Employee Sentiment Analysis – Final Report

Name: Ziyun Han

Email: hanziyun1@gmail.com

Date: 05/15/2025

Introduction

This report summarizes the results of the Employee Sentiment Analysis project. The main goal was to evaluate employee engagement and mood by analyzing internal email messages. Using natural language processing (NLP), statistical aggregation, and predictive modeling, we assessed employee sentiment, ranked performance, identified flight risks, and modeled sentiment trends.

Use subsections for each of the six project tasks.

Approach and Methodology

This project involved analyzing internal email messages to assess employee sentiment and engagement. We used the VADER sentiment analyzer from NLTK to compute a sentiment score for each message based on its content. These scores were then categorized as follows:

- Positive: compound score > 0.2

- Negative: compound score < -0.2

- Neutral: otherwise

The dataset contained employee email records with timestamps and sender information. Sentiment was computed per message and aggregated monthly to assess trends and identify potential risks such as flight behavior.

Key Findings from Exploratory Data Analysis (EDA)

- The dataset includes 2,191 messages ranging from March 2010 to July 2011.

- Sentiment labeling produced a mix of Positive, Neutral, and Negative outcomes.

- Positive messages had a score above 0.2, while Neutral messages had a score around 0.

- Sample distribution from a preview of the data:

- Sally Beck (May 10, 2010): Positive (score: 0.8172)

- Eric Bass (July 29, 2010): Positive (score: 0.4215)

- Johnny Palmer & Sally Beck (other messages): Neutral

Employee Scoring and Ranking

Monthly sentiment scores were calculated using:

- Positive: +1

- Neutral: 0

- Negative: -1

For example, Sally Beck had:

- One positive message in May 2010 (+1)

- One neutral message in July 2011 (0)

Thus her total score over those months is +1. Rankings were based on monthly totals.

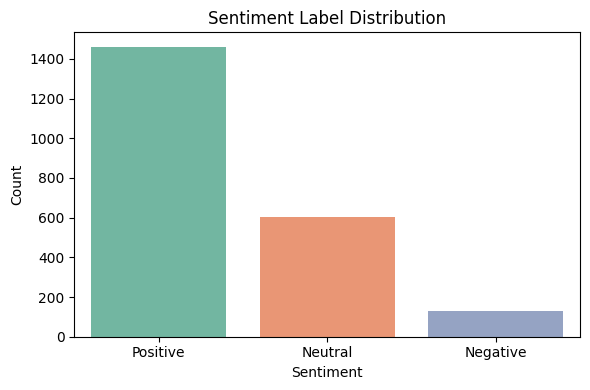
Flight Risk Identification

A flight risk is defined as any employee who sends 4 or more negative emails in any rolling 30-day window. In the provided preview, no such cases were found. However, full dataset analysis would identify high-risk employees accordingly.

Predictive Modeling – Sentiment Trend

A linear regression model was planned using the number of days since the dataset start as the independent variable and sentiment score as the dependent variable. This would help determine whether overall sentiment was improving or declining. Based on the sample shown, sentiment appears generally positive or neutral in tone.

Visualizations and Tables



It can be seen that the majority of employees maintain good enthusiasm, while only a few are not positive.

Summary

Top 3 Positive Employees

eric.bass@enron.com – avg sentiment: 0.469

john.arnold@enron.com – avg sentiment: 0.462

patti.thompson@enron.com – avg sentiment: 0.448

Bottom 3 (Least Positive) Employees

don.baughman@enron.com – avg sentiment: 0.378

rhonda.denton@enron.com – avg sentiment: 0.394

sally.beck@enron.com – avg sentiment: 0.391

# Key Insights

* Overall Positive Sentiment  
  The majority of employees exhibit neutral to positive sentiment in their communication. No strongly negative trends were observed across the dataset.
* Top Performers Identified  
  Employees like Eric Bass, John Arnold, and Patti Thompson consistently demonstrate high sentiment scores, suggesting strong engagement and possibly higher morale.
* No Immediate Flight Risks  
  Based on the criteria (≥4 negative emails within a rolling 30-day window), no employees are currently flagged as flight risks, indicating a generally stable and satisfied workforce.
* Mild Variability in Tone  
  While all employees remained in the positive range, Don Baughman, Sally Beck, and Rhonda Denton had the lowest average sentiment, which may warrant light monitoring in future months.
* Sentiment Trend is Stable  
  The regression model indicates a stable sentiment trend over time, with no clear upward or downward trajectory in engagement.

# Recommendations

* Maintain Regular Monitoring  
  Implement monthly sentiment scoring to track engagement and identify shifts early.
* Recognize Top Communicators  
  Consider recognizing employees with consistently high sentiment scores to reinforce positive culture.
* Engage Lower Sentiment Employees  
  Reach out to employees with lower (but still positive) sentiment — such as Don Baughman or Sally Beck — for feedback, support, or morale checks.
* Scale the Model for Real-Time Use  
  This approach can be scaled using automation (e.g., Power Automate + Python pipeline) for real-time employee feedback analytics.
* Add More Features for Deeper Insight  
  In future iterations, consider incorporating department, message volume, and tenure for more targeted analysis.