Churn Analysis Report

AI-Powered Virtual Internship , Excelerate

1. Introduction

In large-scale online programs, such as Excelerate's AI-Powered Virtual Internship, participant engagement and retention are critical markers of program success. While acquiring new learners is essential, retaining them through to completion is equally—if not more—valuable for ensuring impact, satisfaction, and future enrollment. This study investigates **student churn**, defined as learners who signed up for the program but failed to begin or complete it.

Our goal was to examine key behavioral, temporal, and demographic factors that influence churn. We analyzed cleaned, structured student data using exploratory visualizations and feature engineering to identify patterns. Though predictive modeling was attempted, the nature of the dataset led to a pivot toward descriptive analytics and insights-driven recommendations.

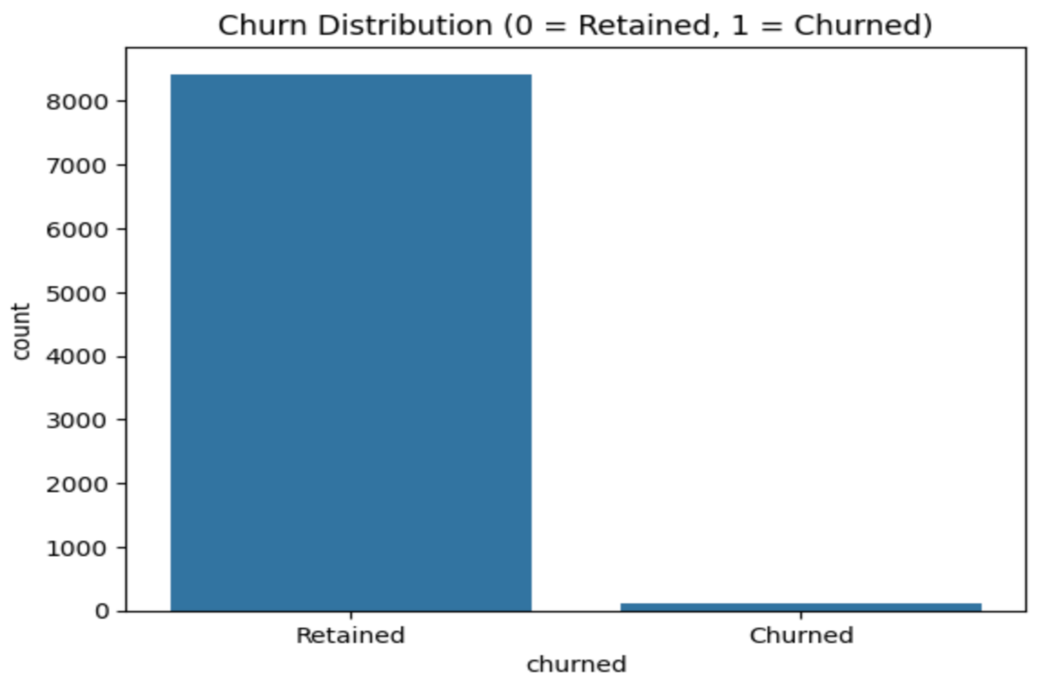
2. Dataset Overview

The dataset included records from **8,529 unique learners**, encompassing their personal demographics, academic context, signup/application timelines, and engagement indicators. From the original dataset, new features were derived to enhance interpretability:

- `CHURNED`: A binary indicator where 1 represents students labeled "Not Started" or "Applied" (indicating dropout), and 0 indicates successful progression or completion.
- `APPLICATION_LAG_DAYS`**:** A continuous variable representing the number of days between signup and actual application.
- `DAYS_SINCE_OPPORTUNITY_START`: A numeric encoding of when each opportunity began, calculated relative to the earliest start date.

This enriched feature set enabled us to conduct a detailed exploratory analysis across behavioral, temporal, and demographic axes.

3. Churn Distribution Overview



To begin, we assessed how churn was distributed across the population. The countplot visualization revealed that **only 105 learners (1.23%) churned**, while the vast majority—**8,424 learners (98.77%)**—remained active or completed the internship.

Although the churn percentage is low, in absolute terms, **over 100 students dropped off**, which is significant in terms of lost opportunity, resources, and learning outcomes. In such high-volume environments, even small percentages reflect large groups of users who can be better engaged through data-informed strategies.

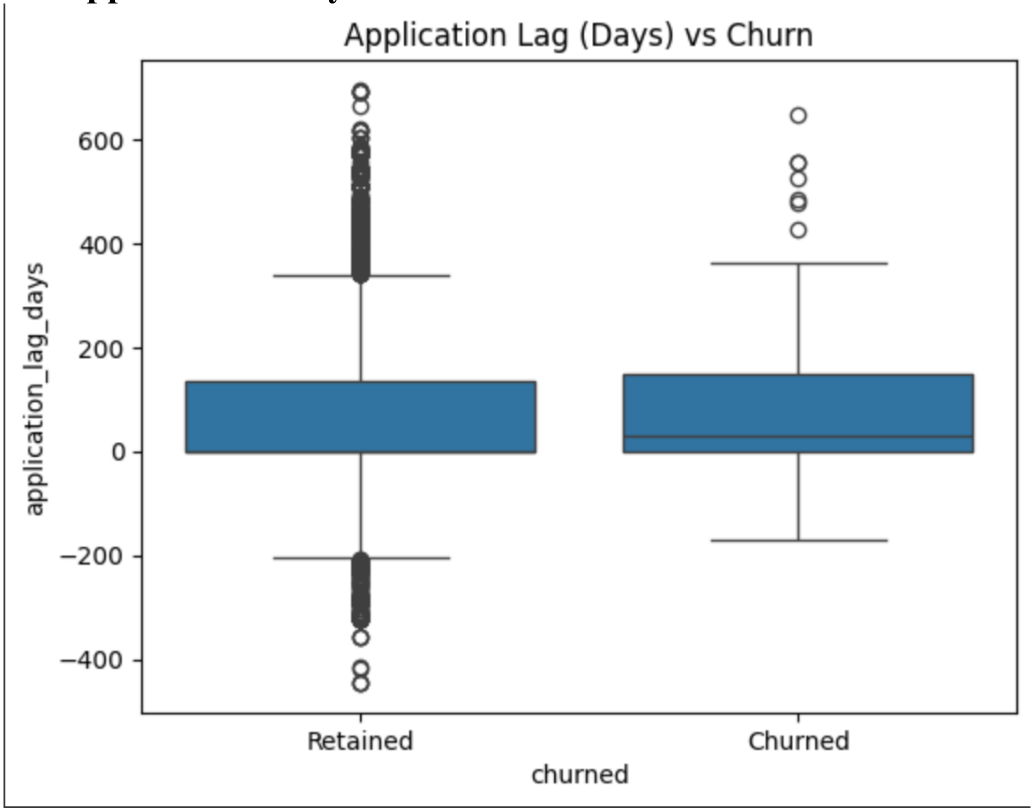
4. Exploratory Analysis of Key Factors

4.1 Age and Churn

Age was explored as a potential risk factor for churn. The boxplot visualization showed that churned learners tended to be **slightly older**, with an average around **26.2 years** compared to **25.0 years** for retained students.

This subtle but consistent difference may suggest that older participants face unique barriers—such as career obligations, academic workloads, or family responsibilities—that interfere with prompt engagement or program continuity.

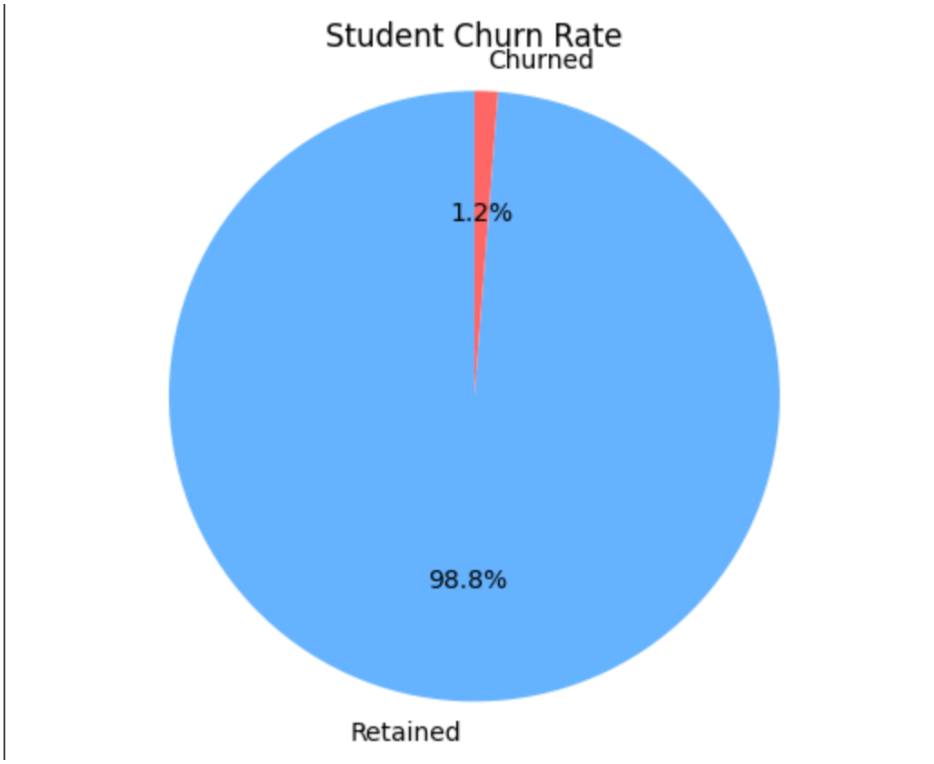
4.2 Application Delay and Churn



One of the most striking signals emerged from **application timing**. When comparing `application_lag_days` across groups, churned users exhibited a **mean delay nearly twice as long** as retained users.

This finding strongly supports the behavioral insight that **early commitment drives retention**. Students who apply promptly after signing up are likely already invested or have fewer decision barriers, while delayed applications correlate with indecision, lower urgency, or external distractions—all of which increase dropout risk.

5. Churn Rate Visualization



The pie chart visualization reinforces the numerical finding that churn is a **minority class**—only **1.2%** of the total learner population. However, from a program design perspective, these 105 churned students represent a missed opportunity for engagement. Proactively addressing the causes behind this small segment could yield outsized returns in retention improvement.

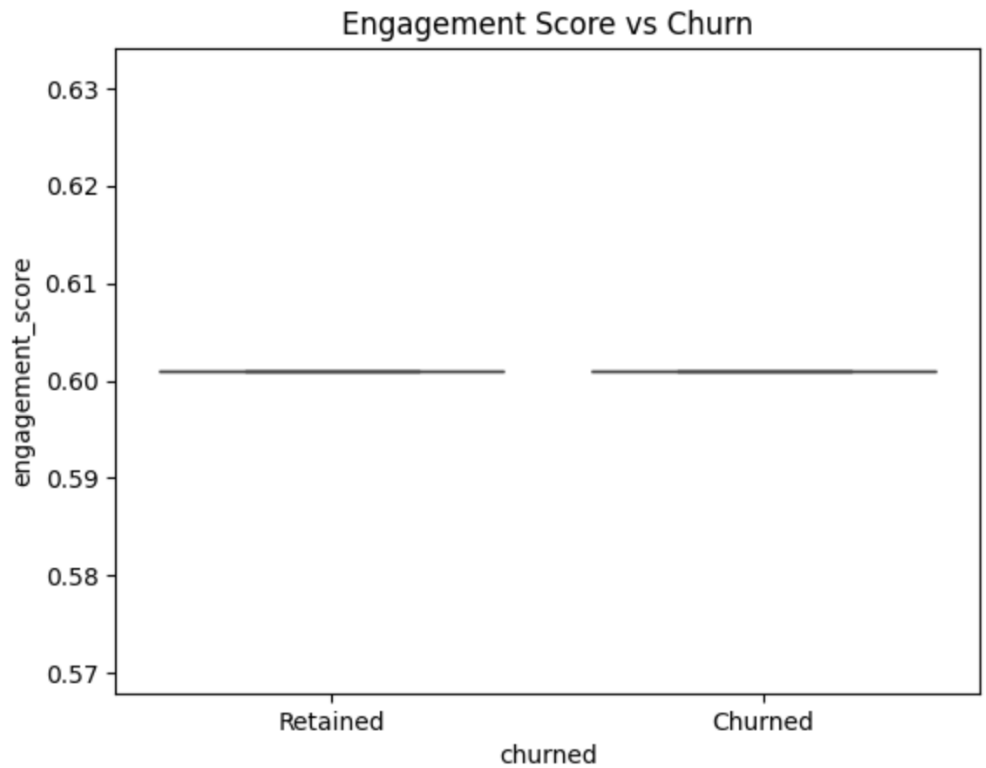
6. Comparative Statistics by Churn Status

Beyond visual exploration, we conducted statistical comparisons of key variables across churned and retained groups:

Metric	**Retained (mean)**	**Churned (mean)**	**Insight**
Age	25.0	26.2	Older students churn slightly more
Application Lag Days	15.1	29.1	Delay is a strong churn predictor

These results complement the visualizations and emphasize the significance of early behavioral cues, particularly **delayed application** as a precursor to disengagement.

7. Feature Removal: Engagement Score



Initially, we expected `engagement score` to be a core predictor. However, upon inspection, the variable was found to be **completely static**—every student received the same engagement score of `0.601`. This lack of variation rendered the feature useless for modeling or visualization. The flat boxplot confirms the uniformity.

This limitation underscores the need for richer behavioral data, such as login frequency, module activity, or message interactions, to quantify true engagement in future cohorts

9. Recommendations

Based on the findings above, we propose the following interventions:

- **Trigger early nudges post-signup**: Send personalized reminders within 24–72 hours to drive immediate application behavior.
- **Improve engagement tracking**: Replace the current engagement metric with multidimensional measures (e.g., time-on-task, activity streaks).
- **Monitor older learners more closely**: Tailor communication or support for older students who may have external time constraints.
- **Continue churn tracking longitudinally**: Re-analyze churn with each cohort and refine strategies using more granular behavioral data.

10. Conclusion

This churn analysis highlights the importance of **application timing** and **learner demographics** in predicting drop-offs. While most students complete or remain active, early disengagement still affects a measurable portion of the population. By encouraging fast action, enriching engagement data, and targeting at-risk profiles, Excelerate can further enhance the success of its AI virtual internship program.