Excelerate

DATA VISUALIZATION REMOTE INTERNSHIP

Week-4 Deliverable

FINAL DASHBOARD & STRATEGIC INSIGHTS SUMMARY REPORT

Team-8

Prepared By:

Sumaiya Tasnim
Zainab Fatima
Faizan
Arop Kuol
Shri Vidya

CONTENTS

Strategic Insight Report: Learning Analytics, Prediction & Pathway Optimization

1. Executive Summary:

This strategic insight report presents a comprehensive analysis of learner engagement, progression, and success patterns on the Excelerate platform, conducted over a four-week internship project. Using a combination of exploratory data analysis (EDA), key performance indicator (KPI) evaluation, predictive modeling, pathway optimization, and dashboard-based visualization, the study aims to uncover actionable insights that can improve learner outcomes and platform effectiveness.

Across the project timeline, we analyzed learner demographic data, opportunity characteristics, engagement timelines, and status transitions to identify critical friction points in the learner journey. The findings reveal high initial learner intent, significant filtering during competitive opportunities (particularly internships), and strong variation in outcomes based on opportunity category, geography, and learner background.

Predictive models were developed to estimate learner progression likelihood, and pathway optimization strategies were proposed to reduce drop-offs and align learners with suitable opportunities. The final Power BI dashboard consolidates these insights into an interactive decision-support tool for stakeholders. Overall, the project provides data-backed recommendations to improve learner retention, optimize opportunity design, and support evidence-based decision-making.

2. Overview of Datasets Used:

The analysis is based on the SLU Opportunity Wise Dataset, sourced from the Excelerate platform. The dataset captures learner interactions with opportunities at an individual level, allowing detailed examination of engagement timelines, outcomes, and demographic influences.

Dataset Characteristics:

· Initial size: 8,558 records × 16 variables

· Final cleaned dataset: 8,246 records × 16 variables

· Granularity: One record per learner per opportunity

Data Dimensions Covered:

· Learner demographics (age derived from DOB, gender, country)

· Academic background (institution, intended/current major)

· Attributes (category, start date, end date, duration)

· Engagement behavior (time to apply, application lead time)

· Outcome indicators (status description and status code)

A rigorous data cleaning process was applied using Python to handle missing values, remove duplicates, standardize categorical variables, correct inconsistencies, and validate data types. The cleaned dataset served as the foundation for all subsequent analysis, modeling, and visualization.

Column Name	Meaning / Description	Datatype
Learner Sign Up Date Time	Date and time when the learner registered in the system	Date/Time
Opportunity Id	Unique identifier for each opportunity (Primary key)	Categorical
Opportunity Name	Name of the opportunity applied for	Categorical
Opportunity Category	Type or category of the opportunity	Categorical
Opportunity End Date	Scheduled end date of the opportunity	Date/Time
First Name	Learner’s first name	Categorical
Date of Birth	Learner’s date of birth	Date/Time
Gender	Learner’s gender	Categorical
Country	Learner’s country of residence	Categorical

Institution Name	Name of the learner’s institution	Categorical
Current/Intended Major	Learner’s field of study or intended major	Categorical
Entry created at	Date and time when the entry was recorded in the system	Date/Time
Status Description	Description of the learner’s application status	Categorical
Status Code	Numeric code representing the application status	Numerical
Apply Date	Date when the learner applied for the opportunity	Date/Time
Opportunity Start Date	Scheduled start date of the opportunity	Date/Time

This table summarizes the structure of the learner opportunity dataset used throughout the analysis, highlighting key demographic, engagement, and outcome-related variables.

3. Summary of Key Findings (Week 1 to Week 3):

Week 1: Foundational Learning Analytics & EDA

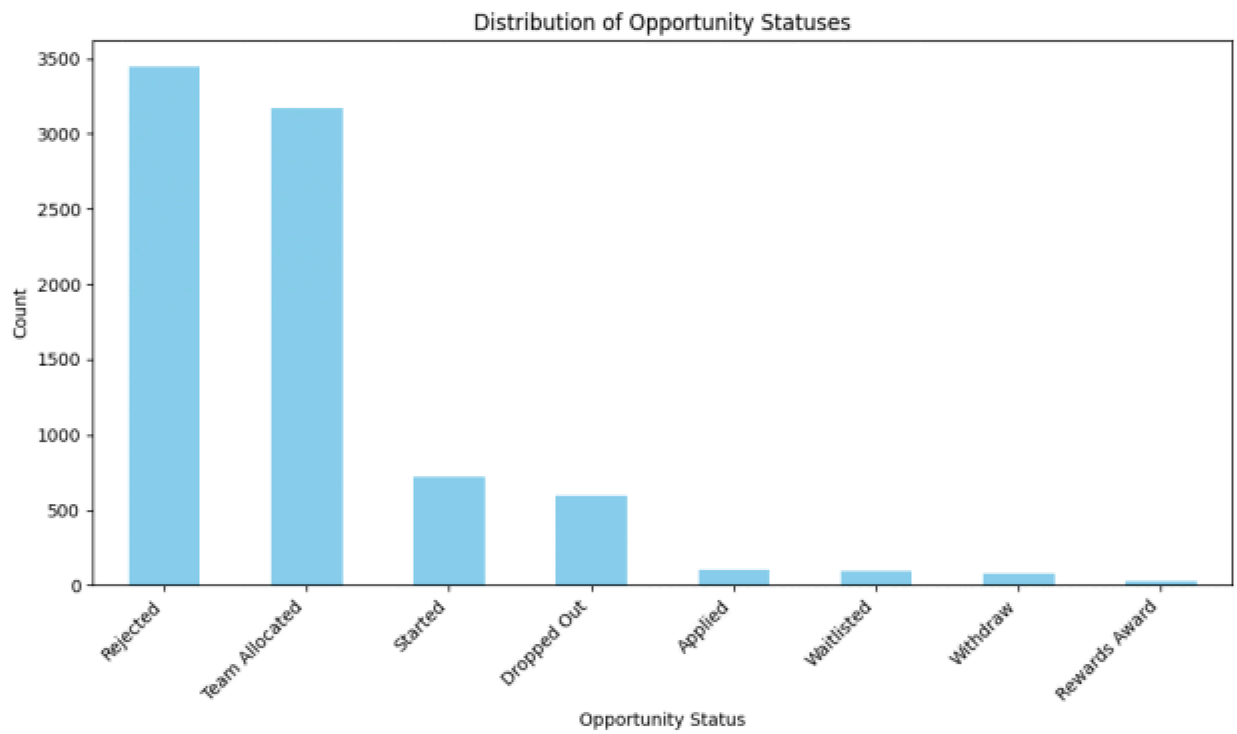
Week 1 focused on understanding learner behavior and platform dynamics through descriptive analysis and KPI exploration.

Key Insights:

1. Learners show very high initial intent, with a median sign-up to application time of ~4 days.
2. Engagement is heavily concentrated in Internships (64%) and Courses (23%).

3. Completion outcomes are rare, with Rewards Awarded at only ~0.35%, highlighting high program selectivity.
4. The United States and India together account for nearly 80% of platform activity.
5. Most learners fall within the mid-20s age range, indicating early career dominance.

This week established baseline KPIs, learner funnels, and early hypotheses about engagement and drop-offs.



This visual shows learner progression across opportunity statuses, highlighting significant filtering at the rejection and allocation stages and the very low completion rate.

Week 2: Behavioral Analysis & Predictive Modeling

To understand and reduce learner drop-offs, we adopted a dual-model analytical approach using **Logistic Regression** and a **Decision Tree Classifier**. Logistic Regression was selected to quantify *probability-based progression patterns* and reveal how timing variables (Apply Delay, SignUp Month, Opportunity Duration) influence completion likelihood. This model allowed us to identify early disengagement clusters and learners at highest risk of stagnation. In parallel, the Decision Tree was trained to capture *stage-by-stage structural bottlenecks* and reveal the exact pathway splits where learners either progress or drop off. While Logistic Regression provides broad, probability-driven risk insights, the Decision Tree exposes clear behavioral choke points within specific opportunity types and journey stages. Together, the two models allow a comprehensive view: **Logistic Regression predicts who is likely to drop off**, while the **Decision Tree shows where and why drop-offs happen**, enabling targeted pathway optimization.

Week 2 extended the analysis into behavioral segmentation and predictive modeling, focusing on identifying factors that influence learner progression.

Key Contributions:

1. Engineered predictive variables such as:

· Time from sign-up to application

· Opportunity duration

· Application lead time

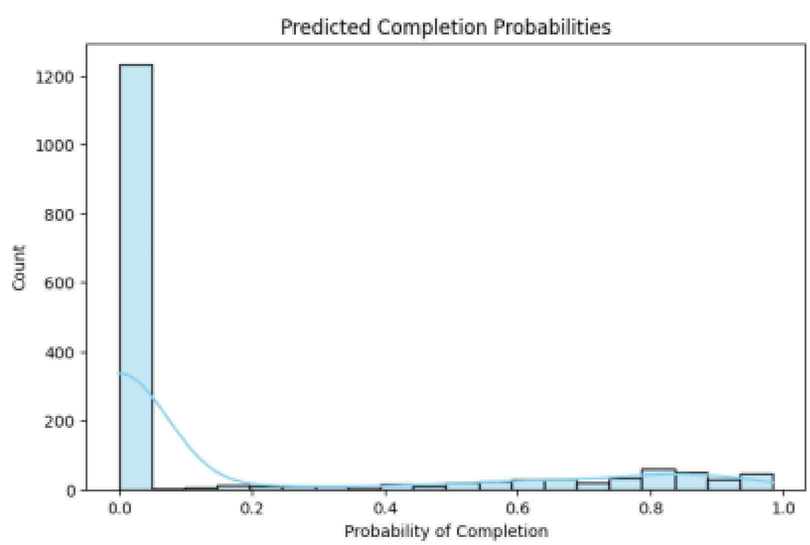
· Learner age at sign-up

2. Built classification models (Logistic Regression and Decision Tree) to predict learner outcomes (e.g., progression vs. rejection/drop-off).

3. Identified time-based engagement variables as the strongest predictors of learner success.

4. Confirmed that opportunity category is a major determinant of outcomes, often out weighing demographic factors.

These models provided a data-driven basis for anticipating learner behavior and identifying high-risk segments.



The above visual outlines the predictive modeling approach used to estimate learner progression outcomes based on engagement, demographic, and opportunity-related variables.

Week 3: Learning Pathway Optimization & Bottleneck Analysis

Week 3 shifted focus from prediction to intervention, analyzing where learners drop off and how pathways can be optimized.

Key Insights:

1. Internships exhibit the highest friction:

· ~66% rejection rate

· Highest dropout after allocation

2. Courses act as a low-friction entry point:

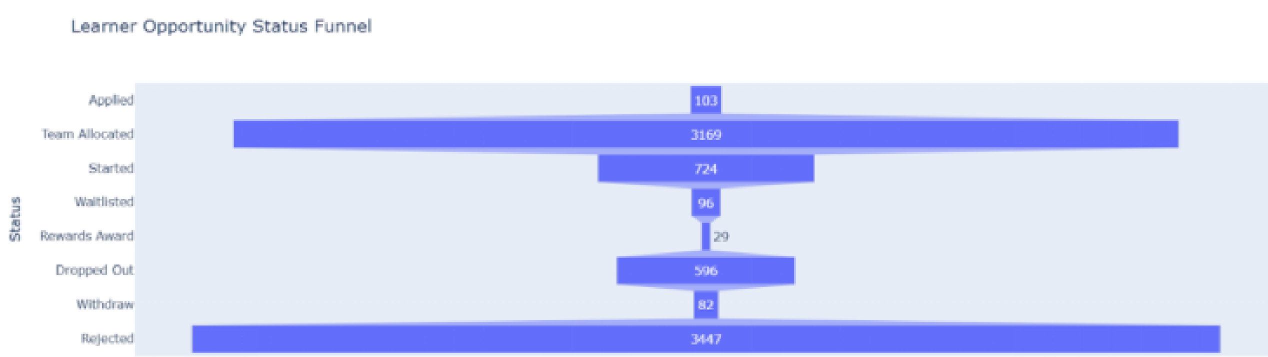
· Near-zero rejection

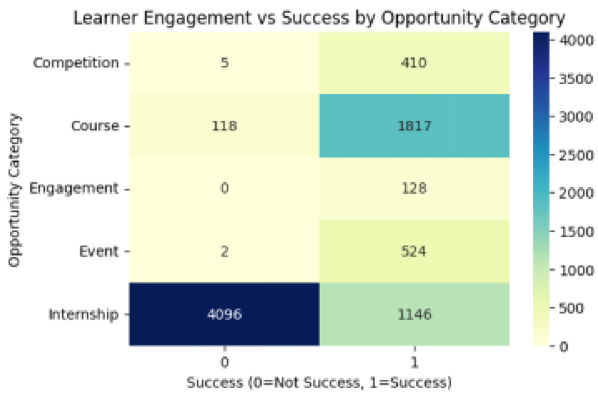
· High progression from allocation to started

3. Country-level differences in "performance" are primarily driven by opportunity preference, not capability.

4. "Team Allocated" emerges as a critical transition stage with inconsistent interpretation across categories.

Strategic pathway recommendations were developed, such as redirecting high-risk internship applicants toward preparatory courses and improving clarity around post-allocation expectations.





The diagram maps a typical learner journey, highlighting key drop-off points and rapid progression stages.

4. Pathway Optimization:

The optimized learner pathway clearly differentiates between high-stakes and low-friction learning journeys, allowing the platform to better align learner expectations with opportunity requirements.

High-Stakes Path (Internships):

- This pathway is characterized by intense competition, strict eligibility requirements, and higher rejection rates.
- Learners following this path require stronger pre-screening mechanisms, clearer communication of eligibility criteria, and access to preparatory learning content to improve readiness.
- Without structured preparation, this pathway leads to higher attrition and disengagement after allocation.

Nurture Path (Courses):

- This pathway supports gradual skill development, lower entry barriers, and smoother progression.
- Courses enable learners to build confidence, acquire relevant competencies, and remain engaged with minimal friction.
- Learners on this path demonstrate higher retention and more consistent progression outcomes.

5. Prediction Model Results:

Model Performance Comparison

Metric	Logistic Regression	Decision Tree	Interpretation Summary
Accuracy	80%	92%	Decision Tree achieves higher overall correctness; Logistic model influenced by class imbalance.
Precision (Class 1)	0.02	0.94	Logistic model predicts very few true completions; Decision Tree provides highly reliable success predictions.
Recall (Class 1)	1.00	0.88	Logistic captures all actual completions (high sensitivity); Decision Tree misses a small portion but remains strong.
F1-Score (Class 1)	0.03	0.91	Decision Tree offers balanced effectiveness; Logistic shows imbalance-driven trade-offs.
ROC-AUC	0.97	0.95	Both models demonstrate excellent discriminative ability; Logistic slightly higher due to probability calibration.

Model Performance Summary:

- The Logistic Regression model provided stable performance and strong interpretability, making it suitable for understanding how individual variables influence learner outcomes.
- The Decision Tree model captured non-linear relationships and interaction effects between engagement and opportunity-related variables.
- Across both models, engagement timing variables such as time from sign-up to application and opportunity category consistently ranked as the strongest predictors of learner success.

Key Takeaway

Learner behavior specifically when learners engage and how quickly they act is more predictive of success than static demographic attributes. This reinforces the importance of behavior driven analytics for early intervention and opportunity matching.

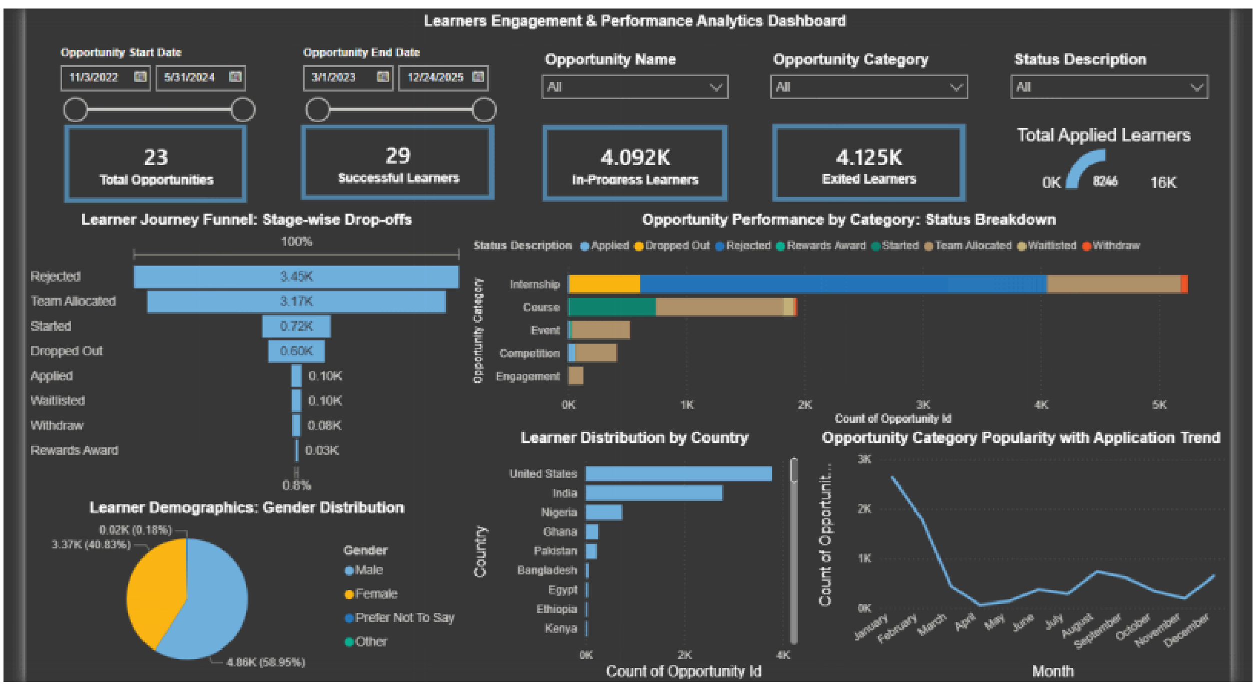
6. Dashboard & Visualization Overview:

The final Power BI dashboard consolidates all analytical insights into a single interactive interface, designed to support both operational monitoring and strategic decision making.

Through this dashboard, stakeholders can:

- · Monitor key KPIs such as allocation, rejection, and drop-off rates in real time
- · Compare performance across opportunity categories to identify high- and low-friction programs
- · Track engagement trends over time, including seasonal surges and declines
- · Analyze geographic and demographic distributions to understand market concentration and diversity

Overall, the dashboard transforms complex analytical outputs into clear, actionable intelligence, enabling faster and more informed decisions.



The dashboard presents an integrated view of learner engagement, progression status, geographic distribution, and time-based trends across the Excelerate platform.

1. KPI Cards

KPI cards provide a high-level summary of platform performance metrics for quick assessment.

2. Status Distribution Chart

This chart shows how learners are distributed across different opportunity statuses.

3. Monthly Trend Line

The trend line highlights seasonal variations in learner engagement over time.

7. Final Recommendations:

· **Introduce Pre-Internship Readiness Pathways**
Use courses as structured preparation funnels to improve learner readiness and reduce rejection rates.

· **Refine Opportunity Matching**
Leverage predictive models to guide learners toward opportunities aligned with their engagement behavior.

· **Clarify Status Transitions**
Standardize the meaning of “Team Allocated” to improve transparency across categories.

· **Address Seasonal Demand**
Prepare operational capacity in advance for peak intake periods such as January.

· **Close the Feedback Loop**
Introduce a clear “Completed” or “Outcome Achieved” status to track learner success more accurately.

8. Future Direction Suggestions:

· Deploy predictive models into production to enable real-time learner recommendations

· Conduct A/B testing to validate the effectiveness of optimized learning pathways

· Incorporate learner feedback and assessment results to enrich outcome measurement

· Expand dashboards to support cohort-level and longitudinal analysis

· Explore personalization through adaptive and behavior-driven learning paths

Final Note:

This strategic insight report demonstrates how learning analytics can evolve from descriptive reporting into strategic decision-making and optimization. By integrating analytics, predictive modeling, pathway design, and visualization, the framework provides a scalable approach to improving learner success while aligning platform strategy with data driven insights.

Our Project's Mandatory All Deliverable's Links:

1. **Dashboard Report:**
   Learner's Engagement & Performance Analytics Dashboard Report.pdf

2. **Our Project Presentation (PPT Slides) Link: Presentation DV Team-8.pdf**

3. **Our Virtual Presentation Recording Link:**
   DV Team-8 Virtual Presentation Recording.mp4

4. **To view our every weekly deliverables files (All Weekly Reports, Cleaned dataset, code files→ .ipynb, Power BI Dashboard file → .pbix) click in this Github link:**
   https://github.com/sumaiya-tasnim-18/IIT--Data-Visualization-Remote-Internship