



Data Visualization Remote Internship

Project Title: Learner Pathway Optimization & Predictive Analytics



Team 8

Presented By

Name	Email	Role
Sumaiya Tasnim	sumaiyaa.tasnim.18@gmail.com	Team Lead
Faizan	faizanlal563@gmail.com	Project Manager
Arop Kuol	Kuolbdka@gmail.com	Project Lead
Shri Vidya	vidyashri750@gmail.com	Project Scribe
Zainab Fatima	malikzainab1901@gmail.com	Team Member

Our 4 week Internship's journey

WEEK-1

Data Exploration & Foundations

- Explored the dataset, cleaned and performed initial EDA.
- Identified key variables for prediction.
- Mapped early learner pathways.

WEEK-2

Prediction Model & Behavior Visuals

- Built our first prediction models - Logistic Regression & Decision Tree
- Created visuals showing user behavior and drop-offs.
- Analysed engagement trends

WEEK-3

Pathway Analysis & Optimization

- Identified bottlenecks in learner pathways.
- Proposed optimization ideas.
- Refined our model and visuals with new insights.

WEEK-4

Final Dashboard & Strategic Reporting

- Built the final interactive dashboard in POWER BI.
- Prepared the strategic insights report.
- Created the final stakeholder presentation.

01

02

03

04

Next >

Dataset Overview

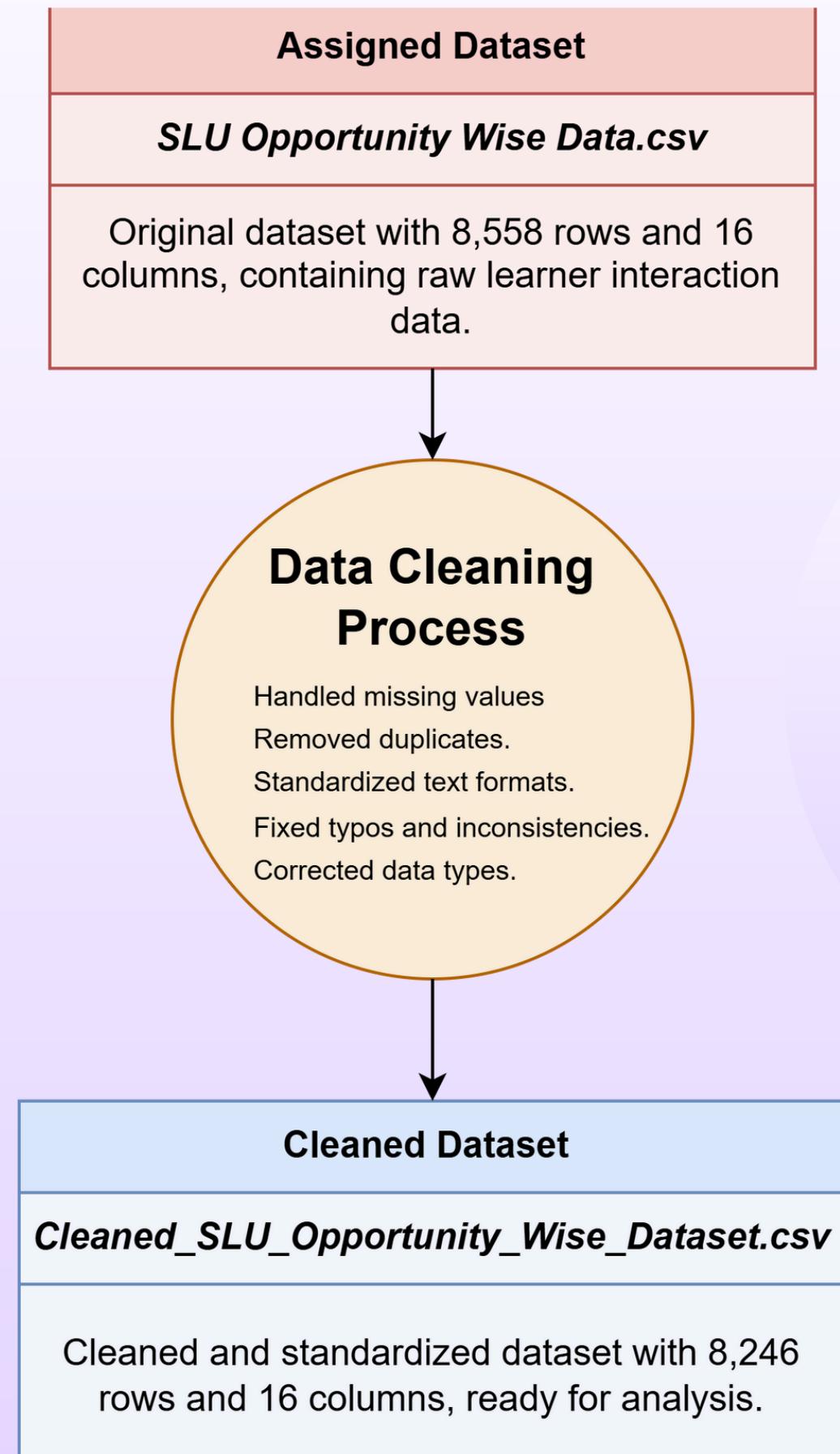
The dataset includes learner and opportunity details across multiple fields.

Key columns are **Learner SignUp DateTime**, **Opportunity Id**, **Opportunity Name**, **Opportunity Category**, **Opportunity Start Date**, and **Opportunity End Date** for **opportunity tracking**.

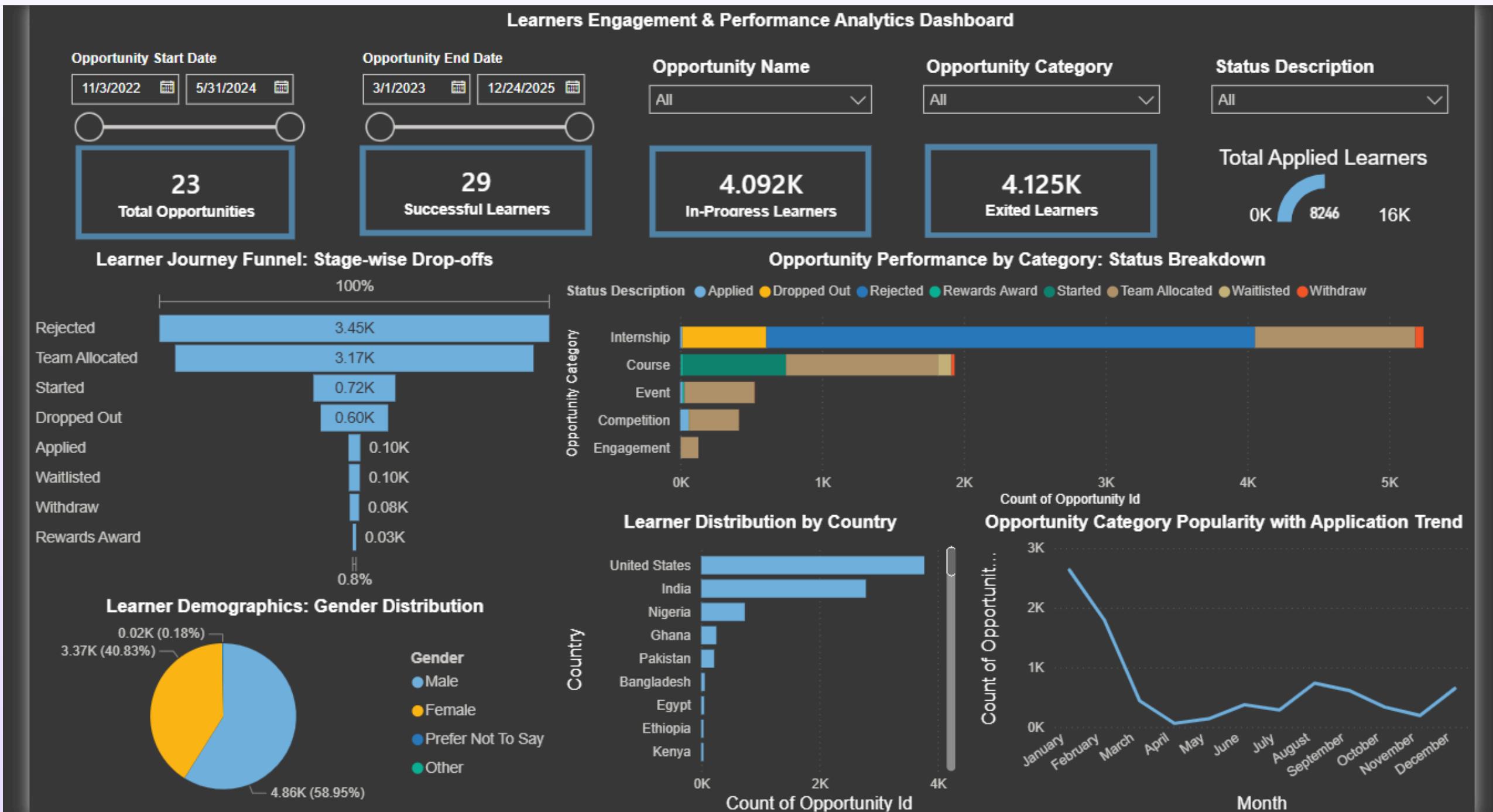
Learner attributes include **First Name**, **Date of Birth**, **Gender**, **Country**, **Institution Name**, and **Current/Intended Major**.

Application details are captured in **Apply Date**, **Status Description**, and **Status Code**, while Entry created at records system timestamps.

These date/time, categorical, and numerical columns provide a structured view for analyzing learner profiles, engagement, and opportunity progression.



Learners Engagement & Performance Analytics Dashboard



KPIs

- **Total Opportunities** → 23 → Large pool, low engagement
- **Successful Learners** ("Rewards Award") → 29 → Extremely low completion
- **In-Progress Learners** (Applied, "Team Allocated", "Started", "Waitlisted") → 4.092K → Huge mid-journey friction
- **Exited Learners** ("Dropped out", "Rejected", "Withdraw") → 4.125K → Attrition equals in-progress
- **Total Learners** → 8,246 → Shows overall funnel reach

Filters

The applied filters enable focused and flexible analysis by examining seasonal effects through **Opportunity Start Date**, duration impact via **Opportunity End Date**, and performance differences across **Opportunity Category**. Drilling down by **Opportunity Name** helps identify underperforming programs, while **Status Description** isolates key outcomes such as drop-outs, team allocations, and rewards, supporting targeted diagnostics and optimization.

Dashboard's Visual Insight: The platform attracts strong interest with 8,246 applicants, but only 29 complete, showing a major conversion gap. The largest drop occurs between "Team Allocated" → "Started" (~3K learners), with Internships having high volume but high dropouts, while Courses show better retention. Regional and seasonal trends reveal access barriers and early enthusiasm fading quickly, highlighting the need for improved onboarding, sustained engagement, and operational support.

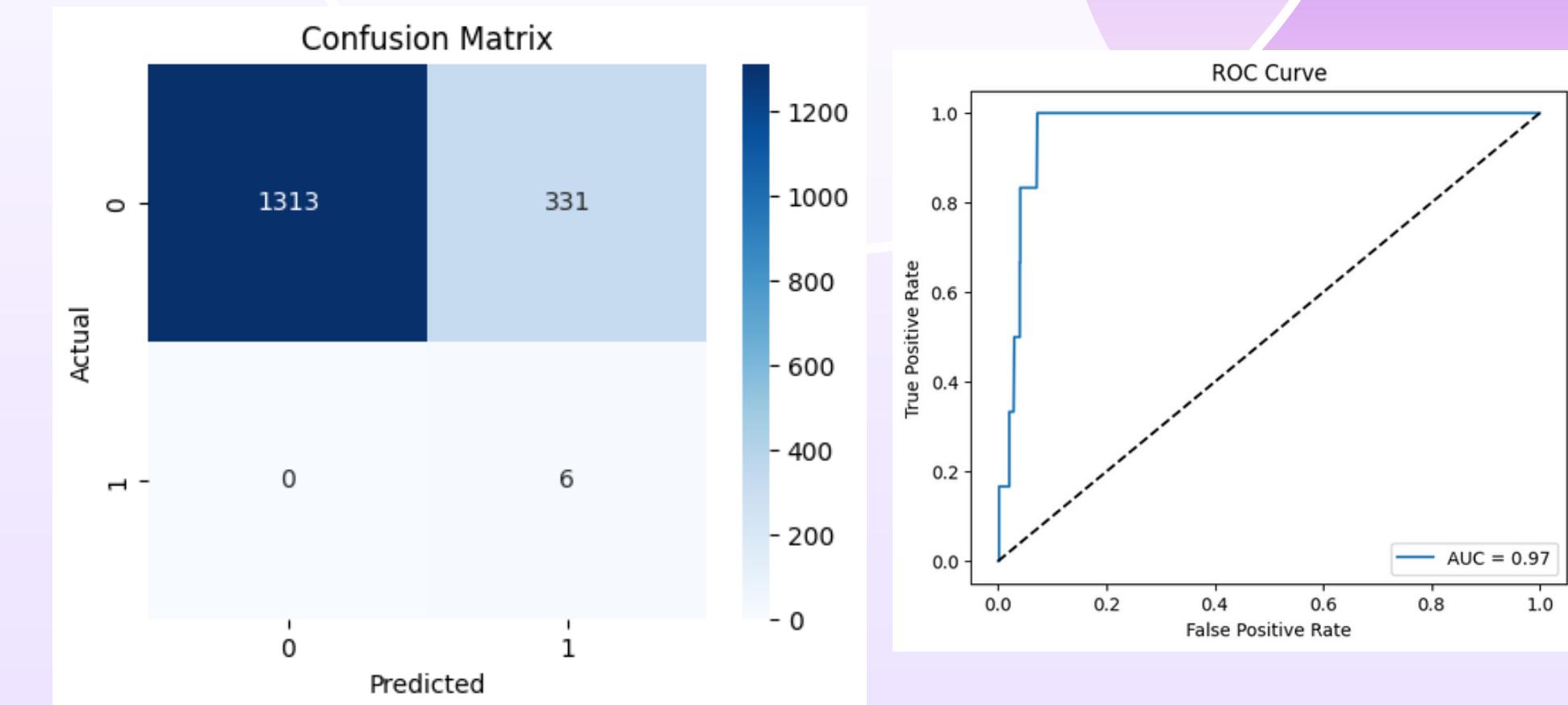
Learners Drop-Offs Solution Approach by Predictive Models

An overview of our **key performance metrics of both Models**

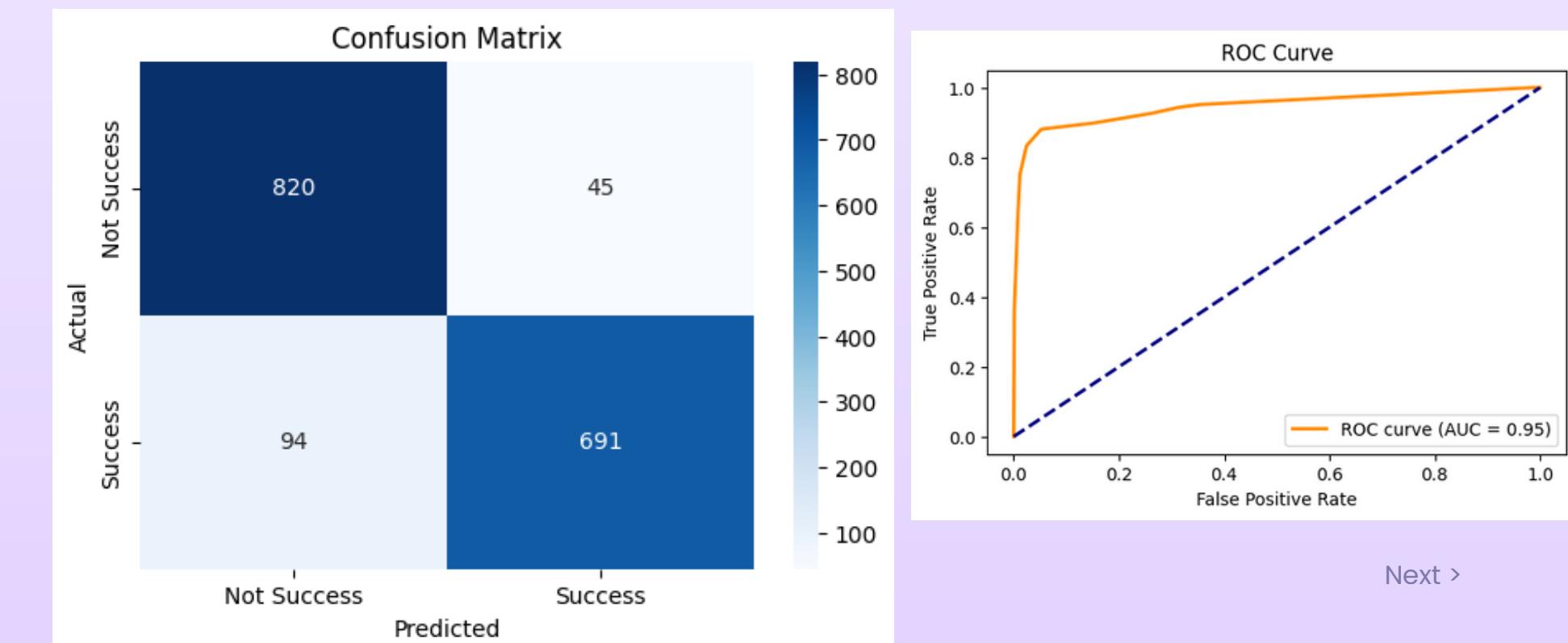
Metric	Logistic Regression	Decision Tree
Accuracy	80 %	92 %
Precision	0.02	0.94
Recall	1.00	0.88
F1-Score	0.03	0.91
ROC-AUC	0.97	0.95

- Logistic Regression emerged as the strongest predictive model, offering stable accuracy and reliable probability outputs for learner completion.
- It outperformed Decision Tree in consistency and generalization, avoiding overfitting and providing clearer separation between high- and low-success learners.
- Its probability-based insights aligned better with intervention planning, enabling early risk detection and more targeted learner support strategies.

Logistic Regression Model's performance Visuals

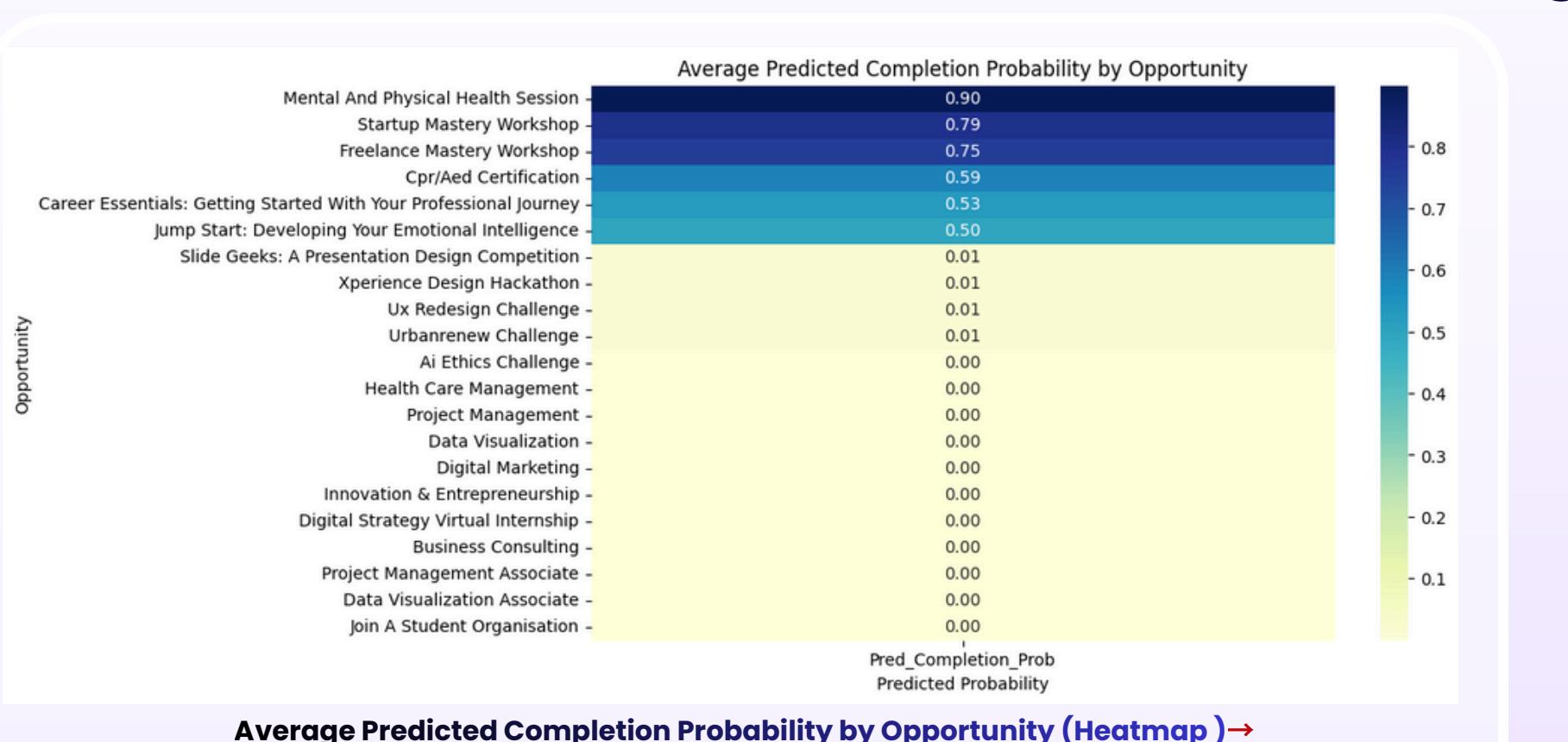


Decision Tree Model's performance Visuals



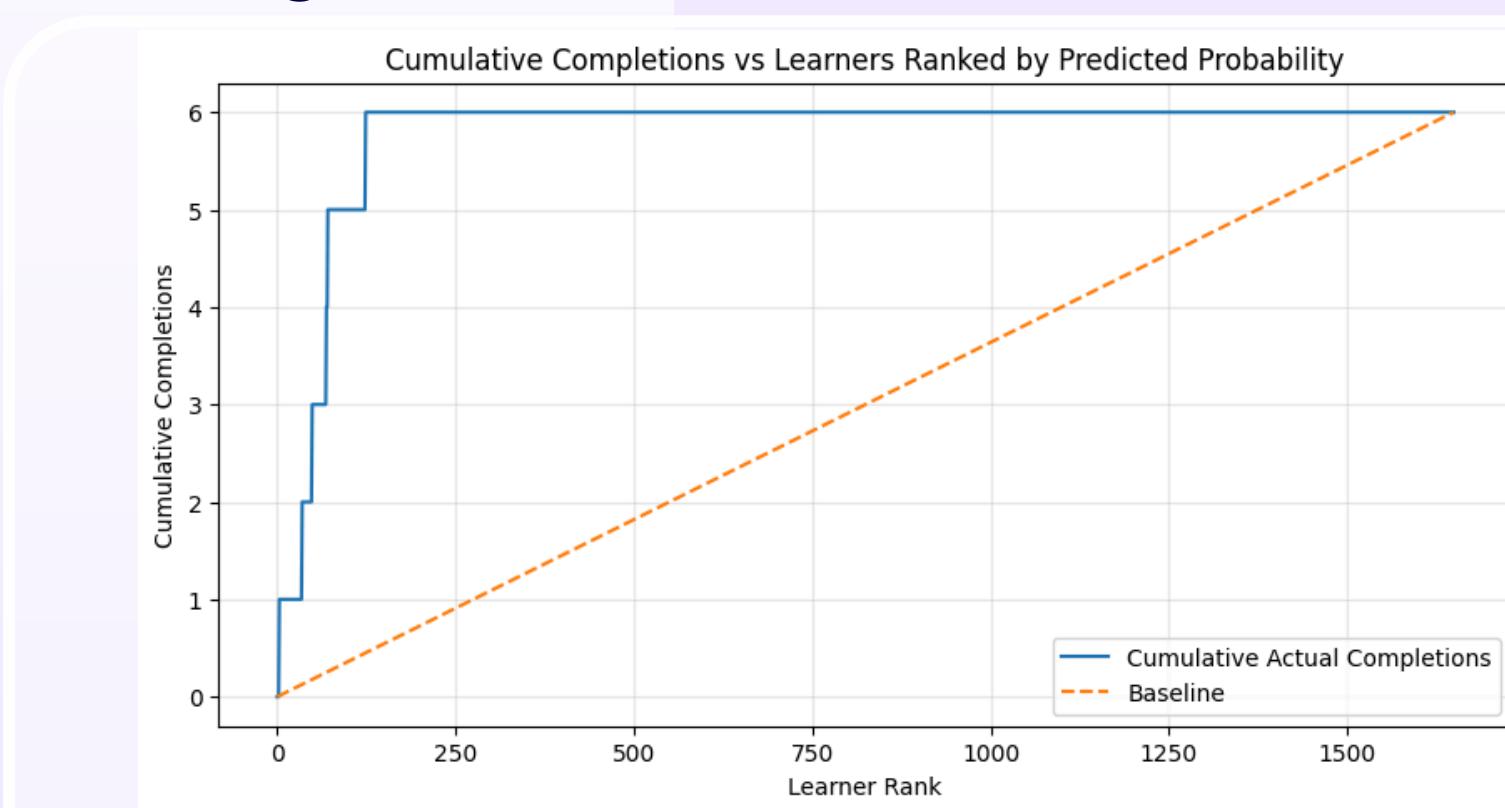
Next >

Predicted Visuals from Logistic Regression Model



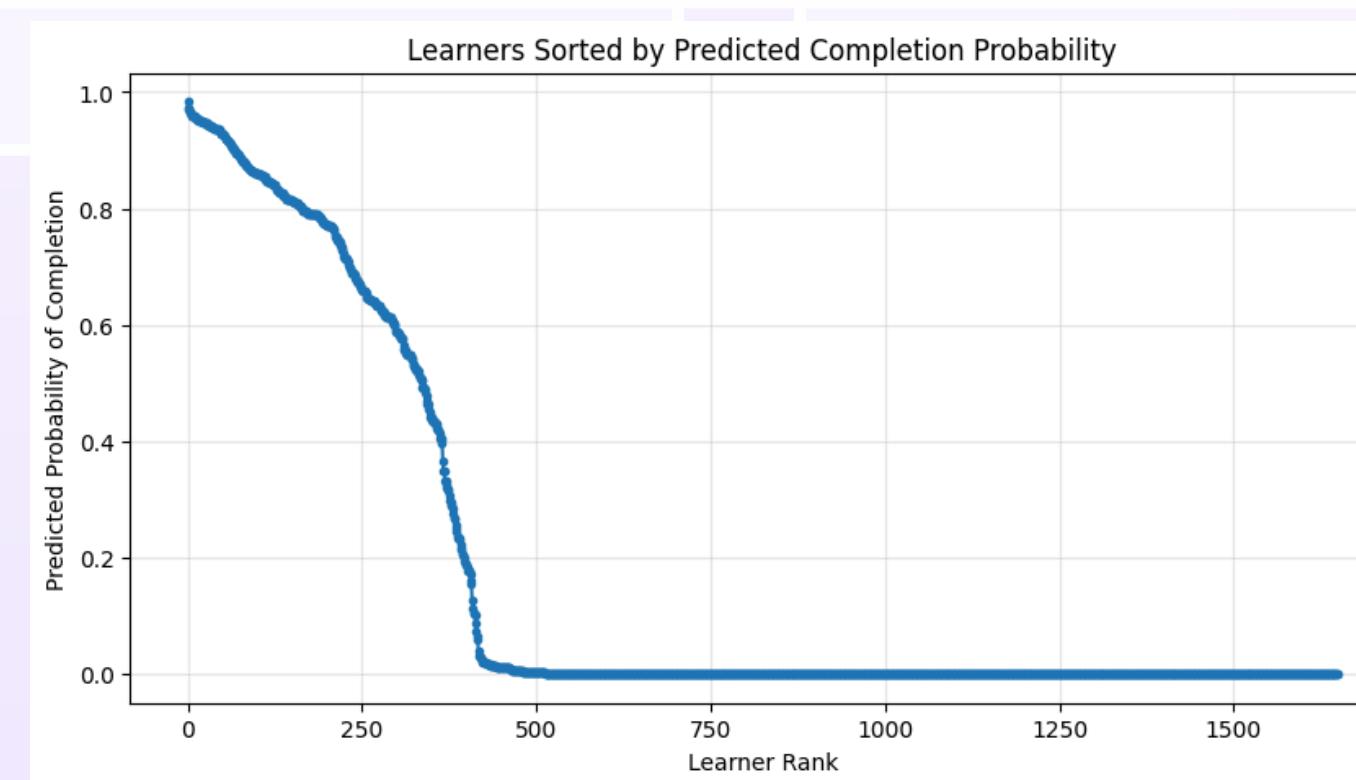
Average Predicted Completion Probability by Opportunity (Heatmap) →

- **Reveals:** Variation in predicted success across opportunity types.
- **Insight:** Certain opportunities consistently show higher completion likelihood, indicating where learners perform best.



Cumulative Completions by Predicted Probability / Lift Plot →

- **Reveals:** How model-ranked learners contribute to overall completions.
- **Insight:** The model effectively prioritizes high-value learners, boosting intervention efficiency.

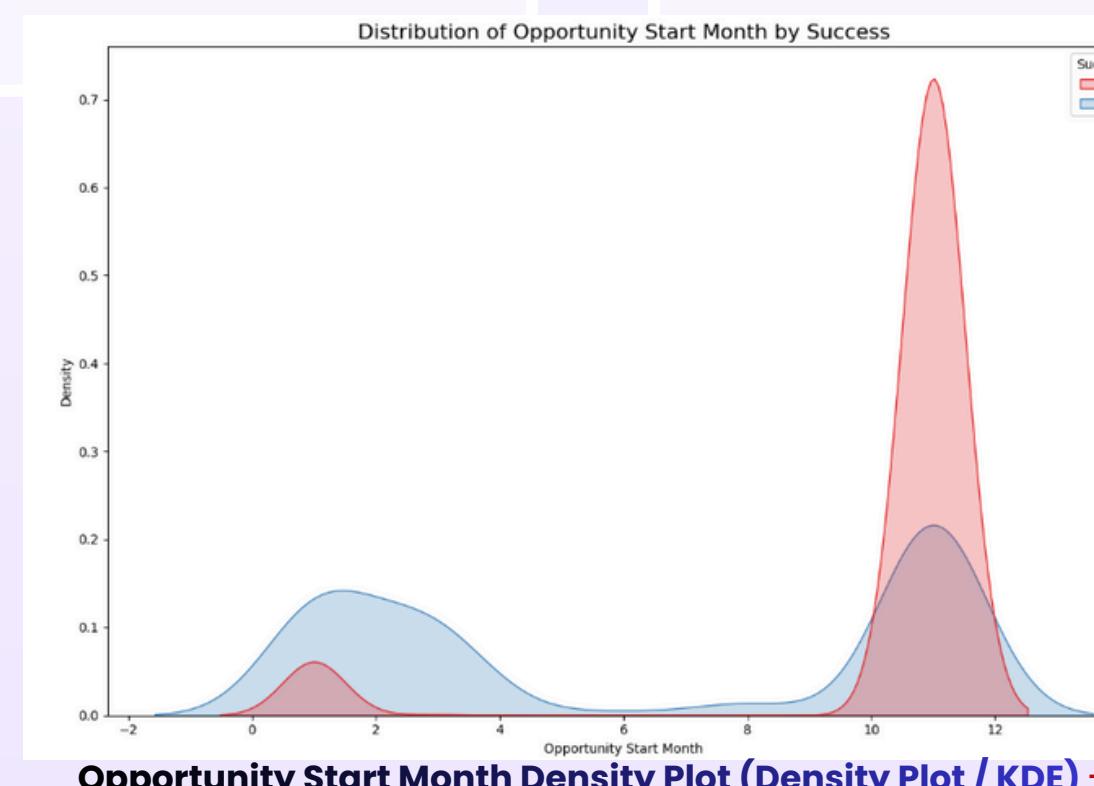
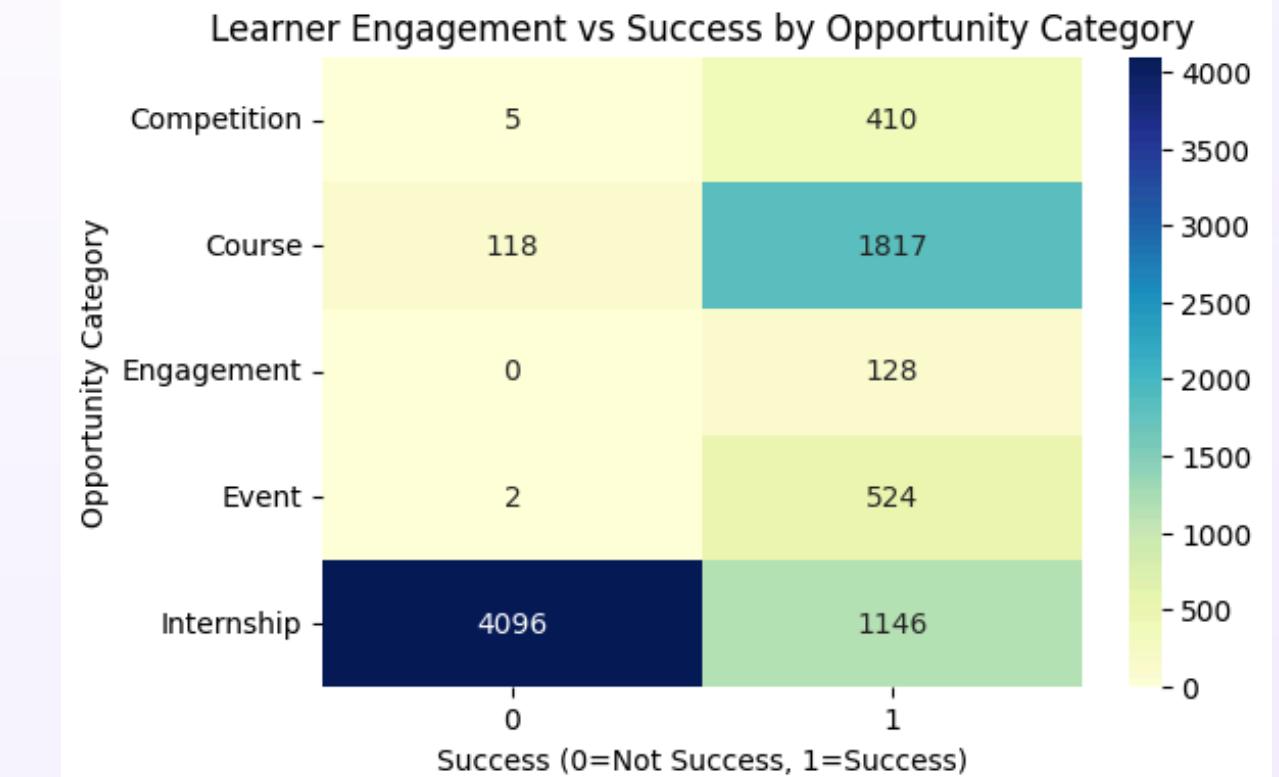
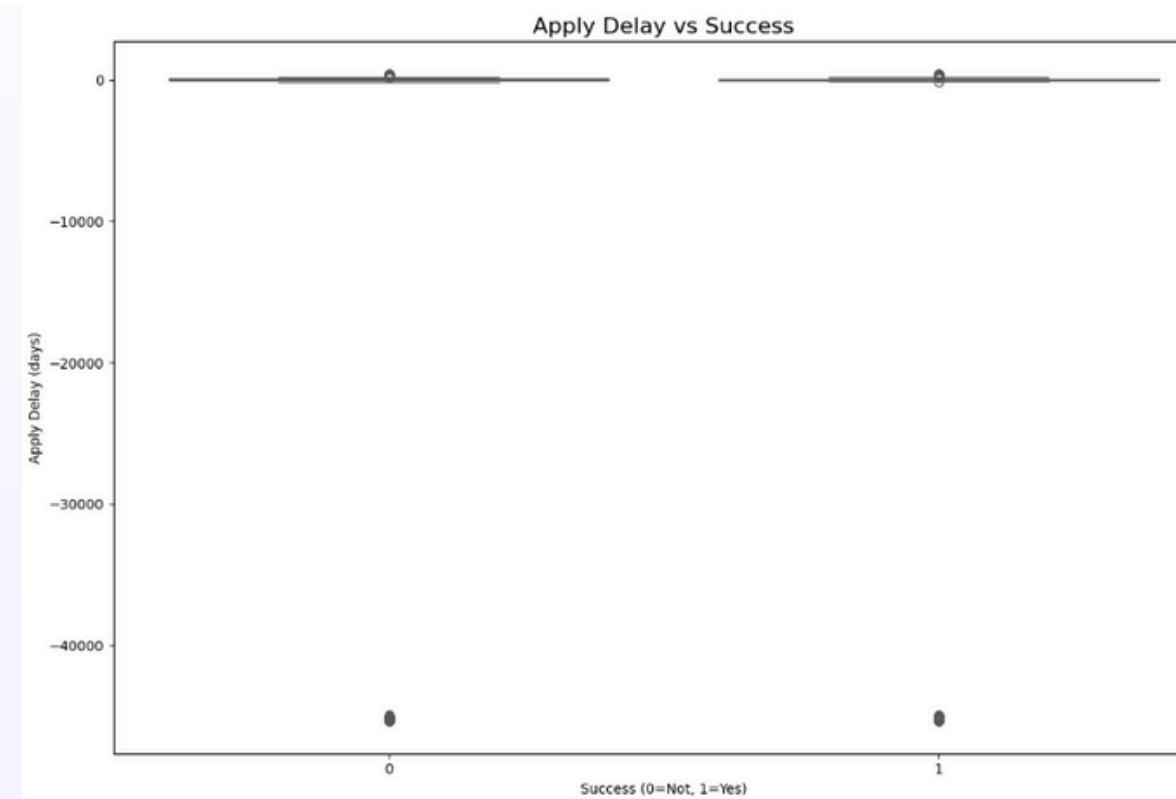


Learners Sorted by Predicted Completion Probability (Scatter Plot) →

- **Reveals:** High-to-low distribution of learner success probability.
- **Insight:** Helps identify top-risk learners early for targeted support.

Next >

Predicted Visuals from Decision Tree Model



Next >

Key Recommendations & Future Direction Suggestions

- **Model signal:** Apply Delay is a strong predictor; longer delays sharply reduce predicted success.
- **Recommendation:** Trigger automated onboarding flows and alerts for learners with low early success probabilities to accelerate "Started" conversion.

- **Model signal:** Predicted success varies widely by Opportunity Category and Opportunity Name.
- **Recommendation:** Scale high-probability programs and redesign or pause low-probability opportunities identified by the models.

- **Model signal:** Success probabilities peak for mid-year sign-ups and early opportunity start months.
- **Recommendation:** Schedule critical programs in high-success months and deploy retention nudges during low-probability periods.

- **Model signal:** Country-level features show uneven success probabilities across regions.
- **Recommendation:** Use predicted risk scores to deliver region-specific support and personalized interventions for high-risk learners.

Onboarding Drop-off & Delayed Engagement

Low Completion & Opportunity Misalignment

Seasonal Motivation & Mid-Journey Friction

Geographic Barriers & Limited Personalized Support

Our Acknowledgement & Conclusion



Key Learnings from the Internship:

- We gained hands-on experience in learner journey analytics and data cleaning.
- We built predictive models (Logistic Regression & Decision Tree) for learner completion and success.
- We learned to visualize insights effectively via dashboards, heatmaps, lift plots, and probability distributions.
- We developed skills in identifying bottlenecks, optimizing learning pathways, and recommending data-driven solutions.

This internship provided an end-to-end experience of analyzing learner behavior, predicting outcomes, and designing actionable interventions to improve engagement and completion. Combining analytics, modeling, and visualization allowed us to uncover key patterns, address bottlenecks, and propose practical solutions for optimizing learner success.

A heartfelt thanks to Excelerate, our team, and our associates for their guidance and support throughout this internship opportunity.

Our Note: "Success is a journey, not a destination for learners. Every milestone brings us closer to greatness."

Thank You