Excelerate Churn Analysis Project Overview

Insights and strategies for reducing churn

PRESENTED BY: TEAM D

Team members

- ANSH BUGRA
- Faiza Bashir
- Faridat Adewole
- Fouzia Ashfaq
- Fridaos Akorede
- Anubhav Rohilla

EXECUTIVE SUMMARY

The Challenge:

A significant portion of students (91.7%) are classified as churned, posing a risk to revenue and community health.

The Opportunity

Implementing a data-driven retention strategy can save an estimated $34,600 annually from just a 10% reduction in churn and strengthen the learning community

Key Drivers of Churn

- Inactivity >126 days → 23.6x more likely to churn .
- Low Engagement Score (<1.2) → 18.2x more likely to churn .
- Each additional opportunity type reduces churn by 63% .

INTRODUCTION

- Excelerate is an experiential learning platform offering courses, internships, events, competitions, and engagements.
- High churn threatens platform sustainability and community cohesion.
- This project focuses on:
    - Exploratory Data Analysis (EDA)
    - Churn definition and identification
    - Predictive modeling
    - Strategic interventions

Three-Week Analysis Journey

Week 1
Data Cleaning & Feature Engineering

- Standardized 23+ columns
- Created derived features
- Applied normalization

Week 2
Exploratory Data Analysis

- Seasonal trend analysis
- Demographic insights
- Correlation mapping

Week 3
Churn Analysis & Modeling

- Predictive models
- Risk stratification
- Strategic recommendations

Dataset Overview

Source

Excelerate "SLU Opportunity-wise dataset" (2022-2024).

Size & Structure

The dataset contains 2,847 rows and 54 columns, which was cleaned and processed down to 23+ analysis-ready columns.

Data Structure

The data is structured in a tabular format, enabling easy manipulation and comprehensive examination of relationships among variables.

Key Columns

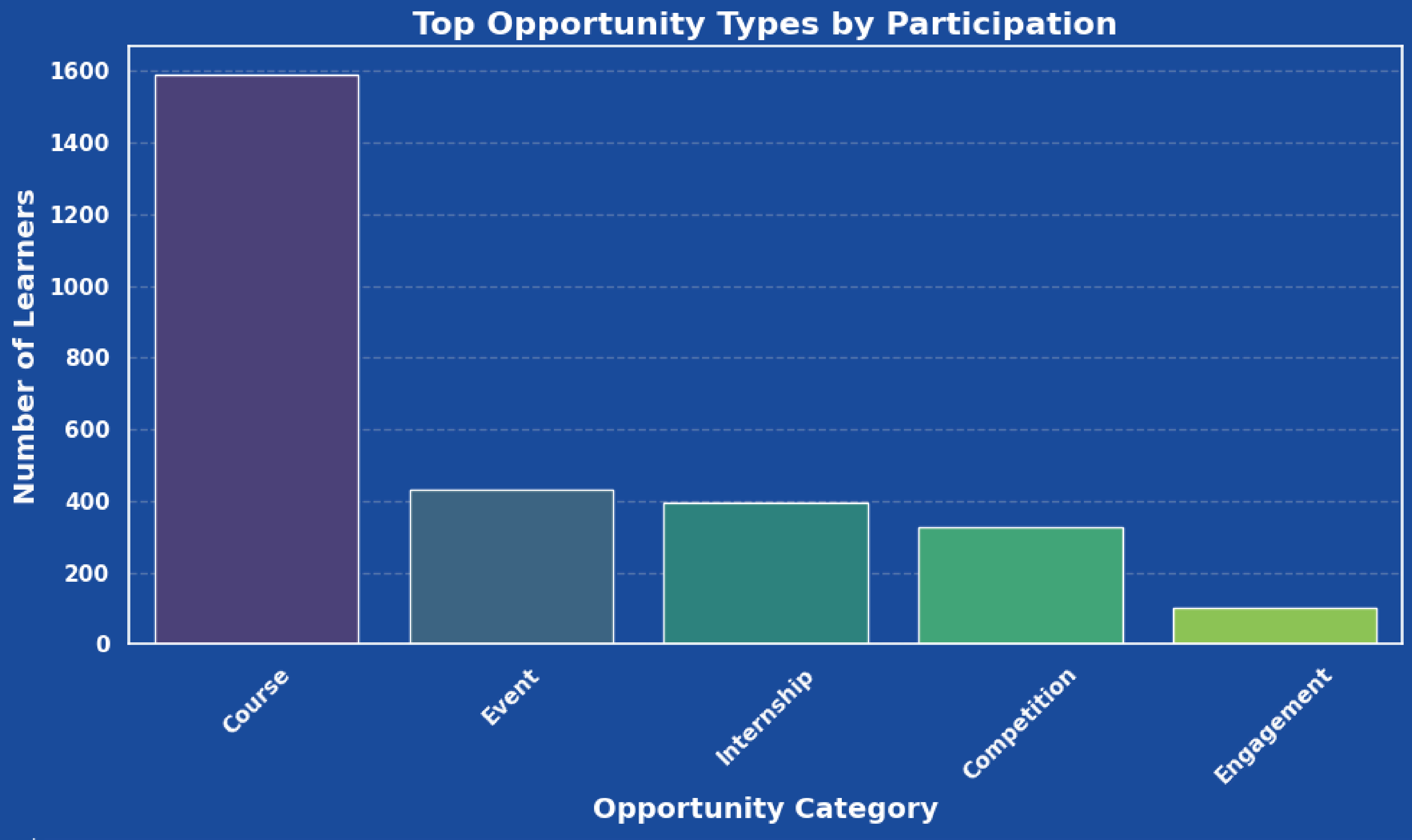
Learner: Age, gender, country, institution, major
Opportunity: Category, duration, start/end date
Engagement: Signup date, apply date, status, engagement duration
Derived: Age, normalized scores, combined score

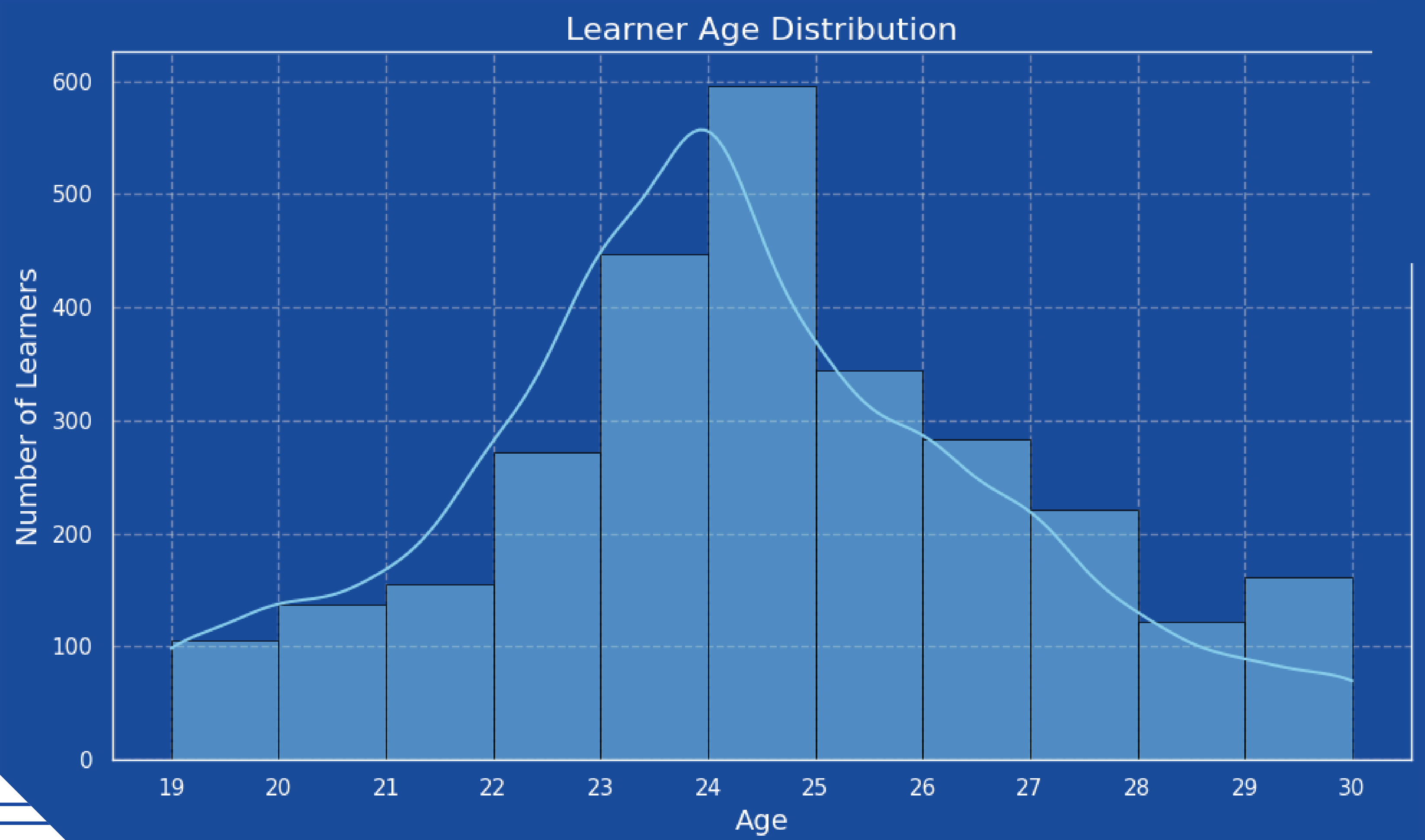
EDA Highlights

Visualizations and Insights

Top Opportunity Types

- Courses: Highest participation (~1,600 learners).
- Engagements: Lowest participation (~100 learners).



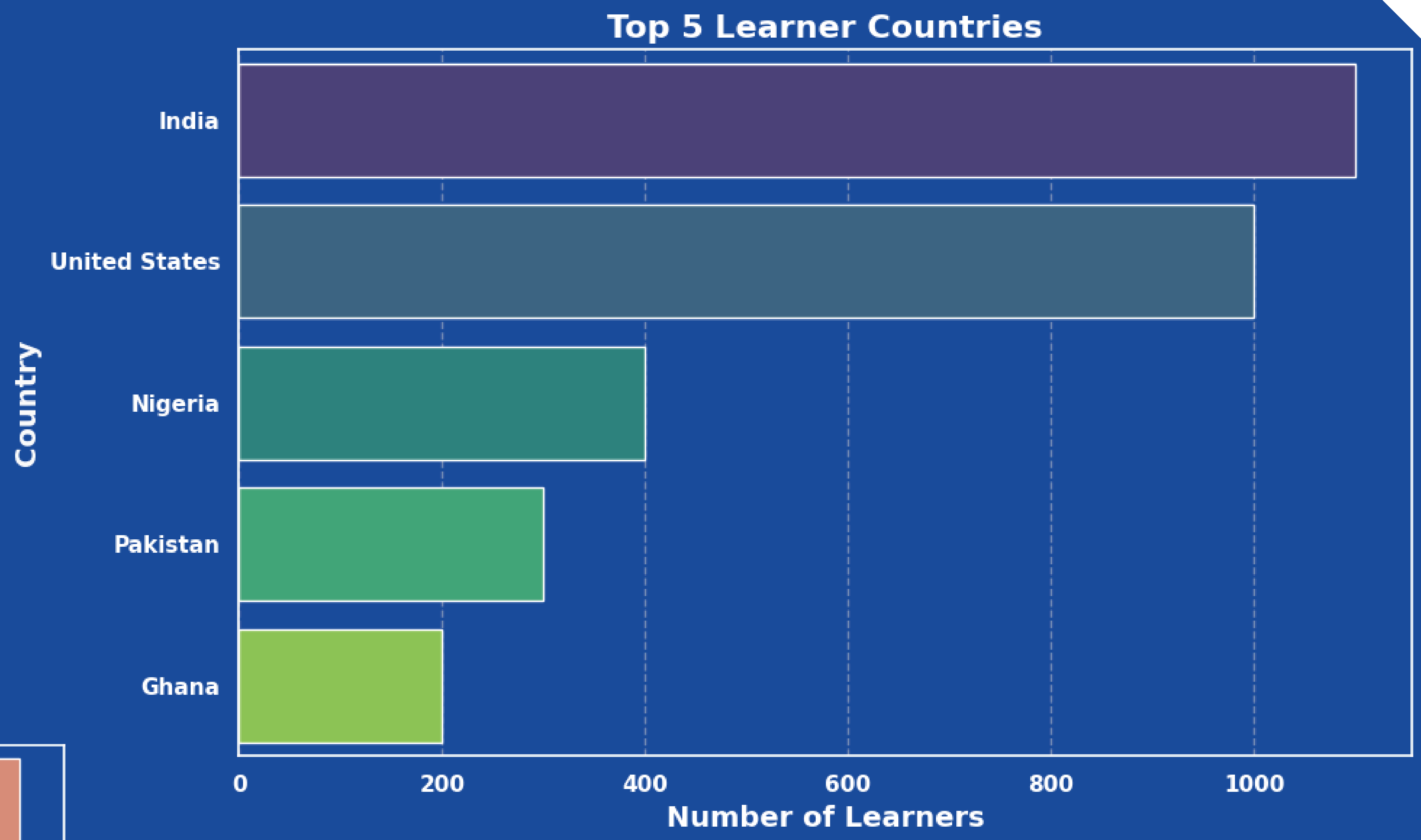


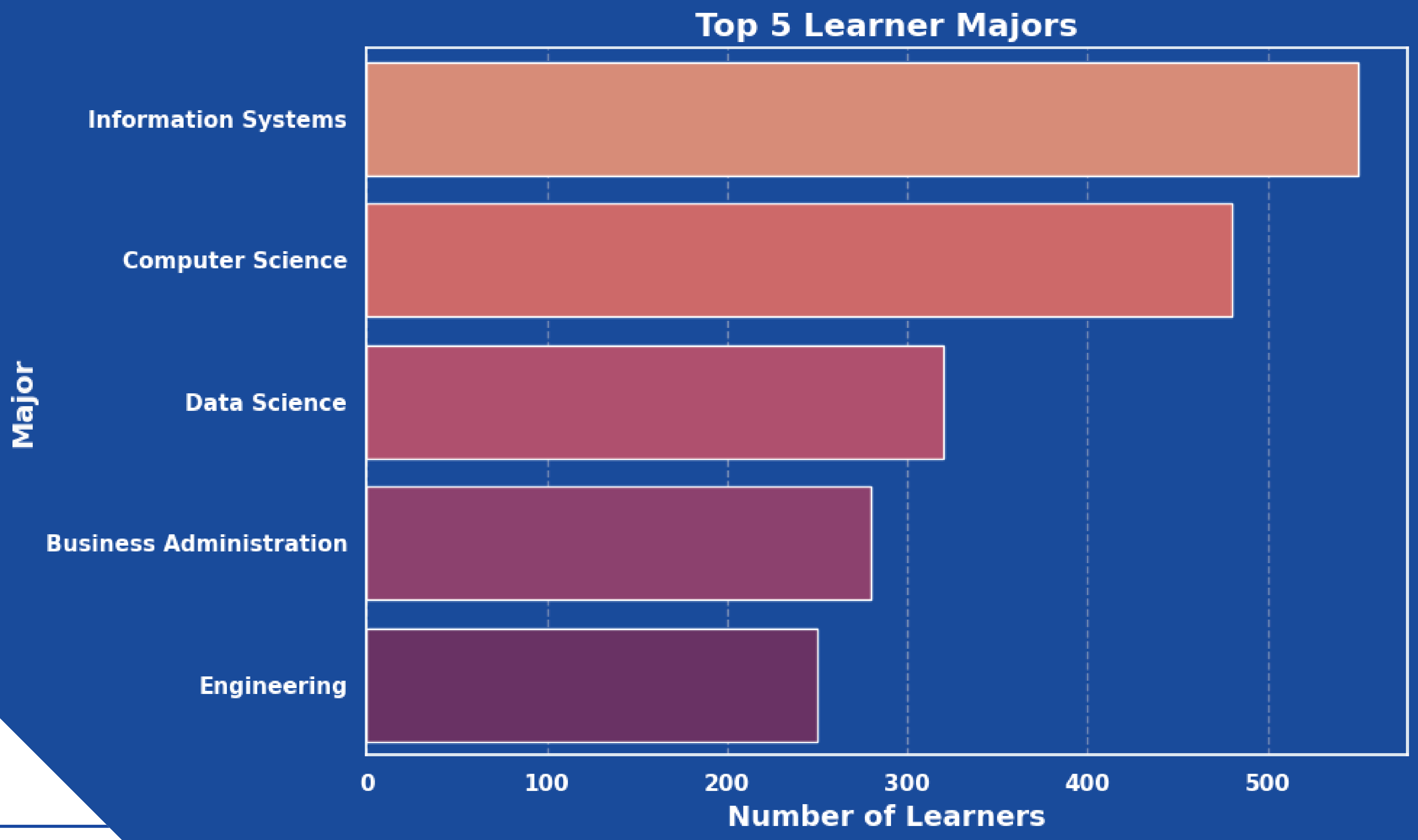
Learner Age distribution

- Peak at age 24
- Majority aged 21–25 (~1750 learners)

Where Are Our Learners From?

- India: ~1,100 learners.
- United States: ~1,000 learners.
- Nigeria: ~400 learners.
- Pakistan: ~300 learners.
- Ghana: ~200 learners.

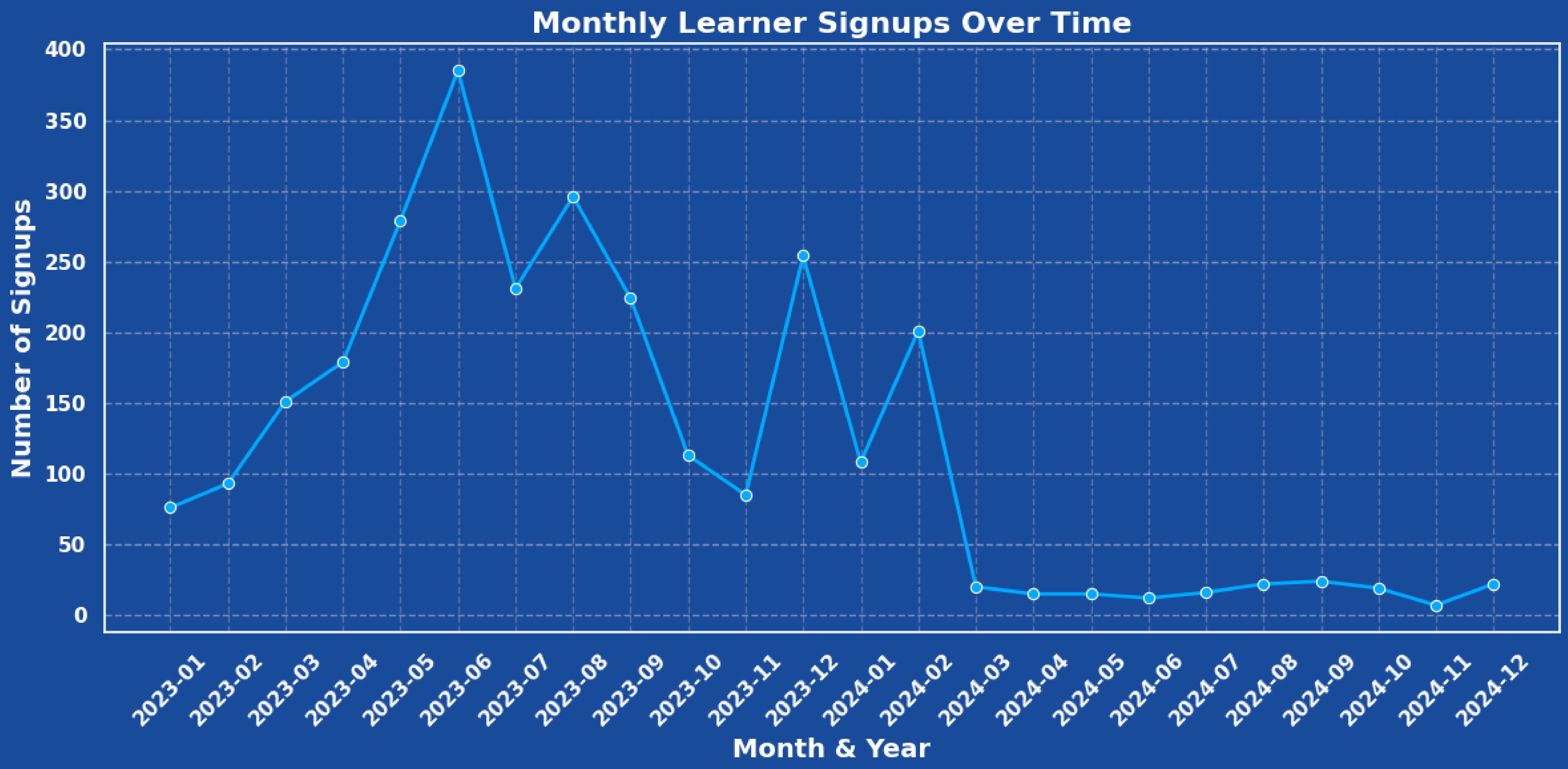




What Do Our Learners Study?

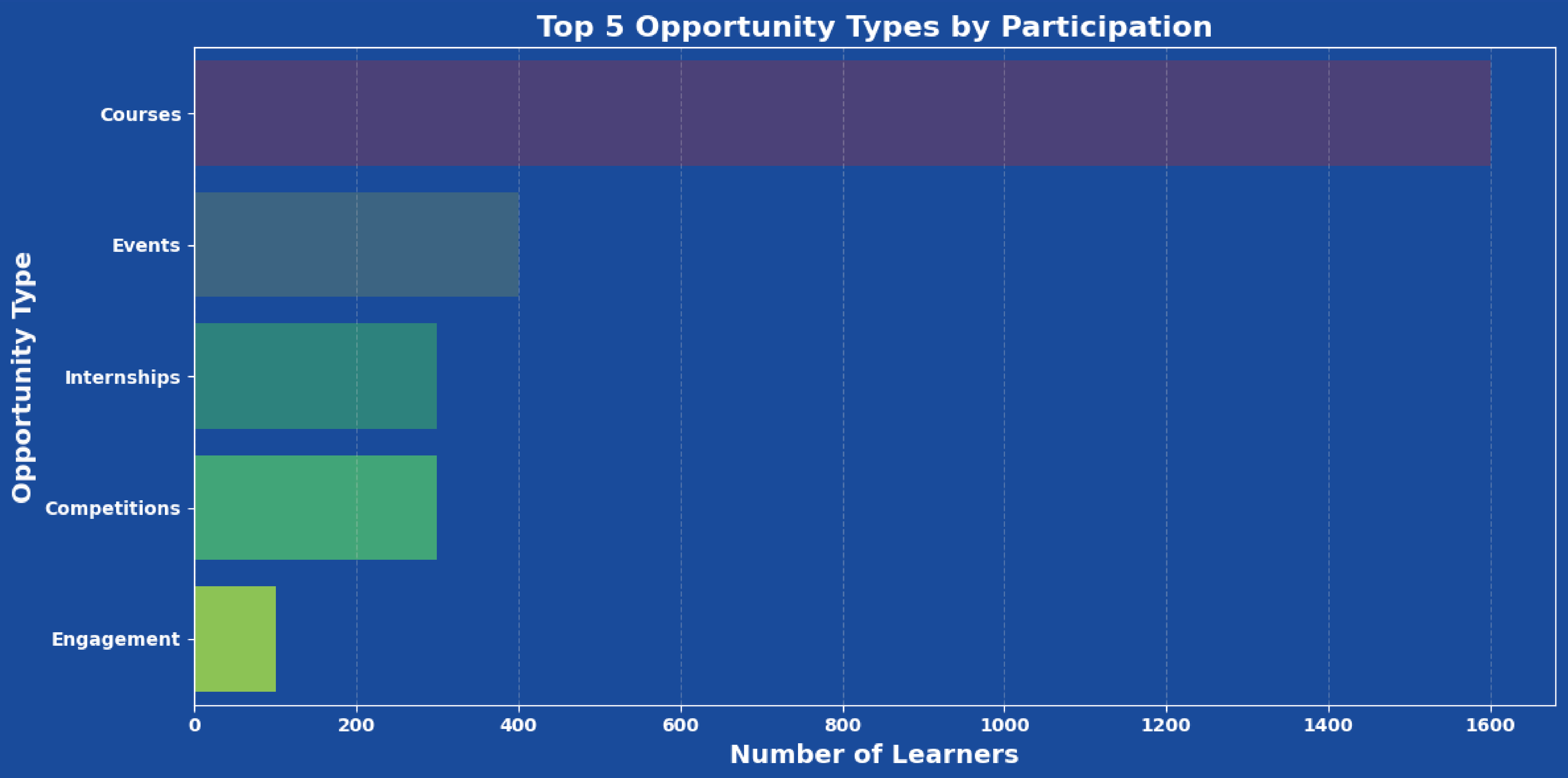
- Information Systems: Most common (~550 learners).
- Computer Science: Second most common (~480 learners).
- Other majors: Data Science, Business Administration, Engineering.

Monthly Signups Over Time



- **Peaks in March, June, September**
- **Sharp drop in early 2024**

Opportunity Engagement Trends



Top 5 Opportunity Types by Participation

- Courses (~1,600)
- Events (~400)
- Internships (~300)
- Competitions (~300)
- Engagement (~100)

Defining and Calculating Churn

- **How Churn Was Defined:**

A learner is flagged as "churned" based on a combination of behavioral indicators rather than a single event.

- **Churn Criteria:**

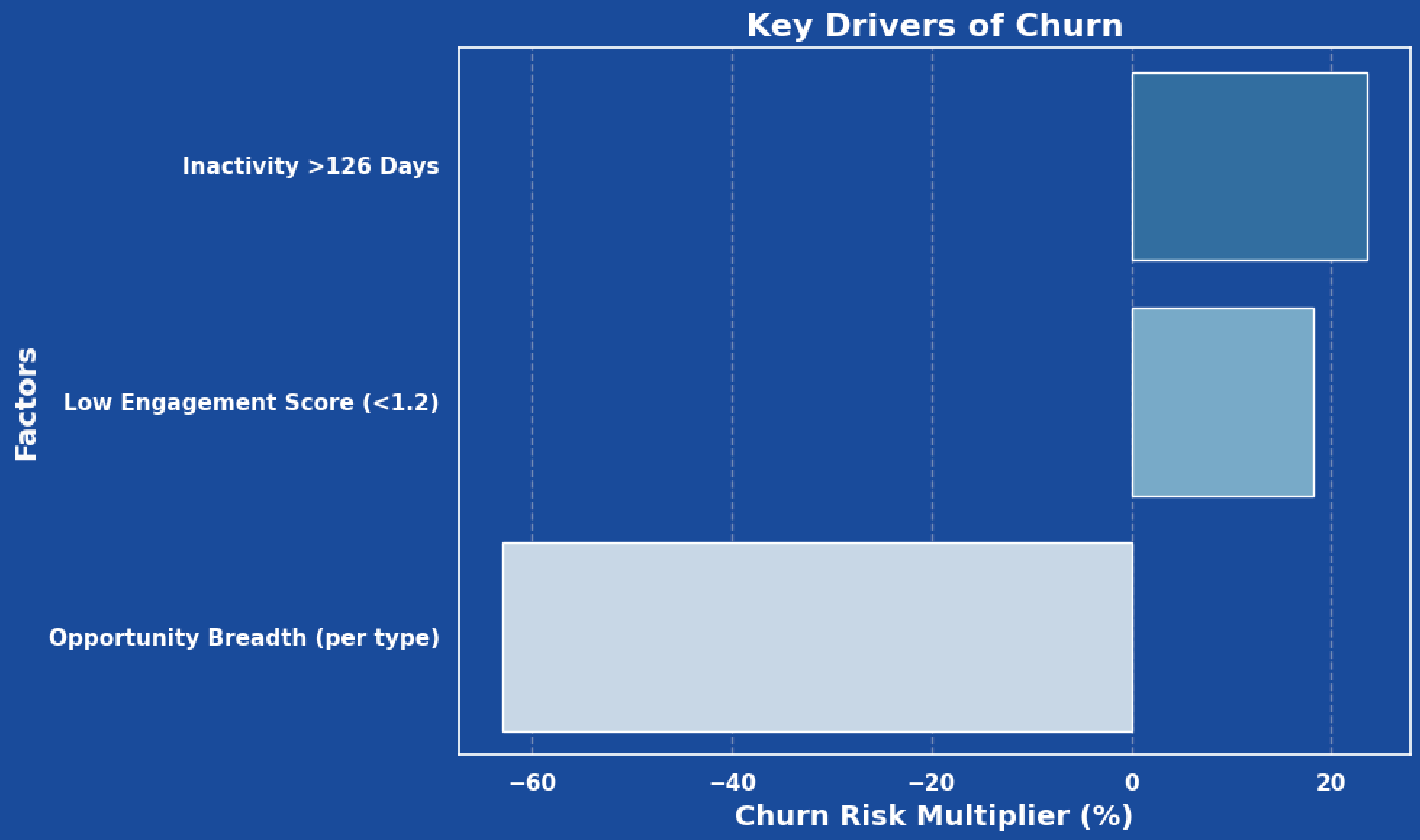
A user is considered churned if they meet the following conditions:

1. **Low Engagement Score:** An "Engagement Score" below 1.2 (representing the bottom 85% of users).
2. **Prolonged Inactivity:** More than 126 days since their last recorded activity (approximating an academic term).
3. **Zero Opportunity Participation**: No recorded participation in any opportunities.

- **Calculated Churn Rate:**

Based on this definition, the churn rate was determined to be 91.7% (2,269 out of 2,847 students).

Key Drivers of Churn



Why Students Drop Out

- Inactivity >126 days → **23.6x more likely to churn.**
- Engagement Score < 1.2 → **18.2x more likely to churn.**
- Opportunity Breadth: Each additional type reduces churn by 63%.

Predictive Modeling Approach

Machine Learning for Churn Prediction Content

Models Used

- Logistic Regression
- Random Forest
- XGBoost

Performance

All models achieved ~50% test accuracy

Struggled with recall due to class imbalance

Visuals: Table comparing model performance metrics, feature importance plot

Approach

Binary classification (churned = 1, active = 0)

SMOTE used for class imbalance

Feature selection: Age, Engagement Duration, Opportunity Category, etc.

Model Performance

How Well Did Our Models Work?

Cross-Validation Accuracy

The models achieved high accuracy in cross-validation, ranging from 91% to 94%.

Test Accuracy

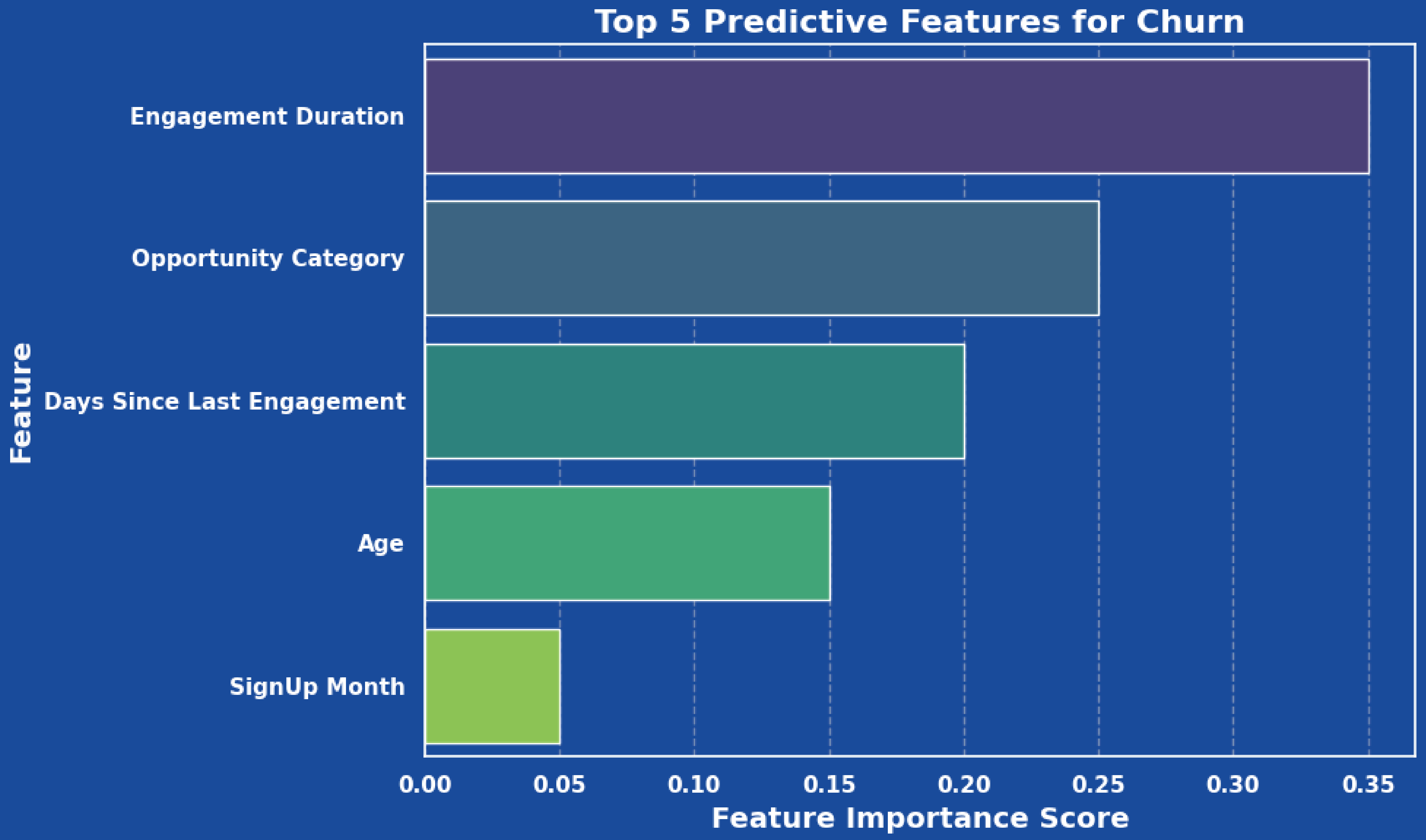
The accuracy on the hold-out test set was approximately 50%.

Key Challenge - Recall:

The models struggled to correctly identify the minority (churn) class, achieving a recall of 0.45. This indicates that while the model is often accurate, it misses a significant portion of the actual churners.

Model	Test Accuracy	Recall (Class 1)	F1-Score
Logistic Regression	49.93%	0.45	0.75
Random Forest	49.93%	0.45	0.75
XGBoost	49.93%	0.45	0.75

Top Features Identified



Top predictors of churn include Engagement Duration, Opportunity Category, and Days Since Last Engagement.

Business Impact of Churn Analysis

Financial & Operational Implications

Savings Potential

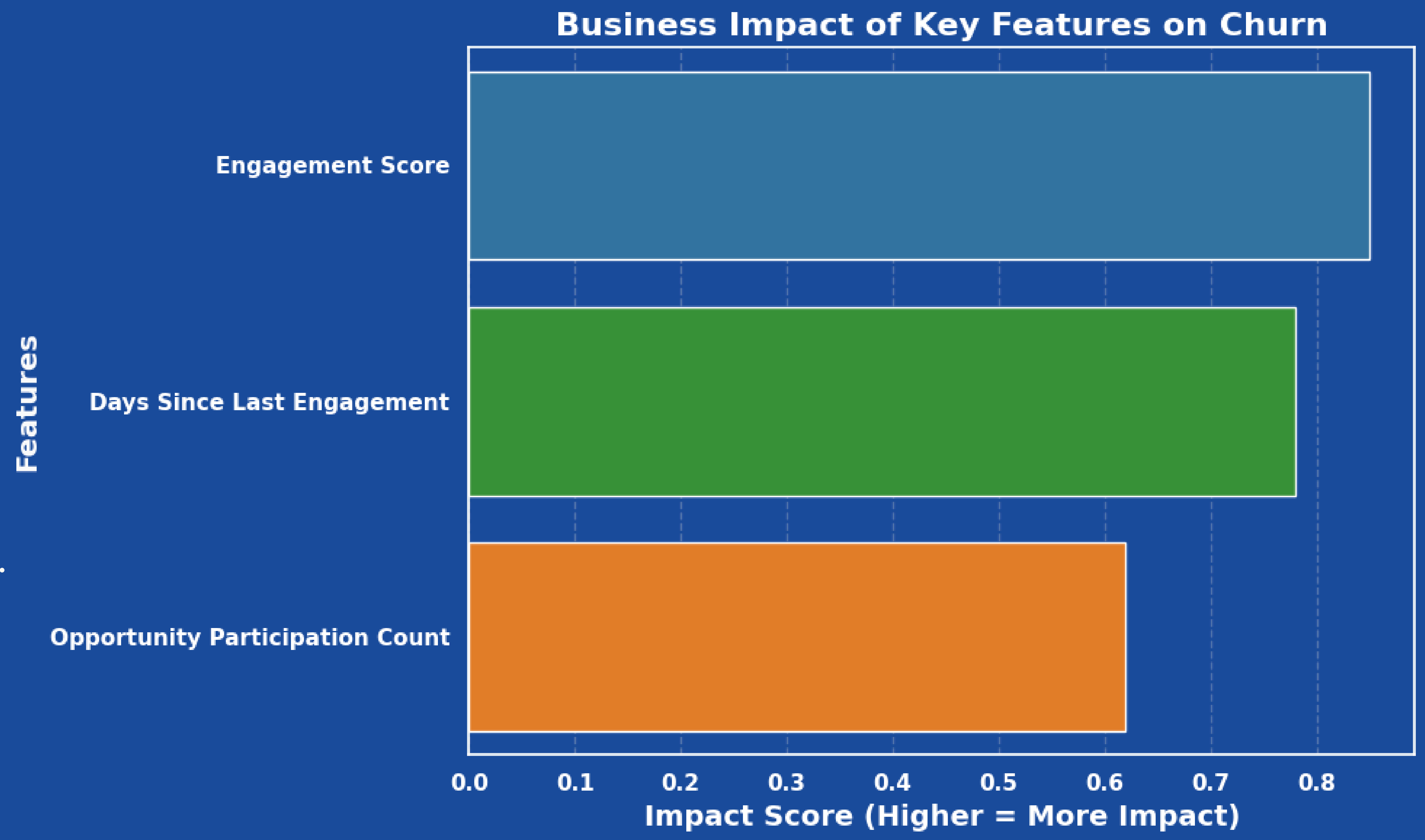
$34,600 saved/year with 10% churn reduction.

Community Impact

High churn reduces peer-to-peer learning by 28%.

Revenue Loss

$152 per churned student annually.



Actionable Strategies for Churn Reduction

Automated Re-engagement Campaigns:

Trigger personalized email nudges and offer micro-rewards (e.g., badges) for users who have been inactive for 90 days.

Expand High-Engagement Opportunities

Increase the number of competitions, as they show a 34% lower churn rate compared to courses. Introduce features like leaderboards and peer mentoring.

Promote Cross-Opportunity Bundling:

Create and market bundles of different opportunity types (e.g., a course followed by a related competition) to encourage broader participation.

Develop a Real-Time Monitoring Dashboard:

Create a dashboard to track at-risk students (Engagement Score < 1.2) for proactive intervention.

Gender-Segmented Communication:

Leverage the insight that females show 19% higher median engagement to tailor communication and potentially boost overall engagement.

The Excelerate EngagePro System

What It Does

- Monitors real-time engagement.
- Flags at-risk students.
- Sends personalized nudges.
- Recommends next-step opportunities.

Expected Benefits

- Proactive identification of at-risk students.
- Increased diversity in opportunity participation.
- Enhanced peer interaction and satisfaction.

Implementation Steps

Deploying the Recommendation System

Phase 1: Foundation (6 weeks)

- Build enhanced student profiles with behavioral tags
- Develop opportunity similarity matrix
- Implement basic hybrid recommender (content + collaborative)

Phase 2: Intelligence Layer (4 weeks)

- Integrate churn prediction model
- Develop seasonal adjustment engine
- Create recommendation audit dashboard

Phase 3: Intervention Systems (4 weeks)

- Build automated mentorship pipeline
- Implement engagement boosters for at-risk categories
- Develop completion incentive framework

Future Work on Retention Strategies

Next Steps for Improvement

Enhanced Feature Engineering

Incorporate sentiment analysis from user feedback

Real-Time Monitoring

Build live dashboard to flag at-risk users

Model Optimization

Try advanced balancing techniques (e.g., ensemble methods)

Longitudinal Studies

Track effectiveness of implemented strategies over time

We appreciate your time and welcome any questions or discussions regarding our findings and recommendations.

Thank You