AI-Driven Insight Extraction from Quarterly Earnings Data for Strategic Decision-Making

Problem

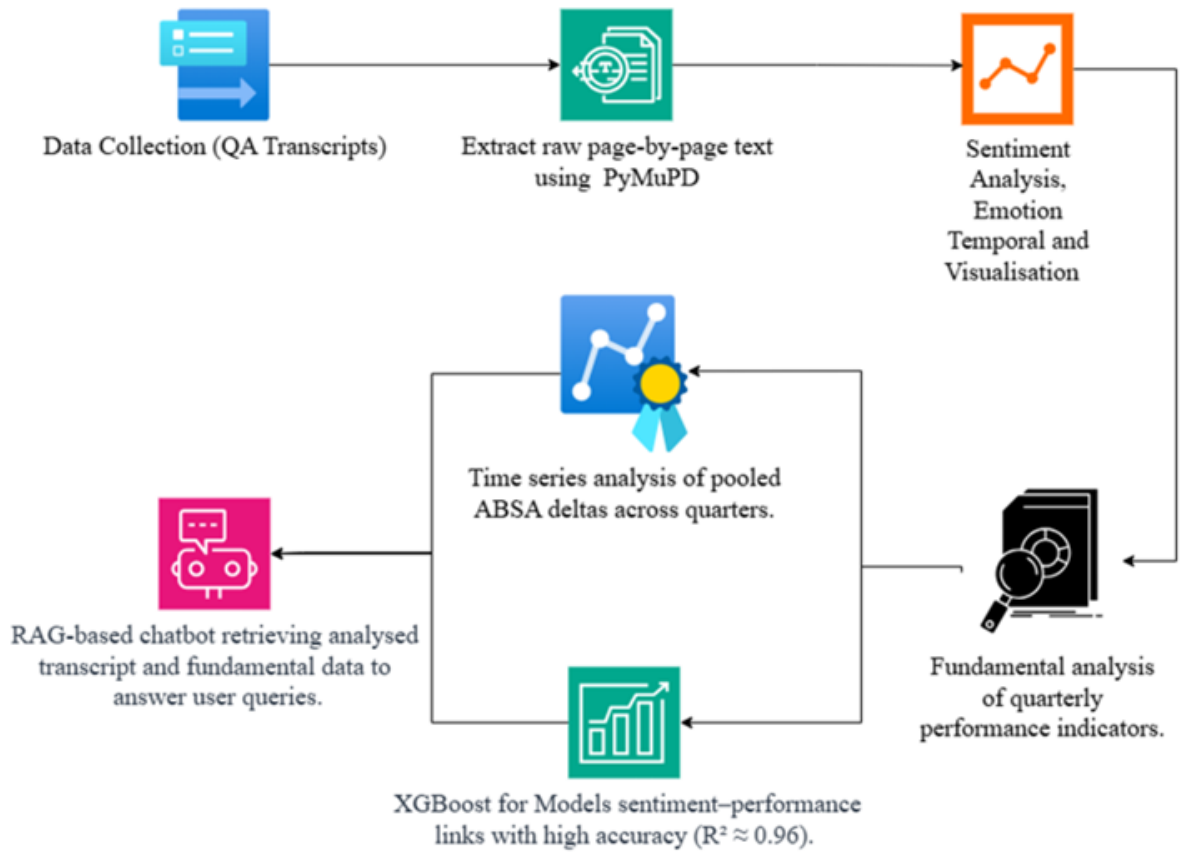
Manual review of earnings calls is time-intensive and inconsistent

Key Innovations

- Topic Modelling comparing Executive Sentiment with Analyst Intensity
- Multi-dimensional disagreement quantification through statistical analysis of analyst-executive sentiment discrepancies using ABSA model

Impact

Enhanced regulatory oversight and systematic investor protection



Technical Overview and System Development

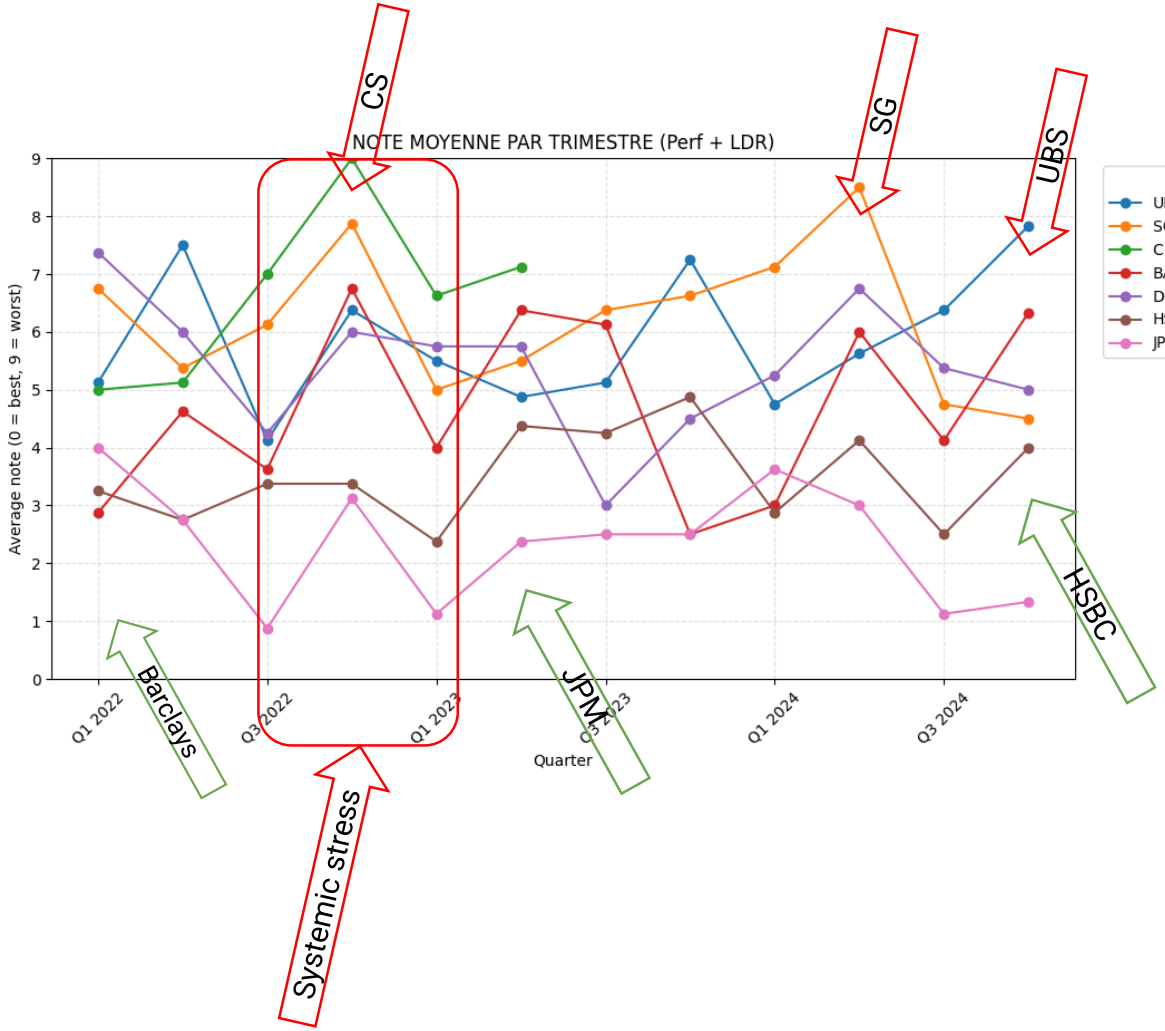
Text Extraction and Processing

- Preprocessing pipeline to extract financial report transcripts from **5 major banks;**
- Implementation of **scraping techniques** to download reports directly from banks' websites;
- Custom **parsing logic** using Python's regex and recursive structures;
- Challenges with **Deutsche Bank transcripts** (no clear speaker separation
- **"exchange_id" column** to track new discussion topics during Q&A sessions;
- Enabled analysis of **analyst questions vs. C-level responses;**
- Delivered a **structured, topic-aware dataset** to support deeper insights for data scientists;

company	call_date	fiscal_quarter	speaker_name	speaker_role	text_clean	exchange_id
UBS	26 April 2022	Q1 2022	Kian Abouhossei n	Analyst	Yes. Thank you very the best in your futur So coming to the qu Asia Pacific, can you lockdown, etcetera . unwinding , net new that context, you cle and how we should t also through the sec The second question guidance on cost ex	UBS-Q1-2022-001
UBS	26 April 2022	Q1 2022	Ralph Hamers	Executive	Okay. Thank you, Kia as your last question side, clearly there's i US, a bit in Asia. An know, I think the – or part of our cost base not the cheapest pla manageable at this r of managing our cos Secondly, as you kn delivered 200 millio well. So, we feel con variable comp depe income ratio of 70% On Asia -Pacific, on clients from last quar investments. Positive the quality of our ad We have seen certai patterns. And clearly the ammunition for i So, that I truly think t be more confidence	UBS-Q1-2022-001
UBS	26 April 2022	Q1 2022	Kirt Gardner	Executive	Yeah. Thank you, Ra sentiment and also ju regions and we starte with the launch of cc going to continue de Regarding the 1 billi what we expect is be interest rates. And as rise and we've obviou the upcoming three I would also highligh although as well that row. So, clearly that	UBS-Q1-2022-001

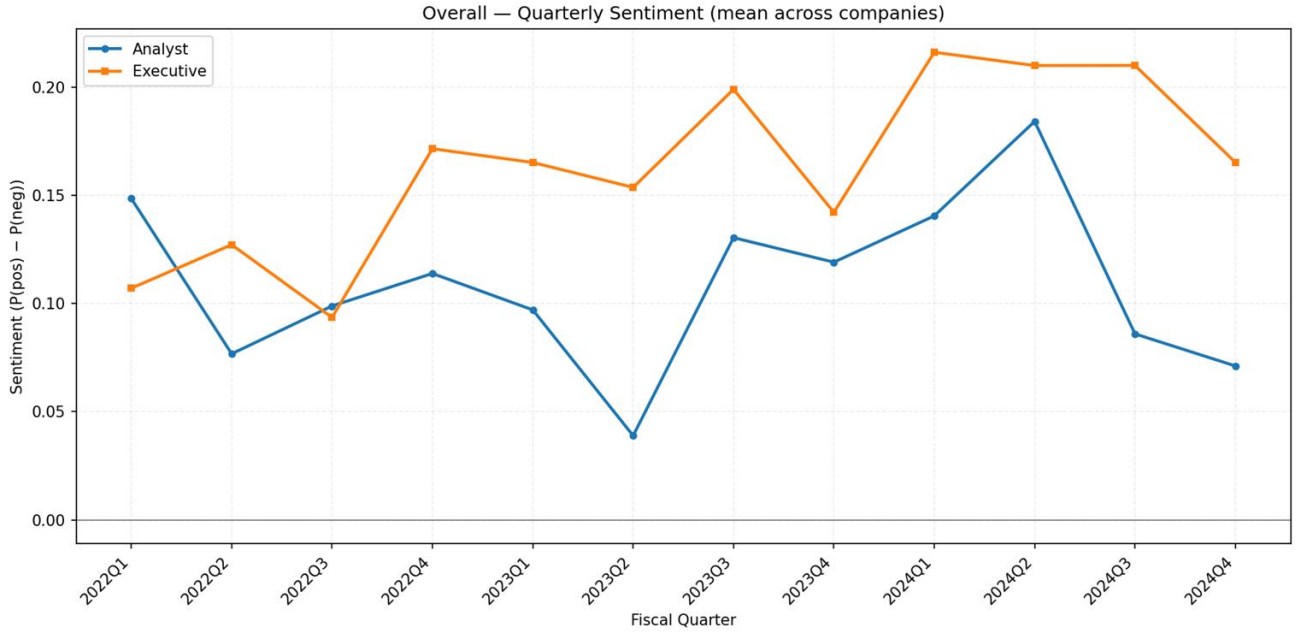
Fundamental Analysis

- Seek for the fundamental performance over 1 day, 1 week, 1 month and 3 months
- Then compute the fundamental strength of the bank derived from the LDR (Loan Deposit Ratio)
- Average the finding to seek for a final ranking from 1 to 9
- 1 strongest vs 9 weakest
- The ranking can be used as a target variable to seek the correlation between the finding in the report and the actual perceived strength of the bank at the moment of the quarterly analysis.
- JPM Strongest over the period
- Credit Suisse weakest
- Market perception is volatile over time and sometime autocorrelated between banks (Q4 2022 and Q2 2024)



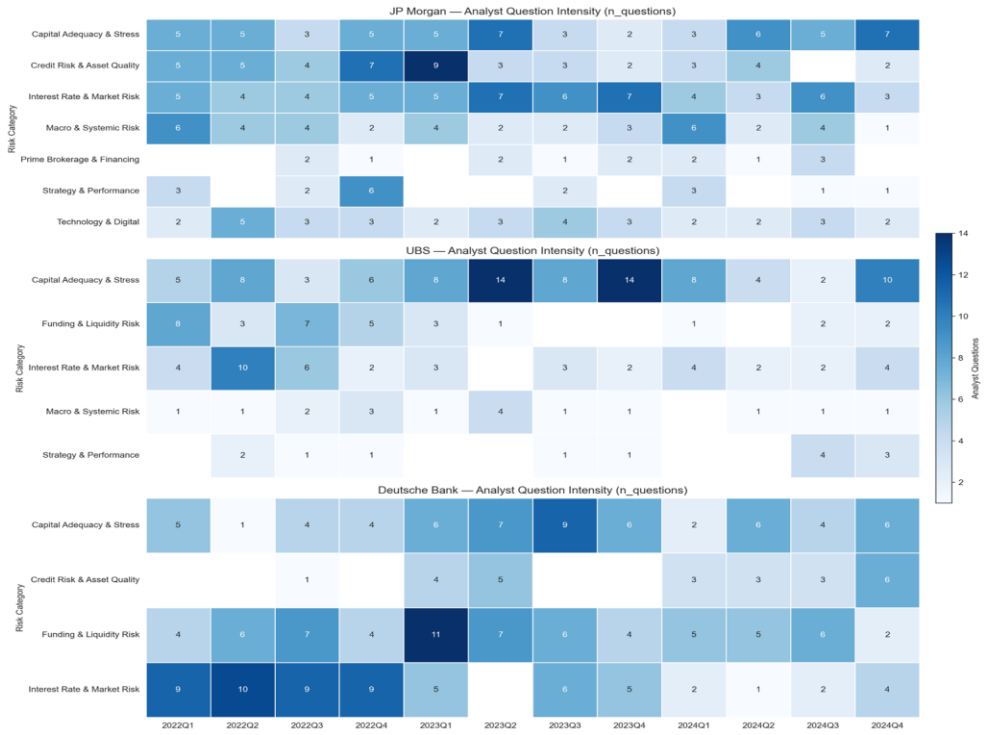
Sentiment and Emotion Analysis

- Applied FinBERT to quarterly earnings call transcripts to measure tone across executives and analysts
- Executives consistently showed higher sentiment
- Emotional analysis added behavioural context to tone
- Negative analyst sentiment often appeared one quarter before periods of balance-sheet caution
- Indicates that linguistic tone can provide early signs of changing prudential conditions



Topic Modelling

Dashboard – Management Sentiment and Analyst Pressure



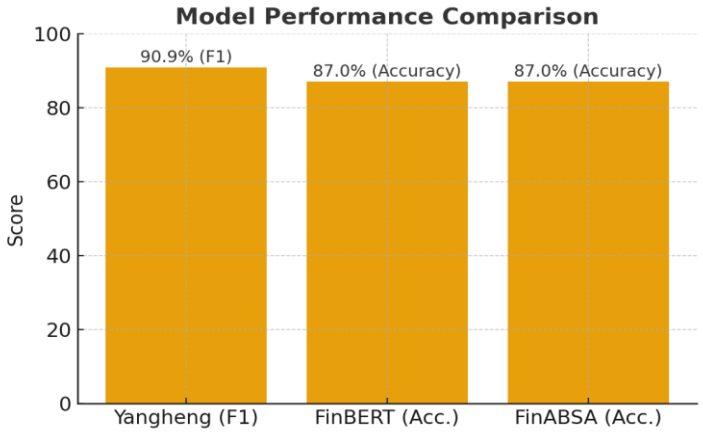
Workflow: Preprocessed Q&A transcripts → Q&A pairing → topic discovery on analyst questions → risk categorisation → sentiment scoring on management responses → quarterly aggregation → dashboard

Cross-bank Comparison and Risk Focus

Bank	Key Risk Focus	Why It Matters
Barclays	Interest Rate & Market Risk Strategy & Performance	Tied to UK rate dynamics; heightened scrutiny of execution and profitability under the changing rate environment.
HSBC	Interest Rate & Market Risk Macro & Systemic Risk	Analysts focus on global macro exposures (Asia, China property, geopolitics).
JP Morgan	Credit Risk & Asset Quality Interest Rate & Market Risk Macro & Systemic Risk	Broad analyst questioning reflects diverse global exposures. Analysts test resilience across multiple dimensions, esp. systemic shocks (CRE, regional banks).
Deutsche Bank	Interest Rate & Market Risk Funding & Liquidity Risk	Persistent questions on structural sensitivities. Analysts remain cautious given DB's trading-heavy model and reliance on wholesale funding.
UBS	Funding & Liquidity Risk Capital Adequacy & Stress	Sharp scrutiny around liquidity and capital buffers, particularly after the Credit Suisse rescue. Analysts flag vulnerability to sector-wide funding stress and integration risk.
Credit Suisse	Funding & Liquidity Risk Securitised & Structured Credit	Intense pre-collapse scrutiny: liquidity concerns flagged early, compounded by structured credit underperformance.

Deberta Model Selected for Aspect Based Sentiment

Model	Strengths	Weaknesses
yangheng/deberta-v3-base-absa-v1.1 (~90.9% F1, ~80–81% F1 on ABSA tasks) ✓	• **Multi-aspect ABSA (24 aspect classification grid)** • **ABSA-specific fine tuning** • **Financial domain adaptation**	• Higher computational cost
ProsusAI/FinBERT (~86–88% accuracy on finance sentiment tasks)	• Financial language training • Regulatory compliance focus	• **General sentiment only** • **No aspect extraction**
amphora/FinABSA (~87% accuracy on finance ABSA test split)	• Trained on financial ABSA dataset (SEntFiN) • Domain-specific sentiment mapping • Handles finance jargon well	• Finance domain • **Usually single aspect–sentiment pair**

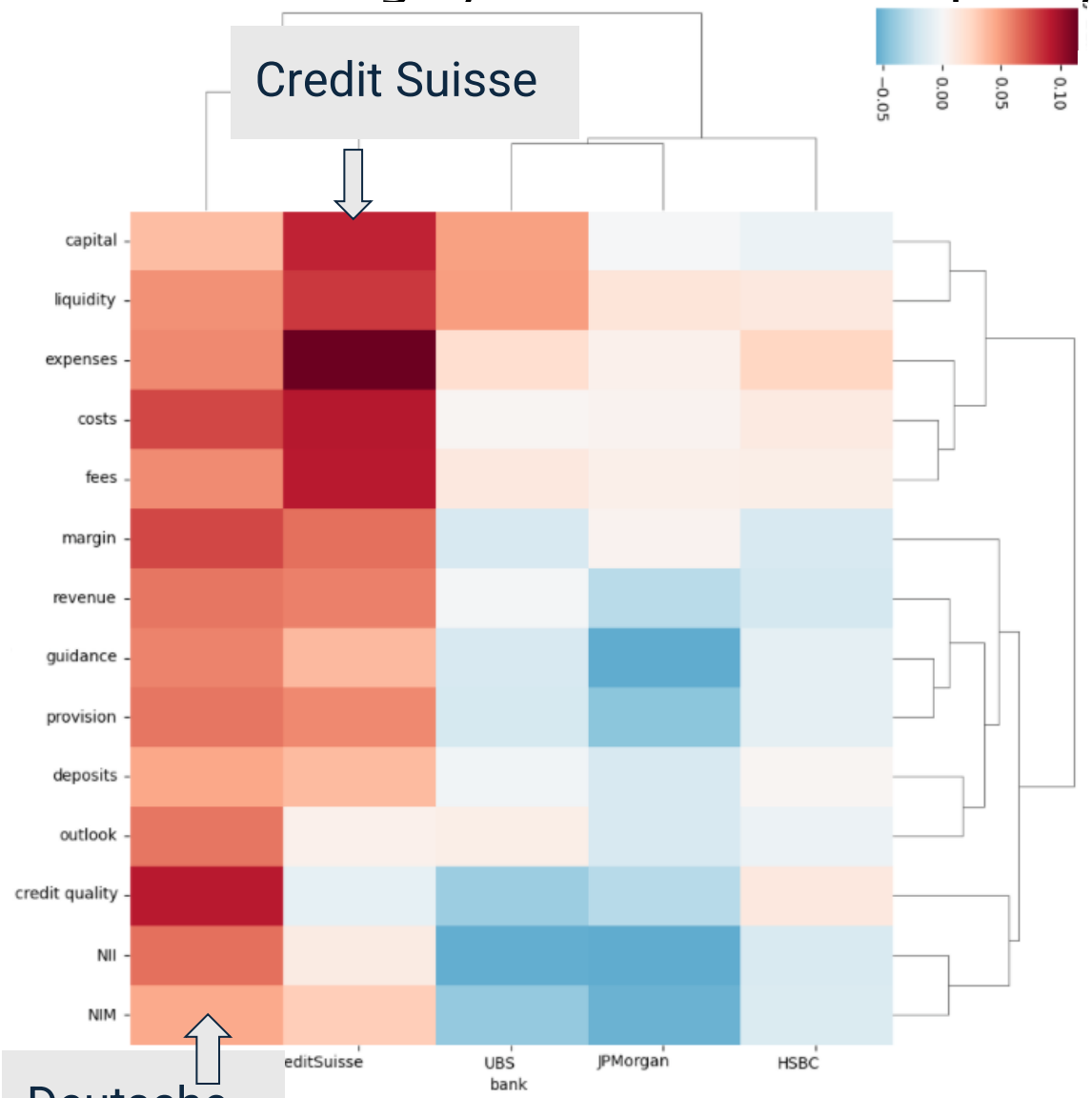


(Accuracy metric not directly comparable to F1)

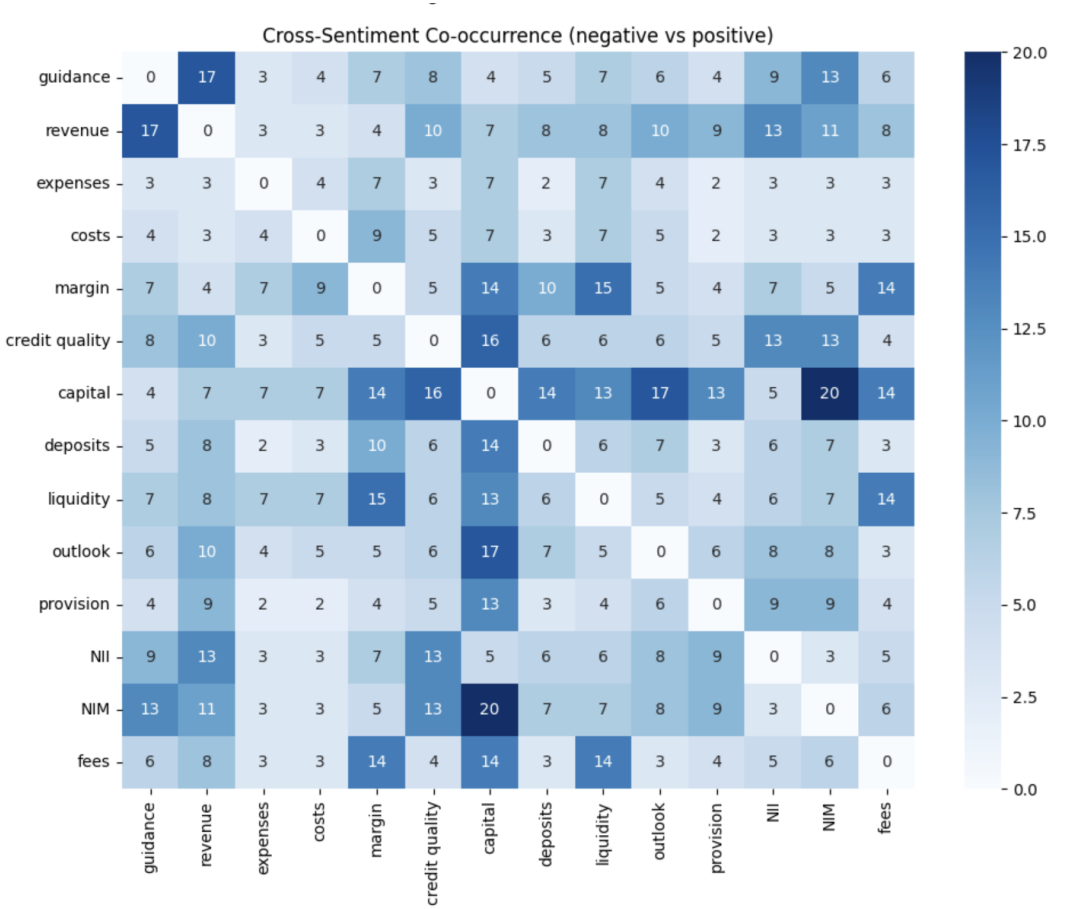
SELECTION RATIONALE:

DeBERTa-v3 ABSA selected for superior aspect-based classification

Bank Clustering by Sentiment Discrepancy detects multiple aspect issues at DB and CS



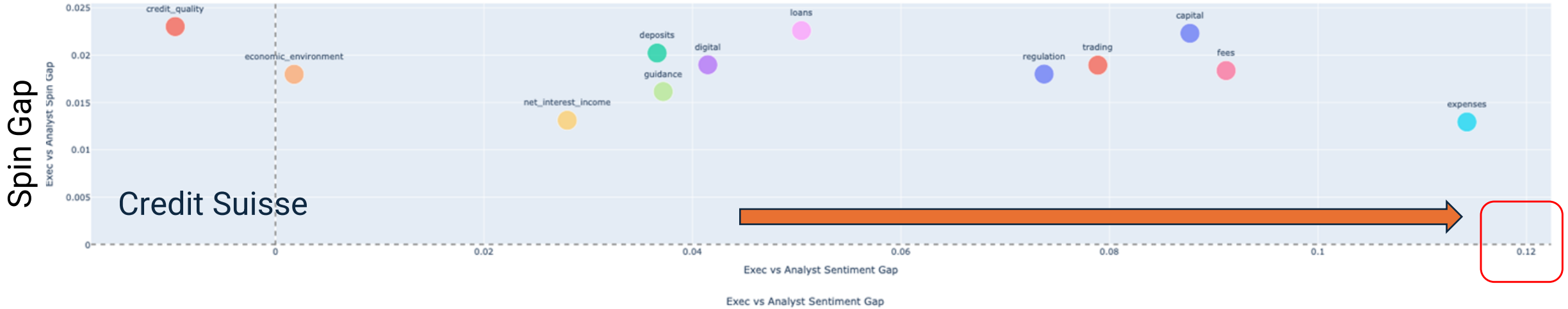
Deutsche Bank



ASPECTS:
Credit Quality
Provision
Capital
Fees
Liquidity
Deposits
Outlook
Costs
Margin
Revenue
Expenses
NII
NIM
Guidance

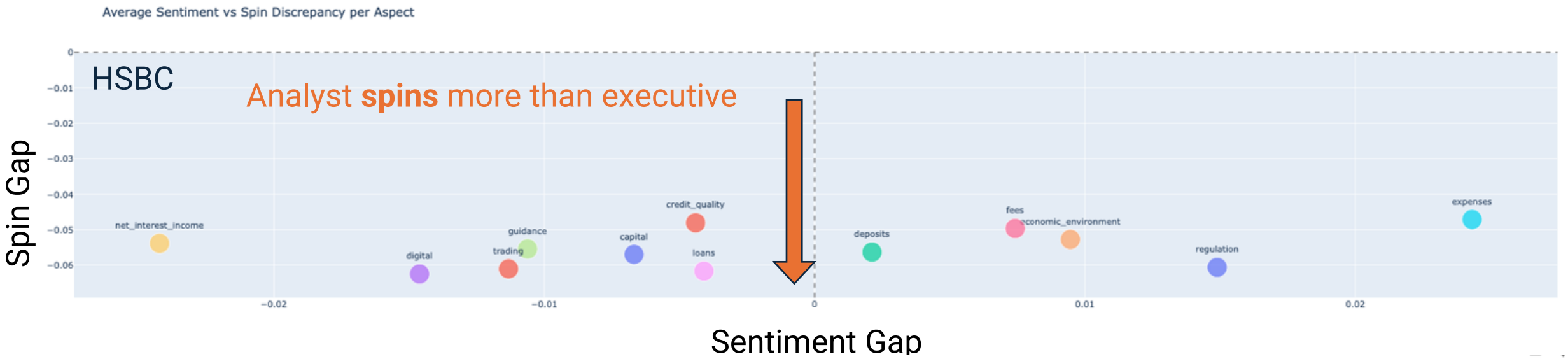
Average Sentiment vs Spin Discrepancy

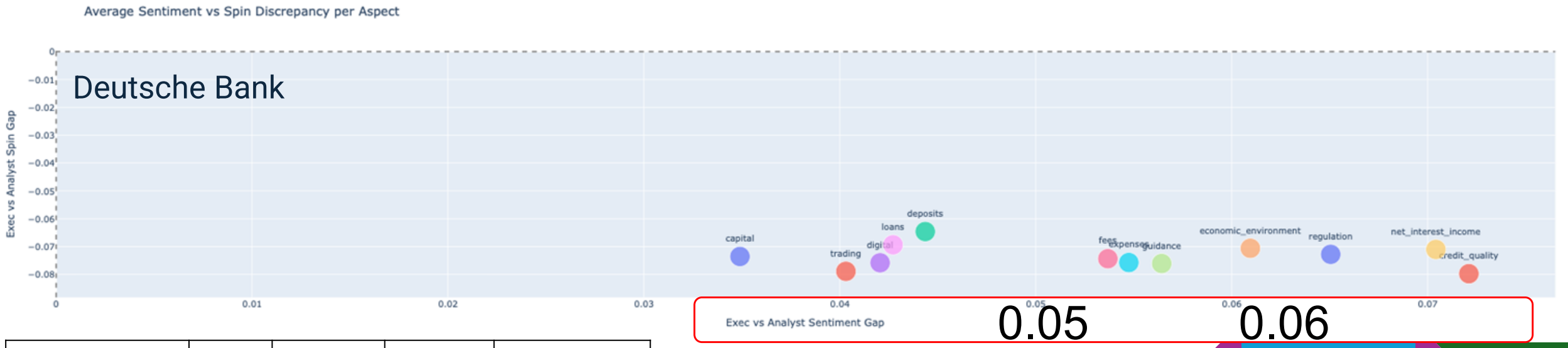
Metric	Trivial	Small	Moderate	Large / Flag
spin_gap	< 0.01	0.01–0.02	0.02–0.04	≥ 0.04
sentiment_gap	< 0.02	0.02–0.04	0.04–0.06	≥ 0.06



Exec vs Analyst Sentiment Gap

→ Time series (all aspects)





Metric	Trivial	Small	Moderate	Large / Flag
spin_gap	< 0.01	0.01–0.02	0.02–0.04	≥ 0.04
sentiment_gap	< 0.02	0.02–0.04	0.04–0.06	≥ 0.06

0.05

0.06

0.07

ABSA Delta Spin vs Basel III Metrics

Bank	LCR alignment	NSFR alignment	LDR alignment	Notable episode	ABSA verdict (vs Basel)
UBS	Strong, coincident (both climb post-Q3'23)	Strong, coincident (up with spin)	Strong inverse (LDR spike coincides with spin trough)	Post-Mar'23 recovery	Good: captures liquidity/funding resilience story
JPM	Strong, coincident	Decouples after '23 (NSFR dips, spin stays high)	Positive (moderate–strong)	Franchise strength overrides NSFR dip	Good but bank-specific: reflects trusted deposit base
HSBC	Decoupled/weak (LCR rises while spin stays low)	Decoupled/weak	Inverse (episodic at spike)	Sentiment negative despite robust liquidity	Mixed: captures market narrative not explained by Basel
Deutsche Bank	Positive, coincident	Positive, leads (~1–2q)	Strong inverse (de-risking with sentiment repair)	Post-CS contagion then normalisation	Good: signals confidence cycle; liquidity/funding drive
Credit Suisse (Q1'22–Q1'23)	Weakening then partial rebound as spin keeps falling	Strong, coincident (down)	Strong, inverse (LDR surges)	Pre-resolution deterioration	Good (early warning): strong negative alignment to funding stress

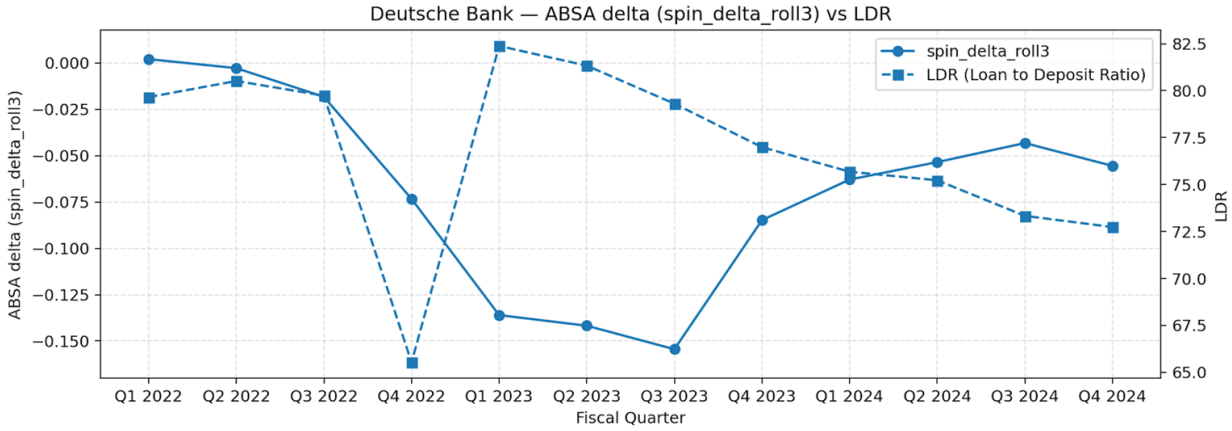
Key Takeaways:

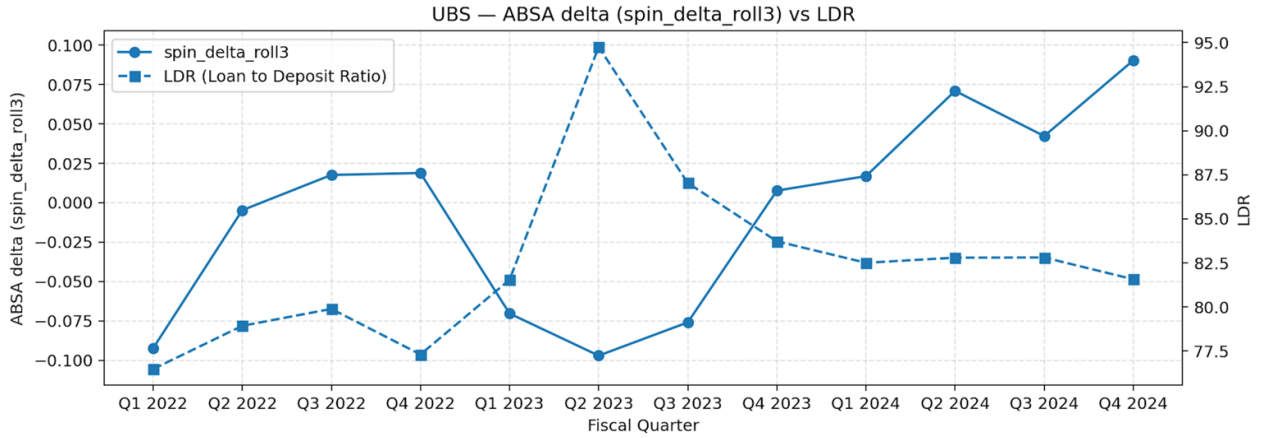
- ABSA spin complements Basel metrics as a forward-looking sentiment indicator.
- Strongest predictive power: NSFR & LCR movements; inverse relation to LDR spikes.
- CET1 acts as baseline capital stability, not sentiment driver.

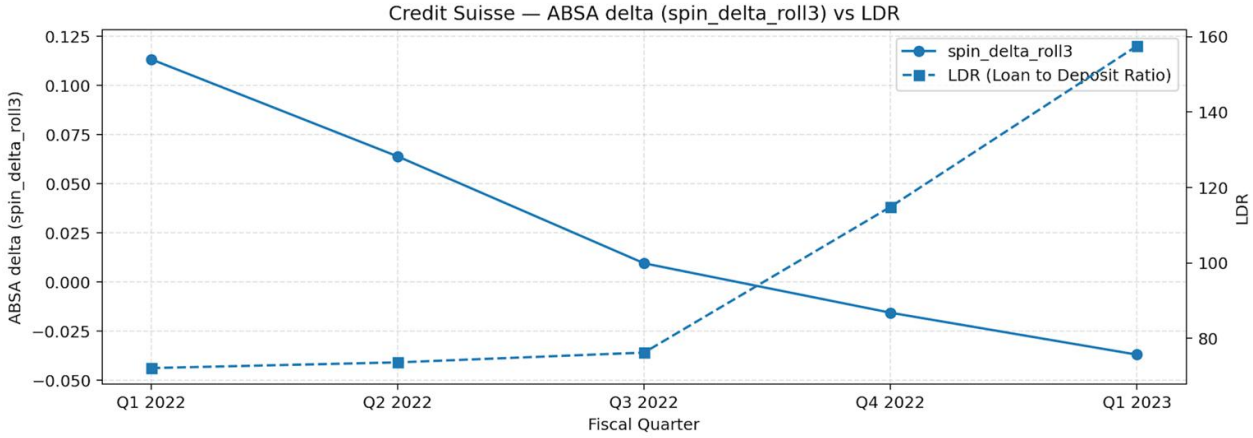
Regulatory Implications:

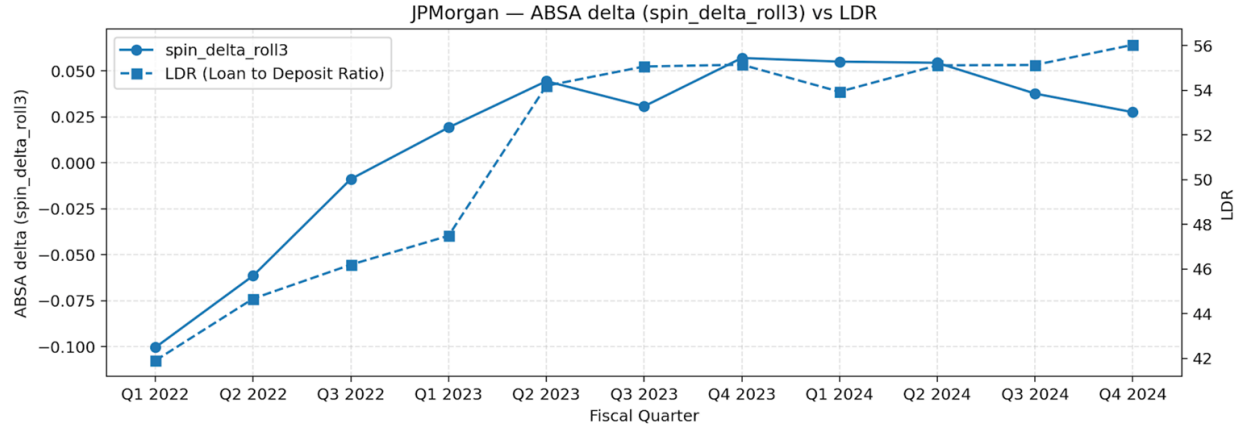
- Integrate ABSA sentiment into supervisory dashboards for early stress detection.
- Use as qualitative overlay for funding and liquidity risk assessment.
- Differences across banks highlight that market confidence depends on structure and franchise strength.

Delta Spin and LDR - Inverse Relationship

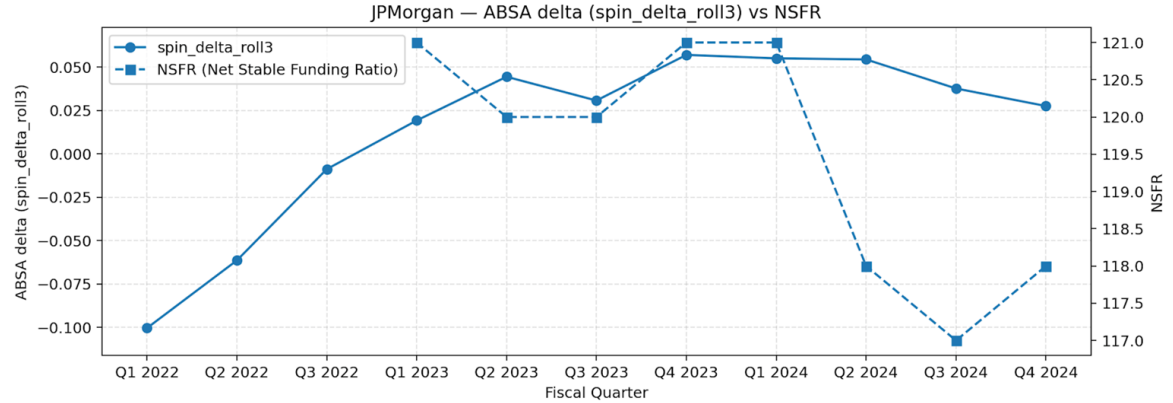


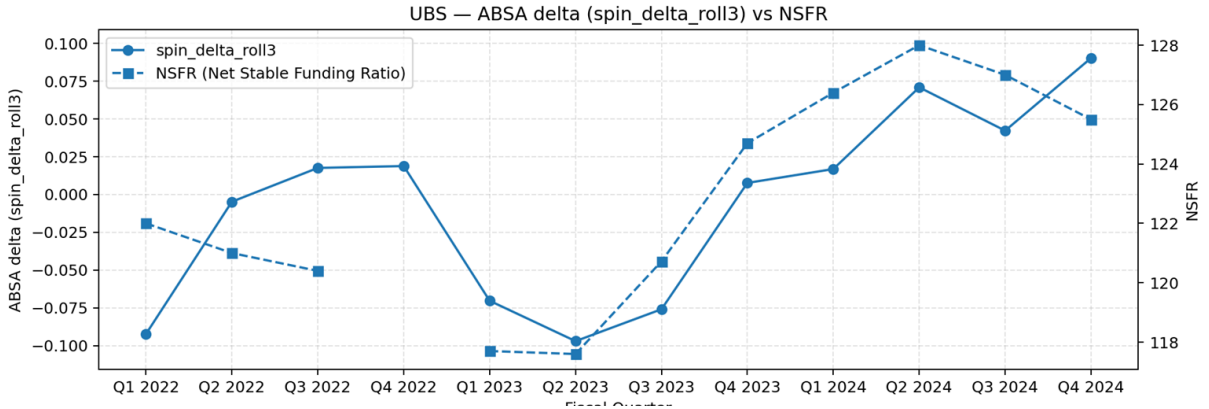


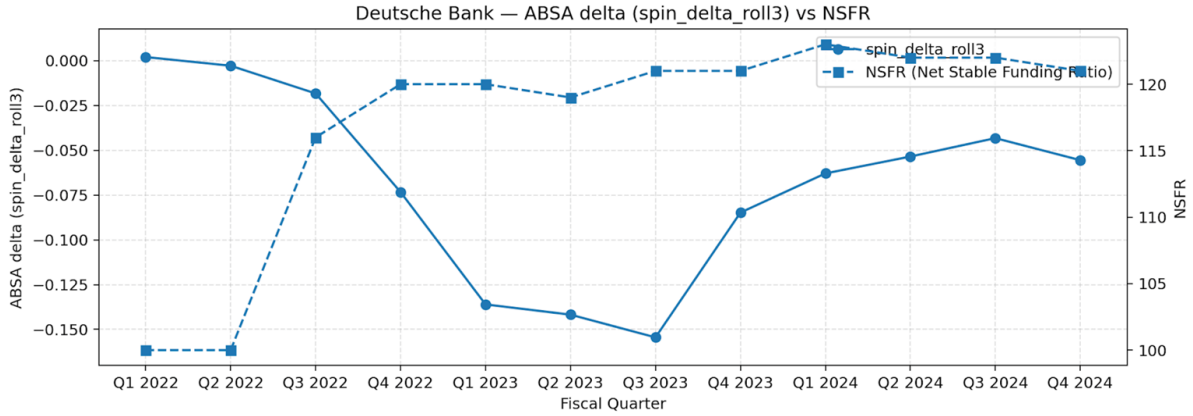




Individual Bank Delta Spin vs NSFR

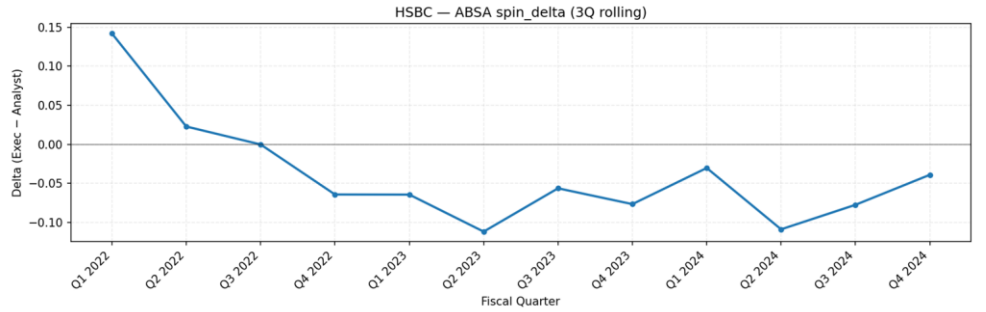


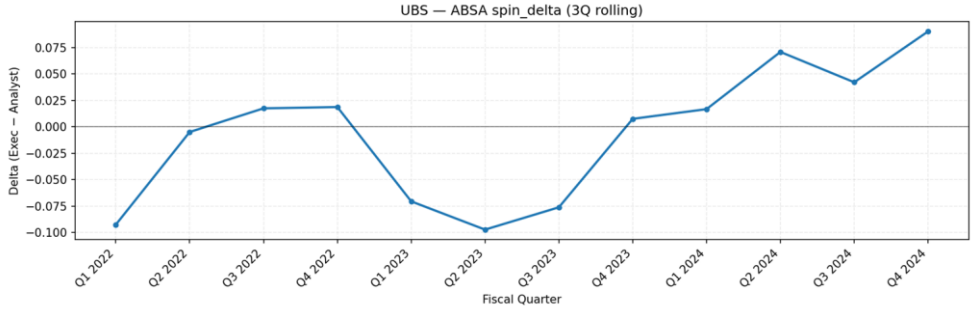


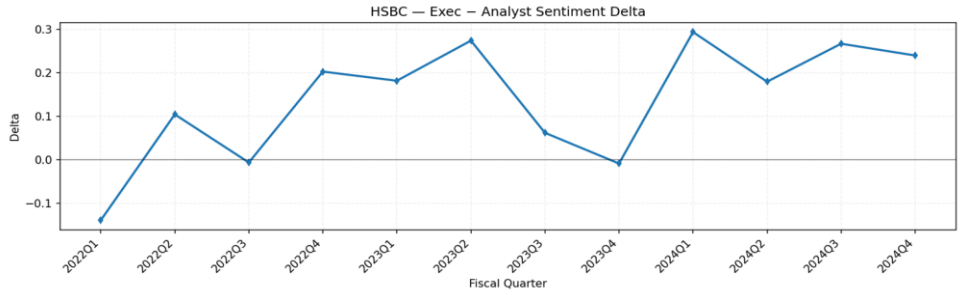


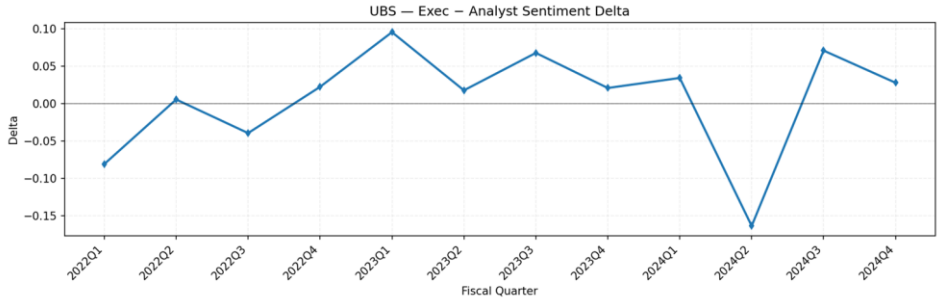
Time-Series Modelling

- Combined pooled FinBERT and ABSA sentiment with key prudential indicators (CET1, NSFR, LCR, LDR)
- Positive correlations with NSFR and LDR suggest tone improves with funding stability and lending activity
- Negative link with LCR indicates more cautious tone during high liquidity buffers
- HSBC: Strong negative CET1 correlations reflect capital-building and cautious tone
- UBS: Positive NSFT correlation signals executive confidence during funding stability
- Narrative tone demonstrates potential as an early warning signal for prudential movements









Query-Driven Interaction (RAG)

Layer	Component	Justification
Embedding Model	Sentence-transformers/all-mpnet-base-v2 (HuggingFace)	Top-performing sentence embedding model for capturing semantic similarity, suitable for Q&A and chatbot RAG. Balance between quality and efficiency.
Vector Database + Retriever	Chroma	Compatibility with RAG and LangChain
LLM (Generator)	Phi-4-mini-instruct	Used for its efficiency and strong mathematical and logical reasoning performance for complex Q&A and multi-step problem solving
Orchestration Layer	LangChain RetrievalQA or ConversationalRetrievalChain.	Enables semantic search and comparative Q&A within PDFs. Conversational RAG allows interactive exploration of the transcripts and comparison with indicators and scores.

PDF → Chunking → Embedding → CSVs → Embedding →
Chroma DB → Phi-4 Mini instruct → RetrievalQA → ConversationalRetrieval

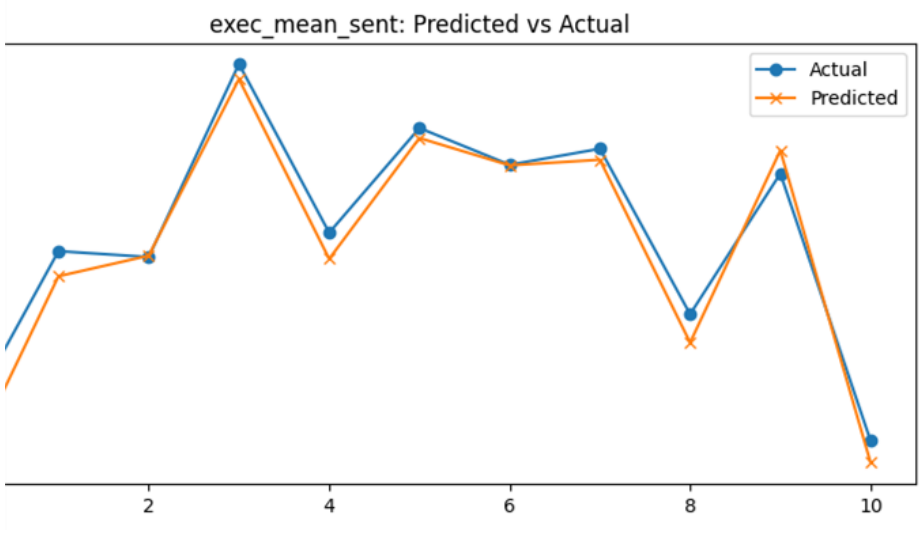
RAG Summary

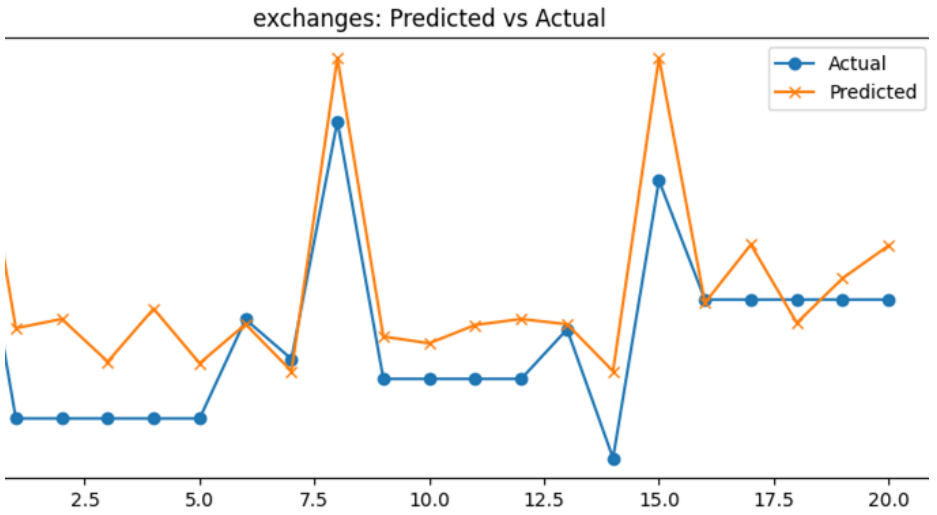
Comparison	Observation	Evasion Rating
CS Q1 → Q2 2022	Shift from concrete → vague answers	2 → 4
CS vs JPM Q3–Q4 2022	JPM clear, CS evasive	CS: 4/5 JPM: 1/5
UBS 2022	Broad, repetitive answers, avoiding specifics	4/5
UBS Q1-Q4 2024	Generalised affirmations, lacking depth, repetitive answers without providing details	2/5

Key Insight:
Linguistic evasiveness preceded observable financial stress — a potential **language-based early warning signal**.

Predictive Modelling

- XGBoost achieved near-perfect accuracy for *executive sentiment ($R^2 \approx 0.96$)* and *discrepancy magnitude ($R^2 = 0.979$)*, showing strong alignment between language and short-term performance.
- Executive and analyst tone were the dominant predictive factors, confirming that management sentiment provides measurable insight into corporate outlook.
- Predictive power declined for aggregated or lagged indicators ($R^2 < 0.45$), indicating influence from broader contextual factors.
- ML and NLP models are highly effective for real-time decision support but need integration with financial and temporal data for longer-term forecasting.





Conclusion

The integration of **FinBERT, ABSA, and prudential metrics** revealed statistically coherent links between **linguistic tone and financial health**, confirming sentiment as a reliable prudential indicator.

XGBoost's high predictive performance validated that management and analyst tone explain most short-term prudential variability across banks.

JP Morgan's stable tone corresponded to sustained prudential strength, while **Credit Suisse's negative tone and high spin gaps** matched its declining market confidence—demonstrating that **language mirrors real financial trajectories**.

The inverse relationship between **ABSA spin and LDR** highlights that **positive tone often coincides with reduced lending activity**, offering a nuanced behavioural signal missed by numeric ratios.

Overall, our results prove that **AI-driven linguistic analytics can transform earnings calls into quantifiable prudential intelligence**, enabling **early detection of stress, improved supervisory insight, and more informed financial decision-making**.

Next Steps

- Train ABSA model on FiQA to improve performance
- Carry out Grid Search on Weights for parametric spin_score
- Improvements to the Chatbot model - retrieval and memory