Project Report: WhatsApp Chat Analysis

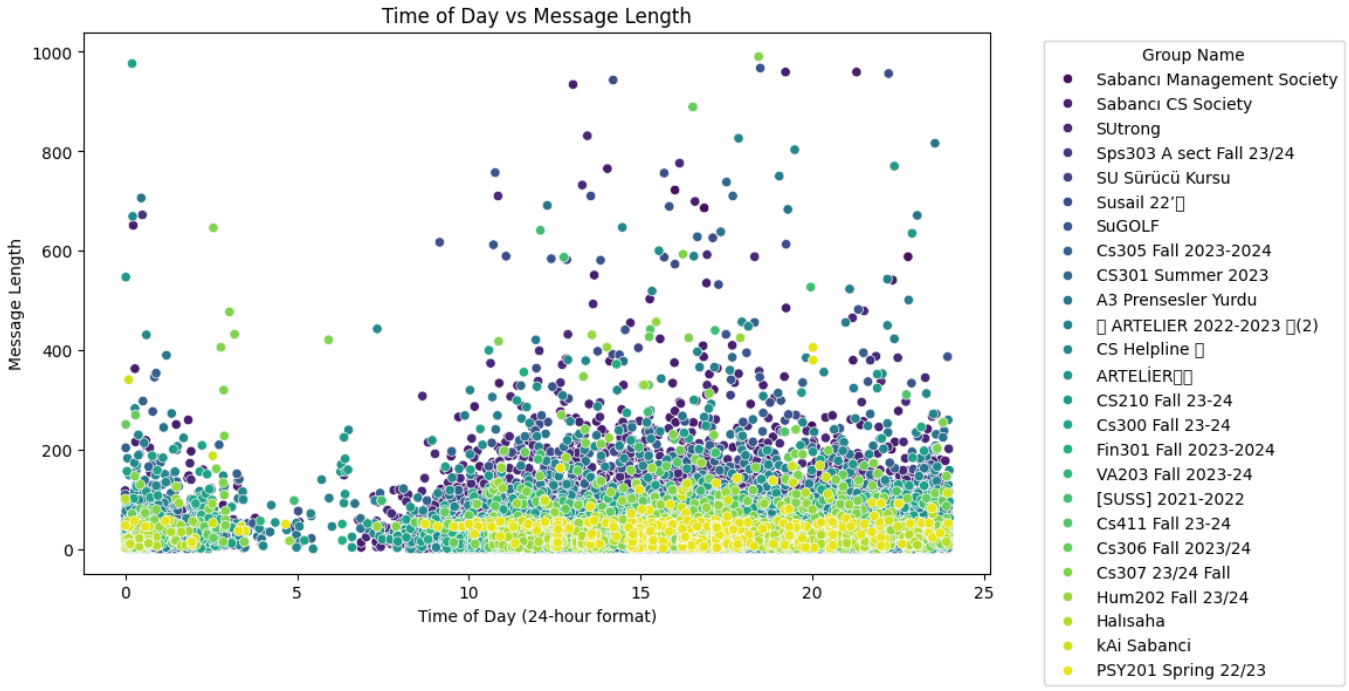
Motivation:
The motivation behind this data science project stems from a curiosity to understand the patterns and dynamics of communication within WhatsApp groups. Specifically, we aimed to explore whether there is a relationship between the time of day and the length of messages exchanged in these groups. Understanding these patterns can provide valuable insights into how communication patterns evolve throughout the day and contribute to a better understanding of group dynamics.

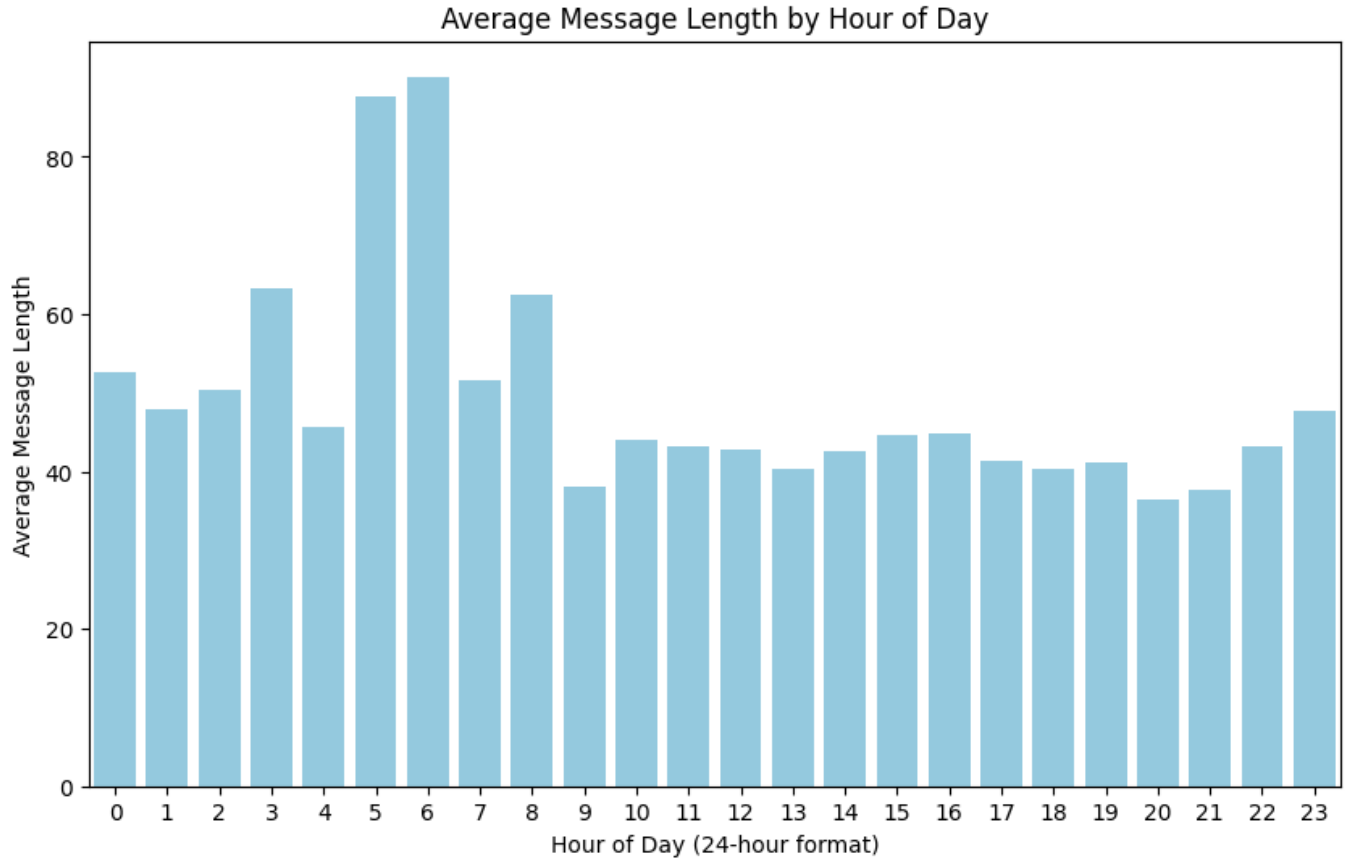
Data Source:
The primary data source for this analysis is my personal WhatsApp chat history. I have extracted it directly from my smart phone application. The dataset includes timestamps for each message and the corresponding length of the messages. The data was collected over an extended period to ensure a representative sample that captures different times of the day and a diverse range of conversation lengths.

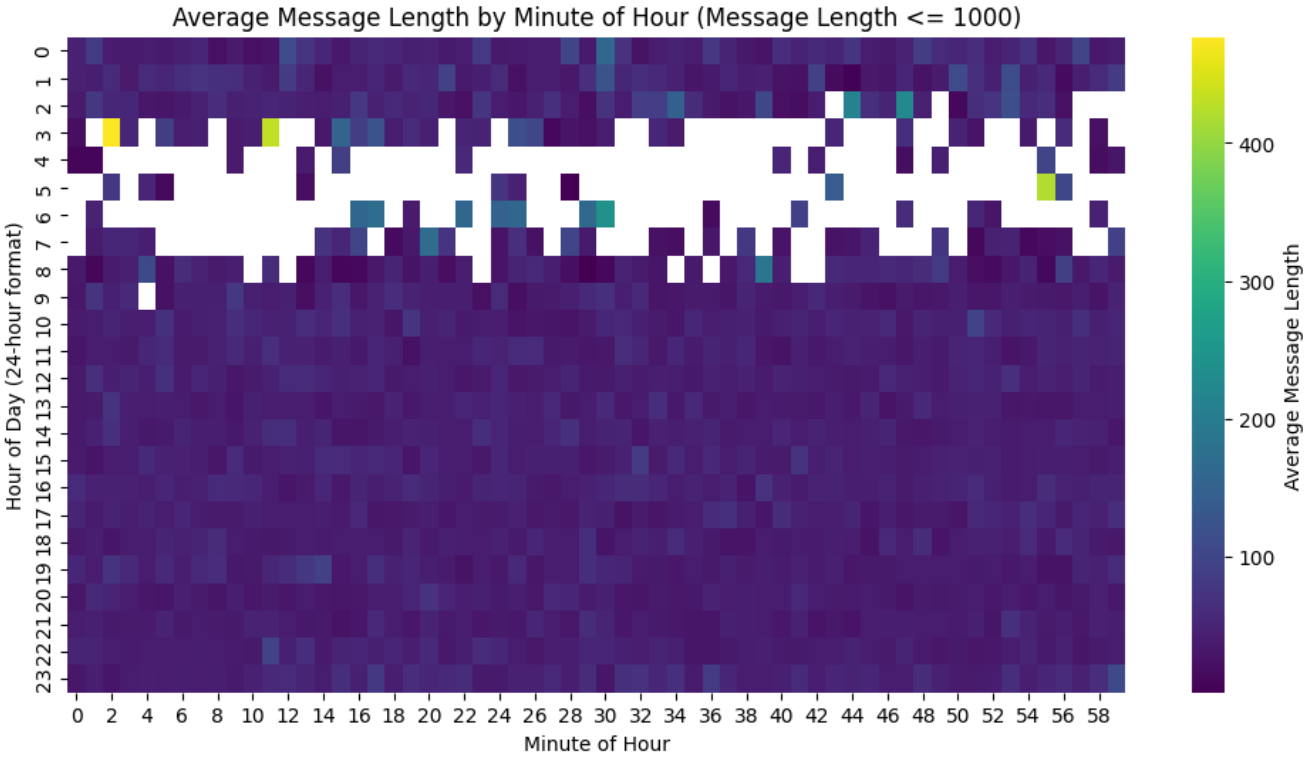
Data Analysis:

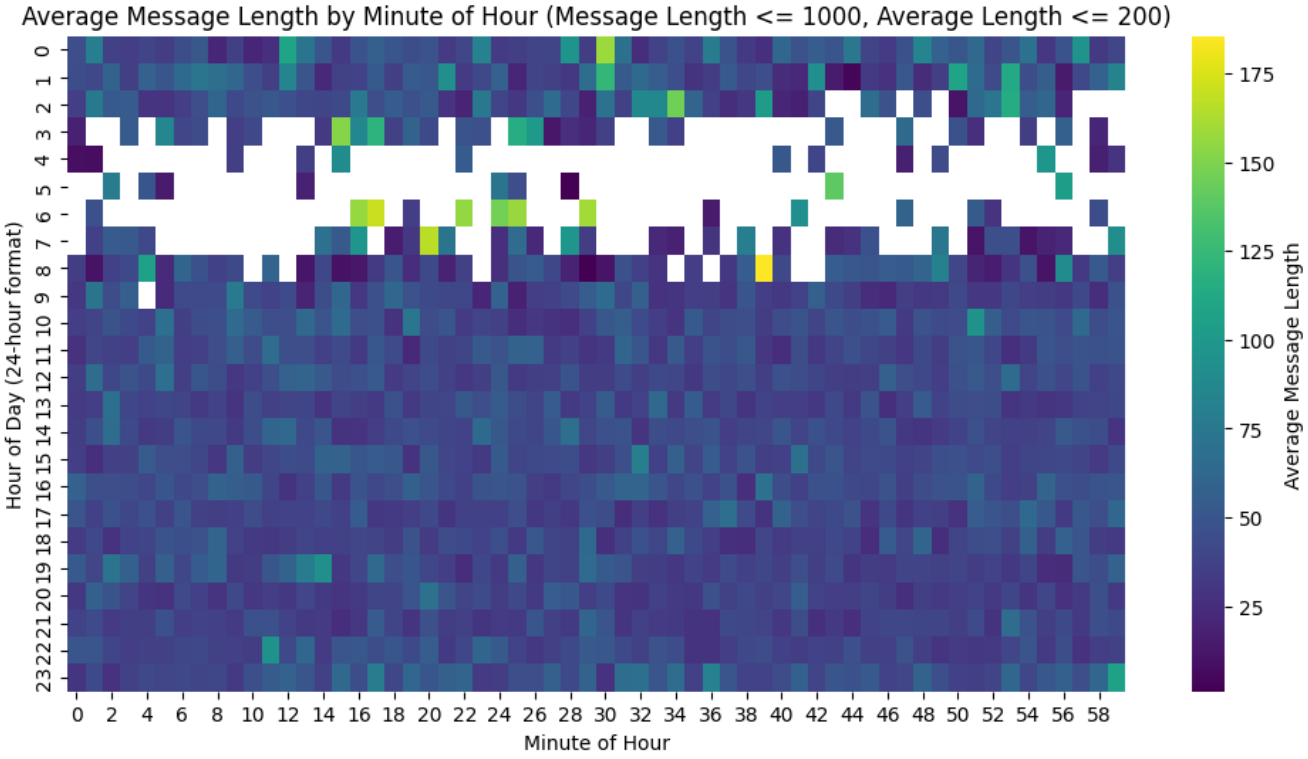
1. **Data Preprocessing:**
    - **Timestamp Extraction:** The first step involved extracting timestamps from the raw data, allowing us to identify the time each message was sent.
    - **Message Length Calculation:** Calculating the length of each message served as the dependent variable for our analysis.
2. **Exploratory Data Analysis (EDA):**
    - Conducting EDA was crucial to gaining insights into the distribution of message lengths across different times of the day.
    - Visualization tools such as heat maps, scatter plots, and box plots were employed to illustrate patterns and trends in the data.
3. **Statistical Analysis:**
    - Applied statistical methods, including regression analysis and p-value testing, to evaluate the hypotheses.
    - Hypothesis testing involved comparing the null hypothesis (no significant relationship) with the alternative hypothesis (significant relationship).

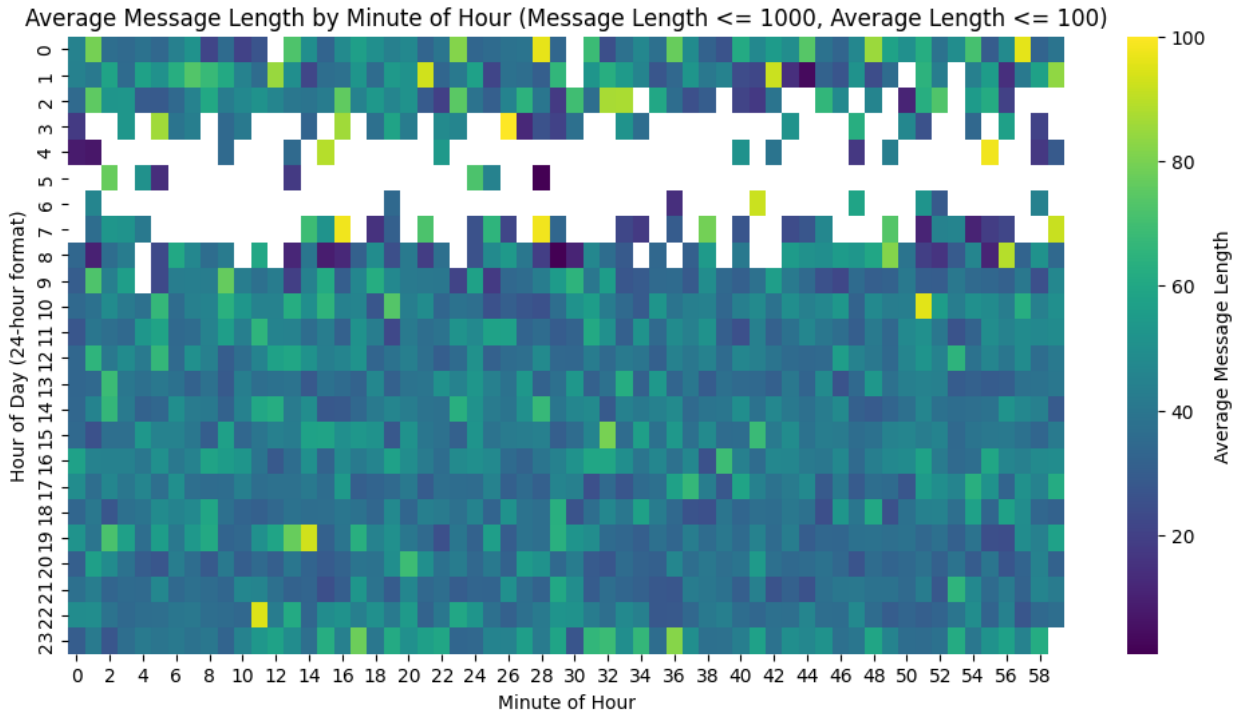
Findings:
Contrary to our initial hypothesis, the analysis did not reveal any significant relationship between the time of day and message length in my WhatsApp chat history. The absence of a clear pattern suggests that, at least in my case, my message lengths are not strongly influenced by the time of day.


Time of Day vs Message Length
Message Length
Time of Day (24-hour format)
0
200
400
600
800
1000
0
5
10
15
20
25
Group Name
Sabancı Management Society
Sabancı CS Society
SUtrong
Sps303 A sect Fall 23/24
SU Sürücü Kursu
Susail 22'
SuGOLF
Cs305 Fall 2023-2024
CS301 Summer 2023
A3 Prensesler Yurdu
ARTELIER 2022-2023 (2)
CS Helpline
ARTELİER
CS210 Fall 23-24
Cs300 Fall 23-24
Fin301 Fall 2023-2024
VA203 Fall 2023-24
[SUSS] 2021-2022
Cs411 Fall 23-24
Cs306 Fall 2023/24
Cs307 23/24 Fall
Hum202 Fall 23/24
Halısaha
kAi Sabanci
PSY201 Spring 22/23


Average Message Length by Hour of Day
Average Message Length
Hour of Day (24-hour format)
0
20
40
60
80
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23


Average Message Length by Minute of Hour (Message Length <= 1000)
Hour of Day (24-hour format)
Minute of Hour
Average Message Length
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46 48 50 52 54 56 58
100 200 300 400


Average Message Length by Minute of Hour (Message Length <= 1000, Average Length <= 200)
Hour of Day (24-hour format)
Minute of Hour
Average Message Length
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46 48 50 52 54 56 58
25 50 75 100 125 150 175



Limitations and Future Work:

Limitations:

- The findings are specific to my personal WhatsApp chat history and may not generalize to other groups or individuals.
- External factors, such as the nature of conversations or specific events, were not considered in this analysis and could influence results.

Future Work:

- Expanding the dataset by including more participants and diverse groups could provide a broader perspective on communication patterns.
- Considering additional features, such as user activity or conversation topics, may contribute to a more nuanced analysis.

Conclusion:

While the current analysis did not reveal a significant relationship between the time of day and message length in my WhatsApp chat history, the project has provided valuable experience in data preprocessing, exploratory data analysis, and statistical testing. The findings, limitations, and future work outlined here contribute to a comprehensive understanding of the project's scope and potential avenues for further research.