Stock Sentiment Analysis Project

Predicting Stock Price Movements Using News Sentiment Analysis

PROJECT OVERVIEW

Comprehensive analysis of the relationship between news sentiment and stock price movements across 8 major technology companies (AAPL, AMZN, GOOGL, META, MSFT, NFLX, NVDA, TSLA) over a 12-month period. This end-to-end data analytics project demonstrates proficiency in SQL, Python, statistical analysis, and data visualization.

KEY FINDINGS

Metric	Result	Insight
Prediction Accuracy	84% average	Sentiment successfully predicts price direction on strong signal days
Best Predictor	TSLA (r=0.61)	Tesla shows strongest correlation between sentiment and price
Impact Asymmetry	+2x positive	Positive news has twice the impact of negative news on prices
Sample Size	2,000+ days	8 stocks × 250 trading days with comprehensive coverage

METHODOLOGY & TECHNICAL APPROACH

Data Pipeline: Extracted financial data and sentiment scores using complex SQL queries with window functions, CTEs, and joins. Processed 2,000+ records with pandas for statistical analysis and correlation testing.

Analysis Techniques: Correlation analysis (Pearson), prediction accuracy modeling, time series analysis with rolling averages, error analysis (false positives/negatives), and monthly trend identification.

Validation: Tested sentiment predictions against actual next-day price movements. Calculated accuracy metrics across all 8 stocks with detailed error classification.

TECHNICAL SKILLS DEMONSTRATED

- **SQL:** Complex queries, window functions, CTEs, aggregations, and multi-table joins
- **Python:** pandas (data manipulation), matplotlib & seaborn (visualization), numpy (statistical operations)
- **Statistics:** Correlation analysis, accuracy metrics, trend analysis, confidence evaluation
- **Data Visualization:** 25+ publication-ready charts including time series, scatter plots, bar charts, heatmaps
- **Business Analysis:** Translated technical findings into actionable insights for stakeholders

BUSINESS RECOMMENDATIONS

1. Trading Strategy: Use sentiment as a confirming indicator, not a sole signal. Focus on TSLA, AAPL, and NFLX where correlations exceed 0.5.
2. Risk Management: Exercise caution during earnings seasons when sentiment reliability decreases by ~15%.
3. Expansion Opportunities: Extend analysis to additional sectors (finance, healthcare) to test model generalizability.

PROJECT DELIVERABLES

- 11 SQL queries covering correlation, accuracy, trend analysis, and error detection
- Interactive Jupyter notebook with complete analysis pipeline and 25+ visualizations
- Comprehensive dataset (CSV) with 2,000+ records ready for further analysis

• Executive summary with key findings and business recommendations

LIMITATIONS & NEXT STEPS

Current Limitations: Analysis limited to tech sector; doesn't account for broader market movements; sentiment timing may not capture intraday news impact.

Future Enhancements: Add machine learning classification models (Random Forest, XGBoost), expand to 500+ stocks across all sectors, incorporate market indicators (VIX, sector ETFs), backtest trading strategies with risk metrics.

This project showcases end-to-end data analysis capabilities from data extraction through insight generation. All code, queries, and visualizations available upon request.