SLU Opportunity Wise Dataset - Summary

Data Overview

- **Scale**: 8,227 clean records from 8,558 original entries
- **Coverage**: 68 countries, 2,864 learners, 22 programs (Jan 2023 - Dec 2024)
- **Success Rate**: 48.1% overall with demographic variations
- **Peak Activity**: August 2023 with 946 signups

Key Features Created (Feature Engineering)

- **Behavioral**: Quick applicant flags, same-day applications, engagement scores
- **Temporal**: Age groups, signup patterns, application delays
- **Advanced**: Age-Major combinations, Country-Gender pairs, competition metrics
- **Engagement Score**: Composite metric (40% success + 25% quick + 20% popularity + 15% same-day)

EDA Key Insights

- **Learner Behavior:** 52% quick applicants, 39% late applicants, 40% same-day applications
- **Demographics:** Male learners dominate both retention and churn; Young adults highest risk group
- **Regional Patterns:** North America & South Asia show highest churn volumes
- **Engagement Impact:** Retained learners have significantly higher engagement scores vs churned
- **Application Patterns:** Churned users average more applications but lower success rates

Strategic Recommendations

Target Demographics

- **Gender**: Increase female participation (41.4% volume but 50.2% success rate)
- **Age**: Focus on teens (70.2% success) while supporting young adults (44% success)
- **Regions**: Scale African success model (60.9%) to other regions
- **Majors**: Diversify beyond Tech dominance (65.6% volume, 41.3% success)

Operational Fixes

- **AI Matching**: Reduce 37.8% rejection rate through better alignment
- **Engagement**: Target 25.5% low-engagement users with gamification
- **Seasonal**: Capitalize on Aug-Sep peaks, address Nov-Mar lulls
- **Retention**: Combat 9.9% dropout with mentorship programs

Limitations & Future Work

- **Current Gaps**: Missing socioeconomic data, motivation factors, European representation (0.3%)
- **Future Plans**: Deep learning models, real-time analytics, A/B testing, social network analysis

Model Results & Business Impact

Best Model: Gradient Boosting Classifier

- **Performance**: 91% accuracy, 0.86 F1-score
- **Top Predictors**:
    - Signup-to-Apply delay (41% importance)
    - Course Duration (36% importance)
- **Churn Detection**: 31.8% churn rate identified through behavioral clustering

Key Success Factors

- **Engagement Level**: Strongest predictor (70.6% success for high vs 0% for low)
- **High-Performing Segments**: Teen-Health learners, African participants
- **Geographic Patterns**: Regional/cultural factors significantly impact outcomes

Overall Conclusion

Business Transformation: The analysis successfully converts raw learner data into actionable intelligence, enabling data-driven decision making across all operational areas.

Predicted Impact: Implementation of recommendations can potentially **increase success rates by 15-20%** through:

- Personalized opportunity matching
- Regional optimization strategies
- Engagement-driven retention programs
- Proactive intervention instead of reactive responses

Strategic Value: The model shifts the organization from reactive to predictive, allowing early identification of at-risk learners and targeted interventions to maximize program effectiveness and learner success.