Who to Target?

bumble | MIT | MIT MANAGEMENT BUSINESS ANALYTICS

Defining and Understanding Passive LTV in an Online Dating Network

Project Sponsors: Laura Mitchell, Luca Cerone, Nicola Ghio, William Fletcher • Faculty Advisor: Dr. Amy Farahat • MBAn Students: Amy Ho & Siqi Wu

Problem

Bumble operates on a freemium model and generates revenue through its premium subscription service and paid premium features. Across Bumble's active users, approximately 1.8 million are payers and the total value of these payments can be attributed to 'passive users' – users who are shown as options to swipe on during an app session.

Therefore, Bumble needs to understand the value that passive users bring to the platform. Our problem statement can be split into 2 parts:

- How should Bumble define an accurate measure of passive LTV?
- What factors impact passive LTV and how can these findings impact Bumble's future work?

Timeline

Feb | Mar | Apr | May | Jun | Jul | Aug

Admin: Data Access; Handover

Data Exploration: EDA

Defining PLTV: Initial Definition; Definition Iteration

Understanding PLTV: Baseline Clustering; Initial Regression; Final Modelling

Two Week Break

Data & EDA

The data we used can be divided into 3 parts:

User Data
Includes demographics, lifestyle habits and personal interests

Activities Data
Daily aggregation of user activities, event log of voting results

Revenue Data
Transactions on Bumble subscription packages and consumables

Criteria for Sample Scope:

- Registered between 2021.3.1 and 2021.4.30
- Profile location in Los Angeles
- Non spam/test users

Final Sample Size: **32.4k** users

Methodology

Defining Passive LTV

Step 1

We proposed the idea that **passive value** can be defined as the proportion of an active user's investment on the passive user:

Active User → Investment → Passive User

Step 2

After iterating through 4 **candidate definitions,** we chose the one that included regular Yes Votes, SuperSwipes and Beeline matches. The steps for calculating passive LTV are:

1. Identifying Feature Values as different users value features differently → **2. Allocating Payer LTV** based on influential ties between users → **3. Aggregating Passive LTV** by adding up total passive value received

Users **A** and **B** each paid **$20**

A: 20 Votes — $1/Vote

B: 40 Votes — $0.50/Vote

Payer LTV ($$$$)

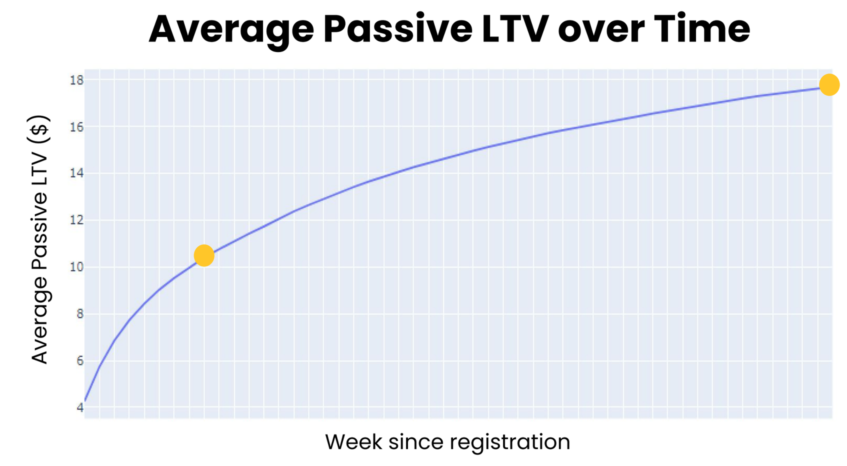
Passive LTV ($$$$$)

Step 3

We **validated our definition** through:

1. Stakeholder agreement
2. Correlation with internal popularity metrics
3. looking at average PLTV over time

Average Passive LTV over Time



Understanding Passive LTV

Model 1. Predicting future Passive LTV with early activity indicators

Features:
User demographics
Early activities indicators
Current passive LTV

Day 0 (Registration) — Day X (Now) — Day Y (Future)

Input — Target

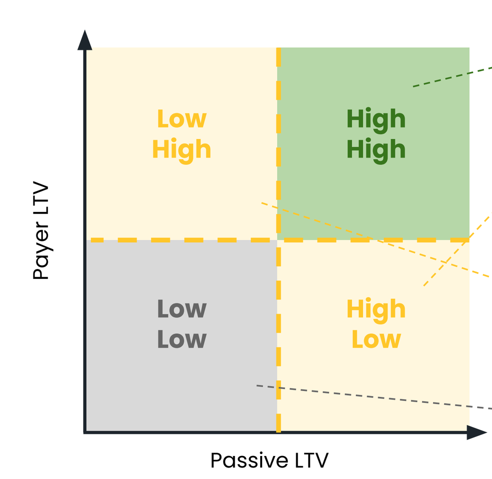
Results:
Across a user's lifetime, out-of-sample model performance is consistent

Time Window	R2	RMSE
Day 1 to Predict Day 7	0.70	7.40
Day 7 to Predict Day 28	0.84	9.43
Day 28 to Predict Day 56	0.93	7.13

Model 2. Identifying key drivers of the most valuable user group

Target Creation

Multiclass labels



Priority of Target

Best User Group
High influence & likely to pay

Rising Star
High influence & unlikely to pay

Loyal Member
Low influence & likely to pay

Limited Opportunity
Low influence & unlikely to pay

Predictive Modeling

Best Model: **XGBoost**

Metrics	Score
Precision	0.58
AUC	0.79

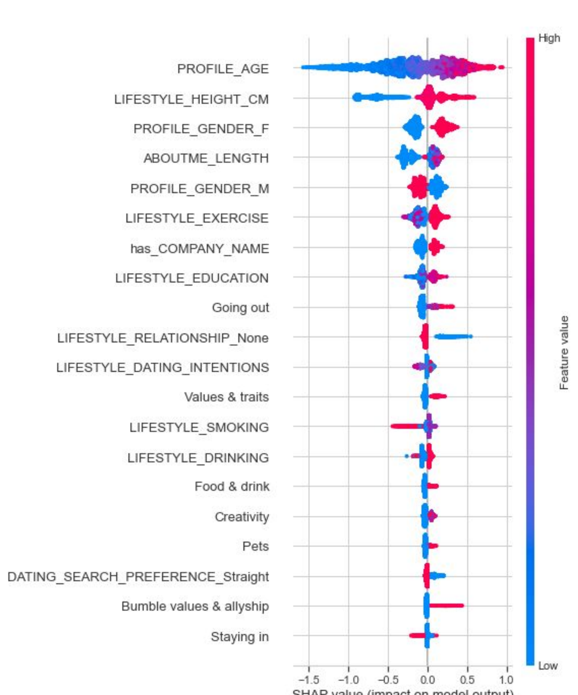
Model Explainability
Shapley Values

Results

We established a **company-wide definition** of Passive LTV for Bumble users, agreed by both the data science team and stakeholders

We provided the product revenue team with the ability to predict which users will have high potential PLTV for **more sophisticated product pricing strategies**.

We also confirmed that **Age, Height, Gender, Exercise, Company** are extremely important to a user being valuable to Bumble's platform for use in future marketing campaigns.



> 0.70
Out-of-sample R-squared in predicting passive LTV

58%
Of users correctly classified by their Payer and Passive LTV

2.3x
Increase in accuracy compared to baseline marketing strategy

Deliverables

We provided Bumble with the following deliverables:

Passive LTV Documentation
Full research process on Passive LTV definition

Modularized Code/Models
3 Python Scripts with a README file for reproducibility

Experimentation Recommendations
Potential A/B testing experiments for the future

Next Steps

Following our project, Bumble will continue to validate our Passive LTV definition and test if our findings are generalizable. This includes:

- Introducing more premium interactions into passive LTV definition
- Performing A/B testing on different users to track their responses to product pricing strategies
- Productionizing our pipeline with appropriate data ingestion