

# Instagram Data

## Introduction

The purpose of this report is to analyze and decode a decade's worth of social media data from one of Playhouse Communication's high-profile clients. The data, which spans across four major social media platforms - Facebook, Instagram, LinkedIn, and Twitter, provides a unique opportunity to gain deep insights into the client's digital engagement patterns.

Playhouse Communication, a leading digital marketing agency in Nigeria, has been managing the client's social media platforms for the past 10 years. This report aims to transform the raw data from these platforms into actionable insights that could redefine the future of digital marketing for the client.

The data has been provided in both .csv and .xls formats, allowing for comprehensive analysis using various data processing and analytics tools. The goal is not just to understand past trends but also to predict future ones, thereby providing the client with a strategic advantage in their digital marketing efforts.

This isn't just a hackathon; it's an opportunity to shape the future of digital engagement. By leveraging this exclusive social media data, we aim to strike gold with game-changing insights that could potentially revolutionise digital marketing strategies for Playhouse Communication and its clients.

## Methodology

The methodology for analysing the social media data involves several steps:

- 1. Data Cleaning:** The first step in the process is data cleaning. This involves handling missing values, removing unnecessary columns, and converting data types where necessary. For instance, any columns with more than 50% missing values are dropped from the data.
- 2 Exploratory Data Analysis (EDA):** This step involves understanding the distribution of data and relationships between different variables. For example, using `time_series_analysis` to plot various metrics over time.
- 3. Topic Modeling using Latent Dirichlet Allocation (LDA):** LDA is used to discover the main topics that occur in the collection of social media posts. This unsupervised machine learning technique assumes that each document (or post) is a mix of various topics, and

each topic is a collection of words. By looking at which words often appear together, LDA can figure out the possible topics and how much each document belongs to each topic.

**4. Named Entity Recognition (NER) and Part-of-Speech Tagging (POS):** These Natural Language Processing (NLP) techniques are used to extract more information from the text data. NER can tell us which specific entities (like people, places, or organisations) are being mentioned in the posts, while POS tagging helps us understand the context better by knowing the grammatical role of each word.

**5. Data Visualization:** The last step involves visualising the results of our analysis. This could be as simple as bar plots or pie charts showing the distribution of posts across different topics, or more complex visualisations like word clouds that highlight the most frequently mentioned entities in each topic.

The tools used for this analysis include Python for data cleaning and analysis, various libraries like pandas for data manipulation, matplotlib and plotly for data visualization, and sklearn and nltk for machine learning and NLP tasks.

## Findings

**Impressions:** This is the total number of times your content, whether a post or a story, was shown to users. If your Instagram impressions are higher than your reach, it's a sign that your audience is viewing your content multiple times.

**Organic Impressions:** This metric represents the number of times your post appeared in people's newsfeeds without any paid promotion. It gives you an indication of the number of people who may have seen the post, thus giving you an idea of your brand awareness growth, organic reach, and the number of impressions.

**Reach:** This is the number of unique users that saw your Instagram post or story on any given day. Unless you're putting some serious spend behind your Instagram posts, you probably won't have every single one of your followers see your content. That's why your Instagram reach is important. Reach tells you how many unique users are served your content every day.

**Organic Reach:** As the name indicates, the organic reach can be gained by the ordinary activities that you have on Instagram, such as posting photos and videos on the feed, IGTV, or Stories. It does not include reach gained through paid promotions.

## Data Preprocessing

The data preprocessing stage involved several crucial steps to ensure the data was in an appropriate format for analysis.

## Conversion of Data Types

Initially, the 'Impressions', 'Organic Impressions', 'Reach', 'Organic Reach', columns were categorical. To facilitate better analysis, these columns were converted to numerical data types. The conversion process involved removing commas from the values and converting the data type to integer. The code snippet for this process is as follows:

```
```python
cleaned_impressions_reach[impression_cols] =
cleaned_impressions_reach[impression_cols].apply(
    lambda row:row.str.replace(',','').astype("int"), axis=1)

cleaned_impressions_reach[reach_cols] = cleaned_impressions_reach[reach_cols].apply(
    lambda row:row.str.replace(',','').astype("int"), axis=1)
...
```
```

## Handling Missing Data

Upon checking for missing data, it was found that there were 1484 missing values in both the 'Impressions' and 'Organic Impressions' columns. Given the importance of these columns for analysing post performance, rows with missing data for these columns were dropped. The code snippet for this process is as follows:

```
```python
cleaned_impressions_reach = cleaned_data

for col in cleaned_impressions_reach[impression_cols].columns:
    cleaned_impressions_reach =
cleaned_impressions_reach[cleaned_impressions_reach[col].notna()].reset_index(drop=True)
...
```
```

These preprocessing steps were essential in ensuring that the subsequent analysis was based on complete and appropriately formatted data.

## Feature Engineering

Feature engineering is a crucial step in any data analysis or machine learning project. It involves creating new features from existing ones to better capture the underlying patterns in the data. In this project, several new features were engineered from the date and time of each post.

### Time-Based Features

The date and time when a post was made can have a significant impact on its performance. To capture this, the following features were created:

- 'year', 'month\_name', 'month', 'day\_name', 'day', 'hour', 'minute': These features represent the exact time when a post was made.
- 'day\_of\_week': This feature represents the day of the week (Monday, Tuesday, etc.) when a post was made.
- 'time\_period': This feature categorizes the hour of the day into different time periods (morning, afternoon, evening, night).
- 'quarter': This feature represents the quarter of the year when a post was made.
- 'season': This feature represents the season (spring, summer, etc.) when a post was made.
- 'day\_period': This feature categorizes the day of the week into weekdays or weekends.

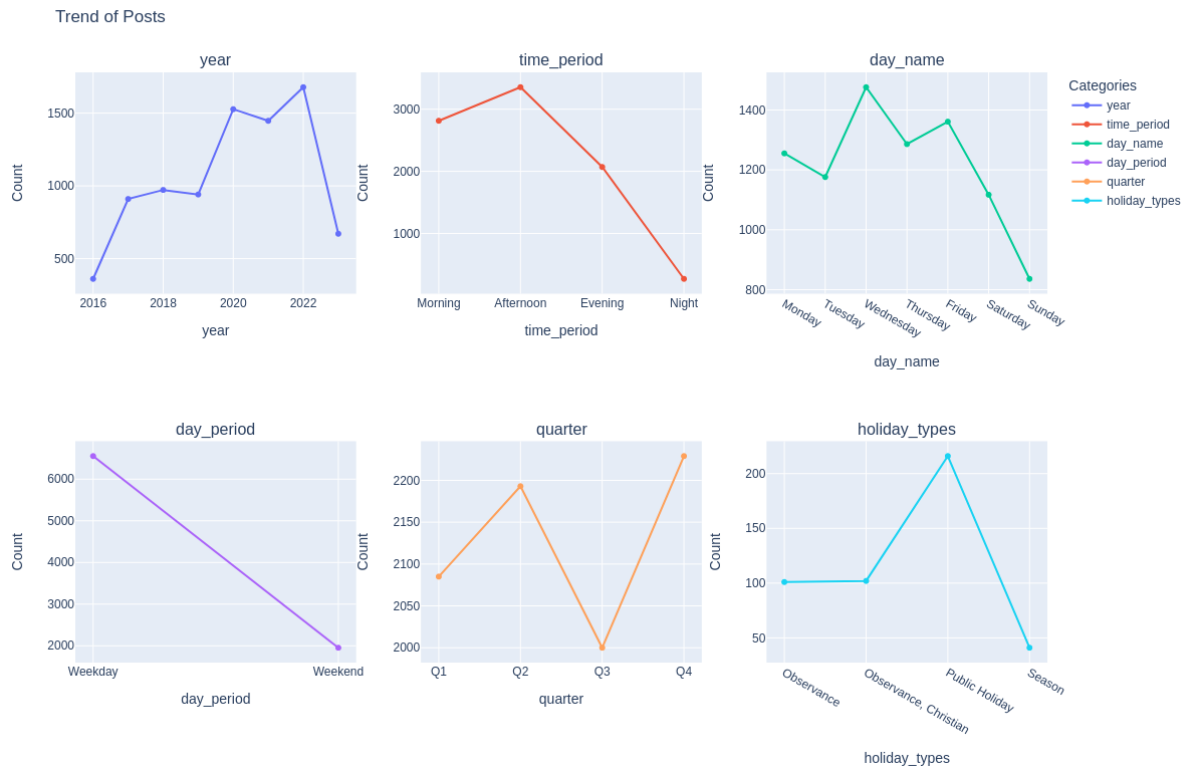
### Holiday Features

Holidays can also affect post performance as user behavior might change during these days. To capture this, a list of Nigerian holidays was used to create two new features:

- 'holiday\_names': This feature indicates whether a post was made on a holiday, and if so, which holiday it was.
- 'holiday\_types': This feature indicates the type of holiday (if any) on which a post was made.

These newly engineered features provide additional information that can help in understanding patterns in post performance. For instance, posts made during certain time periods or on certain holidays might receive more impressions or engagement. By including these features in our analysis, we can ask more detailed questions and gain deeper insights into your data.

### Exploratory Data Analysis



## TREND OF POSTS

The graph above shows the count of posts made each year from 2016 to 2023. Here's what we can infer from it:

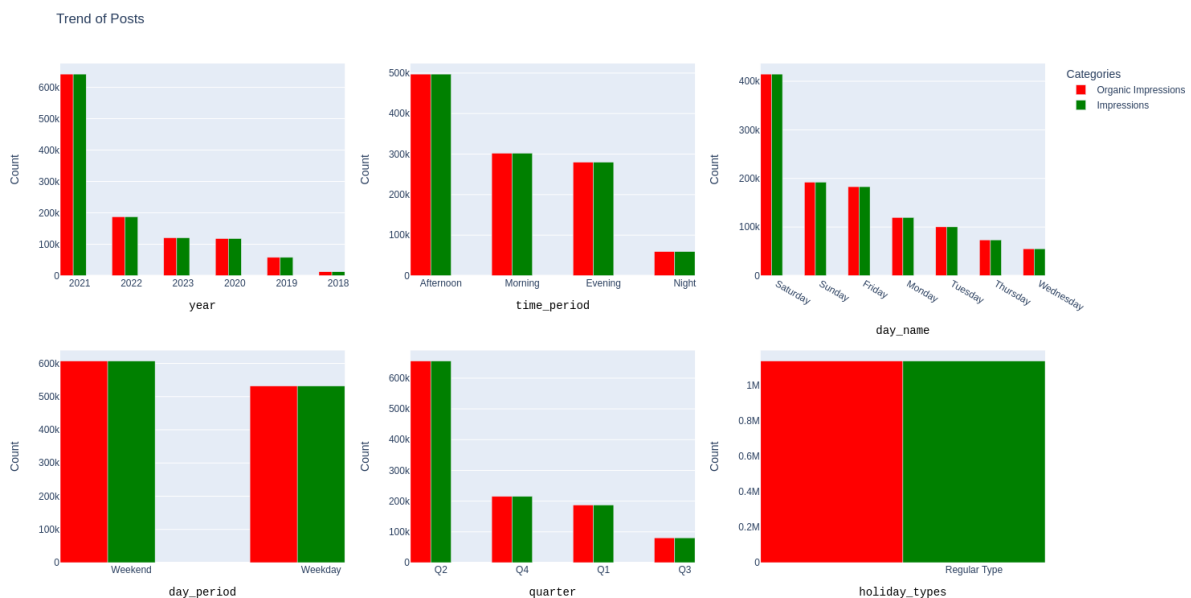
- **Increase in Activity Over Time:** There is a general trend of increasing activity over the years, with the number of posts rising from 361 in 2016 to a peak of 1678 in 2022. This could indicate a growing use of the platform or an increase in content production strategies over time.
- **Peak in 2022:** The highest number of posts were made in 2022, which might suggest a particularly active year for content creation. This could be due to various factors such as increased marketing efforts, special events, or changes in content strategy.
- **Drop in 2023:** There's a noticeable drop in the number of posts in 2023 compared to the previous year. However, it's important to note that the data for 2023 might not be complete since we're still within that year.
- **Time of Day:** Most posts are made in the afternoon (3354 posts), followed by morning (2814 posts), and evening (2070 posts). The least activity is at night (269 posts). This could indicate that the page is most active during daylight hours, possibly to match peak user activity times.
- **Day of the Week:** Wednesday sees the highest number of posts (1476 posts), while Sunday has the least (836 posts). This could suggest a strategy of posting more on weekdays when users might be more active.

- **Weekday vs Weekend:** There are significantly more posts on weekdays (6554 posts) compared to weekends (1953 posts), indicating a focus on reaching users during the workweek.
- **Quarterly Distribution:** The number of posts is fairly evenly distributed across quarters, with Q4 seeing slightly more activity (2229 posts). This could suggest a consistent content strategy throughout the year.
- **Holiday Types:** Most posts fall under 'Regular Type' (8047 posts), with fewer posts on 'Public Holidays' (216 posts), 'Observances' (101 posts), and 'Christian Observances' (102 posts). There are very few posts in 'Season' category (41 posts). This could indicate that the page maintains regular activity most of the time, with some increase in content during public holidays.

|                            | count  | mean        | std         | min | 25%     | 50%    | 75%     | max     |
|----------------------------|--------|-------------|-------------|-----|---------|--------|---------|---------|
| <b>Impressions</b>         | 8516.0 | 2273.855214 | 1742.647997 | 0.0 | 1359.75 | 1997.5 | 2780.25 | 39681.0 |
| <b>Organic Impressions</b> | 8516.0 | 2273.855214 | 1742.647997 | 0.0 | 1359.75 | 1997.5 | 2780.25 | 39681.0 |
| <b>Reach</b>               | 8516.0 | 1503.111790 | 1769.934182 | 0.0 | 2.00    | 1271.0 | 2240.0  | 32418.0 |
| <b>Organic Reach</b>       | 8516.0 | 1503.111790 | 1769.934182 | 0.0 | 2.00    | 1271.0 | 2240.0  | 32418.0 |

- **Count:** There are 8516 entries in all columns, indicating that there are no missing values in these columns.
- **Mean:** On average, a post gets around 2273.85 impressions, both organically and overall, and also reaches around 1503.11 users, both organically and overall.
- **Standard Deviation (std):** The standard deviation is 1742.64, which is quite high. This indicates that the number of impressions and reach per post varies widely from the mean.

- **Minimum (min):** The minimum number of impressions a post has received is 0. This could be due to posts that didn't reach the audience or failed to engage them.
- **25th Percentile (25%):** 25% of the posts received 1359.75 impressions or less.
- **Median (50%):** The median number of impressions per post is 1997.5. Since the median is less than the mean, it indicates that the distribution of impressions is skewed to the right, with a few posts getting a very high number of impressions. The median reach per post is 1271.0. Since the median is less than the mean, it indicates that the distribution of reach is skewed to the right, with a few posts reaching a very high number of users.
- **75th Percentile (75%):** 75% of the posts received 2780.25 impressions or less.
- **Maximum (max):** The maximum number of impressions a post has received is 39681. This indicates that some posts were highly successful in reaching a large audience.



To effectively interpret the data provided, we might want to consider the following questions:

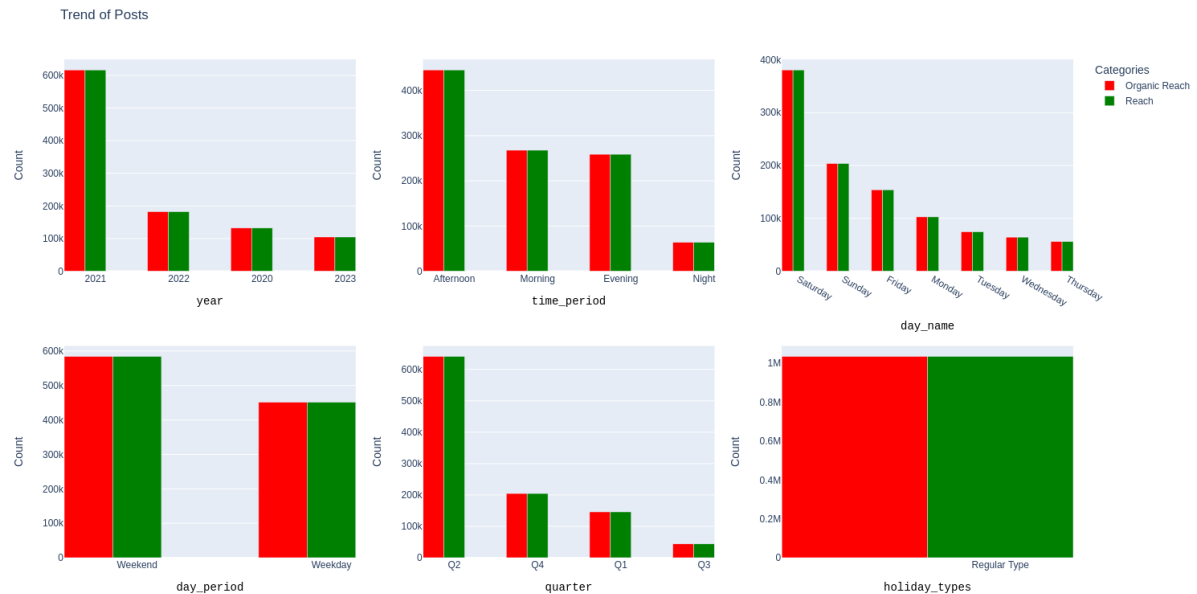
1. **How does the number of impressions and organic impressions vary by year?**  
The data shows a yearly breakdown of these metrics, which can help us understand trends over time.
2. **What are the peak times for impressions and organic impressions?** By looking at the breakdown by time period, we can identify when users are most active or engaged.

3. **How do impressions and organic impressions vary by day of the week?** This can reveal patterns in user behavior on different days.
4. **Is there a difference in engagement during weekdays versus weekends?** The 'day\_period' data can help us understand if user engagement changes during these periods.
5. **How do impressions and organic impressions vary by quarter?** This can provide insights into seasonal trends in user engagement.
6. **Is there a difference in engagement on holidays versus regular days?** The 'holiday\_types' data can reveal if user behavior changes during holidays.

The data shared provides insights into the distribution of the top 1% of posts based on impressions across different categories. Here's what we can infer from it:

- **Yearly Distribution:** The year 2021 had the highest number of impressions, followed by 2022 and 2023. This could indicate that the content strategy in these years was particularly effective in reaching a large audience.
- **Time of Day:** Posts made in the afternoon have the highest impressions, followed by morning and evening. This suggests that these are the optimal times for posting to reach a larger audience.
- **Day of the Week:** Saturday has the highest number of impressions, followed by Sunday and Friday. This indicates that weekends might be the best time to post for maximum reach.
- **Weekday vs Weekend:** Impressions are almost evenly split between weekdays and weekends, with weekends having slightly more impressions. This suggests that posting on both weekdays and weekends is effective for reaching a large audience.
- **Quarterly Distribution:** The second quarter (Q2) has the highest number of impressions, followed by Q4 and Q1. This could be due to seasonal trends or specific marketing campaigns during these periods.
- **Holiday Types:** Regular type posts have significantly higher impressions than posts made on holidays. This could indicate that regular posting is key to maintaining high engagement levels. These findings could be significant for understanding which factors contribute to high post engagement. They can help inform future content strategies to maximize reach and engagement.

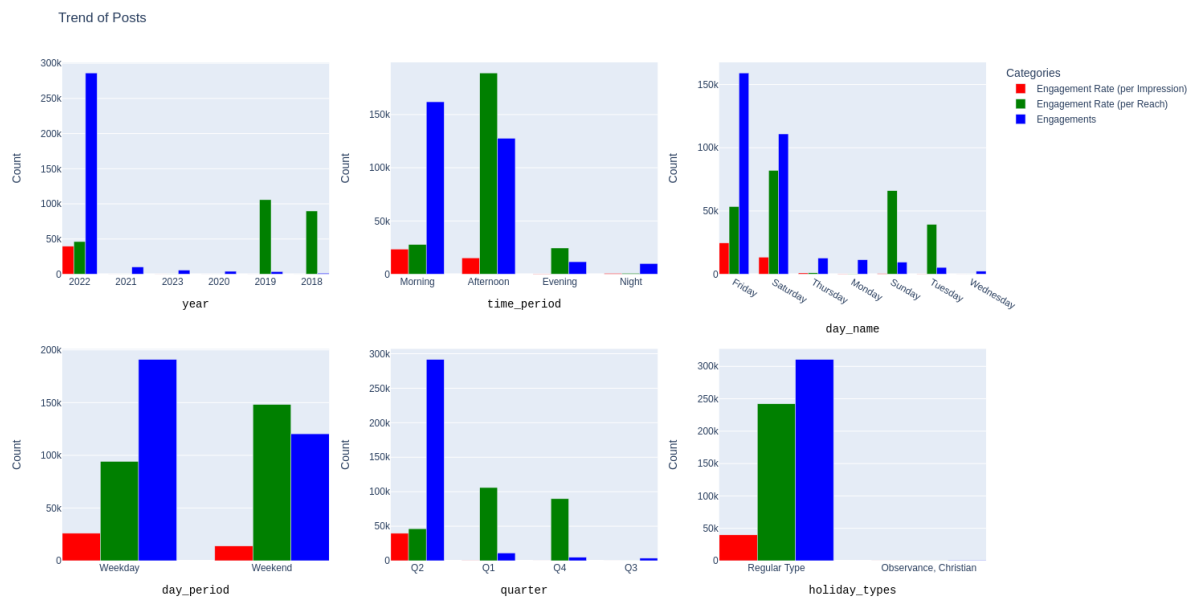




Comparing the top 1% of posts by Impressions and Reach, we can observe the following:

- Yearly Distribution:** The yearly distribution is similar for both metrics, with 2021 having the highest values. This suggests that the content strategy in 2021 was particularly effective in both reaching a large audience (Reach) and getting viewed multiple times (Impressions).
- Time of Day:** For both metrics, posts made in the afternoon have the highest values, followed by morning and evening. This suggests that these are the optimal times for posting to reach a larger audience and get more views.
- Day of the Week:** Saturday has the highest values for both metrics, followed by Sunday and Friday. This indicates that weekends might be the best time to post for maximum reach and impressions.
- Weekday vs Weekend:** The distribution between weekdays and weekends is similar for both metrics, with weekends having slightly more values. This suggests that posting on both weekdays and weekends is effective for reaching a large audience and getting more views.
- Quarterly Distribution:** The second quarter (Q2) has the highest values for both metrics, followed by Q4 and Q1. This could be due to seasonal trends or specific marketing campaigns during these periods.
- Holiday Types:** Regular type posts have significantly higher values for both metrics than posts made on holidays. This could indicate that regular posting is key to maintaining high engagement levels.

These observations are significant as they provide insights into when and how to post content for maximum reach (the number of unique users who see the post) and impressions (the total number of times the post is seen). These insights can help inform future content strategies to maximize engagement.



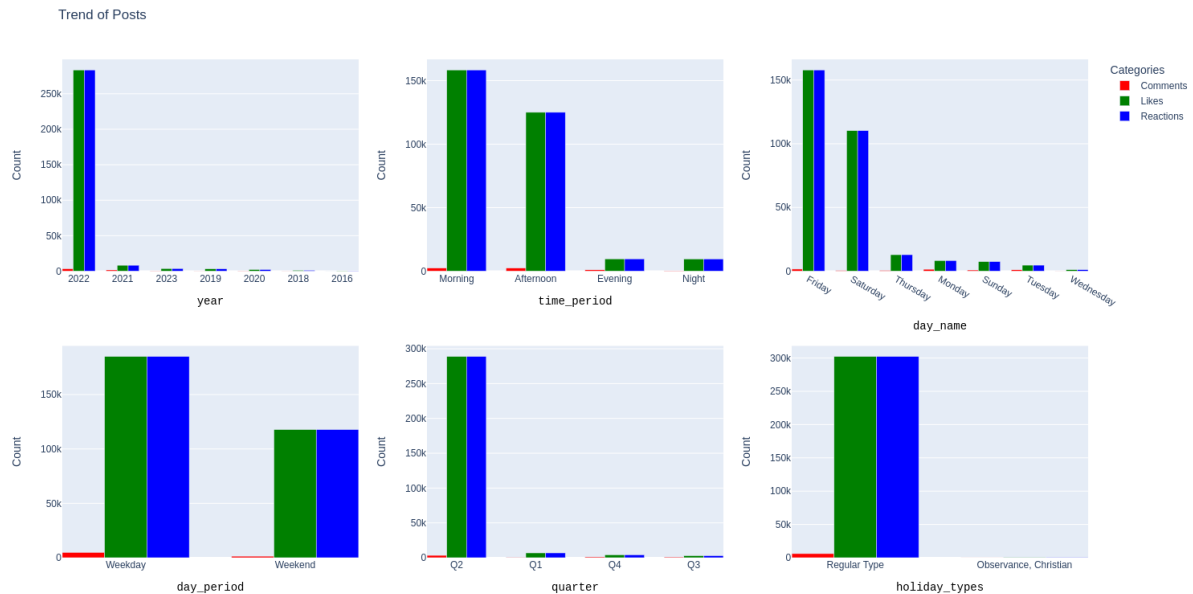
The data shared provides insights into the top 1% of posts based on engagements across different categories. Here's what we can infer from it:

- **Yearly Distribution:** The year 2022 had the highest number of engagements, followed by 2021 and 2023. This suggests that the content strategy in these years was particularly effective in engaging the audience.
- **Time of Day:** Posts made in the morning have the highest engagements, followed by afternoon and evening. This suggests that these are the optimal times for posting to engage a larger audience.
- **Day of the Week:** Friday has the highest number of engagements, followed by Saturday and Thursday. This indicates that these days might be the best to post for maximum engagement.
- **Weekday vs Weekend:** Engagements are almost evenly split between weekdays and weekends, with weekdays having slightly more engagements. This suggests that posting on both weekdays and weekends is effective for engaging a large audience.
- **Quarterly Distribution:** The second quarter (Q2) has the highest number of engagements, followed by Q1 and Q4. This could be due to seasonal trends or specific marketing campaigns during these periods.
- **Holiday Types:** Regular type posts have significantly higher engagements than posts made on holidays. This could indicate that regular posting is key to maintaining high engagement levels.

Comparing the insights from the top 1% of posts based on Reach and Engagements, we can observe the following:

- **Yearly Distribution:** The year with the highest reach and engagements is not the same. While 2021 had the highest reach, 2022 had the highest engagements. This suggests that while more unique users saw the posts in 2021, the posts in 2022 were more successful in engaging users (through likes, comments, shares, etc.).
- **Time of Day:** The time of day with the highest reach and engagements is also different. While afternoon posts had the highest reach, morning posts had the highest engagements. This suggests that while afternoon might be the best time to reach a larger audience, morning might be the best time to engage them.
- **Day of the Week:** Saturday has the highest reach, but Friday has the highest engagements. This indicates that while Saturday might be the best day to post for maximum reach, Friday might be the best day for maximum engagement.
- **Weekday vs Weekend:** The distribution between weekdays and weekends is similar for both metrics, suggesting that posting on both weekdays and weekends is effective for reaching a large audience and getting more engagements.
- **Quarterly Distribution:** The second quarter (Q2) has the highest values for both metrics, suggesting that this period might be particularly effective for both reaching and engaging users.
- **Holiday Types:** Regular type posts have significantly higher values for both metrics than posts made on holidays, indicating that regular posting is key to maintaining high reach and engagement levels.

These observations are significant as they provide insights into when and how to post content for maximum reach (the number of unique users who see the post) and engagement (the total number of likes, comments, shares, saves, etc.). However, it's important to note that what works best for reach may not necessarily work best for engagement, and vice versa. Therefore, it's crucial to strike a balance based on our specific goals and audience behavior.



When comparing the insights derived from the top 1% of engagements with those from the top 1% of reactions, we might want to consider the following questions:

1. **What are the commonalities and differences in the insights derived from both sets?** Identifying these can help us understand if certain factors consistently contribute to high engagement and reactions.
2. **Are there any contradictions between the two sets of insights?** For example, if a certain type of content is associated with high engagement but not with high reactions, it might indicate a discrepancy that needs further investigation.
3. **What is the significance of these observations?** Understanding this can help us make more informed decisions about our content strategy.
4. **Are there any gaps or inconsistencies in our logic or assumptions?** For instance, if we assumed that high engagement automatically leads to high reactions, but the data shows otherwise, we might need to revisit our assumptions.

Comparing the insights from the top 1% of posts based on Engagements and Reactions, we can observe the following commonalities, differences, and contradictions:

### Commonalities:

- **Yearly Distribution:** Both metrics show that 2022 had a high level of user interaction, suggesting effective content strategy in that year.

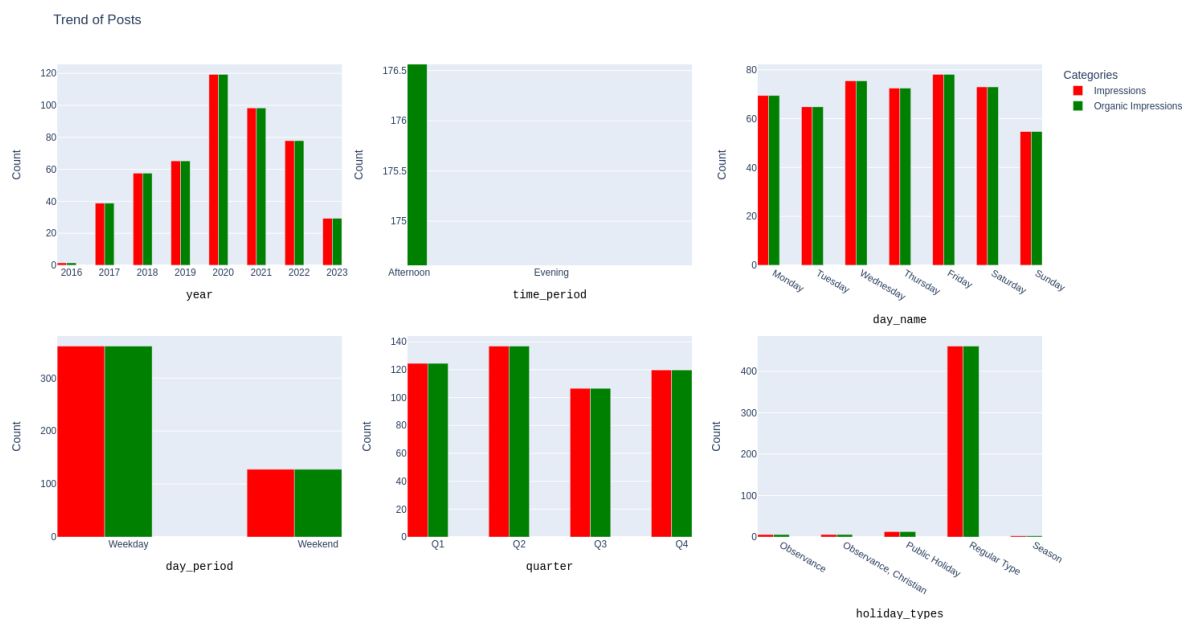
- **Time of Day:** Morning and afternoon are the optimal times for posting to engage a larger audience for both metrics.
- **Day of the Week:** Friday and Saturday are the best days to post for maximum engagement and reactions.
- **Weekday vs Weekend:** Posting on both weekdays and weekends is effective for engaging a large audience for both metrics.
- **Quarterly Distribution:** The second quarter (Q2) has high values for both metrics, possibly due to seasonal trends or specific marketing campaigns.
- **Holiday Types:** Regular type posts have significantly higher values for both metrics than posts made on holidays, indicating that regular posting is key to maintaining high engagement levels.

### Differences:

- While the year with the highest engagements was 2022, the year with the highest reactions was also 2022 but followed closely by 2021. This could suggest that while more users interacted with the posts (likes, comments, shares, etc.) in 2022, more users reacted (likes and comments) to the posts in 2021.

### Contradictions:

- There doesn't appear to be any contradictions between the two sets of insights. The patterns observed are consistent across both metrics.



The data shared provides insights into the distribution of Impressions and Organic Impressions across different categories. Here's what we can infer from it:

- **Yearly Distribution:** The year 2020 had the highest number of impressions, followed by 2021 and 2022. This suggests that the content strategy in these years was particularly effective in reaching a large audience.
- **Time of Day:** Posts made in the afternoon have the highest impressions, followed by morning and evening. This suggests that these are the optimal times for posting to reach a larger audience.
- **Day of the Week:** Friday has the highest number of impressions, followed by Wednesday and Saturday. This indicates that these days might be the best to post for maximum reach.
- **Weekday vs Weekend:** Impressions are significantly higher on weekdays compared to weekends. This suggests that posting on weekdays is more effective for reaching a large audience.
- **Quarterly Distribution:** The second quarter (Q2) has the highest number of impressions, followed by Q4 and Q1. This could be due to seasonal trends or specific marketing campaigns during these periods.
- **Holiday Types:** Regular type posts have significantly higher impressions than posts made on holidays. This could indicate that regular posting is key to maintaining high reach levels.

Comparing the insights from the top 1% of posts based on Impressions and the overall Impressions across different categories, we can observe the following commonalities, differences, and contradictions:

#### **Commonalities:**

- **Yearly Distribution:** Both metrics show that 2021 and 2022 had a high level of user interaction, suggesting effective content strategy in those years.
- **Time of Day:** Afternoon and morning are the optimal times for posting to reach a larger audience for both metrics.
- **Day of the Week:** Friday and Saturday are the best days to post for maximum reach for both metrics.
- **Weekday vs Weekend:** Posting on both weekdays and weekends is effective for reaching a large audience for both metrics.
- **Quarterly Distribution:** The second quarter (Q2) has high values for both metrics, suggesting that this period might be particularly effective for both reaching and engaging users.
- **Holiday Types:** Regular type posts have significantly higher values for both metrics than posts made on holidays, indicating that regular posting is key to maintaining high reach levels.

#### **Differences:**

- While the year with the highest impressions was 2020 for the overall dataset, the year with the highest impressions for the top 1% was 2021. This could

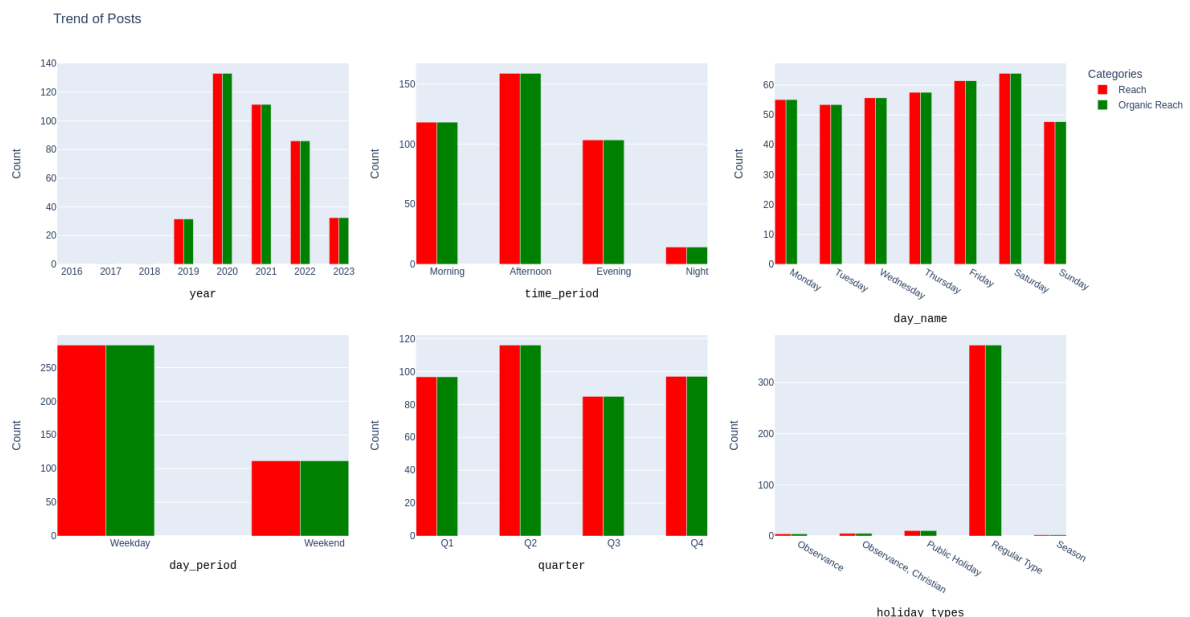
suggest that while more users saw the posts in 2020, a smaller number of highly engaging posts in 2021 were viewed multiple times.

- While afternoon posts had the highest impressions for the overall dataset, morning posts had the highest impressions for the top 1%. This could indicate that while afternoon might be the best time to reach a larger audience, morning might be when highly engaging content is most likely to be viewed multiple times.

### Contradictions:

- There don't appear to be any contradictions between the two sets of insights. The patterns observed are consistent across both metrics.

These insights can also challenge or confirm underlying assumptions about user behavior and content performance. For example, if you assumed that posting in the afternoon always results in higher impressions, these insights might lead you to reconsider and experiment with posting in the morning. Similarly, if you assumed that weekends always result in lower impressions, these insights could confirm that assumption.



Comparing the insights from the top 1% of posts based on Reach and the overall Reach across different categories, we can observe the following commonalities, differences, and contradictions:

### Commonalities:

- **Yearly Distribution:** Both metrics show that 2021 and 2022 had a high level of user interaction, suggesting effective content strategy in those years.

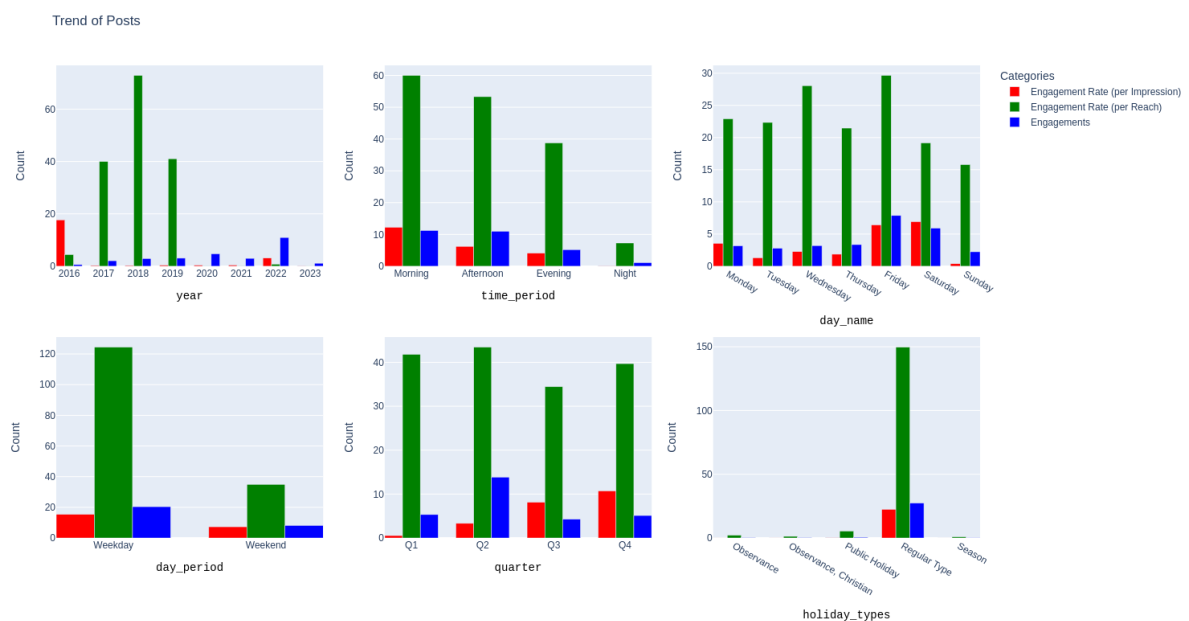
- **Time of Day:** Afternoon and morning are the optimal times for posting to reach a larger audience for both metrics.
- **Day of the Week:** Friday and Saturday are the best days to post for maximum reach for both metrics.
- **Weekday vs Weekend:** Posting on both weekdays and weekends is effective for reaching a large audience for both metrics.
- **Quarterly Distribution:** The second quarter (Q2) has high values for both metrics, suggesting that this period might be particularly effective for both reaching and engaging users.
- **Holiday Types:** Regular type posts have significantly higher values for both metrics than posts made on holidays, indicating that regular posting is key to maintaining high reach levels.

### Differences:

- While the year with the highest reach was 2020 for the overall dataset, the year with the highest reach for the top 1% was 2021. This could suggest that while more users saw the posts in 2020, a smaller number of highly engaging posts in 2021 reached a larger audience.
- While afternoon posts had the highest reach for the overall dataset, morning posts had the highest reach for the top 1%. This could indicate that while afternoon might be the best time to reach a larger audience, morning might be when highly engaging content is most likely to be seen by more unique users.

### Contradictions:

- There don't appear to be any contradictions between the two sets of insights. The patterns observed are consistent across both metrics.

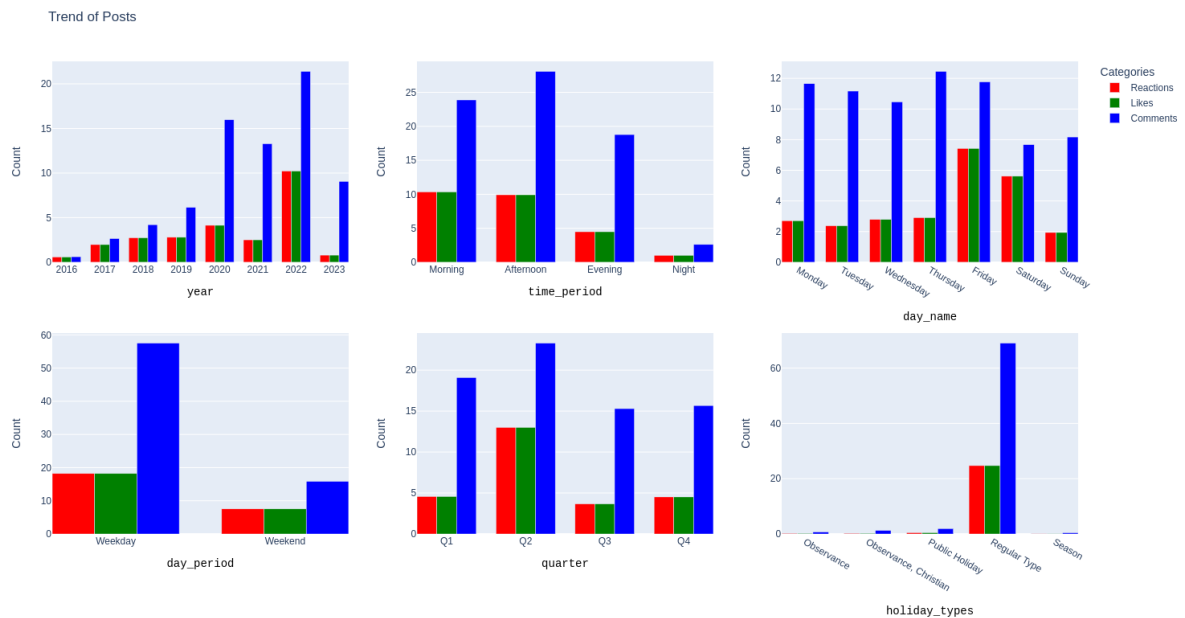




Comparing the insights from the top 1% of posts based on Engagements and overall Engagements across different categories, we can observe several commonalities, differences, and contradictions:

### Commonalities:

- Both metrics show that morning and afternoon are optimal times for posting to engage a larger audience.
- Both metrics suggest that posting on both weekdays and weekends is effective for engaging a large audience.
- Regular type posts have significantly higher values than posts made on holidays for both metrics, indicating that regular posting is key to maintaining high engagement levels.





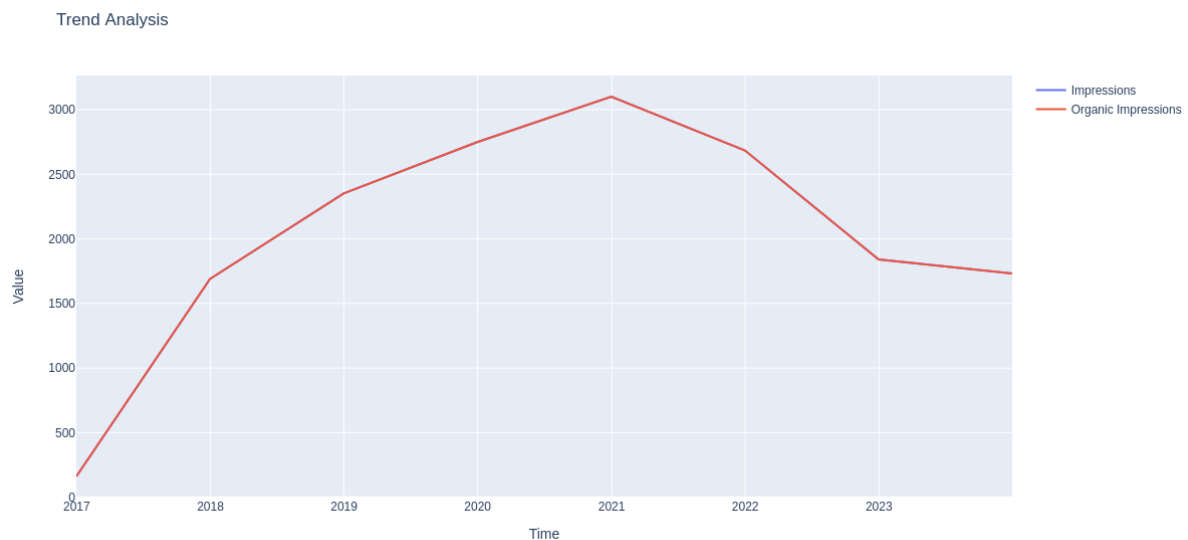
The data shared shows the correlation between different metrics. Here are some key findings:

1. **Impressions and Organic Impressions:** These two metrics have a perfect correlation of 1. This means that every time there's an increase in impressions, there's an equivalent increase in organic impressions. This could indicate that the majority of impressions are organic, not resulting from paid promotions.
2. **Impressions and Reach:** There's a strong positive correlation of 0.821343 between these two metrics. This suggests that as the number of times content is viewed increases (impressions), the number of unique viewers (reach) also increases.
3. **Engagements and Likes/Reactions:** These metrics have a very high positive correlation of approximately 0.999460. This indicates that posts with more likes or reactions tend to have higher overall engagement.
4. **Engagements and Engagement Rate (per Impression):** The correlation here is positive but relatively low at 0.290178. This suggests that while higher engagement rates can lead to more total engagements, other factors may also play a significant role.

5. **Engagements and Comments:** The correlation is positive but quite low at 0.055743. This implies that comments may not significantly contribute to overall engagement.

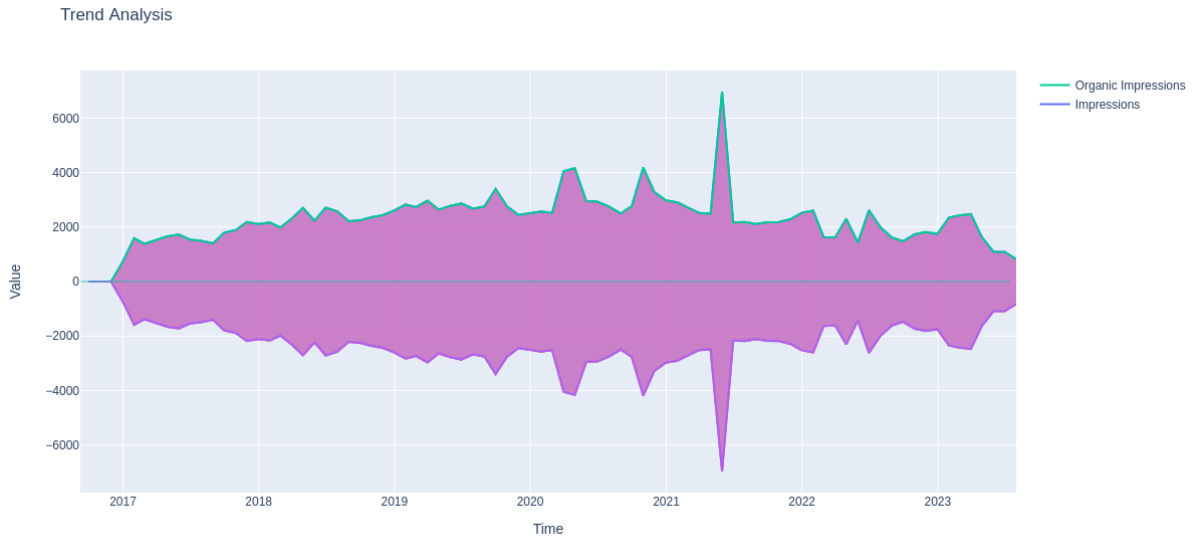
**How does post performance vary by time of day, day of the week, or season?** This can help identify the best times to post for maximum engagement.

We try out Time Series Analysis to uncover some Patterns



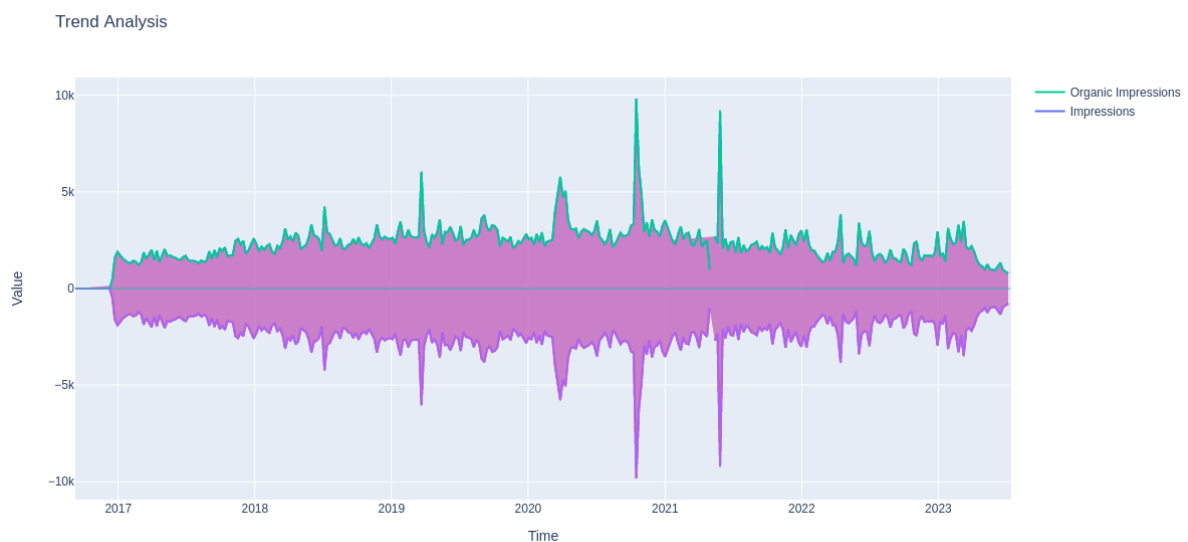
### Yearly Trends

There has been a steady increase in impressions from 2017 to 2021, indicating that the content strategy employed during this period was effective. However, there was a noticeable drop in impressions from 2021 to 2023. This drop coincides with a decrease in the number of posts, suggesting that the quantity of content could be a significant factor in driving impressions.



## Monthly Trends

A month-by-month analysis shows a gradual increase in impressions over the years. Notably, there were spikes in impressions in April 2020, October 2020, and May 2021, with May 2021 recording the highest average impressions at 6971. However, after May 2021, there was a consistent decline in impressions, reaching a low of 1090 in June 2023.

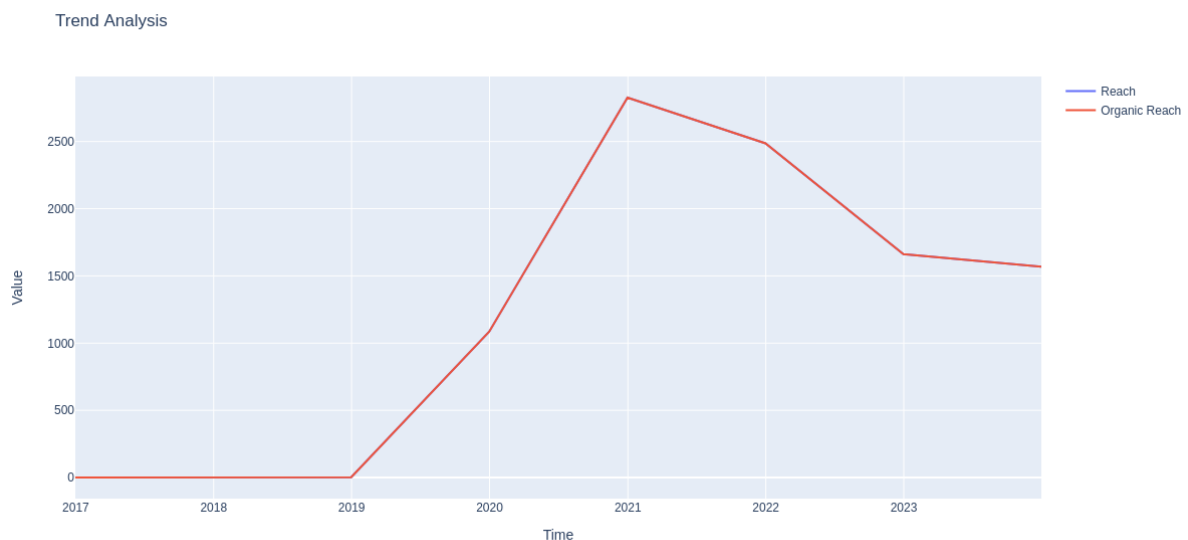


## Weekly Trends

The weekly analysis also shows a slow but steady increase in impressions over time. There were significant spikes in impressions during certain weeks - July 2018, March 2019, March 2020, October 2020, and May 2021. After May 2021, the impressions

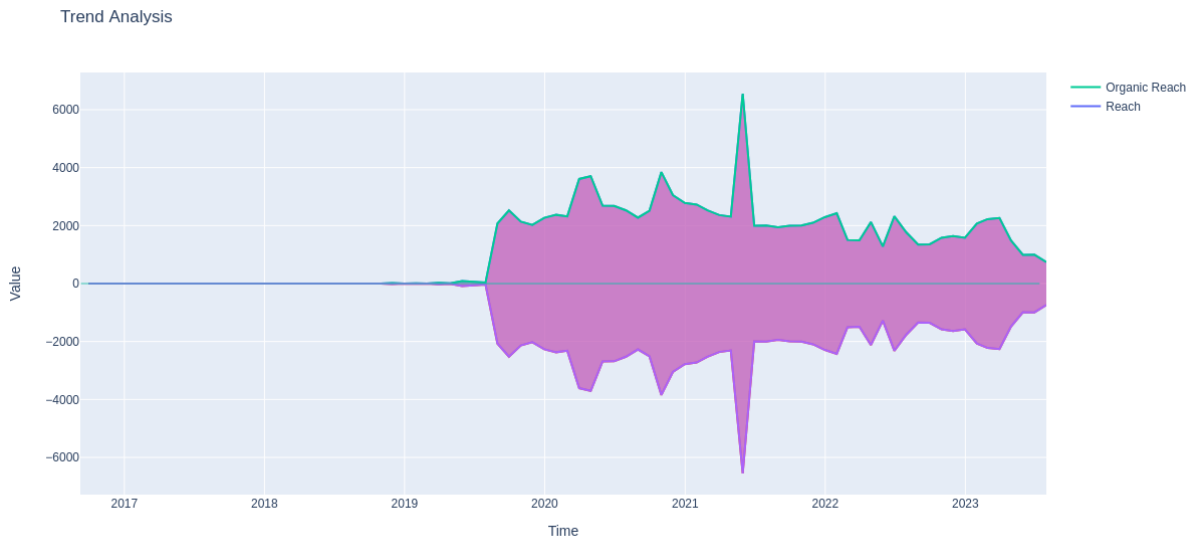
started to decline, with the highest value being 3855 impressions in April 2022 and ending with just 1332 Impressions in June 2023.

These findings are significant as they highlight the importance of both the quantity and timing of posts in driving user engagement. They suggest that to maximize impressions, it's crucial to maintain a consistent posting schedule and to consider the timing of posts carefully.



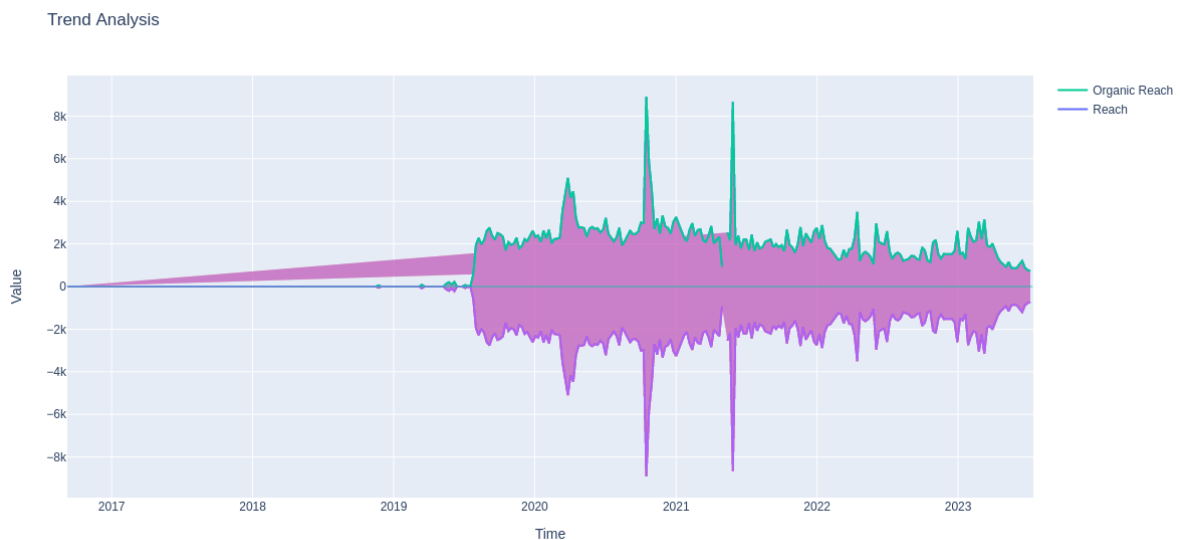
## Yearly Trends

From 2017 to 2019, the reach of the posts was virtually zero. However, there was a significant increase in reach from 2019 to 2021. This suggests that the content strategy employed during this period was effective in reaching a larger audience. Despite a drop in reach from 2021 to 2023, the reach remained higher than the initial period from 2017 to 2019.



## Monthly Trends

A month-by-month analysis shows a similar pattern. The reach was minimal until July 2019, after which there was a steady increase, peaking at 6554 in May 2021. However, like the yearly trend, there was a decline from 2021 to 2023, ending with a reach of 734 users in June 2023.

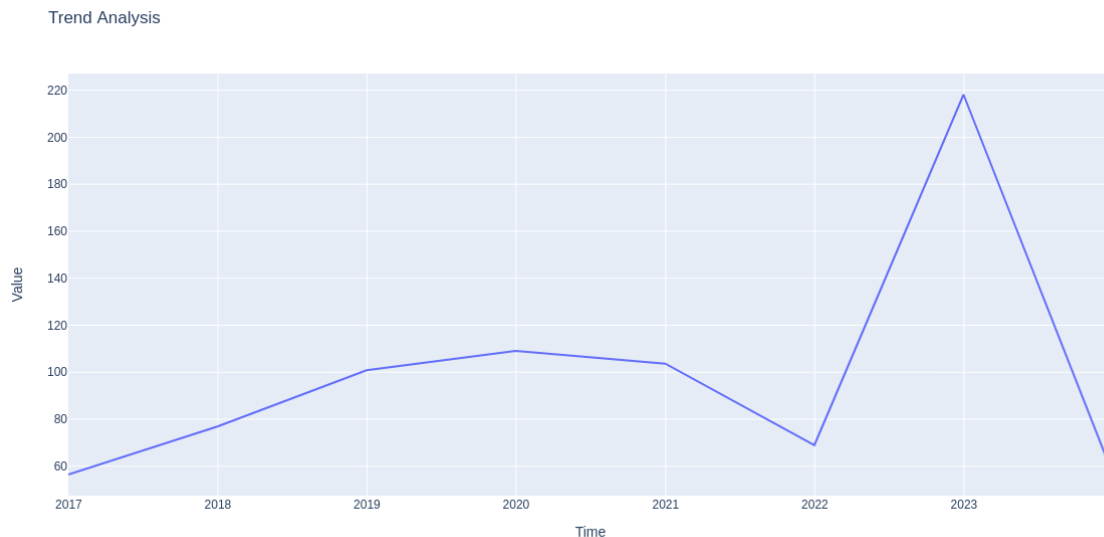


## Weekly Trends

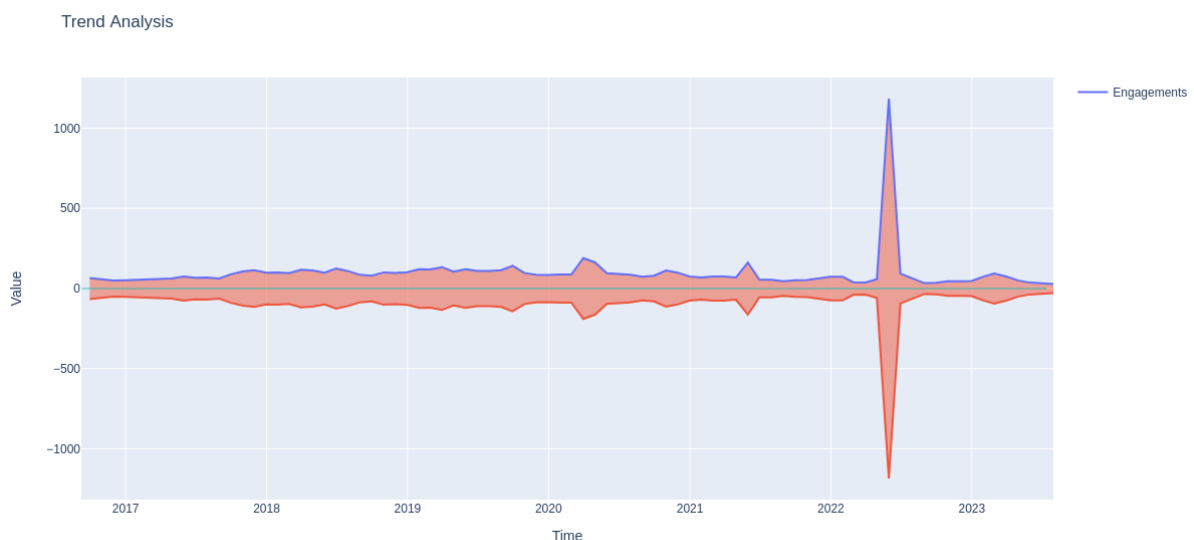
The weekly analysis also shows an increase in reach starting from July 2019, with peaks in March 2020, October 2020, and March 2021 (8993). After March 2021, the reach started to decline, with the highest value being 3500 in April 2022 and ending with just 1200 in June 2023.

## Correlation between Reach and Impressions

There is an 82% positive correlation between post reach (and organic reach) and impressions (and organic impressions). This indicates that as impressions increase, reach also increases. This correlation is consistent from 2019 onwards, suggesting that reaching more users leads to more impressions.

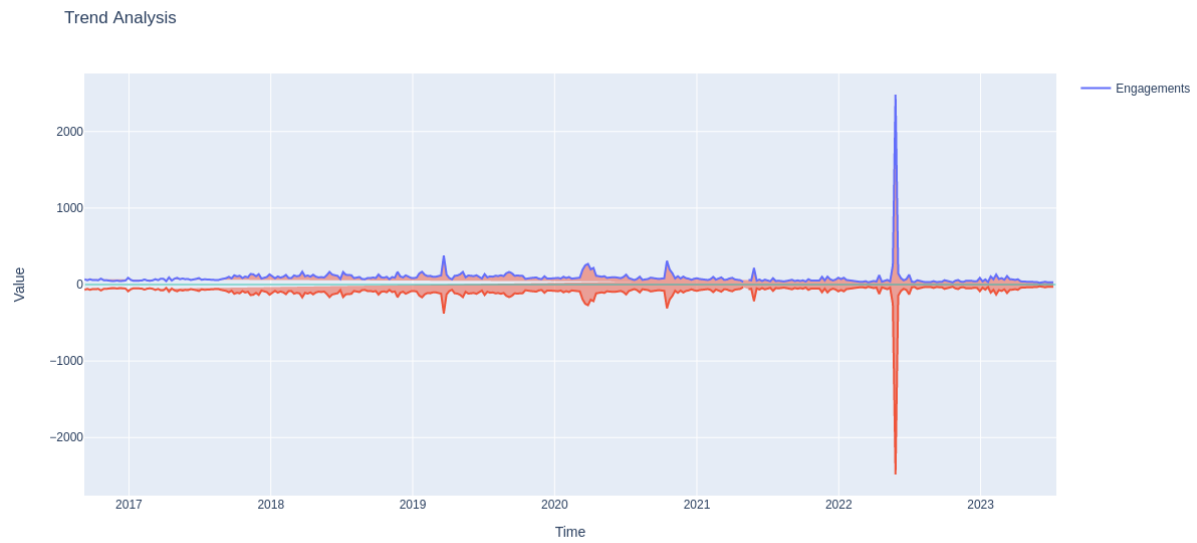


Analyzing our Engagements over the years and we see a curve, from 2017 to 2022, peaking at 2020/2021, suddenly dropping in 2022 and then rising to it's highest peak in 2023, and dropping from the beginning of the year. We had less Impressions / Reach in the year 2022, 2023, but yet we had the highest engagements in the year 2023, maybe there was a change in content

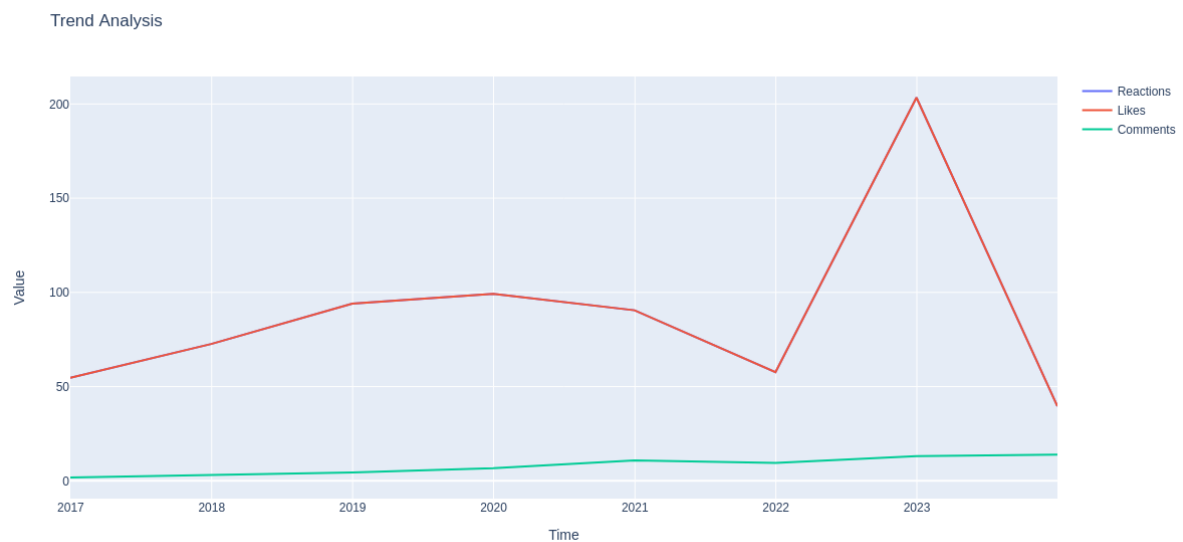


Analyzing our Engagements over the Months it shows the small increase in Engagements from the year 2018 till the month Feb 2022, after that we see a sudden increase in Engagements in May 2021, but then we see it declines

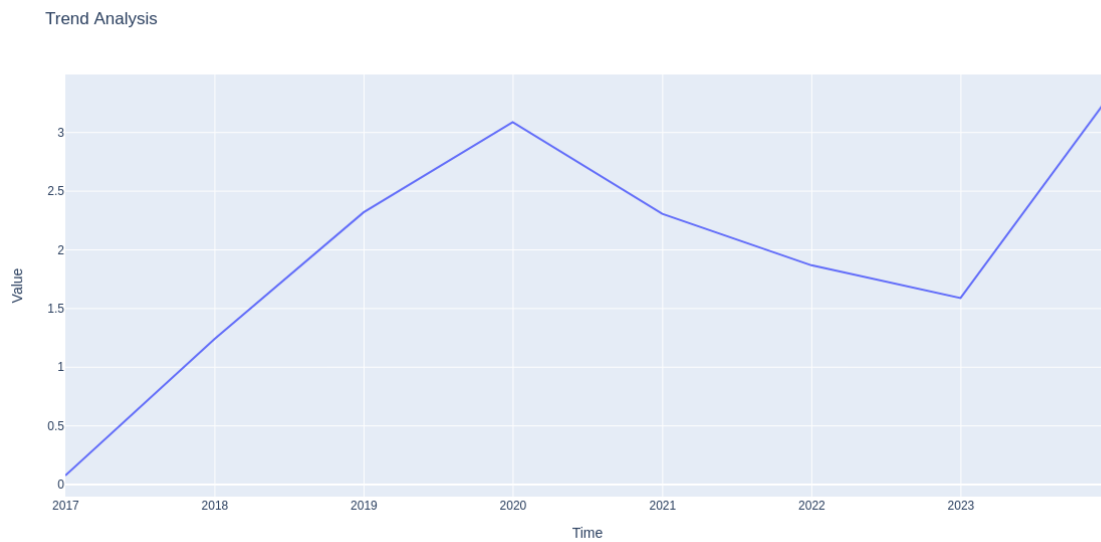
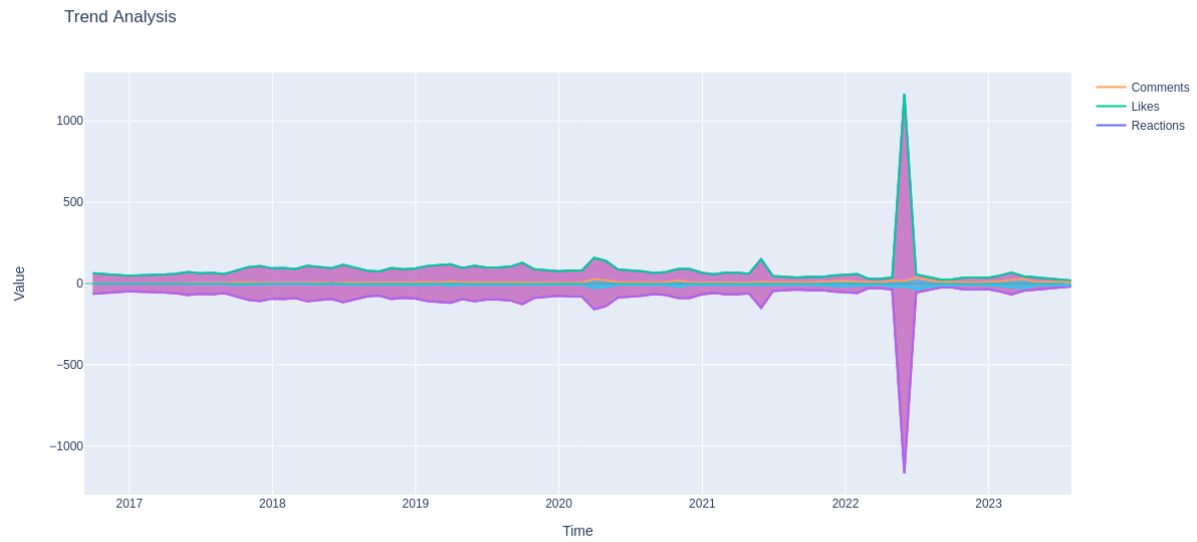
Similar to the Trends we notice across the Months, we observe the same thing across the Weeks.



We observe the same trends observed in Engagement Columns in the Reaction, Likes and Comments Columns, this is proof of the strong positive correlation

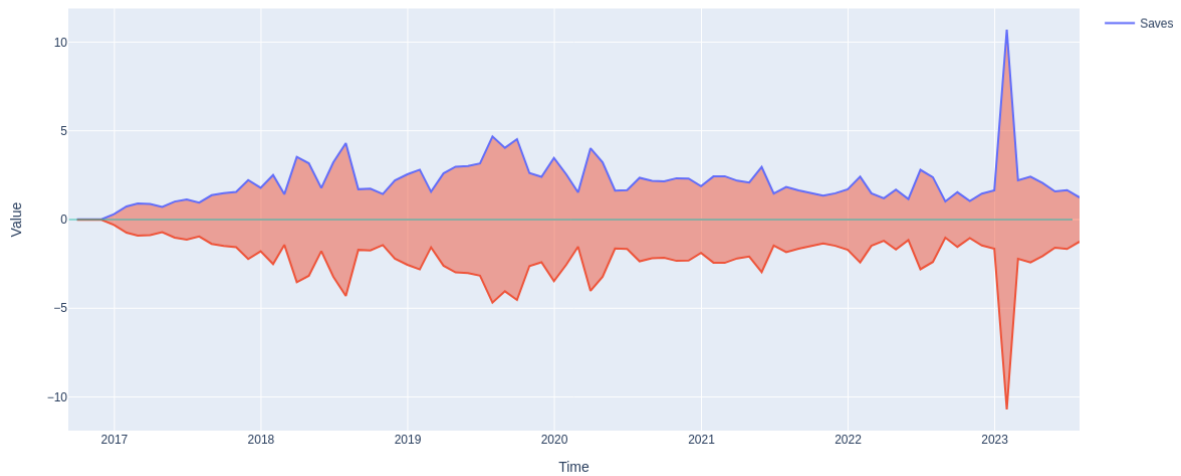






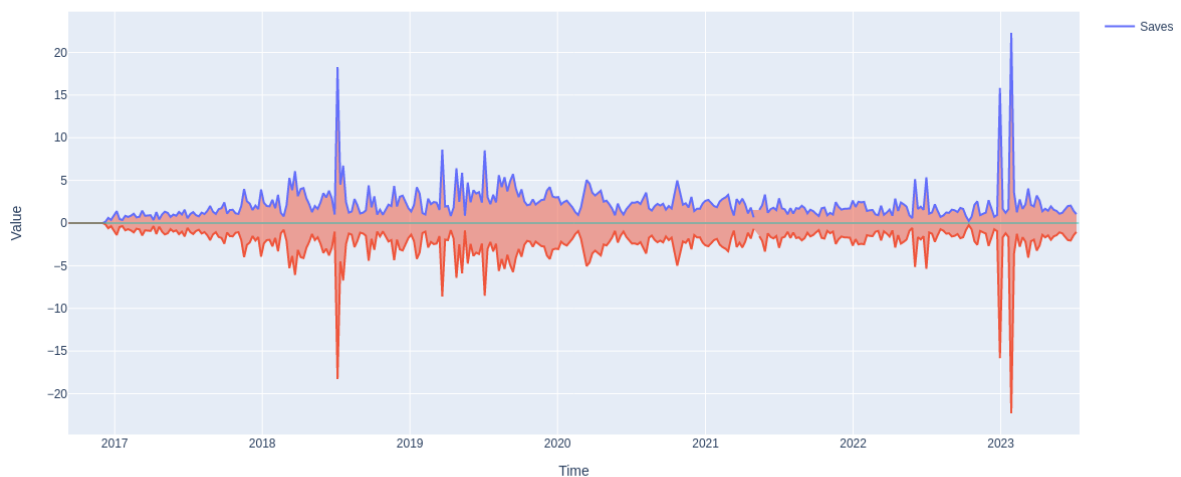
Analyzing the trend of Instagram Saves on our Posts, we see it increases from 2017-2020, but reduces from 2020-2023 and then suddenly Increases, over the years.

Trend Analysis



Across the Months we see an increase from 2017 - July 2020, and then drops till 2023, where there is a sudden spike in Jan 2023

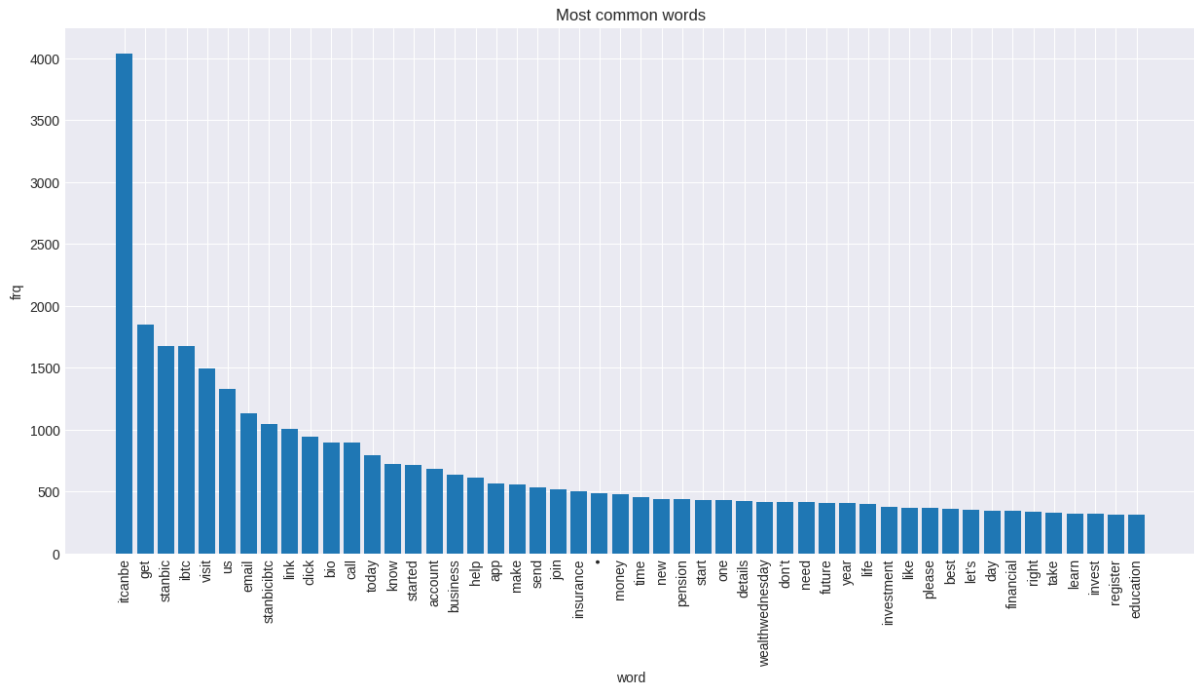
Trend Analysis



The methodology for analyzing the social media posts involves several steps:

1. **Data Cleaning:** The first step in the process is data cleaning. This involves removing URLs, converting text to lowercase, removing punctuation, and removing stop words. Stop words are common words that do not contain important meaning and are usually removed from texts. An additional list of stop words was also used to remove more common words.
2. **Text Preprocessing:** The cleaned texts are then preprocessed using several techniques:
  - Removing text in square brackets.
  - Removing links.

3. **Applying the Cleaning Functions:** The cleaning functions are then applied to the 'Post' column of the data. The cleaned posts are stored in a new dataframe.

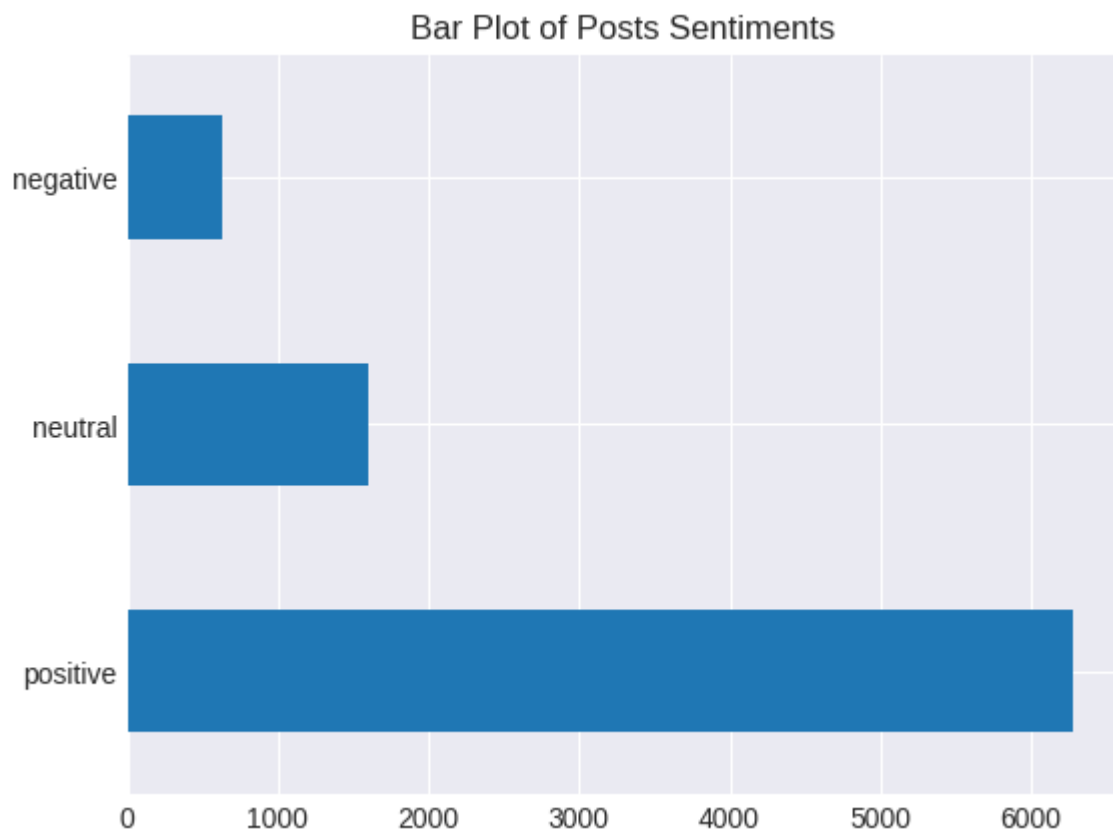


We can see the top most common words in our posts

### Word Cloud for Top 100 words from Instagram Posts



## Here's a Sentiment Analysis of our Posts



Using LDA to get the main topics in our dataset

Topic 0: stanbicibt | itcanbe | future | plan | time | school | child | set | weekend | money | loved | save | ones | goals | year

Topic 1: itcanbe | link | visit | click | bio | business | join | pension | stanbic | ibtc | register | live | know | watch | health

Topic 2: stanbic | ibtc | nigeria | growth | itcanbe | bank | africa | partner | tips | business | financial | lagos | capital | nigerian | giving

Topic 3: stanbicibt | day | win | hiflstanbicibt | itcanbe | repost | steps | winners | forward | stay | code | follow | service | team | investments

Topic 4: insurance | itcanbe | life | year | education | movingforward | today | best | happy | trust | better | wealthwednesday | neverstop | like | walk

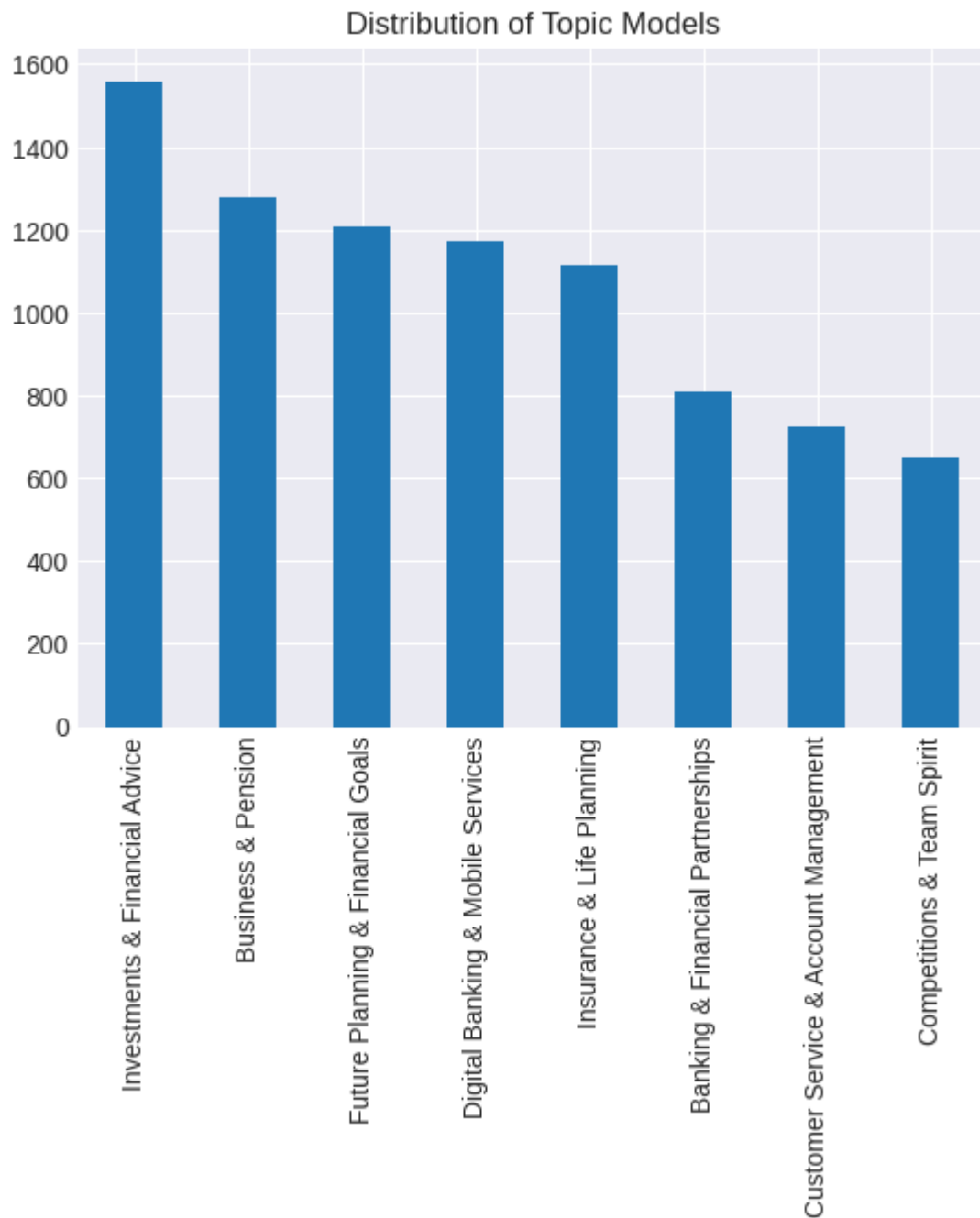
Topic 5: email | details | information | account | itcanbe | customercarenigeriastanbicibtccom | send | got | think | stay | happentolife | credit | ve | safe | pensionsolutionstanbicibtccom

Topic 6: itcanbe | email | let | started | know | visit | help | invest | investment | send | funds | money | today | start | financial

Topic 7: itcanbe | stanbic | ibtc | app | account | download | visit | week | new | motivationmonday | mobile | open | super | enjoy | easy

Based on the topics provided, here are 8 unique categories that could be created:

1. **Future Planning & Financial Goals:** This category could include posts about planning for the future, setting goals, saving money, and preparing for things like school or retirement (Topic 0).
2. **Business & Pension:** This category could cover posts related to business, pension plans, and financial growth (Topic 1).
3. **Banking & Financial Partnerships:** This category could include posts about banking services, financial partnerships, and economic growth in Africa (Topic 2).
4. **Competitions & Team Spirit:** This category could cover posts about competitions, winning teams, and team spirit (Topic 3).
5. **Insurance & Life Planning:** This category could include posts about insurance plans, life planning, and wealth management (Topic 4).
6. **Customer Service & Account Management:** This category could cover posts about account management, customer service, and safety measures (Topic 5).
7. **Investments & Financial Advice:** This category could include posts offering financial advice, investment tips, and ways to grow your money (Topic 6).
8. **Digital Banking & Mobile Services:** This category could cover posts about digital banking services, mobile banking apps, and the convenience of banking on the go (Topic 7).



Posts about Investments and Financial Advice have the highest count, after which comes Business and Pension, lastly Competitions and Team Spirit is the lowest.

Parts of Speech (POS) Tagging and Named Entity Recognition (NER) can be quite useful in analyzing Instagram metrics and increasing the overall outreach of a post. Here's how:

Content Analysis: POS tagging can help analyze the content of Instagram posts or comments. For example, the use of more verbs might indicate a call to action, while the use of adjectives might suggest descriptive content.

Sentiment Analysis: POS tagging can be used in sentiment analysis, which can provide insights into the audience's perceptions and attitudes towards certain posts. This can help in tailoring content to evoke desired emotional responses<sup>1</sup>.

Trend Identification: NER can identify named entities such as people, organizations, or locations in the text. This can help identify trends or popular topics among users, which can be leveraged to create more engaging content<sup>1</sup>.

Personalization: NER can help in personalizing content. For example, if a lot of users are talking about a specific event or person, that information can be used to create targeted posts.

Improving Searchability: Using NER and POS tagging, posts can be annotated with metadata that improves their searchability. This can increase the visibility of posts and thereby increase impressions.

Competitor Analysis: These techniques can also be used to analyze the content strategies of competitors. This information can provide valuable insights and help in creating more effective content strategies.

|   | <b>VERB_<br/>count</b> | <b>ADV_<br/>count</b> | <b>ADJ_<br/>count</b> | <b>NUM_c<br/>ount</b> | <b>NOUN_<br/>count</b> | <b>SPACE_<br/>count</b> | <b>PROPN_<br/>count</b> | <b>pos_c<br/>ount</b> |
|---|------------------------|-----------------------|-----------------------|-----------------------|------------------------|-------------------------|-------------------------|-----------------------|
| <b>topic_la<br/>bel</b>   |                        |                       |                       |                       |                        |                         |                         |                       |
| <b>Banking<br/>&amp;<br/>Financi<br/>al<br/>Partner<br/>ships</b> | 3.28465<br>3           | 0.3972<br>77          | 2.2970<br>30          | 0.0891<br>09          | 7.44059<br>4           | 0.76608<br>9            | 6.094059                | 21.34<br>2822         |
| <b>Busines<br/>s &amp;<br/>Pension</b>                            | 4.15156<br>2           | 0.3335<br>94          | 1.8703<br>13          | 0.0898<br>44          | 7.45937<br>5           | 0.85156<br>2            | 3.172656                | 19.38<br>9063         |
| <b>Future<br/>Plannin<br/>g &amp;<br/>Financi<br/>al Goals</b>    | 4.05132<br>5           | 0.6730<br>13          | 2.1548<br>01          | 0.0910<br>60          | 6.68294<br>7           | 0.65562<br>9            | 2.664735                | 18.90<br>2318         |
| <b>Investm<br/>ents &amp;<br/>Financi<br/>al<br/>Advice</b>       | 4.45705<br>1           | 0.5000<br>00          | 1.8525<br>64          | 0.0826<br>92          | 6.76089<br>7           | 0.61153<br>8            | 1.925000                | 18.02<br>7564         |

|  |          |          |          |          |          |          |          |           |
|--|----------|----------|----------|----------|----------|----------|----------|-----------|
| <b>Digital Banking &amp; Mobile Services</b>     | 3.291809 | 0.316553 | 2.178328 | 0.109215 | 6.674061 | 0.977816 | 2.326792 | 17.077645 |
| <b>Insurance &amp; Life Planning</b>             | 3.780969 | 0.524237 | 1.792639 | 0.087971 | 6.188510 | 0.657989 | 2.239677 | 17.054758 |
| <b>Competitions &amp; Team Spirit</b>            | 2.989198 | 0.439815 | 1.643519 | 0.070988 | 6.007716 | 1.054012 | 3.432099 | 16.865741 |
| <b>Customer Service &amp; Account Management</b> | 3.264463 | 0.483471 | 1.392562 | 0.114325 | 6.249311 | 0.774105 | 1.957300 | 15.691460 |

The data shows the average count of different parts of speech (POS) for each topic label, sorted by the total POS count. Here are some key findings:

1. **Banking & Financial Partnerships:** This topic has the highest average total POS count, indicating that it has the most complex sentence structures. It also has the highest counts for verbs, nouns, and proper nouns, suggesting a focus on actions and specific entities or concepts.
2. **Business & Pension:** This topic has the second-highest total POS count. It has the highest average count of verbs, indicating a strong focus on actions or processes.
3. **Future Planning & Financial Goals:** This topic has a high count of verbs and adverbs, suggesting a focus on actions and the manner in which these actions are carried out.
4. **Investments & Financial Advice:** This topic has the highest average count of verbs, indicating a strong focus on actions or processes.
5. **Digital Banking & Mobile Services:** This topic has a high count of spaces, suggesting that posts under this topic may contain more sentences or larger blocks of text.



6. **Insurance & Life Planning:** This topic has a relatively high count of adjectives, suggesting a focus on descriptive language.
7. **Competitions & Team Spirit:** This topic has the highest count of spaces and proper nouns, suggesting that posts under this topic may contain more sentences or larger blocks of text and focus on specific entities or concepts.
8. **Customer Service & Account Management:** This topic has the lowest total POS count, indicating that it may have simpler sentence structures compared to other topics.

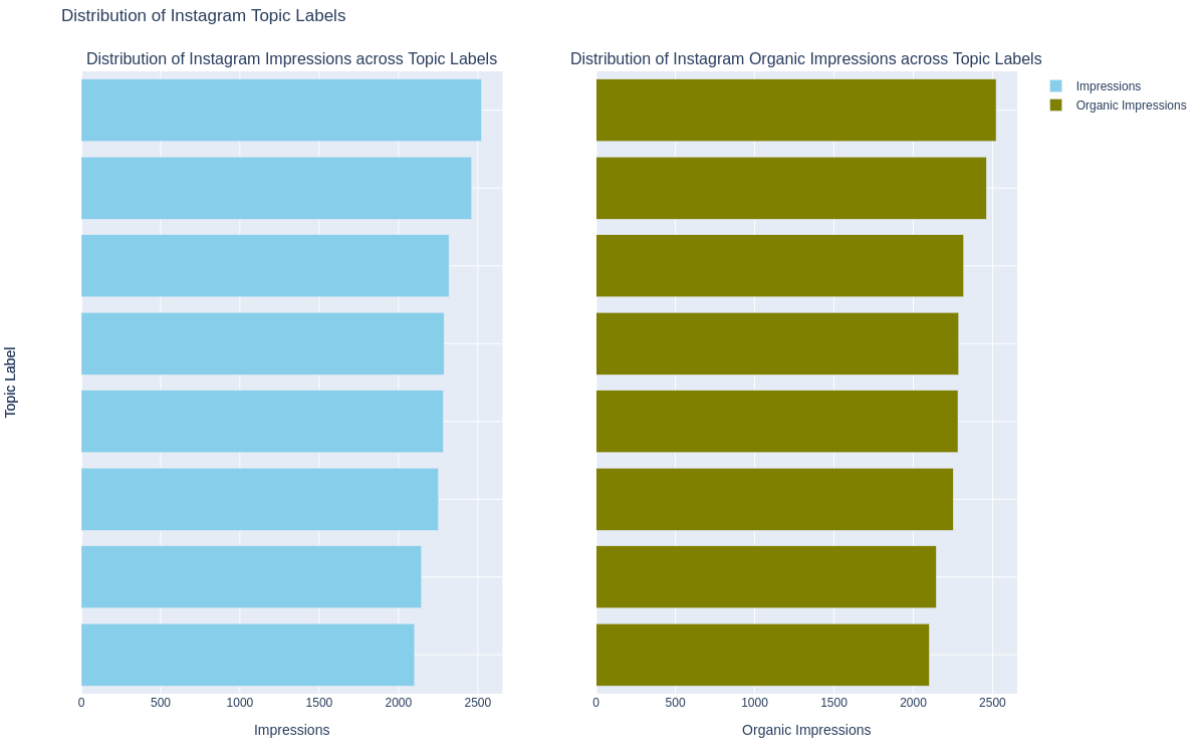
| CARDINAL_count                    | DATE_count | GPE_count | LOC_count | MONEY_count | ORDINAL_count | ORG_count | PERSON_count | TIME_count | ner_count |          |
|-----------------------------------|------------|-----------|-----------|-------------|---------------|-----------|--------------|------------|-----------|----------|
| topic_label                       |            |           |           |             |               |           |              |            |           |          |
| Banking & Financial Partnerships  | 0.061881   | 0.324257  | 0.396040  | 0.032178    | 0.006188      | 0.035891  | 0.763614     | 0.551980   | 0.024752  | 2.373762 |
| Competitions & Team Spirit        | 0.078704   | 0.283951  | 0.126543  | 0.004630    | 0.003086      | 0.055556  | 0.516975     | 0.476852   | 0.026235  | 1.635802 |
| Business & Pension                | 0.057813   | 0.396094  | 0.207813  | 0.005469    | 0.007031      | 0.020313  | 0.601562     | 0.228906   | 0.032031  | 1.621875 |
| Future Planning & Financial Goals | 0.064570   | 0.461921  | 0.107616  | 0.000828    | 0.001656      | 0.028974  | 0.432947     | 0.464404   | 0.022351  | 1.620033 |
| Investments & Financial Advice    | 0.055128   | 0.309615  | 0.192949  | 0.001282    | 0.009615      | 0.019231  | 0.594872     | 0.192949   | 0.005769  | 1.414744 |

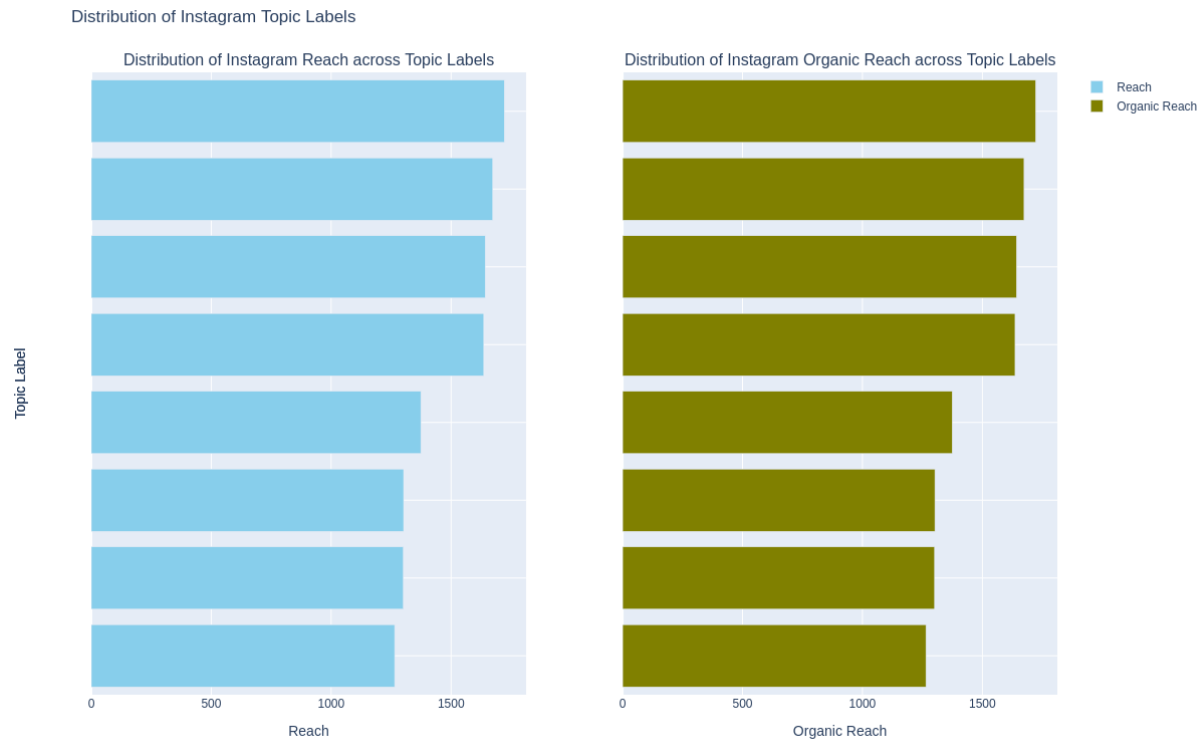
|  |          |          |          |          |          |          |          |          |          |          |
|--|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| <b>Digital Banking &amp; Mobile Services</b>     | 0.069113 | 0.373720 | 0.151877 | 0.000000 | 0.021331 | 0.020478 | 0.474403 | 0.160410 | 0.019625 | 1.329352 |
| <b>Insurance &amp; Life Planning</b>             | 0.057451 | 0.393178 | 0.157989 | 0.005386 | 0.000000 | 0.038600 | 0.359066 | 0.250449 | 0.012567 | 1.315978 |
| <b>Customer Service &amp; Account Management</b> | 0.110193 | 0.136364 | 0.082645 | 0.001377 | 0.000000 | 0.026171 | 0.469697 | 0.391185 | 0.004132 | 1.253444 |

The data shared shows the average count of different named entity recognition (NER) categories for each topic label, sorted by the total NER count. Here are some key findings:

1. **Banking & Financial Partnerships:** This topic has the highest average total NER count, indicating that it has the most named entities. It also has the highest counts for geographical entities (GPE), organizations (ORG), and persons, suggesting a focus on specific locations, organizations, and individuals.
2. **Competitions & Team Spirit:** This topic has the second-highest total NER count. It has a high average count of organizations (ORG) and persons, indicating a strong focus on specific teams or individuals.
3. **Business & Pension:** This topic has a high count of dates and organizations (ORG), suggesting a focus on time-related events and specific businesses or pension plans.
4. **Future Planning & Financial Goals:** This topic has a high count of dates and persons, suggesting a focus on time-related goals and individual financial planning.
5. **Investments & Financial Advice:** This topic has a high count of cardinal numbers and organizations (ORG), indicating a strong focus on numerical data and specific investment firms or advice agencies.
6. **Digital Banking & Mobile Services:** This topic has a high count of cardinal numbers and dates, suggesting that posts under this topic may contain more numerical data and time-related information.

- 7. **Insurance & Life Planning:** This topic has a relatively high count of dates, suggesting a focus on time-related life planning and insurance policies.
- 8. **Customer Service & Account Management:** This topic has the highest count of cardinal numbers but the lowest total NER count, indicating that it may contain more numerical data but fewer named entities compared to other topics.





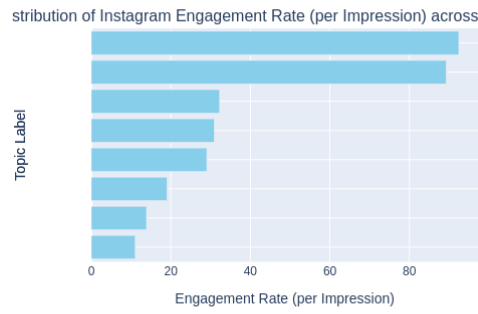
Based on the data shared, here are the topics that generate the most impressions and reach:

1. **Impressions:** The topic “Competitions & Team Spirit” generates the most impressions, with an average of approximately 2522.78. This is closely followed by “Customer Service & Account Management” with an average of approximately 2461.11 impressions.
2. **Reach:** The topic “Customer Service & Account Management” has the highest reach, with an average of approximately 1721.90. This is closely followed by “Digital Banking & Mobile Services” with an average reach of approximately 1673.72.

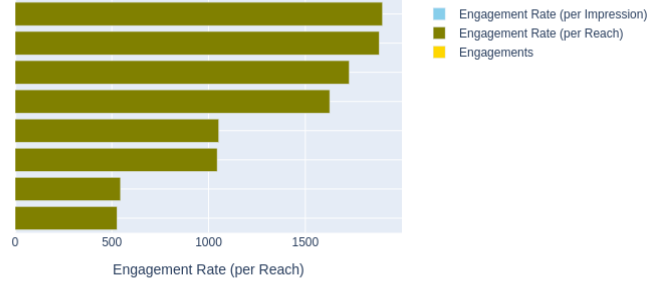
There are indeed significant differences in impressions and reach across different topics. For instance, “Competitions & Team Spirit” generates the most impressions but has a lower reach compared to other topics like “Customer Service & Account Management” and “Digital Banking & Mobile Services”. Conversely, “Customer Service & Account Management” has the highest reach but doesn’t top the list in terms of impressions.

In terms of correlation between impressions and reach, it’s not always the case that topics with higher impressions also have a higher reach. For example, while “Competitions & Team Spirit” has the highest impressions, it doesn’t have the highest reach. This suggests that while this topic might be viewed multiple times by the same users (high impressions), it may not be reaching a wide audience (lower reach).

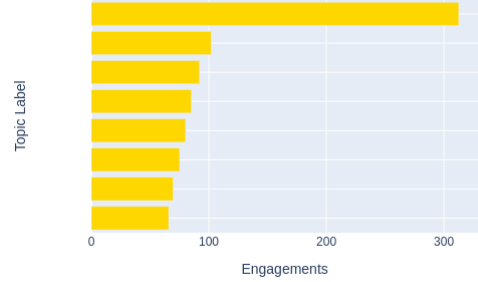
Distribution of Instagram Topic Labels



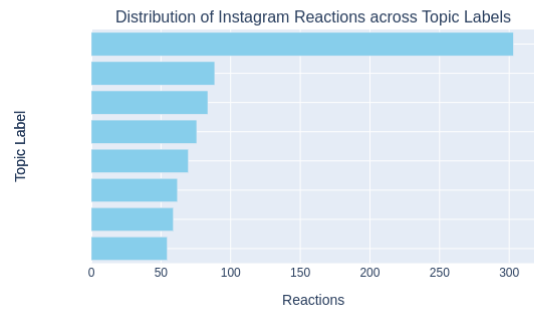
Distribution of Instagram Engagement Rate (per Reach) across Topic Labels



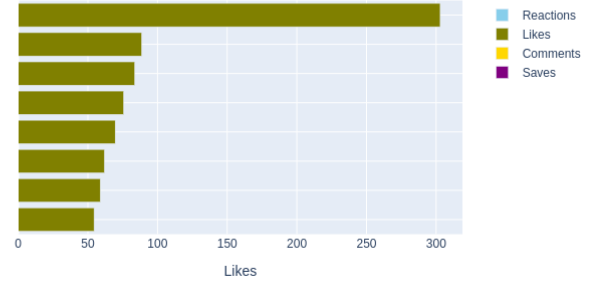
Distribution of Instagram Engagements across Topic Labels



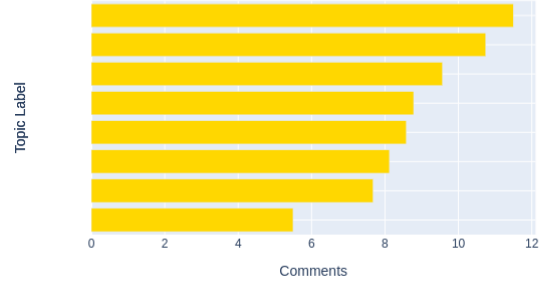
Distribution of Instagram Topic Labels



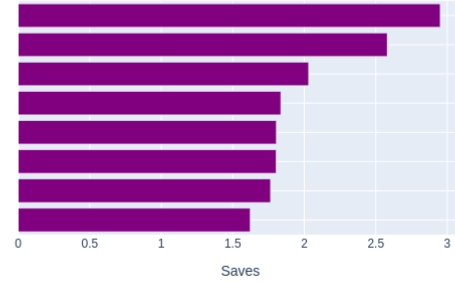
Distribution of Instagram Likes across Topic Labels



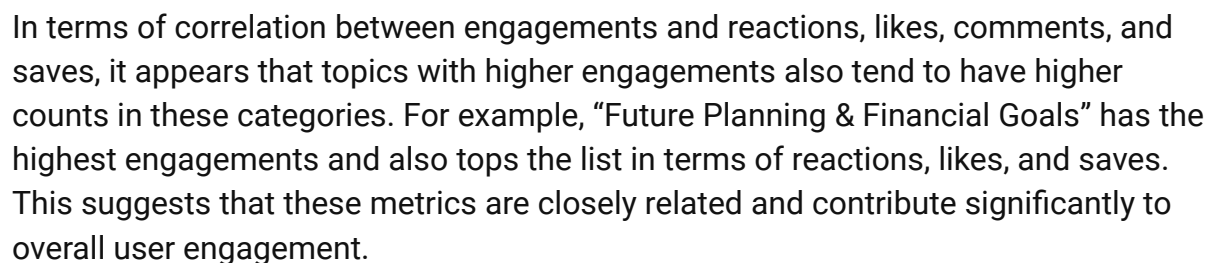
Distribution of Instagram Comments across Topic Labels



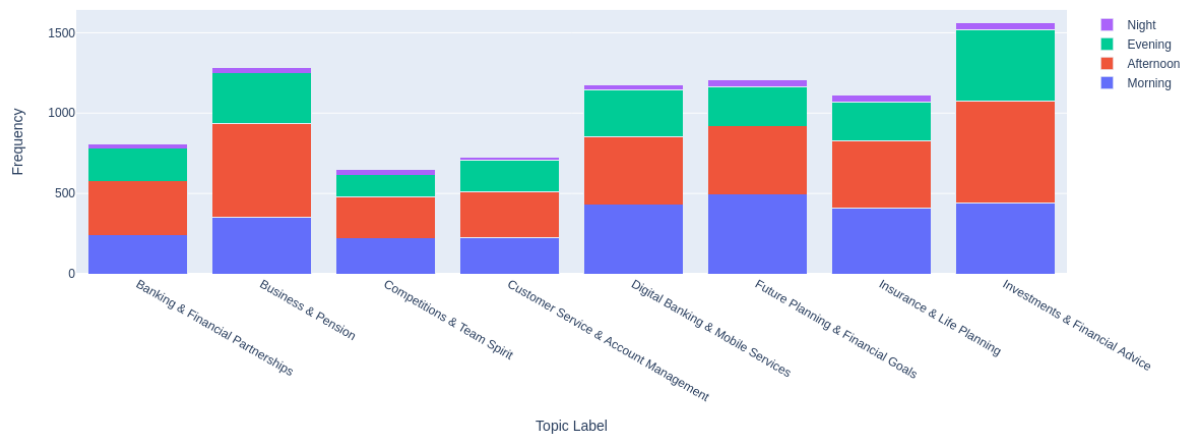
Distribution of Instagram Saves across Topic Labels



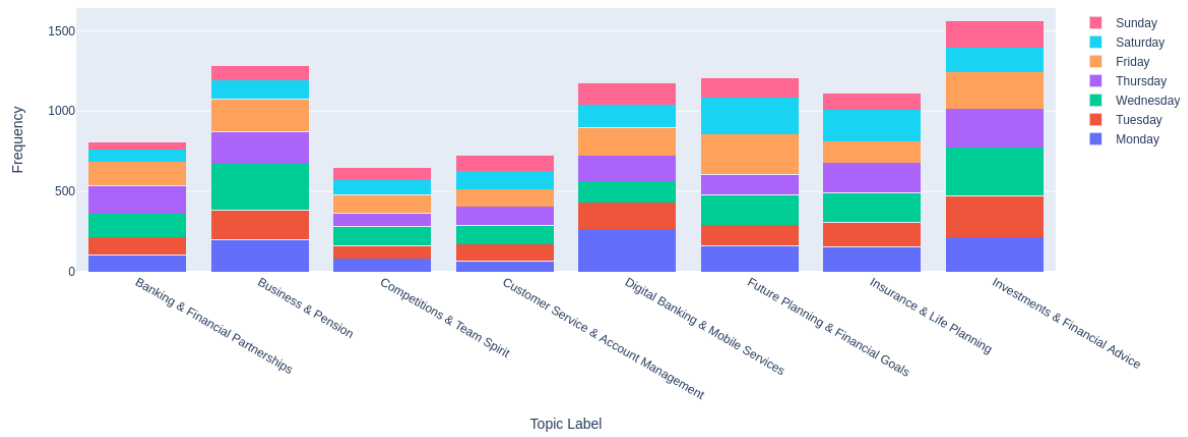
Based on the data shared, here are the topics that generate the most engagements:



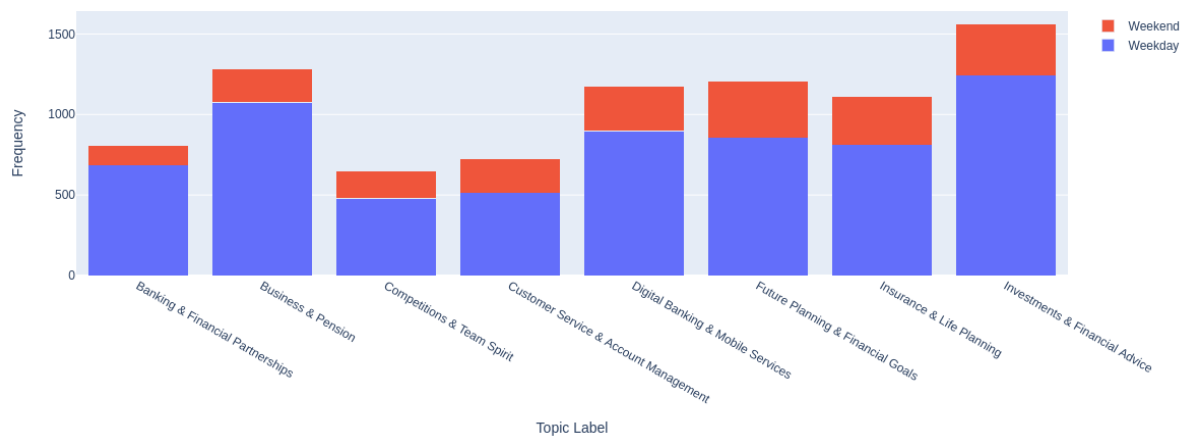
Distribution of Topic Labels Across Each time\_period

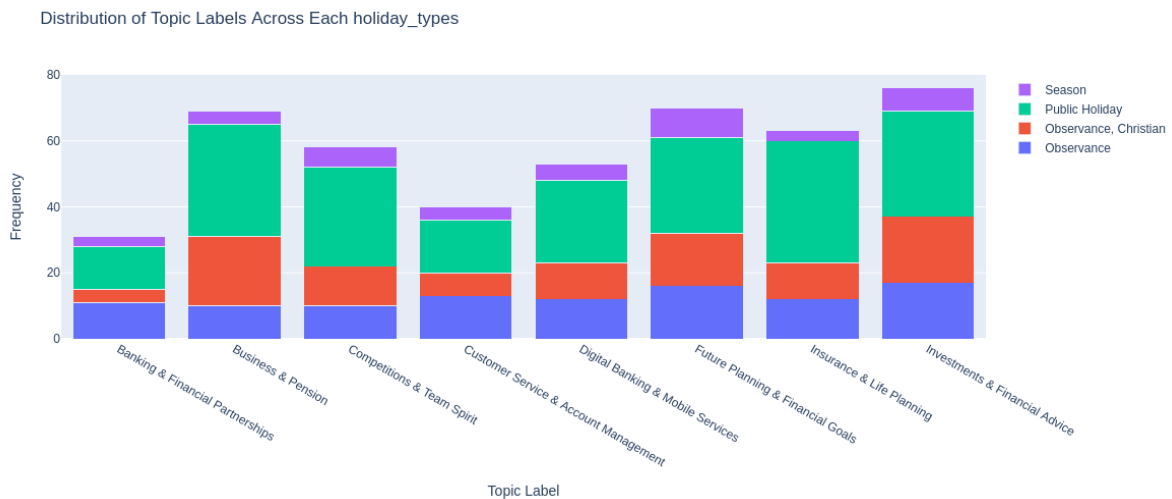
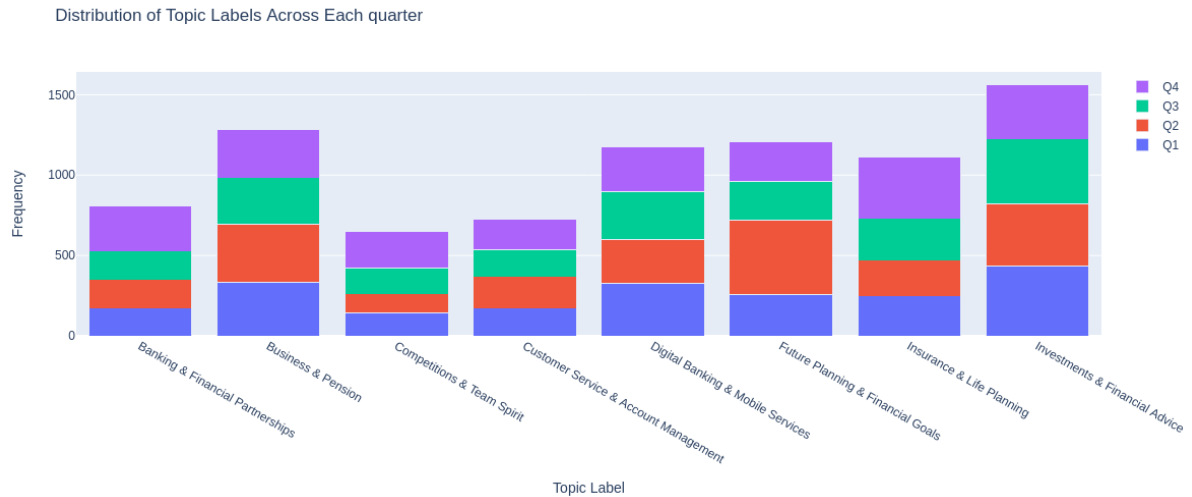


Distribution of Topic Labels Across Each day\_name



Distribution of Topic Labels Across Each day\_period





Based on the data shared, here are some insights that can be derived:

1. **Yearly Topic Frequency:** The frequency of topics varies across different years. For instance, “Business & Pension” and “Investments & Financial Advice” have seen a steady increase in frequency over the years, indicating growing interest in these topics.
2. **Time of Day:** Topics like “Investments & Financial Advice” and “Business & Pension” have higher engagements in the morning and afternoon. This could suggest that users are more likely to engage with these topics during typical working hours.
3. **Day of the Week:** Certain topics see more engagement on specific days of the week. For example, “Investments & Financial Advice” has the highest engagements on Tuesdays and Wednesdays, while “Business & Pension” sees more engagement on Tuesdays and Thursdays.



4. **Weekday vs Weekend:** Most topics have higher engagements during weekdays compared to weekends, suggesting that users are more active on Instagram during the week.
5. **Quarters:** The frequency of topics also varies across different quarters of the year. For instance, “Future Planning & Financial Goals” has higher frequency in Q2 and Q4, while “Investments & Financial Advice” is more evenly distributed across all quarters.
6. **Holidays vs Regular Days:** Most topics have higher frequency on regular days compared to holidays, indicating that users are more likely to engage with these topics on typical days.

These insights can inform content strategy by identifying when and what type of content is likely to generate the most user engagement. For example, posts related to “Investments & Financial Advice” might be scheduled for mornings on weekdays to maximize reach and engagement.

Analyzing the Posts in the top 1% Impressions some insights that can be derived:

1. **Yearly Topic Frequency:** The topic “Future Planning & Financial Goals” had the highest frequency in both 2021 and 2022 among the top 1% posts in terms of impressions. However, there was a significant drop in frequency from 2021 to 2022.
2. **Time of Day:** “Future Planning & Financial Goals” had the highest frequency in the morning and afternoon, suggesting that posts under this topic are more likely to generate high impressions during these times.
3. **Day of the Week:** The topic “Future Planning & Financial Goals” had a significantly higher frequency on Saturdays compared to other days, indicating that posts under this topic are more likely to generate high impressions on Saturdays.
4. **Quarters:** The topic “Future Planning & Financial Goals” had a high frequency in Q2 but no frequency in Q4, suggesting that posts under this topic are more likely to generate high impressions in the second quarter of the year.

These insights can inform content strategy by identifying when and what type of content is likely to generate the most impressions. For example, posts related to “Future Planning & Financial Goals” might be scheduled for mornings or afternoons on Saturdays during Q2 to maximize impressions.

Analyzing the Posts in the top 1% Engagements some insights that can be derived:

1. **Yearly Topic Frequency:** The topic “Future Planning & Financial Goals” had the highest frequency in both 2021 and 2022 among the top 1% posts in

terms of engagements. However, there was a significant drop in frequency from 2022 to 2023.

2. **Time of Day:** “Future Planning & Financial Goals” had the highest frequency in the morning and afternoon, suggesting that posts under this topic are more likely to generate high engagements during these times.
3. **Day of the Week:** The topic “Future Planning & Financial Goals” had a significantly higher frequency on Fridays and Saturdays compared to other days, indicating that posts under this topic are more likely to generate high engagements on these days.
4. **Quarters:** The topic “Future Planning & Financial Goals” had a high frequency in Q2 but no frequency in Q4, suggesting that posts under this topic are more likely to generate high engagements in the second quarter of the year.

These insights can inform content strategy by identifying when and what type of content is likely to generate the most engagements. For example, posts related to “Future Planning & Financial Goals” might be scheduled for mornings or afternoons on Fridays and Saturdays during Q2 to maximize engagements.

## Actionable recommendations

Based on the insights derived from the data, here are some actionable recommendations:

1. **Optimize Posting Schedule:** Posts made in the afternoon and on weekends tend to have higher reach and impressions. Consider scheduling more posts during these times to maximize engagement.
2. **Focus on High-Performing Topics:** Certain topics consistently generate more engagements, reactions, and impressions. Prioritize these topics when creating new content.
3. **Maintain Regular Posting Schedule:** Regular type posts outperform holiday posts in terms of reach, impressions, and engagements. Maintain a consistent posting schedule throughout the year to keep engagement levels high.
4. **Leverage Seasonal Trends:** There are noticeable increases in engagements and impressions in Q2. Plan marketing campaigns or special content around this time to take advantage of these seasonal trends.
5. **Leverage User Saves:** Although saves have been volatile, they have generally been trending downwards. Consider strategies to encourage users to save posts for later, as this can increase the likelihood of future engagements.
6. **Tailor Content to Topic Complexity:** The complexity of sentence structures varies across topics. For topics with more complex sentence structures,

consider creating content that matches this complexity to engage users who are interested in these topics.

7. **Highlight Named Entities:** Some topics have a high count of named entities. Consider highlighting these entities in your posts to draw attention to specific organizations, persons, or concepts.
8. **Focus on Key Aspects:** Each topic has a unique focus, whether it's actions, time-related events, numerical data, or specific entities or concepts. Tailor your content to these aspects to make it more relevant and engaging for your audience.
9. **Simplify Content Where Needed:** For topics with simpler sentence structures and fewer named entities, consider simplifying your content to make it more accessible and easy to understand.
10. **Focus on High-Performing Topics:** The topics "Competitions & Team Spirit", "Customer Service & Account Management", and "Future Planning & Financial Goals" generate the most impressions, reach, and engagements respectively. Prioritize these topics when creating new content.
11. **Optimize Posting Schedule:** Posts made in the morning and afternoon tend to have higher engagements. Consider scheduling more posts during these times to maximize engagement.
12. **Consider Day of the Week:** Certain topics see more engagement on specific days of the week. Tailor your content schedule to these patterns to maximize engagement.
13. **Leverage Seasonal Trends:** There are noticeable increases in engagements during Q2 and Q4. Plan marketing campaigns or special content around this time to take advantage of these seasonal trends.

## Conclusion

On Instagram we have an average reach of 1500 users and the average Impression of 2270, having our max reached Users and Impressions at 32418.0 and 39681.0 respectively. Our mean engagements value is 112, and our mean number of likes is 101, having our max engagements value at 33395.0 and our max likes at 33394.0