# Preliminary Project Proposal:
# Predicting Stock Market Trends Using Historical Price Data

**Group A**

**This Week's Roles:**

Team Leader: Ross Schanck

Recorder: Auriana Anderson

Spokesperson: Chase Golden

November 3, 2025

## Background and Research Question

Thes stock marked is heavily influenced by many factors like a company's performance, investor behavior and other economic conditions. Even though the market is unpredictable, traders rely on trends to make decisions on investments. We want to determine whether machine learning can identify short-term patterns to help predict if stock prices will increase or decrease the next day based on recent trends.

### Research Question

Can we predict next-day up and down movement of S&P 500 stocks using historical price and trade volume data?

### Why does this question matter:

Even small improvements in predictability could help analysts and investors make better decisions. These small improvements can lead to more profitable gains over time. For us as data scientists, this project helps to combine real-world financial data with machine learning methods.

### Novelty:

Predicting stock market trends is a common challenge, but our focus is on building a simple and clear model using technical indicators. We are not trying to beat the stock market but instead we are trying to use machine learning to uncover trends that may not otherwise be detected in stock data.

## Hypothesis and Prediction

### Hypothesis:

If recent stock trends and trading volumes show consistent patterns, then simple classification models like logistic regression or random forest can predict next-day stock movement more accurately than random guessing.

### Prediction:

We expect that our model will correctly predict stock direction 55-60% of the time. This is slightly better than chance but will prove that short-term patterns contain usable information. Our hope is that more historic data will improve the model's overall accuracy.

## Data and Analysis

### Data Sources:

We plan to use publicly available stock market data sets. In beginning to explore available data we found several sources that may prove useful.

- S&P 500 Historical Data (1927 – 2022)
    - https://www.kaggle.com/datasets/henryhan117/sp-500-historical-data
- S&P 500 Stocks (daily updated)
    - https://www.kaggle.com/datasets/andrewmvd/sp-500-stocks
- The yfinance library allows you to download daily stock prices directly in python
    - https://pypi.org/project/yfinance/

**Response Variable (Target):**

Whether a stock's closing price goes up (1) or down (0) the following day.

**Predictor Variables (Features):**

- Previous day's closing price
- Stock's 5- and 10-day movement averages
- Daily trading volume
- Stock's precent of change from prior day

**Analysis Plan:**

1. Download and clean the data using Python
2. Create new features such as movement averages and percent changes
3. Split the data into training and testing sets – 80/20 split
4. Train simple models like logistic regression and random forest to predict next-day direction
5. Evaluate performance using accuracy and F1-score
6. Discuss results and if the patterns found could generalize

**Potential Pitfalls:**

Stock prices are noisy and often influenced by unpredictable factors like news or other world events. This is not reflected in the stock data. It is also possible that our model will not outperform random guessing. We will handle this by clearly separating training and testing data and reporting results honestly.

**How we will know if the question is answered:**
If our models perform better than random guessing, above 50% accuracy, and show consistent patterns, then we will consider our hypothesis supported.

## Technical Details

- **Language:** Python
- **Libraries:** pandas, scikit-learn, matplotlib, yfinance

- **Resources needed:** Kaggle datasets
- **GitHub Repository:** https://github.com/aurianaanderson/StockMarketTrends

## Summary:

This project will allow our team to apply data cleaning, feature engineering, and machine learning to a real-world dataset while exploring a popular challenge in data science, predicting stock trends. Even if our model's accuracy is modest, the process will deepen our understanding of time series data, model evaluation, and the limitations of predictions in complex systems like financial markets.