# Final Report: Employee Sentiment and Engagement Analysis

## 1. Introduction

This project aims to analyze an unlabeled dataset of employee messages to assess sentiment and engagement levels across the organization. By leveraging natural language processing (NLP), statistical analysis, and predictive modeling, we derive actionable insights into employee morale, identify potential flight risks, and rank employees based on sentiment trends. The ultimate goal is to support HR decision-making with data-driven evidence.

## 2. Methodology

2.1 Sentiment Labeling  
Each message body was analyzed using the TextBlob NLP library to assign a sentiment polarity score. Based on the polarity:  
- Polarity > 0.1 → Positive  
- Polarity < –0.1 → Negative  
- Otherwise → Neutral  
This process added a new Sentiment column to the dataset, enabling downstream analysis.  
  
2.2 Data Preprocessing  
- Converted date column to datetime format  
- Extracted Month from each timestamp  
- Calculated Message\_Length for modeling  
- Mapped sentiment to numeric scores: Positive = +1, Negative = –1, Neutral = 0

## 3. EDA Findings

3.1 Sentiment Distribution  
The majority of messages were Neutral, followed by Positive and Negative. This suggests a generally balanced tone, with occasional spikes in negativity.  
  
3.2 Time Trends  
Line plots of daily sentiment counts revealed:  
- Periodic surges in Negative messages, often clustered  
- Positive sentiment remained relatively stable  
- Neutral messages dominated during quieter periods  
  
3.3 Anomalies  
- A few employees had unusually high message volumes  
- Some months showed sharp sentiment shifts, possibly linked to organizational events  
(Include bar charts and line plots here)

## 4. Employee Scoring & Ranking

4.1 Monthly Sentiment Score  
Each employee’s messages were scored monthly:  
- Positive = +1  
- Negative = –1  
- Neutral = 0  
Scores were aggregated per employee per month.  
  
4.2 Rankings  
For each month:  
- Top 3 Positive Employees: Highest cumulative sentiment scores  
- Top 3 Negative Employees: Lowest (most negative) scores  
(Include tables or charts showing rankings)

## 5. Flight Risk Analysis

5.1 Criteria  
An employee is flagged as a flight risk if they sent 4 or more Negative messages within any rolling 30-day window, regardless of month boundaries.  
  
5.2 Results  
The following employees were flagged as potential flight risks:  
- [Employee A]  
- [Employee B]  
- [Employee C]  
This method ensures early detection of disengagement patterns.

## 6. Predictive Modeling

6.1 Features Used  
- Message\_Count: Number of messages per employee per month  
- Avg\_Message\_Length: Average length of messages  
- Target: Monthly\_Score (sentiment score)  
  
6.2 Model  
A Linear Regression model was trained and evaluated:  
- R² Score: [e.g., 0.62]  
- MAE: [e.g., 1.3]  
  
6.3 Interpretation  
- Message frequency had a stronger correlation with sentiment score than message length  
- The model can help forecast sentiment dips and identify at-risk employees early  
(Include scatter plot of actual vs predicted scores)

## 7. Conclusion

This analysis provides a robust framework for understanding employee sentiment and engagement. Key takeaways:  
- Most employees maintain a neutral or positive tone  
- A small subset shows consistent negativity, warranting attention  
- Predictive modeling offers a scalable way to monitor sentiment trends  
  
Recommendations:  
- Investigate flagged flight-risk employees for potential intervention  
- Monitor sentiment trends monthly to detect shifts early  
- Consider integrating this pipeline into HR dashboards for real-time insights