

# Social Media Usage and Emotional Well-Being: Analysis and Predictions

Orlando Marin, Tatiana Eng

# Topics of Discussion

01 Introduction to the Dataset

02 Data Pre-processing

03 Summary Statistics

04 Regression Model

05 Data Visualizations

06 Conclusion and Takeaways

# 01

## Introduction to the Dataset

# Introduction to the Dataset

## (Background Information)

- The Social Media Usage and Emotional Well-Being dataset explores the relationship between social media usage patterns and emotional well-being
- Why did we choose this dataset?
  - Mental Health Awareness – College students often experience stress, anxiety, and other emotional challenges. Understanding how social media affects emotional well-being can help them make informed choices.
  - Comparing Platform Effects – Different platforms may have varying effects on emotions (e.g., Instagram and self-esteem, Twitter and anxiety). This is useful for students to evaluate their habits.
  - Personalized Well-Being Strategies – Students can use data-driven insights to modify their social media habits in ways that support mental health and overall well-being.

# 02

## Data Pre-processing

# Data Pre-Processing

1. Removed rows that contain empty values
2. Swapped 76 rows where “Age” and “Gender” values were in the wrong column
3. Changed the datatype of “Age” from string to int
4. One-hot encoding: Converted categorical values into numeric values for the purpose of using them in our regression model
5. Created our own variable (activity score) to mitigate multicollinearity

# Dataset Features

## Original:

### Qualitative Features:

- Gender
- Social Media Platform
- Dominant Emotion

### Quantitative Features:

- Age
- Daily Usage Time (minutes)
- Posts per Day
- Likes Received per Day
- Comments Received per Day
- Messages Sent per Day

## After Our Modifications:

### Qualitative Features:

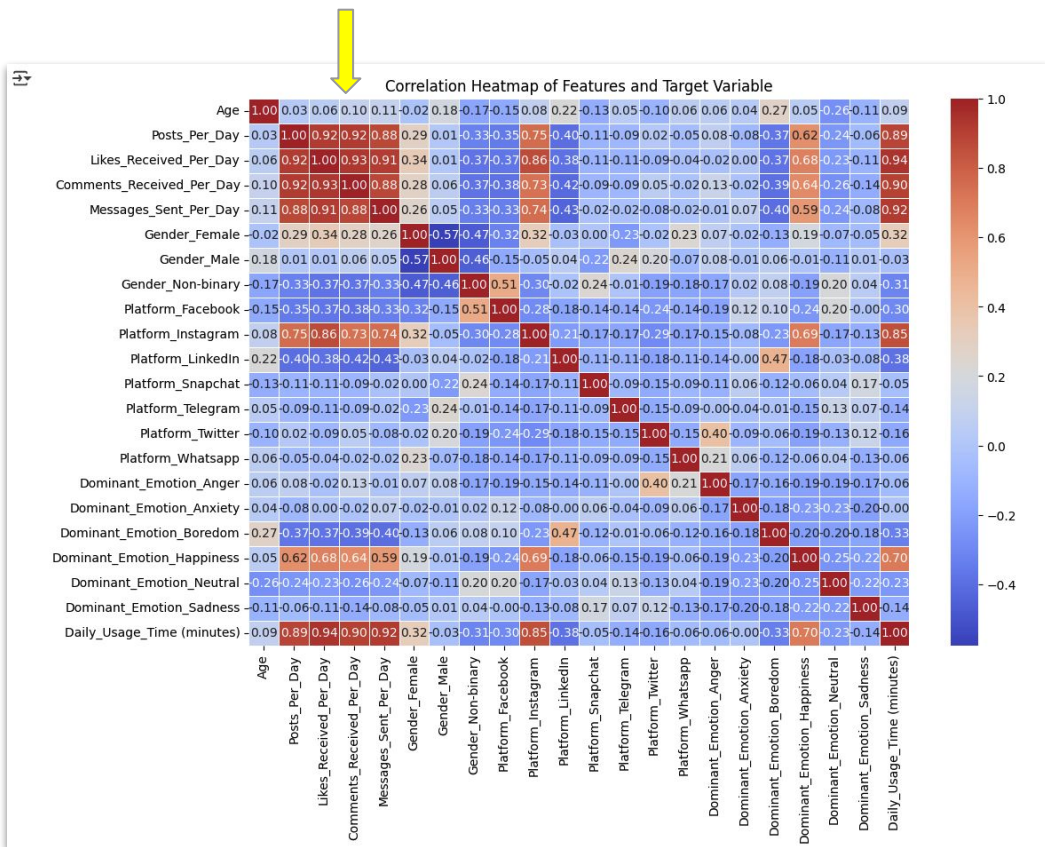
- Gender
- Social Media Platform
- Dominant Emotion

### Quantitative Features:

- Age
- Daily Usage Time (minutes)
- Activity Score
  - Calculated by taking the sum of Posts per Day, Likes Received per Day, Comments Received per Day, and Messages Sent per Day columns

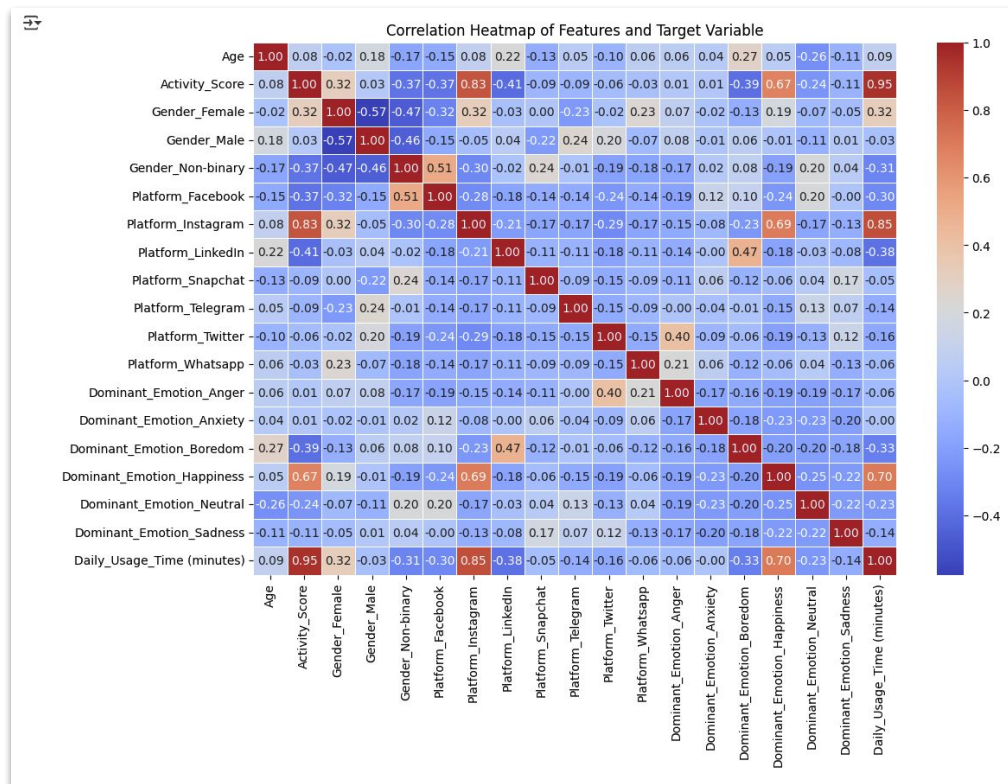
Each user was identified using a unique user ID.

# Data Pre-Processing: Correlation Heatmap Before Adjusting for Multicollinearity





# Data Pre-Processing: Correlation Heatmap After Adjusting for Multicollinearity



Correlation values above 0.75 indicate a likelihood of multicollinearity.

We created our own variable (Activity\_Score) to reduce multicollinearity, as evident in the heatmap.

# Data Pre-Processing: Reducing Multicollinearity

## PROBLEM

- Multicollinearity among the following features:
  - Posts per day
  - Likes received per day
  - Comments received per day
  - Messages sent per day

## SOLUTION

- Created a new feature, Activity Score, which is calculated by taking the sum of the following features:
  - Posts per day
  - Likes received per day
  - Comments received per day
  - Messages sent per day

# 03

## Summary Statistics

# Summary Statistics

	Age	Daily_Usage_Time (minutes)	Posts_Per_Day	Likes_Received_Per_Day	Comments_Received_Per_Day	Messages_Sent_Per_Day	Activity_Score
Mean	27.51	95.95	3.32	39.90	15.61	22.56	81.39
Standard deviation	3.93	38.85	1.91	26.39	8.82	8.52	44.48
Minimum	21	40	1	5	2	8	18
50%	27.50	85	3	33	14	22	73
Maximum	35	200	8	110	40	50	203

# 04

## Regression Model

# Multiple Regression Model

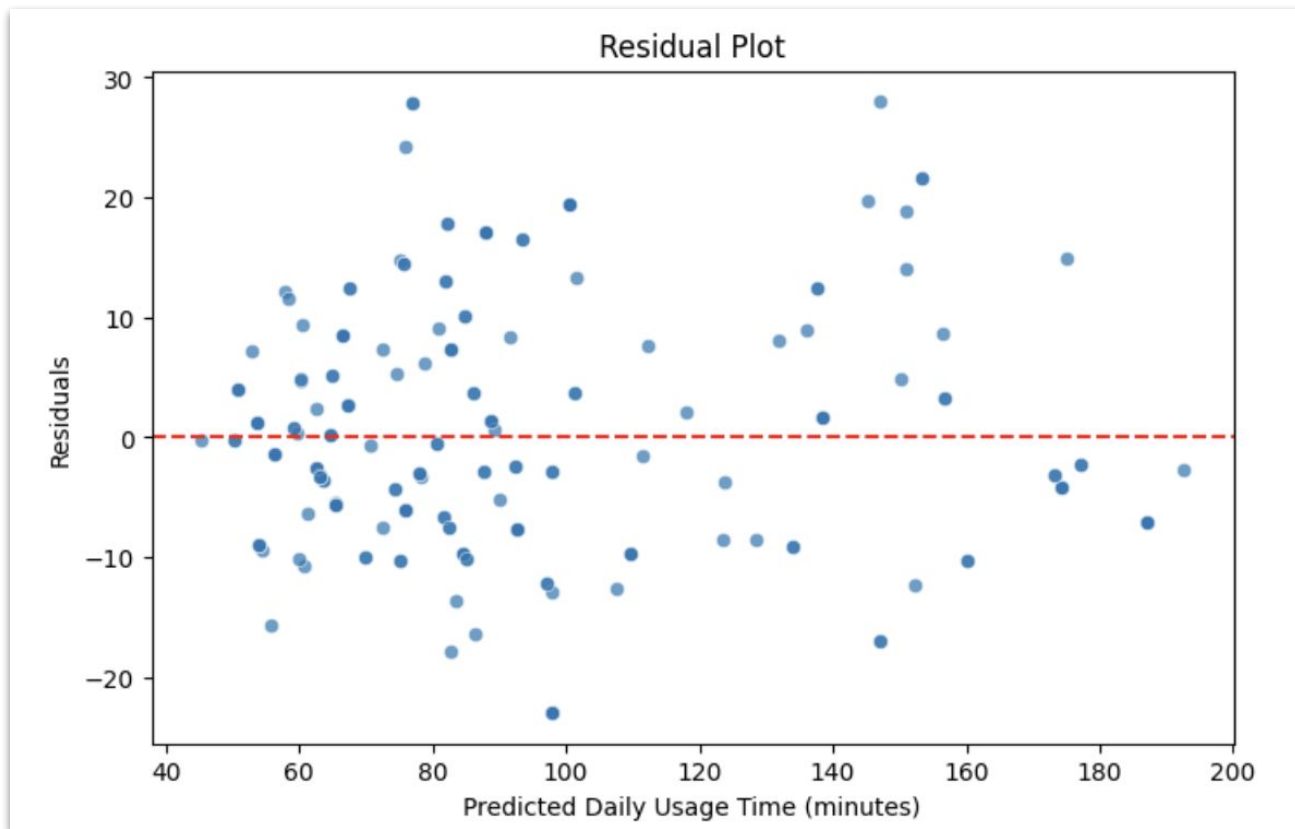
80% training, 20% testing

$$R^2 = 0.92$$

Root mean squared error = 10.35

- On average, the model's predicted Daily Usage Time (minutes) deviates by approximately 10.35 minutes from the actual values in our dataset.

# Analysis of Our Residuals



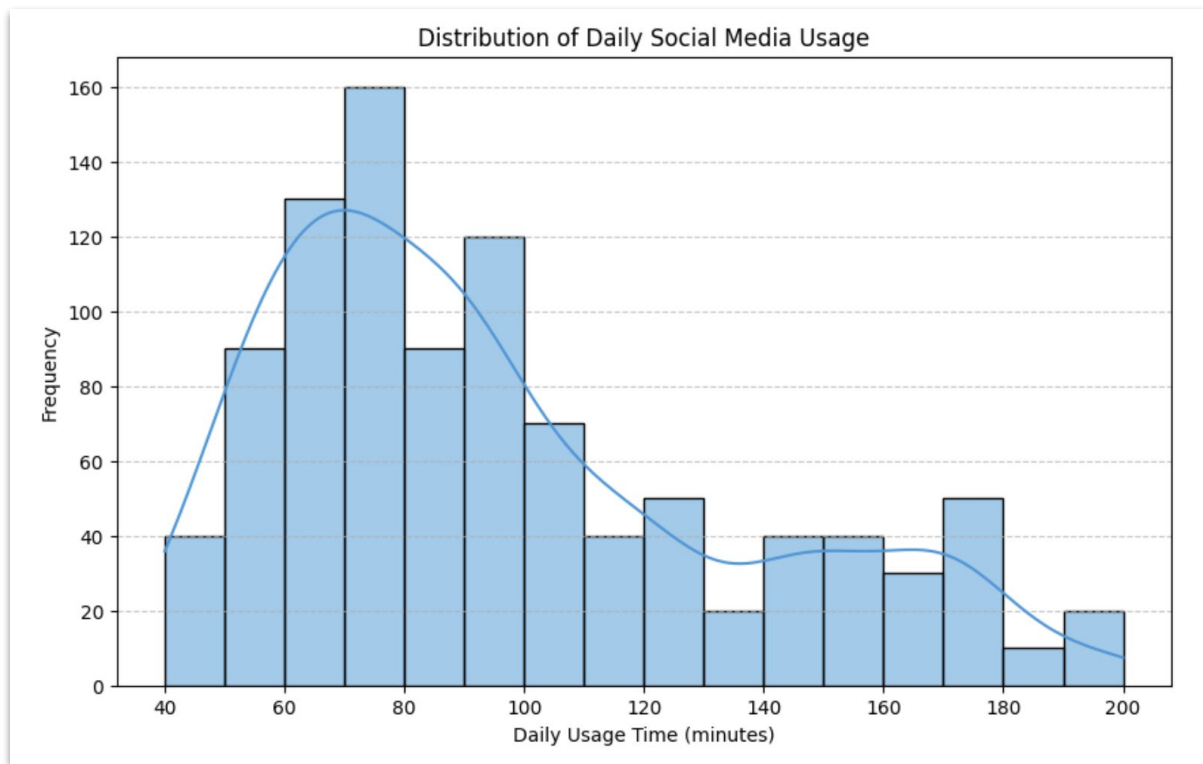
[Source: Code](#)

# 05

## Data Visualizations

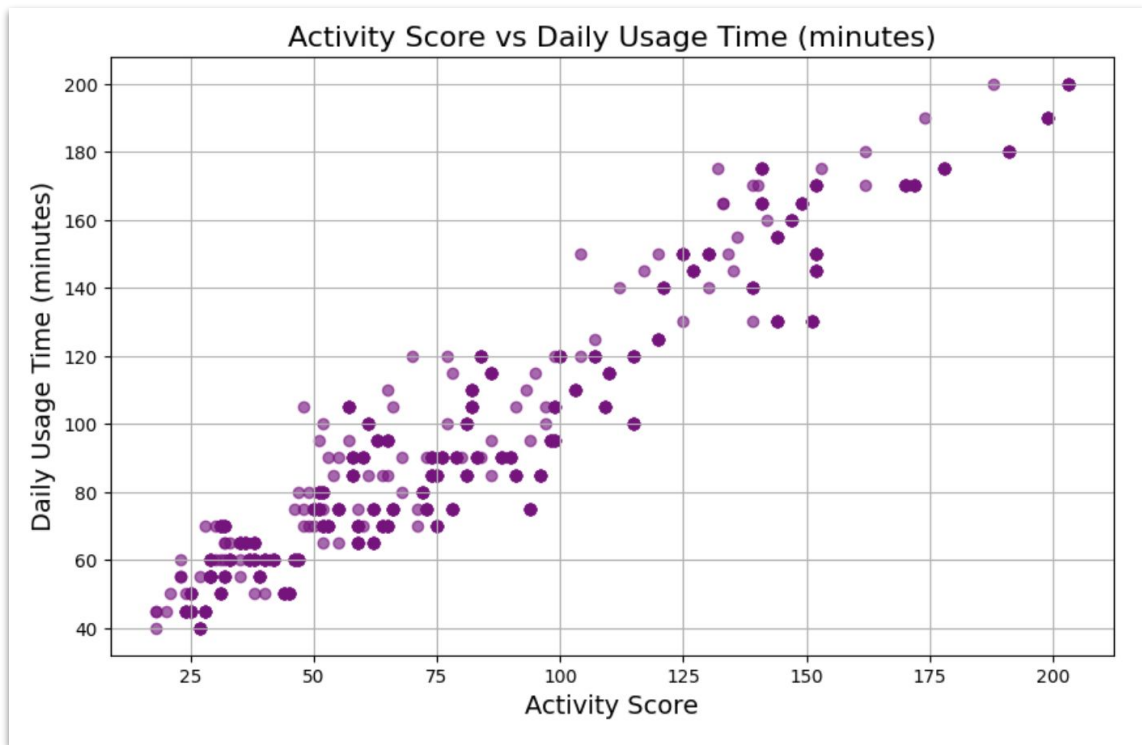


## Social Media Users Average 60 - 100 Minutes per Day



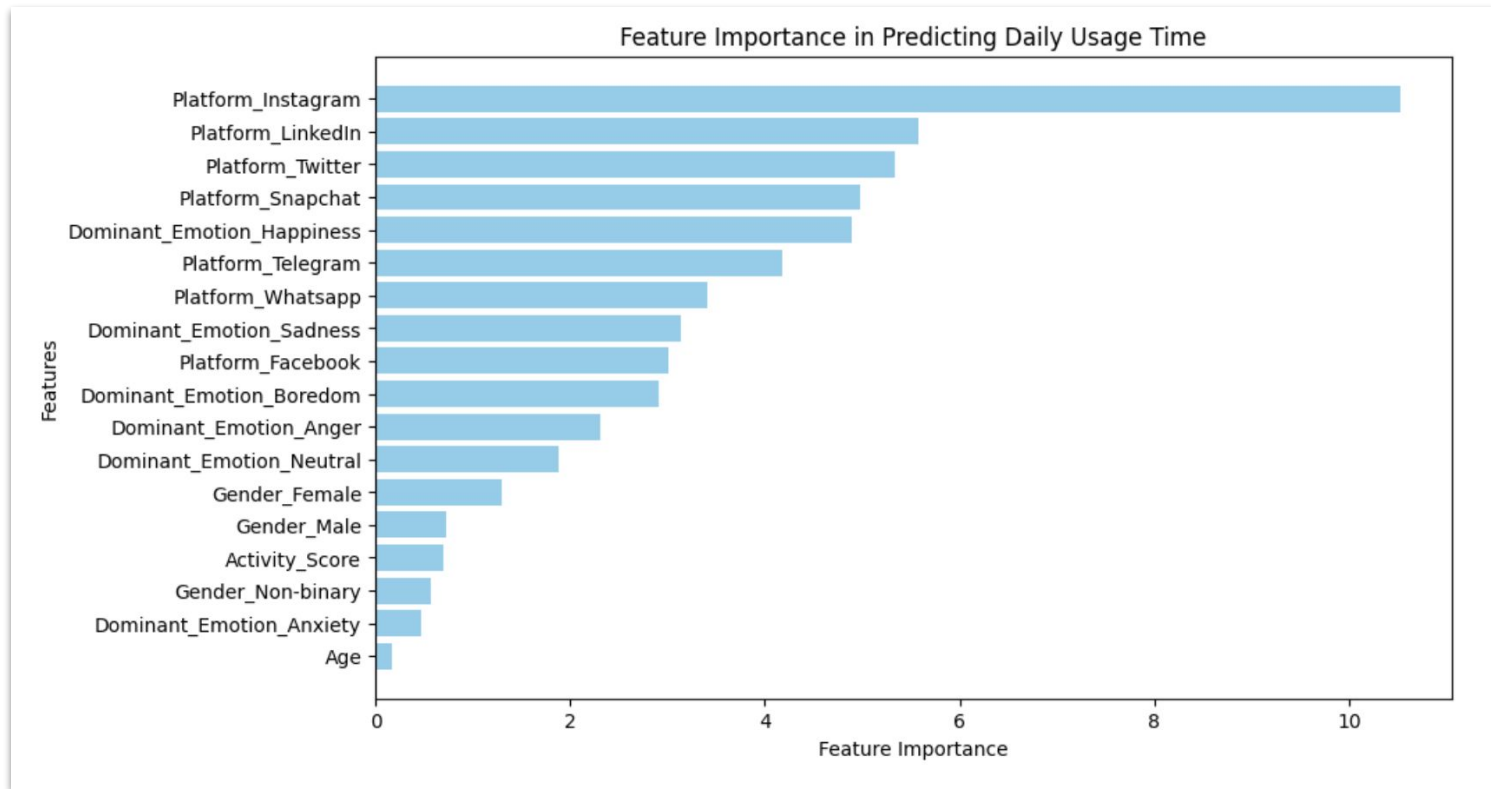
[Source: Code](#)

# Linear Relationship Between Activity Score and Daily Usage Time



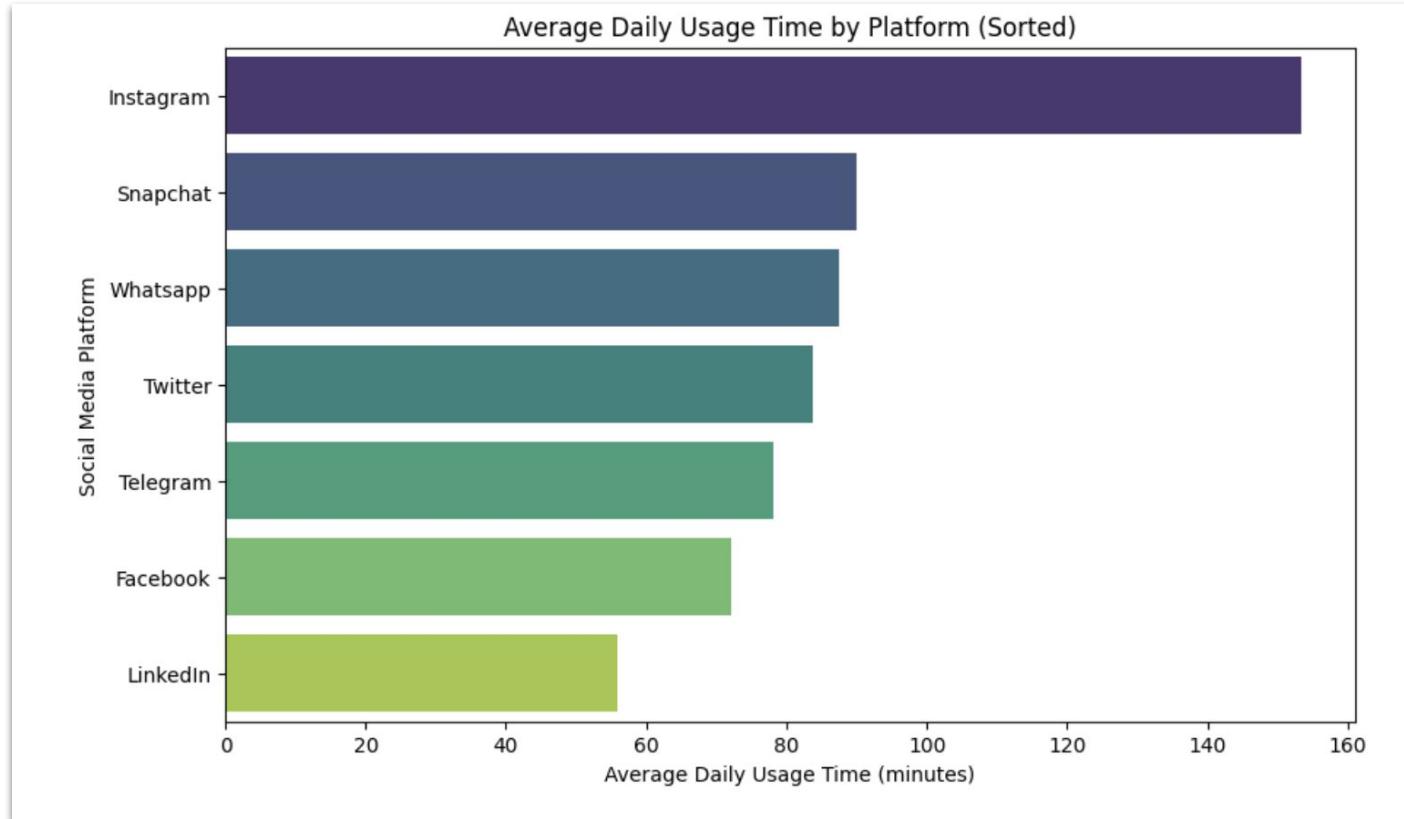
[Source: Code](#)

# Which Features Contribute Most to Predicting Usage Time?



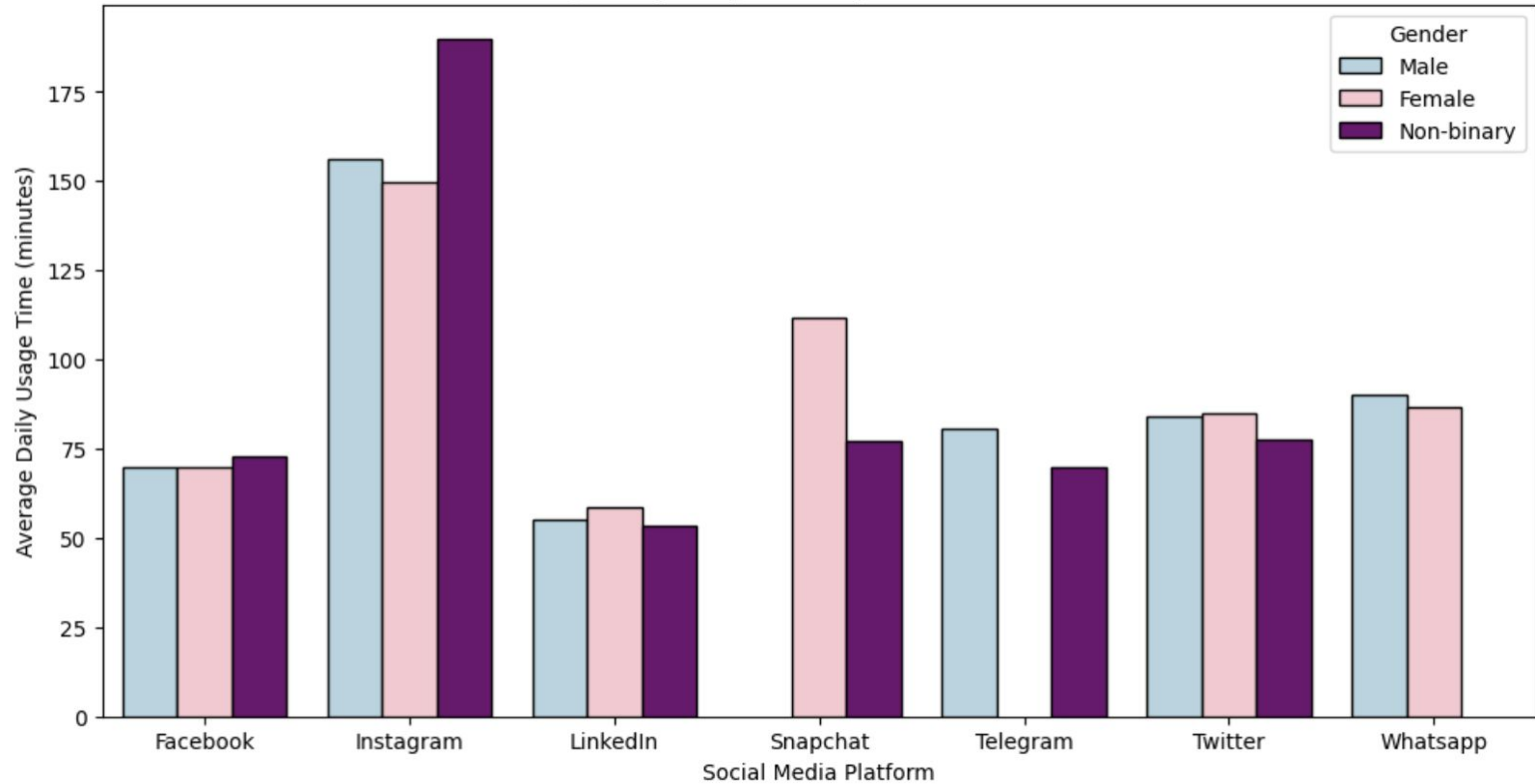
[Source: Code](#)

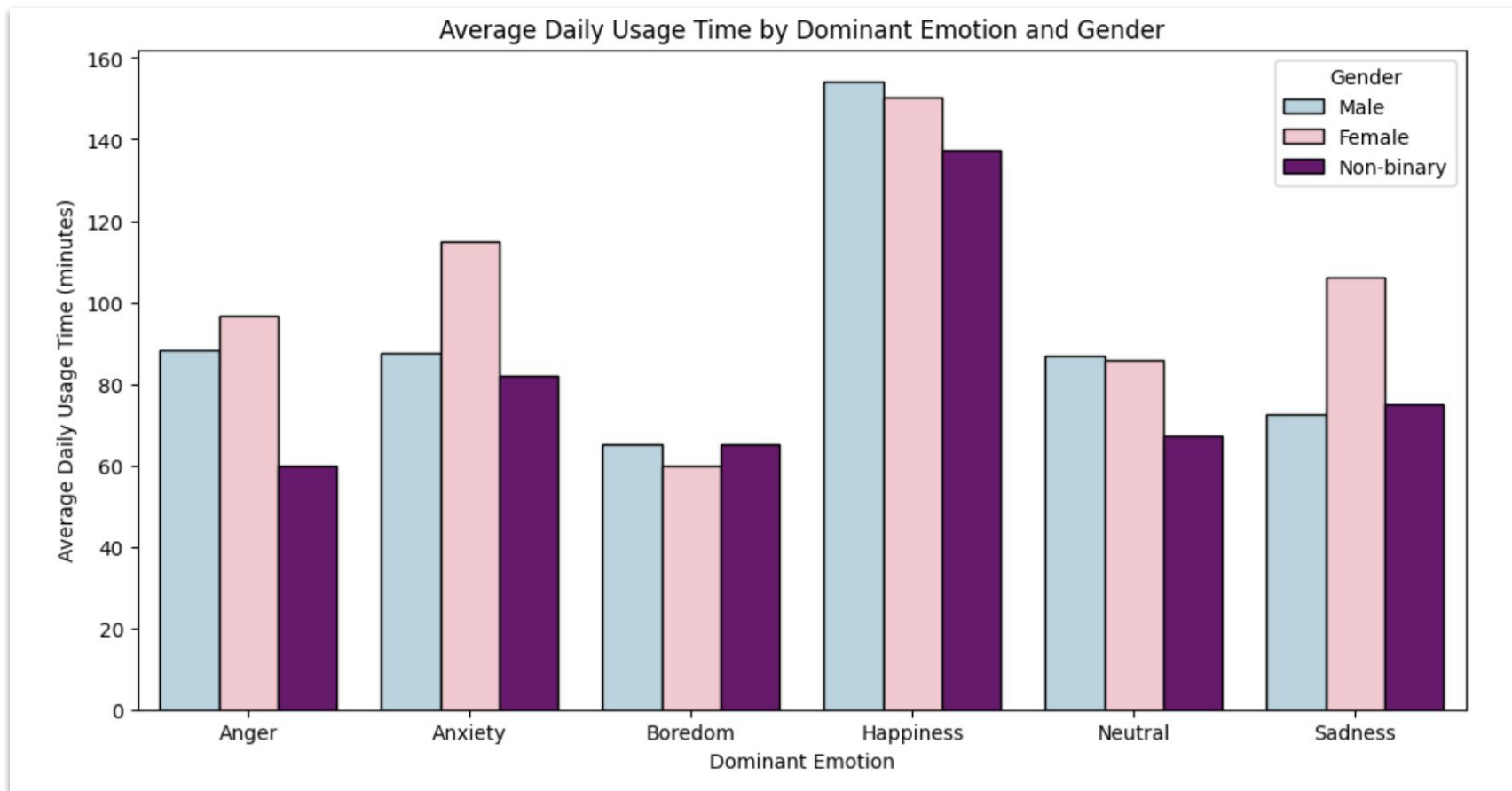
# Instagram and Snapchat Have the Highest Daily Usage



[Source: Code](#)

Average Daily Usage Time by Platform and Gender





# 06

## Conclusion and Takeaways

# Conclusion and Takeaways

## Limiting Factors

- Dataset purely focused on social media usage and didn't consider outside factors
- The mean and median age of the people surveyed was 27.5 years old, which runs the risk of making the results less relevant to students
- Each person surveyed only included social media usage for one platform
- The dataset did not include data on popular social media platforms like TikTok or YouTube, which could impact results

## Suggestions for Future Study

- Gather survey data from college students, including data on multiple social media platforms, and use the data to predict dominant emotion
- Use social media usage data, coupled with data in other areas (education, exercise, sleep) to make further predictions

## Recommendations to Students

- Curate Your Feed – Follow content that positively impacts emotional well-being.
- Take Breaks – Periodic social media detoxing can improve focus and reduce stress.



Thank you!