Online Outcomes

Predicting Success in Virtual Learning

Goal

To implement a model to predict unsatisfactory outcomes and develop recommendations to avoid and/or remedy those outcomes.

Business Understanding

- Pre-pandemic, online education was projected to be a $350 billion industry by 2025.
- Covid-19 greatly accelerated this trend, amplifying opportunities and concerns.
- Online education institutions and students can benefit from improved student outcomes.

Methods


python™


pandas


SQLite


NumPy


scikit
learn

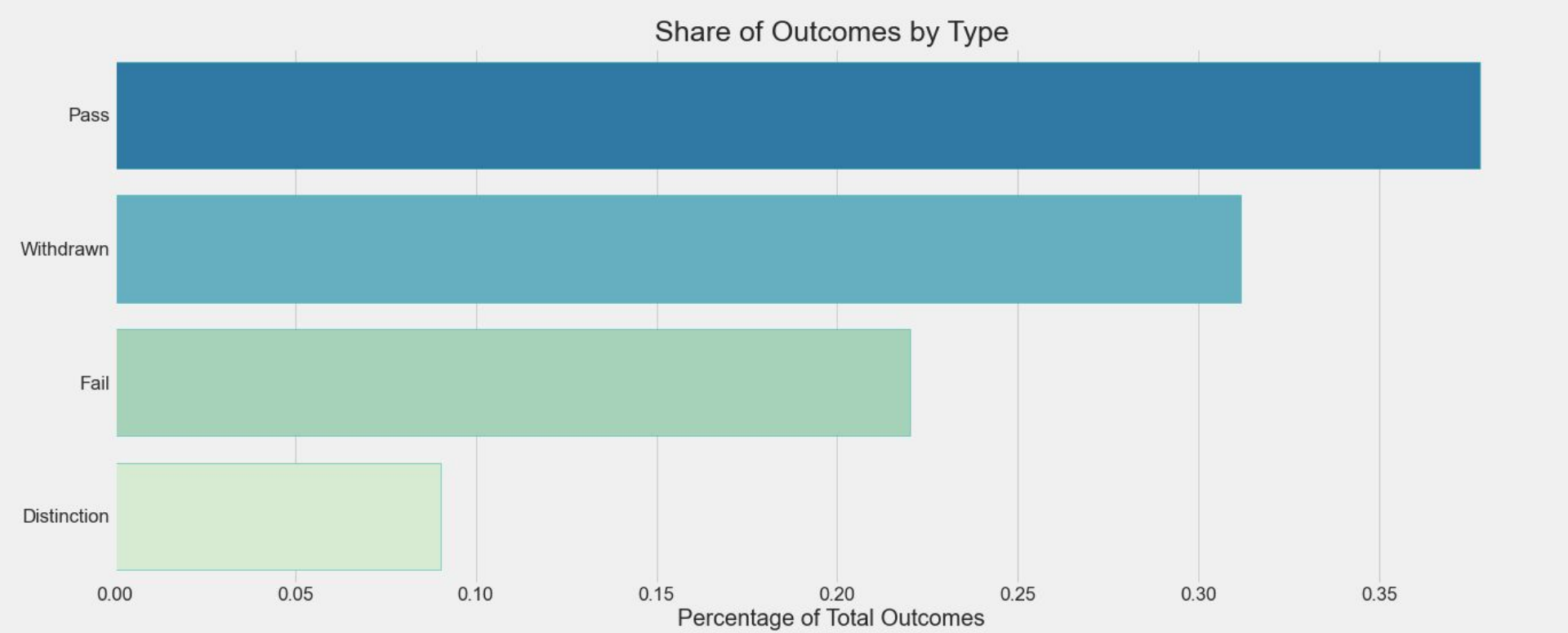

dmlc
XGBoost


imbalanced
learn

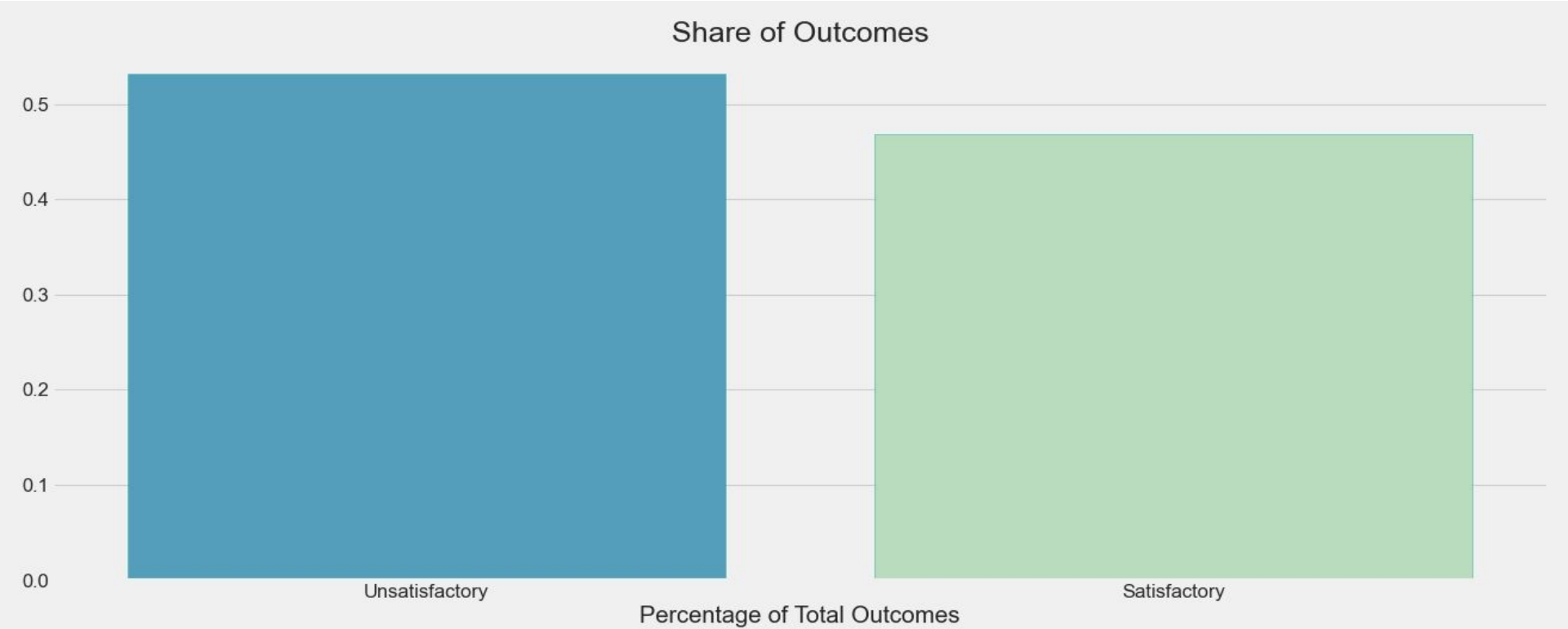
Data

- The Open University Learning Analytics dataset.
- 19,458 rows of tabular data
- 14 columns (13 features plus the target).

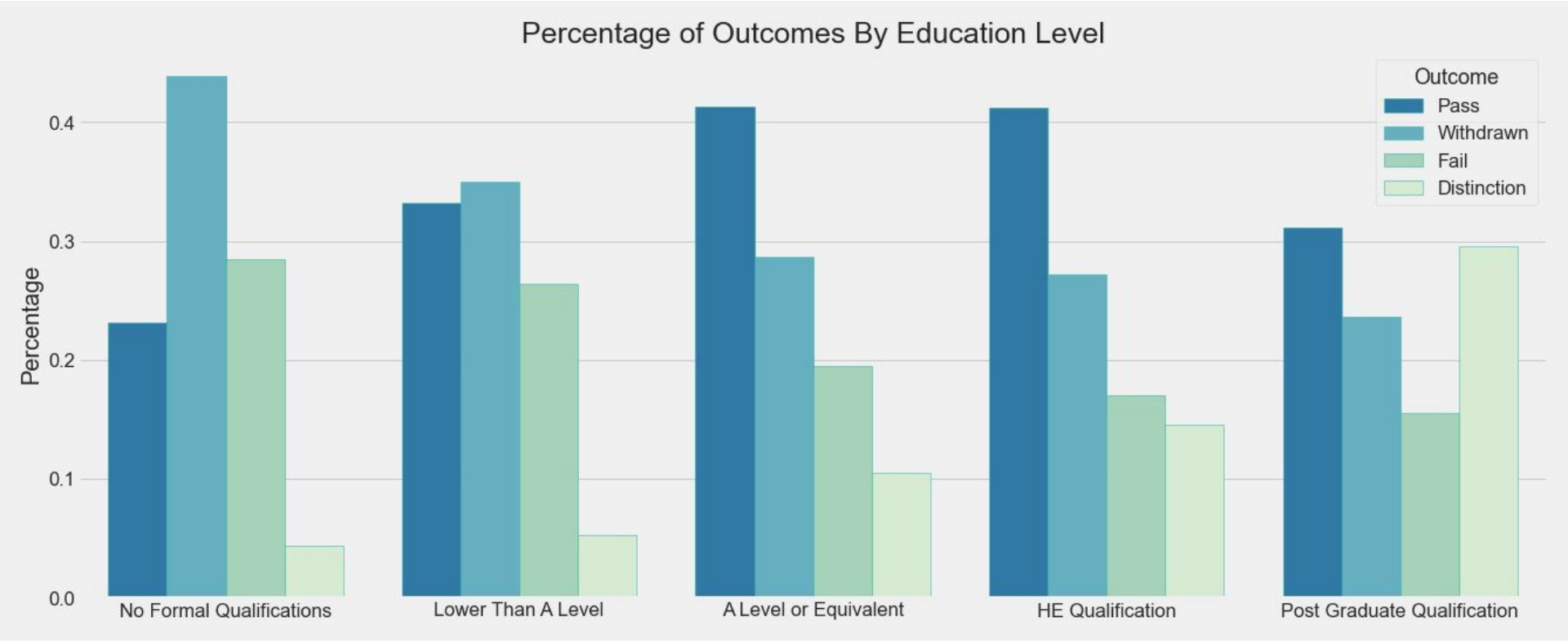
Class Distribution



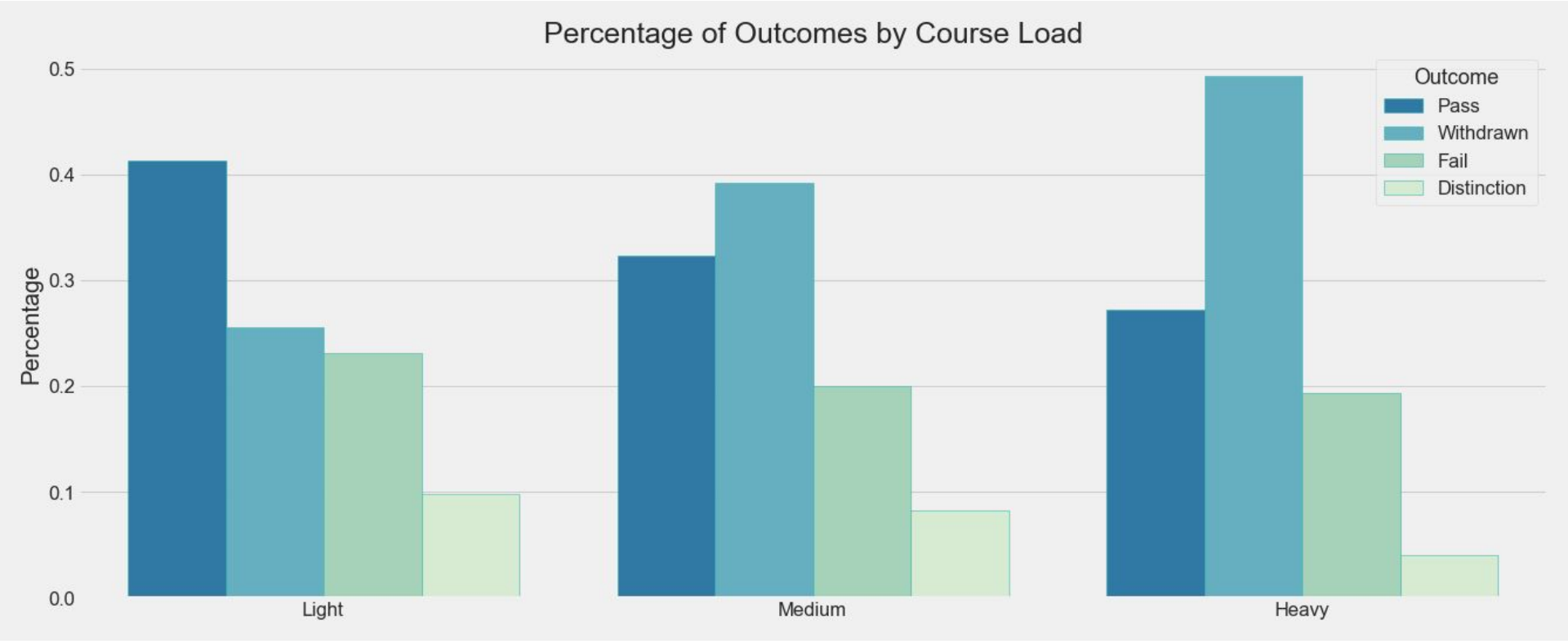
Share of Outcomes



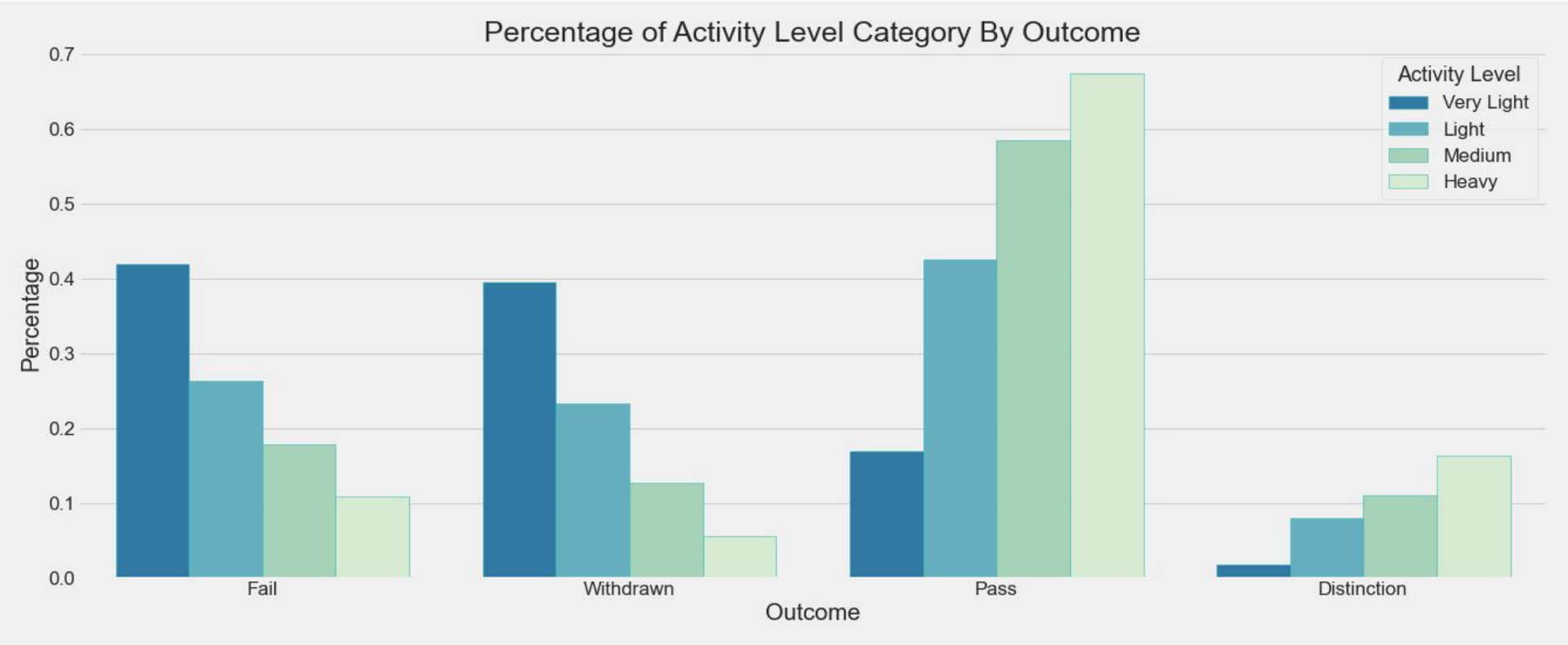
Outcomes by Education Level



Outcomes by Course Load



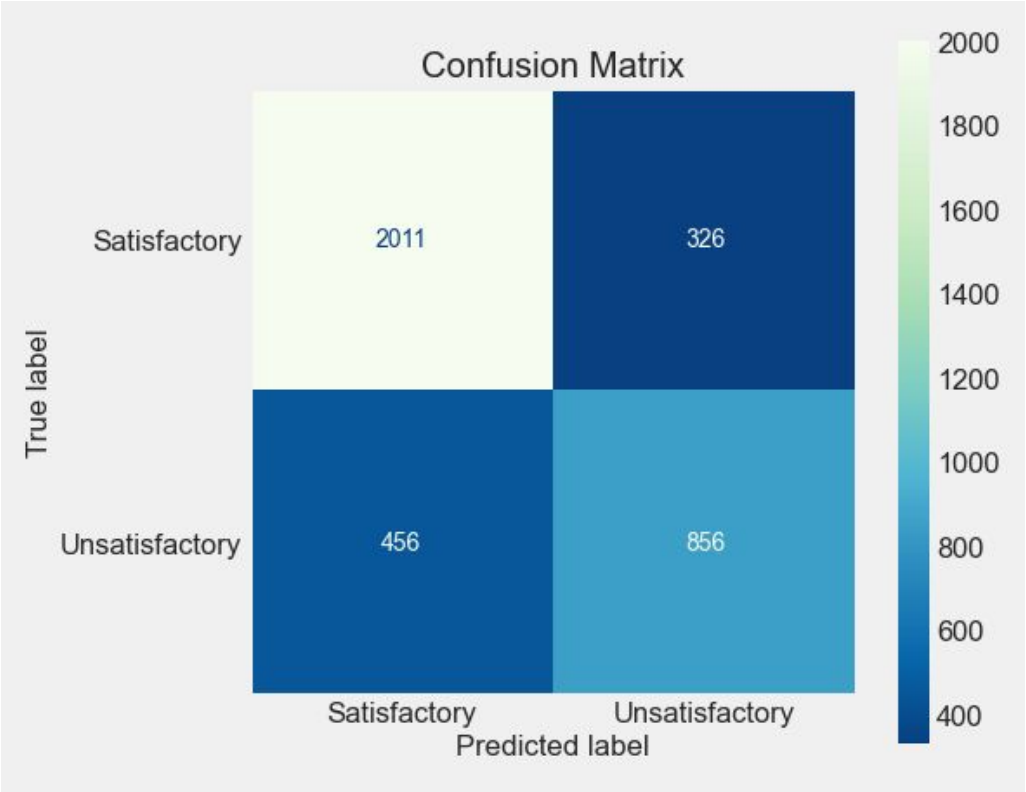
Activity Level & Outcomes



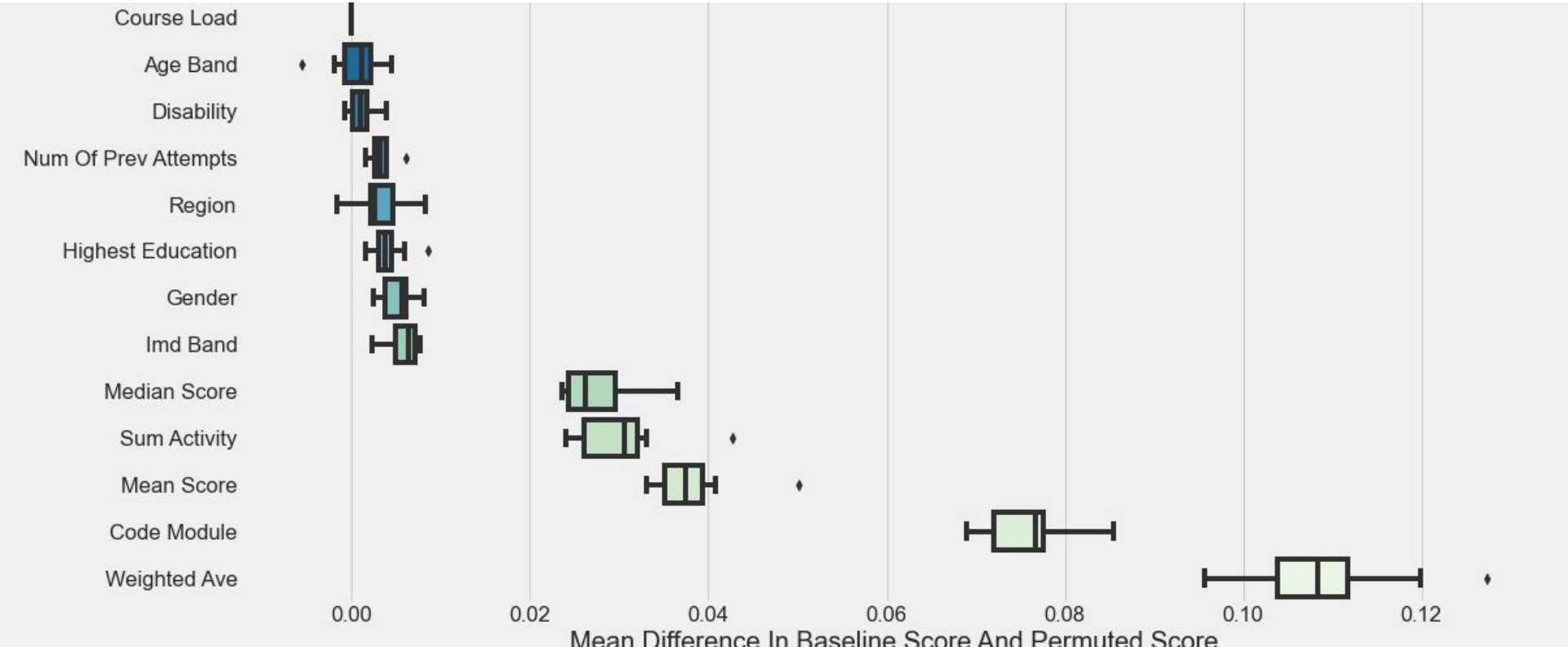
Recommendations

- Implement preemptive outreach and support programs for non-traditional learners.
- Counsel students to maintain modest course loads.
- Advise students to maintain medium or higher online activity levels.

Model: F1=.69 (Recall=.65, Precision=.72)



Permutation Importances



Next Steps

- Developing a multiclass classifier.
- Rescaling the data by individual module.
- Exploring the types of activities that drive outcomes.

Thank You!

Email: jeffrey.h.watson@protonmail.com
GitHub: @jeffreyhwatson