LLM FINAL ASSESSMENT

Submitted by – Kushagra Jain

Data Analytics and LLM Internship

PROJECT OVERVIEW

The goal of the project is to evaluate employee sentiment and engagement using data analysis techniques such as sentiment analysis, exploratory data analysis, sentiment labelling, employee ranking and predictive modeling. I need to analyze employee messages to understand how people feel and how engaged they are at work. I'll be working with raw data that doesn't have labels, so I will be figuring out the patterns myself using text analysis and statistics. The project has several different parts where I will analyze the data and build models to get insights.

PROJECT OBJECTIVE

The main goal is to evaluate employee sentiment and engagement by performing the following:

* **Sentiment Labeling:** Automatically label each message as Positive, Negative, or Neutral.
* **Exploratory Data Analysis (EDA):** Analyze and visualize the data to understand its structure and underlying trends.
* **Employee Score Calculation:** Compute a monthly sentiment score for each employee based on their messages.
* **Employee Ranking:** Identify and rank employees by their sentiment scores.
* **Flight Risk Identification:** A Flight risk is any employee who has sent 4 or more negative mails in a given month.
* **Predictive Modeling:** Develop a linear regression model to further analyze sentiment trends.

SENTIMENT LABELING

For sentiment labelling, I first prepared the text to be analyzed:

* Removing special characters, HTML tags, and extra whitespace.
* Normalizing text (lowercase everything).
* Removing stopwords (common but unimportant words like “the”, “and”, etc.).
* Lemmatization (reducing words to root forms like “running” → “run”).
* Combined subject and body text for full context of the email into a new column called full\_text

For the tool to be used for sentiment analysis, I used Vader as it is a reliable tool commonly used for analyzing sentiment from emails. It is rule-based, fast and reliable, and is great for social texts, emails et. Given the context, TextBlob or Transformers would be either not accurate enough or too resource heavy.

For the sentiment analysis, I used the VADER tool. The labeling criteria in VADER is as follows:

* Compound Score Range >= 0.05: 'Positive'
* Compound Score Range <= -0.05: 'Negative'
* Compound Score Range between -0.05 and 0.05: 'Neutral'

The compound score is a normalized score given to text, calculated using a lexicon of words with known sentiment values and some heuristic rules

Here is a sample of the sentiment labelling:

A screenshot of a computer

AI-generated content may be incorrect.

EXPLORATORY DATA ANALYSIS

I performed exploratory data analysis on the data to understand the trends and patterns in the data regarding the sentiment scores. I examined the overall structure (eg. the number of records, data types, missing values, null values etc.).

A screenshot of a computer

AI-generated content may be incorrect.

I also investigated the distribution of the sentiment labels across the dataset to understand the sentiment trends among the employees.

A bar graph with different colored squares

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

I analyzed the trends of sentiment scores over time to understand if there are any seasonal effects in the sentiments of the employees.

A graph of a number of different colored lines

AI-generated content may be incorrect.

I also investigated the major sentiment for each individual employee, and the respective distributions of sentiments over time.

A graph with different colored bars

AI-generated content may be incorrect.

A summary of the sentiments among the employees and their nature can be found below:

A screen shot of a computer

AI-generated content may be incorrect.

TAKEAWAYS

Additional Patterns and Insights into Employee Engagement:

* Positive Sentiment Dominates: Suggests a generally optimistic culture or model bias (e.g., polite closings).
* Stable Positive Trend: Positive sentiment remains steady across months.
* Fluctuating Neutral/Negative Sentiments: Possible reaction to internal events or pressures.
* Top Senders = Mostly Positive: Likely leaders or HR roles maintaining positive tone.
* Email Length Patterns:

Positive emails → longest subject and body.

Neutral emails → shortest, likely factual or procedural.

* Anomalies:
* Spikes in negative/neutral sentiment may indicate specific stress events.
* Certain senders show higher negative sentiment—worth deeper review.

EMPLOYEE SCORE CALCULATION

I computed a monthly score for each employee based on their messages. The scores were aggregated on a monthly basis for each employee. These scores were reset at the beginning of every new month. The scoring metric is as follows:

* Positive Message: +1
* Negative Message: -1
* Neutral Message: 0 (No Effect)

The result of the task is the creation of the column monthly\_sentiment\_score.

EMPLOYEE RANKING

I generated a list of employees based on their monthly sentiment scores. I then ranked them to understand who the employees were who were most likely to send a positive or a negative message.

I created two distinct lists:

* Top Three Positive Employees: The three employees with the highest positive scores in a given month.
* Top Three Negative Employees: The three employees with the lowest (most negative) scores in each month.

I sorted them first in descending order and then in alphabetical order. I ensured that the ranking is clearly derived from the sentiment score calculation as previously executed.

A screenshot of a computer

AI-generated content may be incorrect.

FLIGHT RISK IDENTIFICATION

This stage involved identifying the employees that are at a risk of leaving the company based on their monthly sentiment scores. A flight risk was established as any employee who had sent 4 or more negative emails in a span of 30 days (irrespective of the score). The 30 day period is a rolling count of days, irrespective of months.

A screenshot of a computer

AI-generated content may be incorrect.

PREDICTIVE MODELING

I developed a linear regression model to analyze sentiment trends and predict sentiment scores using a variety of independent variables that may influence sentiment scores. The idea was to predict the sentiment scores of an employee based on their engagement and sentiment on mails. This would help us to identify potential flight risks.

FEATURE ENGINEERING

I first created the following features for the purpose of prediction in the dataset:

* Subject Length: The total length of the subject of the mail.
* Body Length: The total length of the body text of the mail.
* Word Count: The total word count of the combined text (subject and body text).
* Average Monthly Messages: The average number of messages sent by a particular user in a month.

MODEL DEVELOPMENT

I then proceeded to split the data into training and testing dataset (80%-20%). I then trained the linear regression OLS model on the training dataset and cross validated across the testing dataset. The model output and evaluation metrics are as follows:

A screenshot of a computer

AI-generated content may be incorrect.

A screen shot of a computer error

AI-generated content may be incorrect.

INTERPRETATION

Model Fit:

* R-squared = 0.050: The model explains only 5% of the variation in sentiment score, indicating a very weak fit.
* F-statistic is significant (p < 0.001): Despite the weak fit, the model as a whole is statistically significant, meaning at least one predictor contributes meaningfully.

|  |  |  |  |
| --- | --- | --- | --- |
| Feature | Coefficient | P-Value | Interpretation |
| const | 0.4074 | 0.000 | The baseline sentiment score is ~0.41 when all features are 0. |
| message\_count | -0.0007 | 0.805 | Not significant. Number of messages has **no meaningful impact** on sentiment. |
| subject\_len | 0.000028 | 0.981 | Not significant. Subject length does **not affect** sentiment. |
| body\_len | -0.0014 | 0.015 | Significant and negative. Longer email bodies are associated with **slightly lower** sentiment scores. |
| word\_count | 0.0172 | 0.000 | Highly significant. Emails with more words are linked to **higher** sentiment scores. |

CONCLUSION

* Only two features (body\_len and word\_count) have statistically significant effects.
* Word count positively impacts sentiment, while long body length negatively affects it.
* Overall, the model has low predictive power, suggesting important predictors might be missing or sentiment is influenced by more complex factors (e.g., tone, context, language used).

The model can be improved further by adding more features. As we can see the model has a weak fit on our data, this can be improved by adding more features to the data to make the model more reliable. We can also develop advanced feature engineering techniques such as squaring and interaction terms to further develop the model fit on our data.