Sentiment Analysis through My WhatsApp Messages

Motivation

This project aims to gain insights into my messaging habits and emotional patterns through sentiment analysis. By understanding how my sentiments vary throughout the day, I aim to optimise communication strategies and identify potential patterns that may influence my mood and behaviour.

Data Source

The primary data source for this project is my personal WhatsApp messages. The dataset includes conversational data with timestamps, allowing for a detailed analysis of messaging intensity and sentiment over different periods. The data is collected and processed using Python scripts in a Jupyter Notebook environment.

Data Analysis

1. Data Collection

Conversational data is collected from WhatsApp, ensuring the inclusion of timestamps for each interaction.

2. Data Cleaning

- Missing values are handled.
- Text data is converted to lowercase, and special characters and links are removed.
- Tokenization is performed, and stopwords are removed.
- Spell checking is implemented using the ZEMBEREK Turkish NLP library.

3. Message Intensity Analysis

The intensity of messages is observed over time and date periods. The count of messages is analyzed to understand messaging habits.

4. Sentiment Analysis

The BERT-based Turkish model is used for sentiment analysis. Sentiment scores are calculated for each message.

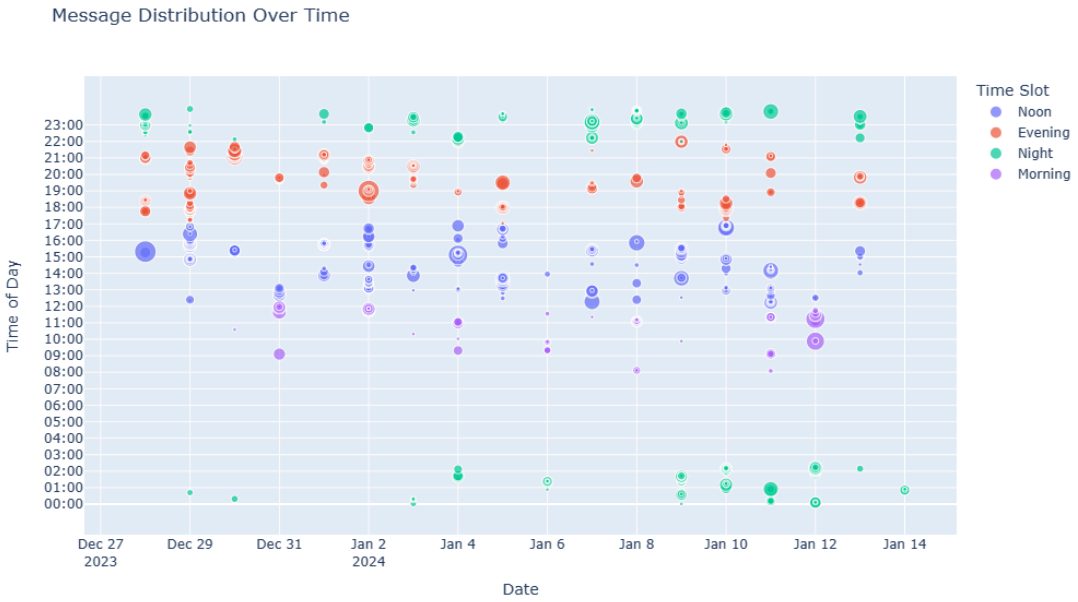
5. Hypothesis Testing

The Pearson correlation test is employed to assess the correlation between sentiment scores during different time slots. Hypotheses are defined and tested to explore patterns and trends.

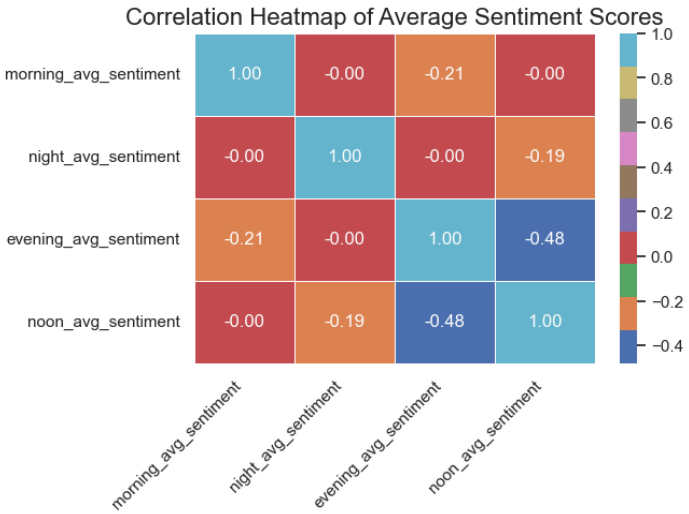
Findings

The analysis provides several key findings:

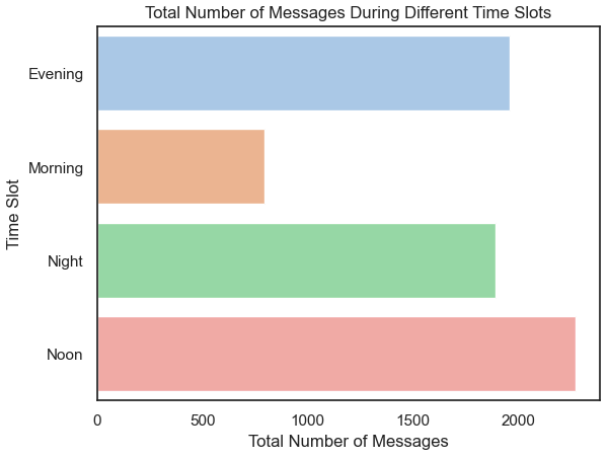
- Messaging Intensity Ranks: The total number of messages during different time slots follows a distinct ranking, revealing patterns in messaging habits.

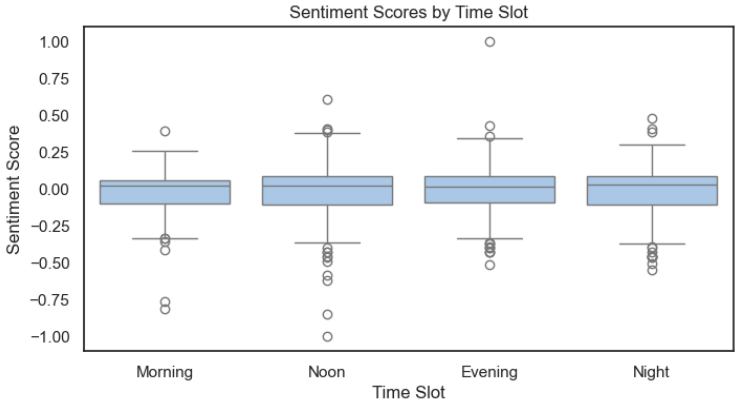


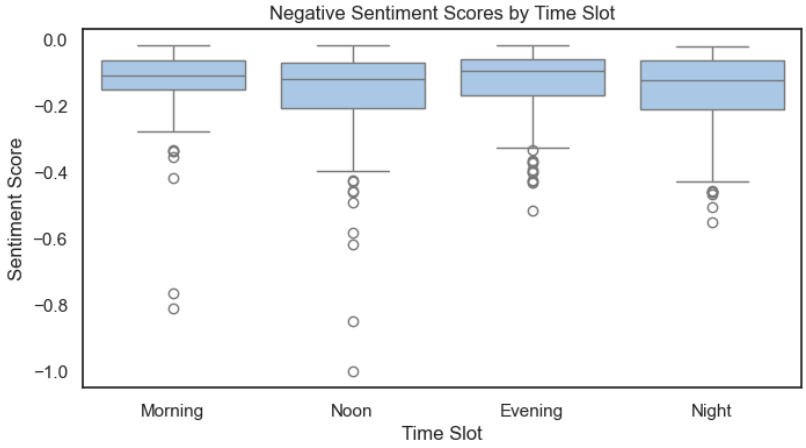
  - Please be aware that the plotted graph is not visible when viewing the .ipynb file on GitHub.

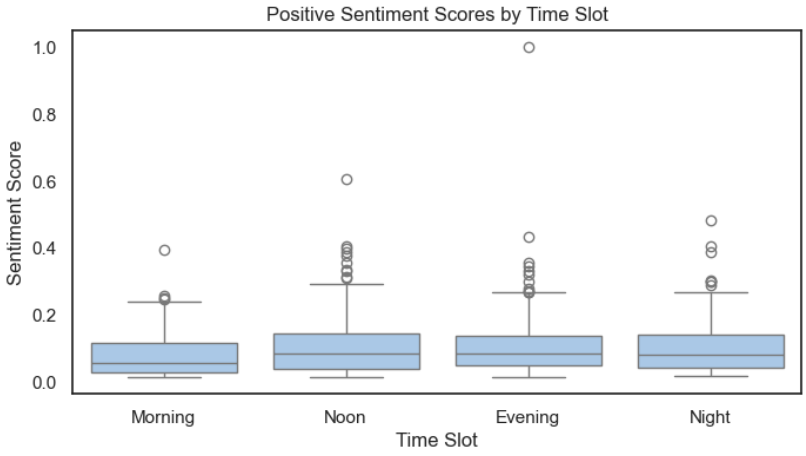


- Sentiment Correlations: Most time slots show no significant correlation in sentiment scores, except for a significant correlation between evening and noon sentiment. This implies a potential pattern or similarity in sentiment during these specific time slots.









Limitations and Future Work

Limitations

- The analysis relies on WhatsApp messages and may not capture sentiments from other communication channels.
- The sentiment analysis model may need to be revised to understand nuanced emotions.

Future Work

- Integration with additional data sources for a more comprehensive sentiment analysis.
- Refinement of sentiment analysis models to capture subtle emotional nuances.
- Long-term analysis to identify trends and changes in messaging habits.

Conclusion

Understanding messaging habits and sentiment patterns provides valuable insights into personal communication dynamics. The project's findings can guide improvements in communication strategies and lay the groundwork for future investigations into emotional well-being and behaviour.