# Final Report: Employee Sentiment Analysis

## 1. Project Overview

This project involved analyzing employee messages from an unlabeled dataset to assess sentiment and engagement. The pipeline included sentiment labeling, exploratory analysis, score aggregation, employee ranking, flight risk identification, and predictive modeling.

---

## 2. Sentiment Labeling

We used a pre-trained sentiment analysis model (e.g., VADER or transformer) to label each message in the dataset as \*\*Positive\*\*, \*\*Negative\*\*, or \*\*Neutral\*\* based on message content. The sentiment scores were stored in a new column named `Sentiment`.

---

## 3. Exploratory Data Analysis (EDA)

- Total messages analyzed: 2,191

- Unique employees: 10

- Sentiment labels distribution was visualized.

- Monthly message activity was tracked and plotted.

Key Insight: Messaging activity varied significantly across months, hinting at differing engagement levels.

---

## 4. Monthly Employee Score Calculation

Each message was scored as:

- Positive = +1

- Negative = –1

- Neutral = 0

Monthly scores were calculated by summing message scores per employee per month.

---

## 5. Employee Ranking

For each month:

- \*\*Top 3 Positive Employees\*\*: highest monthly sentiment scores

- \*\*Top 3 Negative Employees\*\*: lowest monthly sentiment scores

Ranking helped highlight consistently positive contributors and detect potentially disengaged staff.

---

## 6. Flight Risk Identification

An employee was flagged as \*\*Flight Risk\*\* if they sent \*\*4 or more negative messages\*\* in any rolling 30-day period.

Result: Employees like `alice@example.com` were flagged due to concentrated negative activity.

---

## 7. Predictive Modeling

We trained a \*\*linear regression model\*\* using:

- Message count per month

- Average message length

- Average word count

\*\*Model Results (Simulated):\*\*

- MAE: 3.33

- RMSE: 4.48

- R² Score: -19.11 (low due to limited sample)

\*\*Coefficients\*\* revealed that:

- More frequent and wordier messages were associated with positive sentiment

- Longer character count may imply negativity

---

## 8. Conclusion & Recommendations

- Monitoring sentiment trends helps reveal engagement and potential risks.

- Monthly score tracking and ranking support performance evaluation.

- Predictive modeling may inform early interventions, especially for high-risk individuals.

Next Steps:

- Enhance model with more features (e.g., topic modeling, department info)

- Use classification models to predict sentiment categories

- Integrate results into dashboards for HR/management insight