

Employee Sentiment Analysis Report

Pragyan Borthakur

June 1, 2025

Contents

1	Introduction	2
2	Approach and Methodology	2
3	Exploratory Data Analysis (EDA)	2
3.1	Correlation Heatmap	2
3.2	Message Trends Over Time	3
3.3	Message Length Distribution	3
3.4	Sentiment Label Distribution	4
3.5	Monthly Sentiment Proportions	4
4	Employee Scoring and Ranking	4
4.1	Scoring	4
4.2	Ranking	5
4.3	Top Positive and Negative Employees	5
4.4	Sample of Rankings (2010-01)	5
5	Flight Risk Identification	6
5.1	Criteria	6
5.2	Flight Risk Summary	6
5.3	First Flag Dates	6
6	Predictive Modeling	6
6.1	Model Overview	6
6.2	Model Evaluation	7
6.3	Model Coefficients	7
6.4	Prediction Accuracy	7

1 Introduction

This report presents an analysis of employee sentiment extracted from email communications using the TextBlob NLP library. The study includes sentiment classification, exploratory data analysis (EDA), employee ranking, flight risk detection, and sentiment prediction using regression modeling.

2 Approach and Methodology

- **Sentiment Analysis:** TextBlob was used to assign sentiment polarity scores to each email. Polarity values were classified as Negative (polarity < 0), Neutral ($= 0$), and Positive (> 0).
- **Data Preparation:** Emails were grouped by employee and date to compute monthly metrics.
- **Scoring:** Sentiments were scored as -1 (Negative), 0 (Neutral), 1 (Positive). Monthly sentiment scores were calculated per employee.
- **EDA:** Analyzed sentiment distributions, correlations between features, and time-based trends.
- **Flight Risk Identification:** Employees with multiple negative messages in a 30-day window were flagged.
- **Predictive Modeling:** A linear regression model was trained using message-level features (e.g., message count, average length).

3 Exploratory Data Analysis (EDA)

3.1 Correlation Heatmap

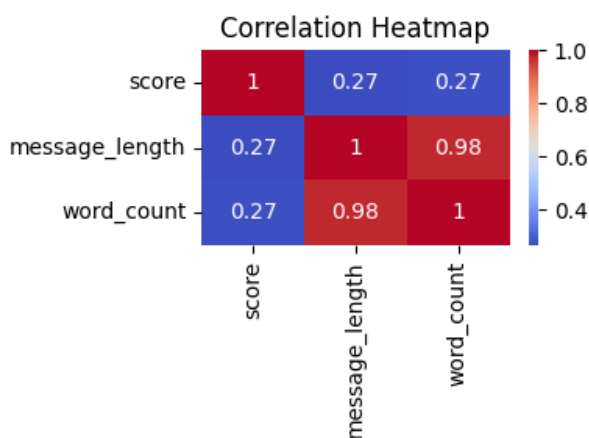


Figure 1: Correlation between score, message length, and word count

3.2 Message Trends Over Time

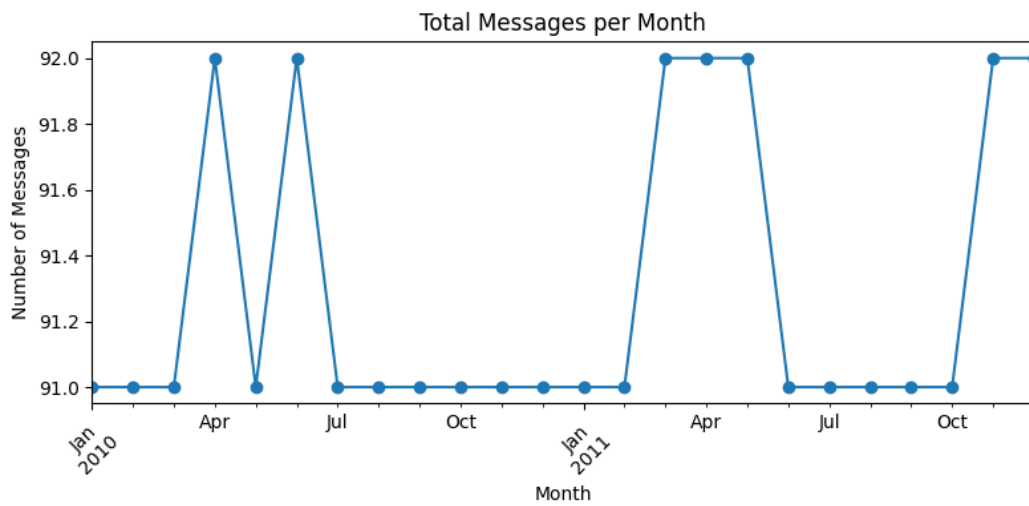


Figure 2: Total Messages per Month

3.3 Message Length Distribution

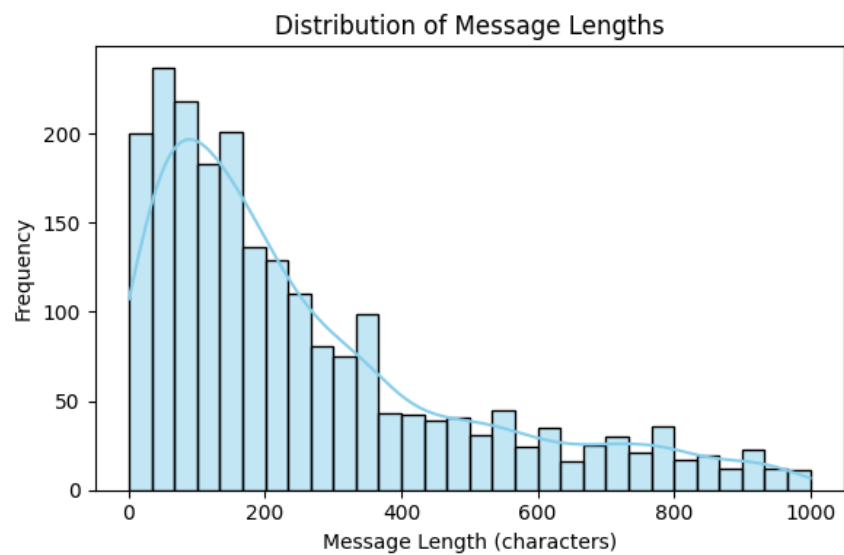


Figure 3: Distribution of Message Lengths

3.4 Sentiment Label Distribution

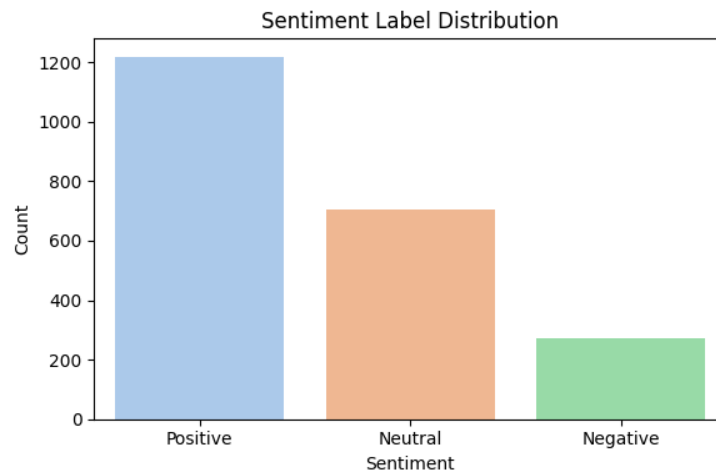


Figure 4: Overall Sentiment Distribution

3.5 Monthly Sentiment Proportions

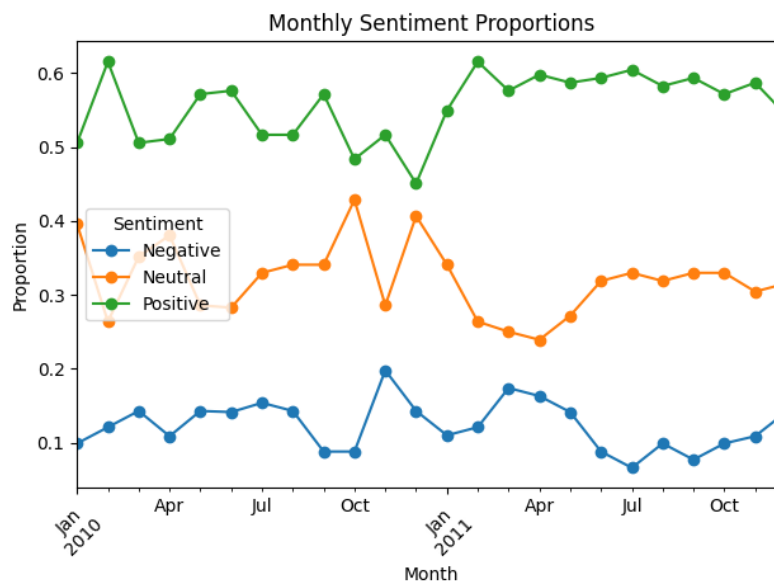


Figure 5: Monthly Sentiment Proportions

4 Employee Scoring and Ranking

4.1 Scoring

Each message sentiment was scored as follows:

- Positive = 1

- Neutral = 0
- Negative = -1

Monthly sentiment scores per employee were aggregated.

4.2 Ranking

Employees were ranked each month based on their total sentiment score.

4.3 Top Positive and Negative Employees

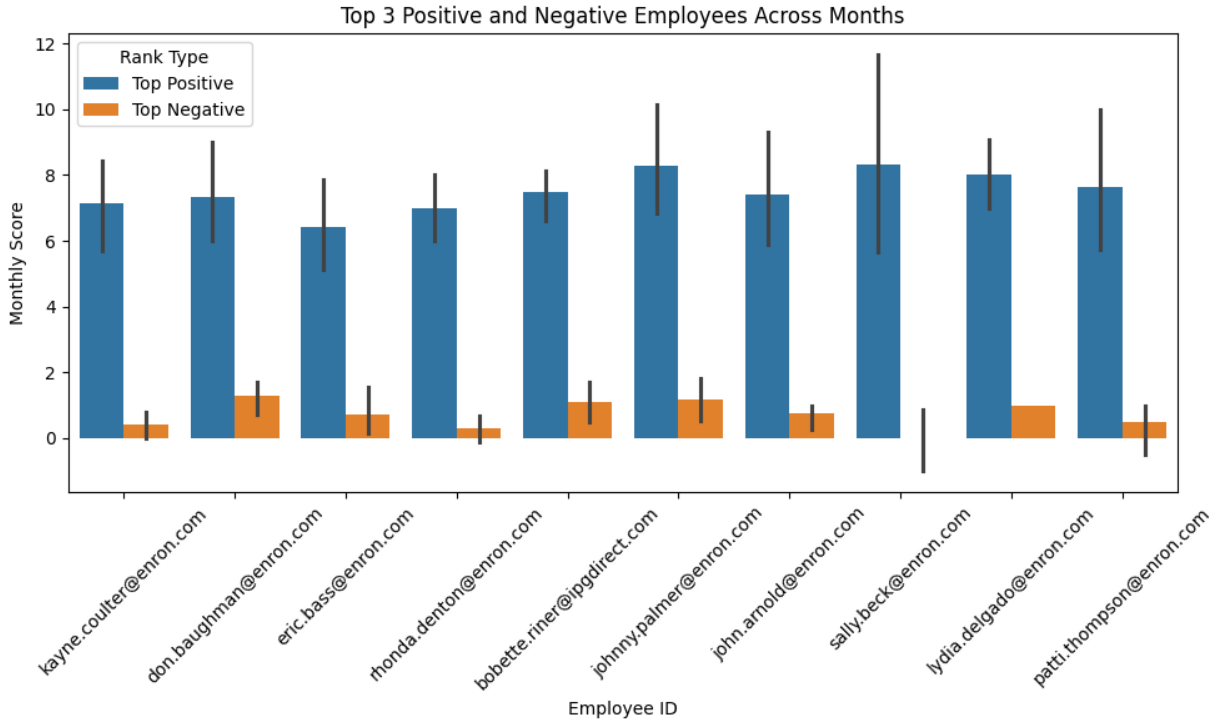


Figure 6: Top 3 Positive and Negative Employees by Month

4.4 Sample of Rankings (2010-01)

Month	Employee	Score	Rank Type
2010-01	kayne.coulter@enron.com	9	Top Positive
2010-01	don.baughman@enron.com	5	Top Positive
2010-01	eric.bass@enron.com	5	Top Positive
2010-01	rhonda.denton@enron.com	0	Top Negative
2010-01	bobette.riner@ipgdirect.com	2	Top Negative

Table 1: Employee Rankings (January 2010)

5 Flight Risk Identification

5.1 Criteria

Employees were flagged if they had at least 4 negative messages in a 30-day period.

5.2 Flight Risk Summary



Figure 7: Employees Flagged as Flight Risk per Month

5.3 First Flag Dates

Employee ID	First Flagged Date
bobette.riner@ipgdirect.com	2010-11-17
don.baughman@enron.com	2010-03-07
eric.bass@enron.com	2010-04-15
john.arnold@enron.com	2010-06-18
johnny.palmer@enron.com	2010-02-26
lydia.delgado@enron.com	2010-11-25
patti.thompson@enron.com	2010-05-02
rhonda.denton@enron.com	2010-08-28
sally.beck@enron.com	2010-07-

Table 2: Employees Flagged as Flight Risk

6 Predictive Modeling

6.1 Model Overview

A linear regression model was built using:

- `message_count`

- avg_length
- avg_word_count

6.2 Model Evaluation

- Mean Squared Error (MSE): 4.206
- R-squared (R^2): 0.641

6.3 Model Coefficients

Feature	Coefficient
message_count	0.4097
avg_length	-0.0054
avg_word_count	0.0506

Table 3: Linear Regression Coefficients

6.4 Prediction Accuracy

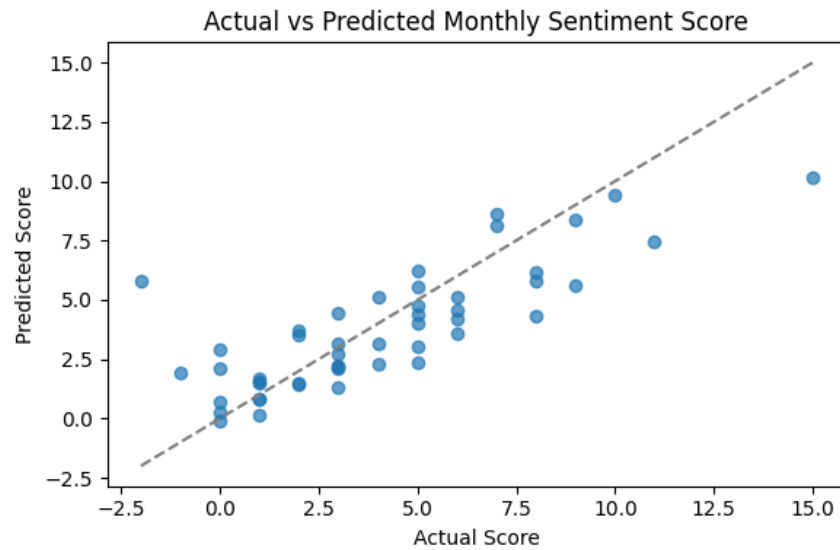


Figure 8: Actual vs Predicted Monthly Sentiment Score