

STAT 385 Final Project

Larry Barrett, Daniel Mroz, Al Pakroshnis

May 5, 2025

1 Question Synthesis

Coming into this project our team wanted to separate analysis into three primary sections; the first two would be

2 Question 1

Q1. How does

2.1 Q1 EDA

3 Question 2

Q1. How does

3.1 Q2 EDA

4 Question 3

4.1 Intro

Which students are most likely to need additional academic support, and what factors best predict this need?

- *Operational Definition:* A student "needs support" if they have at least one past class failure or have received extra school-provided educational support.
- *Goal:* Develop predictive models to classify students who need support, and identify key predictors

Statistical and Machine Learning Solutions:

- Logistic Regression: For interpretable, baseline classification.
- Naive Bayes: For a simple probabilistic approach.
- Random Forest: For robust, non-linear modeling and feature importance.

4.2 Q3 EDA

4.3 Data Preparation Steps:

- Using the vertically merged dataset, created a binary outcome variable `needs_support`, 1 if any failure, 0 otherwise
- Converted categorical variables to factors
- Removed grade columns and identifies for modeling

5 Concluding Remarks