**Employee Sentiment Analysis Report**

**Project by:** Rohit Gudiya  
**Date:** September 2025

**1. Project Overview**

This project analyzes employee email messages to understand sentiment trends, identify top performing employees, detect potential flight risks, and predict future sentiment trends using a Linear Regression model.

**Workflow:**

1. Data Preprocessing
2. Sentiment Labeling
3. Exploratory Data Analysis (EDA)
4. Employee Score Calculation
5. Employee Ranking
6. Flight Risk Identification
7. Predictive Modeling

**2. Dataset Summary**

* **Total Messages:** 2,191
* **Columns:** Subject, Body, Date, From, Message, Sentiment, Sentiment Score, Month, Message Length, Week, Negative Flag
* **Missing Values:** None

**Sentiment Distribution:**

* Positive: 899
* Negative: 1,292
* Neutral: 0

**3. Exploratory Data Analysis (EDA)**

* **Monthly & Weekly Trends:**
  + Line charts show message counts over time for each sentiment.
* **Message Length:**
  + Boxplots and histograms used to detect unusually long or short messages.
* **Key Observations:**
  + Negative sentiment messages are higher than positive.
  + Certain employees have consistent positive or negative trends.

**Visualizations:** *(All stored in visualizations/ folder)*

* sentiment\_distribution.png
* monthly\_message\_counts.png
* weekly\_message\_counts.png
* message\_length\_boxplot.png

**4. Employee Monthly Scores**

* Sentiment mapped to scores: Positive = +1, Neutral = 0, Negative = -1
* Scores aggregated per employee per month

**Example:**

| **Employee ID** | **Month** | **Score** |
| --- | --- | --- |
| johnny.palmer@enron.com | 2010-01 | 0 |
| sally.beck@enron.com | 2010-01 | -3 |
| lydia.delgado@enron.com | 2010-01 | -7 |

**Output File:** output/employee\_monthly\_scores.csv

**5. Employee Rankings**

* **Top Positive Employees:** Employees with highest monthly scores
* **Top Negative Employees:** Employees with lowest monthly scores
* Sorted by score (highest to lowest) and alphabetically if tied

**Example:**

| **Employee ID** | **Month** | **Score** | **Rank Type** |
| --- | --- | --- | --- |
| don.baughman@enron.com | 2010-02 | 5 | Top Positive |
| sally.beck@enron.com | 2010-02 | 2 | Top Positive |
| bobette.riner@ipgdirect.com | 2010-01 | -2 | Top Negative |

**Output File:** output/employee\_rankings.csv

**Visualization:** visualizations/top\_employees\_per\_month.png

**6. Flight Risk Identification**

* Employees sending 4 or more negative emails within a 30-day rolling period are flagged as **Flight Risk**.

**Example Flight Risk Employees:**

* bobette.riner@ipgdirect.com
* don.baughman@enron.com
* eric.bass@enron.com
* johnny.palmer@enron.com

**Output File:** output/flight\_risk\_employees.csv

**Visualization:** visualizations/flight\_risk\_events.png

**7. Predictive Modeling**

* **Model:** Linear Regression
* **Features:** Total messages, average message length, average word count, past sentiment score
* **Target:** Monthly sentiment score

**Performance Metrics:**

* R² Score: 1.00
* MSE: 0.00
* RMSE: 0.00
* MAE: 0.00

**Feature Importance:**

| **Feature** | **Coefficient** |
| --- | --- |
| total\_messages | 1.796583e-16 |
| avg\_msg\_length | 2.775558e-17 |
| avg\_word\_count | -5.898060e-17 |
| past\_sentiment\_score | 1.000000e+00 |

**Visualizations:**

* actual\_vs\_predicted.png
* residuals.png
* feature\_importance.png

**Output Files:**

* output/model\_metrics.csv
* output/feature\_importance.csv

**8. Key Insights & Recommendations**

* High negative message frequency may indicate disengagement or flight risk.
* Recognize or reward top positive employees to boost morale.
* Predictive modeling helps HR proactively manage employee sentiment trends.