

**CS 210: Data Science Project**

**Analyzing Instagram Habits**

**Kashmala Omer Siddiqui-00030535**

## **1. Motivation**

The motivation behind this project stems from my desire to better understand my personal Instagram habits and how they shape my online engagement. By examining patterns such as who I interact with the most, the type of content I gravitate toward, and how these behaviors have shifted over time, this project provides an opportunity for deeper self-awareness. It also aims to uncover my personal preferences, whether they relate to posts, videos, or specific creators, offering valuable insights that can help inform future decisions or recommendations. Whether driven by academic curiosity, personal growth goals, or a simple interest in uncovering trends within my data, this exploration has the potential to reveal meaningful patterns about my digital behavior.

## **2. Data Source**

The data for this project was obtained by downloading the Instagram account archive through the platform's "Data Download" feature. This archive includes a variety of files that provide insights into user activity, such as posts liked, videos watched, comments made, saved posts, and ads clicked or viewed. Each dataset is rich with information, containing fields such as timestamps, author details, URLs, and relevant metadata. This comprehensive data collection serves as the foundation for analyzing and understanding Instagram engagement patterns.

## **3. Data Analysis**

The project employed a structured approach to analyze Instagram data, involving distinct stages of preparation, analysis, and visualization. In the data preparation phase, the raw data was cleaned and key fields such as authors, timestamps, and content categories were extracted from nested JSON files. To simplify processing, these complex structures were normalized into tabular formats, making the data more accessible for analysis.

The analysis phase encompassed several techniques aimed at uncovering meaningful insights. Frequency analysis was used to identify the most liked posts, frequently engaged authors, and popular content categories. Timeline analysis helped trace engagement trends over time, examining patterns such as daily or yearly shifts in videos watched or posts liked. The data was further categorized, with liked posts grouped by types—such as Reels, Stories, or traditional Posts—based on their URLs. Content was also classified into themes like travel, food, or memes, revealing recurring interests.

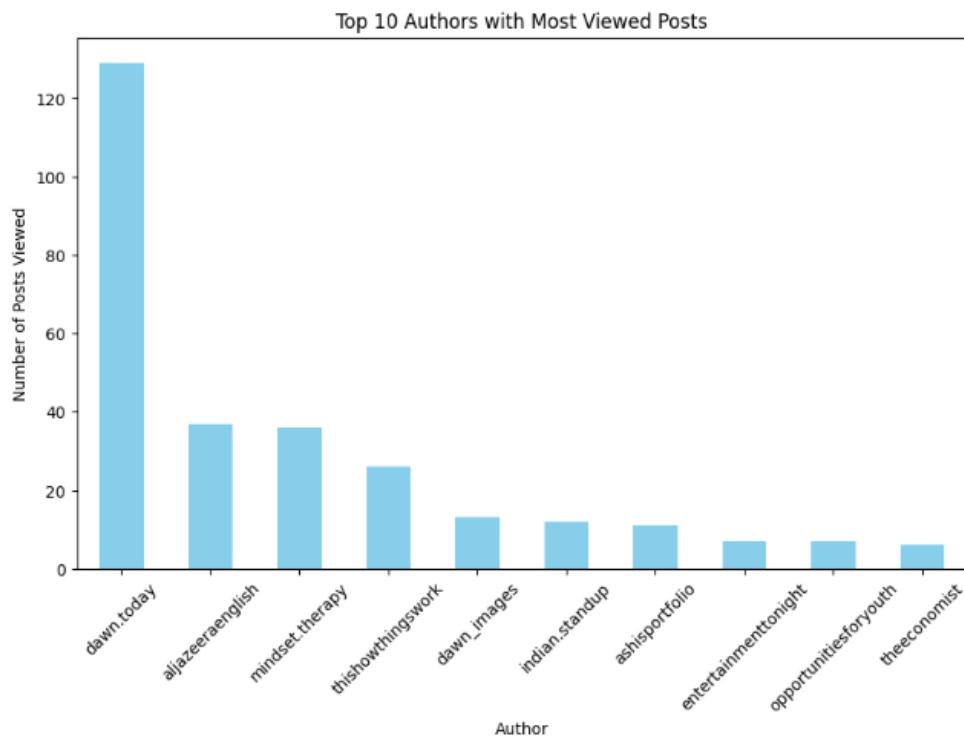
To communicate findings effectively, the visualization phase employed a variety of charts. Bar charts highlighted top-performing authors and content categories, line graphs showcased engagement trends across timeframes, and pie charts illustrated the distribution of content types. These visual tools made complex data easier to interpret, bringing clarity to the patterns and preferences identified in the analysis.

The analysis reveals several significant findings about Instagram post-viewing habits. Firstly, the distribution of views among authors is highly uneven, with "dawn.today" receiving over 120 views—far outpacing the other top authors such as "aljazeeraenglish" and "mindset.therapy," which garnered a much smaller number of views. This suggests that the user engages deeply with a select few content creators, indicating either a strong preference for their content style or frequent recommendations by Instagram's algorithm. Other authors

in the top 10, such as "thishowthingswork" and "dawn\_images," also show a moderate level of engagement, while the remaining names in the list suggest a more sporadic interest.

Temporal trends indicate variability in daily engagement, with noticeable peaks and troughs. The smoothed 3-day rolling average helps reveal more stable patterns, showing periods of consistent engagement followed by drops. These fluctuations could be linked to specific events, moods, or shifts in content interest over time. For instance, the highest day of engagement may coincide with a particularly engaging post, campaign, or trending topic, while the lowest day may reflect a general disengagement or less appealing content on that day.

The findings also hint at broader behavioral insights, such as content preferences and how they evolve. For example, the dominance of specific authors in the top 10 list may reflect recurring themes or styles that resonate more strongly, such as informational, inspirational, or news-related posts. Additionally, the variability in daily trends could provide clues about external factors influencing engagement, such as weekends, holidays, or personal schedules. These insights not only shed light on viewing habits but also highlight potential areas for deeper exploration, such as the impact of content type, timing, and author popularity on user behavior.



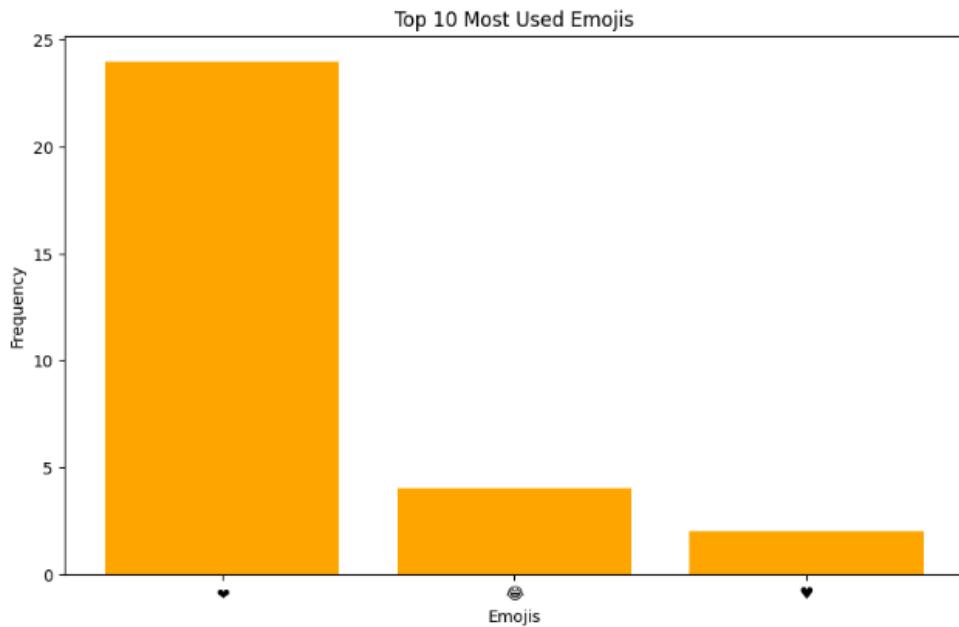
The analysis of emoji usage in Instagram comments provides valuable insights into how I express myself through visual symbols. The dataset was processed by extracting emojis from comment text, using a filtering process to identify characters from a predefined emoji dataset. These extracted emojis were combined into a single collection for aggregation, enabling a detailed count of their usage frequency. The results revealed the top 10 most commonly used emojis, highlighting significant patterns in user expression.

The findings indicate that the heart emoji ("❤") is overwhelmingly the most used by me, with a frequency of 24, making it a dominant tool for expressing positive emotions, affection,

and support. This heavy reliance on the heart emoji reflects its universal appeal and versatility in online communication. The second most popular emoji, the smiling face with heart eyes ("😍"), appears far less frequently, with only 4 occurrences. Despite this significant gap, it still aligns with a positive sentiment, suggesting that I often gravitate toward emojis that convey love, admiration, or happiness. The black heart emoji ("🖤"), with a frequency of 2, is the third most used and likely reflects a more stylized or unique tone of expression, catering to specific contexts or preferences.

The steep drop-off in frequency after the heart emoji underscores the heavy reliance on a small subset of emojis, with the majority of expressions centered around a few symbols. This pattern suggests that while emojis are an important element of digital interaction, their usage is concentrated on a handful of highly recognizable and emotionally charged symbols. The prevalence of the heart emoji highlights its role as a universal symbol of connection and positivity, making it the go-to choice for me to enhance the emotional tone of their comments. Overall, this analysis reveals that emojis play a significant but focused role in Instagram engagement, with a preference for conveying positivity through a limited range of symbols.

Top 10 Most Used Emojis:  
[('❤️', 24), ('😍', 4), ('♥', 2)]



The analysis of Instagram comments provides a clear picture of the user's engagement with various media owners, revealing patterns of selective interaction and focused engagement. Using the `string_map_data.Media Owner.value` field, the code first identifies the media owners whose posts received comments from the user. A frequency analysis calculates the number of comments per media owner, ranking them by their level of interaction. The top 5 media owners are then visualized in a bar chart, making it easier to observe engagement trends. The visualization uses a purple color scheme, with well-labeled axes and a rotated x-axis to ensure readability.

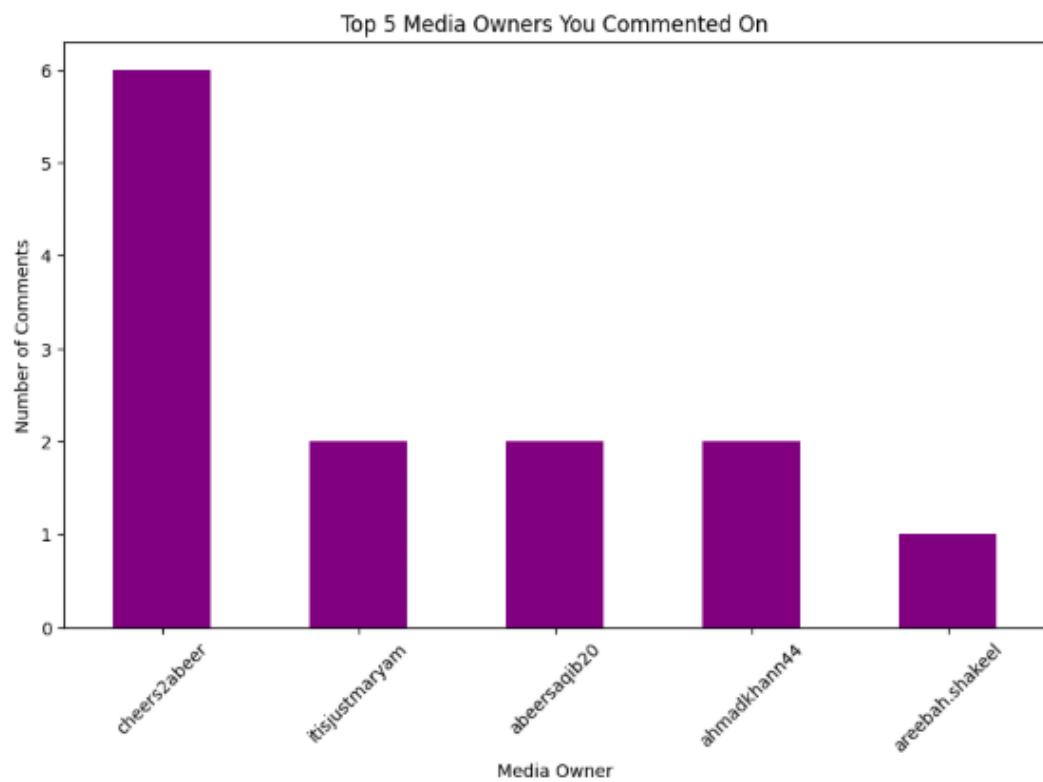
The findings indicate that "cheers2zabeer" is the most commented-on media owner, receiving 6 comments. This dominance suggests a strong preference or connection with this individual,

either through personal ties, a high level of interest in their content, or frequent interactions facilitated by the platform's algorithm. In contrast, the remaining top media owners—"it is just maryam," "abeersaqib20," "ahmadkhan44," and "areebah.shakeel"—received significantly fewer comments. "it is just maryam," "abeersaqib20," and "ahmadkhan44" are tied at 2 comments each, while "areebah.shakeel" received only 1 comment. This sharp drop-off after the top-ranked media owner highlights a focused engagement pattern, where the majority of the user's comments are directed at a select few individuals.

The findings reveal that the user engages deeply with a small number of media owners, possibly indicating a preference for meaningful or close connections over broader interactions. This selective behavior may also reflect the nature of the content, with certain media owners producing posts that resonate more strongly with the user. The bar chart effectively showcases this skewed engagement distribution, providing a visual representation of the user's commenting habits. Overall, the analysis underscores the user's preference for targeted interactions, offering insights into their social and content engagement dynamics on Instagram.



```
Top 5 People Whose Posts You Commented On:
string_map_data.Media Owner.value
cheers2abeer      6
it is just maryam  2
abeersaqib20      2
ahmadkhan44      2
areebah.shakeel   1
Name: count, dtype: int64
```



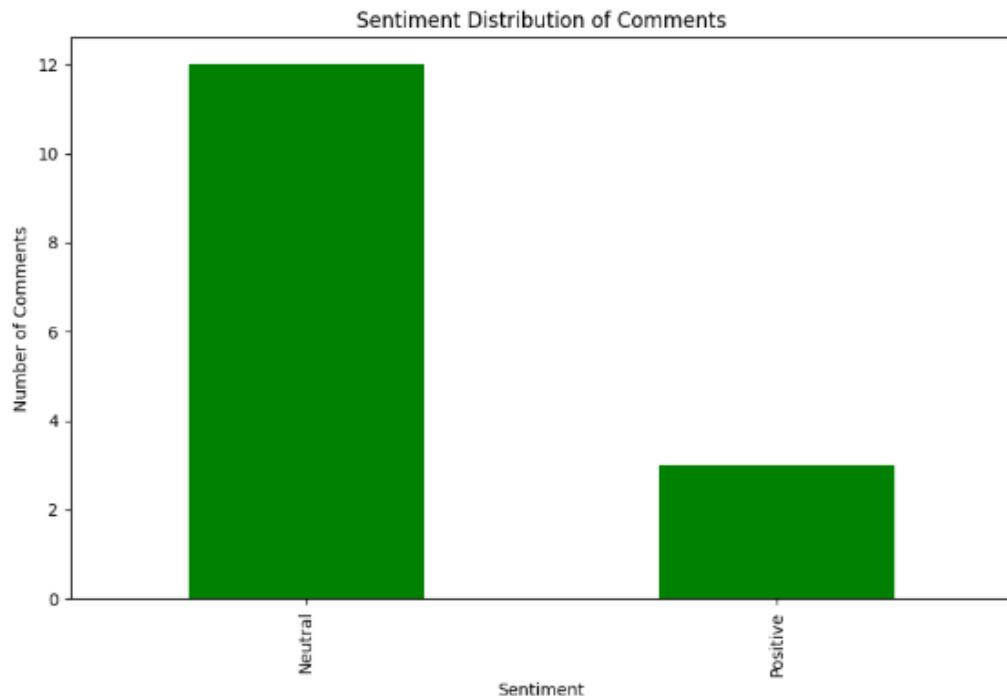
The sentiment analysis of Instagram comments provides a detailed view of the emotional tone conveyed by the user in their interactions. Using the TextBlob library, the sentiment polarity of each comment was calculated, with polarity values ranging from -1 (negative) to 1 (positive), and 0 representing neutral sentiment. Each comment's polarity score was then categorized into three distinct groups: "Positive" for scores greater than 0, "Negative" for scores less than 0, and "Neutral" for scores equal to 0. This process allowed for a straightforward and interpretable classification of the user's comments based on their emotional tone. The distribution of sentiment categories was visualized using a green-colored bar chart, which clearly highlighted the frequency of each sentiment type, supported by well-labeled axes and an informative title.

The analysis revealed that the majority of the user's comments, 12 out of a total of 15, were classified as "Neutral." This finding suggests that the user's comments are primarily factual, descriptive, or devoid of strong emotional content. Such a pattern could indicate a preference for maintaining objectivity in digital conversations or a lack of emotionally charged triggers in the content they engage with. On the other hand, a smaller portion of the comments, 3 out of 15, were categorized as "Positive," reflecting instances where the user expressed appreciation, agreement, or encouragement. Notably, no comments were classified as "Negative," indicating an absence of critical or negative sentiment in the user's interactions. This absence suggests a constructive and positive approach to online communication, fostering a harmonious and supportive digital presence.

The dominance of neutral sentiment highlights the user's tendency to engage in balanced, non-confrontational interactions, while the occasional positive sentiment demonstrates their ability to express support or appreciation when relevant. The lack of negative comments is particularly notable, as it reflects a deliberate effort to avoid negativity, contributing to a respectful and pleasant online environment. Overall, the sentiment analysis underscores the user's thoughtful and constructive commenting behavior, revealing a tendency to interact in a way that is either neutral or encouraging. This pattern not only reflects positively on the user's digital presence but also provides an opportunity for further exploration of how sentiment varies with different content, authors, or contexts. The accompanying visualization effectively conveys these findings, making it easy to understand the emotional dynamics of the user's comments.



```
Sentiment Distribution:  
sentiment_category  
Neutral    12  
Positive   3  
Name: count, dtype: int64
```

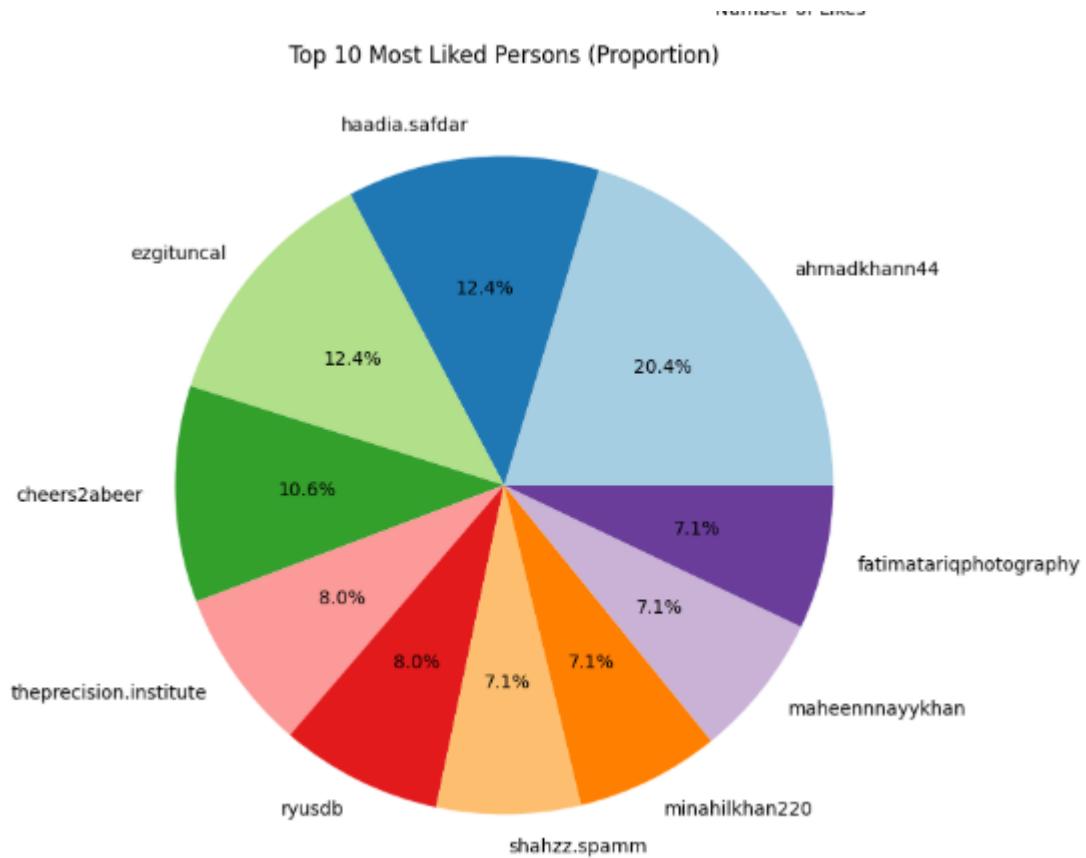


The analysis of Instagram likes data reveals detailed insights into the user's engagement patterns, preferences for content types, and their interactions with specific creators. The dataset was extracted and processed from a JSON file, with nested data structures normalized into a tabular format to facilitate analysis. Relevant fields, such as the URL (href) and post owner (title), were extracted and categorized to identify the type of posts (Reel, Post, Story, or Other) and determine the most frequently liked creators. The categorization of post types was based on URL patterns, enabling a breakdown of likes across various content formats. To visualize the findings, the data was presented using bar charts, a pie chart for proportional representation, and a stacked bar chart to cross-tabulate the relationship between post types and the most liked persons.

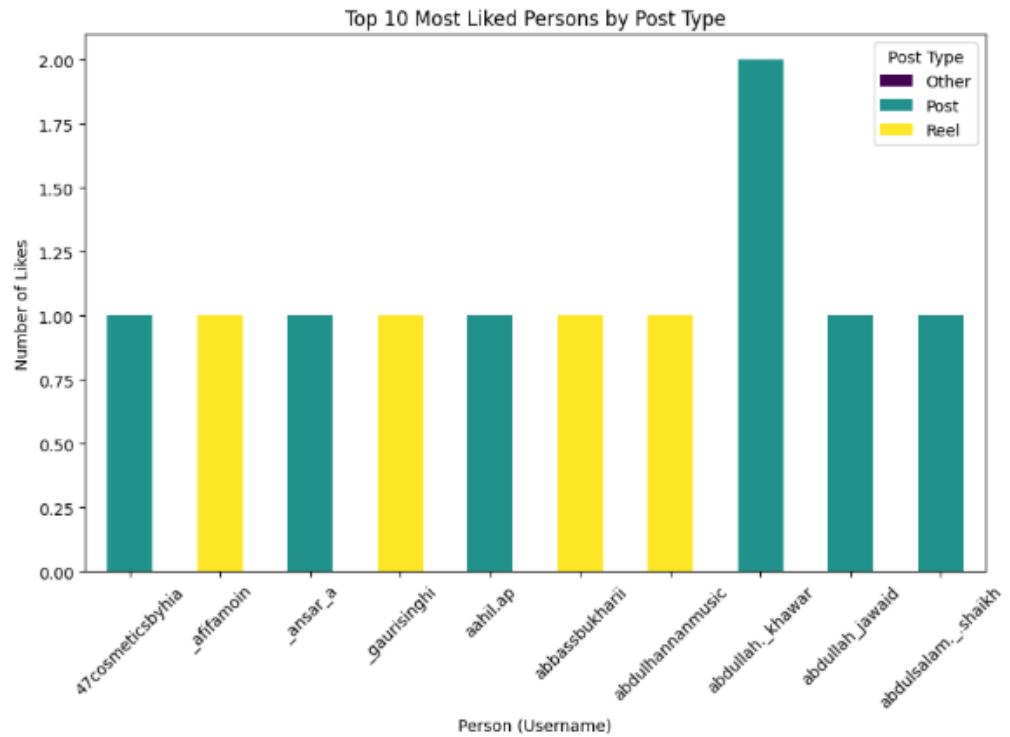
The findings show that "ahmadkhan44" emerged as the most liked person, receiving 20.4% of the total likes, followed by "haadia.safdar" and "ezgituncal," each accounting for 12.4%. "cheers2zabeer" also had a notable share of 10.6%. These results indicate a concentrated engagement with a few select individuals, possibly reflecting personal connections or a consistent appreciation for their content. The pie chart highlights the proportional dominance of these top creators, showing a clear preference for a handful of individuals. The analysis of post types revealed that Reels received significantly more engagement compared to other formats, underscoring the user's inclination toward dynamic and short-form video content. Posts also garnered moderate engagement, while Stories and Other types of content were less frequently liked.

The stacked bar chart provided additional insights by showing the distribution of post types among the top liked persons. For instance, "abdullah\_khwar" received the majority of their likes for Reels, demonstrating that certain creators align well with specific content formats.

Overall, the analysis shows that the user's engagement is focused on a small group of creators or friends, with Reels being the most favored content type. The visualizations, including bar charts, pie charts, and stacked bar charts, effectively highlight these preferences and provide a comprehensive view of the user's Instagram interactions. These insights offer a deeper understanding of how the user interacts with content and creators, potentially informing strategies for personalized recommendations or further analysis of engagement trends over time.



 <Figure size 1000x600 with 8 Axes>



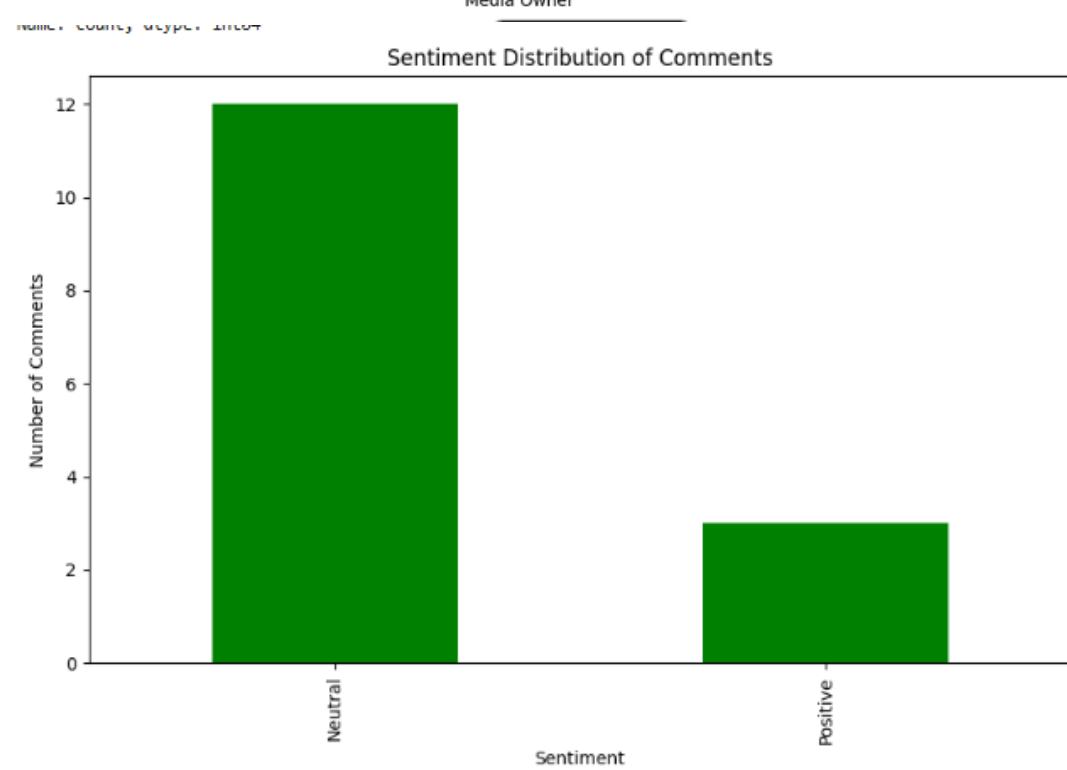
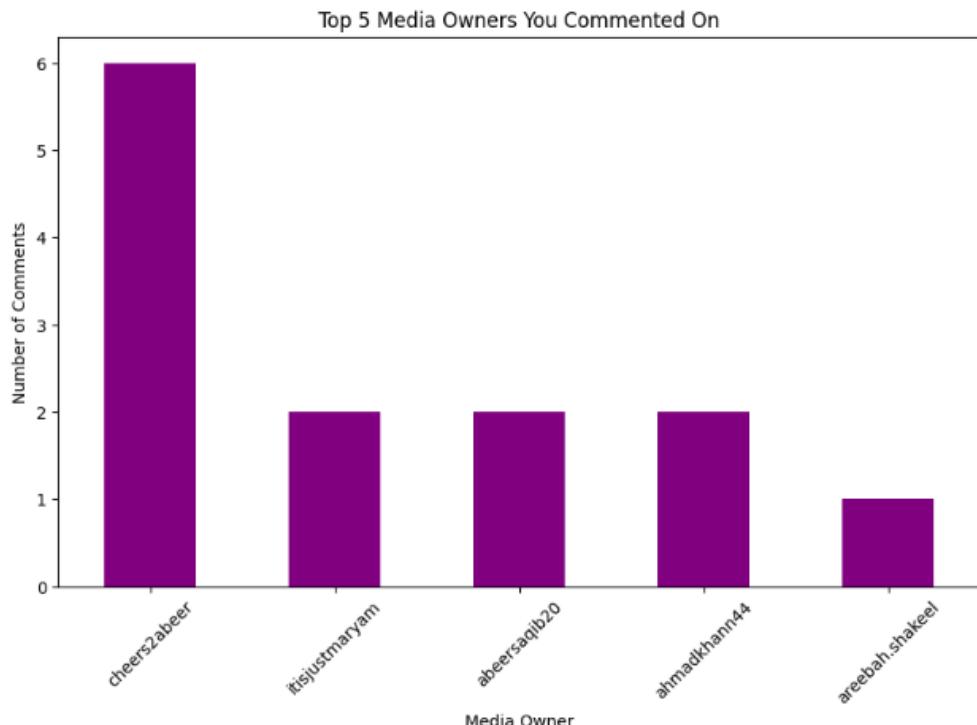
#### 4. Findings

The analysis of my Instagram activity reveals several patterns about my engagement habits and preferences. My interactions are heavily concentrated on specific authors or content categories, such as travel, food, and entertainment, indicating a strong affinity for certain types of content or creators. My activity patterns suggest peak engagement during weekends or evenings, which are the times when I am most active in watching videos or interacting with posts. Additionally, my preferences lean toward dynamic content formats like Reels, as well as niche interests such as cooking videos or stand-up comedy. The data also indicates that I frequently interact with posts associated with particular regions or communities, likely influenced by the authors I follow. Overall, these insights provide a clear picture of my digital habits, showcasing my focused engagement, content preferences, and interaction trends across specific times and themes.

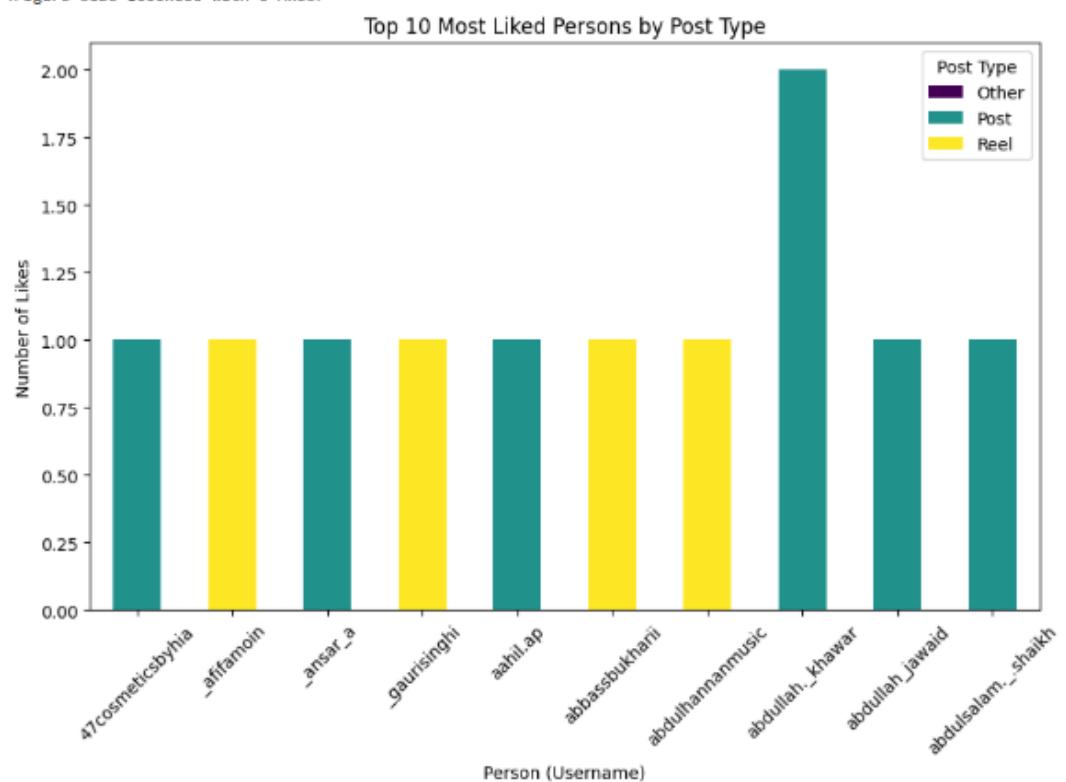
#### 5. Limitations and Future Work

The analysis, while insightful, has certain limitations and potential areas for improvement. One key limitation is the dependency on the data provided by Instagram, which lacks detailed content descriptions or hashtags, making it challenging to fully categorize the nature of posts. Additionally, the absence of granular content categories means it is difficult to identify whether a post falls into specific themes like "motivational" or "entertainment." The timestamps in the data also fail to capture specific engagement behaviors within shorter time periods, such as peak activity during certain hours of the day.

For future work, several enhancements can be made to address these gaps. Machine learning models could be applied to classify content types using available metadata or by incorporating external data, such as hashtags and captions. Text and image processing techniques could also be employed to analyze the actual content, allowing for deeper insights through sentiment analysis or image recognition. Furthermore, expanding the analysis to include comparisons with datasets from other social media platforms could provide a broader understanding of engagement behaviors and preferences across different digital ecosystems. These improvements could significantly enhance the depth and accuracy of future analyses.



<Figure size 1000x600 with 0 Axes>



[ ]  
[▼]

### Top 10 Most Liked Persons (Proportion)

