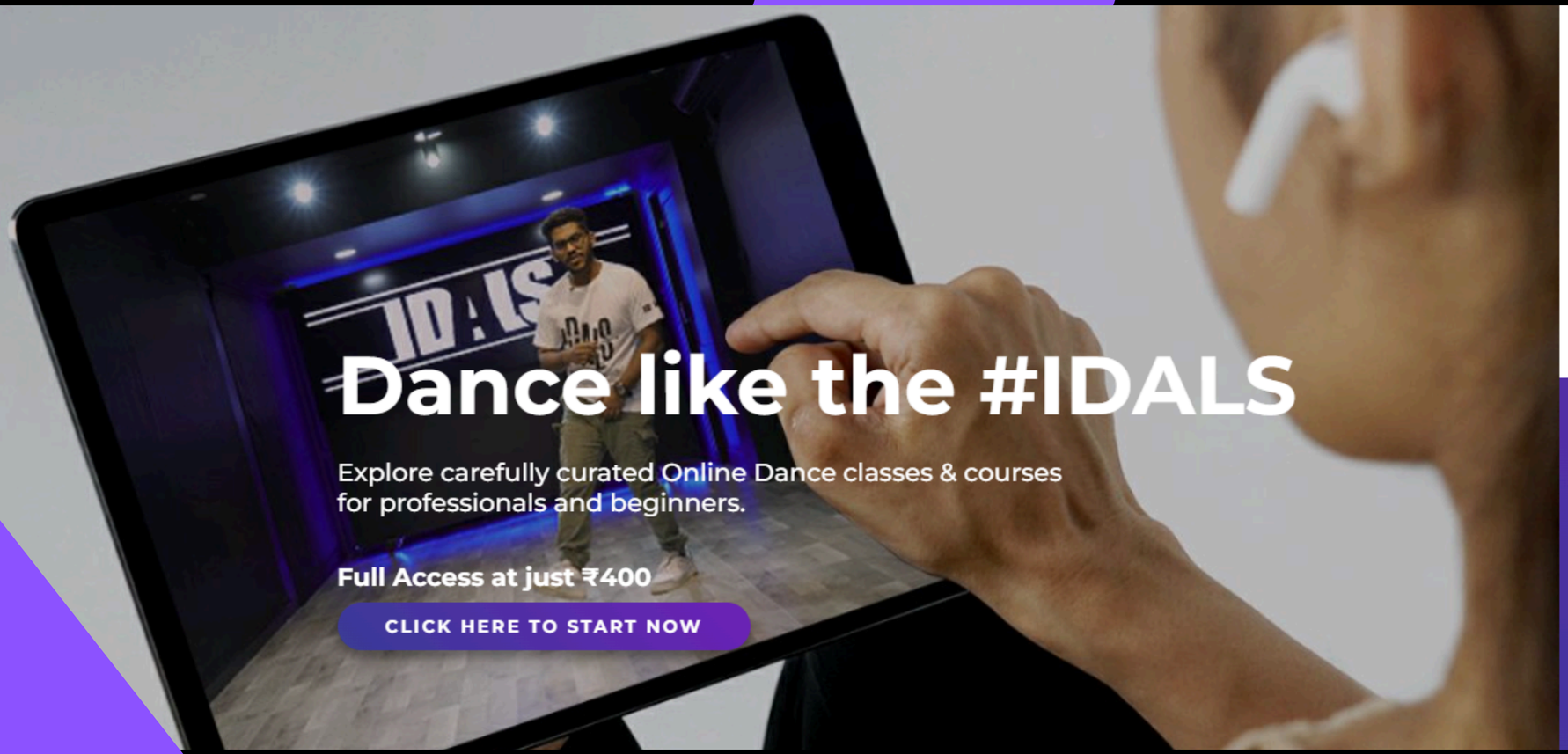
Data Analysis



Watch and learn

- Moves & Steps
- Level Wise Choreography Classes
- Simplified Dance Courses
- Dance Tips & Guidance
- Personality Development Sessions

And much more

Predicting Subscription Behavior on the IDALS EdTech Platform

Shijin Ramesh | Data Analyst | August 5, 2025

Business Problem

Key Question:

Can we predict a user's subscription plan based on profile, dancer type, goals and interests?

Why it matters:

- *Boost conversions from free to paid plans*
- *Personalize offers*
- *Improve retention strategy*

About IDALS Platform

- **Short intro:** Dance-based ed-tech startup
- **Subscription tiers:** Free, Weekly, Monthly, Half-Yearly, Yearly
- **Target audience:** Aspiring dancers, learners during/post lockdown

Data Overview

- **Total records:** 14,842 users
- **Key features:**

  - subscription_plan (Target)

  - type_of_dancer, genres_of_interest

  - purpose_of_learning, future_aspirations

  - repeated_subscriber, signup_month, signup_hour

Data Cleaning & Preparation

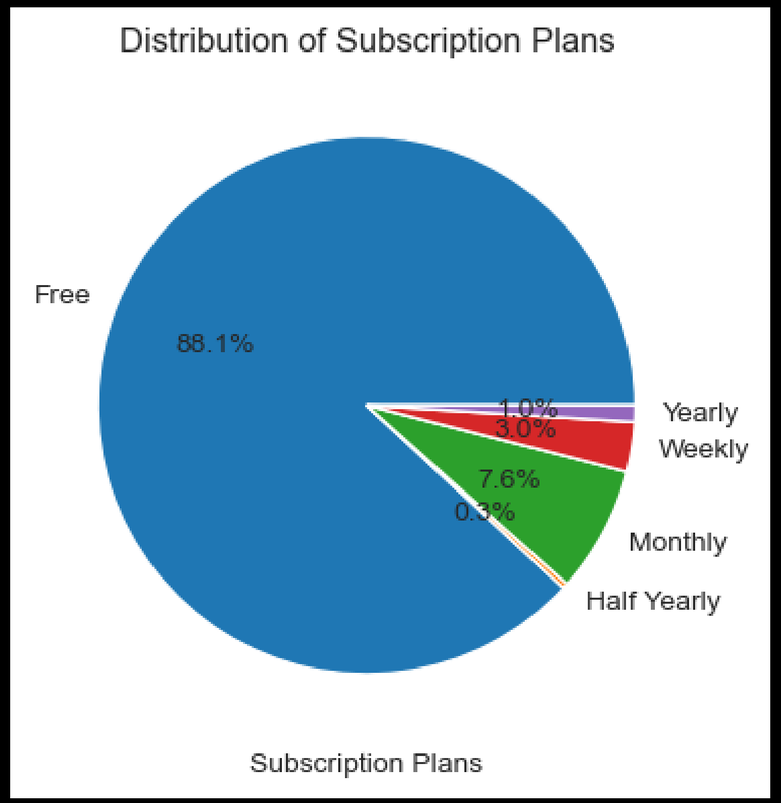
- Cleaned gibberish names, fixed country inconsistencies (e.g., "indai" → "India")
- Merged name fields
- Added columns: *subscriptions, repeated_subscriber*
- *Transformed subscription labels for clarity*

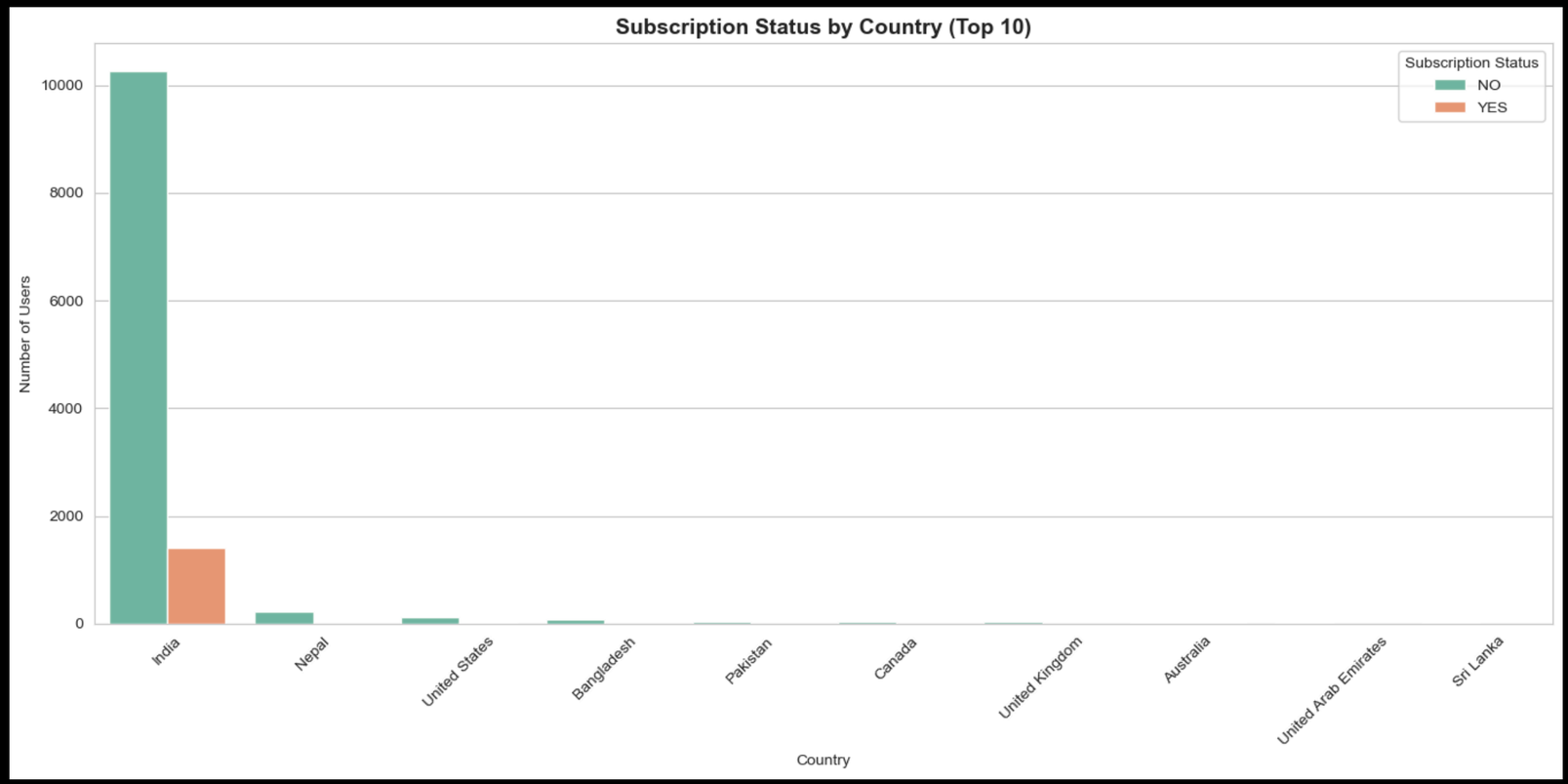
SUBSCRIPTION_PLAN	PLAN_EXPIRES_AT	LOCATION	COUNTRY	CREATED_AT	TYPE_OF_DANCER	GENRES_OF_INTEREST	PURPOSE_OF_LEARNING	FUTURE_ASPIRATIONS	FULL_NAME	SUBSCRIBED	SUBSCRIPTIONS	REPEATED_SUBSCRIBER
Free	null	Bharuch, gujarat	India	1/1/2021 10:11	Advance	Multi Genre	Career	Freelancer or Professional Dancer	Viraj Suthar	NO	null	No
Free	null	Thane	India	1/1/2021 10:20	Beginner	Multi Genre	Career	Freelancer or Professional Dancer	Prarthana Pawar	NO	null	No

Tools used: Excel, Snowflake SQL, Python (Pandas)

Key EDA Insights

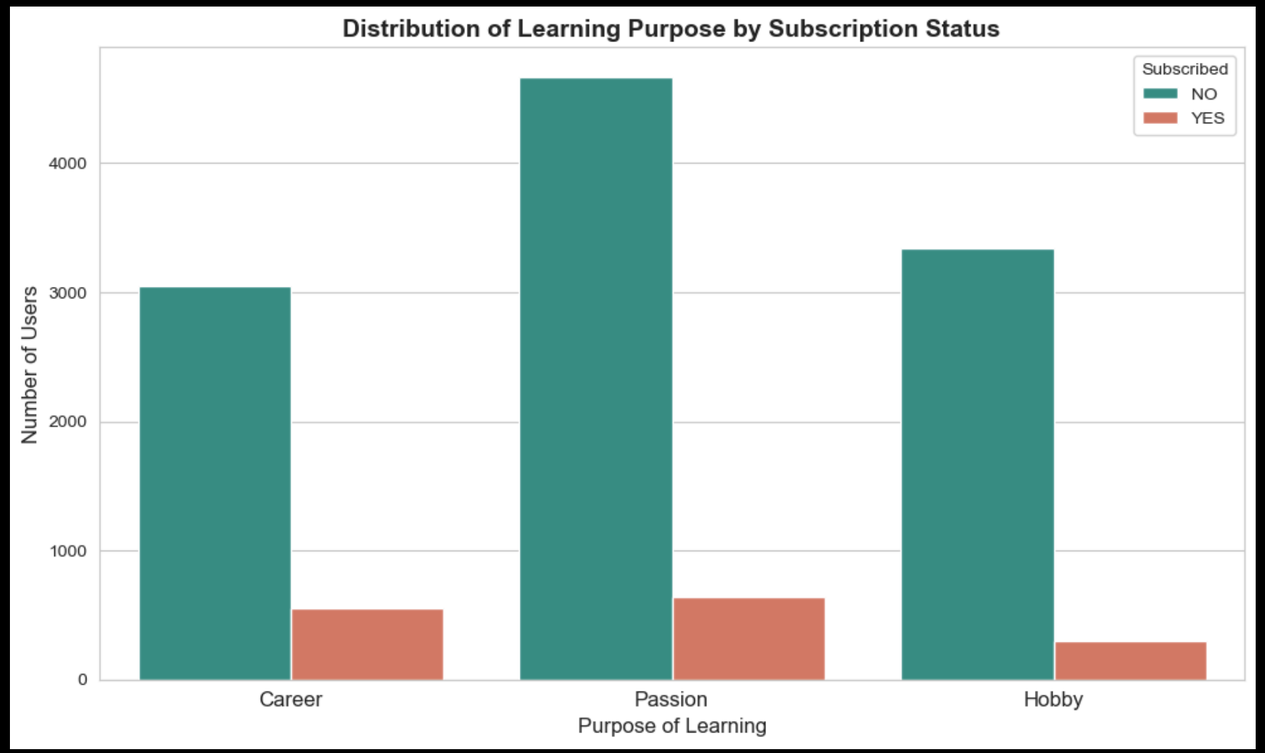
- ~88% users are Free plan users
- Most users from India (90%+)

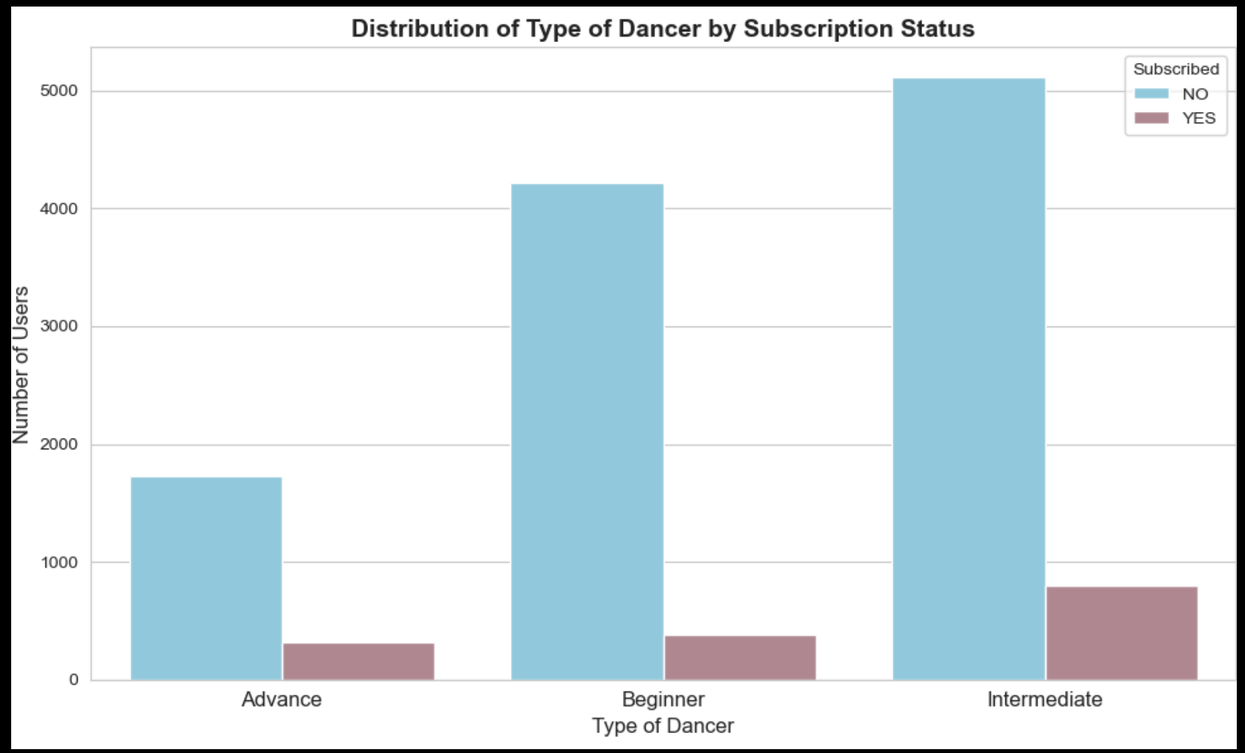




Key EDA Insights

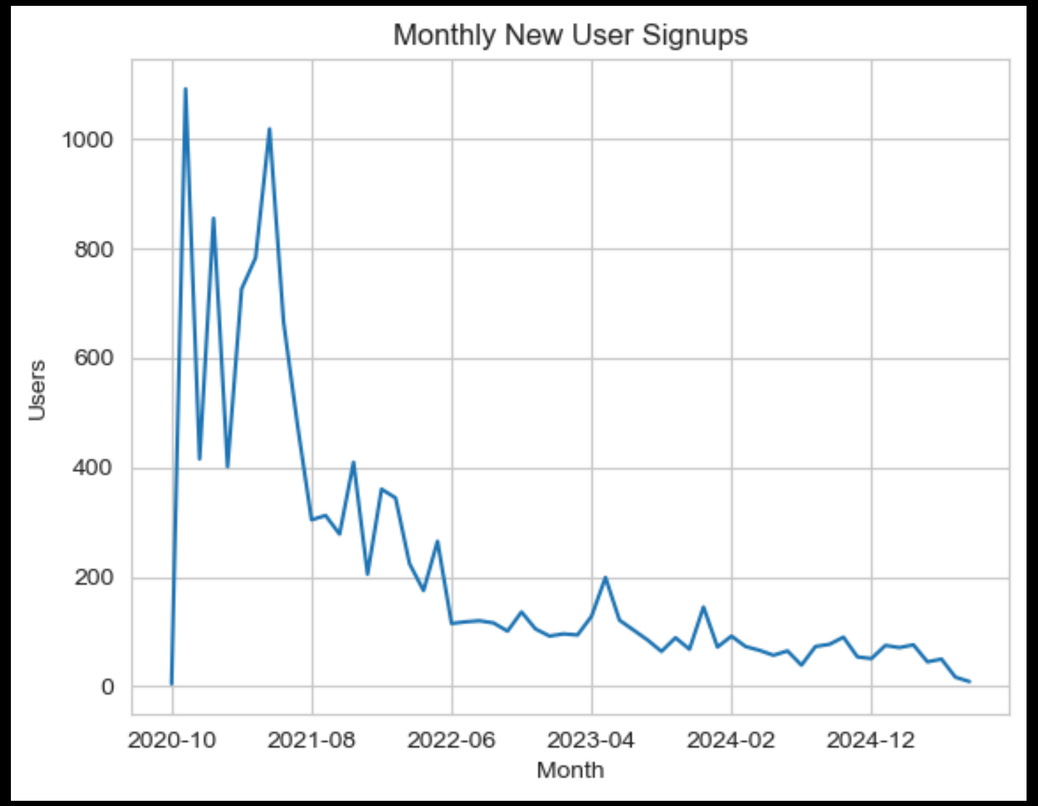
- Intermediate-level dancers and career & based users more likely to subscribe

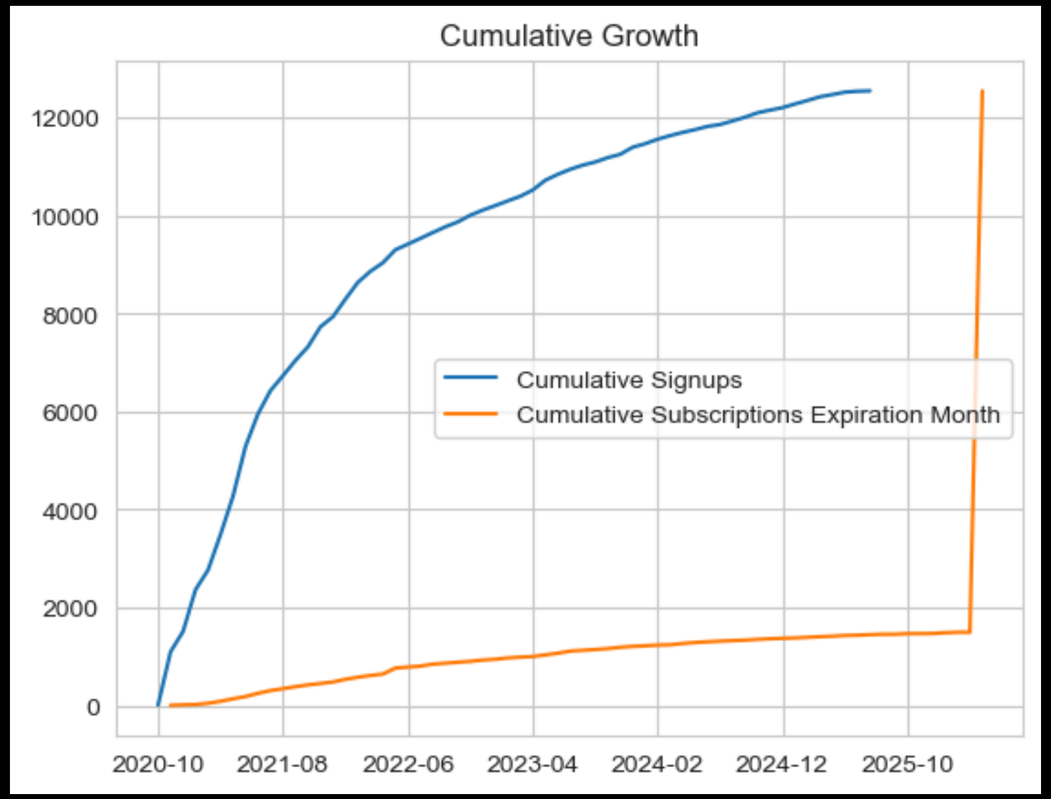




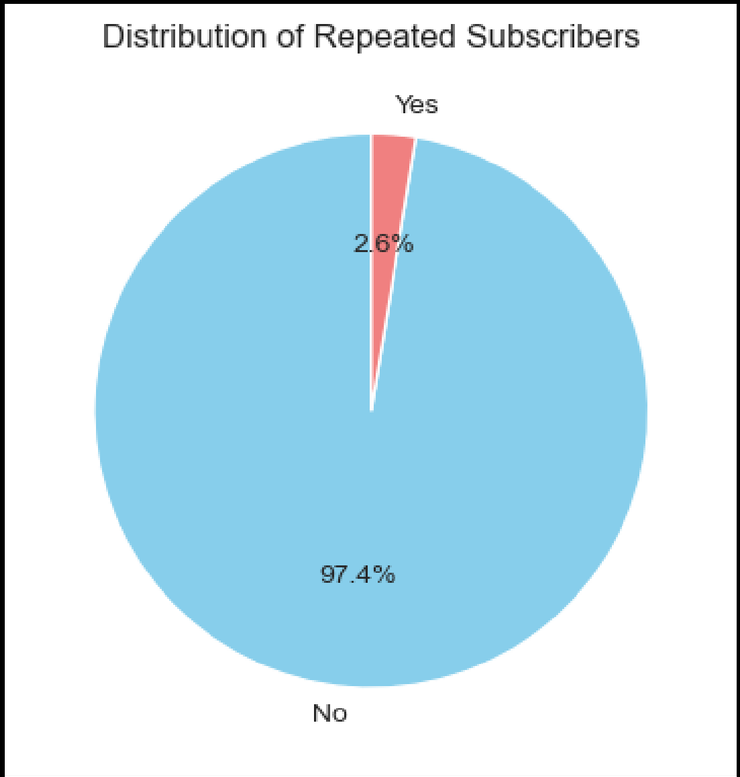
Key EDA Insights

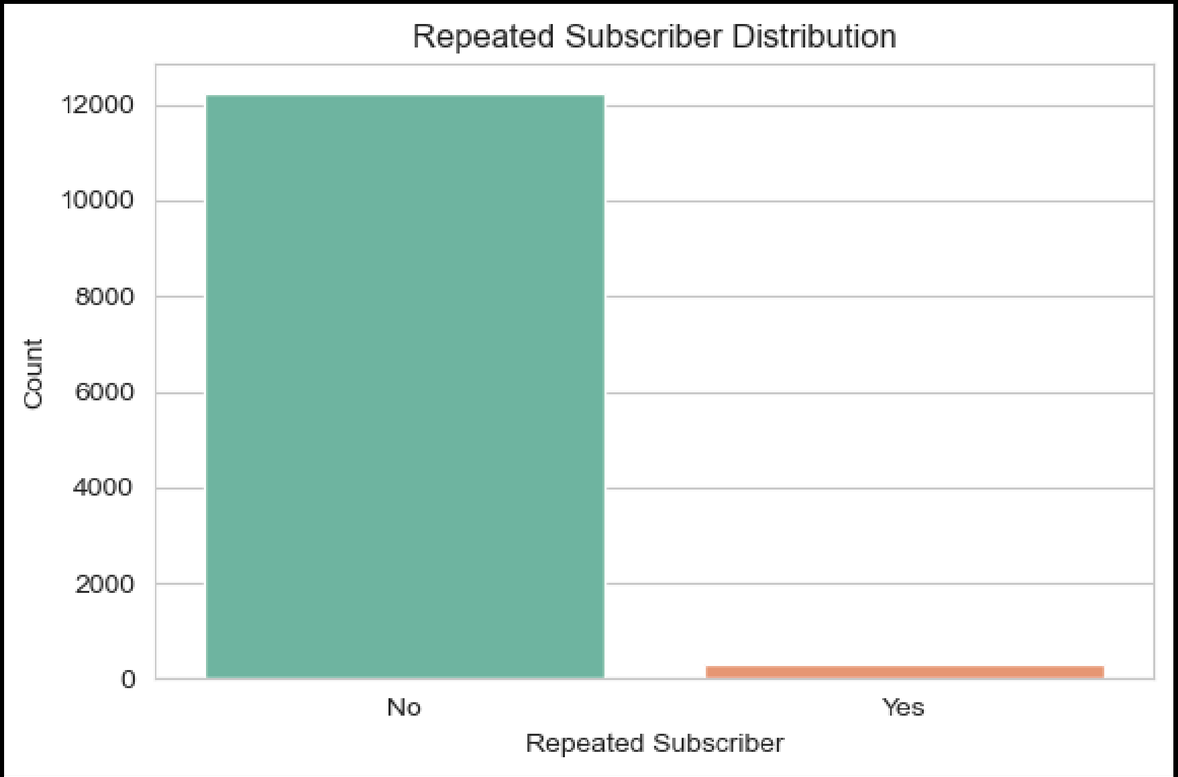
- Peak signups: Nov–Dec 2020 (during lockdown)
- Steady decline after mid-2021 due to offline classes
- Subscription expiries spike in Nov 2025 (anniversary offer)





Key EDA Insights





Modeling Approaches

- Tried multiple models:

  Logistic Regression (baseline)

  Random Forest

  XGBoost Classifier

- Handled class imbalance with class_weight='balanced' and careful feature engineering

Model Performance Overview

Model	Accuracy	Precision (Free)	Recall (Free)	Minority Class F1
Logistic Regression	46%	High	High	Near Zero
Random Forest	88.5%	High	High	Very Low
XGBoost	89.3%	0.9	1.0	Poor for other classes

All models biased toward predicting "Free" class due to severe class imbalance.

Feature Importance (XGBoost)

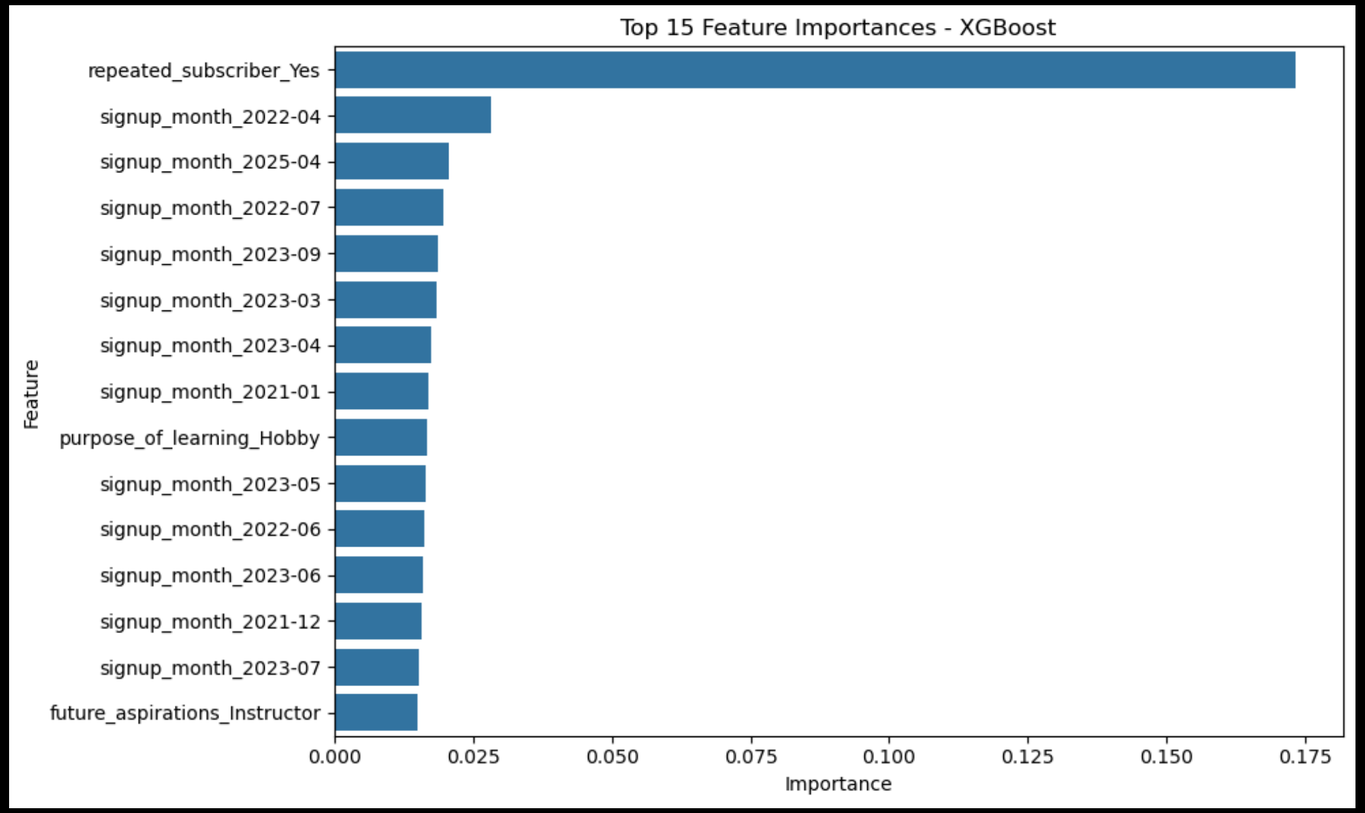
- Top contributing features:

  repeated_subscriber

  purpose_of_learning

  signup_month

- *Still not enough to strongly differentiate between plans*



Conclusions

- High accuracy is misleading due to imbalance
- Existing user data doesn't contain strong predictors
- Advanced models can't fix weak signals in features
- Need to rethink data collection strategy

Recommendations to Stakeholder

A. Boost Conversions

- *Conversion of Free to Paid plan users*
- *Limited-time paid plan offers (anniversary, festive)*

B. Collect Better Features

- *User engagement: time spent, courses completed*
- *Behavioral data: login frequency, drop-offs*
- *Add pricing sensitivity, feedback scores*

C. Segment Intelligently

- *Group users by learning goal, dancer type, and engagement*
- *Target campaigns to each persona type*

D. Retraining Loop

- *Add monthly model retraining with new data*
- *Improve predictions through incremental learning*

Appendix

- *Code Link: https://github.com/shijin/IDALSEdTech_EDA_PredictiveModeling/*