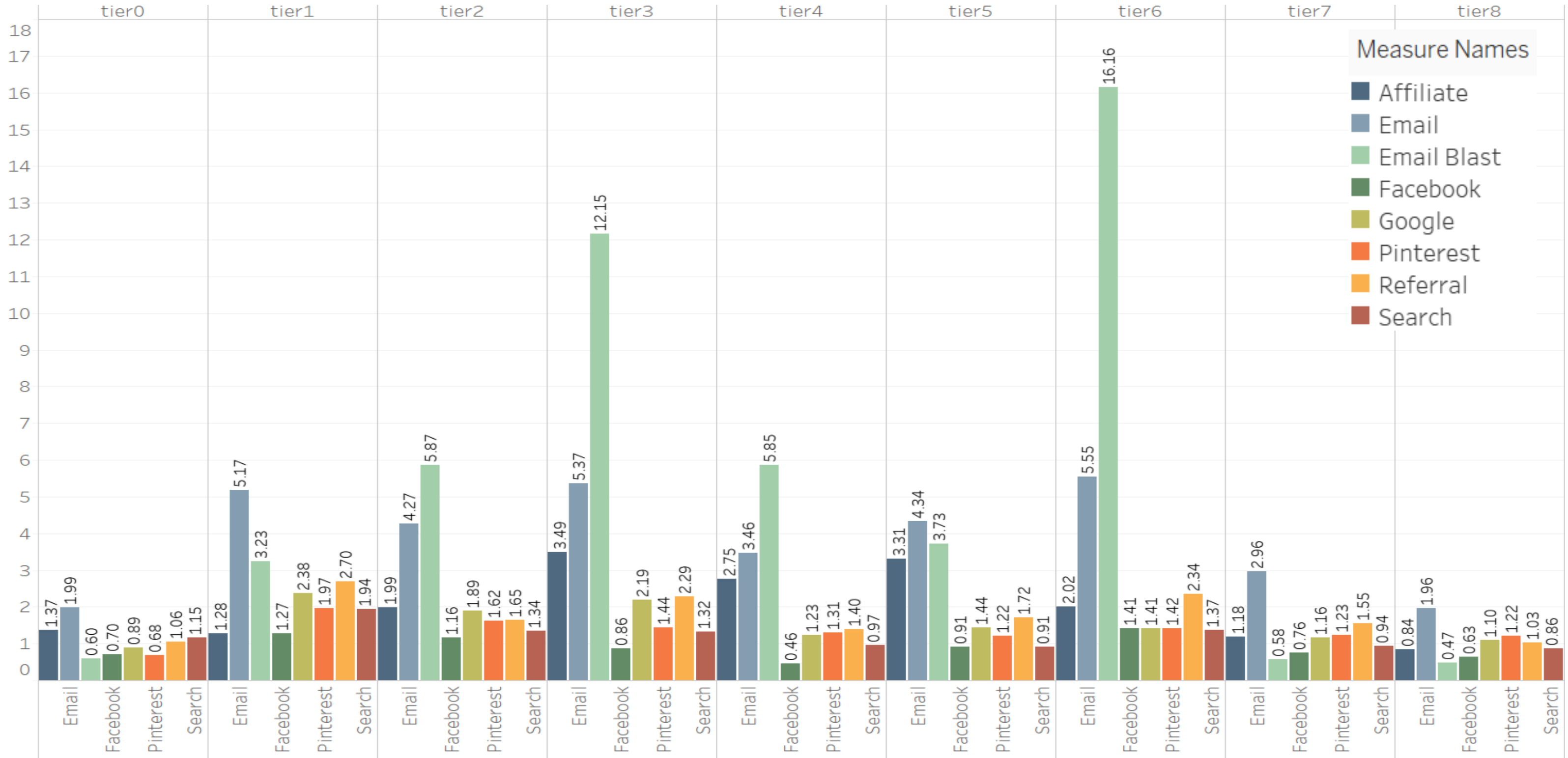
- Advertising channels effectiveness: Facebook > Referral > Search > Google > Pinterest > Affiliate > Email > Email blast
  Facebook, Referral, Search, Pinterest, Google shows lower CAC consistently; these are our trusted marketing channels
- No significance difference between all users and paid users

Appendix C. CAC Calculation

How much we spent to acquire one user?

Marketing Channel Efficiency: Facebook > Search > Pinterest > Google

CAC by tier by channel



Appendix C. CAC Calculation

How much we spent to acquire one paid user?

Marketing Channel Efficiency: Facebook > referral > Search > Google

CAC by tier by channel for paid users