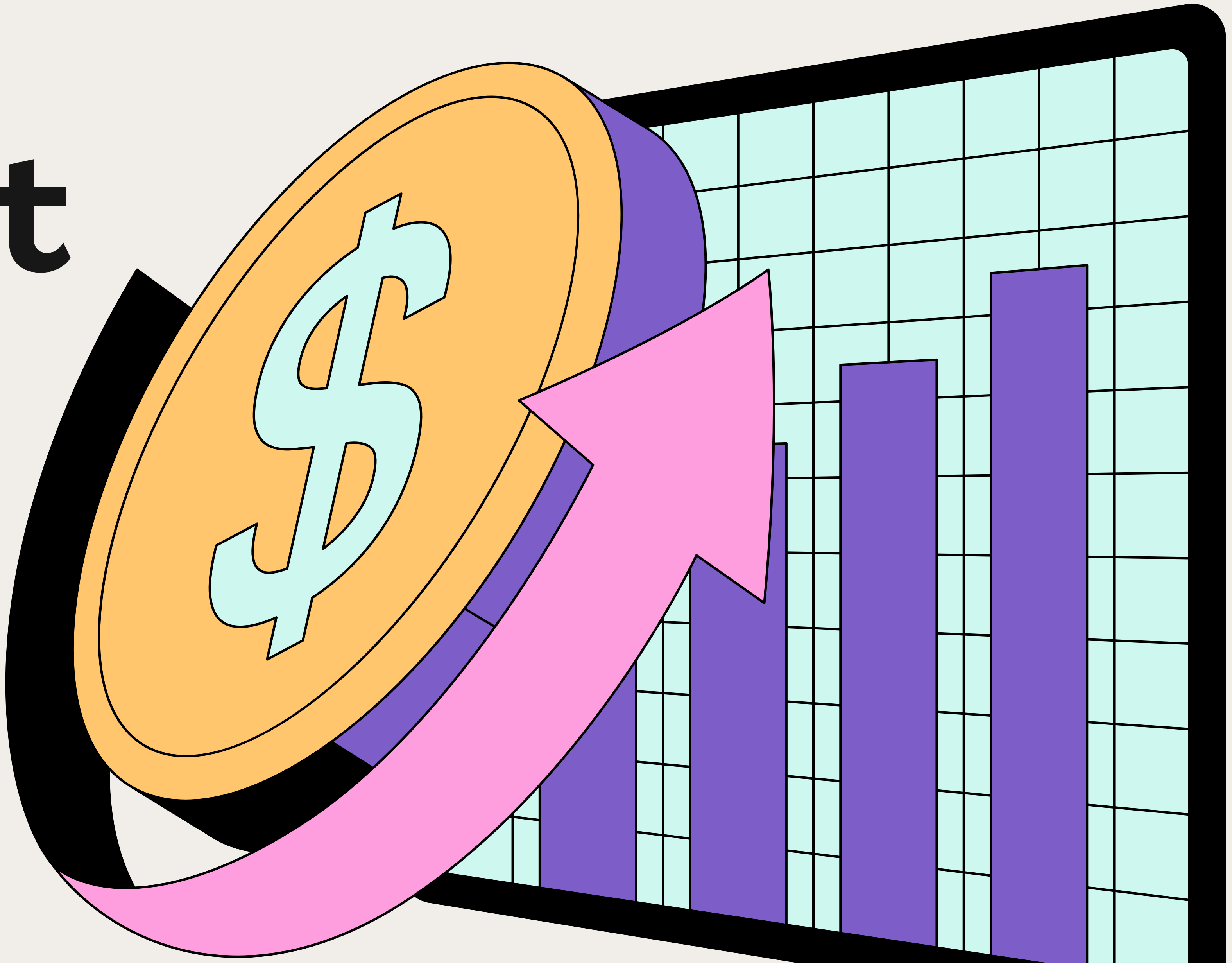
TOMORROW'S PRICE SIGNALS HIDDEN IN TODAY'S LANGUAGE

Predicting Overnight Stock Changes Post-Earnings



ALINA HOTA | PRANIT YADAV | JUSTIN YANG | CONNOR THERRIEN | JOSHUA RINGLER

DECEMBER 1ST , 2025

Problem Statement

Stock Prices used to move based on financial results - that is no longer the case. How do earnings transcripts - specifically forward looking impact the return of a stock?

We turn each transcript into NLP features and use them to forecast:

- Direction of the overnight move
- Which calls may trigger the biggest shifts

Why it matters: Faster signals around earnings help improve risk management, hedging, and analyst prioritization.

Our Dataset

Wharton Research Data Services (WRDS)

- Capital IQ
  - Earnings Calls and Metadata
- Compustat
  - Gather Fundamental Company Data
- CRSP (Center for Research in Security Prices)
  - Market Data (Daily Returns, Mkt Cap, FF12 Industry)
- IBES (Institutional Brokers' Estimate System)
  - Prior Revenue Guidance

wrds WHARTON RESEARCH DATA SERVICES

The Global Standard for Business Research

What are We Predicting

Predicting: **close_to_open_return**

- (Open Price - Close Price) / (Close Price)
- Has both Direction, and Magnitude!

Key Features

- Guidance Surprise %
- NLP
  - Forward Outlook
- Context
  - Industry
  - Market Cap

Features

Guidance Surprise %

- Measure of Expected Revenue-Reported Revenue
- Pull the Expected Revenue from IBES Database
- Reported Revenue from Compustat

Fama-French 12 Industry Classification

- Groups companies into industry Sectors
    - Utilities, Healthcare, Finance, etc.
- One Hot Encoded

Market Cap

- The Market Cap of the company at Earnings Date
- Could size of company impact the guidance surprise?

Word Count

- Length of transcript impact returns?

Ensuring Accuracy

Correct Linkage

- Mapped Company ID (CIQ) → GVKEY (Compustat) → PERMNO (CRSP)

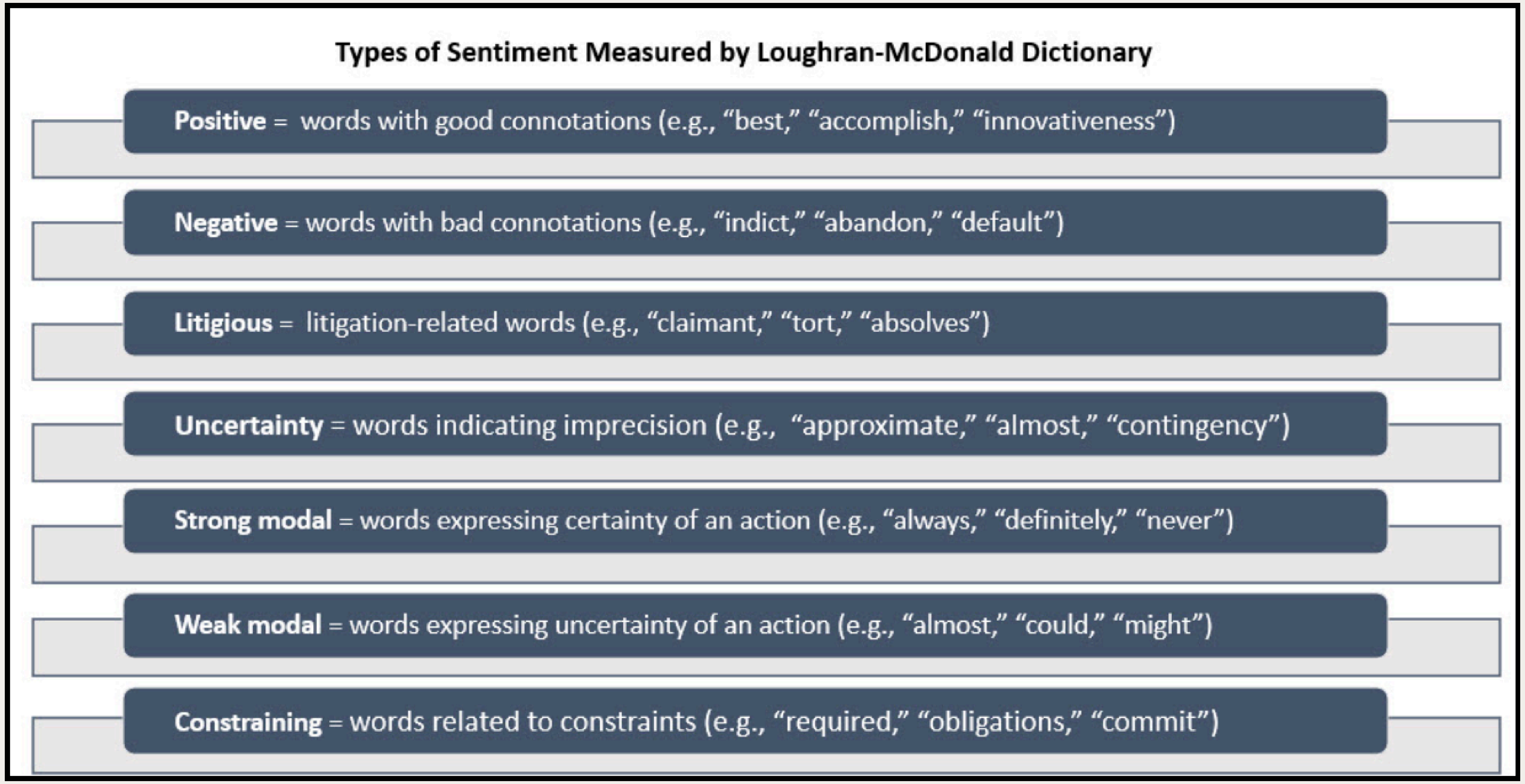
Time-Series

- Ensure that closing price and opening price are from correct dates
- Ensure that reporting was done at night

Data Leakage

- Initial NLP Created Data Leaks
- Had to change how we added certain NLP values
    - Strict 70-30 Split in certain Feature Engineering

Loughran-McDonald Dictionary



Citation: WRDS (Wharton Research Data Services

Why did we use this?

- Finance-specific sentiment dictionary that avoids misclassifying common financial terms.
- Converts earnings-call language into meaningful, structured features for prediction.
- Captures risk, uncertainty, and managerial confidence that often move markets.
- Adds interpretability and reduces noise compared to generic NLP sentiment tools.

Feature Engineering

1. Financial Sentiment Analysis

- lm_positive_count / lm_negative_count: Loughran-McDonald dictionary
- net_sentiment: Balance of positive vs negative financial terms
- sentiment_polarity: Intensity of sentiment direction

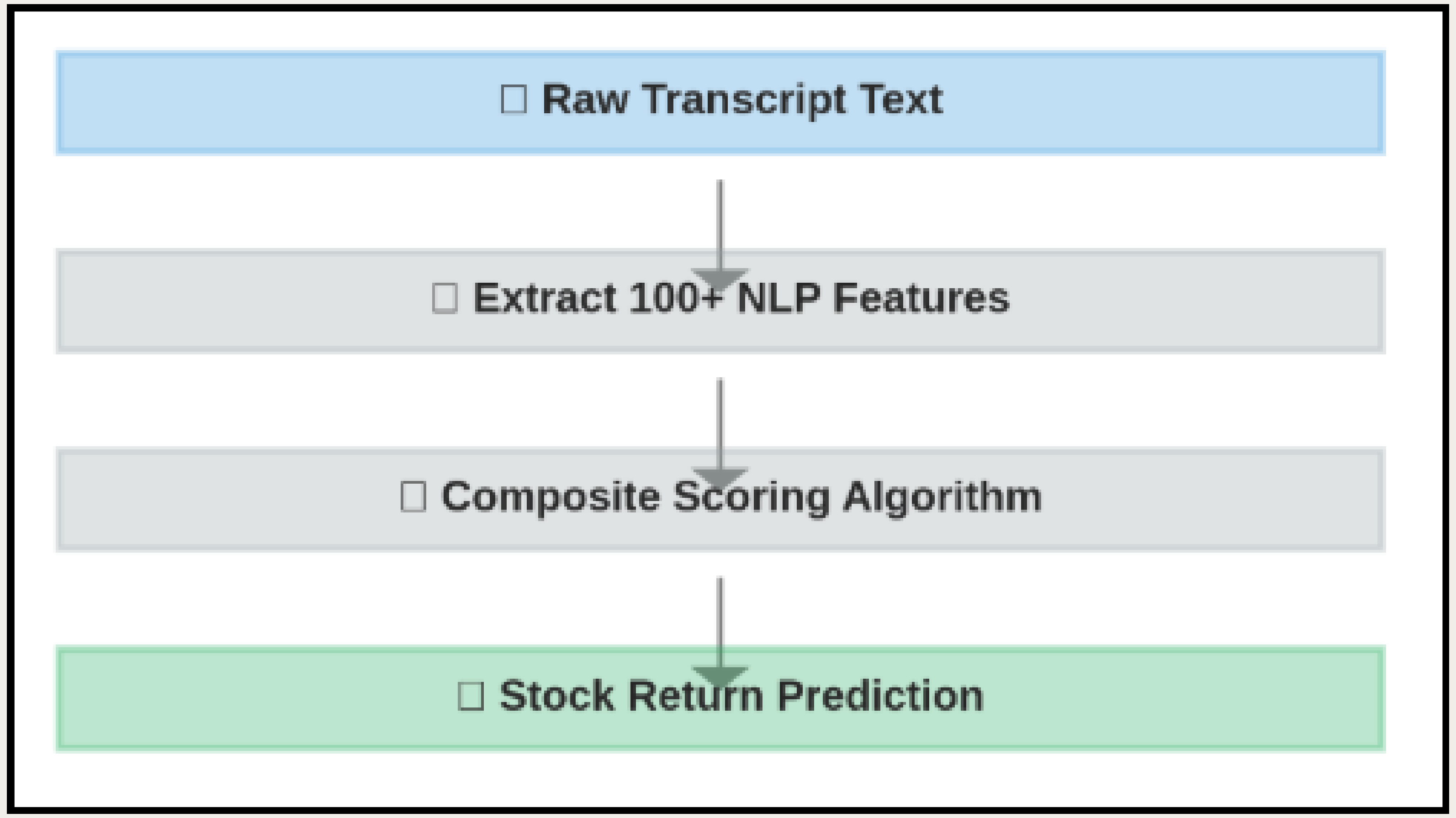
Why It Matters: Financial language differs from general sentiment, "liability" is negative in finance but neutral in general English. The Loughran-McDonald dictionary captures finance specific connotations that predict investor reactions.

2. Management Tone Analysis

- emotional_word_ratio: Emotional vs factual language
- uncertainty_ratio: Hedging and doubt indicators
- evasion_score: Defensive language patterns
- modal_verb_ratio: "Could", "might", "should" usage

Why It Matters: CEOs who use more certain language ("will" vs "might") and less defensive phrasing see better stock reactions. Emotional stability and confidence in delivery predict positive returns.

NLP Pipeline:

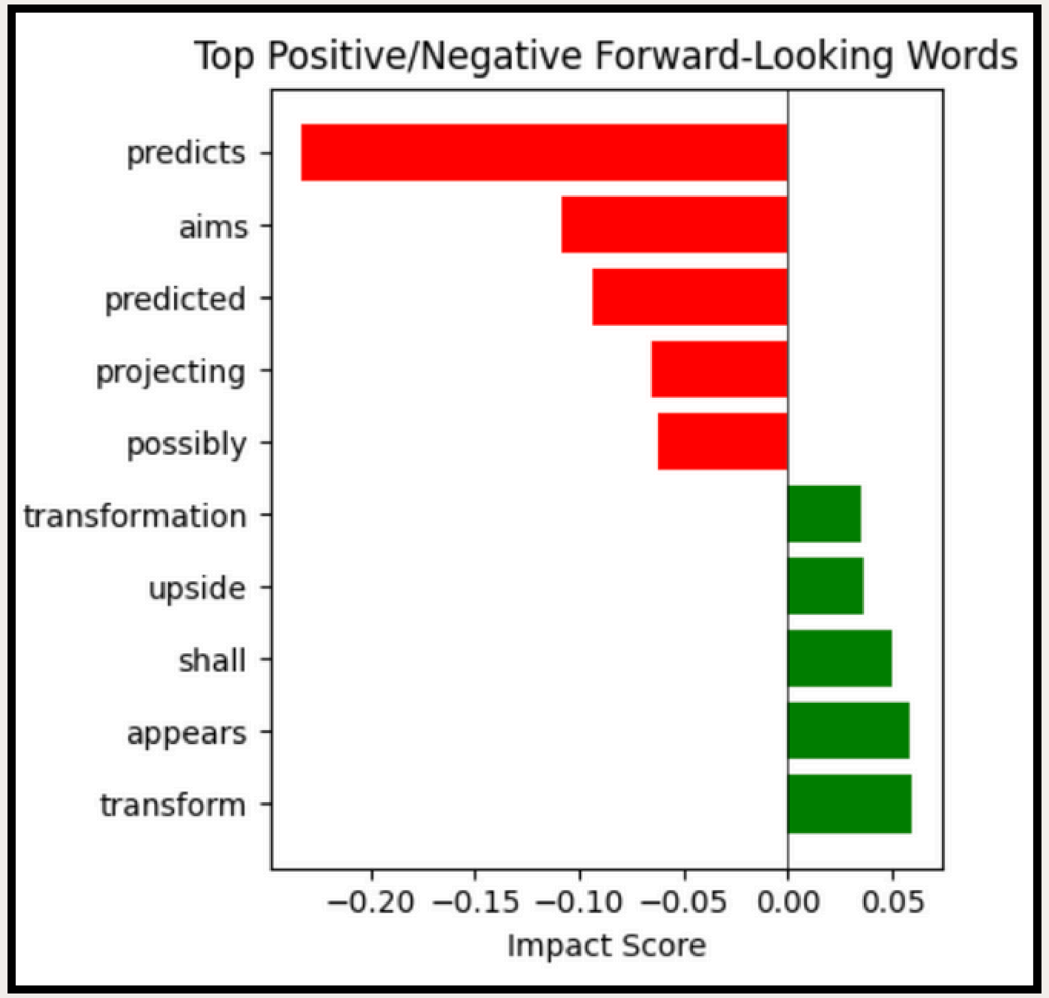


Foward-Looking Language

Features Created:

- *forward_intensity* - Density of future-focused discussion (Count)
- *forward_confidence_score* - Optimistic vs cautious future language (Ratio)
- *forward_certainty_score* - Definiteness of forward statements
- ****forward_outlook_score**** - Composite predictive signal

Why It Matters: We found that how management discusses the future is more predictive than current results.

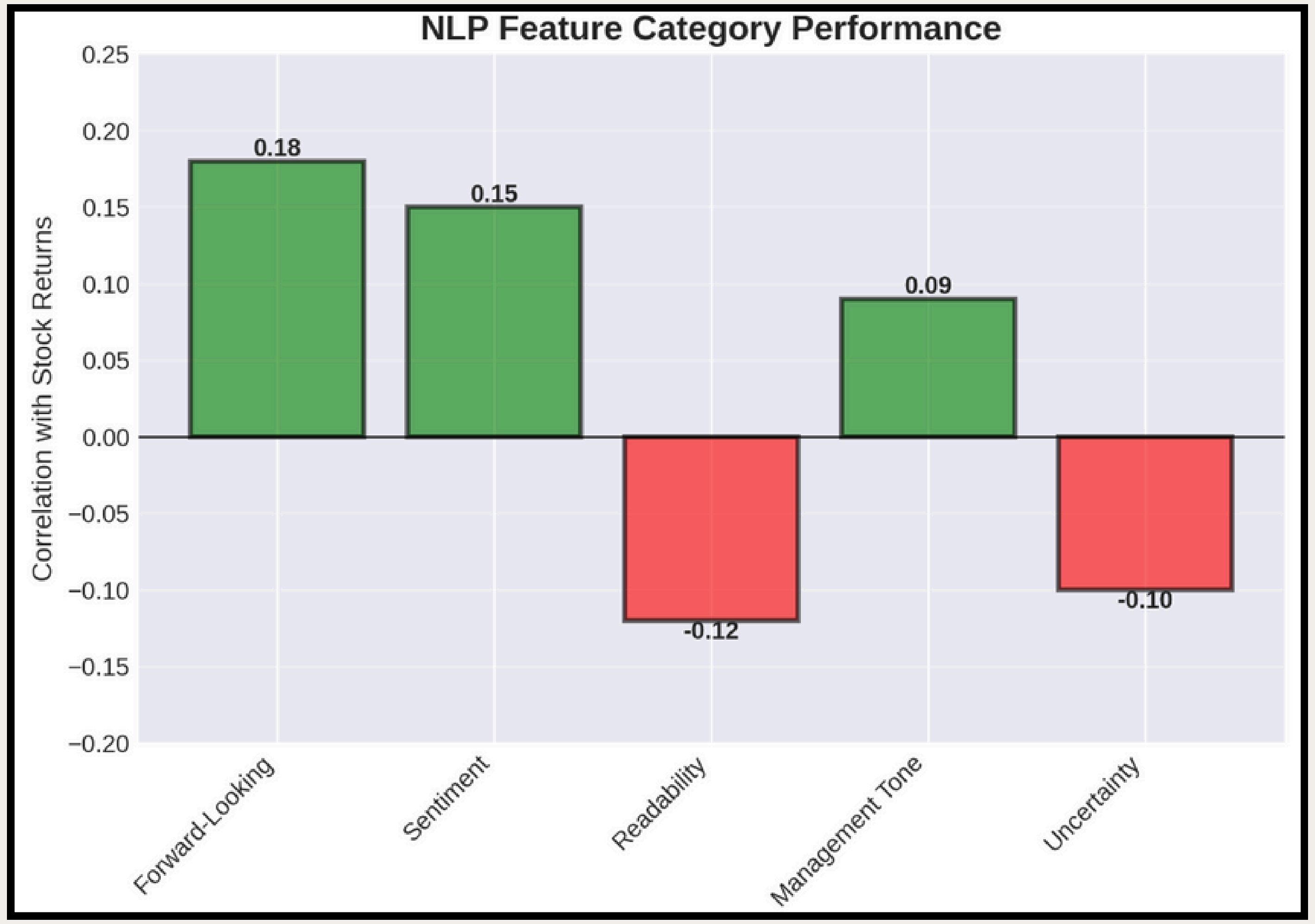


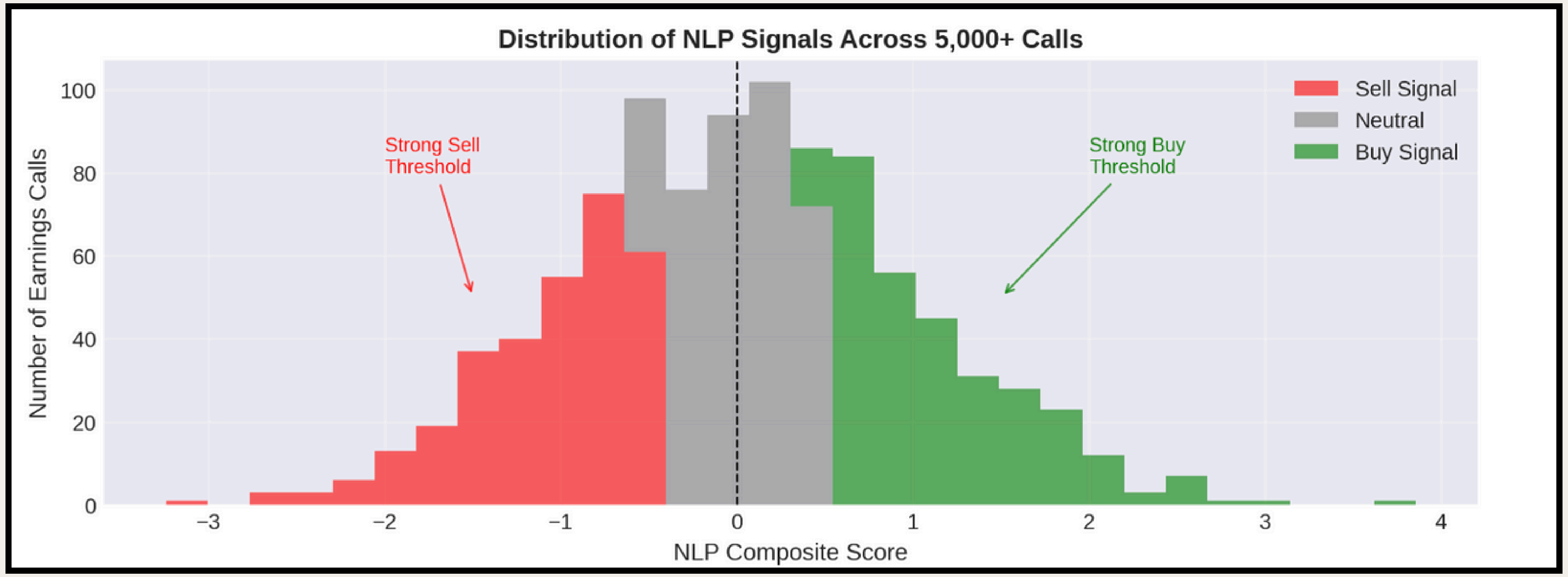
***forward_outlook_score* =**
0.5 × Intensity Score (specific term analysis)
+ 0.3 × Confidence Score (positive vs negative future tone)
+ 0.2 × Certainty Score (definiteness of statements)

NLP Key Findings

Validation & Results:

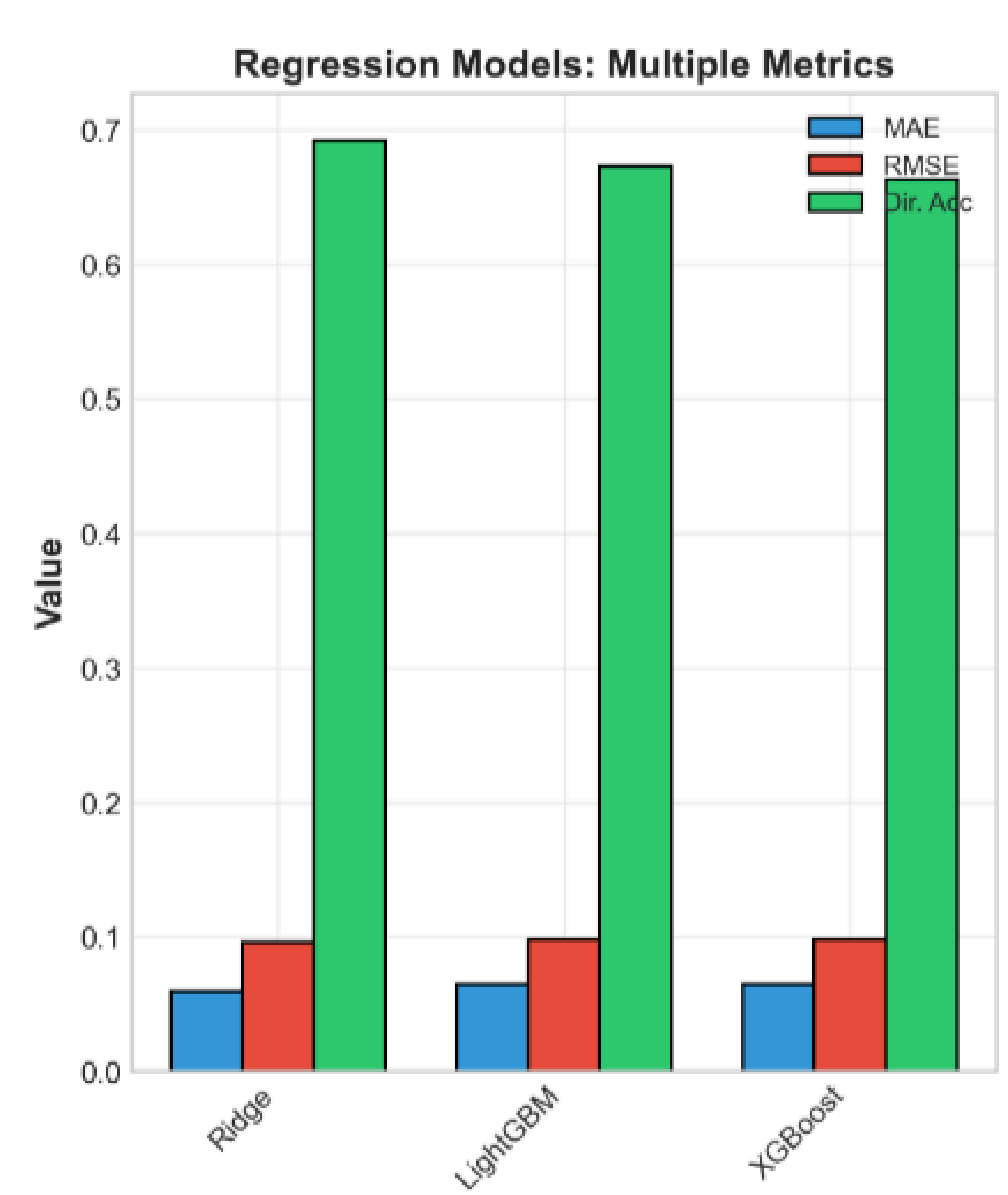
- Forward-looking confidence shows strongest correlation with overnight returns (r = 0.15–0.20)
- Readability complexity negatively correlates with returns (r = −0. 12)
- Composite NLP score improves prediction accuracy by ~7% over financial metrics alone

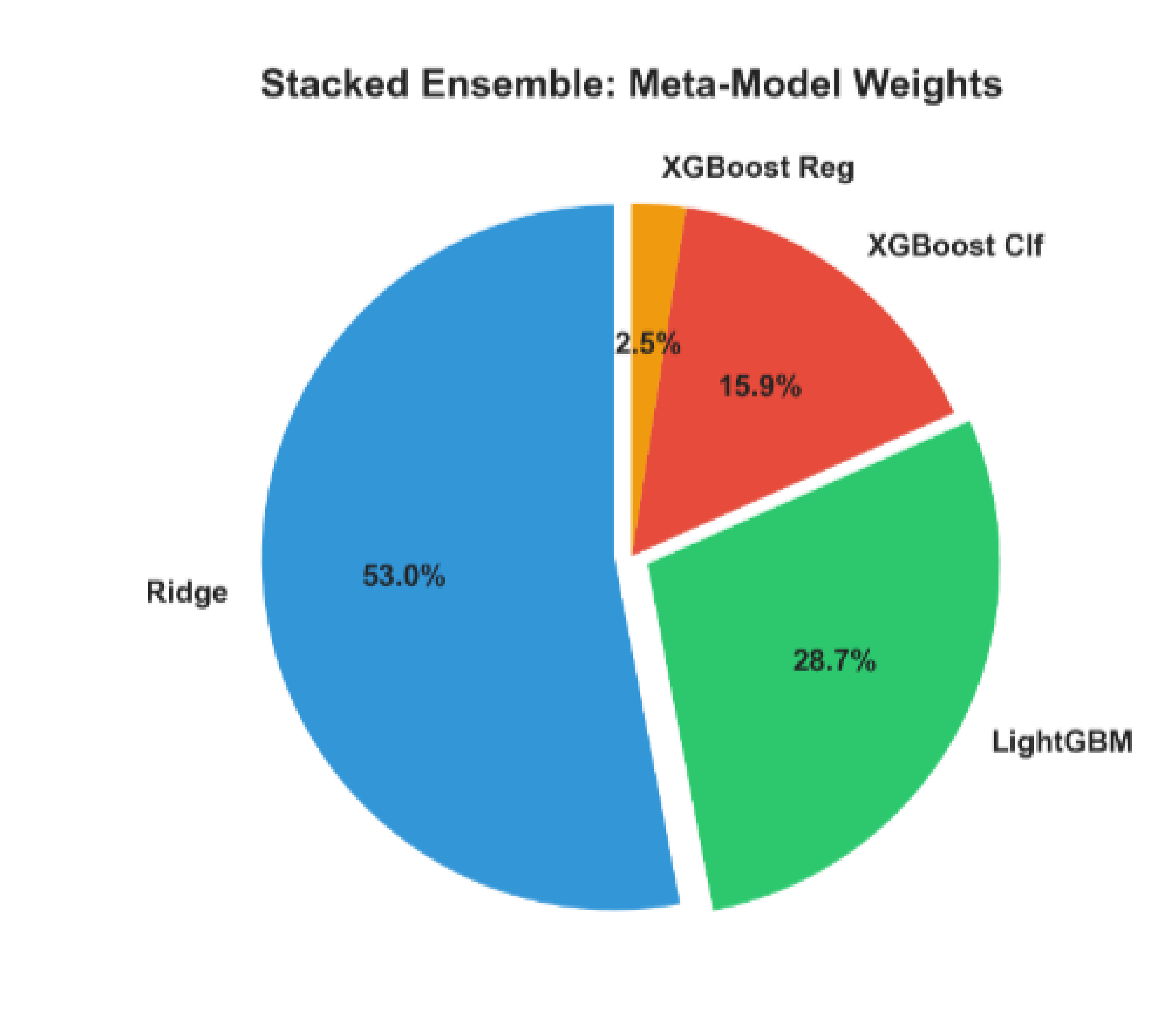




Regression Models

- Ridge Regression (baseline)
- XGBoost Regressor
- LightGBM
- Stacked Ensemble (meta-learning)





Directional Accuracy

- **Directional Accuracy** = The percentage of time your model correctly predicts whether the stock will go UP or DOWN

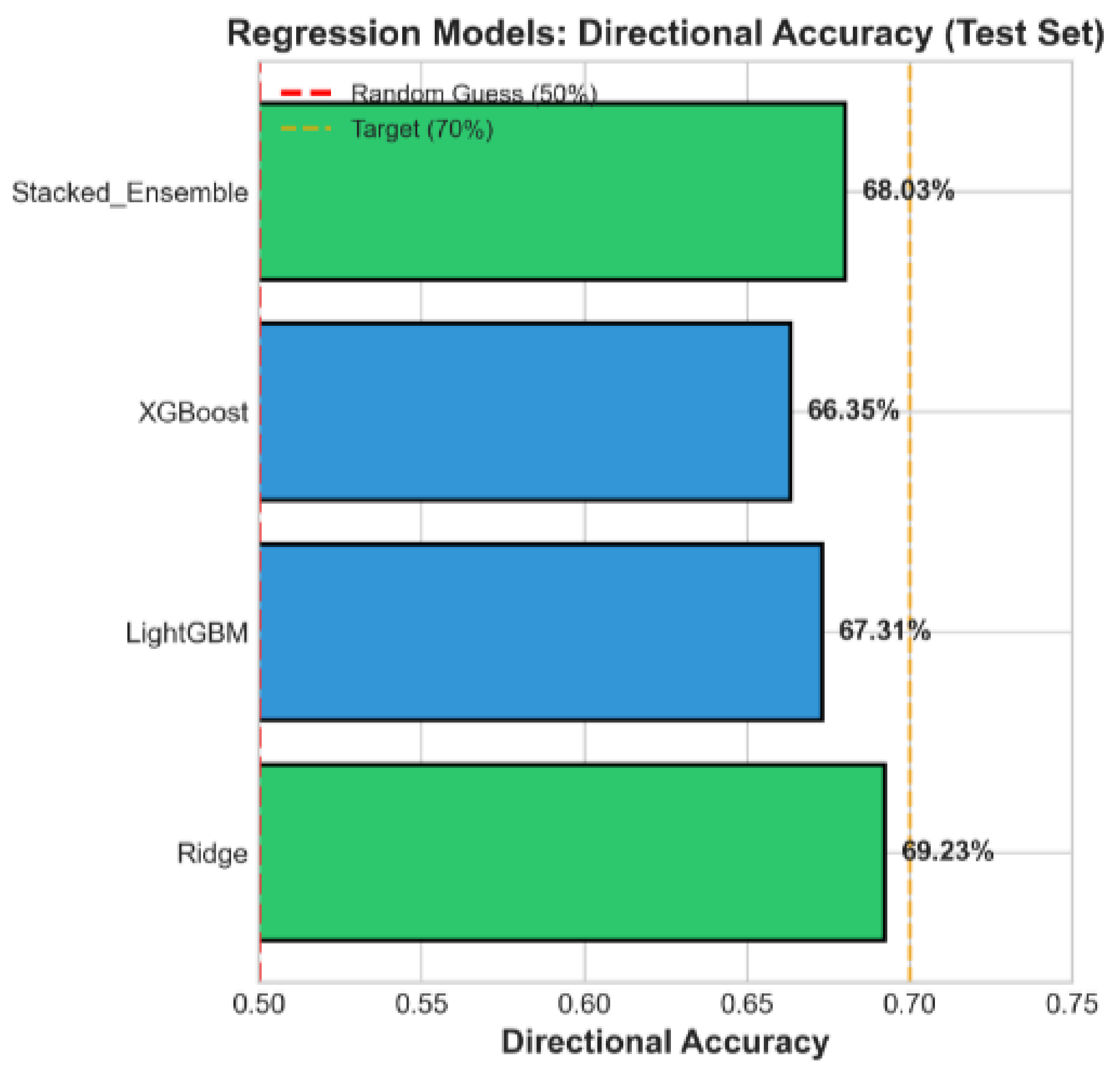
- **Why It Matters for Trading:**
  - Direction matters more than magnitude: Getting UP/DOWN right = profitable trade, regardless of exact percentage
  - Traditional metrics like RMSE or MAE measure prediction error, but they don't tell us if we'd make money
  - Predicting +2% when actual is +5% (3% error) → Still profitable if you bought the stock ✅
  - Predict -1%, actual +2% → 3% error BUT you lose money ❌

- **Business Value:**
  - Converts to 74.7% win rate in our portfolio simulation
  - Shows real-world profitability, not just statistical performance

- **In code:**
  - y_true_direction = (actual_returns > 0) # True if stock went up
  - y_pred_direction = (predicted_returns > 0) # True if we predicted up
  - directional_accuracy = accuracy_score(y_true_direction, y_pred_direction)
  - (actual_return > 0) == (predicted_return > 0). If both positive or both negative, it's correct. Count the correct predictions and divide by total

Regression Models Results

Model	MAE	RMSE	r^2	Directional Accuracy
Ridge	0.059965	0.096101	0.069883	0.692308
Stacked Ensemble	0.062538	0.096613	0.059942	0.680288
LightGBM	0.065218	0.098303	0.026767	0.673077
XGBoost	0.065256	0.097800	0.036697	0.663462



Future Directions & Improvements

1. Enhanced NLP

- Transformer models (BERT, FinBERT)
- Separate management vs. analyst Q&A sentiment
- Topic modeling for key themes

2. Real-World Implementation

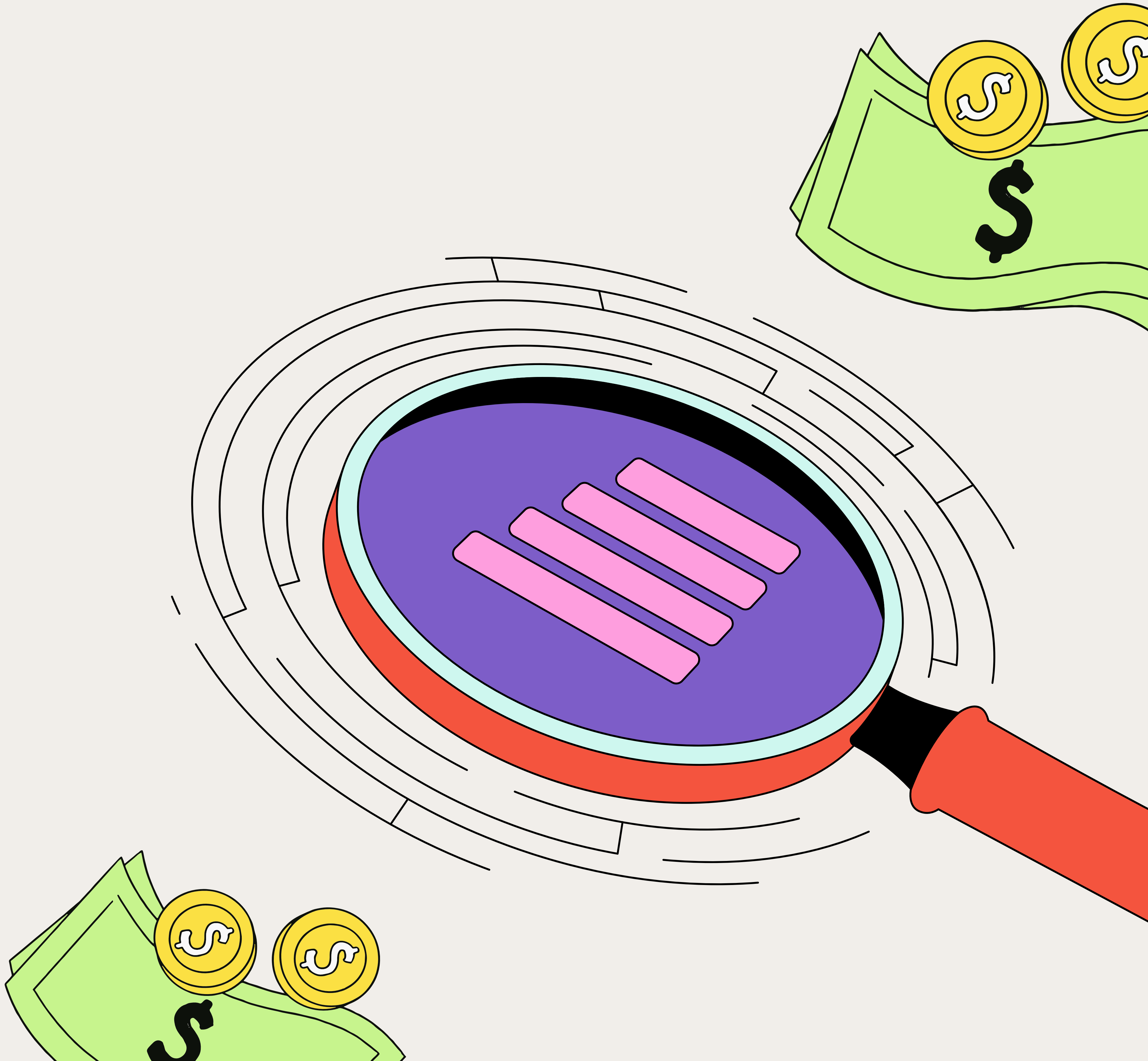
- Transaction costs & slippage
- Risk management & position sizing
- Optimal confidence thresholds

3. Robust Validation

- Walk-forward testing by year
- Bull vs. bear market performance
- Sector-specific models

4. Multi-Modal Data

- Audio features (tone, hesitation)
- Post-call analyst revisions
- Social media sentiment reaction



Thank you.

Questions or Thoughts?