Final Report: Employee Sentiment Analysis & Flight Risk Prediction

1. Introduction

This project analyzes employee emails/messages to understand sentiment trends, rank employees, identify potential flight risks, and build a predictive model for HR strategy. The dataset consists of emails with subject, body, sender, and date.

2. Methodology

The project was divided into the following tasks:

• Data Preprocessing: Combined Subject & Body, cleaned missing values, converted date to datetime.

• Task 1 – Sentiment Analysis: Applied VADER sentiment scoring (Positive, Negative, Neutral).

• Task 2 – Exploratory Data Analysis: Distribution, trends, heatmaps, and word clouds.

• Task 3 – Monthly Scoring: Assigned sentiment scores (+1, -1, 0) and aggregated by month.

• Task 4 – Employee Ranking: Identified top 3 positive and negative employees per month.

• Task 5 – Flight Risk Identification: Flagged employees with ≥4 negative emails in 30 days.

• Task 6 – Predictive Modeling: Built Linear Regression using features (frequency, length, word count).

• Outlier Detection: Identified employees with unusual communication patterns using IQR.

3. Key Findings (EDA)

• Neutral messages dominate overall sentiment distribution.

• Positive sentiment is consistent among some employees, while others show repeated negativity.

• Flight risk employees were flagged based on frequent negative communication.

4. Employee Scoring & Ranking

• Top 3 Positive Employees:

	Month	Type	Rank	Employee \	Monthly_Sentiment_Score
0	2010-01	Top Positive	1	kayne.coulter@enron.com	14
1	2010-01	Top Positive	2	eric.bass@enron.com	9
2	2010-01	Top Positive	3	lydia.delgado@enron.com	9

• Top 3 Negative Employees:

	Month	Type	Rank	Employee \	Monthly_Sentiment_Score
139	2011-11	Top Negative	2	rhonda.denton@enron.com	2
140	2011-11	Top Negative	3	lydia.delgado@enron.com	4
141	2011-12	Top Negative	1	johnny.palmer@enron.com	1

These rankings help HR identify top performers and employees who may need intervention.

5. Flight Risk Analysis

Employees with ≥4 negative emails in a rolling 30-day period were flagged.

List of flight risk employees: [Replace with actual names].

6. Predictive Modeling Results

The Linear Regression model was built to predict Monthly Sentiment Score using:

• Message Frequency
• Avg. Message Length
• Avg. Word Count

Results:

- RMSE: 1.83
- R^2 Score: 0.71

Insights: Communication patterns provide predictive power for employee sentiment trends.

7. Conclusion & Recommendations

- Recognize high positive employees as role models.
- Provide HR support to consistently negative employees.
- Monitor flight risk employees proactively to reduce attrition.
- Use predictive modeling for proactive HR engagement.

8. Appendix

- Tables of employee sentiment scores.
- Visualizations stored in the 'visualizations' folder.
- Code and documentation available in GitHub repository.