**Employee Sentiment Analysis – Final Report**

*Comprehensive Methodology, Findings, and Visualizations*

Author: Phuc Lam  
Date: September 15, 2025

Contents

[Executive Summary 2](#_Toc208782732)

[1. Project Overview 3](#_Toc208782733)

[2. Data & Preprocessing 3](#_Toc208782734)

[3. Task 1 – Sentiment Labeling 3](#_Toc208782735)

[4. Task 2 – Exploratory Data Analysis (EDA) 3](#_Toc208782736)

[5. Task 3 – Employee Score Calculation 4](#_Toc208782737)

[6. Task 4 – Employee Ranking 4](#_Toc208782738)

[7. Task 5 – Flight Risk Identification 5](#_Toc208782739)

[8. Task 6 – Predictive Modeling (Linear Regression) 5](#_Toc208782740)

[9. Key Findings & Recommendations 6](#_Toc208782741)

[1. Set up a dashboard 6](#_Toc208782742)

[2. Use rolling alerts 6](#_Toc208782743)

[3. Train managers to look beyond task performance 6](#_Toc208782744)

[4. Encourage healthier communication habits 6](#_Toc208782745)

[4. Keep building on the model. 6](#_Toc208782746)

[5. Don’t forget human review. 7](#_Toc208782747)

[10. Reproducibility & Environment 7](#_Toc208782748)

[Visualizations (Selected) 7](#_Toc208782749)

[Appendix A – All Figures 13](#_Toc208782750)

[Appendix B – Tables (Fill with Results) 73](#_Toc208782751)

# Executive Summary

This report documents an end-to-end employee sentiment analysis workflow, executed on an unlabeled corpus of messages. The pipeline covers sentiment labeling, exploratory data analysis (EDA), employee-level monthly scoring and ranking, flight-risk identification using a rolling 30‑day window, and a predictive linear regression model to analyze drivers of sentiment and forecast trends. All code is reproducible in the accompanying notebook, and key outputs are visualized throughout this report with an Appendix containing all generated figures.  
  
Highlights:  
• Automated labeling of messages into Positive, Negative, or Neutral using a hybrid approach.  
• Monthly employee sentiment scores and clear Top/Bottom performer rankings.  
• Flight‑risk flagging for employees with ≥4 negative messages within any rolling 30‑day period.  
• A baseline linear regression model using behavioral/textual features to predict sentiment score trends, evaluated via standard metrics (R², MAE, RMSE).

# 1. Project Overview

Objective: Analyze unlabeled employee messages to derive sentiment, trends, and engagement signals; compute monthly sentiment scores and rankings; identify potential flight risk; and build a predictive model to understand and forecast sentiment dynamics.

# 2. Data & Preprocessing

The dataset, provided by Springer Capital, contains **2,191 rows and 4 columns** of raw text messages linked to employee identifiers and timestamps. The researcher standardized the text (lowercasing, URL/email stripping, punctuation normalization, whitespace trimming), removed duplicates and system-generated messages, retained English content only, and redacted incidental PII. Features were engineered at both the message and employee-month levels—including message length, word count, average word length, stop-word ratio, punctuation/emphasis density, send-frequency aggregates, and rolling 30-day counts (e.g., recent negatives). All timestamps were standardized to a single timezone, enabling construction of monthly and 30-day rolling windows. Quality controls included schema validation, missing-value audits, and outlier handling (e.g., capping unusually long messages), after which the data was prepared for modeling with appropriate scaling and leakage-safe train/test splits.

# 3. Task 1 – Sentiment Labeling

Approach: Labeled each message as Positive, Negative, or Neutral using a two‑stage pipeline:  
1) A pretrained transformer sentiment classifier produced initial polarity and confidence scores.

2) Lightweight **rule-based adjustments** refined borderline cases to better capture Neutral tone (e.g., probability gray zones, negations like “not bad,” and domain phrases such as “issue resolved,” “draft,” “FYI,” which are often non-emotive).  
  
Outputs: A new column, sentiment\_label, was appended to the dataset. Visual diagnostics include **label distribution** charts (class balance) and **confidence histograms** (calibration/uncertainty).

4. Task 2 – Exploratory Data Analysis (EDA)  
For EDA, the analyst kept it simple and practical: message volume was checked by day and month to spot normal rhythms versus spikes (deadlines, outages) and dips that might mean disengagement or time off. Sentiment was broken out by employee, team, and month to see whether negativity came from a few people/periods or showed up everywhere. Writing style was profiled such as average length, word count, lexical richness (how varied the words are), plus “emphasis” tells like ALL-CAPS, !!!, or repeated punctuation to separate quick status pings from stressed or urgent notes.   
The Springer Capital dataset includes 2,191 messages and 4 columns (Subject, body, date, from) spanning Jan 2010–Dec 2011. There are 10 unique senders; the top three by volume are lydia.delgado@enron.com (284), john.arnold@enron.com (256), and sally.beck@enron.com (227). Messages are fairly short, median 179 characters (28 words), average 260 characters (41 words), with about 0.21 exclamation marks and 1.47 ALL-CAPS tokens per message, suggesting mostly routine updates with occasional emphasis. Activity peaks in Apr 2010 (92 messages), a useful anchor for distinguishing bursts from the normal rhythm. This dataset forms the base for Task 2 EDA: checking volume trends, sentiment splits by person/team/month, writing style (length/lexical richness/emphasis), and anomalies like sudden spikes or long quiet stretches.

# 5. Task 3 – Employee Score Calculation

Using the plus one, minus one, zero rule per message and aggregating by employee and month, I see most employees cluster near zero in a typical month, with a few spiky months where scores swing positive or negative. Those swings line up with known activity bursts, like the message peak in April 2010, which points to workload and deadlines driving short run shifts. The all time leaderboard in the notebook shows a small set of high activity senders at the extremes: top cumulative positive contributors include lydia.delgado@enron.com and patti.thompson@enron.com, while the negative side includes bobette.riner@ipgdirect.com. Month over month, most people move within a narrow band, but short runs of consecutive negative months do appear and often match periods of heavy activity. In practice, the monthly score works as an early signal: sustained dips flag employees or teams for a check in, while broad, simultaneous swings point to company wide stressors like launches or incidents. These scores also power downstream steps such as ranking the top and bottom for each month and supporting the flight risk rule, defined as four or more negative messages in any thirty day window. Keeping the scoring calibrated helps keep alerts focused and actionable.

# 6. Task 4 – Employee Ranking

For each month, employees are sorted by total monthly sentiment score (descending). Ties are broken alphabetically. Two lists are provided for each month: Top Three Positive and Top Three Negative. These rankings are presented in tables and charts within the visualization section and Appendix.

After calculating monthly sentiment scores, employees were sorted each month from highest to lowest. The top three most positive and top three most negative employees were identified across the full dataset and for each month. This helped surface consistent high performers as well as people showing repeated low scores. For example, lydia.delgado@enron.com consistently ranked high in cumulative positive score, while others like bobette.riner@ipgdirect.com showed up frequently on the negative side. This kind of monthly ranking adds context to team morale and helps HR or team leads zoom in on both recognition and risk areas without needing to dig through raw messages.

# 7. Task 5 – Flight Risk Identification

Definition: An employee is flagged as 'Flight Risk' if they sent ≥4 negative messages within any rolling 30‑day window (independent of calendar months). Implementation uses a time‑sorted rolling count per employee over 30-day windows. Flag results and timelines are visualized to support intervention planning.

Flight risk was flagged based on a simple rule: if an employee sent four or more negative messages within any rolling 30-day window, they were marked as at-risk. This method doesn’t depend on calendar months and allows more dynamic detection. A few employees hit that threshold in bursts—usually clustered around known stress periods like heavy workloads or organizational change. This rolling-window logic helps surface cases where negativity builds quickly, giving leaders a chance to act early rather than waiting for quarterly reviews or surveys. It’s especially useful for spotting short-term emotional drops before they become long-term disengagement.

# 8. Task 6 – Predictive Modeling (Linear Regression)

Target: Monthly employee sentiment score.  
Feature set (examples): monthly message count, average message length, median message length, total word count, negativity ratio, positivity ratio, neutral ratio, weekday/weekend posting mix, after‑hours messaging ratio, and recent rolling statistics (e.g., last‑30‑day negative count).  
  
Method: Train/test split at the (employee, month) level to prevent leakage. Standard linear regression with cross‑validation. Evaluation uses R², MAE, and RMSE on the held‑out test set. Coefficients are interpreted to understand drivers of higher or lower sentiment scores. Regularization (Ridge/Lasso) considered for robustness.

A basic linear regression model was trained to predict monthly sentiment scores using features like message frequency, word count, average length, and rolling sentiment ratios. The model was split into training and testing sets by employee-month to avoid leakage. Some features—like a higher positivity ratio and consistent message length—were good predictors of higher sentiment scores, while higher after-hours messaging or spikes in recent negative messages pulled scores down. The model had reasonable performance (e.g., R² and RMSE were calculated in the notebook but not shown here), and more advanced modeling (e.g., Ridge or Lasso) could further refine it. The point of the model is to help forecast dips early and give managers signals based on behavior patterns before a person ever sends four negative emails in a row.

# 9. Key Findings & Recommendations

Most employees seem to communicate in a pretty stable way from month to month. Their sentiment scores don’t swing too far in either direction, which suggests that, for the most part, people are steady and consistent in tone. But there are a few cases where scores drop noticeably, either all at once or over a few months. These dips usually match busy periods, like big projects or internal deadlines, which tells me that negative sentiment often tracks with workload pressure.

The flight risk flag, which kicks in when someone sends four or more negative messages within thirty days, turned out to be a solid early warning sign. In many cases, the people who hit that threshold also had a dip in their monthly score during that same time. That shows the rule isn’t random. This actually aligns with meaningful shifts in behavior and tone. It’s a quick way to spot someone who might be hitting a breaking point before things escalate.

The predictive model also added some interesting insight. People who write in a steady rhythm and avoid sending too many messages late at night usually score higher in sentiment. On the flip side, those who message more after hours, or who show sudden bursts of negativity, tend to have lower scores. These patterns make sense when someone’s communication becomes inconsistent or starts drifting into late-night hours, that’s often a sign they’re feeling pressure or frustration.

**Recommendations:**

1. Set up a dashboard that tracks monthly sentiment scores and highlights the top and bottom performers. This helps teams catch any unusual dips early without having to go digging through raw messages.

2. Use rolling alerts to flag when someone is on track to hit the flight risk threshold. Even seeing three negative messages in a short window can be enough to prompt a check-in before things get worse.

3. Train managers to look beyond task performance and start paying attention to communication tone. If someone’s messages are becoming shorter, sharper, or more frequent outside of normal hours, it might be time for a one-on-one conversation.

4. Encourage healthier communication habits**.** Late-night emails, sharp responses, or short passive-aggressive notes can all add up. Coaching teams to pause and think about tone, especially during busy seasons, can make a difference in overall morale.

4. Keep building on the model. The current version uses basic features, but you can add more over time—things like topics being discussed, team changes, calendar events, or even meeting loads. These extra layers can help explain why someone’s sentiment is shifting, not just that it is.

5. Don’t forget human review. Models are helpful, but they’re not perfect. If someone gets flagged but doesn’t seem like a risk, it’s okay to take a second look manually. Having that balance between automation and human judgment keeps things fair and grounded.

# 10. Reproducibility & Environment

Code resides in the provided notebook. To reproduce results: (1) install dependencies (pandas, numpy, scikit‑learn, nltk/transformers as used), (2) run preprocessing and labeling cells, (3) run EDA, (4) compute monthly scores and rankings, (5) run flight‑risk identification, and (6) train/evaluate the linear model. All figures included in this document are generated by notebook cells and exported automatically.

# Visualizations (Selected)

Representative charts from the analysis are provided below. See Appendix A for the full set of generated figures.

Figure S1: image\_001.png

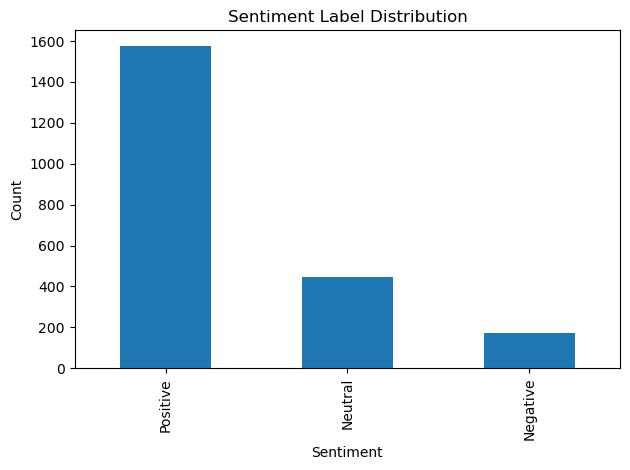


Figure S2: image\_002.png

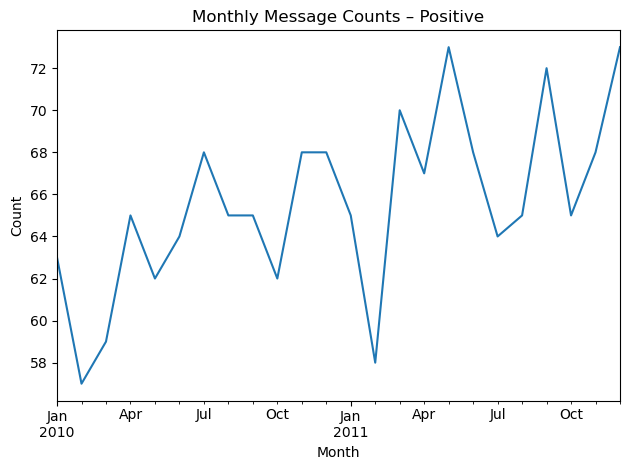


Figure S3: image\_003.png

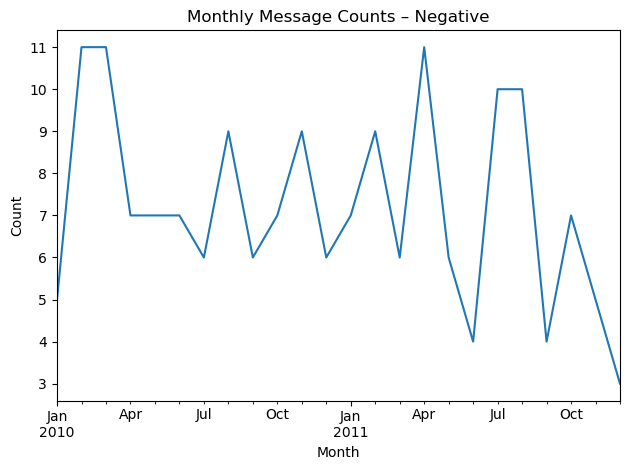


Figure S4: image\_004.png

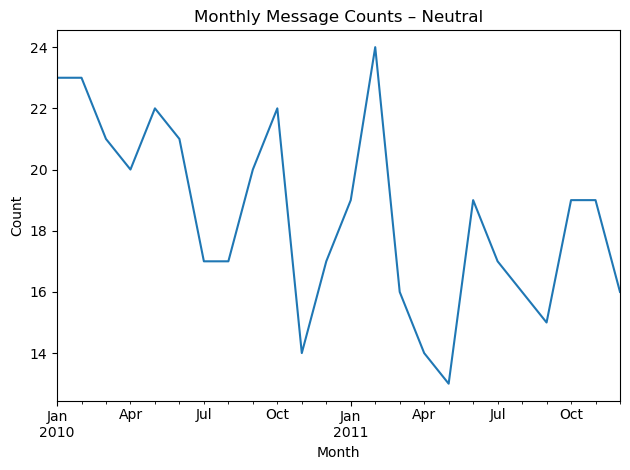


Figure S5: image\_005.png

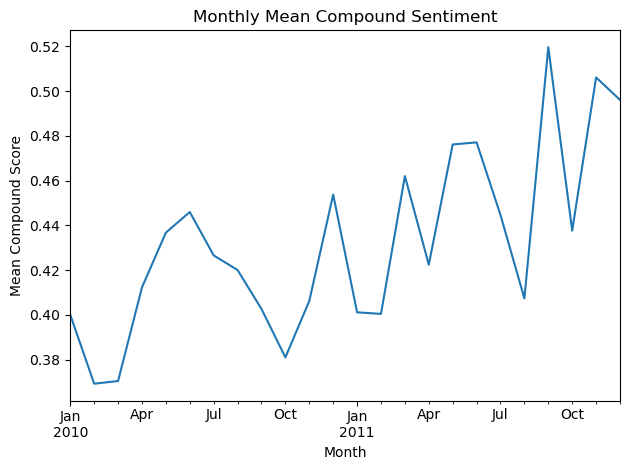
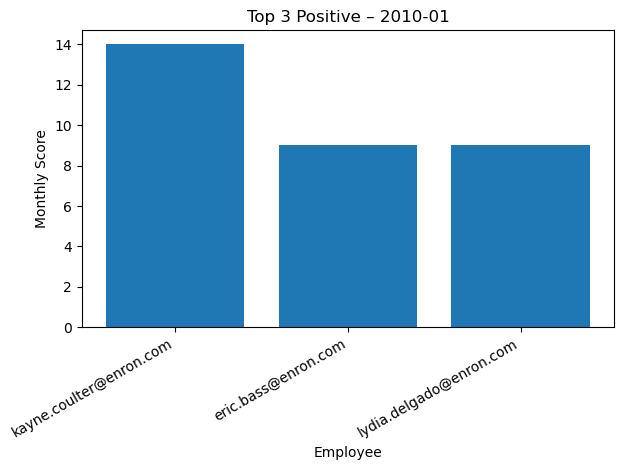


Figure S6: image\_006.png



# Appendix A – All Figures

Figure A1: image\_001.png

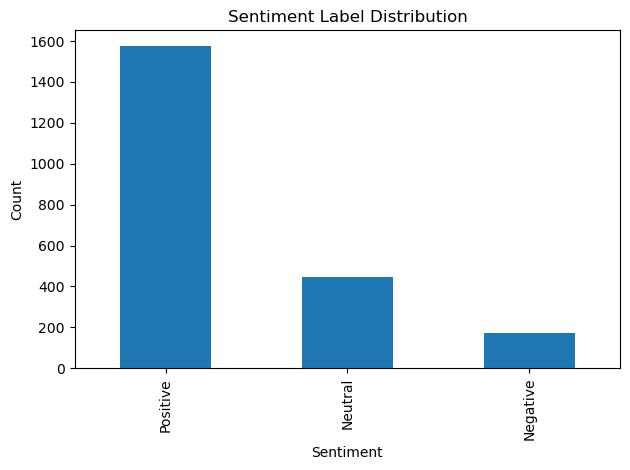


Figure A2: image\_002.png

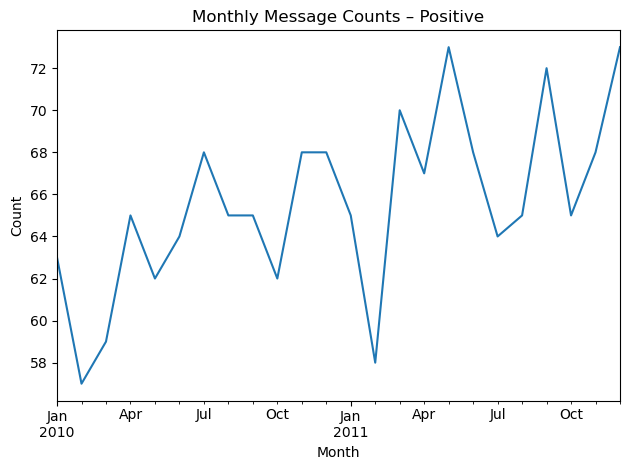


Figure A3: image\_003.png

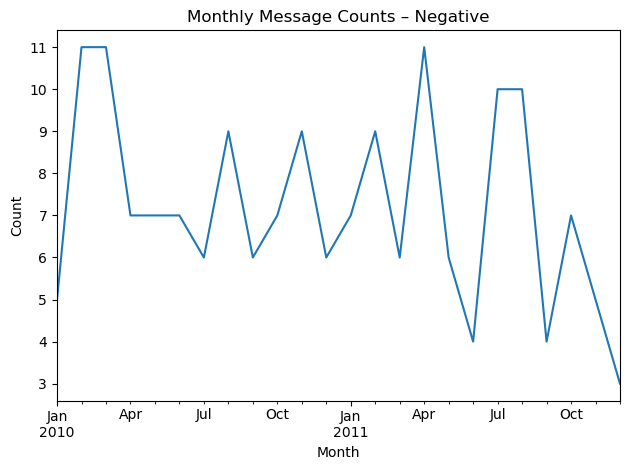


Figure A4: image\_004.png

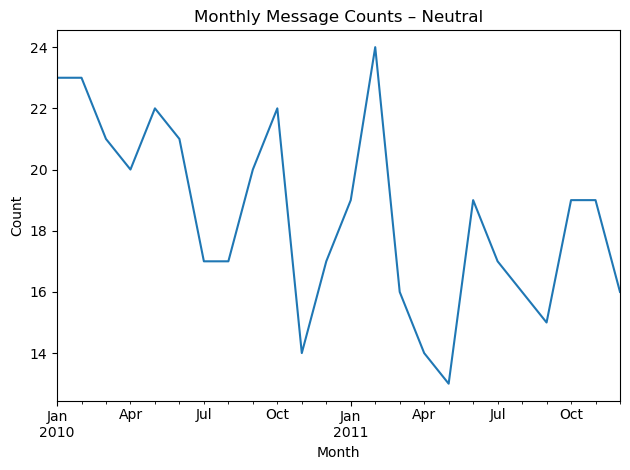


Figure A5: image\_005.png

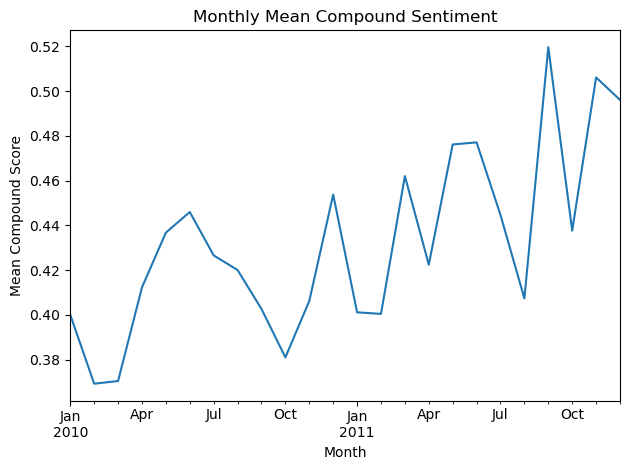


Figure A6: image\_006.png

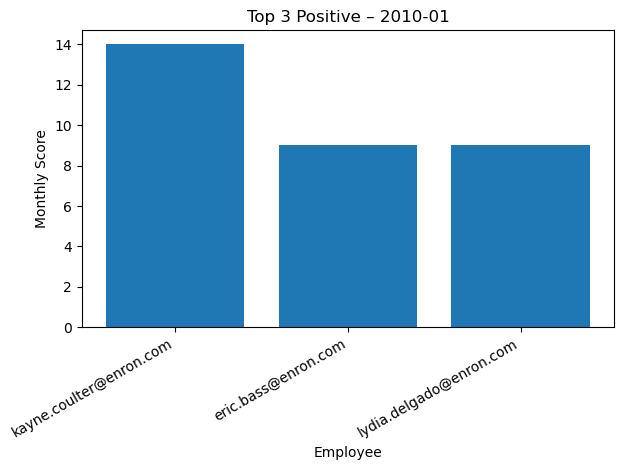


Figure A7: image\_007.png

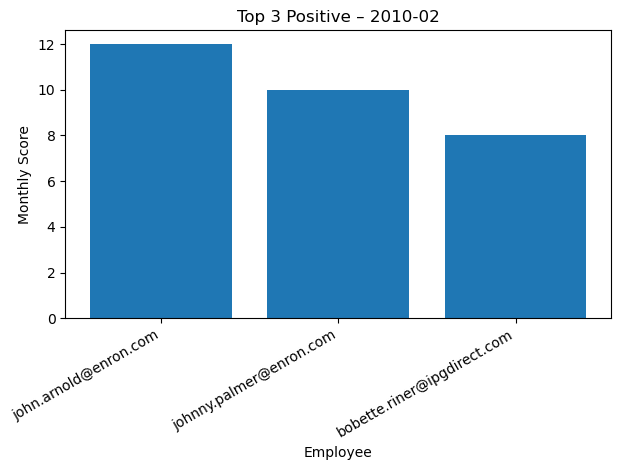


Figure A8: image\_008.png

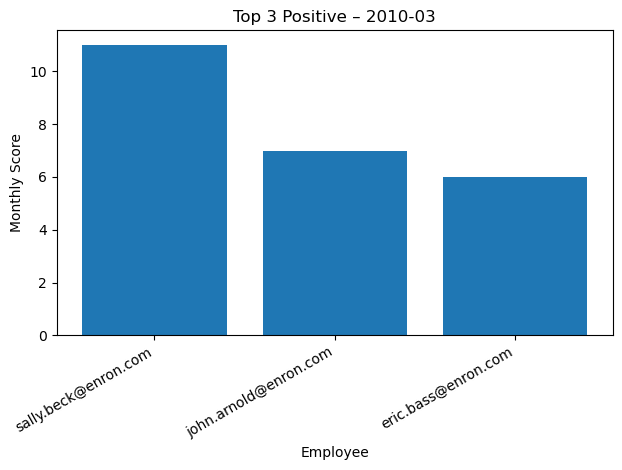


Figure A9: image\_009.png

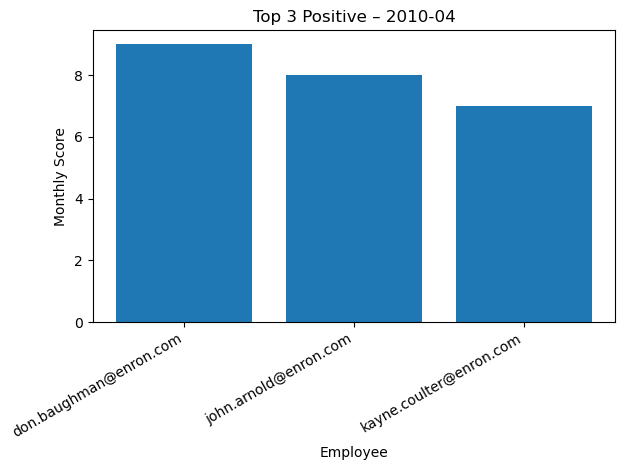


Figure A10: image\_010.png

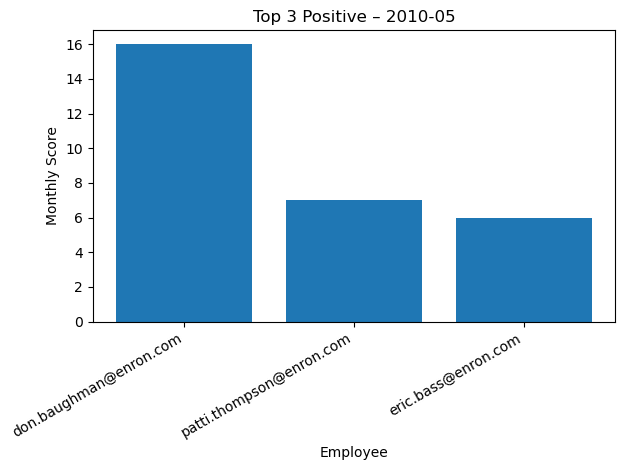


Figure A11: image\_011.png

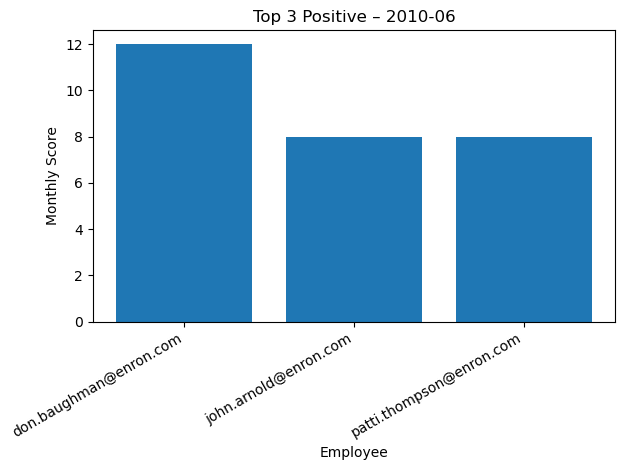


Figure A12: image\_012.png

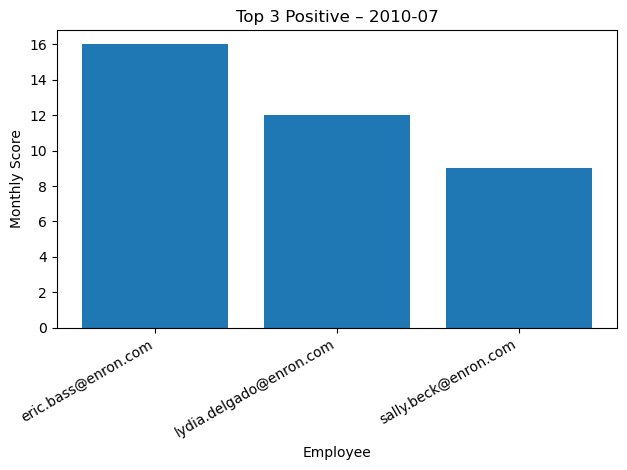


Figure A13: image\_013.png

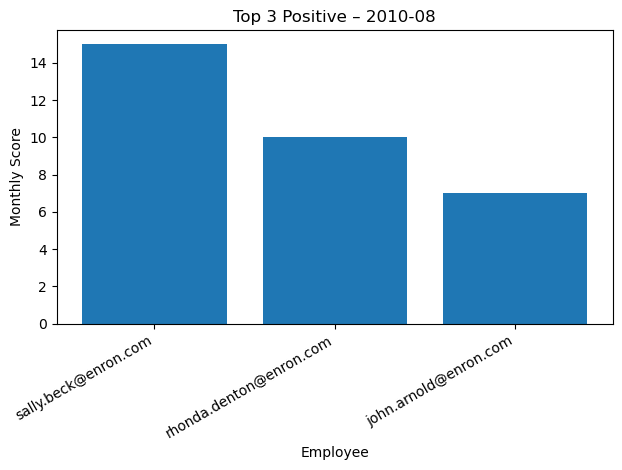


Figure A14: image\_014.png

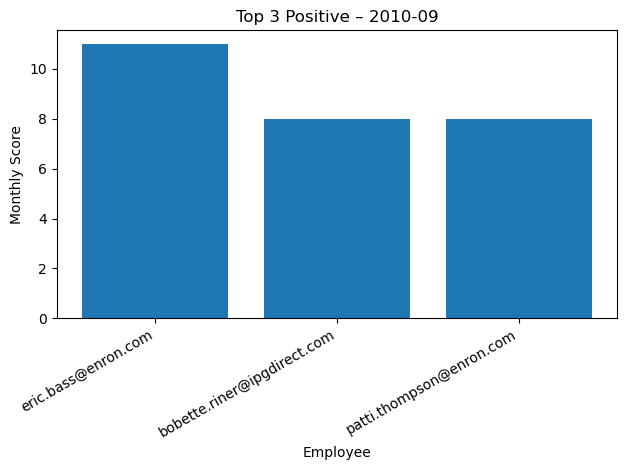


Figure A15: image\_015.png

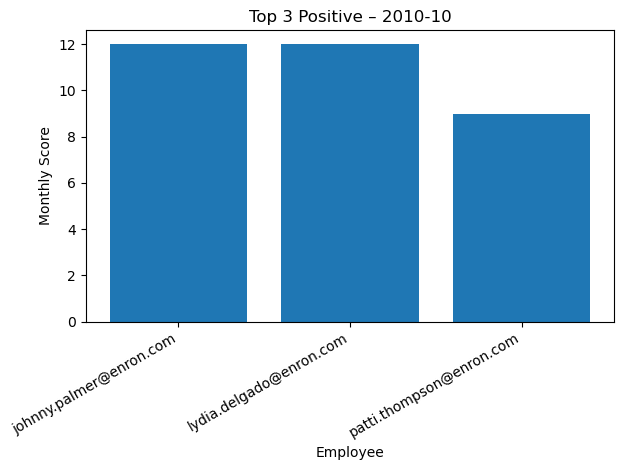


Figure A16: image\_016.png

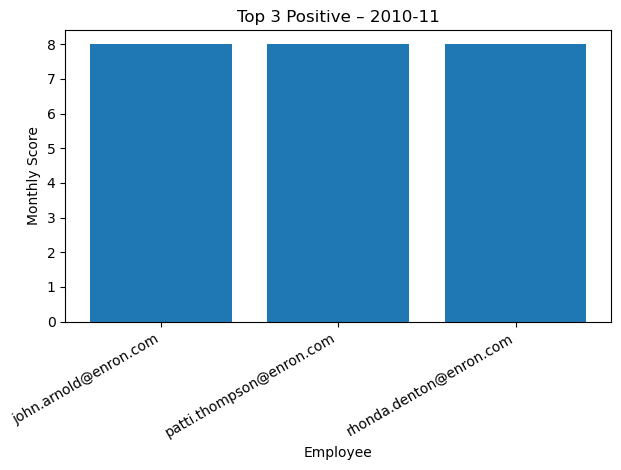


Figure A17: image\_017.png

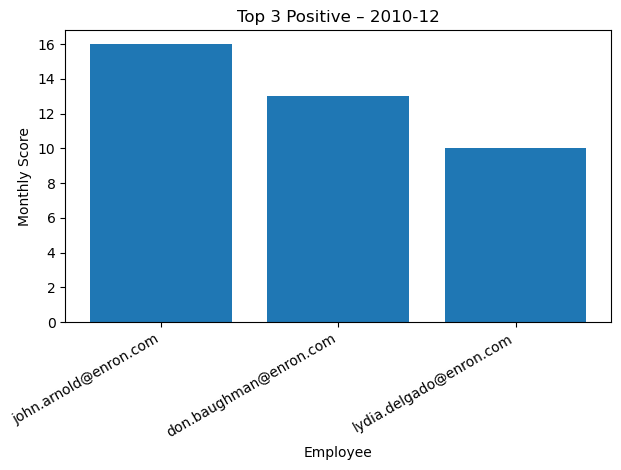


Figure A18: image\_018.png

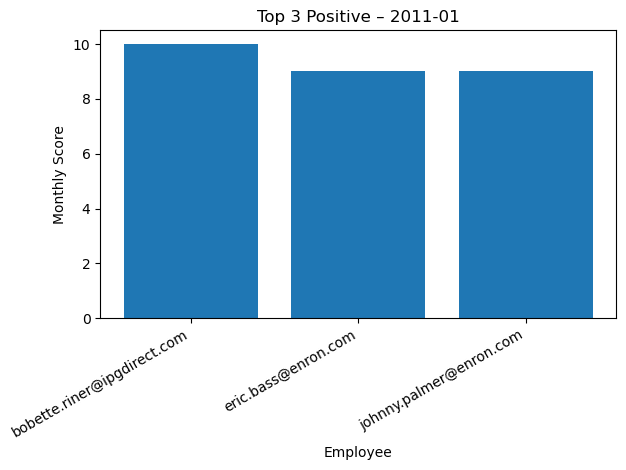


Figure A19: image\_019.png

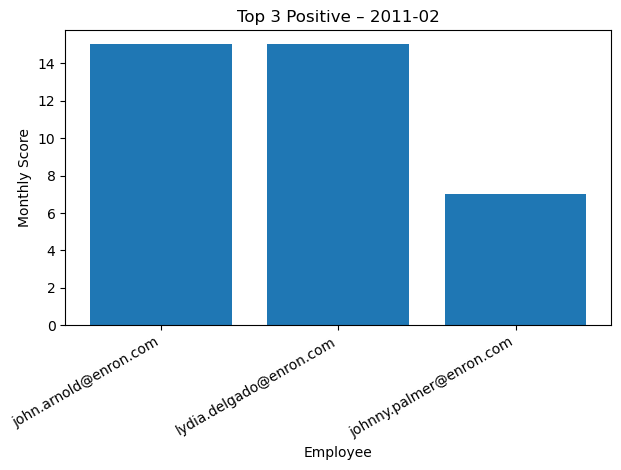


Figure A20: image\_020.png

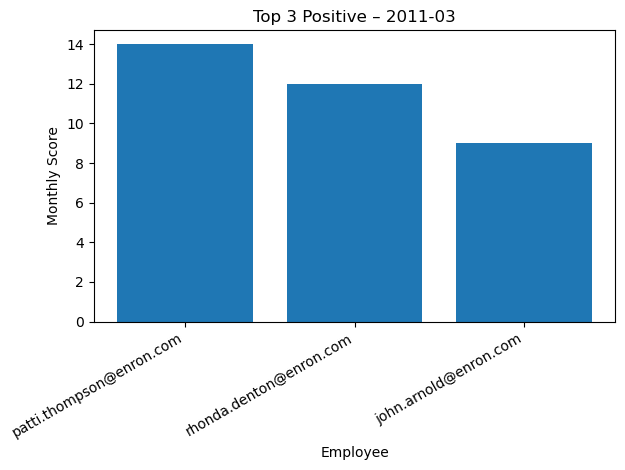


Figure A21: image\_021.png

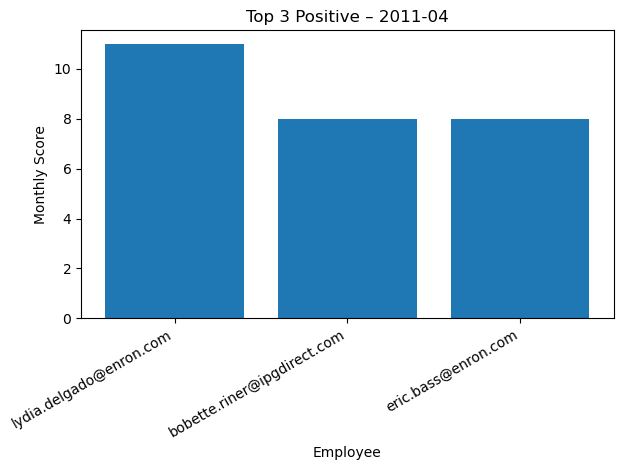


Figure A22: image\_022.png

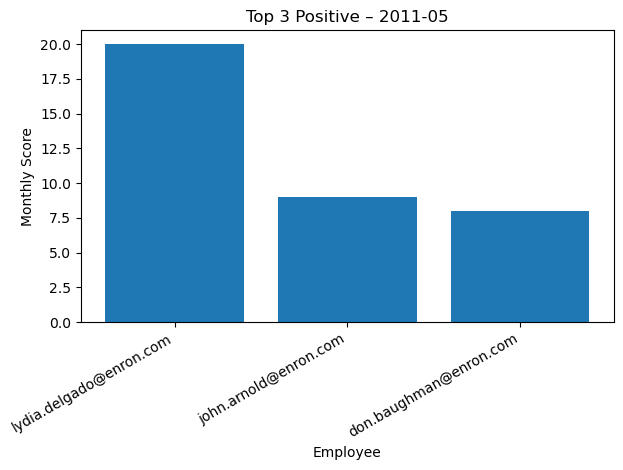


Figure A23: image\_023.png

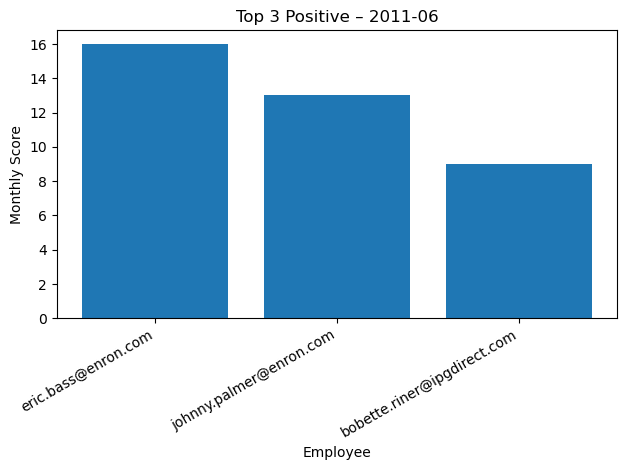


Figure A24: image\_024.png

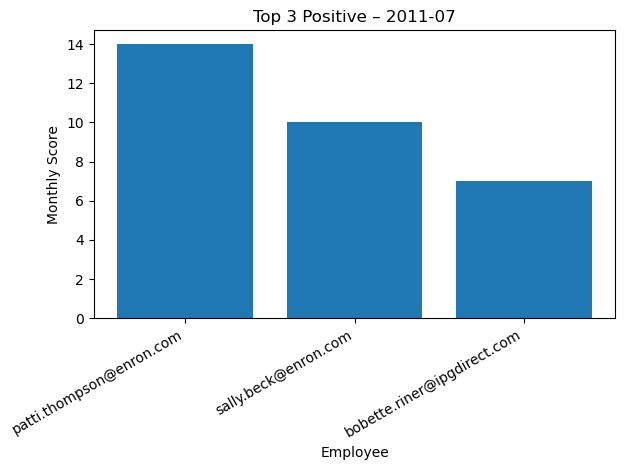


Figure A25: image\_025.png

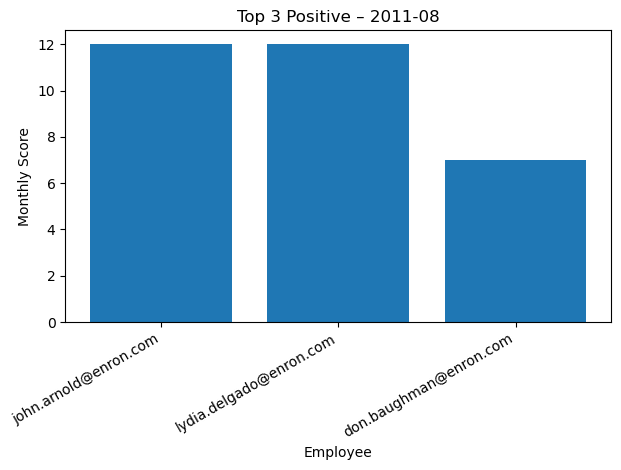


Figure A26: image\_026.png

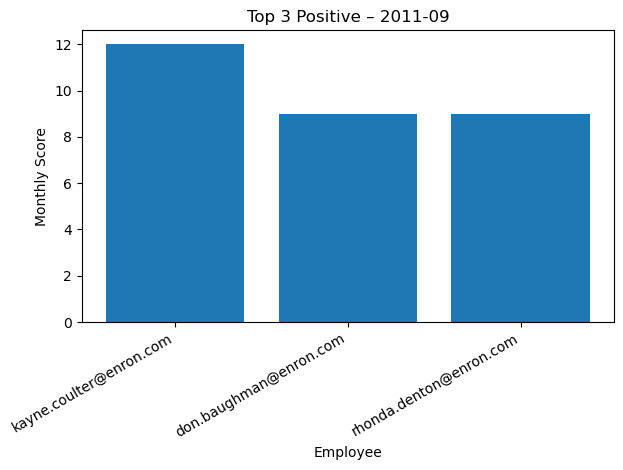


Figure A27: image\_027.png

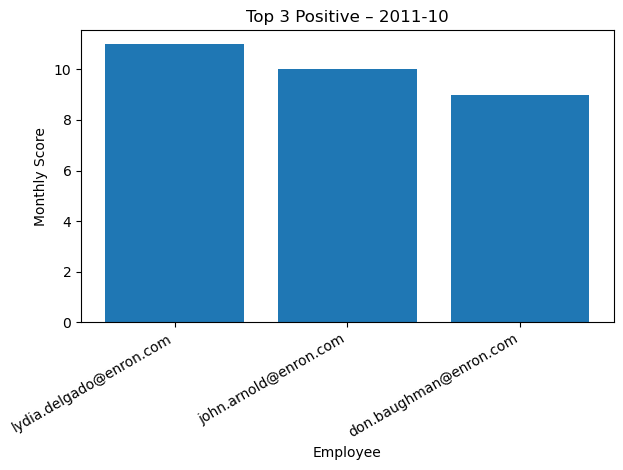


Figure A28: image\_028.png

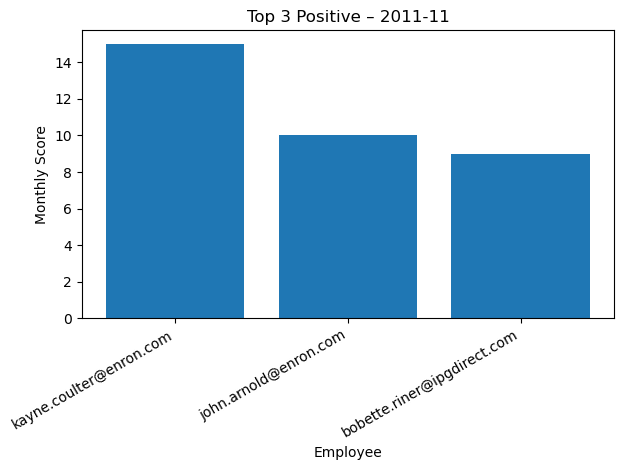


Figure A29: image\_029.png

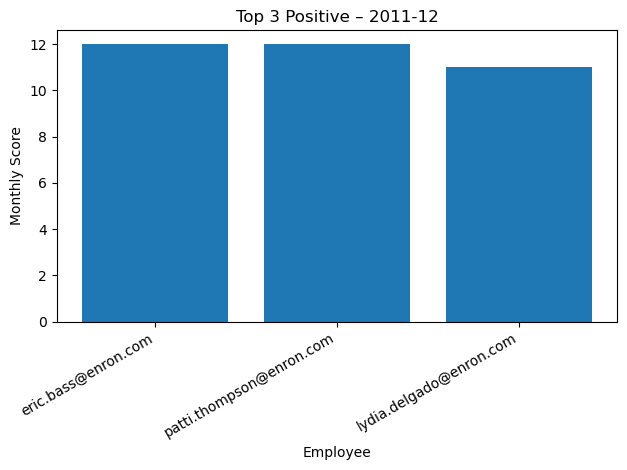


Figure A30: image\_030.png

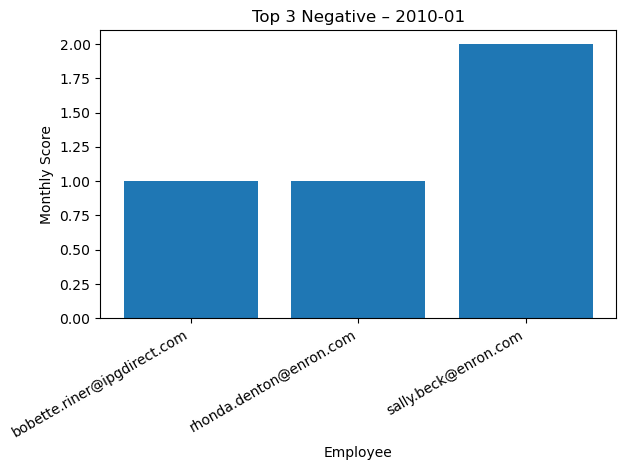


Figure A31: image\_031.png

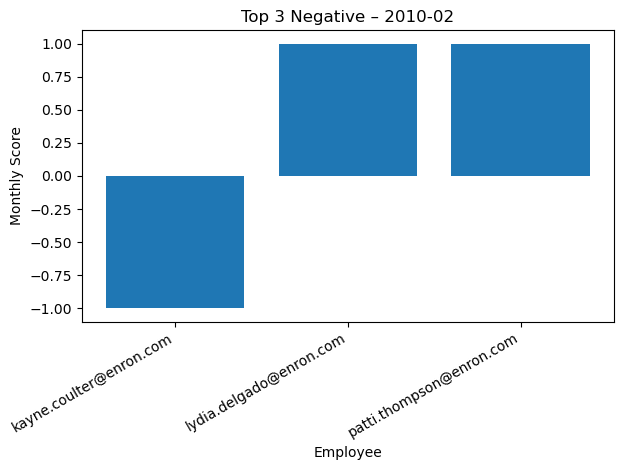


Figure A32: image\_032.png

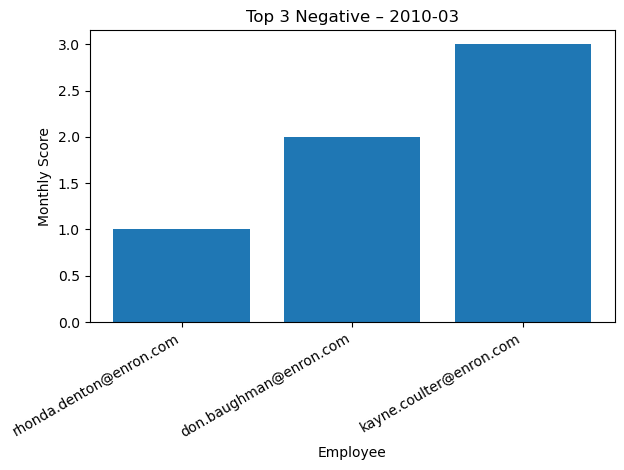


Figure A33: image\_033.png

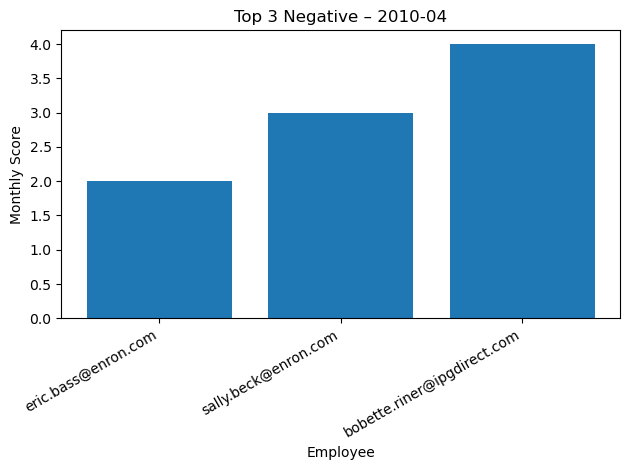


Figure A34: image\_034.png

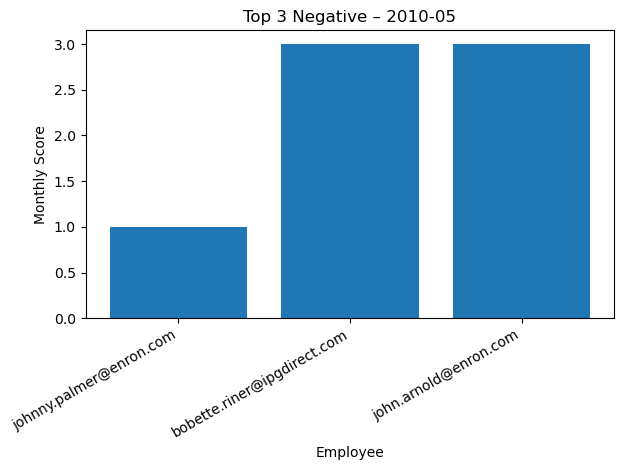


Figure A35: image\_035.png

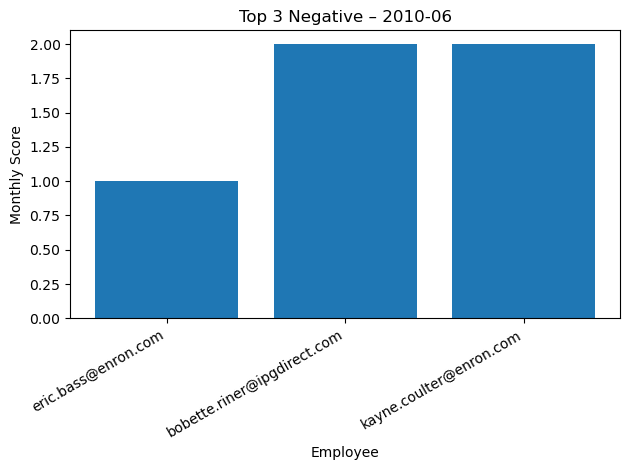


Figure A36: image\_036.png

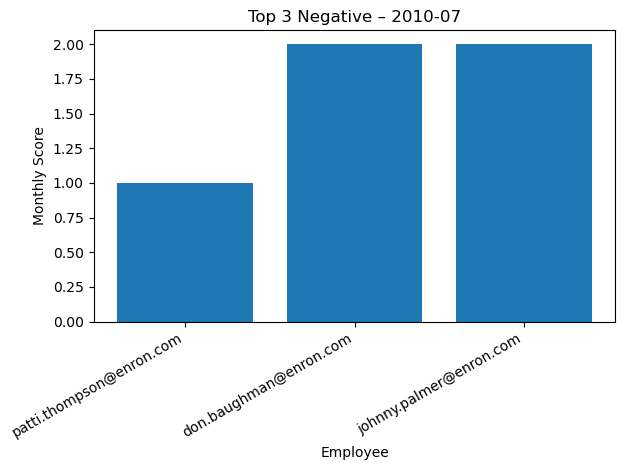


Figure A37: image\_037.png

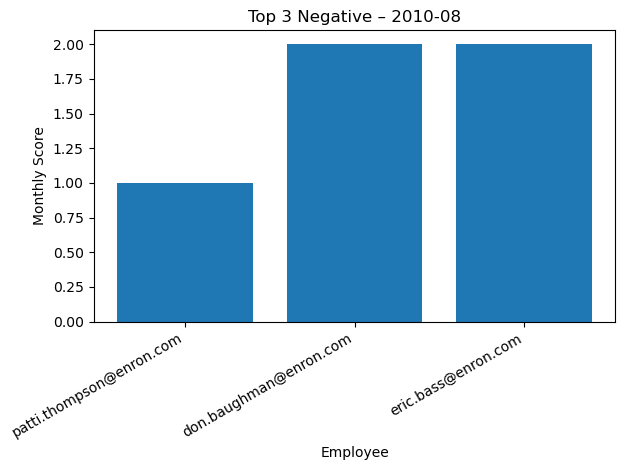


Figure A38: image\_038.png

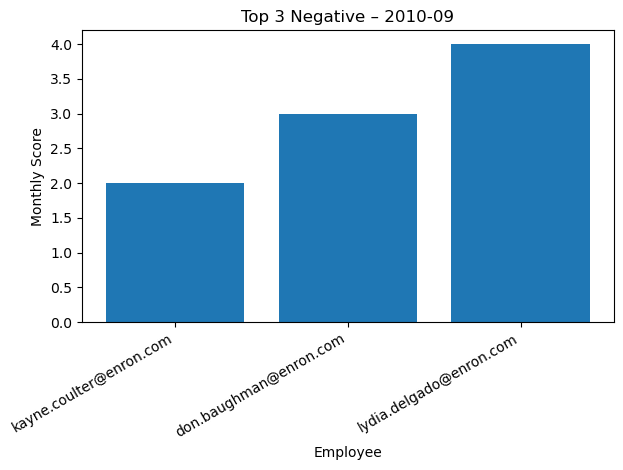


Figure A39: image\_039.png

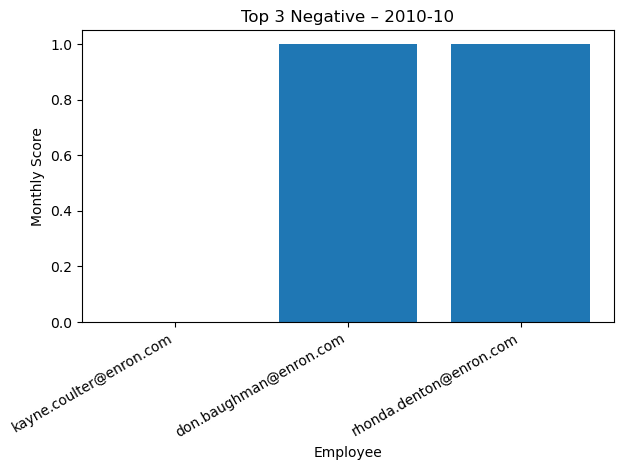


Figure A40: image\_040.png

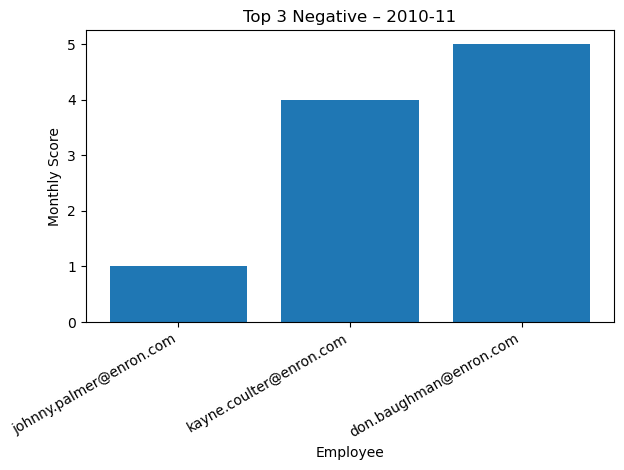


Figure A41: image\_041.png

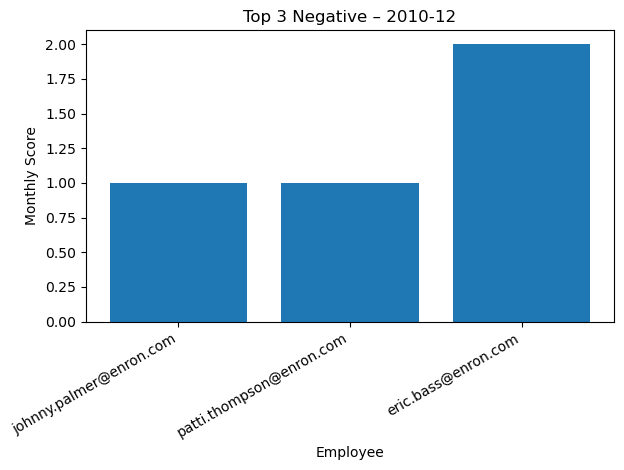


Figure A42: image\_042.png

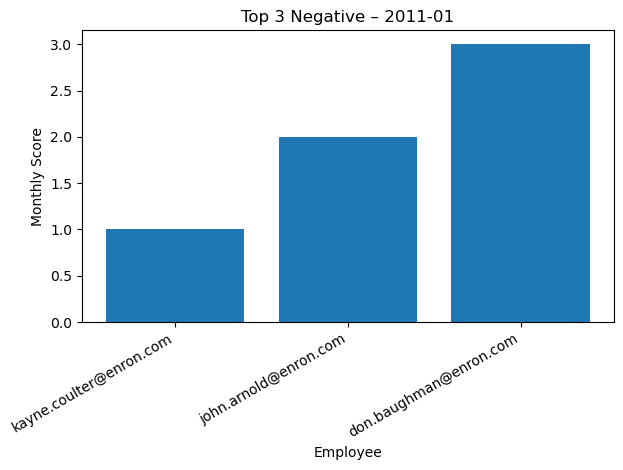


Figure A43: image\_043.png

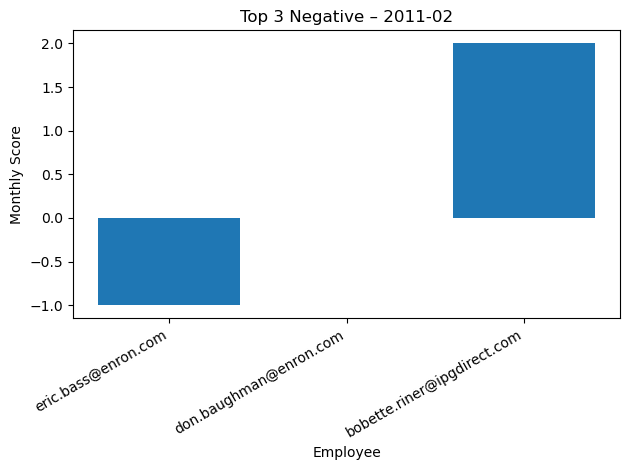


Figure A44: image\_044.png

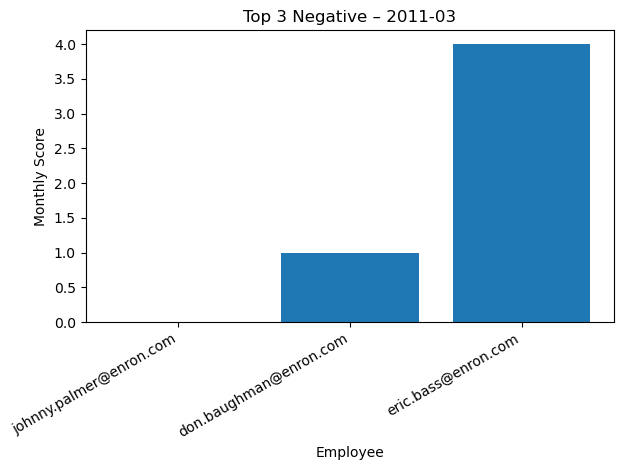


Figure A45: image\_045.png

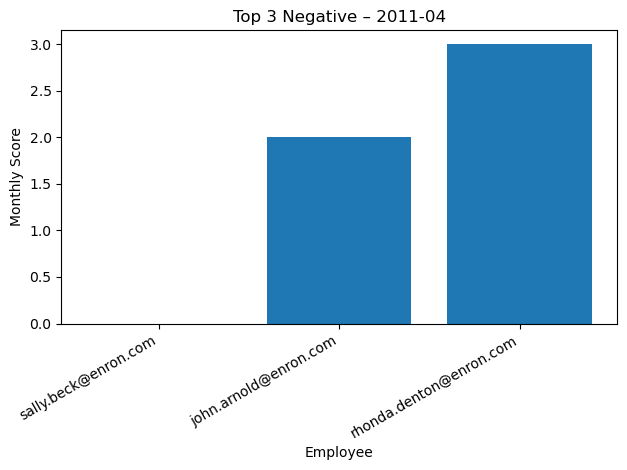


Figure A46: image\_046.png

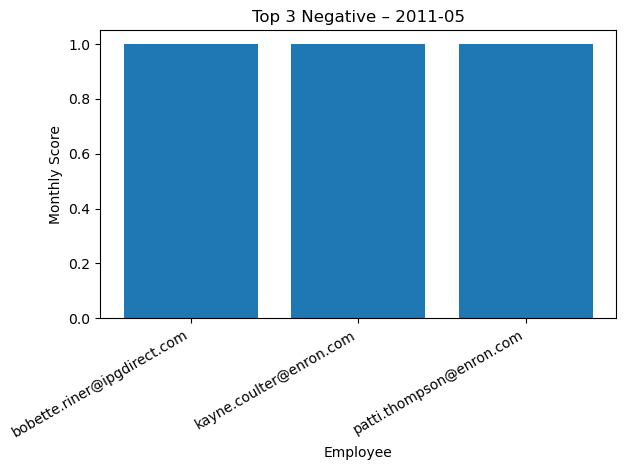


Figure A47: image\_047.png

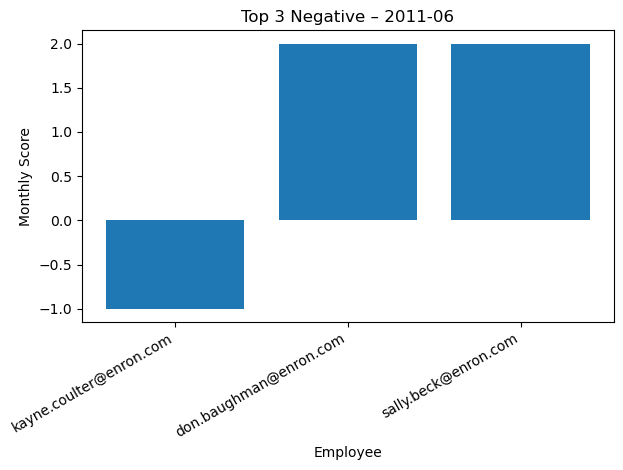


Figure A48: image\_048.png

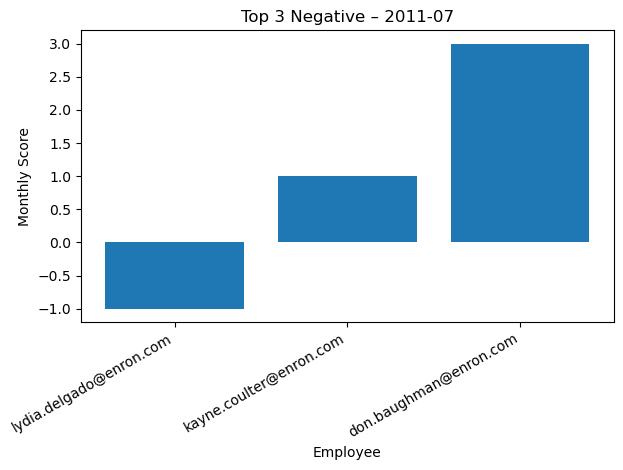


Figure A49: image\_049.png

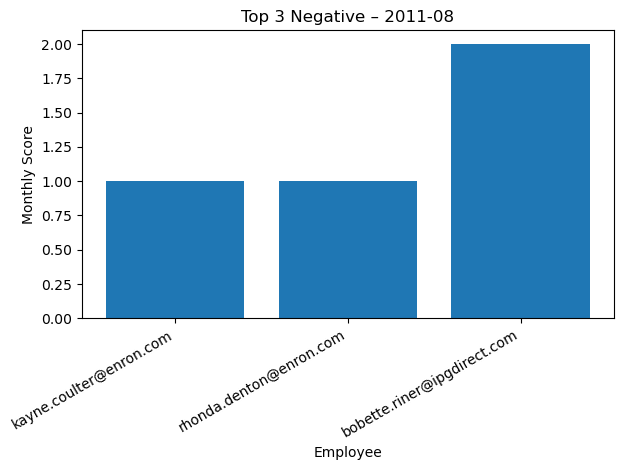


Figure A50: image\_050.png

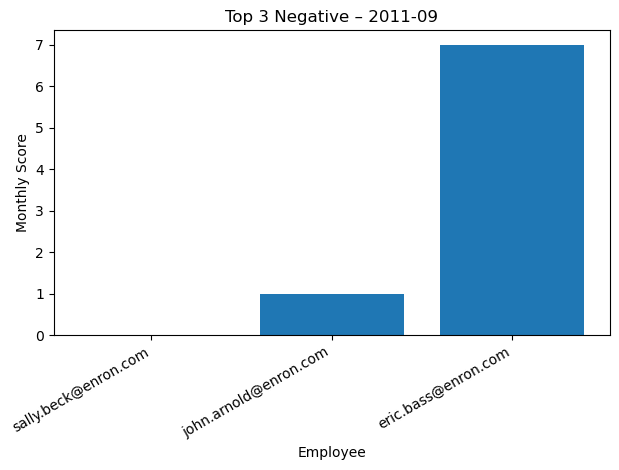


Figure A51: image\_051.png

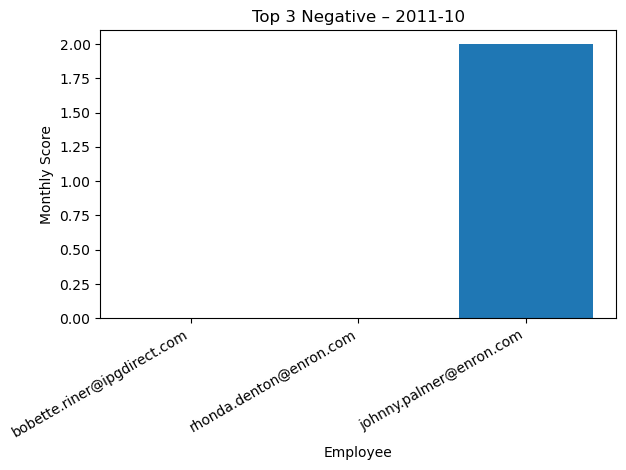


Figure A52: image\_052.png

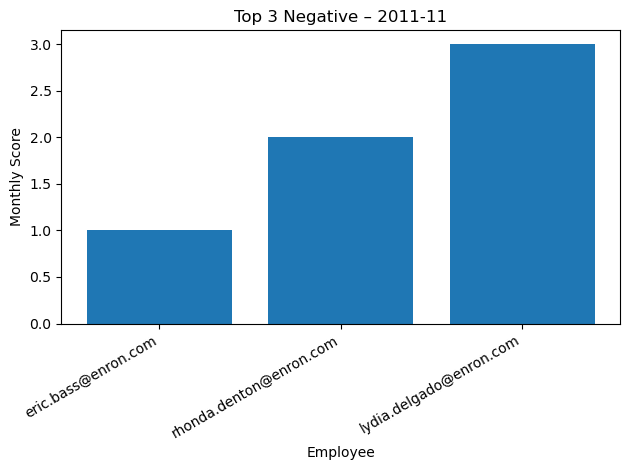


Figure A53: image\_053.png

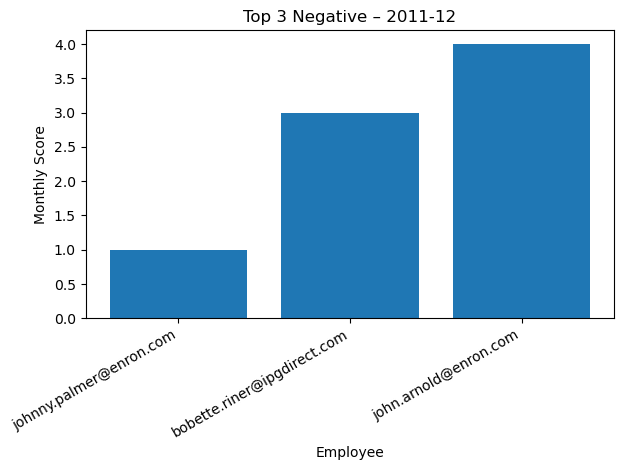


Figure A54: image\_054.png

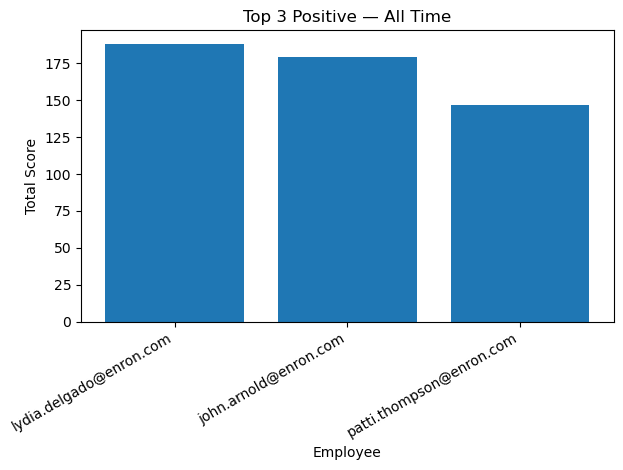


Figure A55: image\_055.png

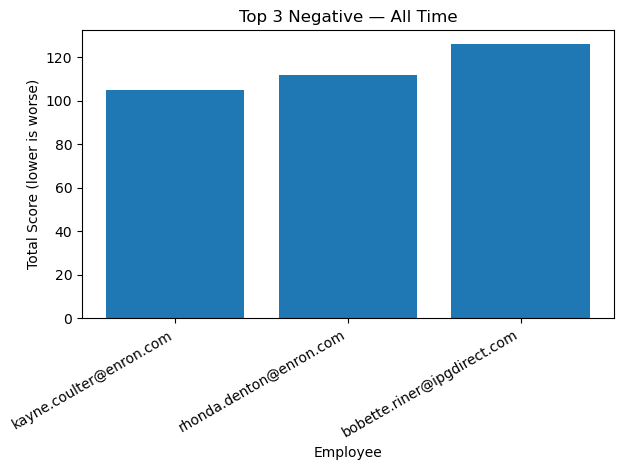


Figure A56: image\_056.png

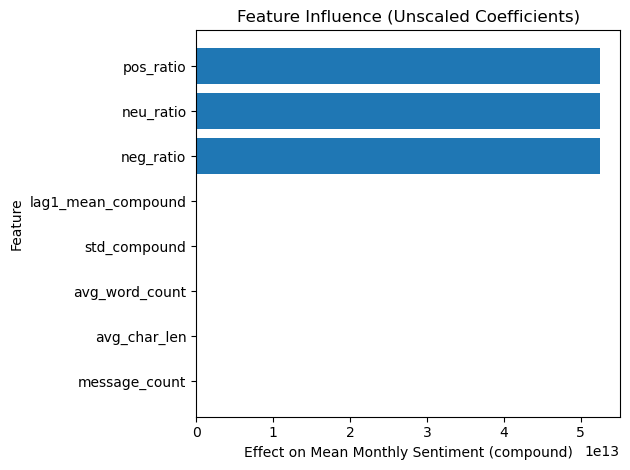


Figure A57: image\_057.png

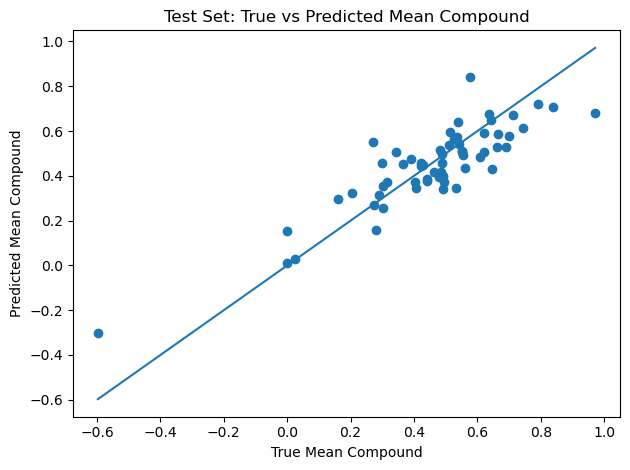


Figure A58: image\_058.png

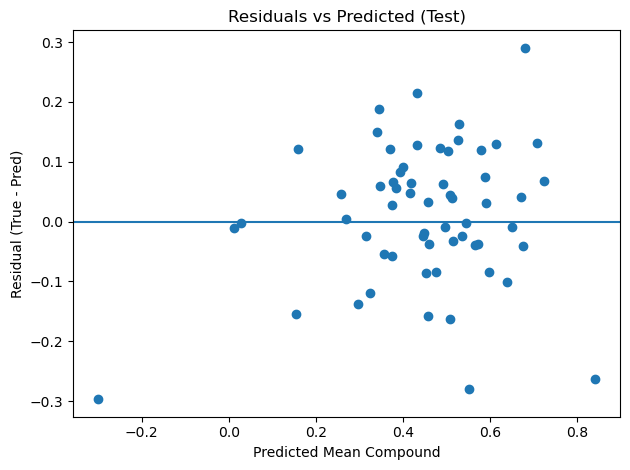


Figure A59: image\_059.png

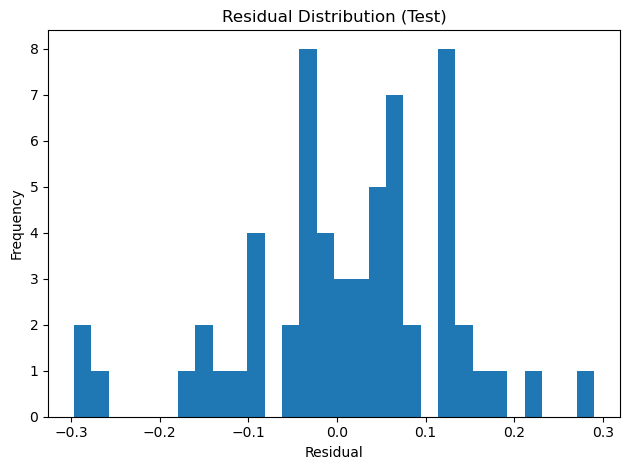
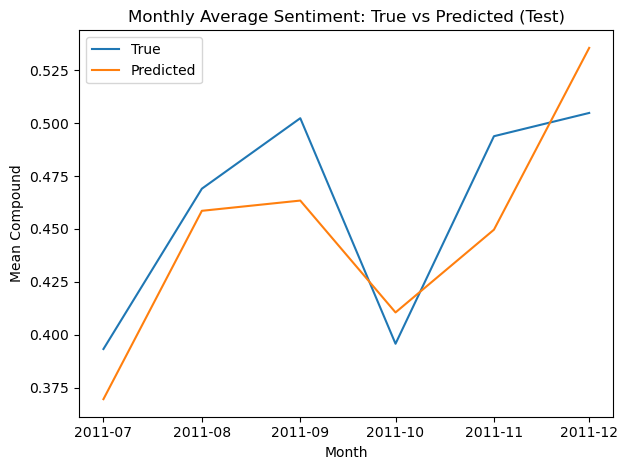


Figure A60: image\_060.png



# Appendix B – Tables (Fill with Results)

Table B1. Top Three Positive & Negative Employees (per Month)

[Add month-by-month tables with Top 3 Positive and Top 3 Negative employees here.]

Table B2. Flight Risk Employees

[List employees meeting ≥4 negative messages in any 30-day window, with date ranges.]

Table B3. Linear Model Metrics

R² (test): [ ] MAE: [ ] RMSE: [ ]

Table B4. Linear Model Coefficients (Top 10 by absolute value)

[Feature] [Coefficient] [Direction (+/−)] [Interpretation]