Facebook 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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9803 entries, 0 to 9802
Columns: 147 entries, Date to Tags
dtypes: float64(78), object(69)
memory usage: 11.0+ MB
```

```
Impressions: This metric tells you how many people have seen your content. If you have a lot of impressions, it means that your content is being seen by a lot of people. However, it's important to note that impressions don't necessarily mean engagement. Just because someone saw your content doesn't mean that they interacted with it in any way.
```

```
Organic Impressions: This metric tells you how many people have seen your content without being promoted. This is a good metric to track to see how well your content is performing on its own. If you have a lot of organic impressions, it means that people are interested in your content and are finding it without you having to pay to promote it.
```

```
Viral Impressions: This metric tells you how many people have seen your content because someone shared it, liked it, or commented on it. This is a good metric to track to see how engaging your content is. If you have a lot of viral impressions, it means that people are liking and sharing your content with their friends.
```

```
Non-viral Impressions: This metric tells you how many people have seen your content without being promoted or shared. This is a good metric to track to see how well your content is performing on its own. If you have a
```

lot of non-viral impressions, it means that people are finding your content but are not choosing to interact with it.

Fan Impressions: This metric tells you how many people who are fans of your Page have seen your content. This is a good metric to track to see how well your content is performing with your existing audience. If you have a lot of fan impressions, it means that your fans are interested in your content and are seeing it.

Fan Organic Impressions: This metric tells you how many people who are fans of your Page have seen your content without being promoted. This is a good metric to track to see how well your content is performing with your existing audience without you having to pay to promote it.

Non-fan Impressions: This metric tells you how many people who are not fans of your Page have seen your content. This is a good metric to track to see how well your content is performing with a new audience. If you have a lot of non-fan impressions, it means that people are finding your content but are not choosing to follow your Page.

Non-fan Organic Impressions: This metric tells you how many people who are not fans of your Page have seen your content without being promoted. This is a good metric to track to see how well your content is performing with a new audience without you having to pay to promote it.

1. Reach: This is the total number of unique users who had any content from your Page or about your Page enter their screen. It includes both people who have liked your Pages and those who haven't.

2. Organic Reach: This metric represents the number of times your post appeared in people's newsfeeds. It gives you a strong 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.

3. Viral Reach: On Facebook, going viral means that a post has generated a great deal of attention in the form of a high number of likes, shares, and comments. So, viral reach would refer to the number of times a post was displayed because it was shared by users and their friends.

4. Non-viral Reach: This is the number of people who had any content from your Page enter their screen. This doesn't include when someone's friend likes or follows your Page, engages with a post, shares a photo of your Page and checks into your Page.

5. Fan Reach: This is the number of times any content from your Page or about your Page entered a person's screen for people who have liked your page.

What is the difference between Impressions and Reach?

Impressions measure the total number of times your content was displayed on a screen, regardless of whether or not it was seen by a unique person. For example, if one person saw your post three times, it would count as three impressions.

Reach measures the total number of unique people who saw your content. So, if the same person saw your post three times, it would only count as one reach.

In general, impressions are always higher than reach because the same person can see your content multiple times.

Here is an example:

Suppose you have a post that is seen by 100 unique people. If each of those people sees the post twice, your post would have 200 impressions but only 100 reach.

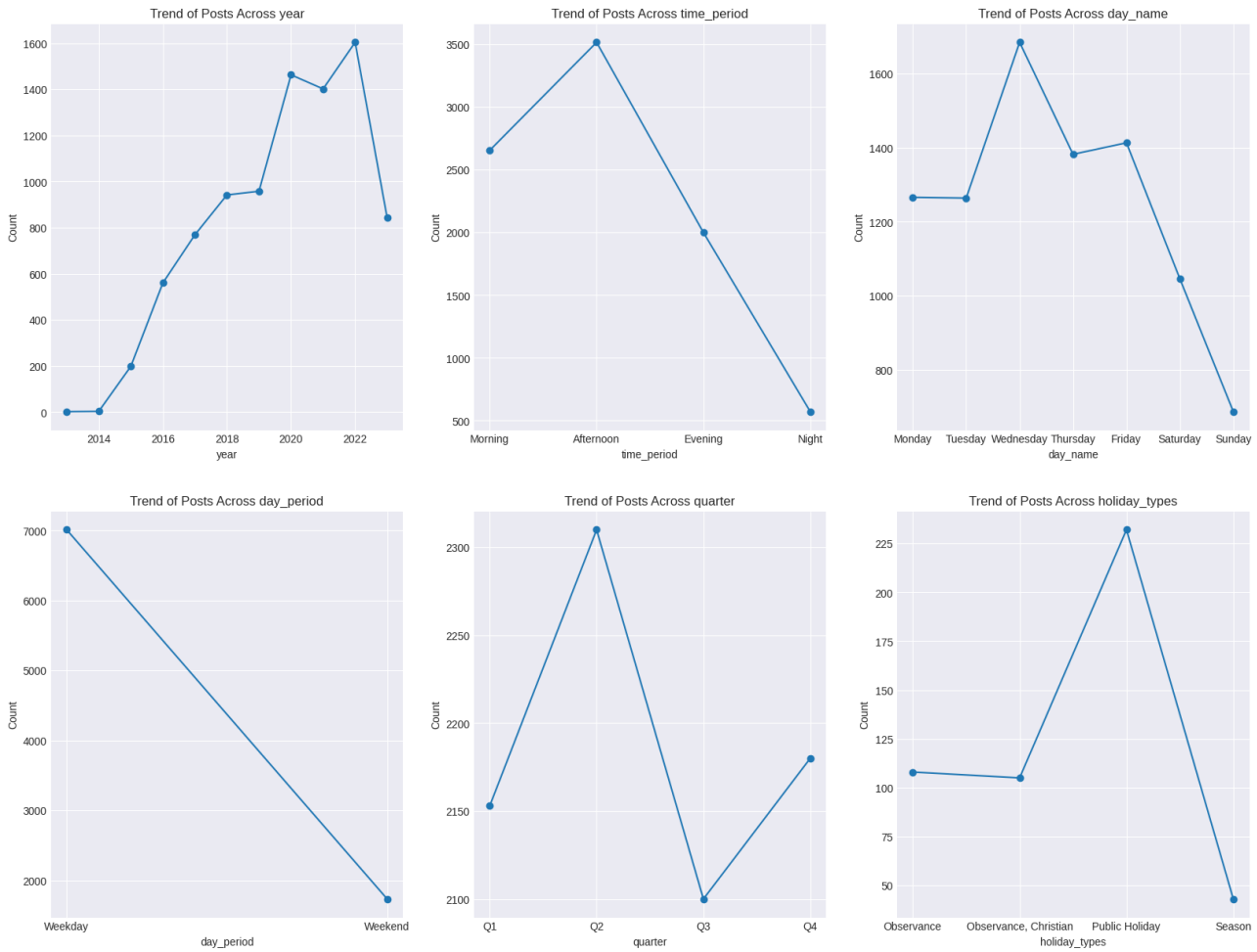
Why are Impressions and Reach important?

Impressions and Reach are both important metrics to track because they can give you insights into how well your content is performing.

Impressions can tell you how often your content is being seen. This can be helpful for understanding how visible your content is and how many people are exposed to your message.

Reach can tell you how many unique people are seeing your content. This can be helpful for understanding how many people you are reaching with your message.

Count of Posts against each feature

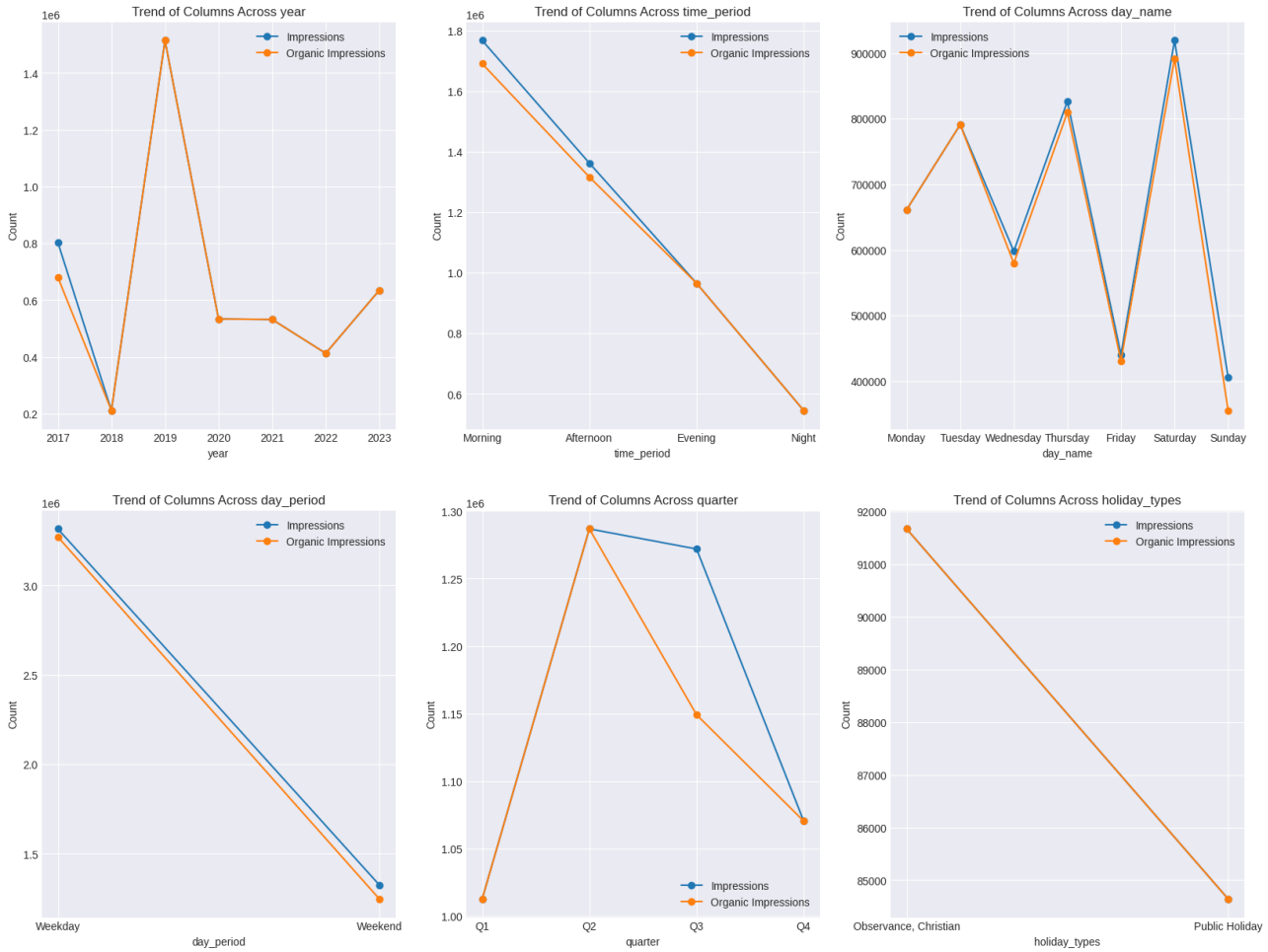



```
cleaned_impressions_reach[impression_cols].describe().T
```

	count	mean	std	min	25%	50%	75%	max
Impressions	8893.0	5857.725177	7295.418176	0.0	2019.0	4082.0	7376.0	207378.0
Organic Impressions	8893.0	5794.522996	7168.084516	0.0	2007.0	4063.0	7292.0	207378.0
Viral Impressions	8893.0	623.155066	1650.699781	0.0	5.0	88.0	544.0	35816.0
Non-viral Impressions	8893.0	5234.547959	6341.260127	0.0	1888.0	3784.0	6564.0	207373.0
Fan Impressions	8893.0	4919.463398	5747.002971	0.0	1678.0	3483.0	6207.0	103287.0
Fan Organic Impressions	8893.0	4919.463398	5747.002971	0.0	1678.0	3483.0	6207.0	103287.0
Non-fan Impressions	8893.0	938.261779	3067.161092	0.0	263.0	466.0	849.0	206231.0
Non-fan Organic Impressions	8893.0	875.059597	3002.945270	-1930.0	242.0	449.0	804.0	206231.0

It seems like we have a lot of Outliers in our Dataset, let's see what is the number of outliers we have

```
last_quantile = cleaned_impressions_reach['Impressions'].quantile(0.99)
#Posts that had Impression in the Top 1%
df = cleaned_impressions_reach[cleaned_impressions_reach['Impressions'] >
last_quantile]
```



There doesn't seem to be a correlation between the number of posts in a year and the number of Impressions gotten across the Year. even when there was a sudden increase in the number of posts in the year 2018 - 2022, there was a large decrease in the sum of Impressions in 2018, and a large decrease between the years 2019 - 2022.

The number of Impressions across the Time period, seems to decrease linearly from Morning till Evening, this may very well be part of the Facebook Algorithm

Thursday and Saturday seem to have the highest number of impressions, compared to wednesday and friday who have the most posts, what makes thursday and saturday so special?

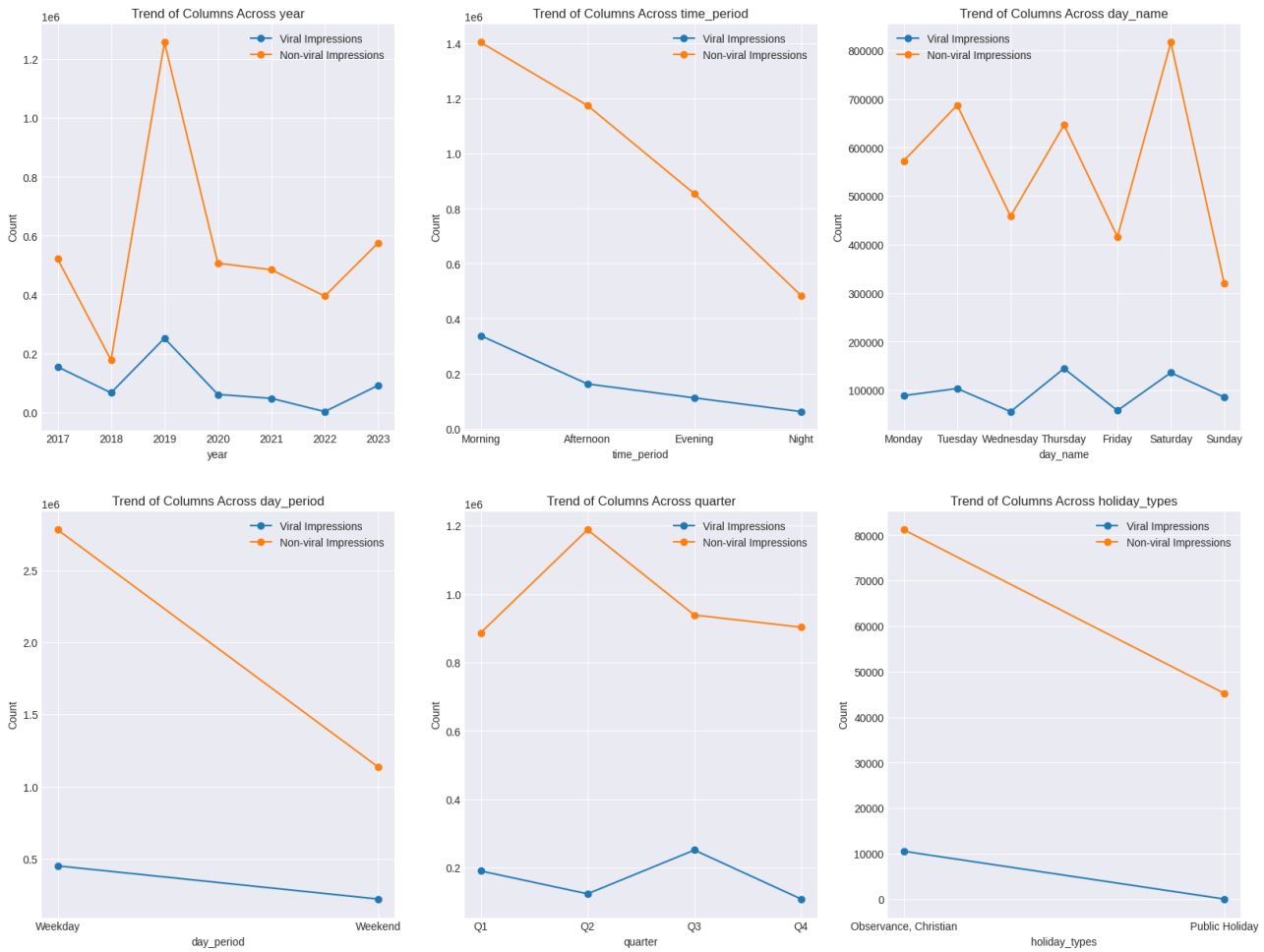
```
The sum of Impressions rises from the first quarter to the second quarter but drops rapidly till the fourth quarter, there seems to be a lot of unorganic impressions after the 2nd quarter

Christian Holidays and other religious Holidays have more impressions than others

The Highest Impressions Scores on the Posts were between the years 2018 and 2020, with 2019 having the highest possible values, there was a spike from 2017 which quickly went down but rose up from 2018 till 2019 and then there was a sharp drop from 2019 till 2020.

Having lowest dips in 2018 and 2022

last_quantile = cleaned_impressions_reach['Non-viral Impressions'].quantile(0.99)
df = cleaned_impressions_reach[cleaned_impressions_reach['Non-viral Impressions'] > last_quantile]
```

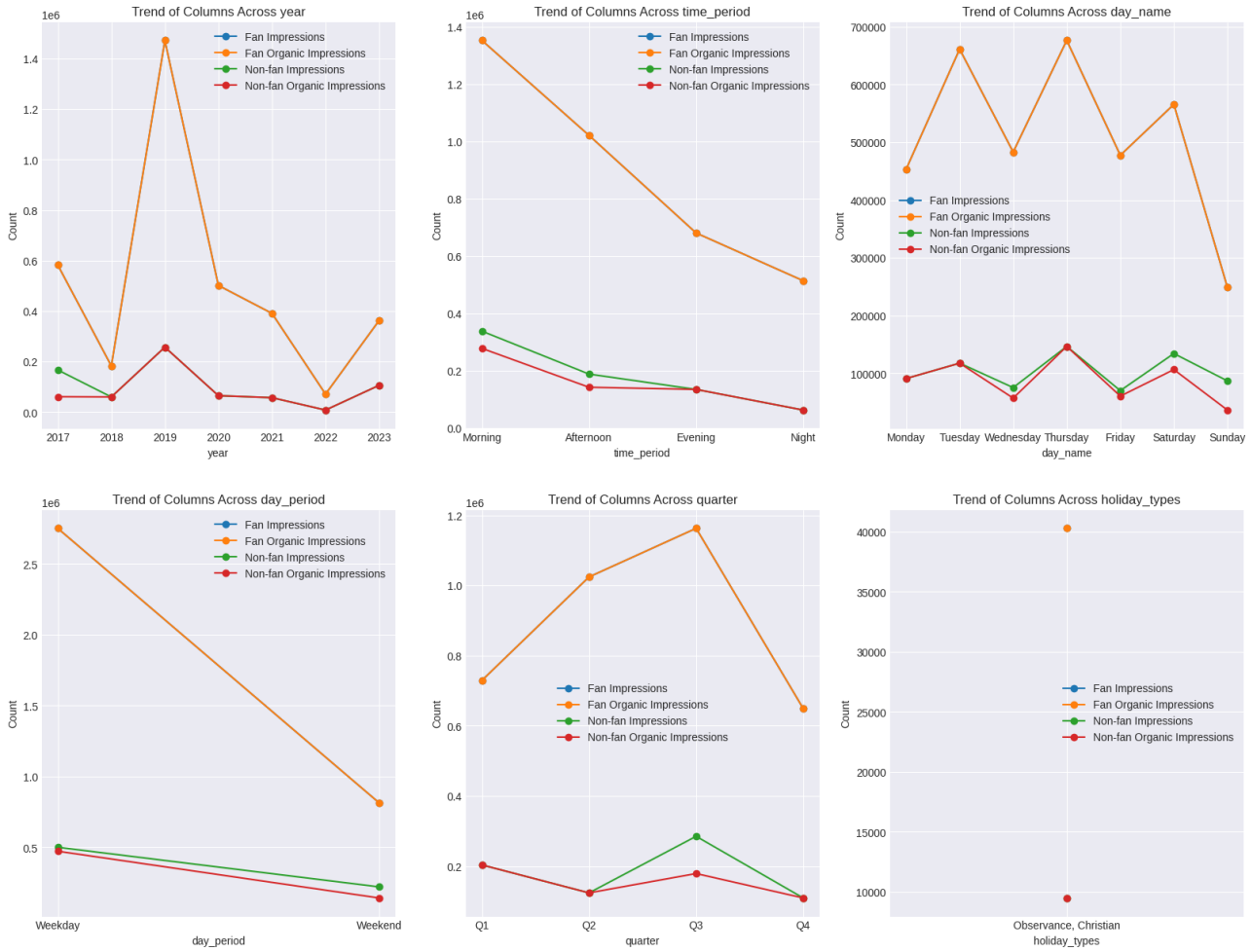


We can see that Viral Impressions and Non Viral Impressions show different results across

the Quarter, were we have the peak viral impressions in the third quarter and the peak non viral impressions in the Second Quarter, The results seem to be the same everywhere else with thursday and friday having the highest peak Viral/Non-viral impressions

```
last_quantile = cleaned_impressions_reach['Fan Impressions'].quantile(0.99)
df = cleaned_impressions_reach[cleaned_impressions_reach['Fan Impressions'] > last_quantile]
```

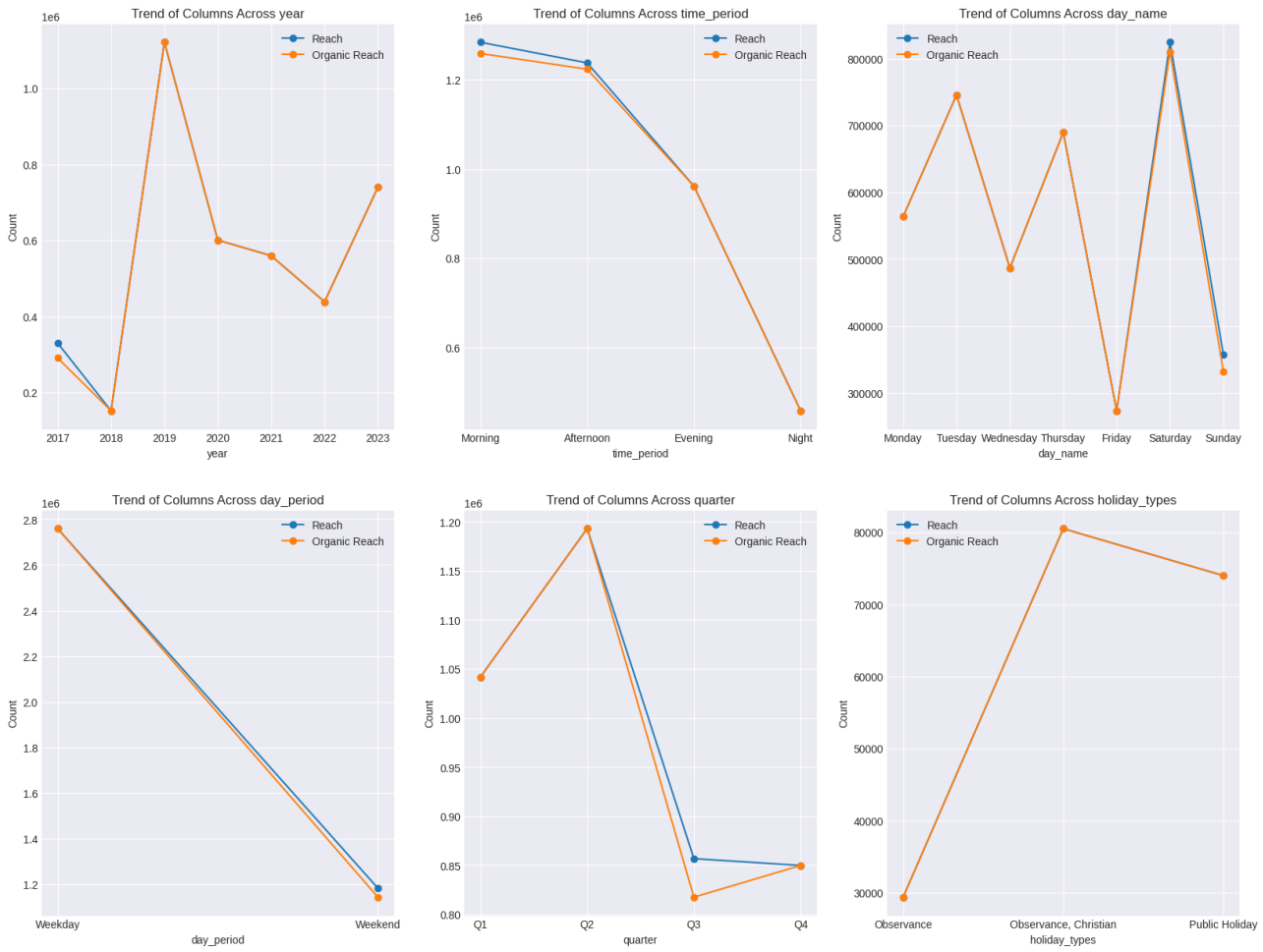
```
df.groupby('year')[impression_cols[4:]].sum()
```



The Same Trends Can be noticed about all the impressions with 2019 been that year, Morning – Night in that order, but seems most fans enjoy our contents on Tuesdays, Thursdays and Saturday, and also seems like the fans impression rises from q1 – q3 but drops in q4, but the non fan impression drops and rises in between quarters.

```
last_quantile = cleaned_impressions_reach['Reach'].quantile(0.99)
df = cleaned_impressions_reach[cleaned_impressions_reach['Reach'] > last_quantile]
```

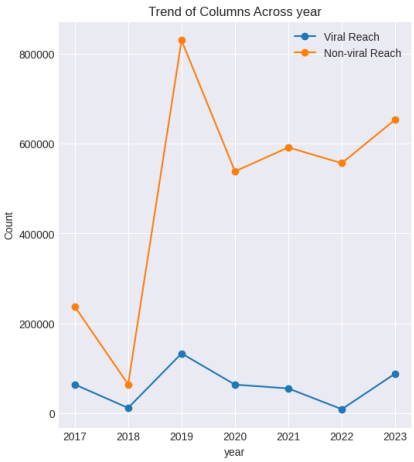
```
df.groupby('year')[reach_cols[:2]].sum()
```

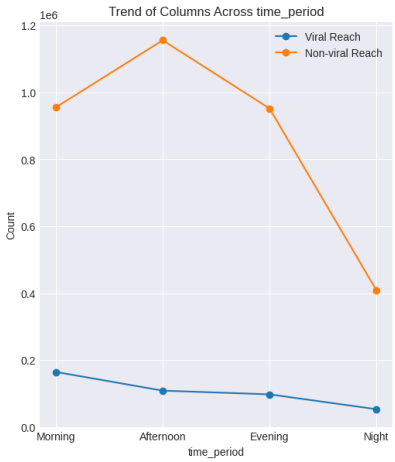


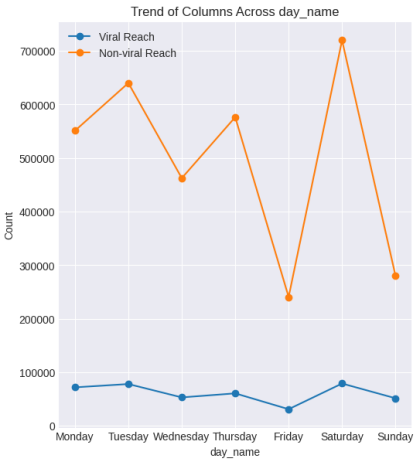
Our Reach and Organic Reach seems to have different result, 2019 is still that year, with our Reach decreasing from Morning – Night, Weekday – Weekend, but based on the day we see Tuesday and Saturday having the highest reach, and Friday and Sunday been so low. We see a rise from q1 – q1 but a sharp drop to q3, Christian Observances have the highest Reach with Public Holidays too.

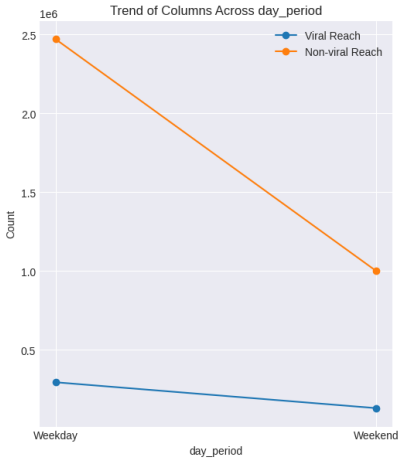
```
last_quantile = cleaned_impressions_reach['Non-viral Reach'].quantile(0.99)
df = cleaned_impressions_reach[cleaned_impressions_reach['Non-viral Reach'] > last_quantile]
```

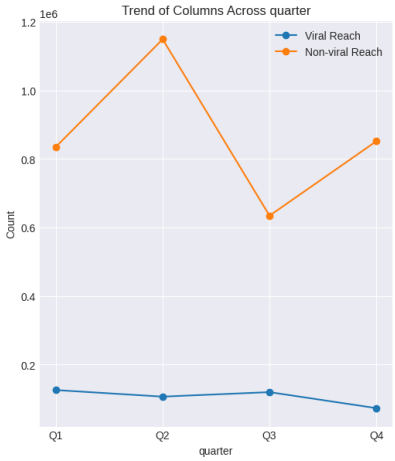
```
df.groupby('year')[reach_cols[2:4]].sum()
```

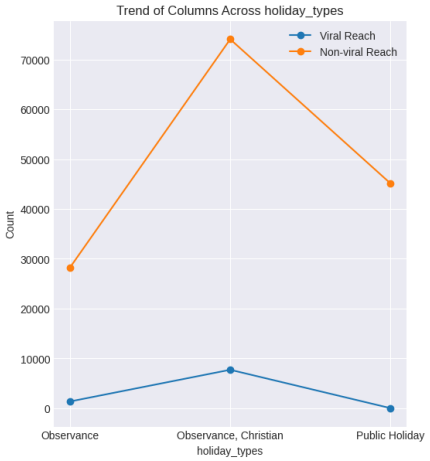

Trend of Columns Across year
Viral Reach
Non-viral Reach
Count
year

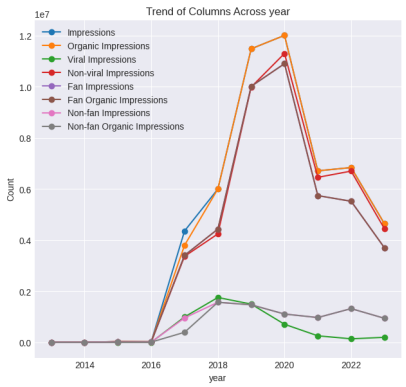

Trend of Columns Across time_period
1e6
Viral Reach
Non-viral Reach
Count
Morning
Afternoon
Evening
Night
time_period

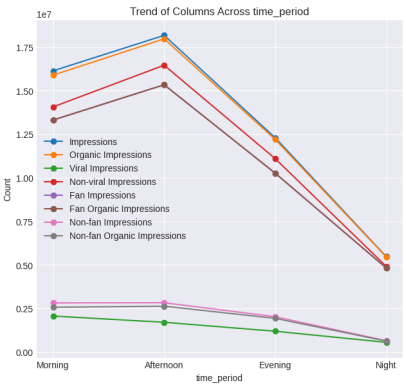

Trend of Columns Across day_name
Viral Reach
Non-viral Reach
Count
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
day_name

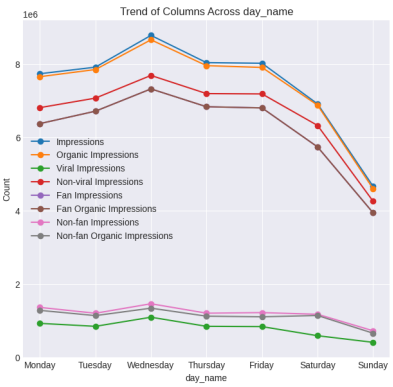

Trend of Columns Across day_period
1e6
Viral Reach
Non-viral Reach
Count
Weekday
Weekend
day_period

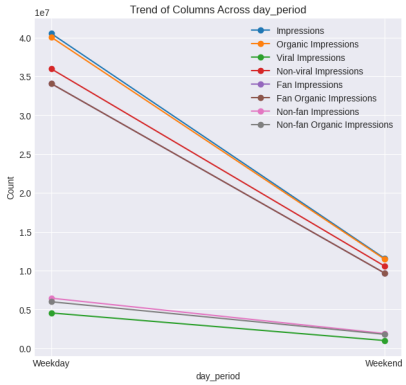

Trend of Columns Across quarter
1e6
Viral Reach
Non-viral Reach
Count
Q1
Q2
Q3
Q4
quarter

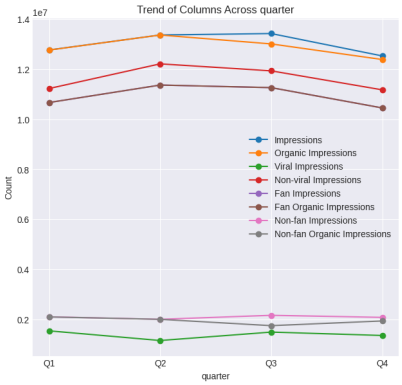

Trend of Columns Across holiday_types
Viral Reach
Non-viral Reach
Count
Observance
Observance, Christian
Public Holiday
holiday_types

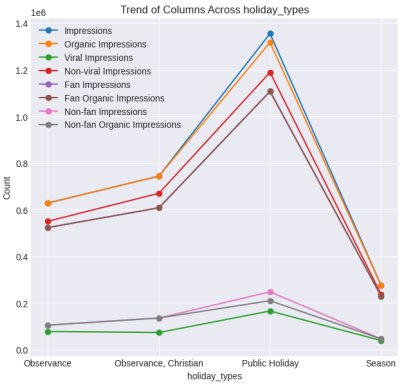

Trend of Columns Across year
1e7
Impressions
Organic Impressions
Viral Impressions
Non-viral Impressions
Fan Impressions
Fan Organic Impressions
Non-fan Impressions
Non-fan Organic Impressions
Count
year

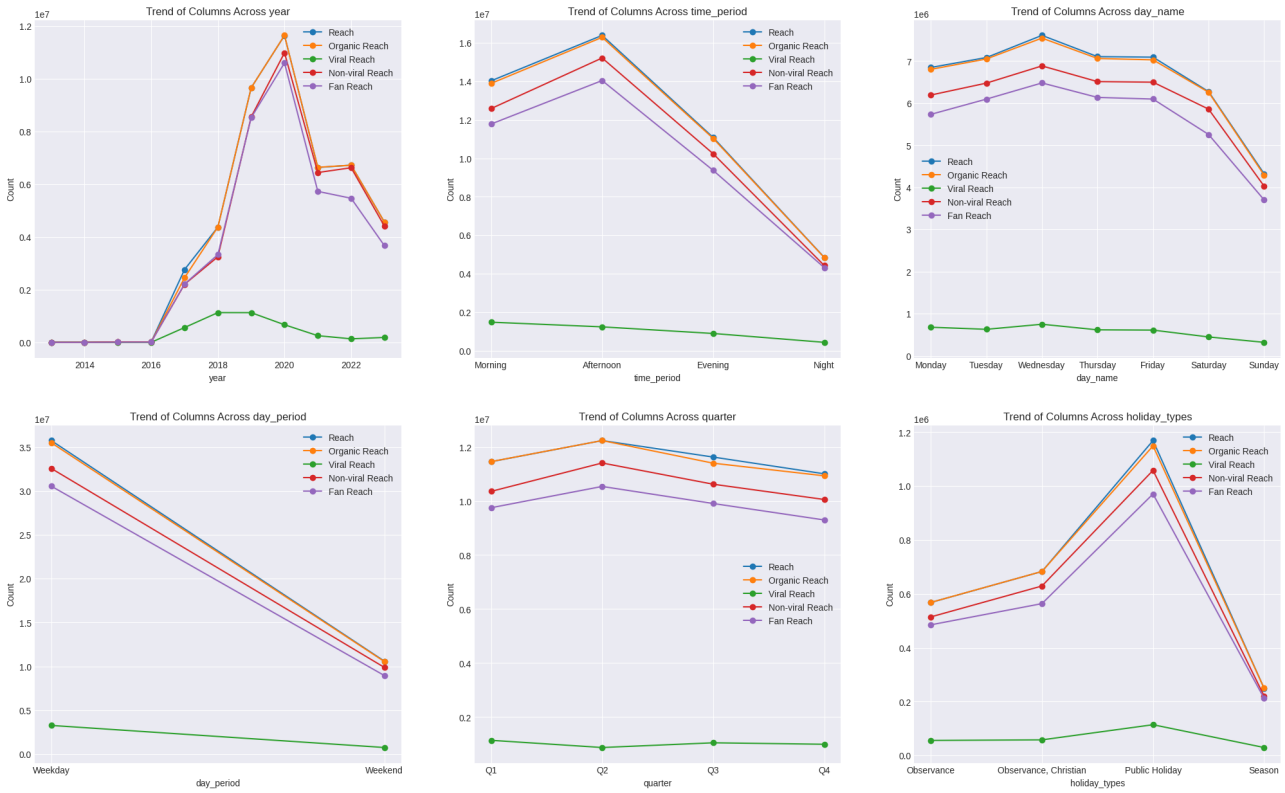

Trend of Columns Across time_period
1e7
Impressions
Organic Impressions
Viral Impressions
Non-viral Impressions
Fan Impressions
Fan Organic Impressions
Non-fan Impressions
Non-fan Organic Impressions
Count
Morning
Afternoon
Evening
Night
time_period


Trend of Columns Across day_name
1e6
Impressions
Organic Impressions
Viral Impressions
Non-viral Impressions
Fan Impressions
Fan Organic Impressions
Non-fan Impressions
Non-fan Organic Impressions
Count
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
day_name


Trend of Columns Across day_period
1e7
Impressions
Organic Impressions
Viral Impressions
Non-viral Impressions
Fan Impressions
Fan Organic Impressions
Non-fan Impressions
Non-fan Organic Impressions
Count
Weekday
Weekend
day_period

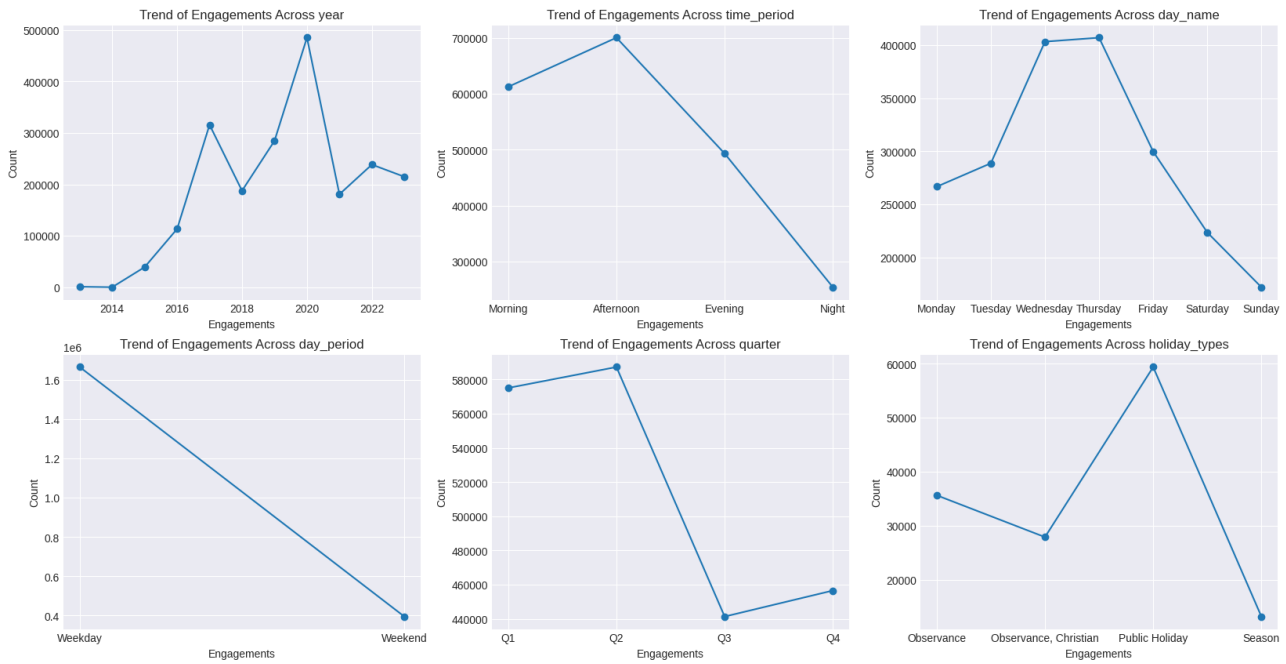

Trend of Columns Across quarter
1e7
Impressions
Organic Impressions
Viral Impressions
Non-viral Impressions
Fan Impressions
Fan Organic Impressions
Non-fan Impressions
Non-fan Organic Impressions
Count
Q1
Q2
Q3
Q4
quarter


Trend of Columns Across holiday_types
1e6
Impressions
Organic Impressions
Viral Impressions
Non-viral Impressions
Fan Impressions
Fan Organic Impressions
Non-fan Impressions
Non-fan Organic Impressions
Count
Observance
Observance, Christian
Public Holiday
Season
holiday_types



We can see that the Outliers have a different distribution from the Normal Distribution
In some cases, we can see that the spread of Reach/Organic Reach is very similar to the spread of the number of posts, implying that the higher number of posts, equate a Higher Reach/Impression

Time_period, Day_period, Day_name, Holiday_type



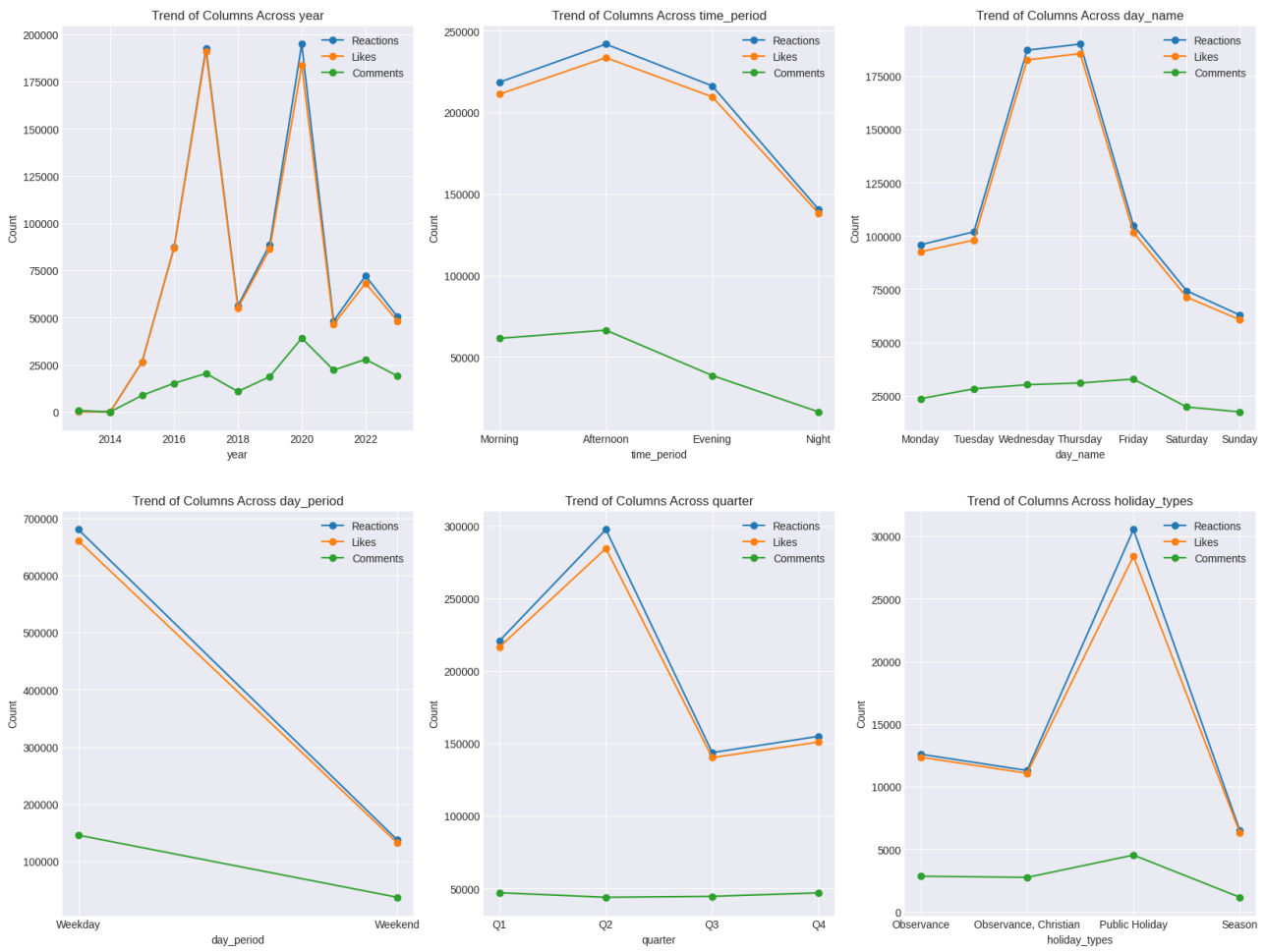
Now we can notice a trend of Engagements following a similar trend to the top 1%

Non-Viral Reach, across the Time Period, A rise from Morning to Afternoon and then a sharp decrease going to Night, Most Engagements gotten come from Wednesday and Thursday, the lowest coming on Sunday.

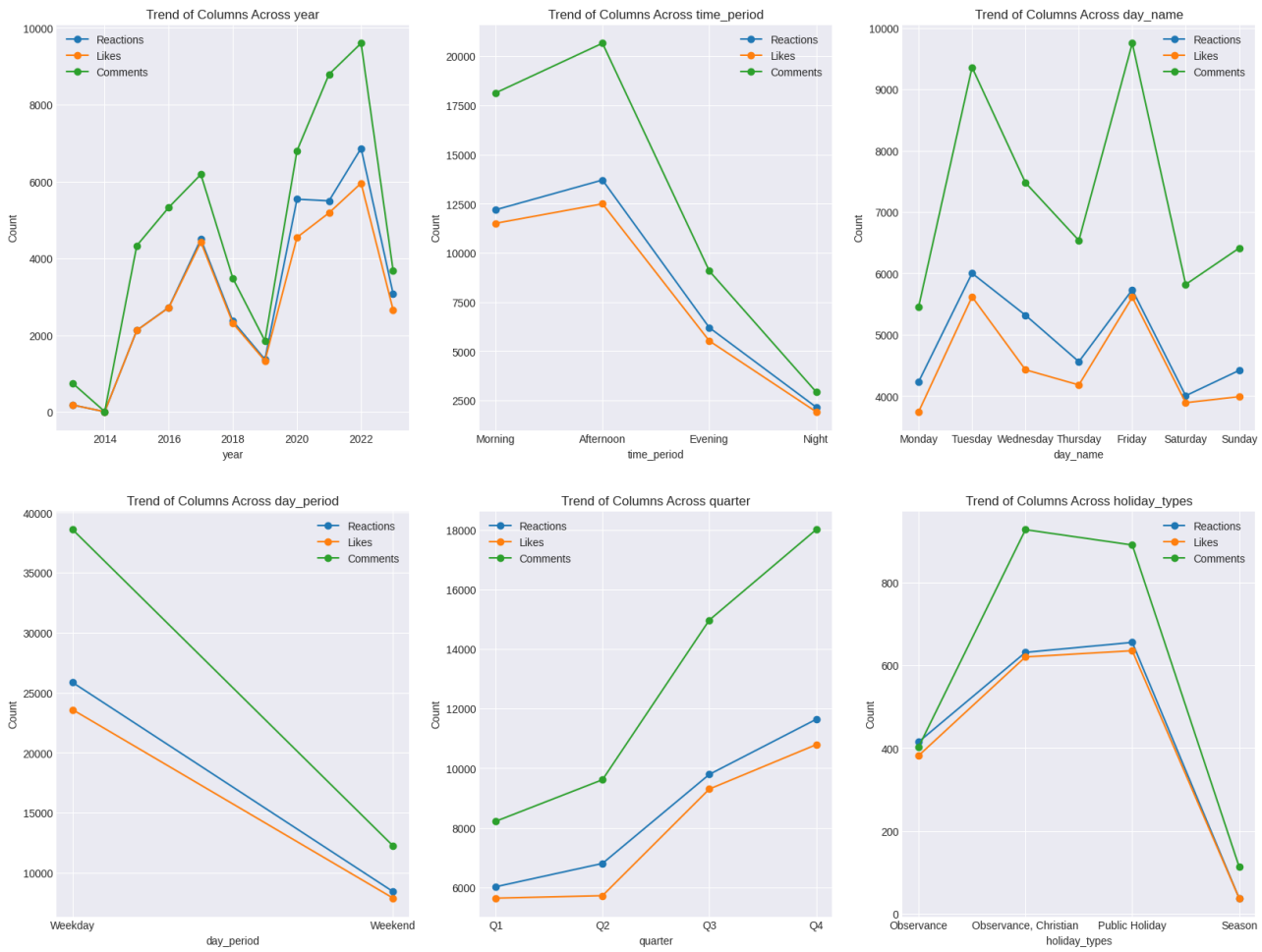
Let us check out the correlations

```
corr_matrix =
cleaned_impressions_reach[impression_cols[:2]+reach_cols[:2]+engagement_cols
+rlc_cols]
heatmap(corr_matrix, length=15, width=10)
```

Impressions
Organic Impressions
Reach
Organic Reach
Engagements
Reactions
Likes
Comments
1 1 0.98 0.97 0.24 0.036 0.034 0.43
1 1 0.98 0.98 0.23 0.036 0.033 0.42
0.98 0.98 1 1 0.23 0.035 0.033 0.42
0.97 0.98 1 1 0.23 0.034 0.032 0.41
0.24 0.23 0.23 0.23 1 0.96 0.96 0.63
0.036 0.036 0.035 0.034 0.96 1 1 0.47
0.034 0.033 0.033 0.032 0.96 1 1 0.46
0.43 0.42 0.42 0.41 0.63 0.47 0.46 1
1.0
0.8
0.6
0.4
0.2


Trend of Columns Across year
Count
year
Trend of Columns Across time_period
Morning
Afternoon
Evening
Night
time_period
Trend of Columns Across day_name
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
day_name
Trend of Columns Across day_period
Weekday
Weekend
day_period
Trend of Columns Across quarter
Q1
Q2
Q3
Q4
quarter
Trend of Columns Across holiday_types
Observance
Observance, Christian
Public Holiday
Season
holiday_types
Reactions
Likes
Comments

What happens when the number of comments is higher than the number of Reactions/Likes



We can notice 2017 and 2020 are the years with the highest record of Reactions and Likes, we can also see 2018, 2021, 2023 having the lowest record of Reactions and Likes, We can also see the trends of Reactions and Likes are very similar to the trends of Engagements. This can be explained by our correlation table as we can Reactions and Likes are closely correlated with Engagements having a score of 0.96

We observe that when the count of comments is higher than likes, this can be categorized as controversial or viral posts. From the plots what we can observe is 2017/2022 had posts with the most comments, with the lowest dates belonging to the year 2014, 2019, 2023. We observe a sharp drop from the time_period Afternoon - Night, the most controversial tweets seem to occur on tuesdays and fridays, with thursday being relatively low, alongside monday and saturday. We also observe a sharp increase from q1-q2-q3-q4 with the highest comments coming from posts made on Christian Observed Holidays and Public Holidays.

Sure, here are the definitions of the columns related to Facebook posts:

1. **Click-Through Rate (CTR):** This is the percentage of impressions that received a link click out of the total number of impressions. It indicates how many link clicks your ad

received compared to the number of impressions it received. It's a common metric used by online advertisers to understand how ads drive traffic to websites and other destinations.

2. **Other Post Clicks:** This metric includes any post click that isn't a link click, photo view, or video play. These clicks can include clicks on people's names in comments, clicks on the like count, or clicks on the time.

3. **Post Clicks (All)**: This metric includes link clicks as well as clicks on other parts of your ad. For example, someone might click on your Page's name. The following clicks may be measured under clicks all: Link clicks, Post likes, comments or shares, Clicks to a Facebook Page or Instagram profile, Clicks to expand a photo or video to full screen.

4. **Post Photo View Clicks:** This is the number of times users clicked on a photo in your post to view it.

5. **Post Video Play Clicks**: This is the number of times users clicked on a video in your post to play it.

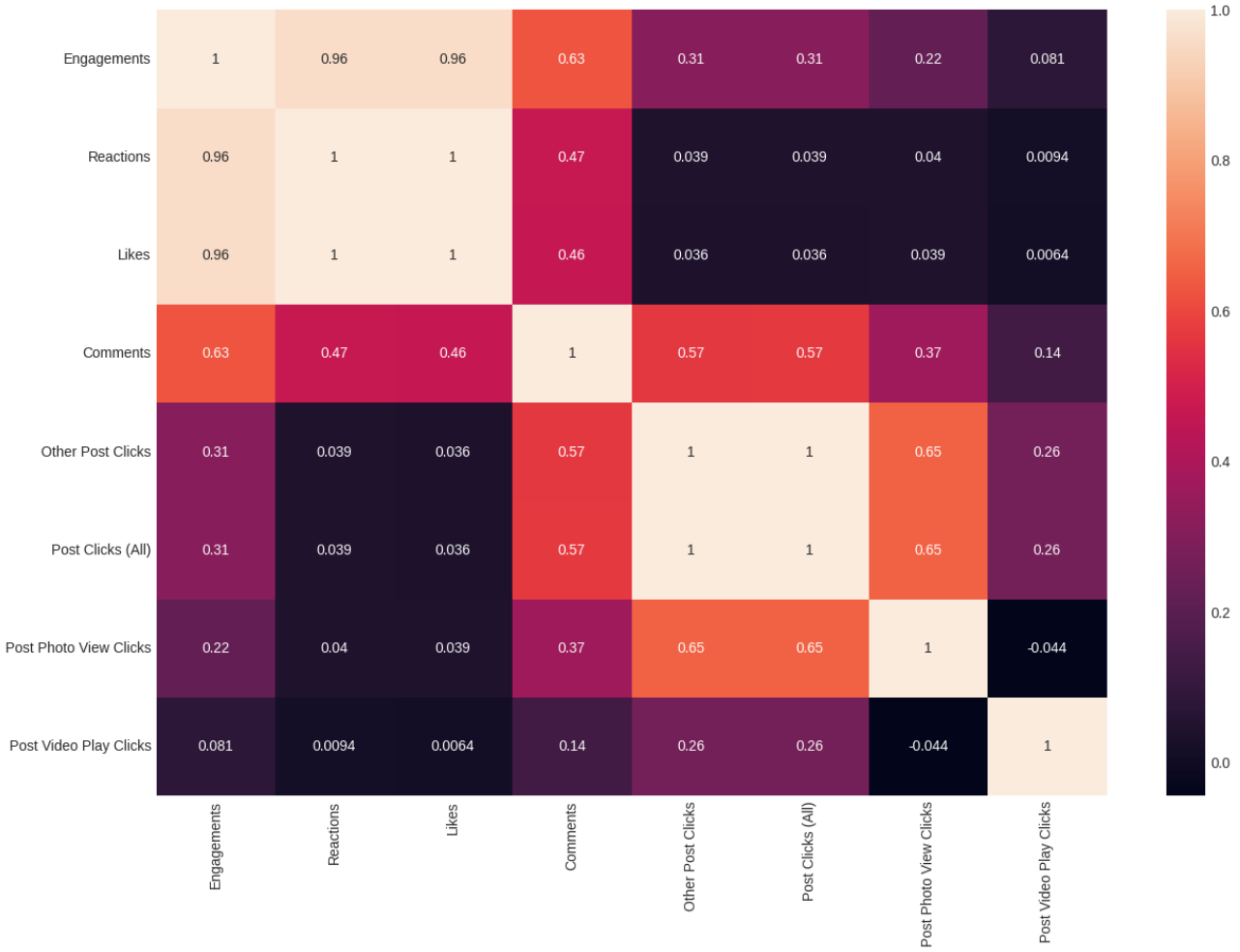
Click-Through Rate (CTR) and **Post Clicks (All)** are both metrics used in social media analytics, but they measure different things:

- **Click-Through Rate (CTR)** is a percentage that represents the number of times a link was clicked divided by the number of times the link was viewed (impressions). It's a measure of how effectively a piece of content is driving users to click on its links. A higher CTR generally indicates that users find the content relevant and are interested in what the link leads to.
- **Post Clicks (All)**, on the other hand, is a broader metric that includes all clicks associated with a post. This includes not only link clicks, but also other interactions such as likes, comments, shares, clicks on the page name or profile picture, and clicks to expand a photo or video to full screen. It's a measure of overall engagement with a post.

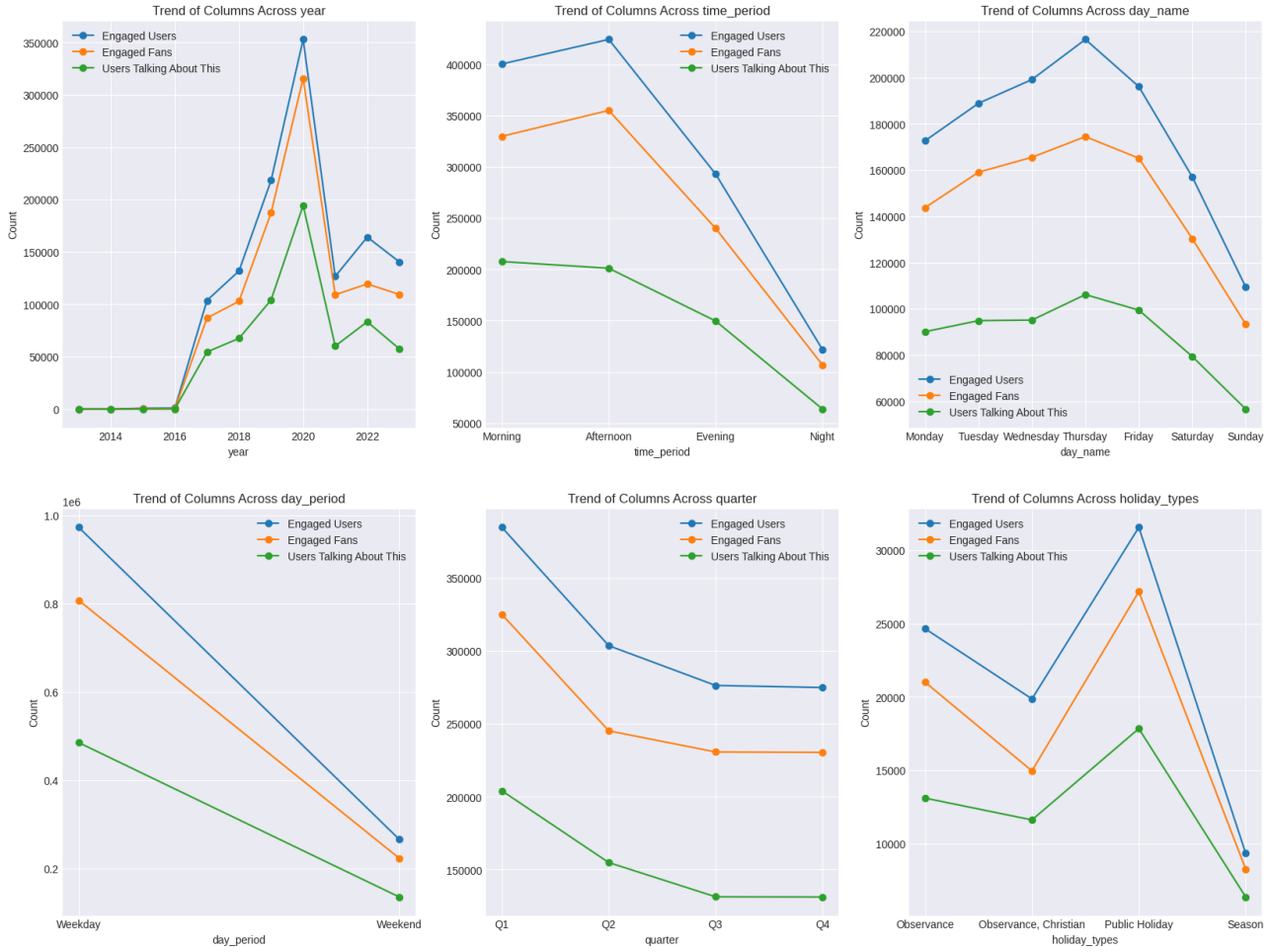
So, while CTR specifically measures how effectively your content drives users to click on links, Post Clicks (All) gives you a broader view of how users are interacting with your content.

When the **Other Post Clicks** count is less than the **Post Clicks (All)** count for a particular post, it means that there are some clicks that are not categorized as "Other". These could be link clicks, photo views, video plays, likes, comments, shares, clicks on the page name or profile picture, and clicks to expand a photo or video to full screen. These are all included in the **Post Clicks (All)** count but not in the **Other Post Clicks** count.

On the other hand, when the **Other Post Clicks** count equals the **Post Clicks (All)** count for a particular post, it means that all the clicks on the post are categorized as "Other". This could include clicks on people's names in comments, clicks on the like count, or clicks on the time. In this case, there are no link clicks, photo views, video plays or other types of clicks that are usually included in the **Post Clicks (All)** count. This might suggest that while users are interacting with your post, they may not be engaging with its main content (like links, photos or videos) as much.

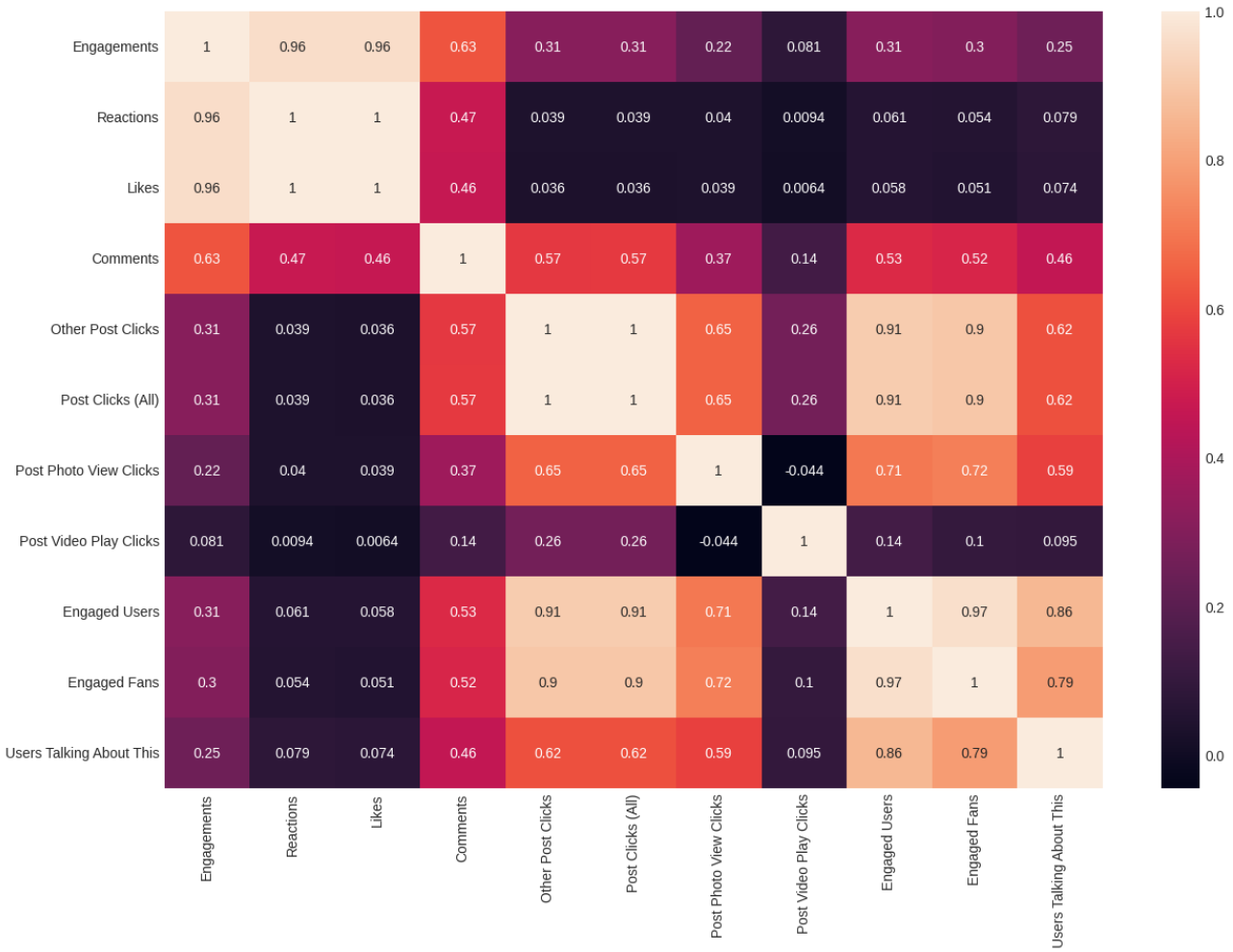


Interestingly, Post Clicks(All) and Other Post Clicks seems to have a relationship with the number of comments.




```
Comparing the Different Plots, when put alongside the Trend of Reactions,
Likes, and Comments, for th Year Column we can see that 2017 and 2020 were
the years with peak Reactions, likes, with comments having a peak at 2020
and 2022, but the trends of Engaged Users, Fans and Users talking about the
post, have their only peaks at the year 2020, shouldn't more Engaged Users
equal more Reactions? Likes and Comments? A sharp drop from the number of
Engaged Users, Fans from Afternoon to Evening, Compared to the drop of
Reactions, Likes and Comments, Are there Engaged Users but no Reactions,
Likes or Comments? The Highest Reactions, and Likes come from Wednesday and
Thursday, but Engaged Users, Fans only arrive on Thursday why is that?
During the Quarters, the Highest Reactions, Likes, come from the 2nd Quarter
with the lowest being in the 3rd Quarter, but here it shows the Engaged
Users/Fans dropping from q1 - q4
```

Trends of Engaged Users/Fans doesn't seem to follow the Trend of Engagements, why is that?



As we can see there is a very close correlation with Engaged Users/Engaged Fans and Other Post Clicks, Post Clicks (All) and Post Photo View Clicks,we can come to the conclusion that our Engaged Users/ Engaged Fans are those which add to the Other Post Clicks and Post Clicks (All), those Users/ Fans who are clicking links, opening photos, and making comments

In Facebook Metrics, the terms “Unique” and “All” are used to differentiate between the total number of actions and the number of unique people who took those actions. Here’s what each of the terms you asked about means:

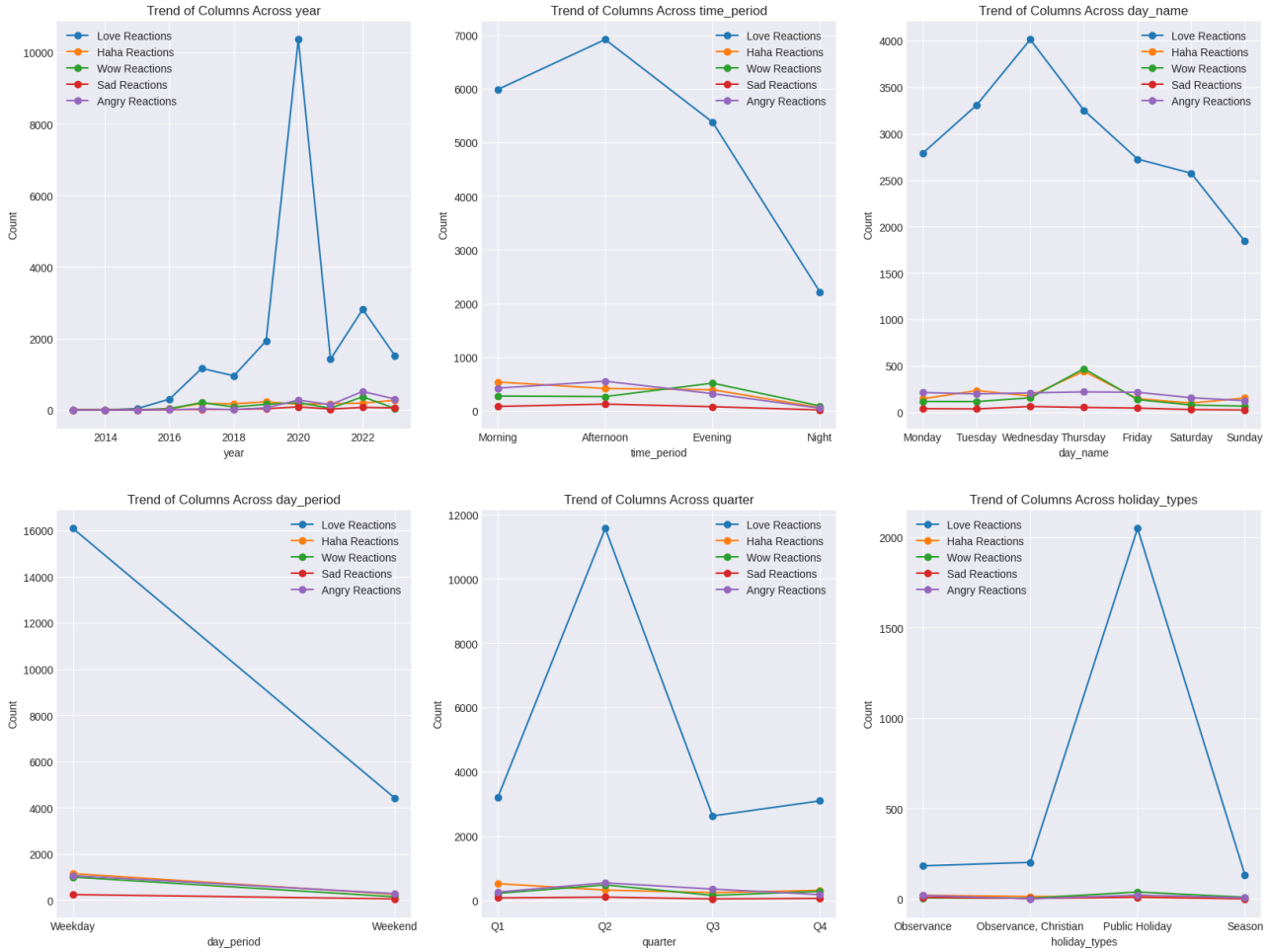
1. **Unique Reactions vs Reactions**: “Unique Reactions” refers to the number of unique people who reacted to your post, while “Reactions” is the total number of reactions your post received. If a single user reacts multiple times, “Reactions” will count each one, but “Unique Reactions” will count that user only once[1].
2. **Unique Post Photo View Clicks vs Post Photo View Clicks**: “Unique Post Photo View Clicks” is the number of unique people who clicked to view photos in your post, while “Post Photo View Clicks” is the total number of times users clicked to view photos[2]. If a single user views the photo multiple times, it will be counted each time in “Post Photo View Clicks”, but only once in “Unique Post Photo View Clicks”.
3. **Unique Post Clicks vs Post Clicks (All)**: “Unique Post Clicks” refers to the number of unique people who clicked anywhere on your post[3]. On the other hand, “Post Clicks (All)” includes all clicks on your ad or post, such as link clicks, post

likes, comments or shares, clicks to a Facebook Page or Instagram profile, and clicks to expand a photo or video to full screen[4].

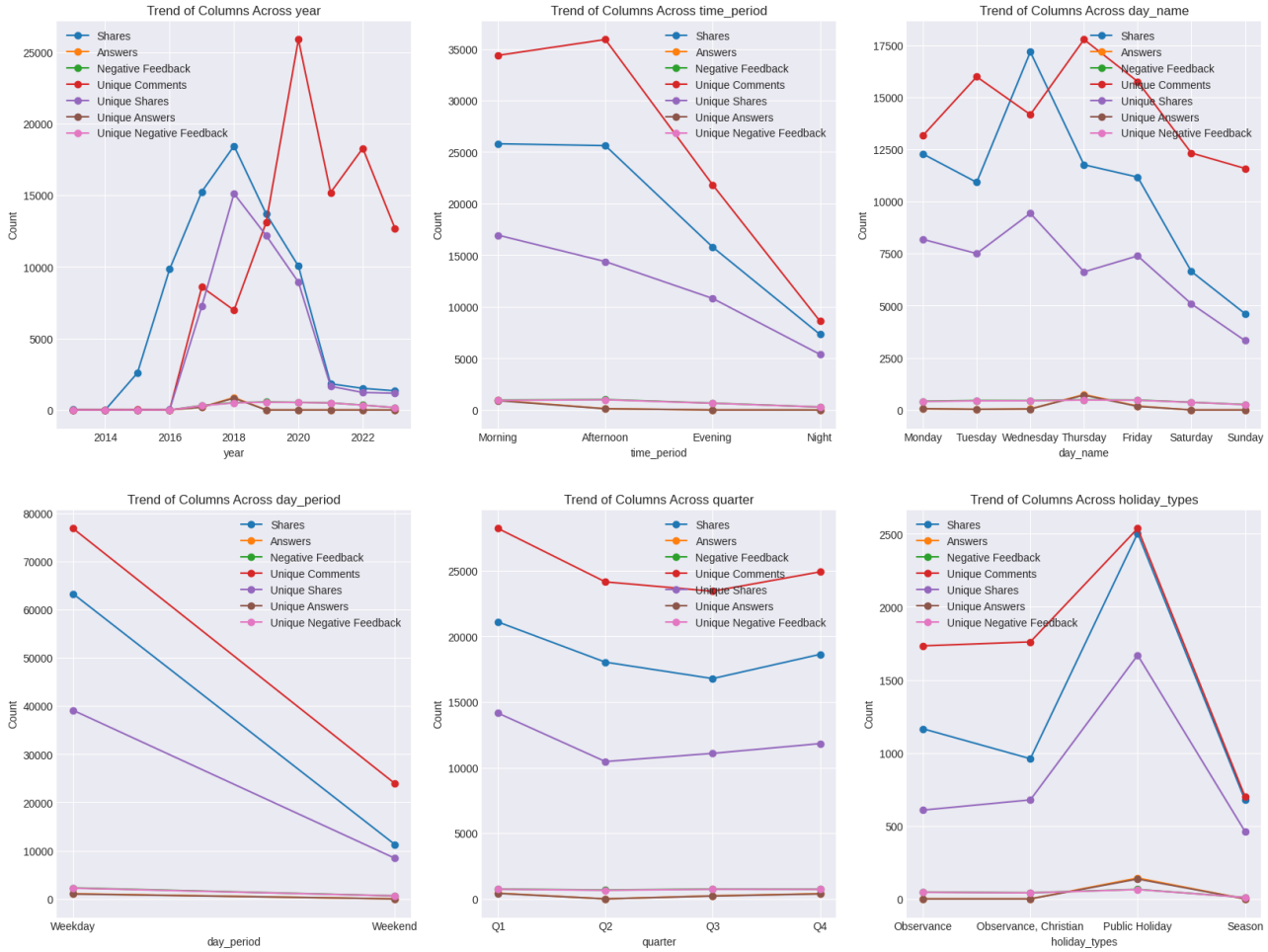
4. **Unique Other Post Clicks vs Other Post Clicks**: “Unique Other Post Clicks” is the number of unique people who clicked on parts of your post that aren’t link clicks, photo views, or video plays. “Other Post Clicks” is the total number of such clicks[5]. This can include clicks on people’s names in comments, clicks on the like count, or clicks on the time.

In summary, metrics with “Unique” count the number of unique people who took an action, while metrics without it count the total number of those actions[6].

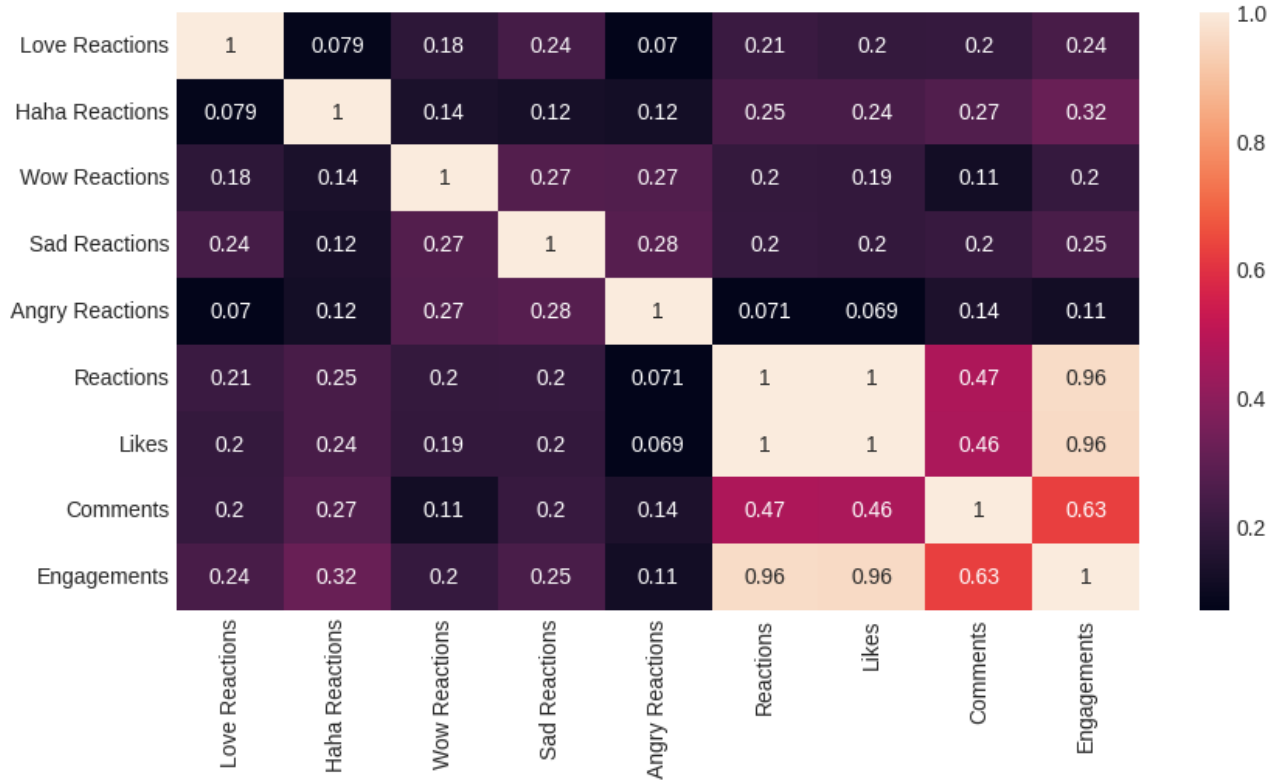
We see a closer correlation of Engaged Users, Engaged Fans and Users Talking About This, Users Talking About This --> Unique Reactions , Engaged Users/Fans --> Unique Post Clicks/Unique Other Post Clicks.



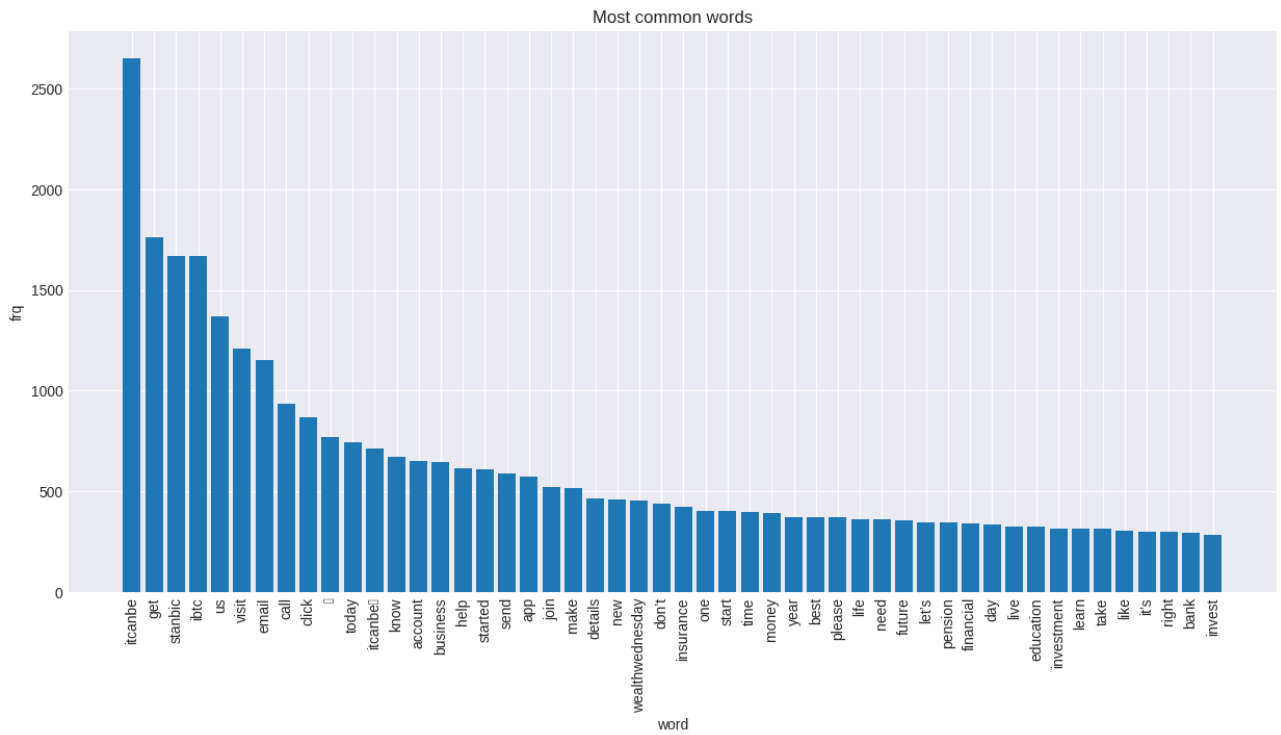
2020 seems to be the year with the most love Reactions compared to the Others, maybe a lot of Sensitive Posts were made then, Afternoon seems to be the period with the most Love Reactions what kind of Posts were made in the afternoon? Wednesday also has a lot of Love Reactions, Thursdays have a mixture of Haha and Wow Reactions, Second Quarter has a lot of love Reactions compared to the other Quarters, Public Holidays also have a lot of love Reactions



2020 has the most Unique Comments, then 2022 before dropping low in 2023, but 2018 has the most shared Posts and Unique shares, Afternoons best Unique Comments, Shares, Wednesday the best day for Shares, Thursday the best day for Unique Comments, Q1 >Q2 >= Q3 < Q4, Public Holidays taking this one home.



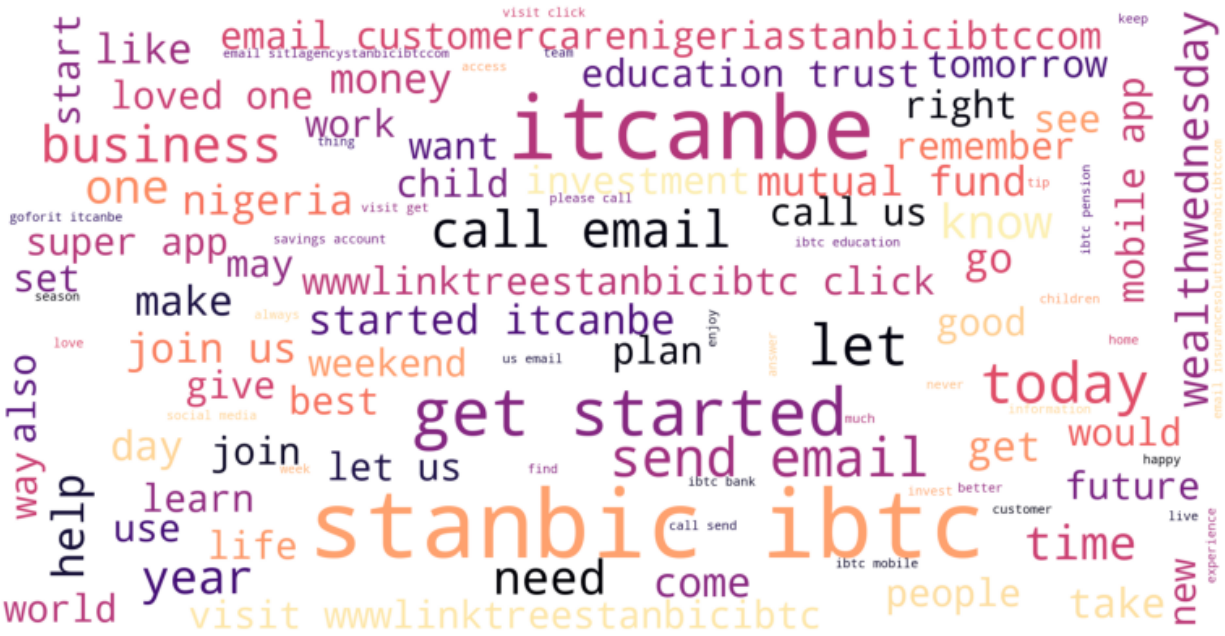
Top 50 words in our Facebook POSTS



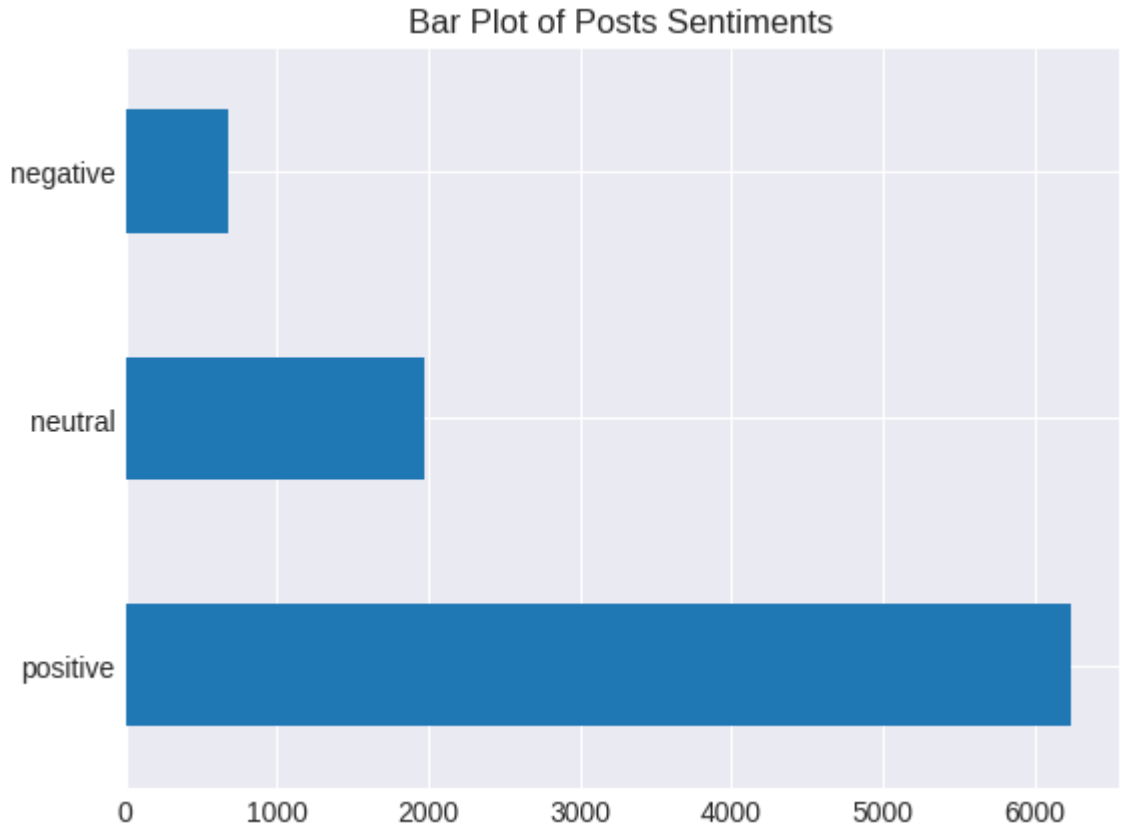
WORDCLOUD FOR TOP 100

The most frequent words in our captions are itcanbe, stanbic, ibtc, visit, us, email, call, click, know, account and business, we are convincing people to join the company, invest in the company, create accounts, and make business decisions with us

Word Cloud for Top 100 words from Facebook Posts



Checking out our Word Cloud distribution we can see the most frequent words, such as customer care nigeria stanbic ibtc education trust wealth wednesday get started, we can see that most of our posts revolve around Creating Accounts, Investments, Personal Finance and Future Planning, We can also see the Sentiment Labels of our posts, A majority of our posts are Positive, or have a Positive outlook, we have about 6000+ Positive Posts, With 2000 Posts being recognized as being Neutral, and < 1000 have a Negative Outlook.



Using Unsupervised Machine Learning Methods such as Topic Modeling:
Topic modeling is a way of abstract modeling to discover the abstract ‘topics’ that occur in the collections of documents. The idea is that we will perform unsupervised classification on different documents, which find some natural groups in topics. We can answer the following question using topic modeling.

What is the topic/main idea of the document?
Given a document, can we find another document with a similar topic?
How do topics field change over time?

Latent Dirichlet Allocation
Latent Dirichlet allocation is one of the most popular methods for performing topic modeling. Each document consists of various words and each topic can be associated with some words. The aim behind the LDA to find topics that the document belongs to, on the basis of words contains in it. It assumes that documents with similar topics will use a similar group of words. This enables the documents to map the probability distribution over latent topics and topics are probability distribution.

Let's break down these concepts:

1. **Topic Modelling using Latent Dirichlet Allocation (LDA)**: Imagine you have a huge pile of books but they don't have any labels or categories. You want to organize them, but reading each one would take forever. Topic Modelling is like a smart librarian who can skim through the books and sort them into different piles based on the main topics they cover. LDA is a popular method for doing this. It assumes that each book (or document) 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.

2. **Named Entity Recognition (NER)**: This is like a smart highlighter. When you read a text, you can usually pick out names of people, places, organizations, dates, etc. NER does the same thing but automatically. It scans through the text and highlights these "named entities". This can be very useful in understanding who or what the text is about.

3. **Part-of-Speech Tagging (POS)**: This is like an English teacher analyzing a sentence. It involves labeling each word with its appropriate part of speech (like noun, verb, adjective, etc.). This helps to understand the grammatical structure of the sentence and can be used in many applications like text-to-speech, grammar checking, and more.

In the context of social media posts analysis:

- **Topic Modelling** can help us understand what are the main topics people are talking about.
- **NER** can tell us which specific entities (like people, places, or organizations) are being mentioned.
- **POS Tagging** can help us understand the context better by knowing the grammatical role of each word.

By combining these techniques, we can get a rich and structured understanding of unstructured social media data.
Using LDA to get the 8 main topics in our dataset

Topic 0: itcanbe | future | education | let | help | trust | set | today | children | time | child | answer | better | email | sitlagencystanbicibtccom
Topic 1: wealthwednesday | life | people | love | like | insurance | share | gift | tips | smwlagos | itcanbe | know | season | stand | care
Topic 2: email | itcanbe | send | account | details | started | invest | visit | investment | savings | make | money | information | start | mutual
Topic 3: stanbic | ibtc | itcanbe | app | visit | pension | started | download | click | account | mobile | banking | plan | super | today
Topic 4: join | stanbic | ibtc | day | itcanbe | win | register | link | loved | ones | nan | click | apply | miss | tomorrow
Topic 5: business | itcanbe | live | growth | nigeria | visit | watch | happy | click | thank | africa | world | session | best | movingforward
Topic 6: itcanbe | year | insurance | best | stay | week | new | hiflstanbicibtc | series | don | motivationmonday | powerupmonday | let | click | health
Topic 7: sdg | itcanbe | sustainable | know | throwbackthursday | covered | focused | sustainabilitysaturday | forget | assets | sustainability | raise | emoji | spend | game

We can create 8 unique categories from this topics

Topic 0 - Future Planning: This topic seems to be about planning for the future, with keywords like "future", "education", "children", and "time".

Topic 1 - Personal Finance: This topic appears to be about personal finance and insurance, with keywords like "wealthwednesday", "insurance", "share", and "gift".

Topic 2 - Investment: This topic seems to be about investment and savings, with keywords like "email", "invest", "account", and "savings".

Topic 3 - Banking Services: This topic appears to be about banking services, with keywords like "stanbic", "ibtc", "app", and "pension".

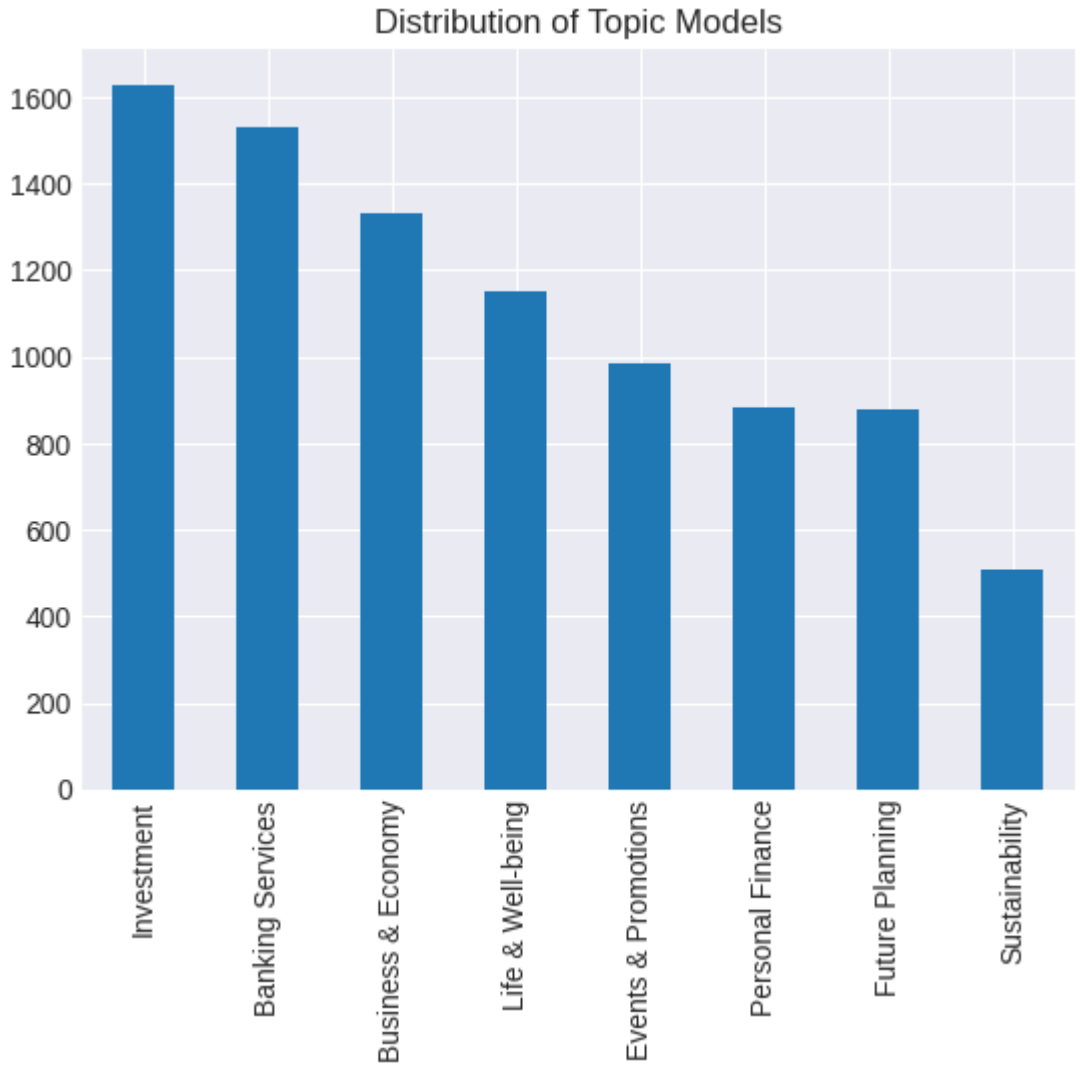
Topic 4 - Events & Promotions: This topic seems to be about events and promotions, with keywords like "join", "win", and "register".

Topic 5 - Business & Economy: This topic appears to be about business and the economy, with keywords like "business", "growth", and "nigeria".

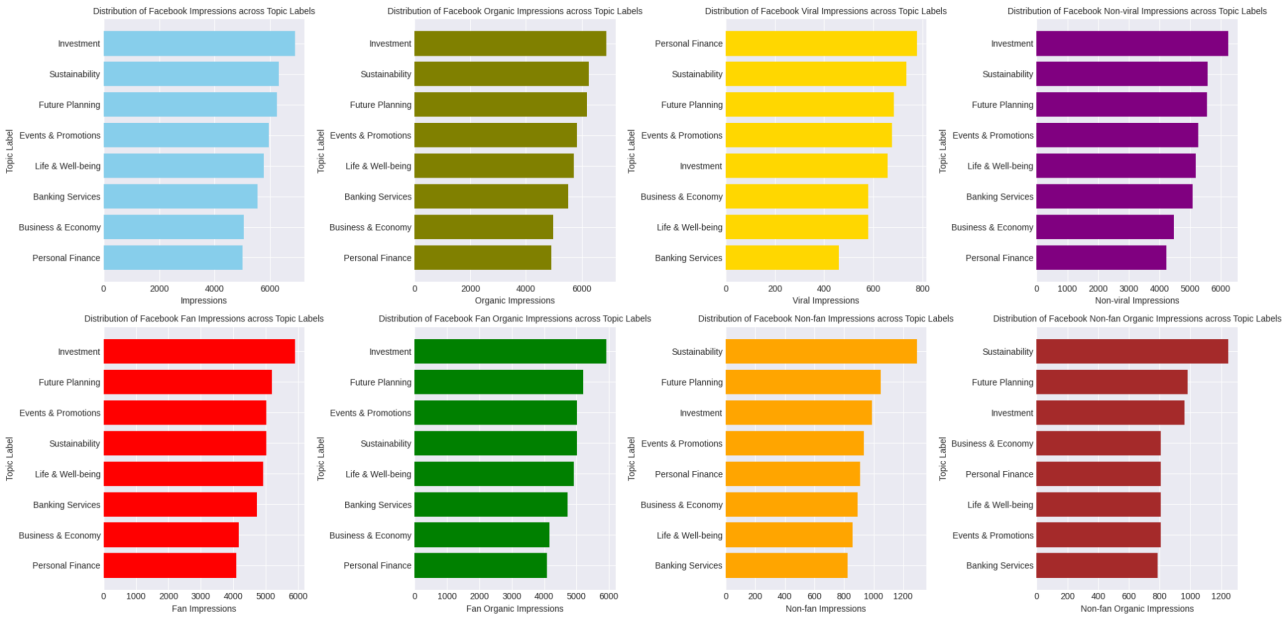
Topic 6 - Life & Well-being: This topic seems to be about life and well-being, with keywords like "year", "insurance", and "week".

Topic 7 - Sustainability: This topic appears to be about sustainability, with keywords like "sdg" (Sustainable Development Goals), "sustainable", and "sustainabilitysaturday".

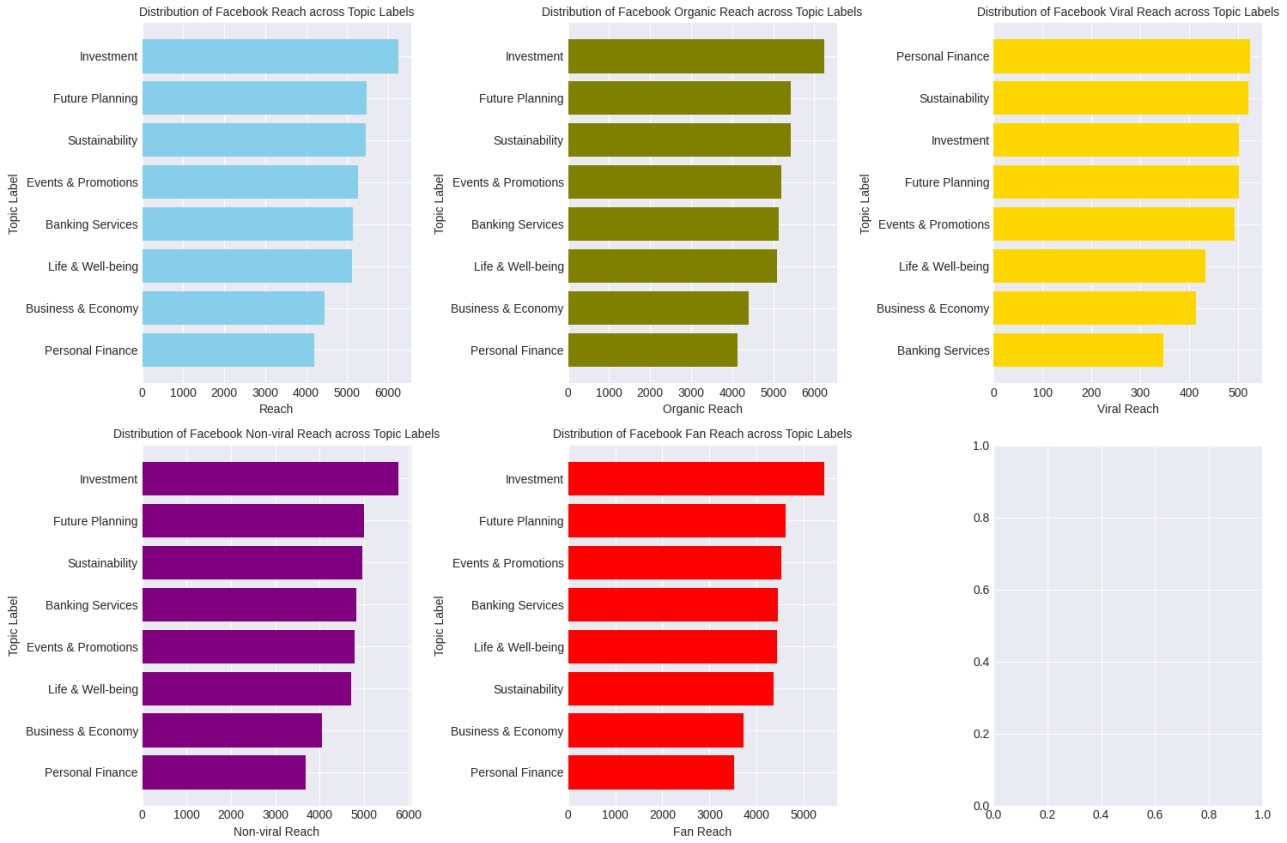
Below we can also see a distribution of this Topic Models in our Dataset based on our Facebook Posts, We see a majority of Posts relating to the topic Investment, and then Banking Services that are offered, and also how to grow Business and get the best of the Economy.



We track the Average Number of Impressions across Our Topics, and we observe the following: Posts about Investments and Sustainability give us a higher number of Impressions, Organic Impressions, Non-viral Impressions. While Posts about Investment and Future Planning give us a higher number of Fan Impressions and Fan Organic Impressions, compared to other topics. Posts about Personal Finance, and Sustainability gives us more Viral Impressions, while Posts about Sustainability and Future Planning give us higher Non-Fan Impressions.



Posts on Investment and Future Planning are the top 2 topics with the highest Facebook Reach, Organic Reach, Non Viral Reach, and Fan Reach, Personal Finance and Sustainability have the most Viral Reach .



Based on the insights from your data, here are some policies you could consider to increase your Facebook Impressions, Reach, and Fan Engagement:

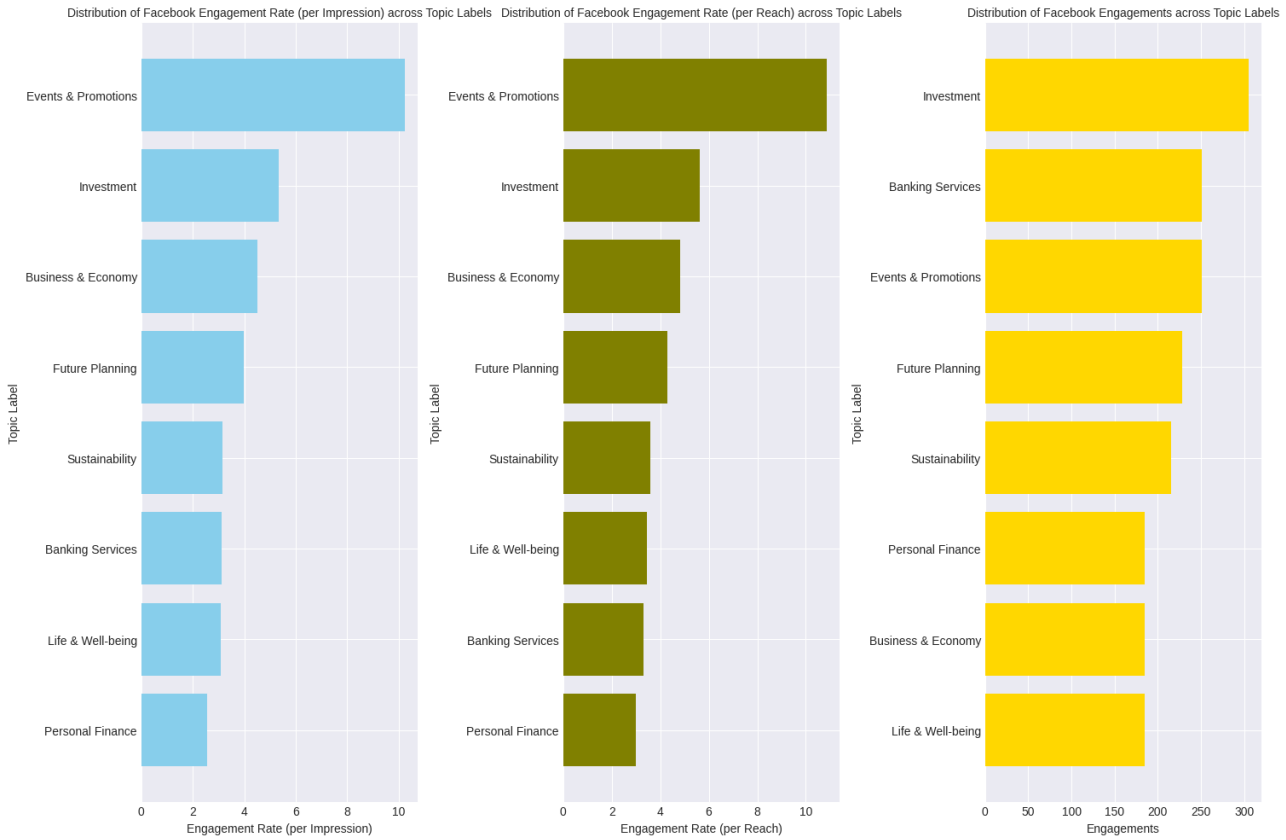
Focus on High-Impact Topics: Increase the frequency of posts about Investments and Sustainability, as these topics seem to generate a higher number of Impressions, Organic Impressions, and Non-viral Impressions.

Engage Your Fans: Posts about Investment and Future Planning seem to resonate with your fans, generating higher Fan Impressions and Fan Organic Impressions. Consider creating more content around these topics to engage your existing audience.

Encourage Virality: Personal Finance and Sustainability posts have more Viral Impressions. Try to create content that encourages sharing, such as infographics or interactive posts.

Reach New Audiences: Sustainability and Future Planning posts have higher Non-Fan Impressions. These topics could be a good way to reach new audiences. Consider using targeted advertising to boost these posts among non-fans.

Across Engagements, Engagement Rate (per Impression)
We have the following order
Events and Promotion → Investment → Business and Economy,
Engagement Rate (per Reach)
We have the following order
Events and Promotion → Investment → Business and Economy,
Number of Engagements
Investment → Banking Services → Events and Promotion



This leads us to ask the question?

Sustainability and Future Planning give us higher Non-Fan Impressions, and Posts about Investments and Sustainability give us higher number of Impressions, Organic Impressions, Non-viral Impressions, But they don't convert to Engagements? Why is that? why do they only attract Impressions and Reach but not Engagements?

A simple Answer:
Content Relevance: While these topics might attract views, the content are not engaging enough for the audience to interact with. The content might not resonate with the audience's interests or needs.

Call to Action: The posts might lack a clear call-to-action that encourages users to engage with the content. This could be in the form of a question, a prompt to share opinions, or an invitation to share the post.

Post Timing: The timing of the posts might not align with when the audience is most active on Facebook. If posts are published when users are less likely to be online, they might get overlooked.

Content Type: Different types of content (text, image, video, link) can have different levels of engagement. It's possible that the content type used for these topics is less engaging.

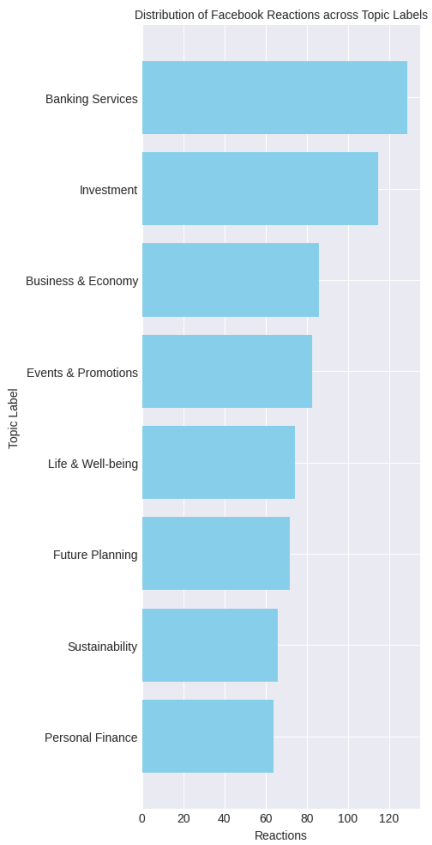
Some Policies which can be formulated based off this data is:

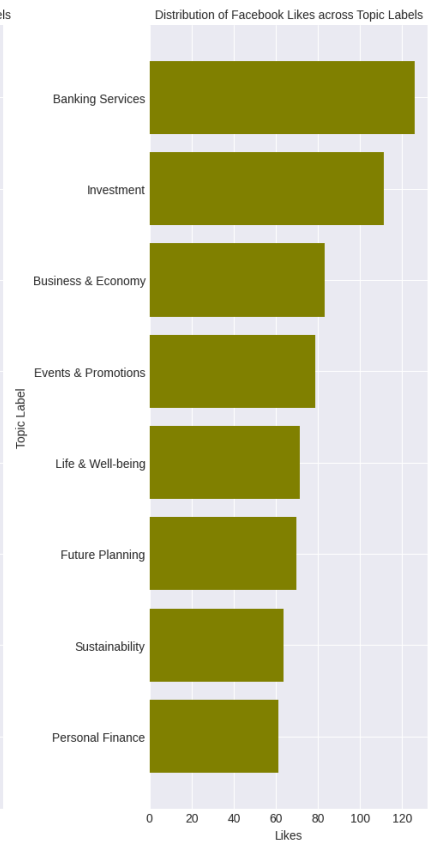
Focus on High-Impact Topics: Increase the frequency of posts about Investment and Banking Services, as these topics seem to generate a higher number of engagements.

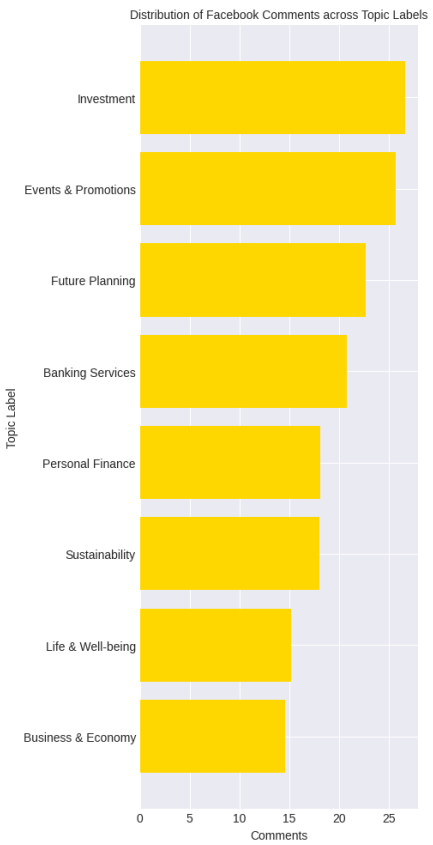
Encourage Virality: Posts about Personal Finance and Sustainability have more Viral Reach. Try to create content that encourages sharing, such as infographics or interactive posts.

Engage Your Fans: Posts about Investment and Future Planning seem to resonate with your fans, generating higher Fan Impressions and Fan Organic Impressions. Consider creating more content around these topics to engage your existing audience.

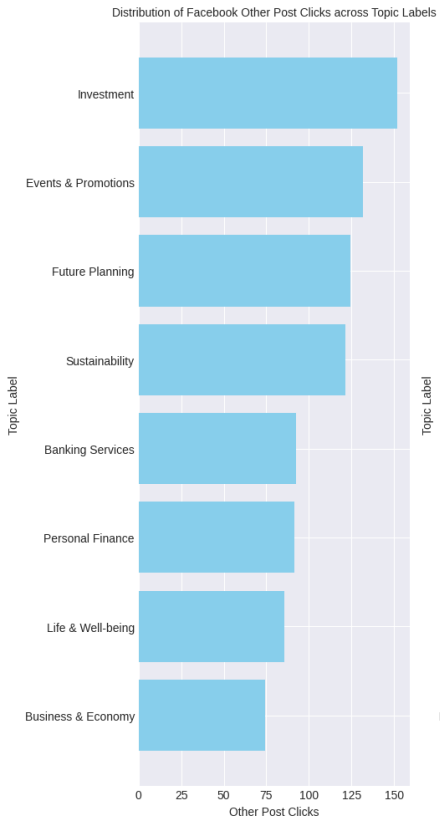
Promote Events and Promotions: These types of posts have the highest Engagement Rate per Impression and per Reach. Consider creating more of these types of posts to increase engagement.

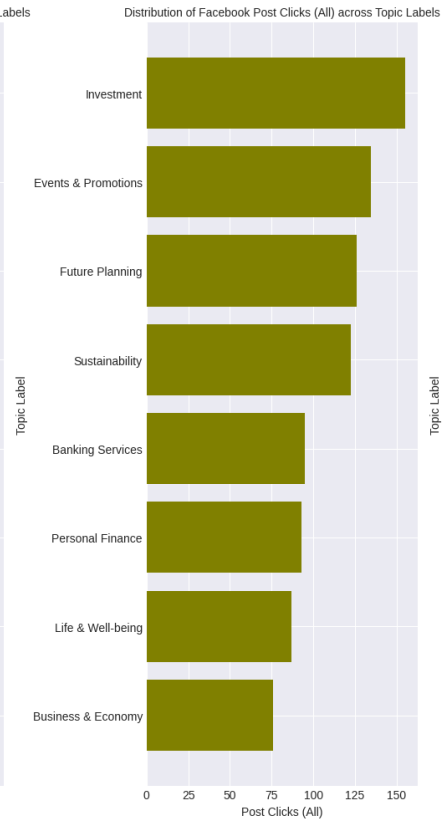


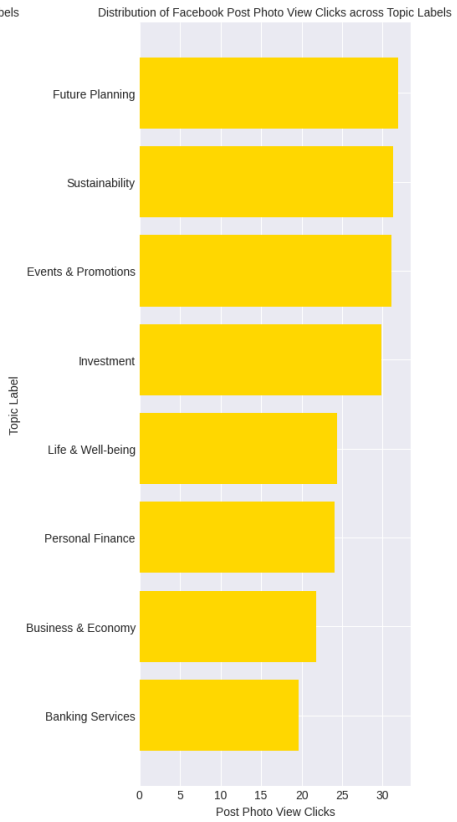




Just as we noticed with Engagements, the topics that seem to have the higher number of Reactions and likes are Banking Services, Investment, Business and Economy,
The Topics that generate more comments are Investments, Events and Promotions and Future Planning.

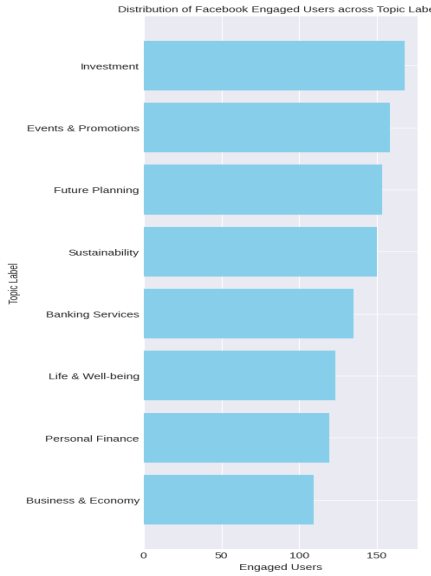


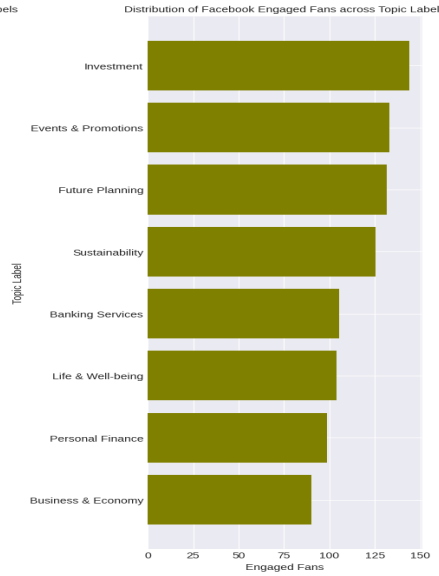


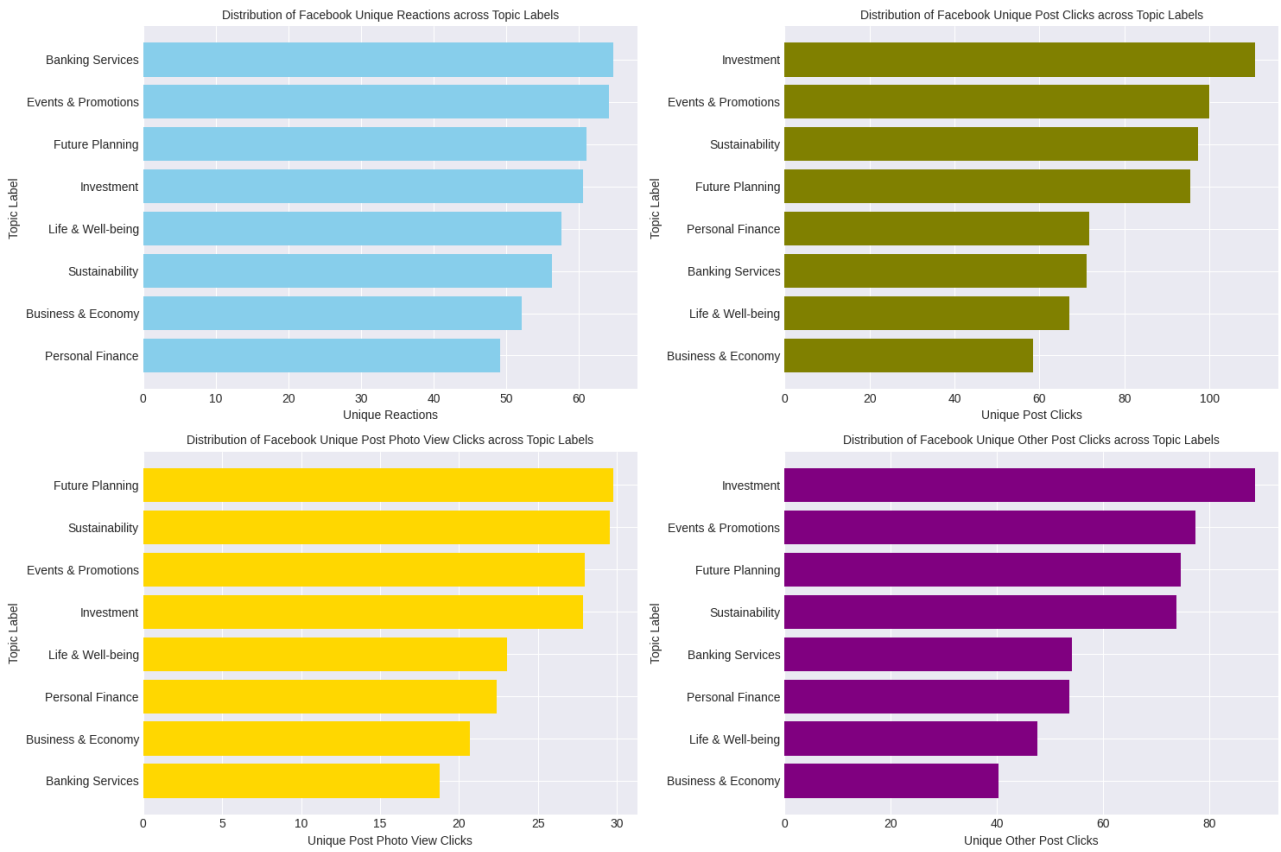


The Topics with the highest Post Clicks are Investment, Events and Promotion, and Future and Planning, Topics with the Highest Photo View Clicks are Future Planning, Sustainability and Events & Promotions, Interestingly as proven before in our EDA, we can also see that the topics with the highest comments, are also the topics with the highest Post Clicks, so users who comment are most likely to click on posts.

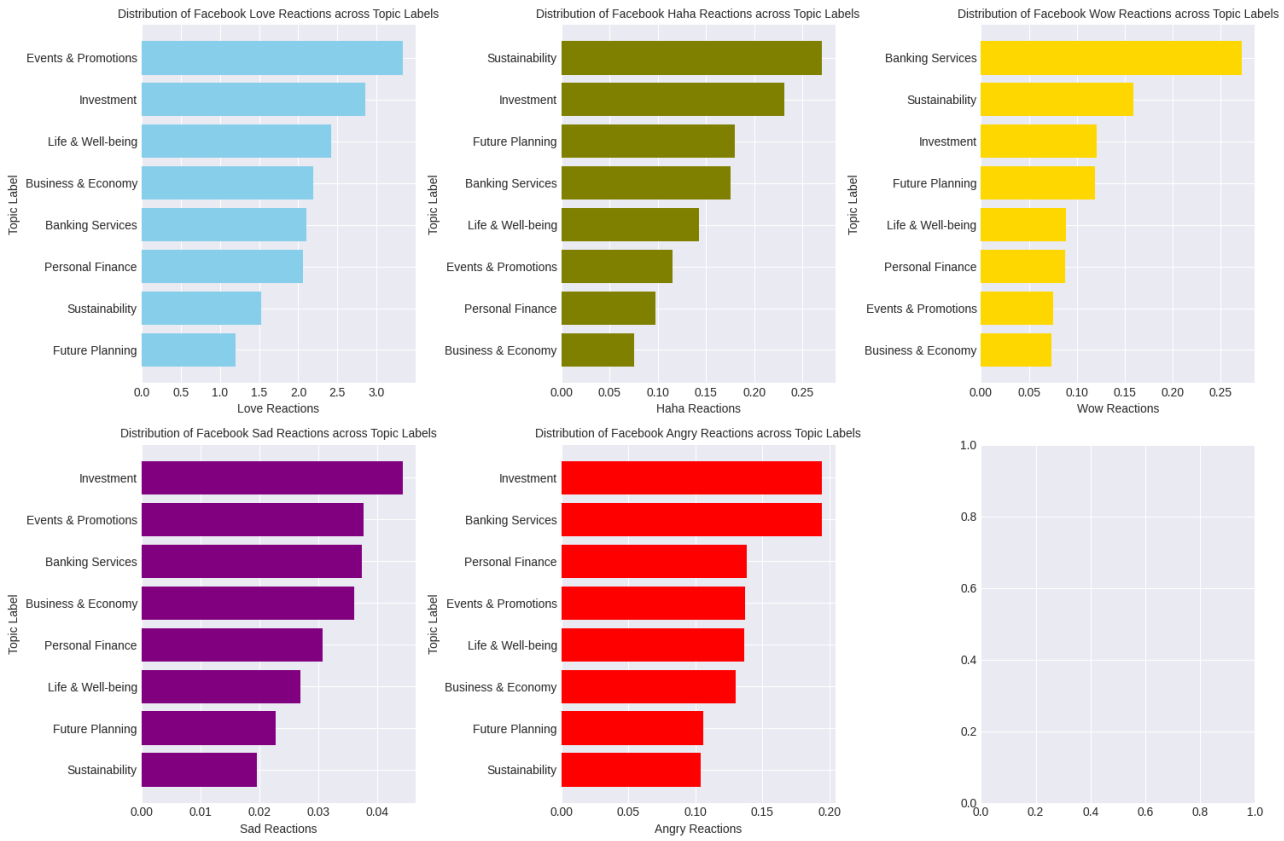
As can be easily seen and observe, Engaged Fans and Engaged Users share the exact same topics with those topics that have the highest Post Clicks and also have the highest comments, which are Investment, Events and Promotion, and Future and Planning, Facebook Users talking about the post, still revolve around our Most Engaging topics, which are Investment , Banking Services, and Events and Promotions.

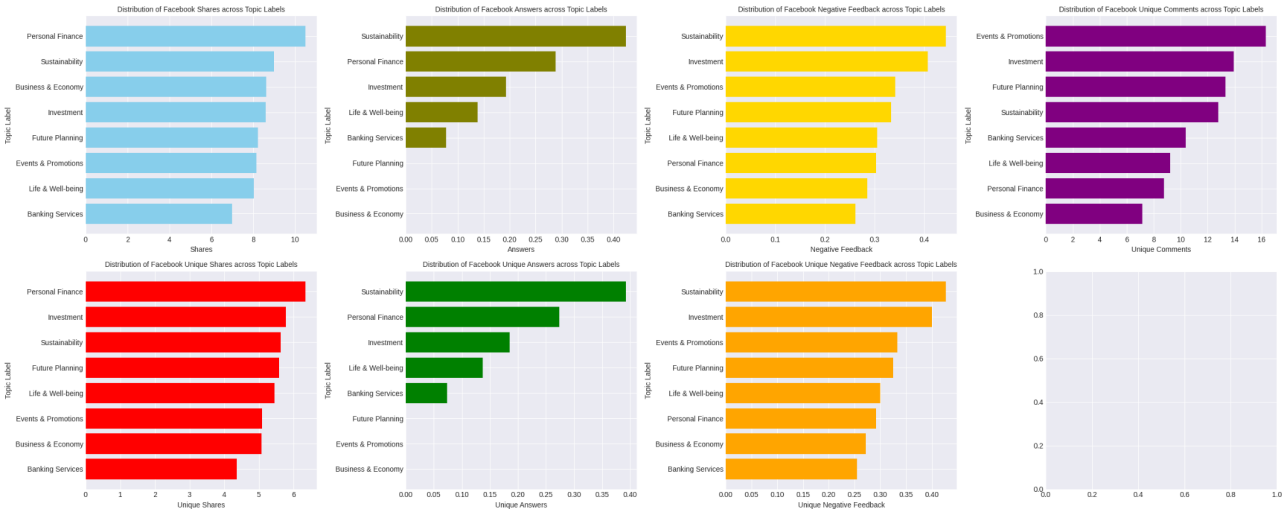






This confirms our Correlation Heatmap
Reactions To each Topic Label





We can see some few uprising patterns

Topic Label with the Highest Values across each Post's Shares are Personal Finance, Sustainability, Business & Economy
Topic Label with the Highest Values across each Post's Answers are Sustainability, Personal Finance, Investment
Topic Label with the Highest Values across each Post's Negative Feedback are Sustainability, Investment, Events & Promotions
Topic Label with the Highest Values across each Post's Unique Comments are Events & Promotions, Investment, Future Planning
Topic Label with the Highest Values across each Post's Unique Shares are Personal Finance, Investment, Sustainability
Topic Label with the Highest Values across each Post's Unique Answers are Sustainability, Personal Finance, Investment
Topic Label with the Highest Values across each Post's Unique Negative Feedback are Sustainability, Investment, Events & Promotions.

Actionable Recommendations

Based on the information above, here are some insights about each topic label:

1. **Investment**: This topic seems to generate a high number of impressions, reach, and engagements. It also leads in fan impressions and fan organic impressions. Investment posts have the highest number of engagements and are among the top topics for comments and post clicks. However, they also receive some negative feedback.
2. **Sustainability**: While these posts get a high number of impressions, they don't convert well to engagements. They do, however, generate a significant number of viral impressions and non-fan impressions. Posts on this topic also receive the highest negative feedback.
3. **Future Planning**: These posts have high reach and impressions, particularly among fans. They also generate a significant number of non-fan impressions. Future Planning is among the top topics for post clicks and photo view clicks.

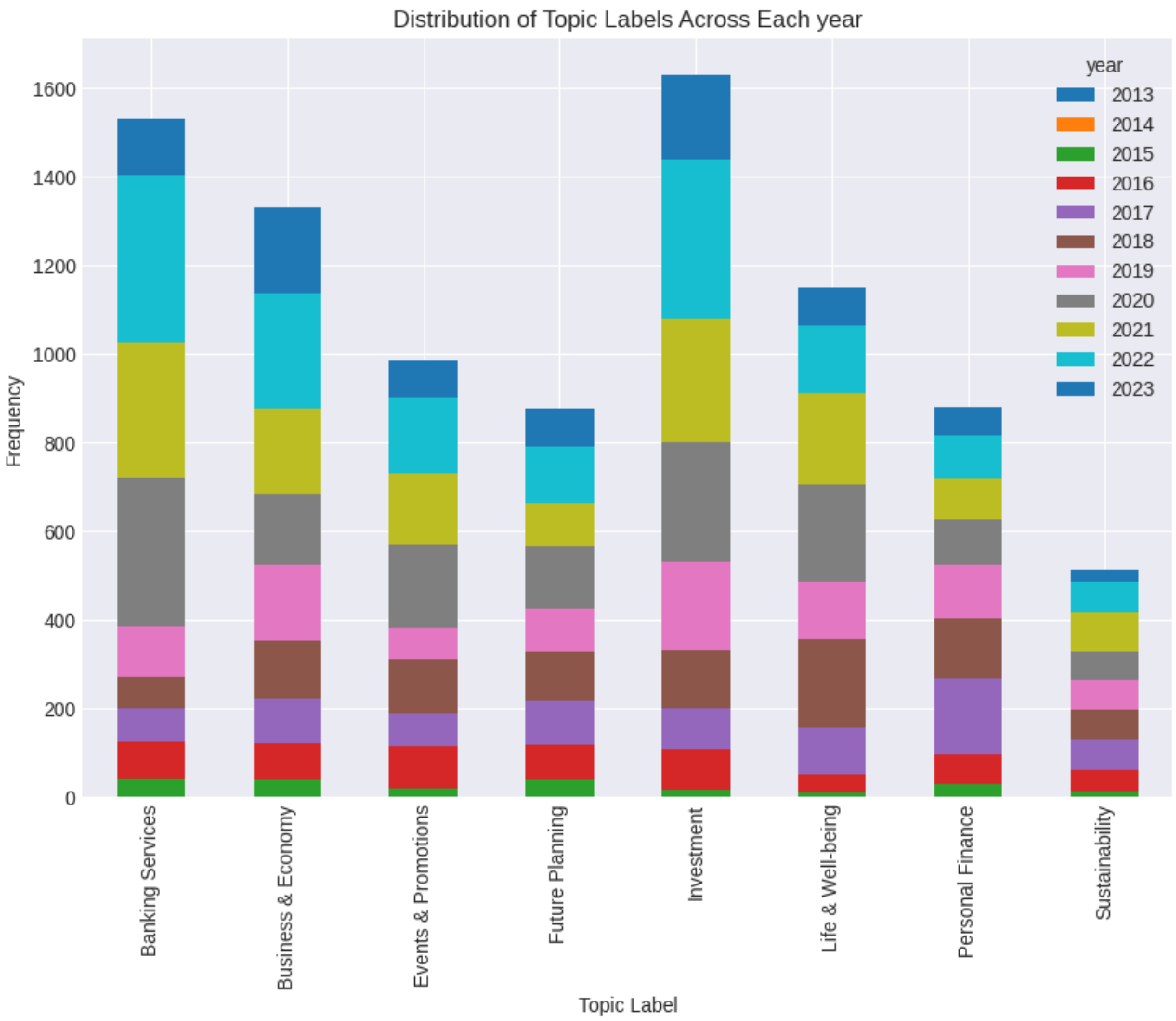
4. **Personal Finance**: Posts on this topic generate the most viral reach. They also lead in shares and unique shares, indicating that users find this content shareable.
5. **Events & Promotions**: These posts have the highest engagement rate per impression and per reach. They are among the top topics for comments, post clicks, and unique comments.
6. **Banking Services**: Posts on this topic generate a significant number of engagements, second only to Investment.
7. **Business & Economy**: This topic is among the top three for engagement rate per impression and per reach.

From these insights, it's clear that different topics resonate differently with your audience. Investment, Events & Promotions, and Banking Services seem to be the most engaging topics, while Sustainability posts get a lot of visibility but less engagement. Personal Finance posts are highly shareable, indicating that users find this content valuable enough to share with their networks.

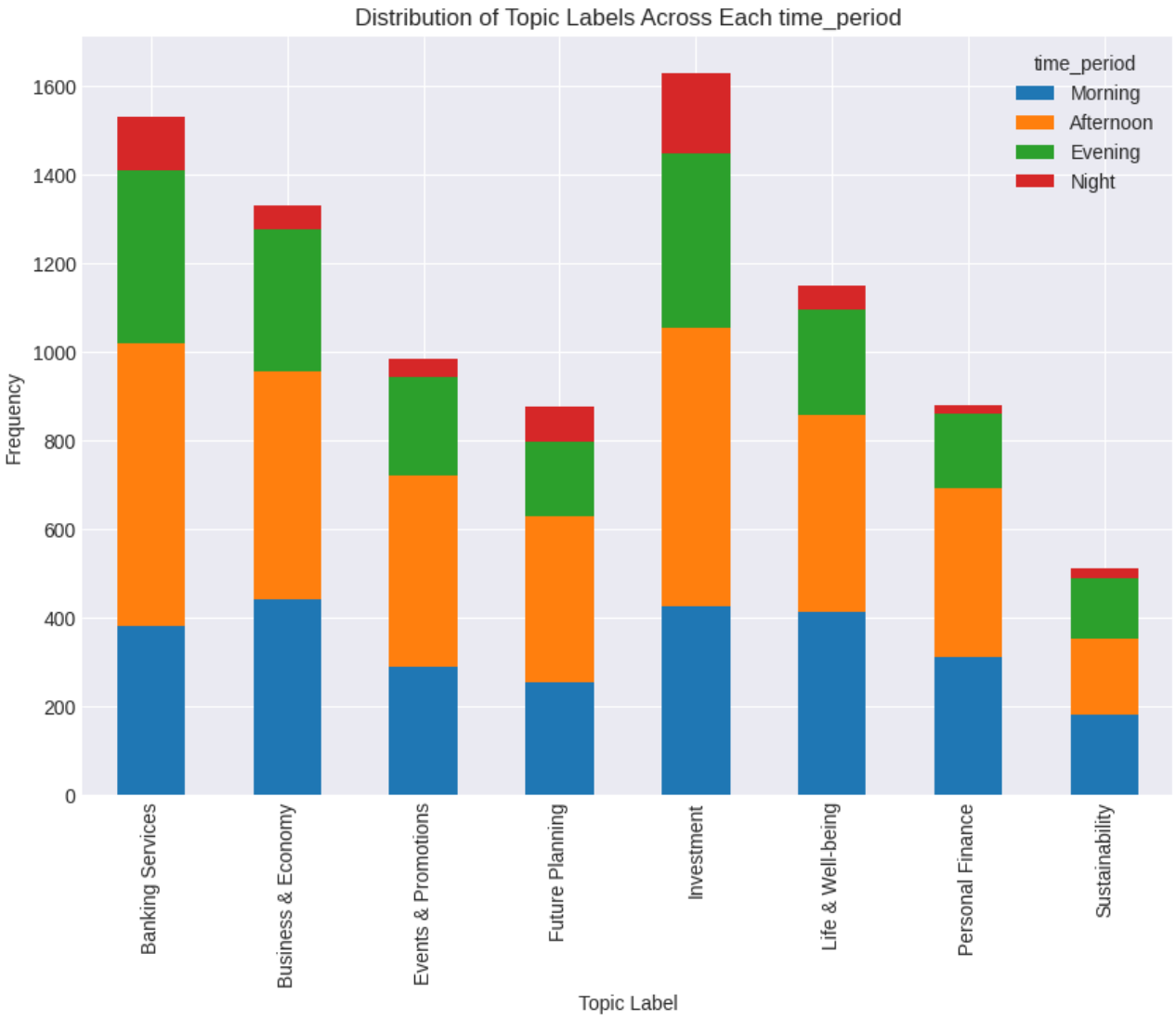
Based on the insights from this data, here are some policies a social media agency could consider to optimize their Facebook posts:

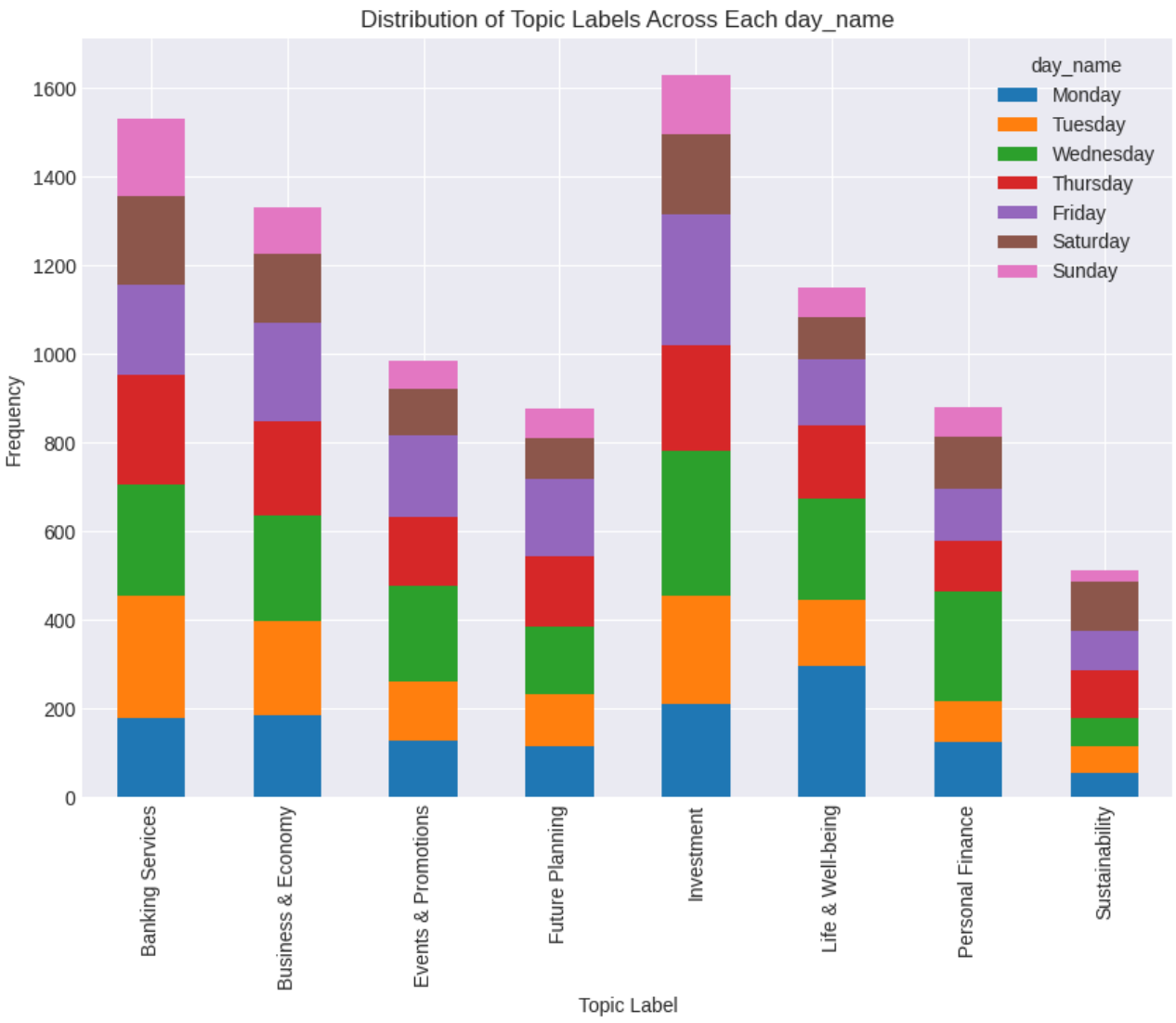
1. **Focus on High-Impact Topics**: Increase the frequency of posts about Investment, Events & Promotions, and Banking Services as these topics seem to generate a higher number of engagements.
2. **Engage Your Fans**: Posts about Investment and Future Planning seem to resonate with your fans, generating higher Fan Impressions and Fan Organic Impressions. Consider creating more content around these topics to engage your existing audience.
3. **Encourage Virality**: Personal Finance and Sustainability posts have more Viral Reach. Try to create content that encourages sharing, such as infographics or interactive posts.
4. **Reach New Audiences**: Sustainability and Future Planning posts have higher Non-Fan Impressions. These topics could be a good way to reach new audiences. Consider using targeted advertising to boost these posts among non-fans.
5. **Promote Engagement**: Encourage users to like, comment on, and share your posts. This could be through contests, giveaways, or simply by asking your followers to engage.
6. **Manage Negative Feedback**: Posts about Sustainability and Investment receive the highest negative feedback. It's important to monitor these comments and address any issues or concerns raised by your audience.

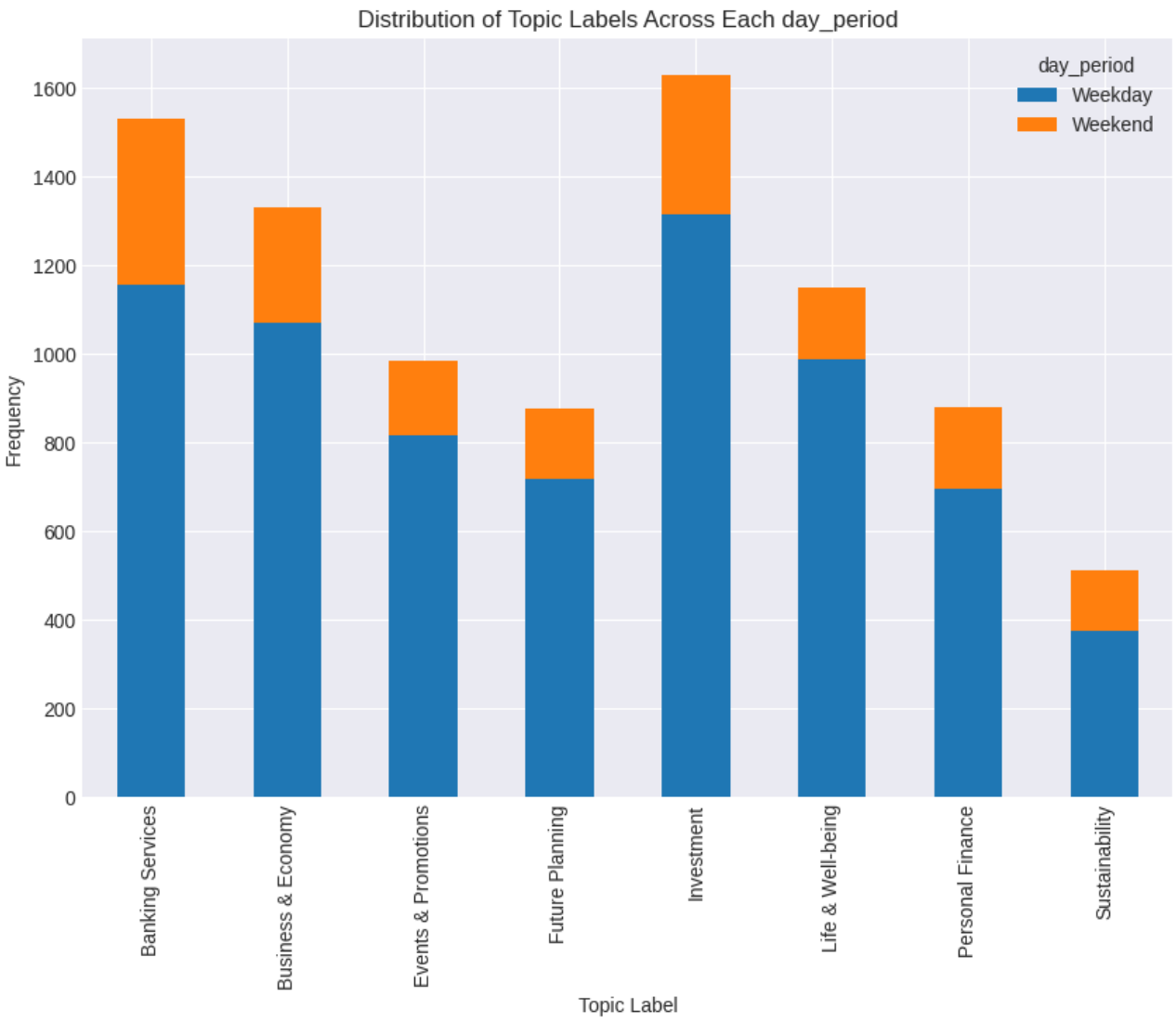
Below, we can see, the distribution of Topic labels across the years we can see Investment being well represented followed by Banking Services and Business and Economy,

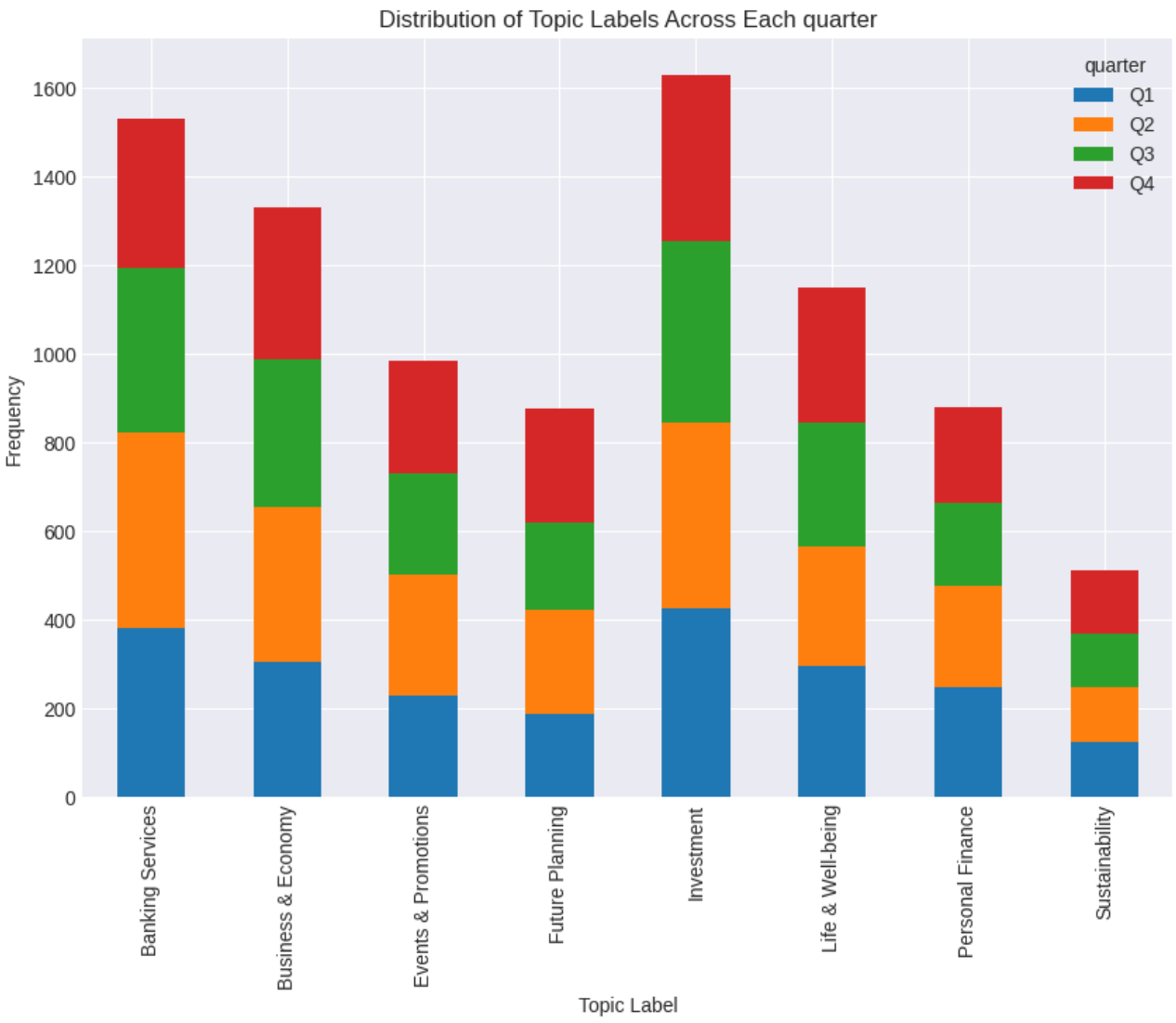


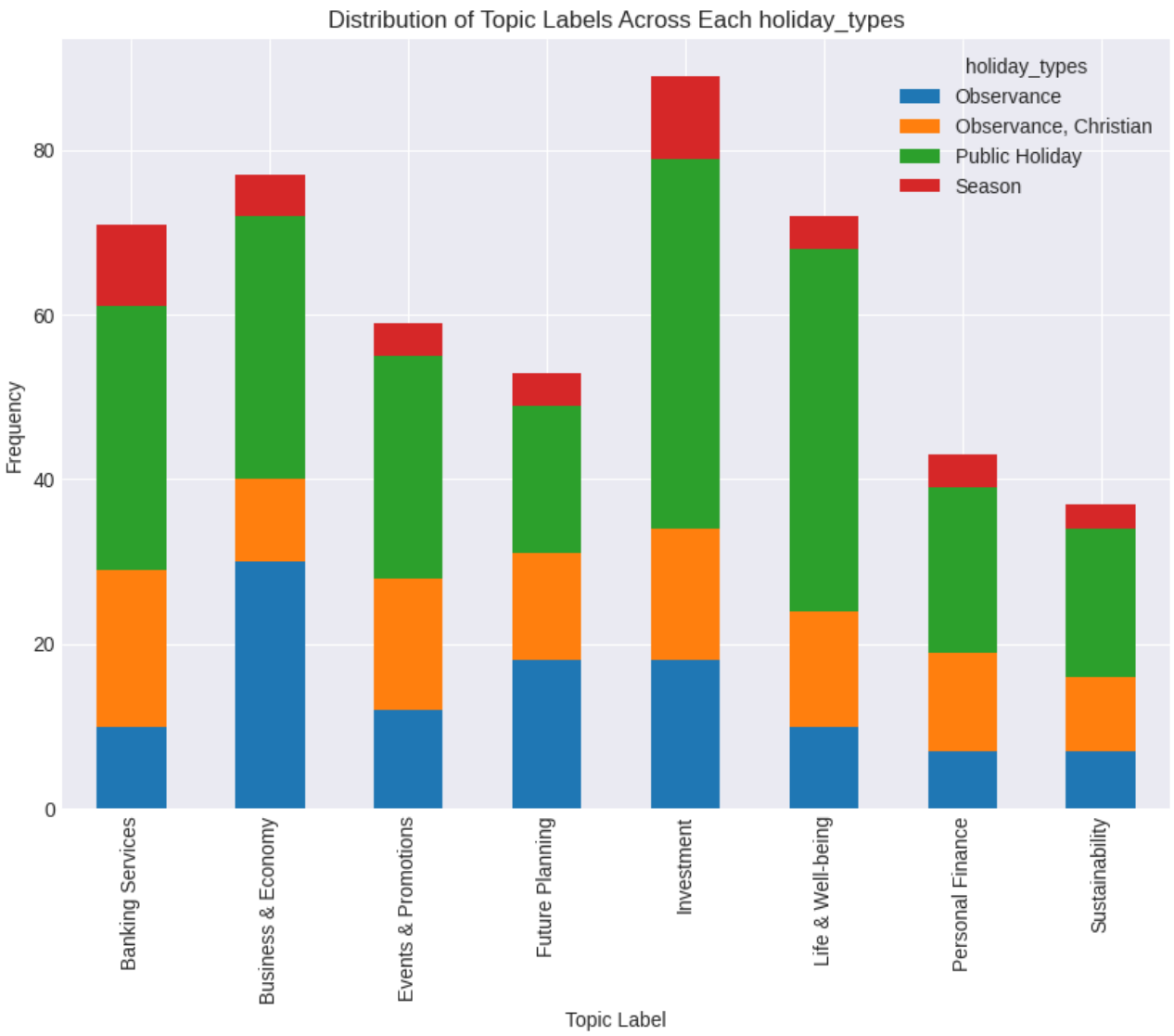
Below, we can also see the same across the Day Periods, Investment Posts are well represented against all Day periods, followed by Banking Services, Business and Economy and then Life and well being


Distribution of Topic Labels Across Each time_period
time_period
Morning
Afternoon
Evening
Night
Frequency
0
200
400
600
800
1000
1200
1400
1600
Banking Services
Business & Economy
Events & Promotions
Future Planning
Investment
Life & Well-being
Personal Finance
Sustainability
Topic Label


Distribution of Topic Labels Across Each day_name
day_name
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
Frequency
0
200
400
600
800
1000
1200
1400
1600
Banking Services
Business & Economy
Events & Promotions
Future Planning
Investment
Life & Well-being
Personal Finance
Sustainability
Topic Label

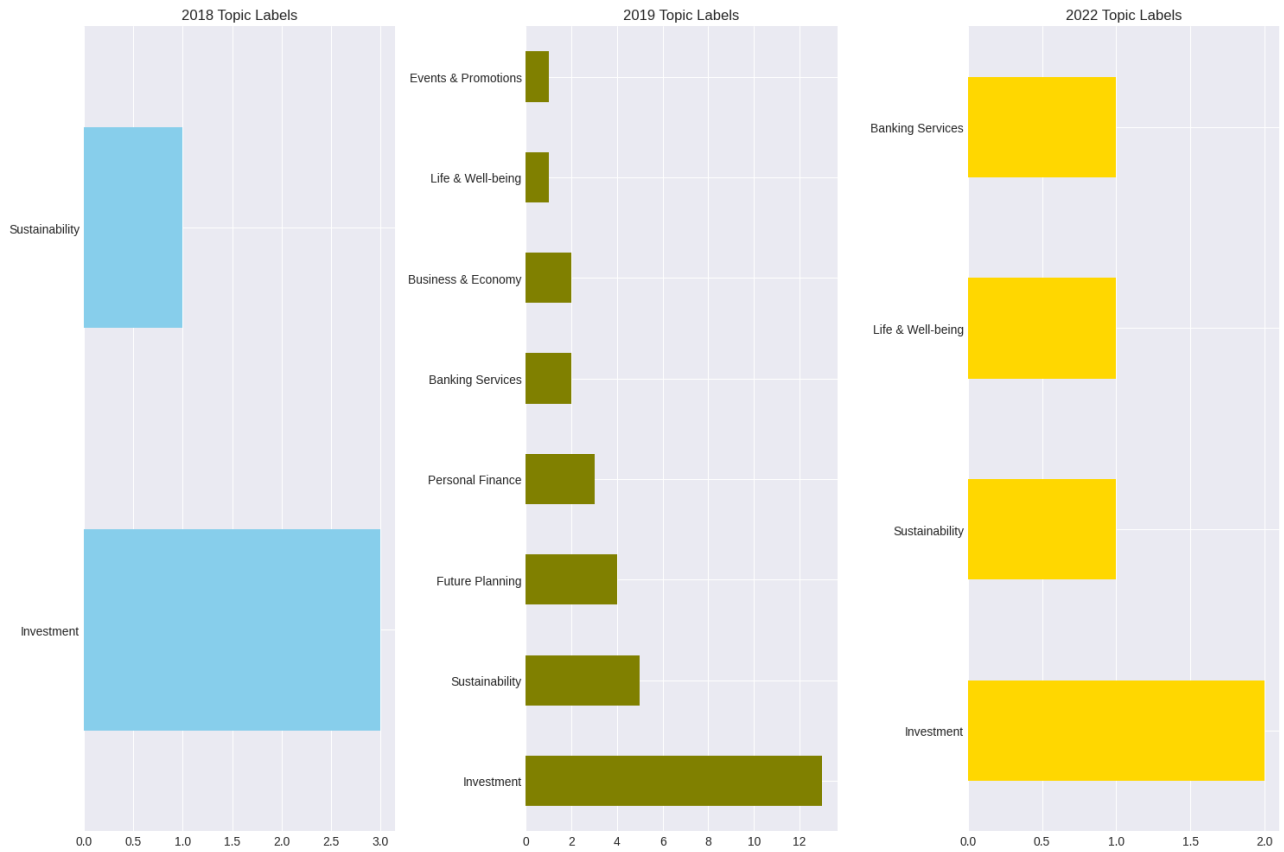

Distribution of Topic Labels Across Each day_period
day_period
Weekday
Weekend
Frequency
0
200
400
600
800
1000
1200
1400
1600
Banking Services
Business & Economy
Events & Promotions
Future Planning
Investment
Life & Well-being
Personal Finance
Sustainability
Topic Label


Distribution of Topic Labels Across Each quarter
quarter
Q1
Q2
Q3
Q4
Frequency
0
200
400
600
800
1000
1200
1400
1600
Banking Services
Business & Economy
Events & Promotions
Future Planning
Investment
Life & Well-being
Personal Finance
Sustainability
Topic Label



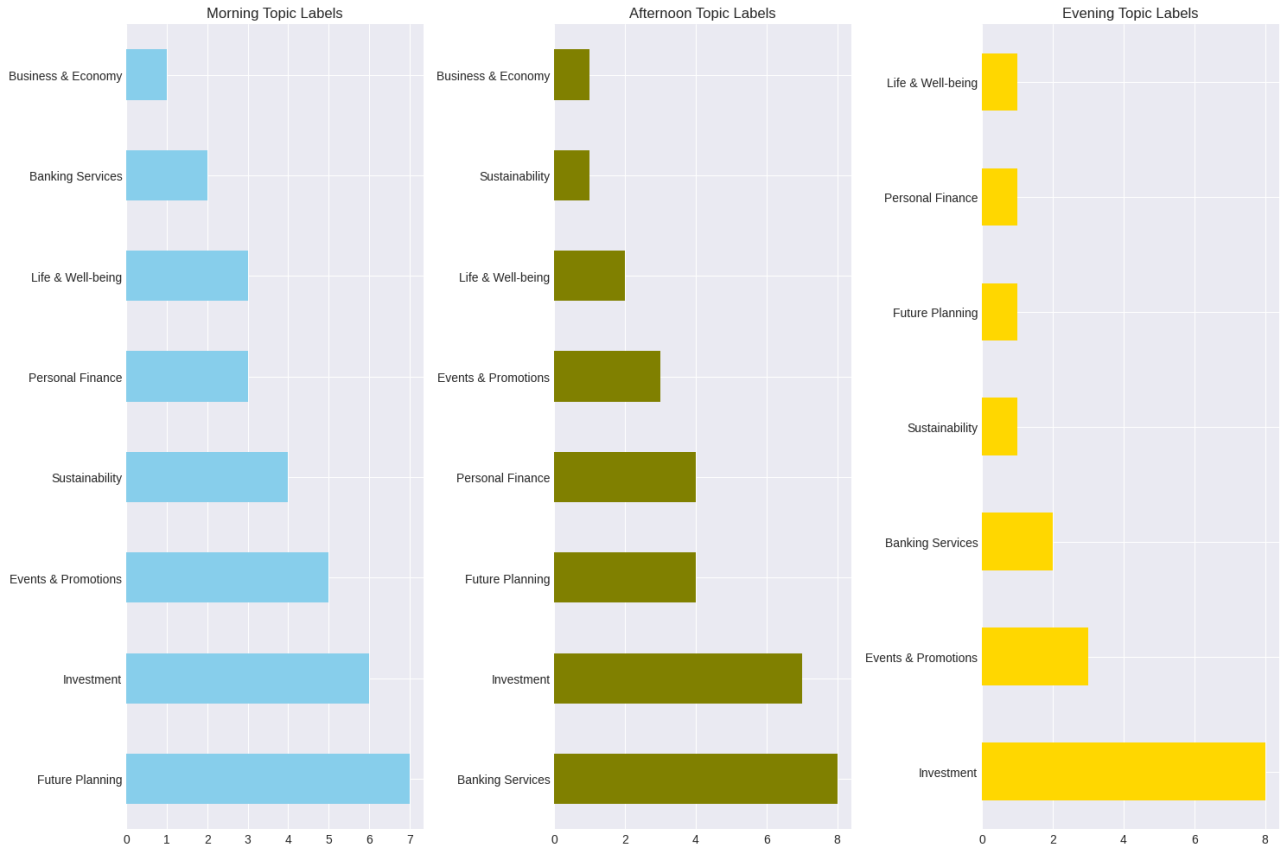
Distribution of Posts in the Top 1% Impression

Our Insights are finally coming together as we can see and have a clearer explanation now, Posts that were in the top 1% Impression in the year 2019, had the highest Verb count, Adverb count, Adjectives count, Noun Count. Using more parts of Speech in our posts helps us to generate more insights, we can also see compared to the others 2019 had the highest number of topics being discussed. When our posts are distributed it helps to generate more insights, the year 2019 also had the most Named Entities Recognized, so we had Important Dates, Organizations mentioned in our Posts which gathered more Impressions

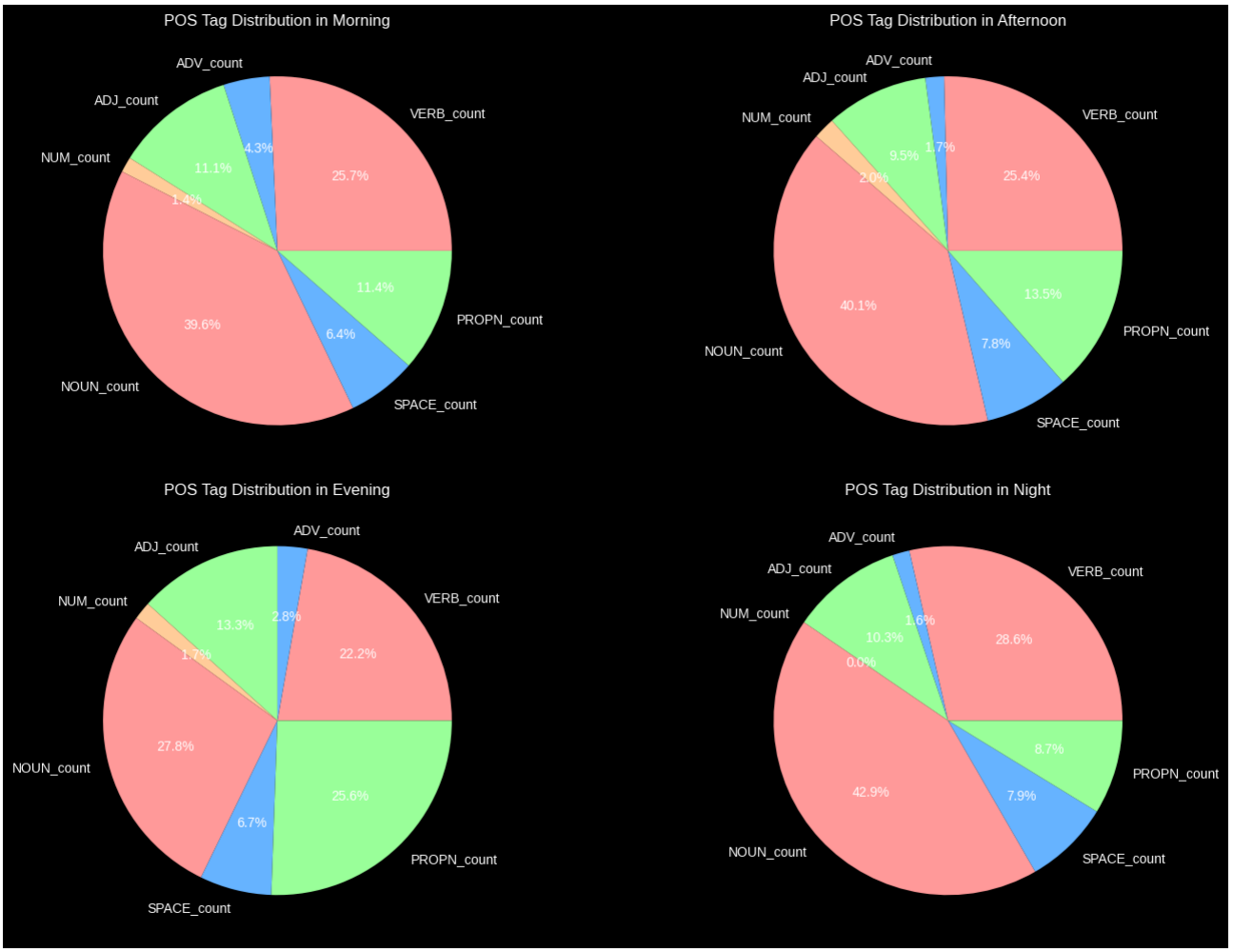


As we can see here the year that had the Top 1% Impressions which is 2019, had more topic labels that the lowest years, 2018 and 2022

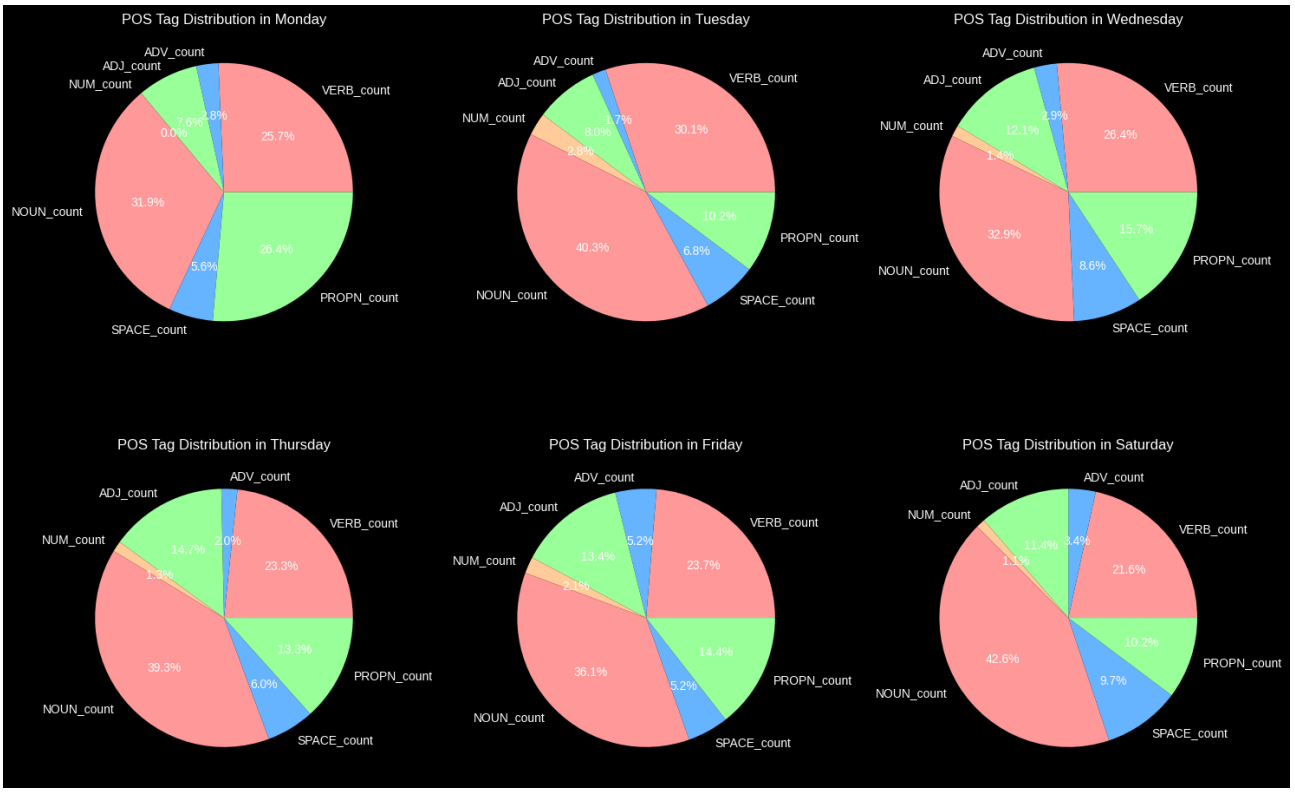
We also notice the same for Named Entity Recognition and Parts of Speech Tags



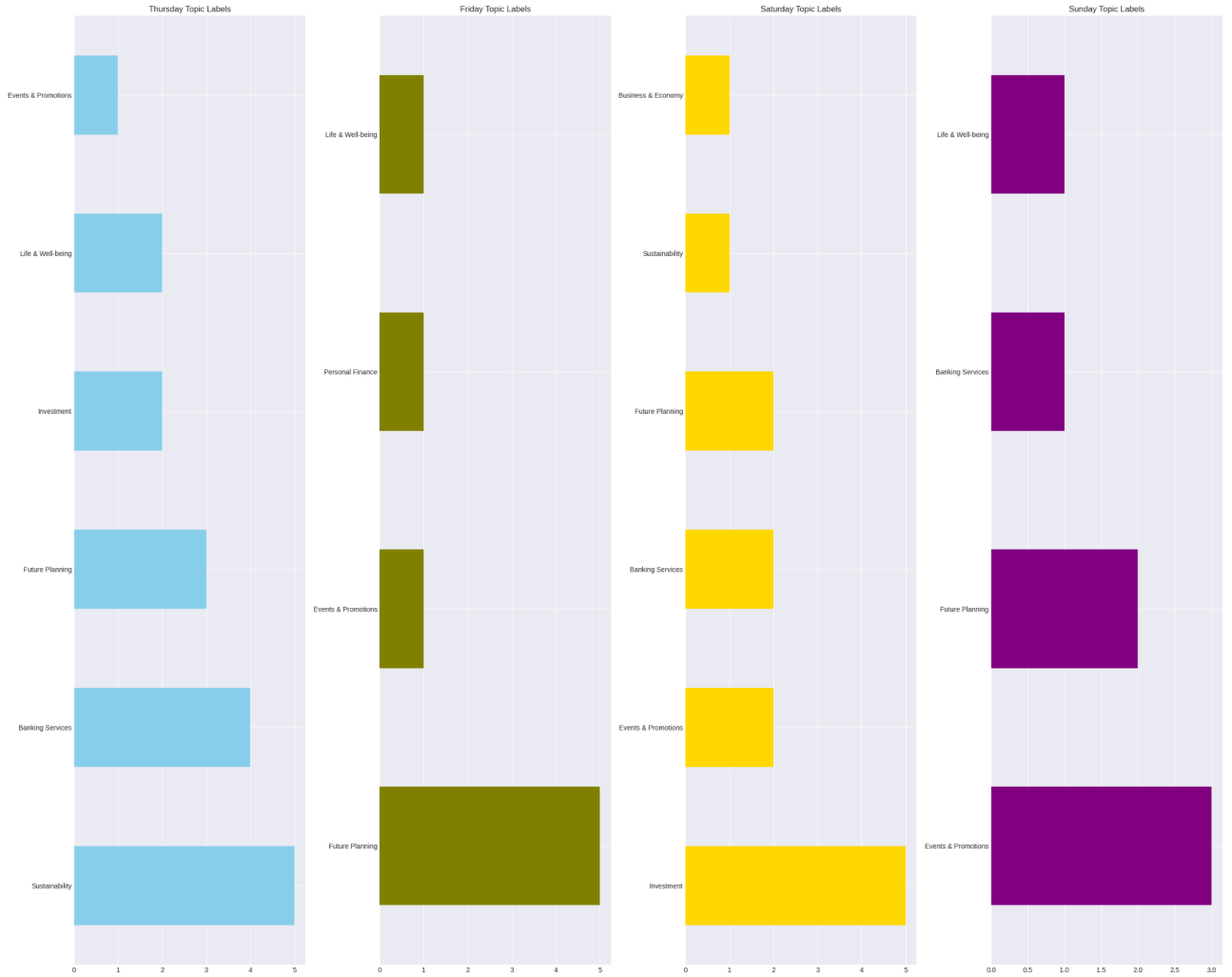
This also confirms our theory showing to us that the Morning and Afternoon Time Periods, Posts that were in the top 1% Impression during the Morning and Afternoon Time Periods, had the highest Verb count, Adverb count, Adjectives count, Noun Count. Using more parts of Speech in our posts helps us to generate more Impressions, we can also see compared to the others Morning and Afternoon had the highest number of topics being discussed. When our posts are distributed it helps to generate more Impressions, so we had Important Dates, Organizations mentioned in our Posts which gathered more Impressions.



The Same theory can be applied to Top Posts with the Highest Impressions, we notice



The highest days with the highest part of speech tags are Tuesday, Thursday, Saturday,



Conclusion

On Facebook we have an average Impression of 5857, having our max Impressions at 207830.0. Facebook offers us more Impressions and Reach and Engagements compared to other social medias, Facebooks offers more Impressions, followed by Twitter and then Instagram and then LinkedIn