Data Driven Marketing Strategy Exploration for Mevod

Yutong Cai
Github Link

About Mevod

Company Background

About Mevod

Current Situation

- Beach head product to move into the OTT space of a regional traditional Telco
- There are limited players currently in the market

Objective

- Become the dominant regional market player

Tactical Plan

- Acquiring and/or producing native content
- Offering superior customer service to establish their committed presence to the region.

Business Goal

Expected Achievement
Current Business Structure
Marketing Analysis Methods

Business Goal

Support market expansion strategy for Mevod

Current Business Model:

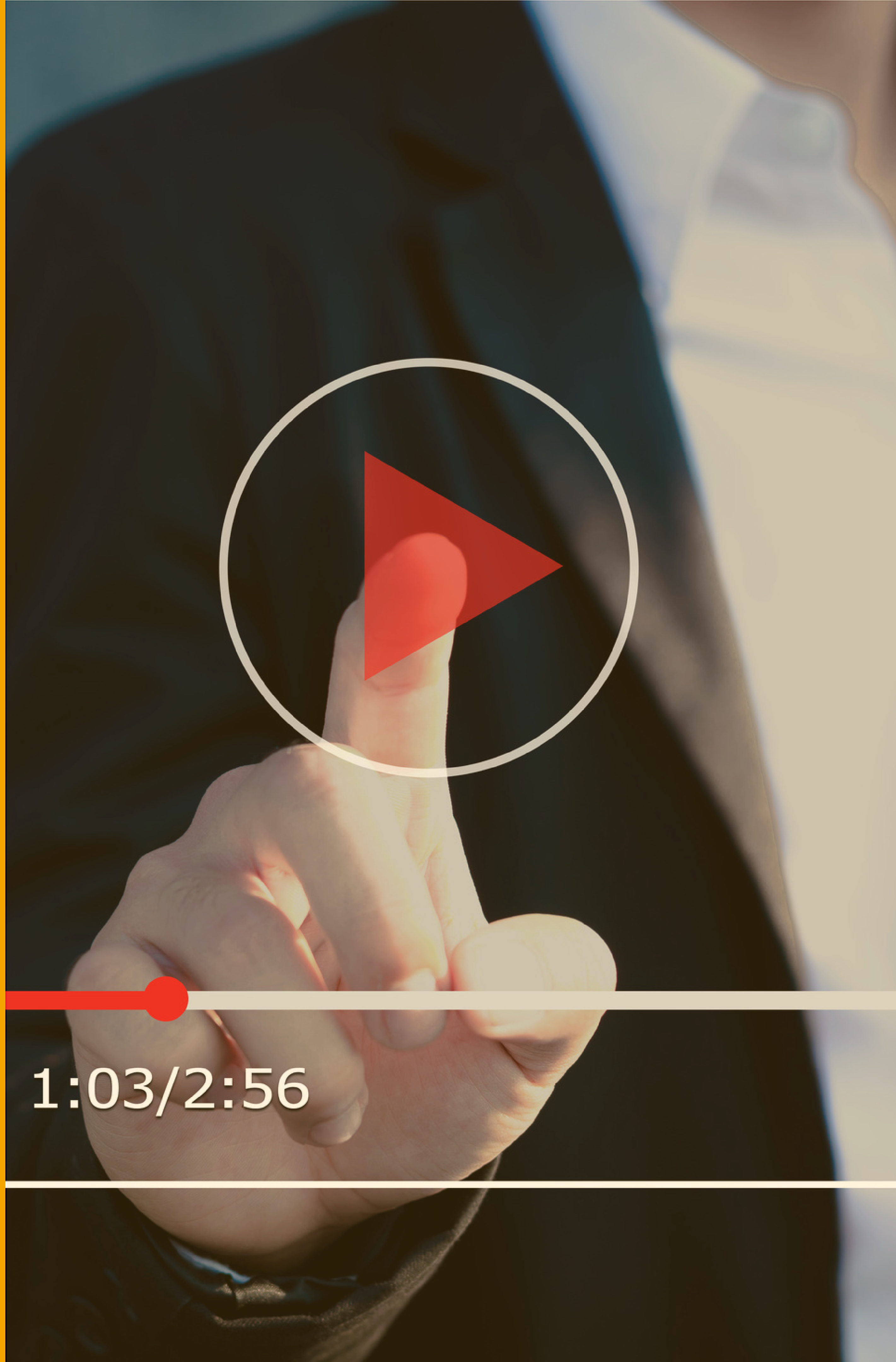
- Subscription business
- Customers pay a monthly fee for access to the service with 4-month contracts
- Customers can watch content through a variety of platforms
- Several pricing schemes
  - No trial fee
  - Discounted trail fee
  - 14 day trial period
  - 7 day trail period
  - High priced monthly plan
  - Low priced monthly plan

Marketing Analysis Method:

- **Customer segmentation:** help the marketing team design acquisition strategies supporting the Executive team's growth objective
- **Allocation:** calculating advertising channel spend efficiency and effectiveness, supporting the advertising team's budget allocation for the upcoming quarter
- **Churn model:** predicting churning probability and develop recommendation(s) on an alternative product structure of pricing and others.

Data Exploration

Data Overview
Data Preprocessing
Early Data Analysis



Data Overview
-- Data Set

SUBSCRIBERS

- 30 columns
- Contains user portrait and information related to the using of the platform
- e.g. age, gender, attribution method, plan_type

CHANNEL SPEND

- 3 columns (Channel, Date, Spend_AED)
- Contains the allocation information of the spending on different channels during different time period (month)

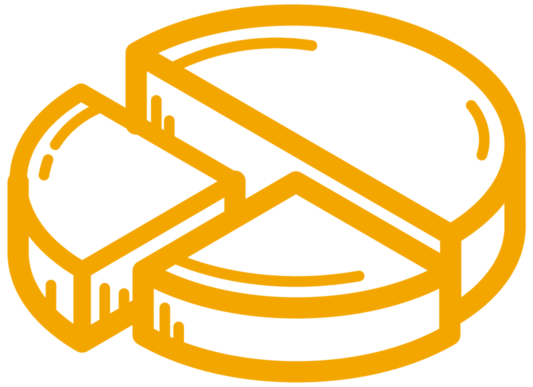
COSTOMER SERVICE

- 13 columns
- Contains users' consumption information correlated with the service
- e.g. whether the user is currently substituted, users' last payment, users' renew situation

ENGAGEMENT

- 9 columns
- Contains users' engagement with the service
- e.g. app open times, number of videos completed

Data Preprocessing



Deal with Missing Value

Since some of the missing values are meaningful, we cannot simply delete the missing values.
(eg. a person may never cancel the service, so the cancel_date should be NaN)
Thus, we fill categorical missing value with 'not_specified' and numerical value with 0 for early data analysis. And some of them will be adjusted when applying methods.

Deal with Outliers

There are some extreme values in the subscribers dataset.
eg. there are negative weekly consumptions, negative subscriber's target number of services, negative join fee, and a user lived for 81720000 year
These values makes the specific row (or the specific user's data) unreliable. Thus, rows with unreasonable values are deleted for data analysis.

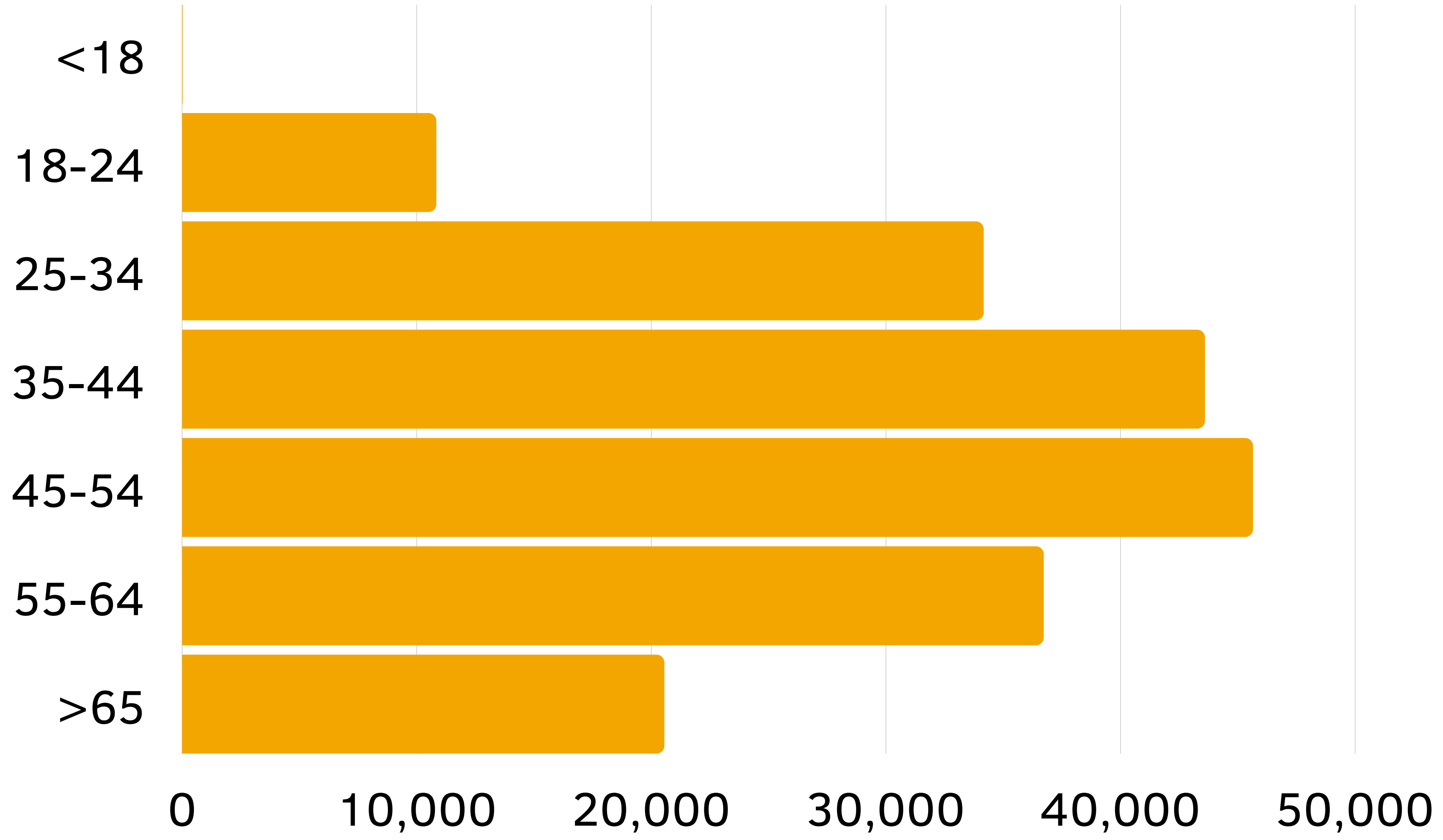
Basic Feature Engineering

Age is classified by range for better user portrait analyzing
Different datetime are created for better clustering

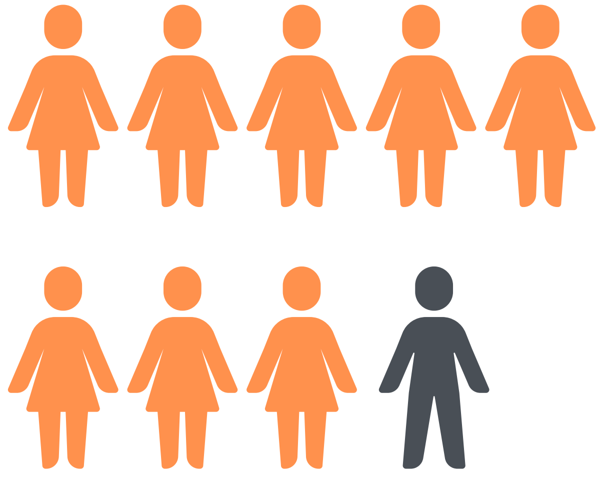
User Portrait

(cleaned data)

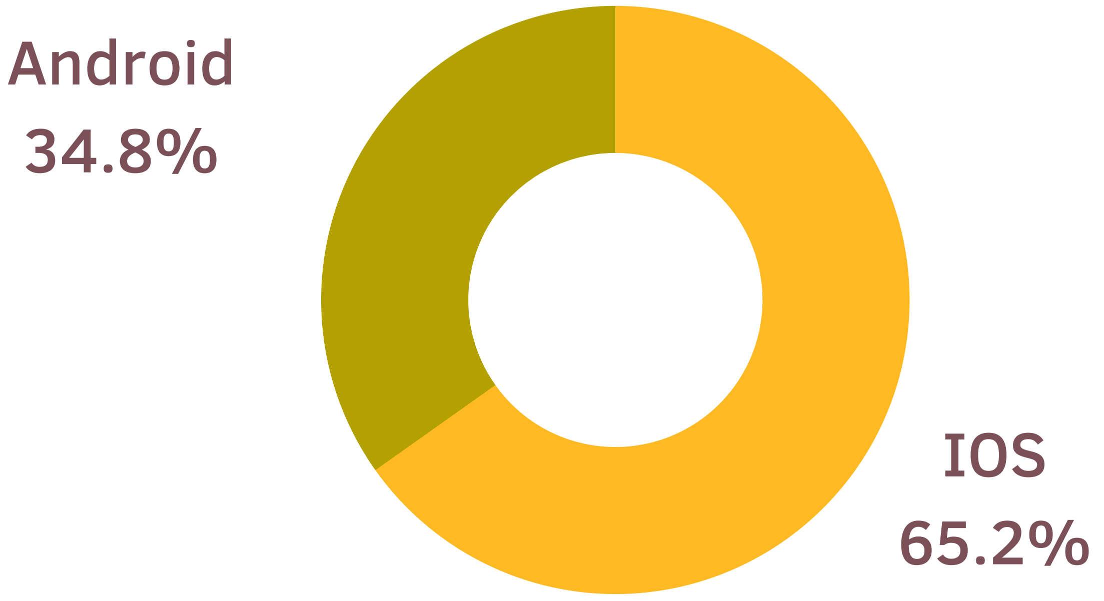
AGE



GENDER



DEVICE



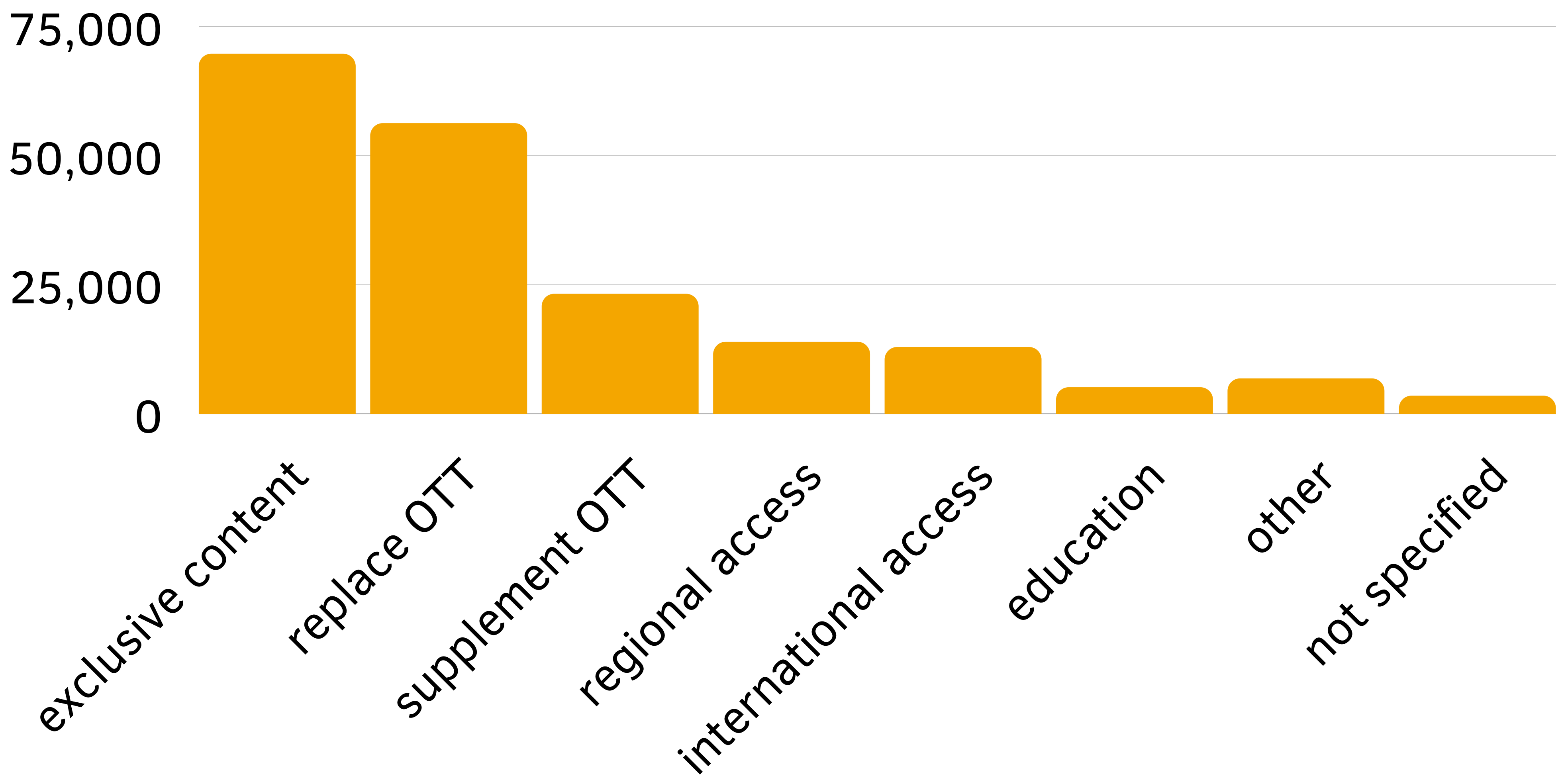
- The distribution of age is very close to normal distribution, most users are between 35-54 years old
- 88.7% of the users are Female
- IOS is the major device used when using this Mevod application (but IOS is not overpowering)
- All of the users are from United Arab Emirates and set Arabic as their primary language

COUNTRY: 100% UAE
LANGUAGE: 100% ARABIC

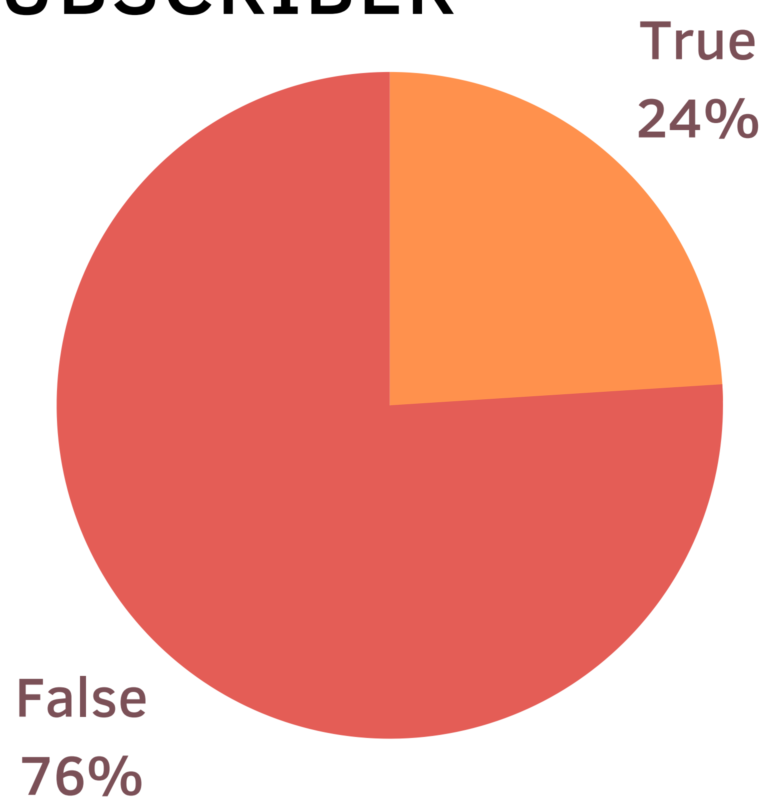
User & Application

(cleaned data)

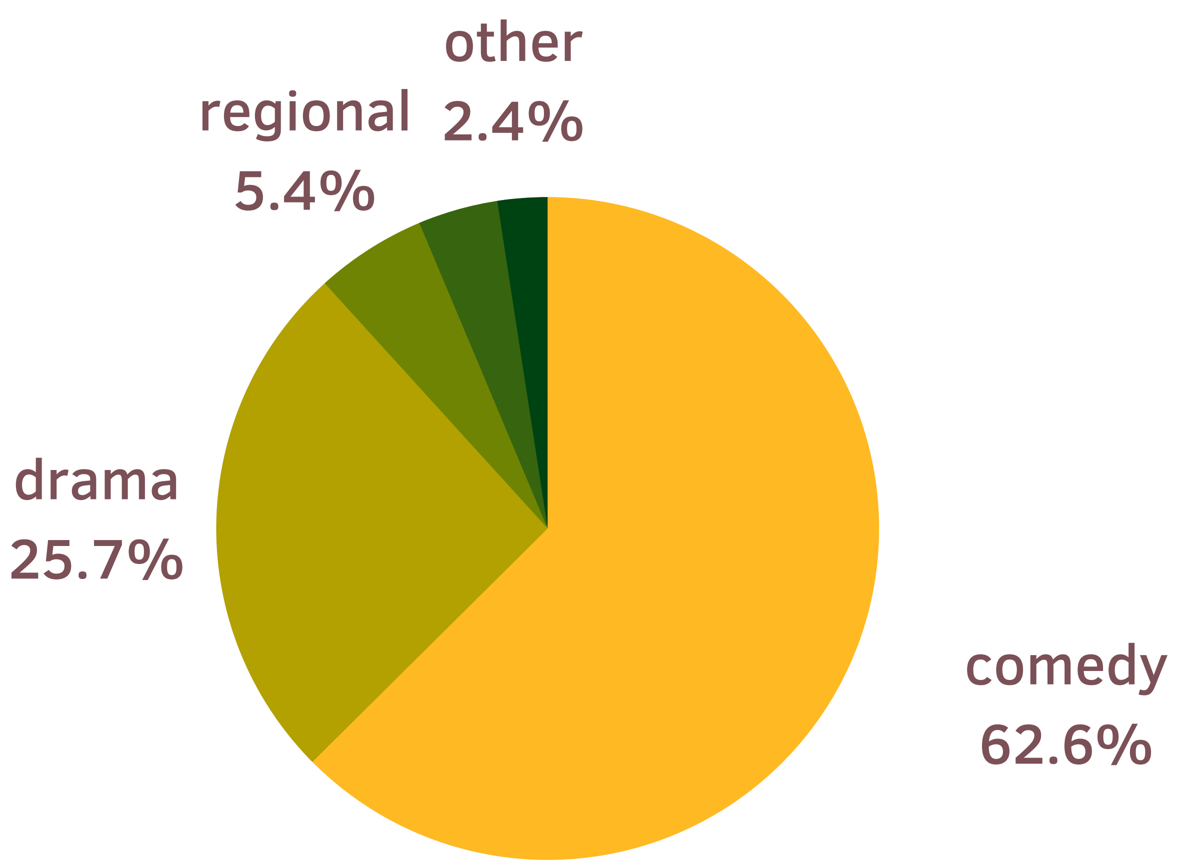
INTENDED USE



CURRENT SUBSCRIBER



PREFERRED GENRE



- The top 1 reason for using the application is "access to exclusive content"
- 24% user are current subscriber, about 76% users left this app
- 88.3% users prefer Comedy and Drama

Segmentation

Dividing customer into groups
Helping the marketing team design acquisition strategies
supporting the Executive team's growth objective

Segmentation

Features

- age
- gender
- preferred genre
- weekly consumption hour
- operation system

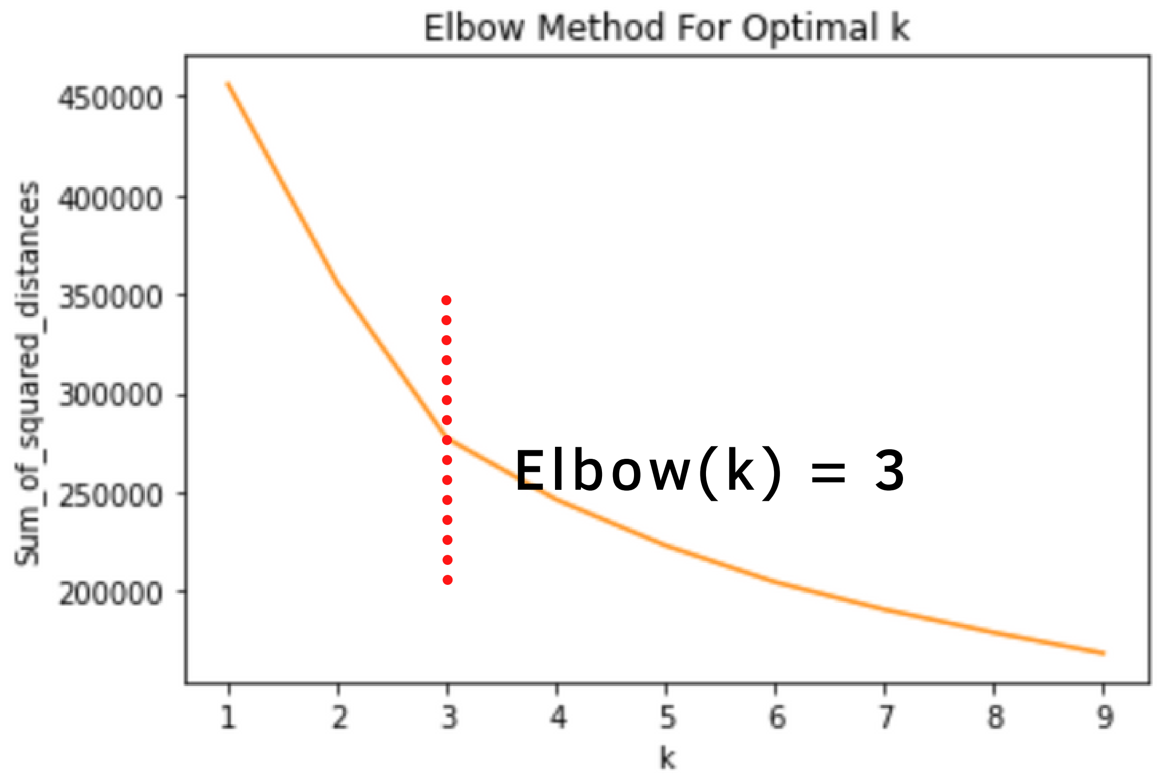
Method for Grouping

- K-means: a model used to cluster different customers
- K-modes: a model used to cluster customer on categorical variables (eg. Location, order time range)

Decision

Elbow Method

- the method used to determine the number of clusters
- finding the 'elbow' point, which diminishing returns are no longer worth the additional cost



Segmentation

Customer Profile

Middle Age

age	34
gender	Female
preferred genre	Comedy
weekly hour	26
operation system	IOS

Elder

age	58
gender	Female
preferred genre	Comedy
weekly hour	26
operation system	IOS

Heavy Users

age	46
gender	Male
preferred genre	Drama
weekly hour	35
operation system	IOS

Strategies from here:

- Male and Female have different using habit and preferred genre, thus need different targeting strategies

Assumptions from here:

- Male prefer Drama more than female
- Male usually spend more time on the app

Segmentation

More on Customer Profile

Assumption I - Supported

Male prefer Drama more than female

Type	Male	Female
comedy	54%	65%
drama	32%	24%
regional	8%	5%
international	4%	3%
other	3%	2%

Assumption II - Supported

Male usually spend more time on the app

Gender	Weekly Hour
Male	35
Female	27

Strategies from here:

- Designing different recommendation from the beginning base on gender if no other information are available since their preferences contain minor differences
- Corresponding to native contents, applying more resource on Comedy and Drama since most users are interested in these two genre
- Investigate more on acquiring male user since 88.7% of the users are female, but male are spending more time on the app

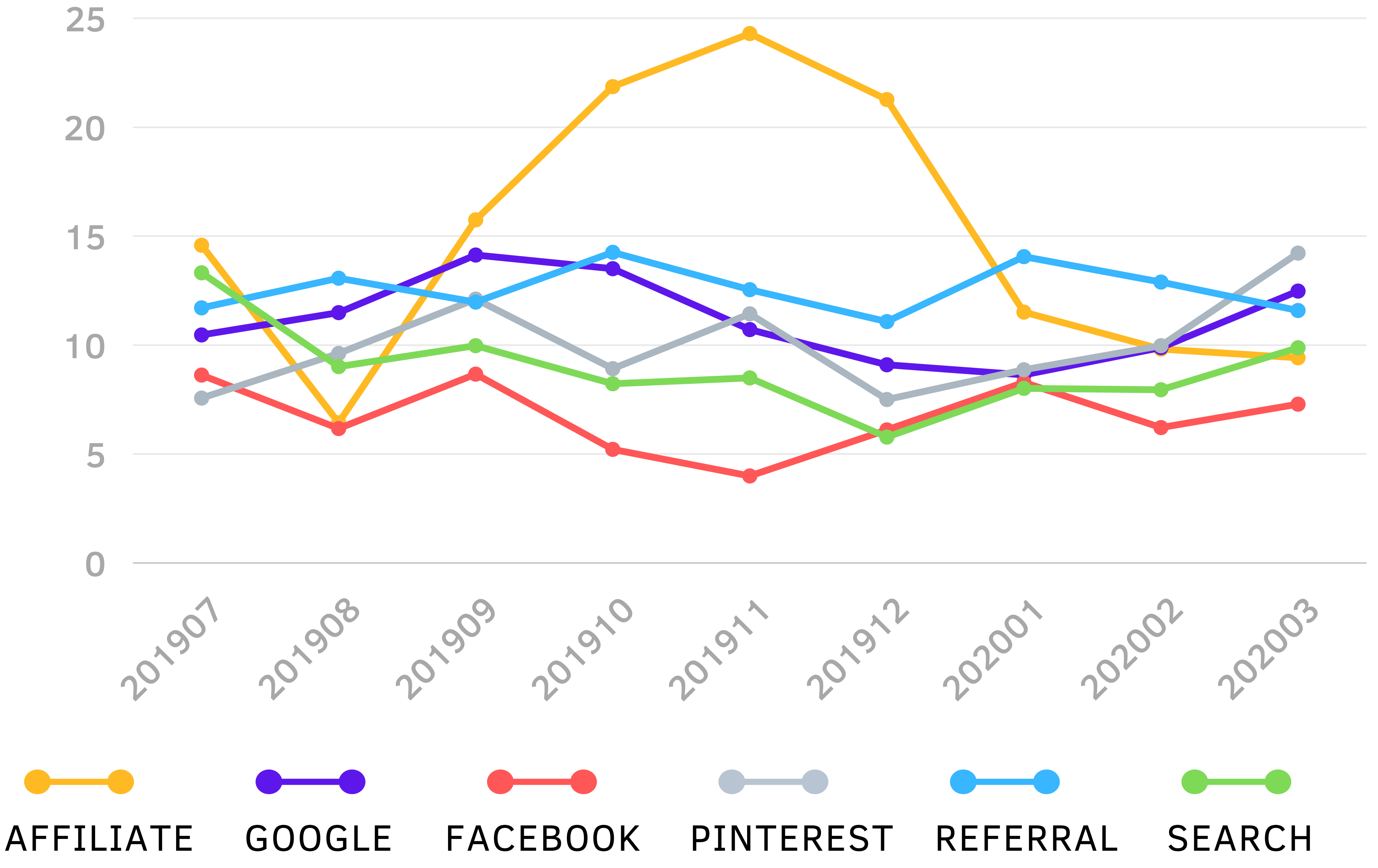
Allocation

Calculating advertising channel spend efficiency and effectiveness
supporting the advertising team's budget allocation for the upcoming quarter

Allocation - Technical & Create Account

Technical Attribution: entrance to sign-up form captured by product

MONTHLY AVERAGE CAC (LOWEST 5)



TOTAL AVERAGE CAC

Facebook	6.97
Search	9.34
Pinterest	10.58
Google	11.90
Affiliate	14.39
Referral	14.45
Email	30.09
Email Blast	31.70

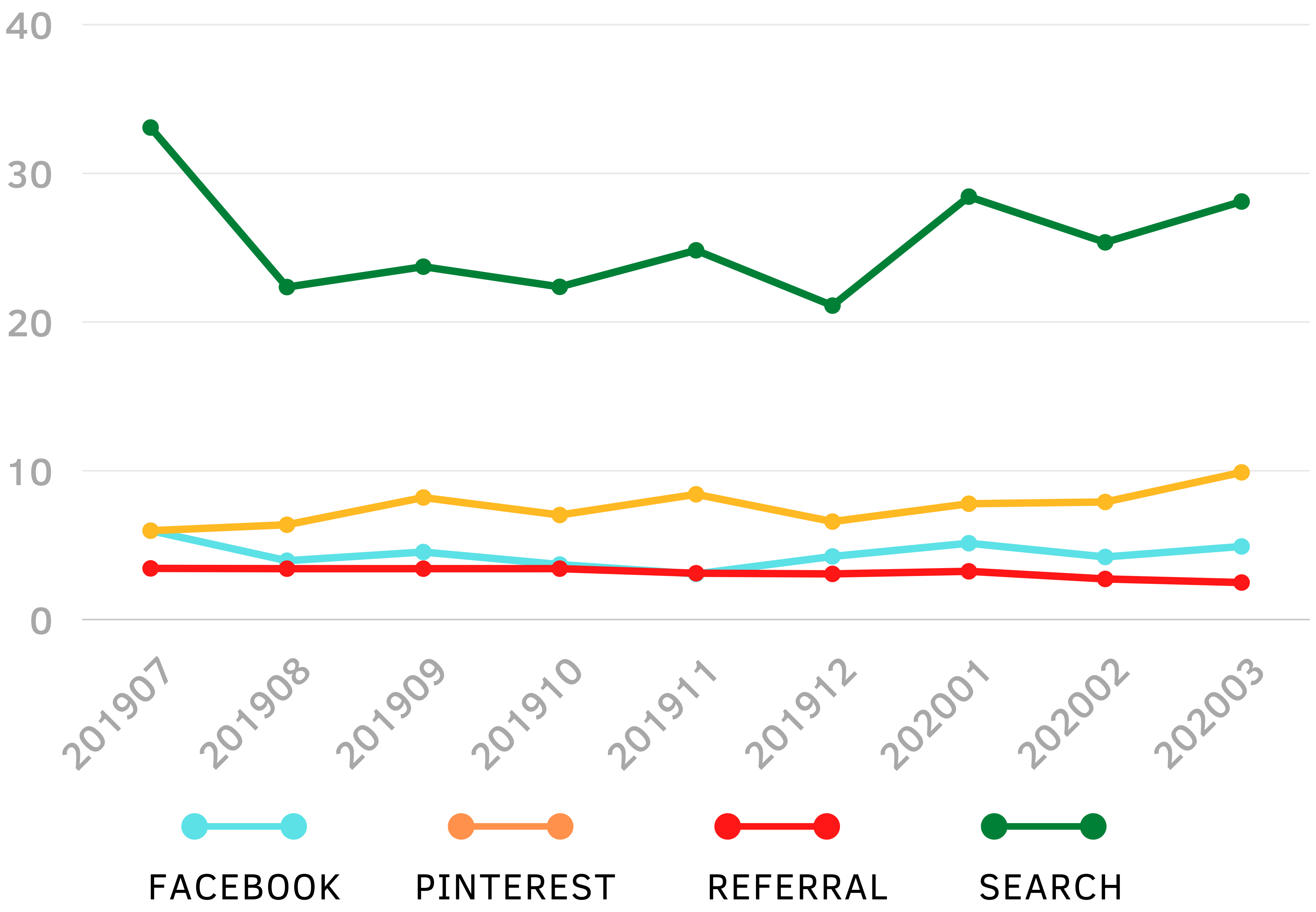
Insights from here:

- No obvious trend between months
- Most effective methods are Facebook, Search, and Pinterest

Allocation - Survey & Create Account

Survey Attribution: customer reported "how did you hear about us"

MONTHLY AVERAGE CAC (LOWEST 5)



TOTAL AVERAGE CAC

Referral	3.57
Facebook	4.70
Pinterest	8.14
Search	27.86
Affiliate	72.13

Insights from here:

- Survey Attribution may show the true reason of user come to this app (but may be biased)
- According to Survey Attribution, the most effective methods are Referral, Facebook, and Pinterest

Allocation - Comparison and Strategy

TOTAL AVERAGE CAC	Technical	Survey
Facebook	6.97	4.70
Search	9.34	27.86
Pinterest	10.58	8.14
Google	11.90	NaN
Affiliate	14.39	72.13
Referral	14.45	3.57
Email	30.09	NaN
Email Blast	31.70	NaN

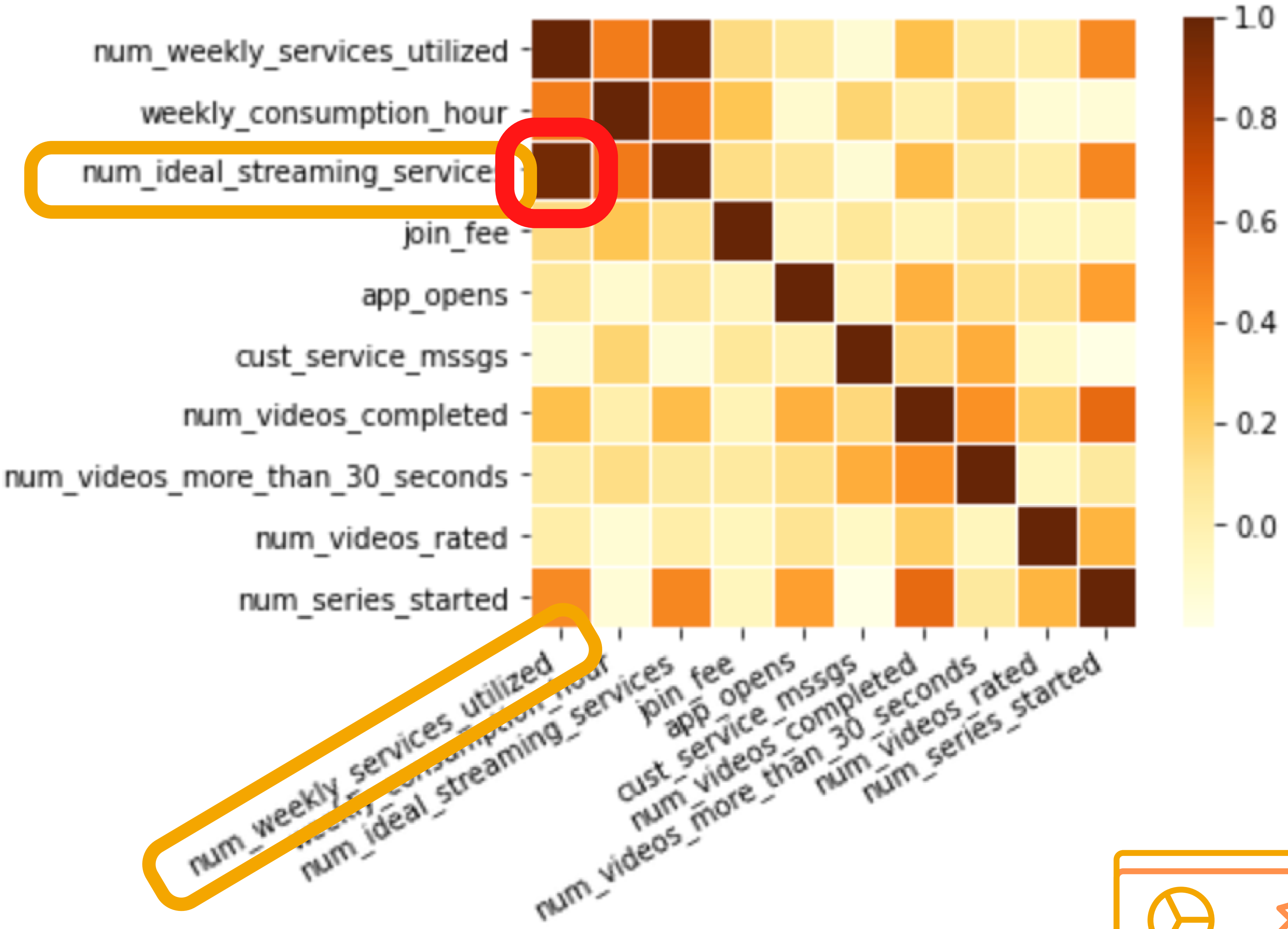
Insights from here:

- The advertising power of Facebook, Pinterest, and Referral might be underestimated
- And the power of Search and Affiliate might be overestimated
- Thus, to avoid unnecessary risk from Search, the major advertising methods should be Facebook and Pinterest

Churn Model

Predicting churning probability and develop recommendation(s) on an alternative product structure of pricing and others

Churn Model



- **Goal**: Predict the churning probability of a user, and apply strategies to prevent churning
- **Data Size**: 92920
- **Models**:
  - Multilinear probability model with backward stepwise method
  - Logistic model
- **Variables**:

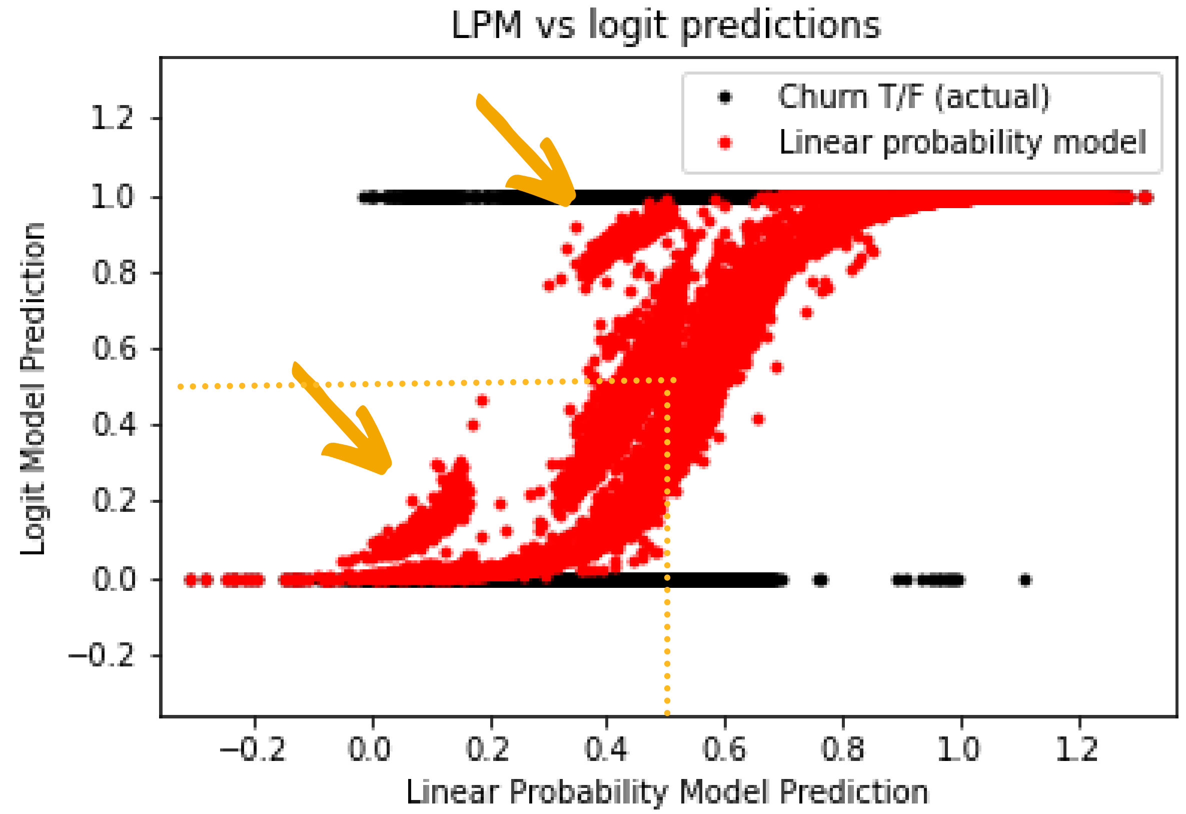
  Y - current substitution situation

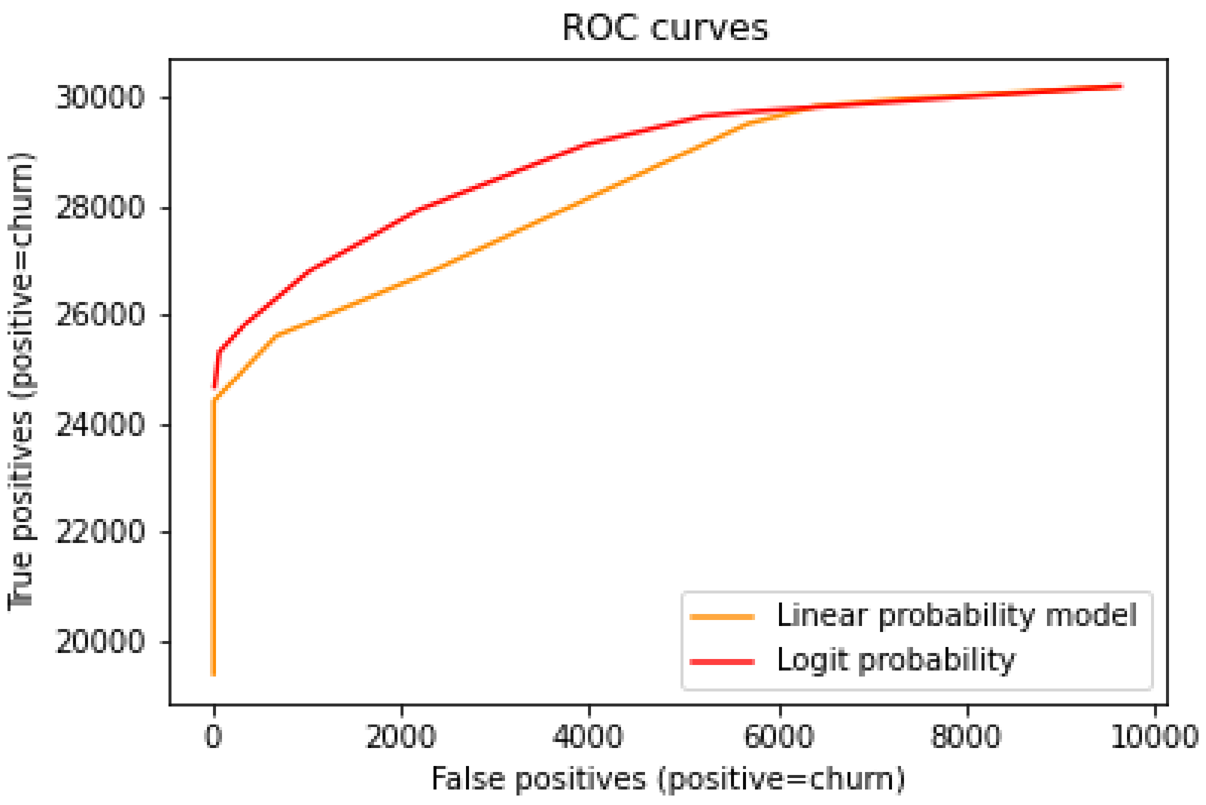
  X - User portraits & Activity of user in the app

  internet package type, weekly services utilized, preferred genre, intended use, weekly consumption hour, retarget situation, gender, operation system, join fee, paid TF, age range, payment period, app opens

Churn - Linear & Logistic Model

- **Linear probability model:**
  - using backward stepwise to reduce variable
  - R-Squared : 0.52
  - Worse ROC performance
- **Logistic model:**
  - R-Squared : 0.60
  - Better ROC performance
- **Conclusion:**
  - In general, the two models have similar results
  - Logistic model have slightly better R-Squared and ROC performance
  - Focus on Logistic model





Churn - Insights from Models

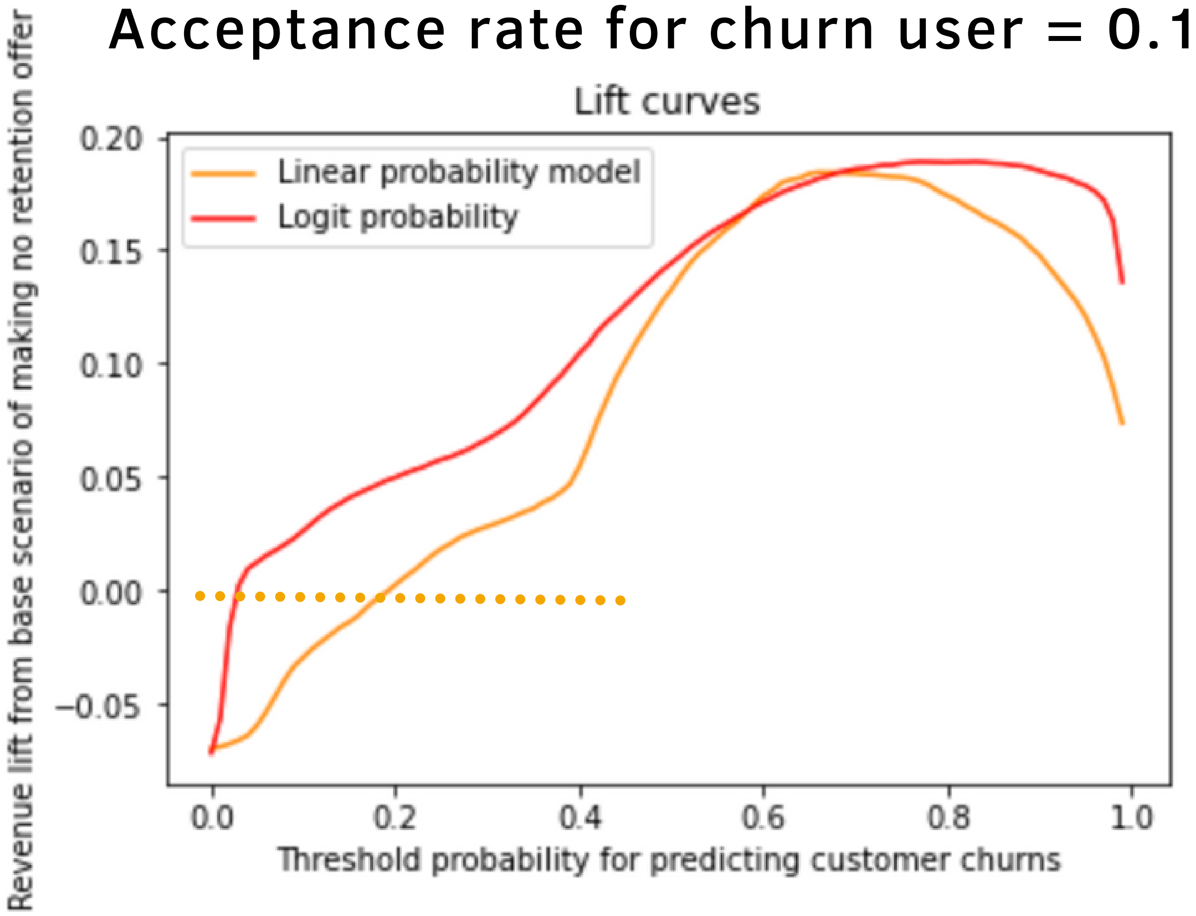
- **These users intend to stay**
  - Users with higher join fee
  - Users with more payment period
  - Users complete more videos
  - Users rate more videos
  - Users start more series
  - Male Users

- **Possible Strategies**
  - Understand 'join fee' (eg. the reason of different join fee, the circumstances of paying high join fee) and targeting the group of user with high 'join fee'
  - Pay more attention on males while doing acquisition

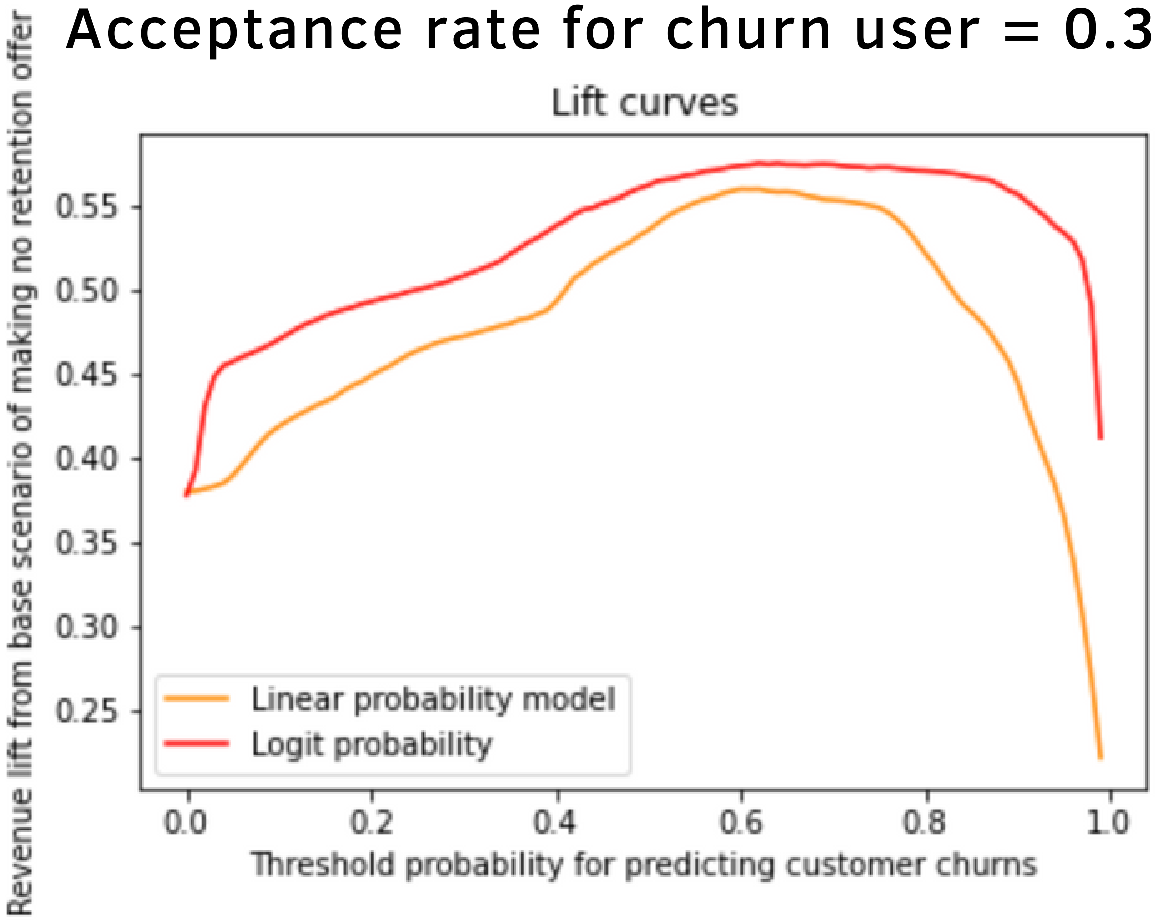
Discount & Prediction Threshold

- Offering a discount to identified potential churners will improve our revenue metric in both models with the model parameters even when only 10% of churn user will accept the offer
- According to the left graph, for the logistic model with 10% of churn users accept the offer, the revenue lift from base scenario of making no retention offer will become positive when the threshold probability for predicting customer churns is greater than 20% and reach the peak when the threshold is around 65%, and the revenue lift will be approx. 17%

Acceptance rate for churn user = 0.1



Acceptance rate for churn user = 0.3



Model Parameters

All non churn users accept offer

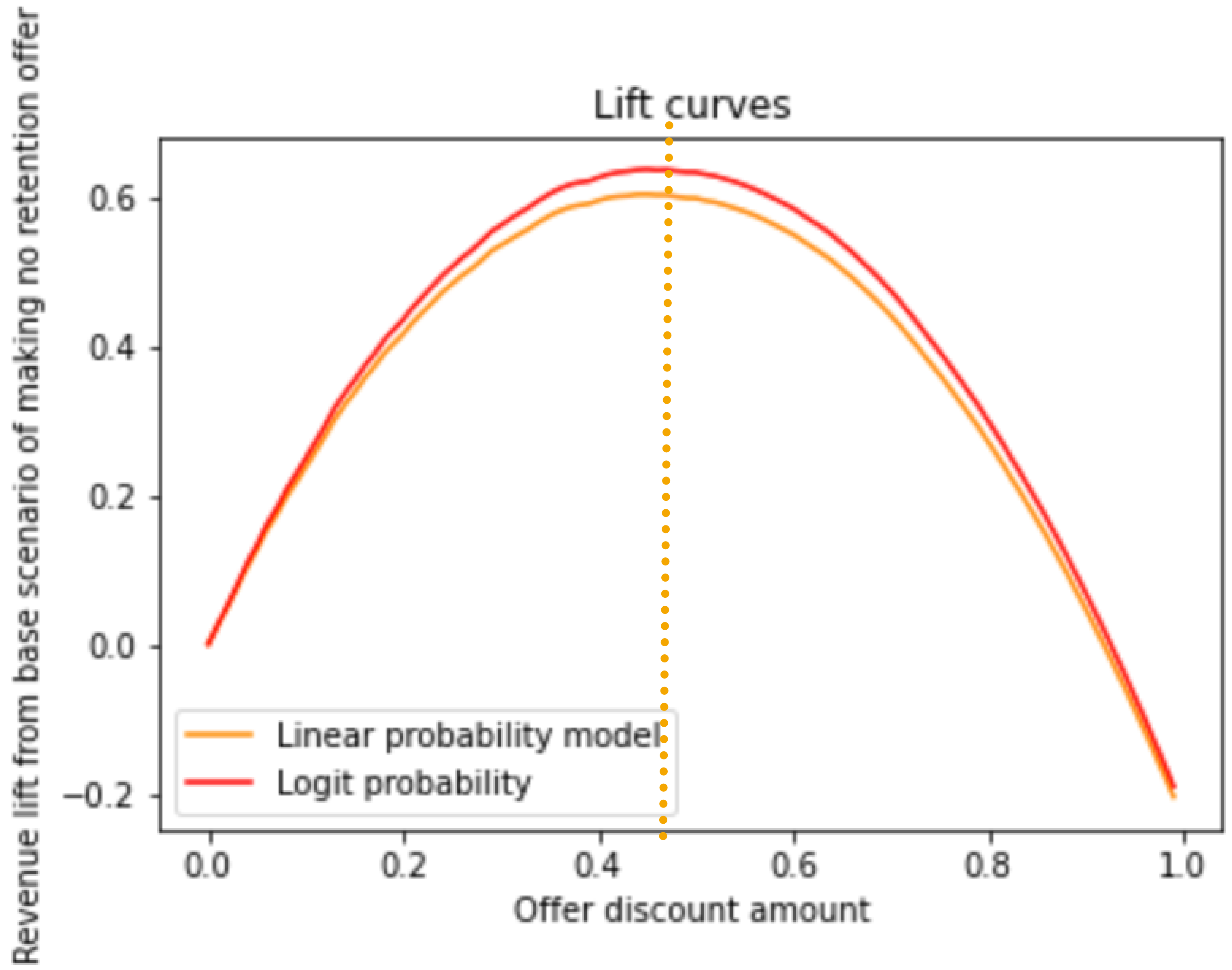
Base price of service = 4.73

Discount rate = 30% off

Optimal Discount

Assumption: churning user's acceptance rate has 1:1 positive linear relationship with discount rate.

- Based on this assumption, the peak of revenue lift occurs when the offer discount rate is around 48%, which gives a 60% lift on revenue.
- This conclusion may be more reliable if reasonable assumptions based on the real situation should be made in order to get more accurate results. To achieve more reasonable assumptions, find the relationship between discount rate and accept rate through A/B test.



Model Parameters

Acceptance rate for non churn user = 1

Base price of service = 4.73

Threshold= 0.6

Marketing Strategy

Targeting Strategy

- Designing different recommendation from the beginning base on gender if no other information are available
- Applying more resource on Comedy and drama since most users are interested in these two genre

Acquisition Strategy

- Spending budget on Facebook and Pinterest to effectively acquire new users
- Paying more attention on males while doing acquisition

Retention Strategy

- Giving potential churners discount rates and using A/B test to find the relationship between acceptance rate and discount rates to get optimal results
- Trying to give more renew choice to potential churners and test the result (e.g. sign 1-month contract instead of 4-month)
- Finding the churners

Thank You