**Team Lambda**

**Preliminary Project Proposal**

**Predicting Investor Stock Purchases**

Team Lead:

Recorder:

Spokesperson:

* **Question**:
  + Are investors like Warren Buffett likely to buy specific stocks next based on historical data and known investment behaviors?
* **Need**:
  + This question addresses the need for predictive investment strategies by utilizing past data to forecast future stock buys made by successful investors. By examining the traits of companies that Warren Buffett usually invests in, we can create models that help spot other stocks that fit his investment approach.
  + This is crucial because renowned investors like Warren Buffett have a proven history, and studying their actions can assist others in enhancing their investment strategies, particularly in value investing.
  + This question holds significance for individual investors, financial analysts, and anyone keen on predictive analytics in the stock market.
* **Why Worth the Time/Effort:**
  + Knowing which stocks Buffett might buy next can help investors make smart choices that follow his successful methods. This can result in better investment returns by using proven value investing strategies.
  + This topic is interesting to investigate because it connects classic investment techniques with new data science, offering useful information for picking stocks.
* **Novelty / Originality**:
  + Unlike traditional stock price predictions, this approach models the decision-making process of influential investors like Warren Buffett.
* **Stakeholder**:
  + **Individual Investors**
  + **Financial Analysts and Advisors**
  + **Academic Researchers**

**Hypothesis and Prediction**

* **Hypothesis**: Historical financial metrics, such as P/E ratio, debt-to-equity ratio, and revenue growth, are good predictors of which stocks Warren Buffett or similar value investors are likely to purchase next.
* **Prediction**: Stocks that show solid financial strength (such as low debt, high dividend yield, and strong revenue growth) and are undervalued (like those with a low P/E ratio) are more likely to be selected by investors such as Warren Buffett in the future.

**Data & Analysis**

* **Data Sets**:
* **Response/Outcome Variable**:
  + The binary variable indicating whether Warren Buffett (or a similar investor) buys a stock in the next quarter or year. It will be 1 if the investor buys the stock and 0 if they do not.
* **Predictor Variables**:
  + Quarter Year
  + Ticker – Name
  + Activity
  + Shares added
  + % change to port
  + P/E Ratio (Price-to-Earnings Ratio)
  + Debt-to-Equity Ratio
  + Dividend Yield
  + Revenue Growth
  + Gross Margin
* **Tentative Analysis Plan**:
  + **Data Preprocessing**:
  + **Exploratory Data Analysis (EDA)**
  + **Feature Selection**
  + **Modeling**
  + **Evaluation** using accuracy, precision, recall, F1-score, and AUC.
* **Pitfalls**:
  + **Data Availability**: Not all of Buffett's investments are publicly available, and some data might be outdated or incomplete.
  + **Feature Selection**: Identifying the correct features that consistently predict Buffett’s investment behavior could be challenging, especially if he follows qualitative factors (e.g., company management) that are harder to quantify.
* **How to Know if the Question is Answered**:
  + If the model can predict with reasonable accuracy whether Buffett is likely to buy a given stock, then the research question is answered.
* **Code:**
  + Python or R?
  + Any other resources, SQL? We used DATAROMA for data pull?
  + Link to GitHub repo: