**Objective:**

The project evaluates whether aggregated earnings announcement activity from big (1% or more weight) and small (less than 1% weight) firms can predict the index price direction (up/down) of the S&P 500 index using data available in **yfinance**.

Methodology:

- Stock and Index Data: 20 years
- Earnings: Quarterly
- Companies split by sectors
- Daily closing prices were used to calculate return
- All big companies are used for Logistic Regression
- 20 logistic regression was done based on number of small companies

Key Findings:

Across most sampling designs, the models achieve modest predictive accuracy, often exceeding a random-guess benchmark, indicating that earnings-related information contains limited but nontrivial predictive content for index direction.

Logistic Regression: Index Up/Down vs Big/Small Earnings Activity
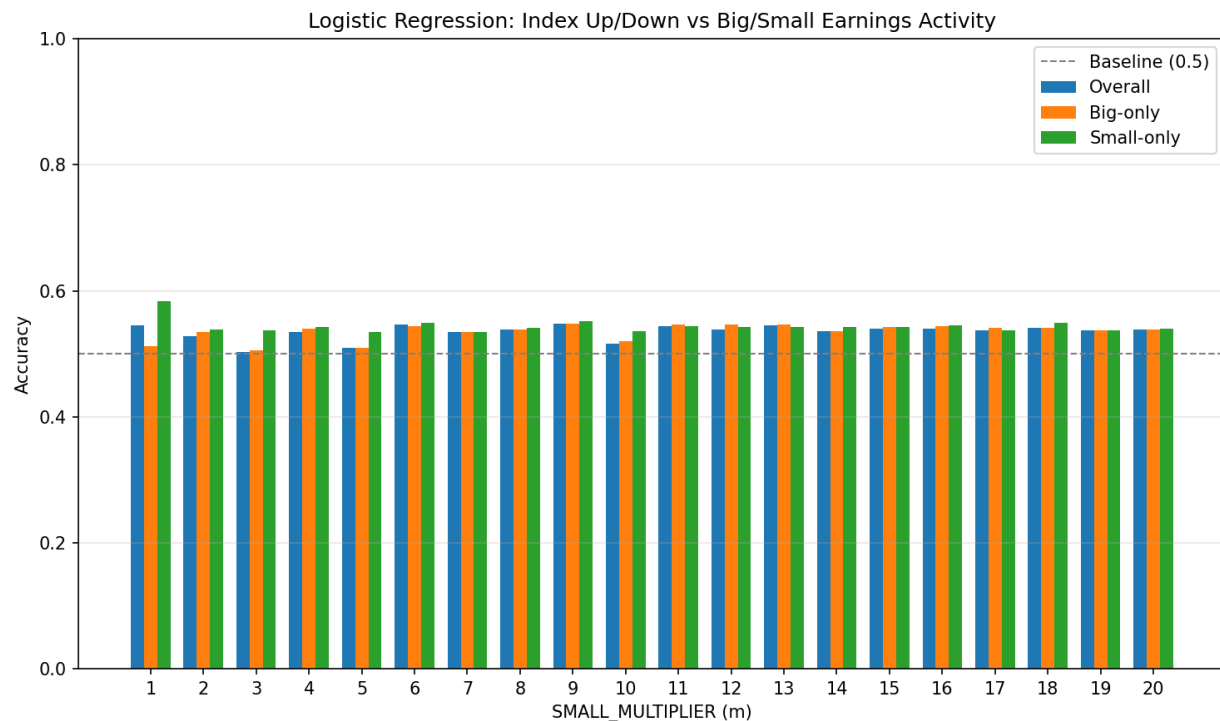
Comparative Insight:

Predictive performance varies across model specifications, highlighting how different compositions of firm-level earnings information affect market-level predictability.

Limitations:

The analysis relies on daily closing prices and simple logistic regression models, which may not fully capture rapid market reactions to earnings announcements or more complex nonlinear relationships.

Scope for Improvement:

Future work could incorporate higher-frequency price data, additional control variables, and more flexible machine-learning models to better capture short-horizon market dynamics and improve predictive performance.