Employee Sentiment Analysis Project

**Documentation**

**1. Process Documentation**

**1.1 Workflow Steps**

- Data ingestion and conversion of date field.

- Monthly aggregation of messages per employee.

- Sentiment labeling using numeric scores (Positive = 1, Neutral = 0, Negative = -1).

- Feature engineering: message count, positive/neutral/negative counts per month.

- Model training using Linear Regression.

- Model evaluation using Mean Squared Error (MSE) and R-squared.

**1.2 Assumptions**

- Sentiment labels are accurate and consistently assigned.

- Monthly aggregation gives meaningful trends for employee sentiment.

**1.3 Tools and Libraries Used**

- Python 3.9

- pandas, numpy

- scikit-learn

- matplotlib

**2.Clarity and Organization**

**2.1 Project Structure**

**- data: Raw and processed data files.**

**- notebooks: Jupyter notebooks for each stage.**

**- scripts: Python scripts for reusable code.**

**- models: Saved model artifacts.**

**- reports: Generated reports and visualizations.**

**2.2 README File**

**- Project overview.**

**- Setup instructions.**

**3. Modeling and Evaluation**

**3.1 Feature Set**

- message\_count

- positive\_count

- negative\_count

- neutral\_count

**3.2 Model Used**

Linear Regression

**3.3 Evaluation Metrics**

- Mean Squared Error (MSE): 0.00

- R-squared (R²): 1.00

**4. Reproducibility**

**4.1 Scripts and Notebooks**

- Separate EDA, preprocessing, modeling, and evaluation notebooks.

- Main pipeline script.

**4.2 Version Control**

- GitHub repository with regular commits.

**4.3 Instructions**

- Step-by-step guide from raw data to final model training and evaluation

**EDA Results:**

Fig 1: Monthly Sentiment Trend

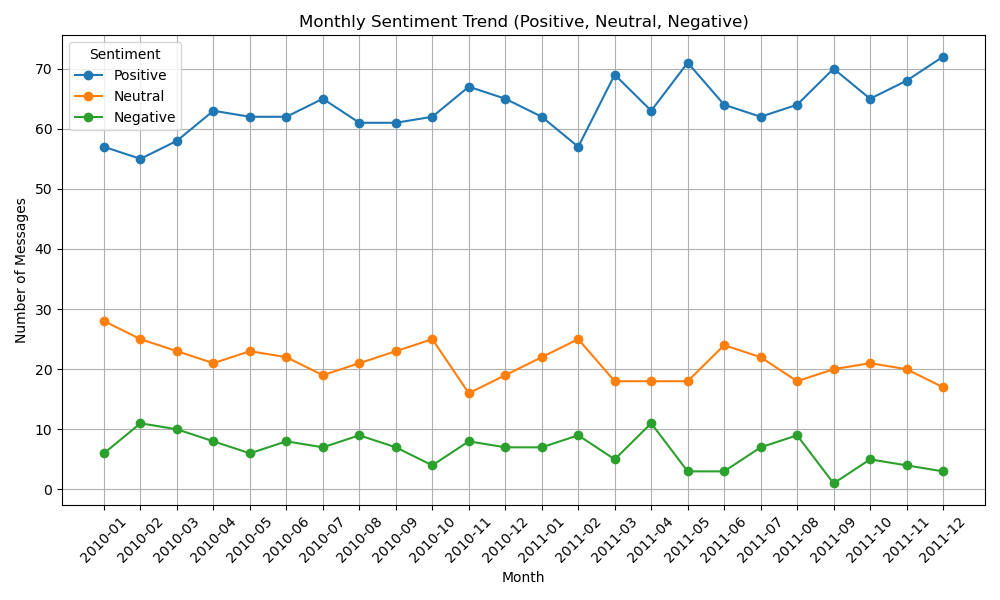


Fig 2: Top 3 Positive Employees

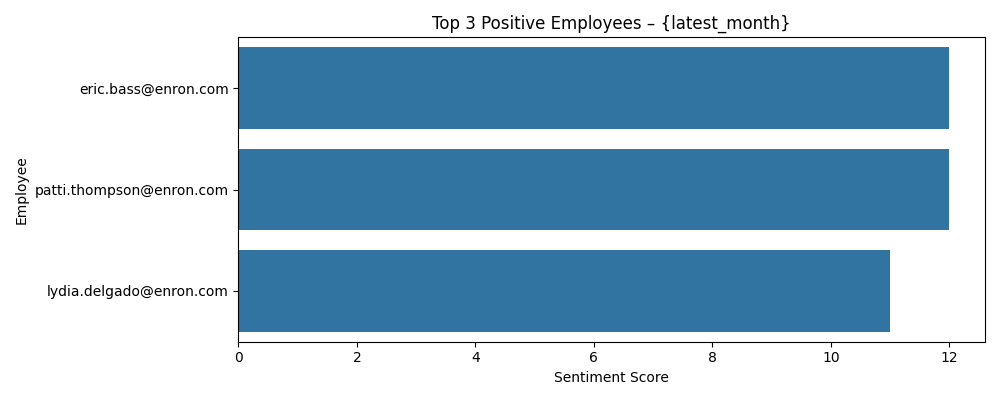


Fig 3: Top 3 Negative Employees

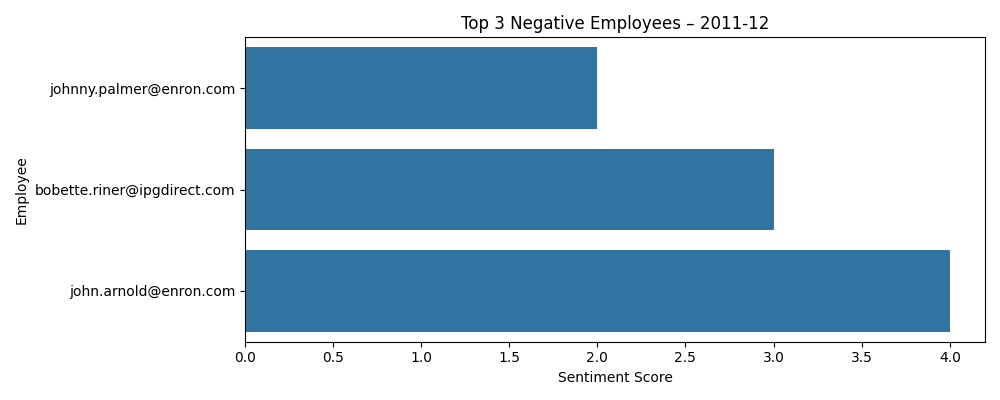


Fig 4: Sentiment Score Predictions (R² = 1.00).

