Employee Sentiment Analysis

# 📌 Overview

This project analyzes employee messages to assess sentiment, engagement, rankings, flight risk, and builds a simple predictive model.

**Methodology**

**Task 1: Sentiment Labeling**

* Combined the Subject and Body of each email into a single text string.
* Used **VADER SentimentIntensityAnalyzer** to classify each message as **Positive (+1)**, **Negative (-1)**, or **Neutral (0)**.
* Stored results in labeled\_dataset.csv.

**Task 2: Exploratory Data Analysis (EDA)**

* Checked for missing values and verified data structure.
* Created visualizations for sentiment distribution, top senders, and monthly sentiment trends.
* Saved charts to the visualization/ folder.

**Task 3: Monthly Sentiment Scoring**

* Mapped sentiment labels to scores (+1, 0, -1).
* Aggregated scores by employee and month (YearMonth) to produce monthly\_sentiment\_scores.csv.

**Task 4: Employee Ranking**

* Sorted employees by monthly score and selected the **top 3 positive** and **top 3 negative** employees for each month.
* Saved rankings to top3\_positive\_per\_month.csv and top3\_negative\_per\_month.csv.

**Task 5: Flight Risk Identification**

* Flagged employees who sent **4 or more negative messages** within any 30-day rolling window.
* Produced flight\_risk\_employees.csv.

**Task 6: Predictive Modeling**

* Created features: monthly message count, average message length, average word count.
* Built a **Linear Regression model** to predict sentiment scores.
* Evaluated with RMSE and R² metrics and saved coefficients to model\_coefficients.csv.

**Setup**

1. **Clone or Download the Repository**
2. git clone <your\_repo\_url>
3. cd <your\_repo\_folder>
4. **Install Dependencies**
5. pip install -r requirements.txt

*(If you don’t have requirements.txt, install the following manually)*

pip install pandas openpyxl nltk matplotlib seaborn scikit-learn

1. **Download NLTK Data**
2. import nltk
3. nltk.download('vader\_lexicon')
4. **Place Dataset**
   * Add your test.xlsx file in the root project directory (or update the path in the script).

# 📊 Key Results

## Sentiment Distribution

Positive: 47.46%  
Neutral: 38.2%  
Negative: 14.34%

## Top 3 Positive Employees

A screenshot of a computer screen

AI-generated content may be incorrect.

## Top 3 Negative Employees — 2011-07

|  |  |
| --- | --- |
|  |  |

## Overall Top 3 Positive (All Months)

|  |  |
| --- | --- |
| Employee | Total Score |
| kayne.coulter@enron.com | 43 |
| sally.beck@enron.com | 41 |
| johnny.palmer@enron.com | 36 |

## Overall Top 3 Negative (All Months)

|  |  |
| --- | --- |
| Employee | Total Score |
| bobette.riner@ipgdirect.com | 1 |
| rhonda.denton@enron.com | 1 |
| sally.beck@enron.com | 2 |

## 🚨 Flight Risk Employees (≥4 negative emails in any 30-day window)

A screenshot of a computer

AI-generated content may be incorrect.

## 🤖 Predictive Model (Linear Regression)

A screenshot of a computer program

AI-generated content may be incorrect.

## 📁 Files

* task1-task6.ipynb
* labeled\_dataset.csv
* monthly\_sentiment\_scores.csv
* top3\_positive\_per\_month.csv
* top3\_negative\_per\_month.csv
* flight\_risk\_employees.csv
* model\_coefficients.csv
* visualization/ (charts)