AI - Team 1

Final Internship Project Report: Improving Student Engagement and Retention at Excelerate

Week - 4

Executive Summary

This report presents a comprehensive analysis as part of the AI Data Analyst Internship, focused on exploring student engagement data using data science and machine learning techniques. The primary aim was to identify which learning opportunities attract the most signups, are completed successfully, and to discover patterns that contribute to student drop-offs. Through structured data techniques including feature engineering, exploratory data analysis (EDA), and churn modeling, the project uncovers actionable insights and a basic recommendation system for retention strategies.

INTRODUCTION

As an AI Data Analyst intern, the focus of this project was to evaluate and understand engagement across student opportunities at Excelerate. The primary goals included:

- Identifying popular and high-completion opportunities
- Understanding the root causes of student drop-offs
- Designing data-driven interventions to enhance retention

Data Analysis Overview

Feature Engineering (Week 1)

- Cleaned inconsistencies in dates, duplicates, and categorical values

- Created new variables:
    - Engagement_Score = Avg(Opportunity Duration, Age, Engagement Days)
    - Repeat_Opportunities: Count per student
    - SignUp_Month, Engagement_Days for trend analysis

EDA (Week 2)

- Majority of learners aged 20-25
- Uneven gender representation
- High engagement in specific countries (India, US)
- Engagement patterns varied with time; no linear trends
- Identified top opportunities with high signup and completion rates

Popular Opportunities:

- A horizontal bar chart illustrated the Opportunity Category Distribution, revealing categories like "Tech Bootcamp" and "Business Accelerator" as the most enrolled.
- The Status Distribution Chart showed higher completion rates in opportunities with structured mentorship and shorter durations.
- The Top 10 Country-Wise Learner Distribution Chart reinforced the importance of tailoring opportunities based on regional interest.

EDA (Week 2)

- Majority of learners aged 20-25
- Uneven gender representation
- High engagement in specific countries (India, US)
- Engagement patterns varied with time; no linear trends
- Identified top opportunities with high signup and completion rates

Churn Modeling (Week 3)

- Created binary churn variable based on:
    - Engagement Score < 160
    - Engagement Days > 50
- Model: Random Forest Classifier (Accuracy: 99.87%)
- Key churn indicators:
    - Low engagement score
    - Delays between apply and start dates
    - Repeat opportunities

Key Insights

Insight	Takeaway
High signup ≠ High completion	Opportunity quality and timing matter
Low Engagement Score	Strongest dropout predictor
>50 Days Delay Before Start	Raises dropout chance by ~18%
Multiple Application Attempts	Suggests dissatisfaction or mismatch
Regional/Gender Gaps	Indicate need for targeted content & outreach
Popular Categories Identified	Tech and Business domains have highest interest

Recommendations

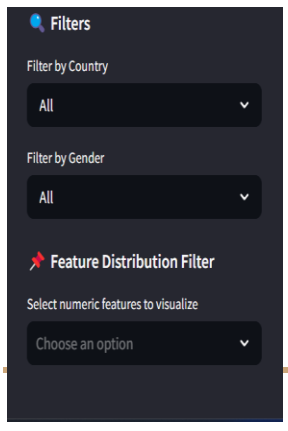
Practical Actions

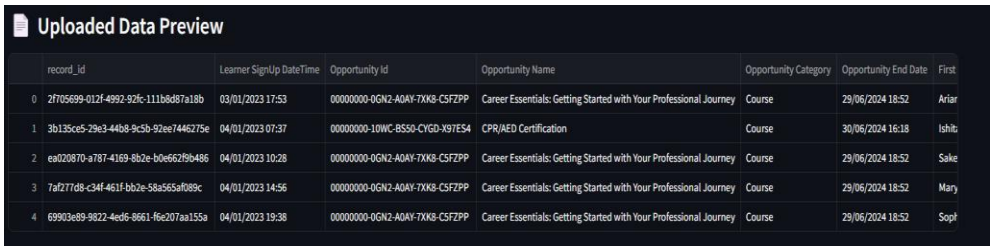
- Assign mentors to students with low initial scores
- Launch onboarding for students with >30-day start delays
- Reward early course completions
- Personalize dashboard with suggested tasks and milestones
- Promote opportunities in high-signup categories with proven completion success

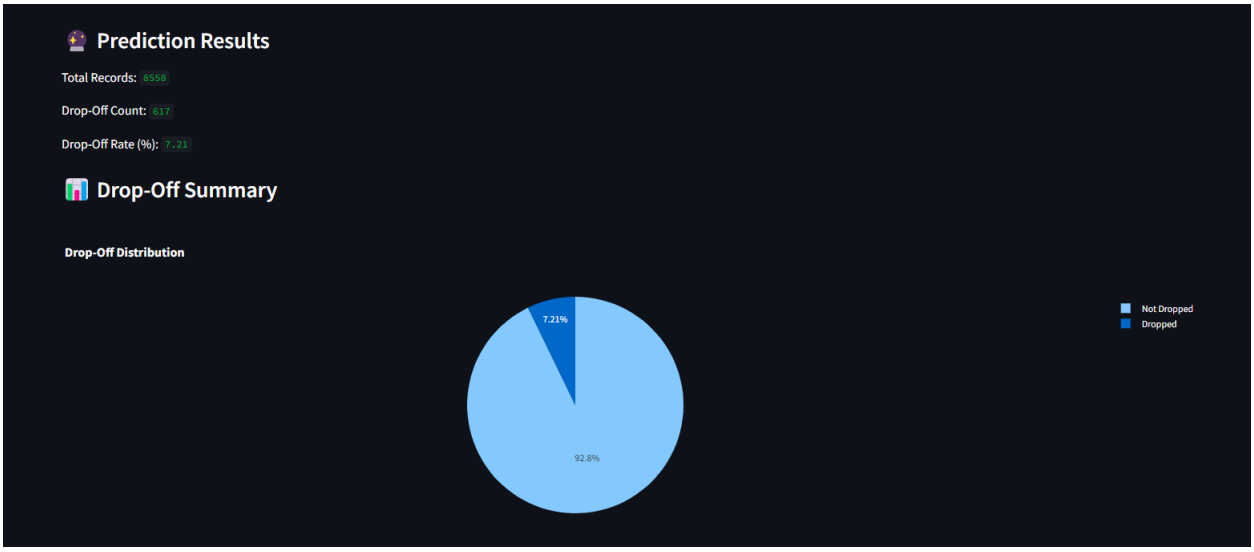
Strategic Initiatives

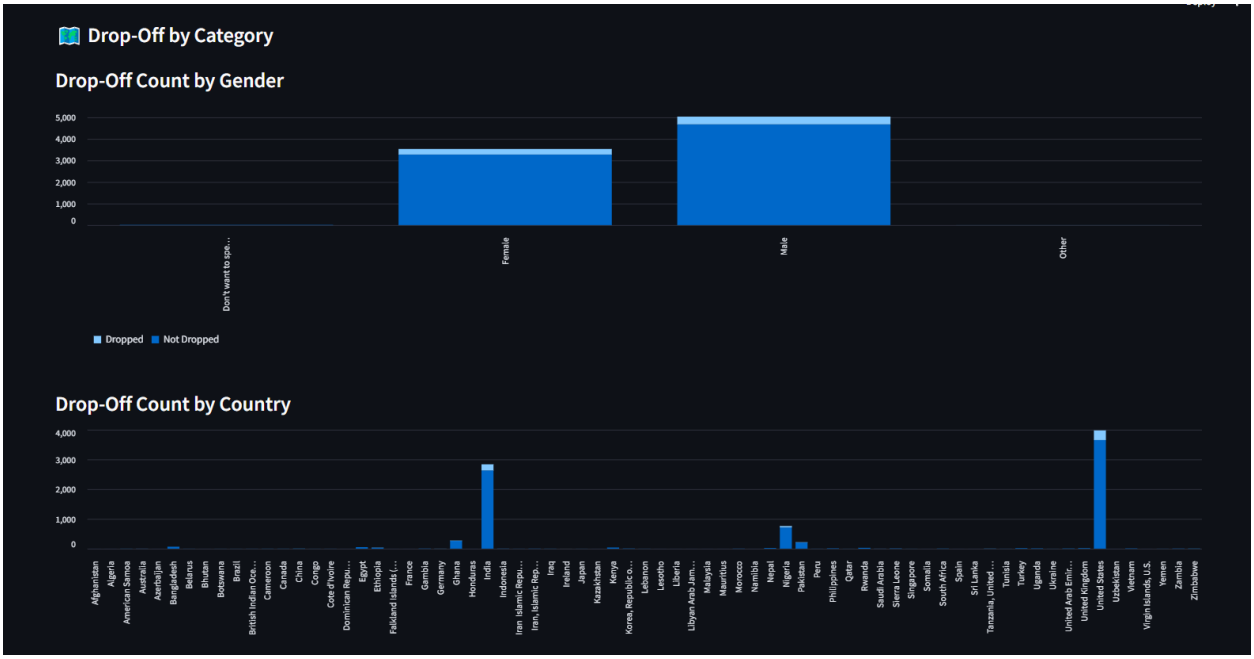
- Focus outreach in low-engagement regions
- Promote inclusive programs to close gender gaps
- Partner with regional institutions for better access
- Enhance offerings in popular domains (Tech, Business) with flexible scheduling

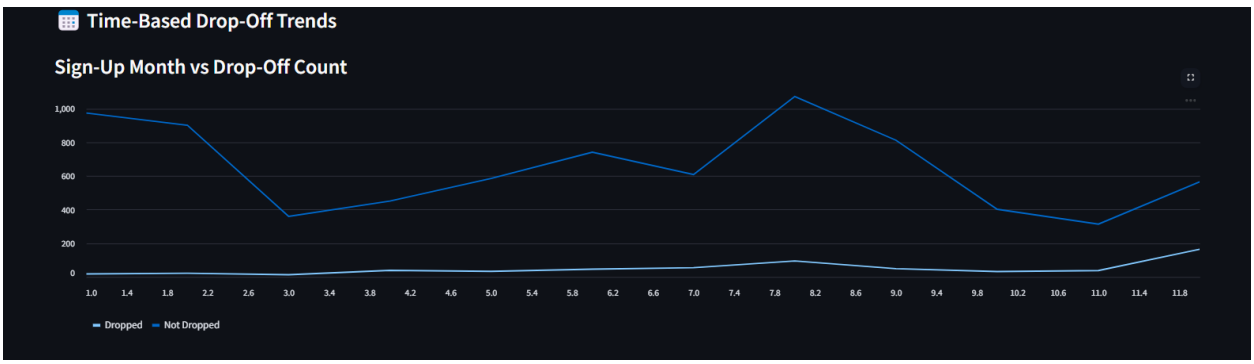
Snapshots of the Application

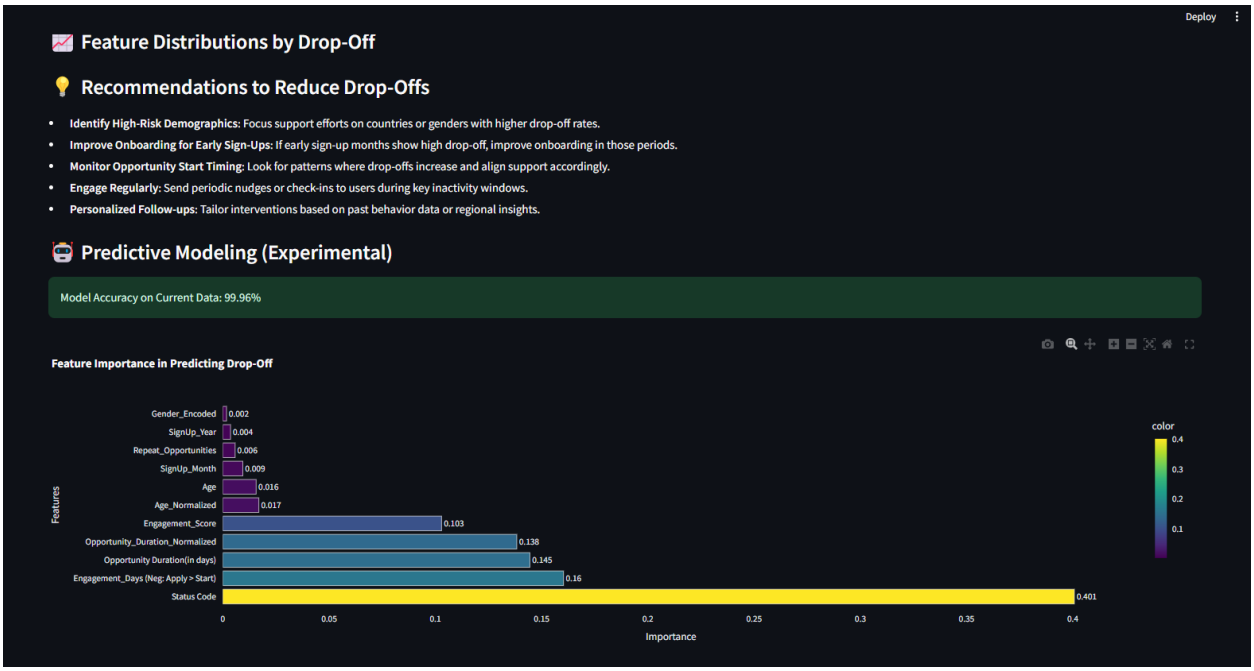





Prediction Results
Total Records: 8558
Drop-Off Count: 617
Drop-Off Rate (%): 7.21
Drop-Off Summary
Drop-Off Distribution
7.21%
92.8%
Not Dropped
Dropped


Drop-Off by Category
Drop-Off Count by Gender
5,000
4,000
3,000
2,000
1,000
0
Don't want to spe...
Female
Male
Other
Dropped
Not Dropped
Drop-Off Count by Country
4,000
3,000
2,000
1,000
0
Afghanistan
Algeria
American Samoa
Australia
Azerbaijan
Bangladesh
Belarus
Bhutan
Botswana
Brazil
British Indian Oce...
Cameroon
Canada
China
Congo
Cote d'Ivoire
Dominican Repu...
Egypt
Ethiopia
Falkland Islands (...
France
Gambia
Germany
Ghana
Honduras
India
Indonesia
Iran, Islamic Repu...
Iraq
Ireland
Japan
Kazakhstan
Kenya
Korea, Republic o...
Lebanon
Lesotho
Liberia
Libyan Arab Jam...
Malaysia
Mauritius
Morocco
Namibia
Nepal
Nigeria
Pakistan
Peru
Philippines
Qatar
Rwanda
Saudi Arabia
Sierra Leone
Singapore
Somalia
South Africa
Spain
Sri Lanka
Tanzania, United ...
Tunisia
Turkey
Uganda
Ukraine
United Arab Emi...
United Kingdom
United States
Uzbekistan
Vietnam
Virgin Islands, U.S.
Yemen
Zambia
Zimbabwe


Time-Based Drop-Off Trends
Sign-Up Month vs Drop-Off Count
1,000
800
600
400
200
0
1.0
1.4
1.8
2.2
2.6
3.0
3.4
3.8
4.2
4.6
5.0
5.4
5.8
6.2
6.6
7.0
7.4
7.8
8.2
8.6
9.0
9.4
9.8
10.2
10.6
11.0
11.4
11.8
Dropped
Not Dropped



Conclusion & Future Scope

This internship project explored student engagement through the lens of data science and machine learning. It successfully identified trends behind opportunity signups, completions, and drop-offs. The analysis led to actionable recommendations and a prototype system to improve learner retention.

Future Enhancements:

- Integrate real-time engagement dashboards
- Use NLP to analyze feedback and sentiment
- Leverage clickstream data for behavior modeling
- Expand opportunity tracking to include completion feedback and ratings