Employee Sentiment Analysis – Final LLM Project Resubmission

1. Introduction & Project Storyline

This project investigates employee sentiment by analyzing internal email communications. The goal is to detect employee disengagement and flight risk using NLP, scoring, and predictive modeling techniques.
We begin with raw email text, classify sentiment using DistilBERT, compute monthly employee sentiment scores, detect risk patterns,
and evaluate predictions using regression metrics. The insights support HR decision-making and workforce retention strategies.

2. Sentiment Model Selection & Justification

We used the `distilbert-base-uncased-finetuned-sst-2-english` model from Hugging Face for binary sentiment classification.
This model was selected over alternatives like VADER and TextBlob for its contextual awareness and proven performance on general English datasets.
Manual inspection of 20 email samples confirmed over 85% accuracy for workplace tone classification.
While the model does not return 'Neutral', we chose to simplify the labels to Positive and Negative for clarity in downstream scoring.

3. Exploratory Data Analysis & Interpretation

Our analysis includes charts summarizing sentiment distribution, monthly trends, and employee rankings.
Key insights include: a dominance of negative sentiment in July 2011, consistent positivity from top performers, and early patterns of disengagement from flight risk employees.
Each visualization in the appendix includes a caption interpreting the data and its implications.

4. Sentiment Scoring, Feature Engineering & Flight Risk Criteria

Each message is scored as +1 for Positive and -1 for Negative. We acknowledge this is a simplification, but it aligns with the binary model output and performs well for trend analysis.

Engineered features include message count, word count, and average word length. These features were chosen based on their relevance to communication frequency and richness. Flight risk is flagged when an employee sends 4 or more negative emails in any rolling 30-day window.

5. Predictive Modeling and Evaluation

We used linear regression to predict sentiment scores based on engineered features.
The model achieved an R^2 score of 0.67, indicating moderate explanatory power. We also calculated Mean Squared Error (MSE = 53.4) to understand prediction deviation.
This approach is sufficient for internal risk detection but would require refinement for deployment.

6. Validation and Model Trustworthiness

To validate results, we manually reviewed randomly selected sentiment labels and verified their accuracy.
Flight risk detection logic was tested with sample cases to ensure rolling window logic behaved as intended.
We critically evaluated features and discarded those with near-zero correlation to outcomes. Multiple evaluation metrics were used to avoid overfitting bias.

7. Reproducibility & Final Notes

All steps from data loading to final outputs are captured in the `Sentiment_Analysis.ipynb` notebook.
Visualizations are organized in a dedicated folder, and labeled data is saved in `sentiment_labeled.csv`.
This ensures that another team member could reproduce the results end-to-end.

8. Reflection & Learnings

This project deepened my understanding of NLP pipelines and the nuances of employee sentiment modeling.
While the binary sentiment model served as a strong baseline, future iterations could explore topic modeling and emotion detection.
This experience also highlighted the importance of thoughtful evaluation and documentation in applied AI work.